



Zhiyuan Yao

Visual Customer Segmentation and Behavior Analysis

A SOM-Based Approach

TURKU CENTRE *for* COMPUTER SCIENCE

TUUCS Dissertations
No 163, October 2013

Visual Customer Segmentation and Behavior Analysis

A SOM-Based Approach

Zhiyuan Yao

To be presented, with the permission of the Department of
Information Technologies at Åbo Akademi University, for public
criticism in the Auditorium Gamma on October 18th, 2013 at noon.

Åbo Akademi University
Department of Information Technologies
Joukahaisenkatu 3-5, FIN-20520 Turku, Finland

2013

Supervised by

Professor Barbro Back
Department of Information Technologies
Åbo Akademi University
Turku, Finland

Docent Tomas Eklund
Department of Information Technologies
Åbo Akademi University
Turku, Finland

Reviewed by

Professor Ranjan Kini
School of Business & Economics
Indiana University Northwest
Gary, Indiana, United States

Professor Jussi Kantola
Department of Production
University of Vaasa
Vaasa, Finland

Opponent

Professor Ranjan Kini
School of Business & Economics
Indiana University Northwest
Gary, Indiana, United States

ISBN 978-952-12-2950-3

ISSN 1239-1883

Abstract

The importance of customer relationship management (CRM), a management principle for transforming organizations from being product-oriented to customer centric, has attracted interest from both academia and industry. Today, customers' behaviors and activities can be easily recorded and stored through Enterprise Resource Planning (ERP) systems and data warehousing. Customers are continuously creating a "customer signature" that lends itself to analysis. It is not uncommon that customers differ in various aspects, and have contrasting preferences and buying behavior. A widely used approach for gaining insight into the heterogeneity of customer behavior and profitability is customer segmentation, i.e., the division of customers into groups based on similarity.

Conventional customer segmentation solutions are often stand-alone analytical models, derived based on a specific time frame, which thereby often disregard the fact that customers' behavior may evolve over time. In order to provide a holistic view of customers' characteristics and purchasing behavior, two dynamics in a customer segmentation model will be examined in this thesis in addition to the conventional segmentation models based on customers' aggregate information of a specific time frame.

The two possible dynamics of a customer segmentation model include changes in segment structures and composition and changes in segment memberships of individual customers. The first dynamic is addressed using temporal customer segmentation where temporal changes in segment structures and profiles are visualized, while the second dynamic is addressed using segment migration analysis, where customers with similar switching patterns among segments are identified visually.

Visualization can assist in the interpretation of the discovered patterns and facilitate the process of knowledge transfer between analysts and decision makers. Visual data mining techniques, e.g., the Self-Organizing Map (SOM) and its extensions, will be used to demonstrate their usefulness in the context of CRM in general and in the task of customer segmentation in particular.

Abstrakt

Ett ämne som väckt intresse både inom industrin och forskningen är hantering av kundförhållanden (CRM, eng. Customer Relationship Management), dvs. en kundorienterad affärsstrategi där företagen från att ha varit produktorienterade väljer att bli mera kundcentrerade. Numera kan kundernas beteende och aktiviteter lätt registreras och sparas med hjälp av integrerade affärssystem (ERP, eng. Enterprise Resource Planning) och datalager (DW, eng. Data Warehousing). Kunder med olika preferenser och köpbeteende skapar sin egen ”signatur” i synnerhet via användningen av kundkort, vilket möjliggör mångsidig modellering av kundernas köpbeteende. För att få en översikt av kundernas köpbeteende och deras lönsamhet, används ofta kundsegmentering som en metod för att indela kunderna i grupper utgående från deras likheter.

De mest använda metoderna för kundsegmentering är analytiska modeller konstruerade för en viss tidsperiod. Dessa modeller beaktar inte att kundernas beteende kan förändras med tiden. I föreliggande avhandling skapas en holistisk översikt av kundernas karaktär och köpbeteende som utöver de konventionella segmenteringsmodellerna även beaktar dynamiken i köpbeteendet.

Dynamiken i en kundsegmenteringsmodell innefattar förändringar i segmentens struktur och innehåll, samt förändringen av individuella kunders tillhörighet i ett segment (s.k migrationsanalyser). Vardera förändringen modelleras, analyseras och exemplifieras med visuella datautvinningstekniker, främst med självorganiserande kartor (SOM, eng. Self-Organizing Maps) och självorganiserande tidskartor (SOTM), en vidareutveckling av SOM.

Visualiseringen anteciperas underlätta tolkningen av identifierade mönster och göra processen med kunskapsöverföring mellan den som gör analysen och beslutsfattaren smidigare.

Tiivistelmä

Asiakkuudenhallinta (CRM) eli organisaation muuttaminen tuotepainotteisesta asiakaskeskeiseksi on herättänyt mielenkiintoa niin yliopisto- kuin yritysmaailmassakin. Asiakkaiden käyttäytymistä ja toimintaa pystytään nykyään helposti tallentamaan ja varastoimaan toiminnanohjausjärjestelmien ja tietovarastojen avulla; asiakkaat jättävät jatkuvasti piirteistään ja ostokäyttäytymisestään kertovia tietojälkiä, joita voidaan analysoida. On tavallista, että asiakkaat poikkeavat toisistaan eri tavoin, ja heidän mieltymyksensä kuten myös ostokäyttäytymisensä saattavat olla hyvinkin erilaisia. Asiakaskäyttäytymisen monimuotoisuuteen ja tuottavuuteen paneuduttaessa käytetäänkin laajalti asiakassegmentointia eli asiakkaiden jakamista ryhmiin samankaltaisuuden perusteella.

Perinteiset asiakassegmentoinnin ratkaisut ovat usein yksittäisiä analyttisiä malleja, jotka on tehty tietyn aikajakson perusteella. Tämän vuoksi ne monesti jättävät huomioimatta sen, että asiakkaiden käyttäytyminen saattaa ajan kuluessa muuttua. Tässä väitöskirjassa pyritäänkin tarjoamaan holistinen kuva asiakkaiden ominaisuuksista ja ostokäyttäytymisestä tarkastelemalla kahta muutosvoimaa tiettyyn aikarajaukseen perustuvien perinteisten segmentointimallien lisäksi.

Nämä kaksi asiakassegmentointimallin dynamiikkaa ovat muutokset segmenttien rakenteessa ja muutokset yksittäisten asiakkaiden kuulumisessa ryhmään. Ensimmäistä dynamiikkaa lähestytään ajallisen asiakassegmentoinnin avulla, jossa visualisoidaan ajan kuluessa tapahtuvat muutokset segmenttien rakenteissa ja profiileissa. Toista dynamiikkaa taas lähestytään käyttäen nk. segmenttisiirtymien analyysia, jossa visuaalisin keinoin tunnistetaan samantyyppisesti segmentistä toiseen vaihtavat asiakkaat.

Visualisoinnin tehtävänä on tukea havaittujen kaavojen tulkitsemista sekä helpottaa tiedonsiirtoa analyysoijan ja päättäjien välillä. Visuaalisia tiedonlouhintamenetelmiä, kuten itseorganisoivia karttoja ja niiden laajennuksia, käytetään osoittamaan näiden menetelmien hyödyllisyys sekä asiakkuudenhallinnassa yleisesti että erityisesti asiakassegmentoinnissa.

Acknowledgements

During the past years, I have received enormous support from the kind people around me, to whom I would like to express my sincerest gratitude.

First and foremost, I would like to thank my supervisors, Prof. Barbro Back and Doc. Tomas Eklund, who have introduced me to the field of data mining and provided me with invaluable academic guidance, continuous advice and encouragement at every stage of my Ph.D studies. I cannot thank you both enough.

I would like to thank the reviewers of my dissertation, Prof. Ranjan Kini from Indiana University Northwest and Prof. Jussi Kantola from University of Vaasa, for their constructive comments on the manuscript. I am especially grateful to Prof. Ranjan Kini for accepting the invitation to be the opponent of my dissertation.

I have also benefited from working with my co-authors, Peter Sarlin, Annika Holmbom and Hongyan Liu. Without your knowledge and collaboration, this dissertation would have been much less than it is. I would also like to thank the other members of the Data Mining and Knowledge Management laboratory, Dorina Marghescu, Minna Tähtinen, Piia Hirkman, Henri Korvela, and Samuel Rönnqvist, whom I have worked closely with during the thesis process.

I acknowledge my gratitude to SOK for providing their data, materials and expertise for this research. Special thanks are due to Hannele Humaloja-Virtanen and Tero Mustonen for their absolute support for the project. I also want thank Prof. Hannu Vanharanta and Dr. Jarmo Toivonen for their assistance and insights given in the TITAN project.

I would like to express my gratitude to my colleagues at the Department of Information Technologies at Åbo Akademi for the financial, academic and technical support. Special thanks must go to Annika Holmbom and Piia Hirkman for translating the abstract of my dissertation. Many thanks are due to the entire staff of TUCS for their financial and administrative support, special thanks are due to TUCS Director Professor Johan Lilius and TUCS Secretary Irmeli Laine. The Nokia Foundation, the Foundation for Economic Education, the Academy of Finland (Kermit-project), the Klingendahl foundation, and the Education Section of the Chinese Embassy in Finland are appreciated for their financial support.

Last but definitely not least, I'm thankful to have the constant support of my family.

Åbo, September 2013.

Table of Contents

1.	Introduction.....	1
1.1	Background.....	1
1.2	Research aim and questions.....	2
1.3	Research methodology.....	3
1.3.1	Research frameworks.....	3
1.3.2	Design science.....	5
1.4	Contribution and publications.....	6
1.5	Overview of the thesis.....	9
2.	Customer Relationship Management.....	11
2.1	Introduction.....	11
2.2	The customer life cycle.....	12
2.3	Customer segmentation.....	14
2.3.1	Segmentation bases.....	15
2.3.2	Methods for conducting customer segmentation.....	17
2.3.3	Evaluation criteria for customer segmentation.....	18
2.4	Dynamics in a customer segmentation model.....	19
2.4.1	Temporal customer segmentation.....	20
2.4.2	Segment migration analysis.....	22
2.5	Summary.....	24
3.	Data Mining and the Knowledge Discovery Process.....	25
3.1	Introduction.....	25
3.2	The KDD Process.....	25
3.3	Data mining tasks.....	28
3.4	Summary.....	29
4.	Data mining methods.....	31
4.1	Visual data mining.....	31
4.2	The SOM.....	32

4.2.1	Training the SOM	35
4.2.2	Evaluation of the SOM	36
4.2.3	Visualizing the SOM.....	37
4.3	The SOTM	37
4.3.1	Training the SOTM.....	39
4.3.2	Evaluation of the SOTM	41
4.3.3	Visualizing the SOTM	41
4.4	Clustering of the SOM and the SOTM	42
4.4.1	Second-level clustering on the SOM and SOTM.....	42
4.4.2	Agglomerative hierarchical clustering	43
4.4.3	Evaluation of the two-level clustering	44
4.5	Transition probability on the SOM	45
4.5.1	Computing the transition probability on the SOM.....	46
4.5.2	Visualizing the transition probabilities	46
4.6	Classification techniques.....	47
4.6.1	The decision tree (DT)	47
4.6.2	The artificial neural network (ANN).....	49
4.6.3	The support vector machine (SVM).....	50
4.6.4	Evaluation	51
4.6.5	Ensemble.....	52
4.7	Summary	53
5.	Data	55
5.1	Data collection and integration	56
5.2	Data cleansing	58
5.3	Data transformation.....	59
5.4	Summary	60
6.	Results.....	61
6.1	Customer segmentation.....	61
6.1.1	Publication 1	62

6.1.2	Publication 2	64
6.2	Temporal customer segmentation	65
6.2.1	Publication 3	65
6.2.2	Publication 4	67
6.3	Segment migration analysis	67
6.3.1	Publication 5	68
6.3.2	Publication 6	69
6.4	Summary	70
7.	Conclusion	71
7.1	Review of the answers to the research questions	71
7.2	Limitations and future work.....	73

List of Original Research Publications

[1] Yao, Z., Holmbom, A., Eklund, T., & Back, B. (2010). Combining unsupervised and supervised data mining techniques for conducting customer portfolio analysis. In *Proceedings of the 10th Industrial Conference on Data Mining (ICDM 2010)*, Berlin, Germany, 12-14 July, 2010, pp. 292-307.

[2] Yao, Z., Eklund, T., & Back, B. (2010), Using SOM-Ward Clustering and Predictive Analytics for Conducting Customer Segmentation, In *Proceedings of the 10th IEEE International Conference on Data Mining Workshops (ICDMW 2010)*, Sydney, Australia, 14-17 December, 2010, pp. 639-646.

[3] Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2012), Temporal customer segmentation using Self-Organizing Time Map. In *Proceedings of the 16th International Conference on Information Visualization (IV 2012)*, Montpellier, France, 10-13 July, 2012, pp. 234-240.

[4] Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2013), Visualizing dynamics in customer behavior with the Self-Organizing Time Map. TUCS Technical Report No. 1085. *Submitted to Business & Information Systems Engineering*.

[5] Sarlin, P., Yao, Z., & Eklund, T. (2012), A Framework for State Transitions on the Self-Organizing Map: Some Temporal Financial Applications. *Intelligent Systems in Accounting, Finance and Management*, Vol.19, No.3, pp. 189-203.

Extends:

Sarlin, P., Yao, Z., & Eklund, T. (2012), Probabilistic Modeling of State Transitions on the Self-Organizing Map: Some Temporal Financial Applications. In *Proceedings of the 45th Annual Hawaii International Conference on Systems Science (HICSS 2012)*, Maui, USA, 4-7 January, 2012, pp. 1128-1137.

[6] Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2013), Combining Visual Customer Segmentation and Response Modeling. Accepted for publication in *Neural Computing and Applications*.

Extends:

Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2012), Combining Visual Customer Segmentation and Response Modeling. In *Proceedings of the 20th European Conference on Information Systems (ECIS 2012)*, Barcelona, Spain, 10-13 June, 2012, Paper 143.

List of Figures and Tables

Figure 1. The framework for paradigmatic analysis (Source: Iivari (1991)).	5
Figure 2. March and Smith's (1995) research framework of design science, modified to reflect the research in this thesis.	6
Figure 3. An overview of the thesis.	9
Figure 4. The customer life cycle progresses through different stages (Source: Linoff and Berry (2011, p. 29)).	13
Figure 5. The evolution of customer relationship in the relationship management stage of CRM.	14
Figure 6. Classification of segmentation methods.	17
Figure 7. Segmentation by migration patterns (Source: Bayer (2010)).	23
Figure 8. An overview of the KDD process (Source: Fayyad, et al. (1996)).	26
Figure 9. The structure of a typical SOM (Source: Demirhan and Güler (2011)).	33
Figure 10. An example of component plane visualization of the SOM model (Source: Yao, et al. (2012a)).	37
Figure 11. The functioning principles of the SOTM (Source: Sarlin (2013b)).	40
Figure 12. A multilayer perceptron with three layers.	50
Figure 13. An example of confusion matrix.	51
Figure 14. An example of ROC curve.	52
Figure 15. The source files containing customer-related information and their relationships.	56
Figure 16. The variables in the MCIF.	57
Figure 17. Computations and visualizations of transition probabilities on the SOM.	69
Table 1. A scheme for analyzing assumptions about the nature of social science (Source: Burrell and Morgan (1979, p. 3)).	4
Table 2. Major segmentation variables for consumer markets (Source: Kotler (2002, p. 149)).	16
Table 3. Related studies demonstrating segment dynamics by measuring changes in segment membership longitudinally.	20
Table 4. Segment migration probabilities (Source: Bayer (2010)).	23
Table 5. Comparison of related work of using the SOM for segmentation tasks.	34

List of Acronyms

ACORN	A Classification of Residential Neighborhoods
ANN	Artificial Neural Network
AUC	Area Under the Curve
BI	Business Intelligence
BMU	Best Matching Unit
CART	Classification and Regression Tree
CHAID	Chi-squared Automatic Interaction Detector
CRM	Customer Relationship Management
DE	Distortion Error
DT	Decision Tree
ERP	Enterprise Resource Planning
FN	False Negative
FP	False Positive
IS	Information Systems
IT	Information Technology
KDD	Knowledge Discovery in Databases
MCIF	Marketing Customer Information File
MDS	Multi-Dimensional Scaling
MLP	Multilayer perceptron
OLAP	Online Analytical Processing
PCA	Principal Component Analysis
QE	Quantization Error
RFM	Recency, Frequency, and Monetary
SOM	Self-Organizing Map
SOTM	Self-Organizing Time Map
SRM	Structural Risk Minimization
SVM	Support Vector Machine
TE	Topographic Error
TN	True Negative
TP	True Positive
VDM	Visual Data Mining

PART ONE: Research Summary

1. Introduction

1.1 Background

For a long time, the focus of modern companies in many industries has been shifting from being product-oriented to customer-centric organizations (Tseng & Piller, 2003). In recent years, this change has been particularly rapid due to the increasing interest in Business Intelligence (BI) in general, and Customer Relationship Management (CRM) in particular. The reason for this trend is two-fold. First, today's customers are continuously creating a "customer signature" that lends itself to analysis. Customers' behavior and activities can be easily recorded and stored through data warehousing, and customers' demographic and lifestyle information is readily available from a variety of public and private sources. Additionally, increasingly affordable and powerful computing resources have also made companies interested in developing their analytical CRM capabilities. Second, customers are more informed and sophisticated by accessing a variety of information sources, and this requires companies to improve their ability to understand customers and to use that understanding to create value for customers in order to achieve better customer satisfaction, market share, and profitability. Moreover, the Pareto Principle or the 80/20 rule that a small percentage of the customers accounts for a high percentage of sales revenue or profit is commonly accepted (Koch, 1999). Hence, companies should gain a thorough understanding of their customers, in particular the high-value customers, to maintain competitiveness. Customer segmentation, an important application of analytical CRM, is widely used to divide customers into distinct and homogeneous segments for collective management of individual segments. This enables companies to better understand the value of customers, their needs and preferences, and their purchasing behavior. This information can be used to predict customers' behavior and to develop differentiated marketing strategies at both the strategic and operational levels.

The application of data mining methods in CRM has been increasing in the industry (Ngai, et al., 2009). With the help of advanced algorithms from research areas such as statistics, artificial intelligence, and machine learning, data mining methods uncover hidden knowledge and patterns in large amounts of data, and provide evidence and support for decision makers (Turban, et al., 2007, p. 305). More importantly, predictive data mining methods can forecast customers' behavior and answer questions that are difficult or time-consuming for traditional approaches, that is, online analytical processing (OLAP), spreadsheets, and statistics (Rygielski, et al., 2002). Data mining is one of the steps in a larger framework: Knowledge Discovery in Databases (KDD) (Fayyad, et al., 1996), which refers to the process of extracting knowledge from data in the context of large databases. Clustering is one of the most popular data mining tasks. It attempts to divide data into groups, such that the differences among clusters are maximized and the variations within each are minimized. Therefore, the

capability of clustering algorithms to reveal natural groupings of data makes it a widely used technique for conducting customer segmentation.

To use the data mining results better, it is important to include subject-matter knowledge in the modeling process and to represent the mined results in a meaningful and effective way. Visualization techniques have been used for visual exploration and mining large datasets (Ferreira de Oliveira & Levkowitz, 2003; Keim et al., 2010).

According to Ferreira de Oliveira and Levkowitz (2003), there are two types of visualization for visual data mining (VDM): *visualization of the data* and *visualization of the data mining result*. Visual clustering, belonging to the latter, consists of techniques capable of simultaneously conducting clustering tasks and providing a multivariate visual display of the clustering results, thereby facilitating the exploration of useful patterns in the data. Nevertheless, few studies discuss the visualization of data mining results within the context of CRM (Ngai, et al., 2009). Therefore, this thesis places greater focus on using VDM techniques, in particular the Self-Organizing Map (Kohonen, 1982) (SOM) and its extensions, for solving customer segmentation problems.

Being a popular tool for exploratory data analysis, visual clustering does not guarantee its usefulness for customer segmentation purposes. The customer segmentation solutions based on summarized data of a particular period of time and traditional cluster analysis often provide only a static overview of customers' average characteristics and their aggregated shopping behavior. Their limitations have also been well documented in the literature in that the validity of a customer segmentation model is limited by various dynamics in the form of either segment instability (Hoek, et al., 1996; Steenkamp & Ter Hofstede, 2002; Blocker & Flint, 2007) or membership migration of individual customers (Calantone & Sawyer, 1978; Farley, et al., 1987) or both (Hu & Rau, 1995). These problems are becoming more important as the external market environment and customers' needs and preferences are experiencing rapid change. However, there is no indication that these dynamics have been extensively studied in the context of visual customer segmentation. In this thesis, the effectiveness of visual customer segmentation will be addressed, as well as different kinds of temporal dynamics.

1.2 Research aim and questions

In this thesis, our aim is to propose a framework to provide a holistic view of multivariate static and temporal patterns of customers' characteristics and purchasing behavior. This aim can be addressed from three perspectives in terms of their corresponding tasks in an analytical CRM system. First, customer segmentation can be used to gain insight into customer behavior and profitability. The solutions are often derived based upon aggregated customer data of a specific time frame, and thereby often disregard the fact that customers' characteristics and behavior may evolve. The second and third perspectives

explore the possible dynamics of a customer segmentation model in terms of segment stability and the change of segment membership of individual customers. The segment stability problem will be addressed using temporal customer segmentation where the change of segment structures and profiles are tracked over time, while the change of segment membership of individual customers is addressed using segment migration analysis, where customers with similar switching patterns among segments are identified visually. Based on the above-mentioned discussion, the research questions for this thesis are formulated as follows:

- RQ1.** How can customer segmentation be conducted and the results visualized?
- RQ2.** How can the changes of segment structures over time be detected and monitored?
- RQ3.** How can segment migration analysis be conducted to explore different migration profiles?

1.3 Research methodology

This section describes the research methodology adopted in this thesis for addressing the research questions in Section 1.2. It starts with an introduction of different research methodologies. Then, the research methodology adopted in this thesis, i.e., design science, is discussed in more detail.

1.3.1 Research frameworks

Burrell and Morgan (1979) proposed a classification framework for conducting social science research, as illustrated in Table 1. The authors divided the research approaches in social science into two categories: *objective* and *subjective*. The framework is based upon four paradigmatic assumptions: *ontology*, *epistemology*, *human nature*, and *methodology*.

According to Burrell and Morgan (1979), ontology deals with the nature of being, existence or reality. There two opposing views regarding the nature of reality. The objective approaches follow *realism*, i.e., reality exists independently of human perception, while the subjective approaches follow *nominalism*, i.e., reality exists as a subjective construction based upon and is influenced by human perception.

Table 1. A scheme for analyzing assumptions about the nature of social science (Source: Burrell and Morgan (1979, p. 3)).

Assumption	Subjectivist approach	Objectivist approach
Ontology	Realism	Nominalism
Epistemology	Positivism	Anti-positivism
Human nature	Determinism	Voluntarism
Methodology	Nomothetic	Ideographic

Epistemology, divided into *positivism* and *anti-positivism*, is concerned with how knowledge is acquired and validated. Positivism aims to explain reality by searching for patterns and relationships. Anti-positivism, on the other hand, claims that reality can only be understood by experiencing it directly, rather than by only observing behavior.

Human nature deals with relationship between humans and their environment, i.e., humans are determined by their environments (*determinism*) or humans affect or create their environments (*voluntarism*). Determinism regards human as being determined by the conditions and the environments they are in. Voluntarism sees humans as being autonomous and reality as the consequence of humans' will.

The methodological assumption deals with the methodologies for understanding reality. *Nomothetic* methodologies are based upon systematic protocol and techniques, aiming for scientific rigor. *Ideographic* methodologies place emphasis on detailed observation of the subject under research and obtaining first-hand knowledge.

The Burrell and Morgan classification framework shows two extreme approaches to social sciences: 1) a realist ontology with a positivist epistemology, a deterministic view of human nature and nomothetic methodologies, and 2) subjective ontology with an anti-positivist epistemology, a voluntaristic view of human nature and ideographic methodologies. However, the dichotomous classification of the sub-dimensions of subjective and objective approaches is considered simplistic when applied to information systems (IS) research (Iivari, 1991). Iivari (1991) extended Burrell and Morgan's framework to make it more suitable for IS research, which is an applied science that is neither regarded as an objective nor subjective approach and thus requires its own set of methods (cf. Figure 1). Compared to Burrell and Morgan's framework, Iivari (1991) included the human nature paradigmatic assumption as a dimension in his ontology paradigmatic assumption, and added *ethics* as a new paradigmatic assumption. Ethics is concerned with the responsibility of a researcher for the consequences of his/her research. In addition, he added the constructive method as the third research methodology to the methodology assumptions of Burrell and Morgan's framework, arguing that the special nature of information technology (IT) and IS

research requires its own set of methods concerned with the engineering of artifacts which include both conceptual models (e.g., models, frameworks and procedures) and practical artifacts (e.g., software). *Design science* and *constructive research* are often referred to as the same type of research. March and Smith (1995) and Hevner, et al. (2004) defined design science as being concerned with building and evaluating artifacts, which corresponds to the constructive methods proposed in Iivari (1991).

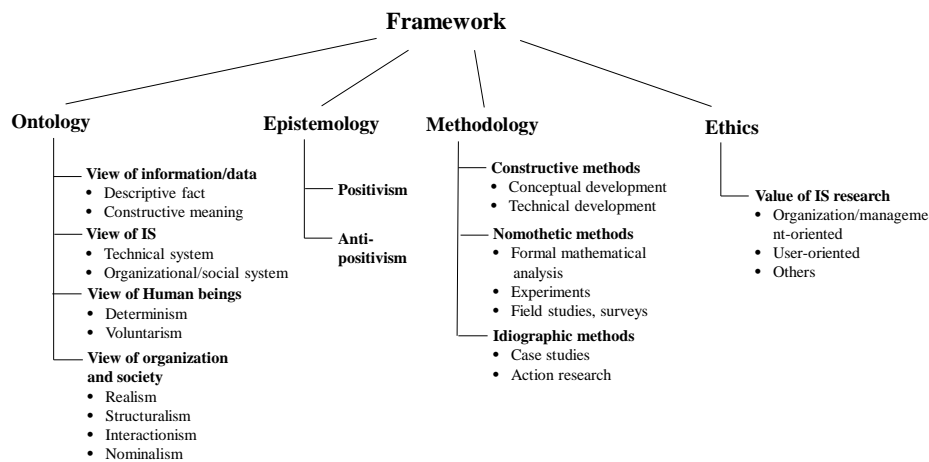


Figure 1. The framework for paradigmatic analysis (Source: Iivari (1991)).

1.3.2 Design science

Based on the arguments in Simon (1996) that IT research studies are artificial as opposed to natural phenomena, and that artificial phenomena can be both created and studied, a number of studies (March & Smith, 1995; Hevner, et al., 2004) identify two research paradigms within IT and IS: *descriptive* and *prescriptive* research. According to March and Smith (1995), descriptive research, or natural science, is a knowledge-producing activity and is explanatory in nature. Prescriptive research, or design science, on the other hand, is a knowledge-using activity that offers prescription for creating artifacts. Despite the inherent distinction between design science and natural science, they are interrelated in such a way that the artifacts created by design science can be the targets of natural science research. On the other hand, design science provides an effective means for testing and validating the claims of natural science research.

Design science consists of two activities: *building* and *evaluation* (March & Smith, 1995). *Building* is the process of constructing an artifact to perform a specific task. *Evaluation* concerns the assessment of the utility, effectiveness, or novelty of artifacts. March and Smith (1995) also defined four types of outputs of design science research: *constructs*, *models*, *methods*, and *instantiations*. *Constructs* form the vocabulary and symbols to describe problems within the

domain and to specify their solutions. They constitute a conceptualization to provide the language in which problems and solutions are defined and communicated (Schön, 1984). *Models* use *constructs* to abstract and represent the connection between problem and solution components (Hevner, et al., 2004). Models use constructs to represent a real world situation to the design problem and its solution space (Simon, 1996). Models aid understanding of problem and solution and their connections, as well as exploration of the effects of design decisions. *Methods* (e.g., mathematical algorithms or description of best practices) provide problem-solving guidance. *Instantiations* realize *constructs*, *models*, and *methods* in a real-world setting.

In this thesis, both design and evaluation of models are conducted. Relevant research efforts for this thesis are listed in Figure 2. The models as introduced in Section 1.2 will be built and evaluated.

Activities	Build	Evaluate
Outputs		
Constructs		
Model	<ul style="list-style-type: none"> ▪ Customer segmentation model ▪ Temporal customer segmentation model ▪ Segment migration model 	<ul style="list-style-type: none"> ▪ Clustering validation measures ▪ Evaluation of classifiers ▪ Quality measures for the SOM
Method		
Instantiation		

Figure 2. March and Smith’s (1995) research framework of design science, modified to reflect the research in this thesis.

1.4 Contribution and publications

The primary contribution of this thesis is to propose a toolset based upon VDM techniques to conduct customer segmentation. VDM techniques are capable of providing a better solution through embedding users in data exploration and the model building process. Given the fact that VDM techniques have not been widely used in CRM (Ngai, et al., 2009), the application of visualization techniques is regarded as a practical problem still open to research. In this thesis, the SOM and its extensions, that is, the self-organizing time map (SOTM) (Sarlin, 2013b), the second-level clustering on the SOTM (Sarlin & Yao, 2013), and transition probability on the SOM (Sarlin, et al., 2012), are applied in analyzing customers’ behavior and demographic characteristics. The created models are closely related and provide a methodological contribution to the literature. The three empirical studies – static customer segmentation, temporal

customer segmentation, and segment migration analysis – could be integrated as a tool to capture the dynamics of the customer base.

The second main contribution of this thesis is to explore both the static structure and the temporal dynamics of the customer base visually. Customers' behavior and demographic characteristics are often not static but change constantly. Dealing with non-static data is considered one of the most challenging problems in data mining research (Yang & Wu, 2006). Visualizing dynamics in data are also identified as one of the most challenging problems in the fields of information visualization and visual analytics (Chen, 2005; Wong, et al., 2012). In this thesis, the time element is incorporated explicitly in the model. Visual temporal clustering is conducted to explore the customer base and tracking customer behavior over time. Also, a visual, data-driven, and efficient framework is provided to illustrate customers' segment migration patterns. The two models combined provide a framework to identify the changes in the customer base at different points in time in a more detailed manner.

The third contribution lies in demonstrating the usefulness of the models in a real-world setting. The data used in this thesis, including customer demographic, lifestyle information, and their purchase history, are from a retailer's customer data warehouse including more than one million customers. Industry data like this are difficult to access and utilize for research purposes. In this sense, the data themselves provide a means for evaluating the suitability and generalizability of methodology proposed in this thesis.

The research work and results have been published in six papers listed on page iv. Each of the studies addresses the research questions posed in Section 1.2.

Publication 1 and *Publication 2* address **RQ1**. Both papers used the SOM to conduct customer segmentation to divide the customer base into different groups, and this enables companies to manage customers in each segment collectively and to target different segments more effectively. Customer segmentation uses historical customer behavior data; it is a retrospective analysis of the customer base. The one step further is to combine the SOM model and classification models to identify customers or segments with development potential. The classification model is built by classifying high- and low-spending customers identified in the segmentation model. In *Publication 1*, I was responsible for collecting and preprocessing data, and building the model. The analysis of the results was a joint effort between the authors. In *Publication 2*, I was again responsible for collecting and preprocessing the data and for building the models and comparing the predictive performances of three classifiers. I am the main author of both papers.

Both the segmentation model and its related classification model are based upon a specific timeframe; hence, these static segmentation models might overlook changes in customers' demographic characteristics and purchasing behavior. **RQ2** aims to explore the changes in segment structures over time, thereby providing a holistic view of multivariate temporal patterns in the customer base.

Publication 3 and *Publication 4* address **RQ2**. *Publication 3* uses the SOTM, a recently introduced adaptation of the SOM model for visual exploration of temporal cluster structures (Sarlin, 2013b), to visually explore and track customers' demographic characteristics and purchasing behavior. This approach is capable of roughly identifying the changes in the profiles of customer segments over time. In *Publication 4*, the SOTM for the temporal customer segmentation model is enhanced by pairing it with a second-level clustering, that is, first using the SOTM to produce the nodes, which are then clustered in the second stage. This enables one to identify the temporal changes in the multivariate cluster structures objectively when data are non-static, which is a normal phenomenon in data describing customer behavior. It also facilitates identification of the patterns of changing, emerging, and disappearing of customer segment structures.

The challenge of visualizing multivariate temporal patterns using SOTM has been introduced in Sarlin (2013b) and Sarlin and Yao (2013). In this thesis, the SOTM and the second-level clustering on the SOTM are applied to demonstrate its applicability in the context of CRM, with emphasis placed on the identification of effects of sales campaigns on customers' shopping behavior. Further, both papers aim to demonstrate the usefulness of the SOTM, especially the adaptability of the SOTM facing abrupt changes in neighboring periods. In both publications, I was responsible for the experimental setup (including collecting and preprocessing the data) and the application of the SOTM model/two-level clustering to the new domain. The enhancement of the SOTM to be applicable to the new domain was a joint effort between Sarlin and me. The analysis was a joint effort on behalf of all of the authors. I am the main author of both papers.

Temporal customer segmentation provides an overview of the changes in segment structures over time. This captures the dynamics of customer segments assuming that the segment structure changes over time. Another type of dynamics in customer segmentation is changes in segment membership of individual customers over time. It is not uncommon for customers in a homogeneous customer segment to exhibit different segment migration patterns, especially during sales events. It is particularly important to provide an intuitive anticipatory analysis to track customers' upward and downward migration patterns. **RQ3** deals with visual segment migration analysis, which is addressed in *Publication 5* and *Publication 6*. *Publication 5* introduces a framework for computing, summarizing, and visualizing transition probability on the SOM for temporal data. The framework includes computing and visualizing node-to-node and node-to-cluster transition matrices, as well as maximum state transitions. This framework is applied in *Publication 6* to combine customer segmentation and segment migration patterns during campaign periods. The paper contributes to the existing literature by proposing a framework for combining customer segmentation and campaign-driven segment migration analysis, which enables one to identify the static customer segment profile and the within-segment heterogeneity in terms of campaign response propensity in one view. In *Publications 5*, I was responsible for preprocessing part of the data, and setting

up and training the model based upon that part of the data. The analysis of the results was a joint effort by all of the authors. In *Publication 6*, I was responsible for preprocessing the data and for building the model. I am the main author for *Publication 6*.

1.5 Overview of the thesis

The remainder of the thesis is organized as follows.

Chapter Two introduces CRM. First, the major phases of the customer relationship life cycle and its related business processes are described, followed by an elaboration on the three data mining applications for enhancing relationship management, namely customer segmentation, temporal customer segmentation, and segment migration analysis.

Chapter Three presents the KDD process for our empirical data mining projects.

Chapter Four introduces the data mining methods utilized in this thesis, as well as the evaluation measures for quantifying model performances.

Chapter Five presents the data and the data preprocessing process for creating the mining datasets.

In Chapter Six, the training and analysis of the models (i.e., customer segmentation, temporal customer segmentation, and segment migration analysis) based on the six publications are summarized and presented.

Finally, in Chapter Seven, conclusions and suggestions for further research are presented. Figure 3 shows an overview of the thesis.

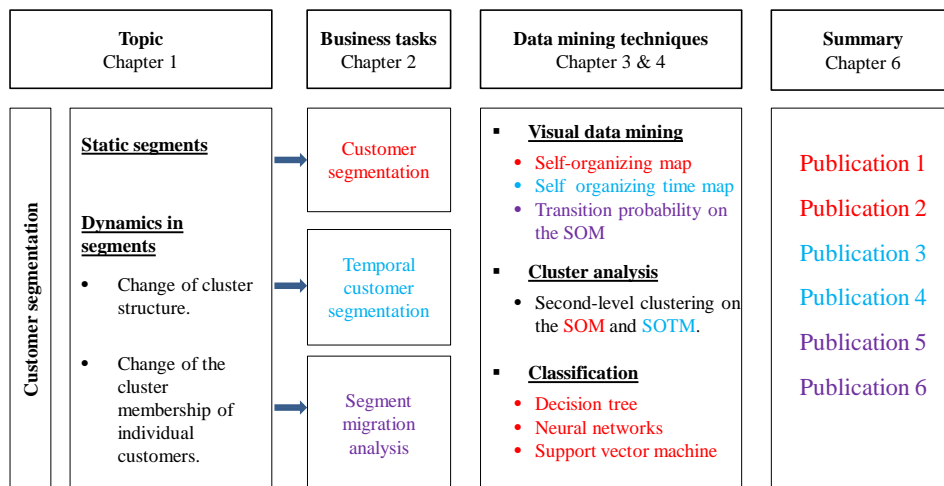


Figure 3. An overview of the thesis.

2. Customer Relationship Management

To gain a thorough understanding of CRM, this chapter begins with a discussion of the major phases of the customer relationship life cycle and its related business processes. This provides an overview of the evolution and progress of a customer relationship and suggests a dynamic approach for effective CRM. The dynamics of customer relationships in the retail industry are more volatile, since the switching cost is much lower than in the financial or telecommunications industries due to the typically non-contractual loyalty program setting in retailing (Reinartz & Kumar, 2000). Therefore, maintaining customer loyalty and satisfaction has more practical significance to retailers. Then, customer segmentation, an important analytical tool for enhancing relationship management, will be elaborated upon. The static approach to customer segmentation is first discussed, followed by the approaches to address the two types of dynamics in customer segments.

2.1 Introduction

CRM has its origin in relationship marketing, which is defined as marketing activities that attract, maintain, and enhance customer relationships (Berry, 1983, p. 25). There is no universal definition of CRM. In academia, the terms “relationship marketing” and “CRM” are commonly used interchangeably. The accumulation of customer data and the development of IT have provided companies with ample opportunities to collect and analyze data about their customers. This has led to the wide adoption of CRM tools in organizations in a range of industries. In the vendor and practitioner communities, CRM refers to technology-based customer solutions or information-enabled relationship marketing (Ryals & Payne, 2001). Boulding, et al. (2005) extended Ryals and Payne’s definition and define CRM using a holistic approach, as the “outcome of the continuing evolution and integration of marketing ideas and newly available data, technologies, and organizational forms.” Payne and Frow (2005) also emphasized the cross-functional nature of CRM, as an “integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications.”

Buttle (2004, p. 3) proposed three levels of CRM: *strategic*, *operational* and *analytical*. Strategic CRM aims to develop a culture, with the customer at the center. Operational CRM, also known as “front-office” CRM, focuses on the areas where direct customer contact occurs and the automation of these customer-facing parts of the business. The focus of this thesis is on the analytical aspects of CRM, that is, how the collected customer and transaction data can be used to support the decision-making process.

2.2 The customer life cycle

A customer life cycle can be both the customer's personal life cycle and that of the customer relationship (Linoff & Berry, 2011, p. 27). In this thesis, the term customer life cycle refers to the latter, which is the stages in the relationship between a customer and a company, as opposed to where they are within their lives. Reinartz, et al. (2004) supported this divisive view of the CRM process for the reasons that customer relationships evolve through distinct phases with different value contribution, and accordingly companies manage the various stages of the relationship systematically and proactively. Therefore, it is important to understand a customer life cycle and its strong association with customer revenue and profitability (Rygielski, et al., 2002).

According to Rygielski, et al. (2002), a customer life cycle framework includes the following four components: *prospects*, *responders*, *active customers*, and *former customers*. Prospects refer to potential customers in the target market. Responders are prospects who have shown an interest in a product or service. Active customers are people who regularly use the product or service. Former customers are defected customers who end their relationships with the company. Linoff and Berry (2011, p. 29) refined the framework by dividing active customers into *new customers* and *established customers*. New customers are responders who have registered or made a commitment with the company. Established customers are people with deepening relationship with the company.

In addition to the customer life cycle in CRM, it is also important to examine the CRM life cycle from companies' perspectives due to the fact that CRM is a dual creation of firm and customer value (Boulding, et al., 2005). Reinartz, et al. (2004) identified three primary dimensions of the CRM process: *initiation*, *maintenance*, and *termination*. Initiation includes customer acquisition and recovery management; the maintenance stage includes customer retention, up-/cross-selling, and referral management; while termination refers to exit management. Similar classification schemes can also be found in Ang and Buttle (2006). Similarly, in Linoff and Berry (2011, p. 29), four CRM dimensions (i.e., *customer acquisition*, *activation*, *relationship management*, and *winback*) are proposed for the five components of the customer life cycle, as illustrated in Figure 4. A customer relationship starts with customer acquisition, which is the process of attracting prospects. Customers who have exhibited an interest will respond and be activated. This provides the initial conditions of the customer relationship. Companies then build relationships with their customers to improve their satisfaction, retention, and profitability. Customers can be lost either in the form of forced churn or voluntary churn. Companies attempt to retain valuable defecting customers because it is argued that retaining existing customers effectively leads to significant improvements in overall profitability (Reichheld & Teal, 2001). For this reason, relationship management (i.e., improving existing customers' satisfaction, retention, and profitability) is considered the

most important stage of the CRM process, and it becomes the main focus of this thesis.

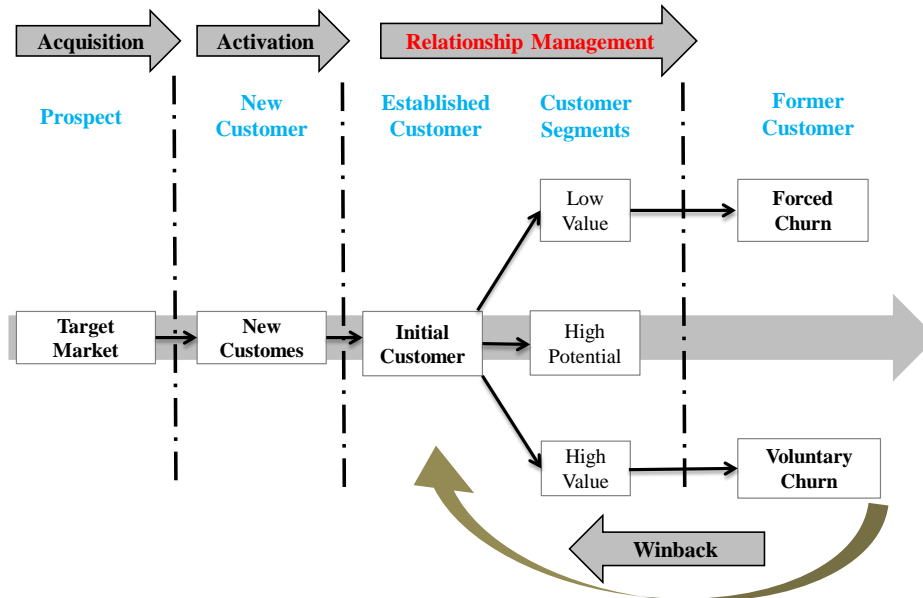


Figure 4. The customer life cycle progresses through different stages (Source: Linoff and Berry (2011, p. 29)).

The primary goal of relationship management is to enhance customers' relationships through customized communications, cross-selling, and customer segmentation (Payne & Frow, 2005). Customer segmentation is an effective approach to evaluating the value of the customers and understanding their behavior. A sound CRM system should address customer heterogeneity in their behavior to understand their preferences for more effective customer targeting and positioning of products and services (Boulding, et al., 2005). Figure 5 provides an example of the evolution of a customer relationship in the relationship management stage of customer life cycle. First, customer segmentation divides the customer base into distinct and internally homogeneous groups, such as low value, high value customers, and low value customers with high spending potential. Effective segmentation enables companies to interact with customers in each segment collectively and target them more effectively. Static customer segmentation solutions, while being effective in customer identification, are not good at capturing the dynamics in the customer base. Customers may evolve and exhibit different characteristics over time, which leads to changes in segment structures at the segment level (e.g., emerging, disappearing, and changing segments) and changes in segment memberships of individual customers. Temporal customer segmentation is capable of capturing the first kind of dynamics while segment migration analysis the latter one. Temporal customer segmentation provides a holistic view of segment evolution; it could identify multivariate temporal patterns of customers' characteristics and

their purchasing behavior. However, when the segments are relatively stable in terms of their sizes and differing characteristics while individual customers change their segment memberships actively, the temporal customer segmentation lacks the capability for tracking customer migration patterns among different segments, and therefore, cannot address this kind of dynamics. For example, customers in the same segment could change their shopping behavior and react in different ways to a sales event. This may result in their switching between different value segments. Segment migration analysis enables the identification of event-driven segment migration patterns and within-segment heterogeneity in terms of event response propensity. In the following sections, customer segmentation is first discussed, followed by a discussion of the methods for detecting the above-discussed dynamics in customer segmentation: temporal customer segmentation and segment migration analysis.

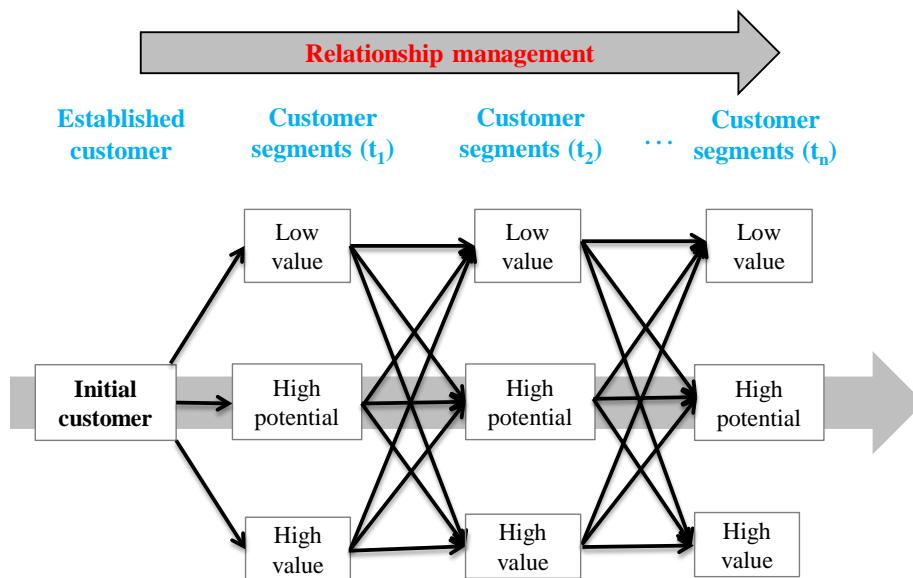


Figure 5. The evolution of customer relationship in the relationship management stage of CRM.

2.3 Customer segmentation

The origin of customer segmentation dates back to the introduction of the concept segmentation in Smith (1956), where he proposed the practices of managing relationships differently within different customer segments to satisfy the specific needs of any target market. Smith (1956) claimed that “market segmentation involves viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desires of customers for more precise satisfactions of their varying wants”.

Customer segmentation divides the customer base into distinct and internally homogeneous groups, which enables effective targeting and management of customers. Nowadays, customer segmentation has been widely adopted as a decision-support tool. The benefits of market segmentation brought to marketers include a deeper understanding of the market, the prediction of customers' behavior, and the capability of finding new market opportunities (Hoek, et al., 1996). In view of the central role of customer segmentation in this thesis, the topic is discussed from the following perspectives. First, the possible bases for segmenting customers are presented, which is followed by a discussion of the techniques for conducting customer segmentation. Finally, the criteria for evaluating the desirability of a customer segmentation model are discussed.

2.3.1 Segmentation bases

Effective customer segmentation depends on choosing the appropriate segmentation bases (i.e., a set of variables or characteristics for assigning customers to homogeneous segments) (Wedel & Kamakura, 2000). Customers differ in their tastes, needs, attitudes, motivations, lifestyles, family size, and so on (Chisnall, 1975, p. 264), and all these characteristics can be used to segment customers.

Kotler (2002, p. 144) made a distinction between the consumer and the business markets for segmentation tasks, in which the author classifies segmentation variables into four bases, i.e., *geographic*, *demographic*, *psychographic* and *behavioral* bases, for consumer market segmentation as illustrated in Table 2. It is noted that profitability, an important measure for customer value, can serve as a variable for behavioral segmentation because of the central importance of profits. The segmentation base chosen to subdivide customer bases depends on a range of factors, such as the type of product, the method of distribution, the media for market communication, and the motivation of buyers (Chisnall, 1975, p. 266). Different segmentation bases provide views with regard to the customer base from different angles. It is not uncommon to use multiple bases for segmentation purposes (Tynan & Drayton, 1987).

Geographic segmentation divides customers or markets nationally, regionally, or locally. Markets can also be subdivided by demographics, such as age, gender, income, and lifestyle. Although demographic data are usually easily available and widely used, the use of demographic variables has long been criticized for its lack of predictive power of customers' future behavior (Frank, 1967; Haley, 1968). Recently, several customer classification systems have been developed by combining geographic information (e.g., postcode data) and demographic data (e.g., census data), for example, Richard Webber's ACORN (A Classification of Residential Neighborhoods) (ACORN, 2013) and the Mosaic system (Experian, 2010) developed by the cooperation of Richard Webber and a UK company called Experian. Both systems classify households according to the type of neighborhood to which they belong. Neighborhood types are recognized and profiled by a range of variables, including customers' demographics,

psychographics, and shopping behavior. The combination of geographic and demographic information makes the system an effective tool in direct marketing. Psychographic data used for segmentation purposes are typically from research instruments, such as questionnaires, polls, and queries. However, customers tend to answer the questions according to their intentions, which might not be in accordance with their actual behavior. For example, one study found that 75% of the respondents claim that they are willing to buy more expensive but environmentally-friendly products, even though only 14% of them actually performed according to their original claims (Johnson & Selnes, 2004).

Table 2. Major segmentation variables for consumer markets (Source: Kotler (2002, p. 149))

Geographic	Region, City, Density, Climate
Demographic	Age, Family size, Family life cycle, Gender, Income, Occupation, Education, Religion, Race, Generation, Nationality, Social economic ranking, Life stage
Psychographic	Lifestyle, Personality
Behavioral	Buying pattern, Usage data, Loyalty status, Channel, Attitude towards product, Profitability

IT has enabled the collecting, storing, and delivering of customer-centric data more efficiently. The availability of large amounts of customer data in data warehouses facilitates analyzing data patterns and extracting knowledge for building better customer relationships. Unlike demographic and psychographic data that are commonly obtained from external sources, data describing customers' purchasing behavior are readily available for decision-support purposes. Not surprisingly, a survey involving 228 Dutch database marketing companies reported that behavioral variables, such as RFM variables (i.e. recency, frequency, and monetary value) and response are found to be the most employed variables for segmentation purposes (Verhoef, et al., 2003). Many studies also report the usefulness of utilizing behavioral data for segmentation purposes in various situations (Kiang & Kumar, 2001). Denny, et al. (2008) reported that purchasing records can predict customers' future behavior more effectively than demographic segmentation. The emphasis of this thesis, therefore, is on using information about customers' purchasing behavior to set up the models, where customers are grouped by behavioral and usage characteristics. Demographic information, an important consideration in delivering marketing strategies, is also incorporated into the analysis, for example, for segment profiling.

2.3.2 Methods for conducting customer segmentation

According to Wedel and Kamakura (2000, p. 17), segmentation methods can be classified based upon two dimensions. First, they can be classified into either *a priori* or *post-hoc* approaches depending upon whether the type and number of segments are predetermined. The other way of classifying segmentation methods divides them into *descriptive* and *predictive* methods depending upon whether they examine the relationship between target and explanatory variables. Based upon these two classification dimensions, segmentation methods can be classified into the following four types as illustrated in Figure 6.

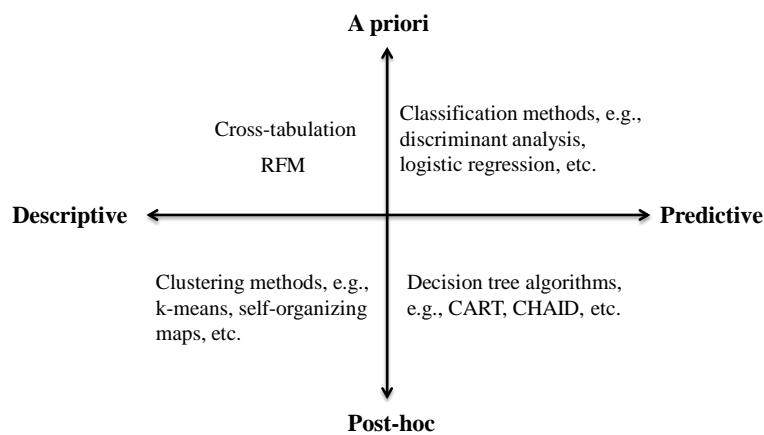


Figure 6. Classification of segmentation methods.

In *a priori descriptive* methods, the type and number of segments are determined before data are collected. Cross-tabulation, a contingency table that provides summary information of categorical data, is a popular tool for *a priori* descriptive segmentation. It is also the most-widely used method in database marketing (Verhoef, et al., 2003). One problem of using cross-tabulation for detecting variable associations is that higher order interactions are difficult to detect and interpret in tables (Wedel & Kamakura, 2000, p. 18). RFM analysis, regarded as the corner stone of direct marketing (Miglautsch, 2000), is a popular method for post-hoc descriptive segmentation. RFM analysis prescribes a segmentation of a customer base upon past behavior using three variables: 1) *recency*, the time elapsed since the last purchase; 2) *frequency*, the number of purchases made over a defined time period; 3) *monetary value*, the monetary value of the past purchases. Hughes (2006) applied RFM analysis to identify segments of customers with high response propensity to a marketing campaign. He divided customers into 125 cells, composed by the five groups for each of the three dimensions. A test mailing was then sent to the sampled subset of each cell to determine the expected response. One of the advantages of this approach is that each cell can be treated as an independent segment and this gives an advantage when the relationship between customer response and the three

explanatory variables is not monotonic (McCarty & Hastak, 2007). Despite many good properties of the RFM analysis, such as simplicity and explainability, the predefined rather than data-driven grouping of customers, the uneven data distribution among the cells (West, et al., 2005; Budayan, et al., 2009), and the high correlation between frequency and monetary value (Kifer, et al., 2004) have imposed limitations on the use of RFM analyses.

In *post-hoc descriptive* methods, internally homogeneous segments are formed based upon a set of measured characteristics. Clustering methods have also been widely used and are regarded as the most popular methods for post-hoc descriptive segmentation (Wedel & Kamakura, 2000, p. 19). As introduced in Section 1.1, clustering analysis attempts to divide data into groups to maximize the inter-cluster differences and minimize the intra-cluster variations in a data-driven way. This property makes clustering algorithms a popular tool for conducting customer segmentation. Reviews of applications of clustering methods in market segmentation were first produced by Punj and Stewart (1983) and Saunders (1993); a more recent review is provided by Wedel and Kamakura (2000, pp. 39-69). The clustering methods applied in this thesis will be introduced in Section 4.1.

Post-hoc predictive methods examine the relationship between a dependent variable and a set of independent variables. The segments formed by *post-hoc predictive* methods are homogeneous in terms of the relationships between dependent and independent variables. DT models – Chi-squared Automatic Interaction Detector (CHAID) and Classification and Regression Tree (CART) – have been widely used for post-hoc predictive segmentation for their simplicity, transparency, and predictive power. Rygielski, et al. (2002) compared the performance of a CHAID and an ANN model in the CRM context, and found that the ANN provided more accurate models, whereas the CHAID was easier to implement and the result of a CHAID model as a set of rules was more intuitive and understandable. McCarty and Hastak (2007) compared the predictive performance of a CHAID and a logistic regression model for direct market segmentation, and found that the CHAID provided more accurate and stable performance. In addition, the CHAID is able to capture the non-monotonic relationship between the dependent and independent variables.

In *a priori predictive* methods, a definition of *a priori* segments is first established, followed by the use of classification models to describe the relationship between segment membership and a set of independent variables. For example, the segments can be first formed by one of the socio-demographic variables, which are then classified using classification models.

2.3.3 Evaluation criteria for customer segmentation

With the appropriate variables selected and the proper techniques applied to the data, different segmentation results are produced. However, not all of the achieved segmentations are meaningful. In order for them to be useful for

developing a proper marketing strategy and to target potential customers, there are five criteria for evaluating the desirability of a customer segmentation model (Kotler, 2002, p. 153):

- *Measurable*: The characteristics of the segments can be measured and identified.
- *Substantial*: The segments should be large and profitable enough, and worth investing in. They should also be feasible to reach with a tailored marketing program.
- *Accessible*: The segments should be meaningful to the company and be effectively reached and served.
- *Differentiable*: Each segment needs to be distinguishable from others and to respond differently to a marketing program.
- *Actionable*: Effective marketing programs can be carried out for the selected segments by considering the objectives and resources of the company.

According to Wind (1978); Thomas (1980); Steenkamp and Ter Hofstede (2002), stability is also an important factor to be considered when evaluating the desirability of a segmentation model.

2.4 Dynamics in a customer segmentation model

One important but often ignored aspect regarding a segmentation model is that the identified segments can change in terms of their characteristics and sizes. Yet, one of the features of a good segmentation model is that the model should be stable, as introduced in the previous section. Since markets are becoming more competitive and customers' needs and preferences are evolving constantly, this may lead to changes in the initially identified segments. If changes in one or more segments are observed, the original targeted marketing strategies should be adjusted accordingly. Wind (1978) summarized three reasons of segment instability: 1) the basis for segmentation, 2) the volatility of the market place, and 3) consumer characteristics. A more specific and less general segmentation basis can lead to instable segments. The more volatile the market is, the less stable the segments would be. The changes in customers' life cycles can also lead to segment instability. In order to address the dynamics for a customer segmentation model explicitly, Lingras, et al. (2005) summarized two types of temporal changes in clustering in the context of CRM: *changes in cluster composition over time* and *changes in cluster memberships of individual objects*.

Changes of cluster composition over time have been addressed by examining segment instability for market segmentation derived from clustering analysis. Assael and Roscoe Jr (1976) and Beane and Ennis (1987) pointed out the problem of segment instability in their reviews of market segmentation techniques. Hoek, et al. (1996) compared the results of previous studies and

conclude that “it seems illogical to expect the size, composition, and behavior of market segments defined in these terms to remain constant.”

The study of variations in cluster memberships of individuals assumes that the cluster structure remains stable over time, but customers can switch between segments. A number of empirical studies (c.f. Table 3) examined the stability over time of identified segments. For example, Calantone and Sawyer (1978) and Brangule-Vlagsma, et al. (2002) showed that the identified segments were stable in terms of their distinguishing characteristics and size, but individual objects (e.g., a household or a customer) showed a more active switching pattern between segments. Farley, et al. (1987) and Hu and Rau (1995) went a step further and provided insights on the triggers of the phenomenon of segment switching based upon external data, such as customer survey. This implies that an analysis should be made following the stable and valid segmentation model derived on the panel data to identify the dynamics in the segment at individual level.

2.4.1 Temporal customer segmentation

The segmentation solutions discussed in Section 2.3.2 are often based upon summarized data of a particular time interval and thus only provide a static snapshot of the underlying customer base. However, customer relationship is a time-centric process (Reinartz, et al., 2004).

Table 3. Related studies demonstrating segment dynamics by measuring changes in segment membership longitudinally.

Work	Segmentation base	Application Area	Conclusion
Calantone and Sawyer (1978)	Desired attributes and benefits	Retail banking market	Segments are stable but individual consumers' segment membership changed significantly.
Farley, Winer, and Lehmann (1987)	Responsiveness to media	Grocery market	There was a high degree of instability in segment membership, with half of the households switching segments.
Hu and Rau (1995)	Customer characteristics and purchasing behavior	Four industries including banking, telecommunication, pharmaceuticals and chemical	Segments are unstable to varying degrees.
Brangule-Vlagsma, Pieters, and Wedel (2002)	Personal values	Consumer survey (no product)	Segment types are stable but there are changes in segment size and cluster memberships for individual customers.

To provide a holistic view of multivariate temporal patterns in data, time series clustering has been widely applied to search for similar shapes in trajectories. In the context of CRM, there are studies related to the direct use and analysis of temporal sequences of raw data. D'Urso and De Giovanni (2008) proposed a number of dissimilarity measures for comparing time sequences and integrating the dissimilarity measures into clustering techniques to conduct market segmentation. West, et al. (2005) applied the SOM and K-means to illustrate temporal variations in customers' loyalty and purchasing behavior. The authors first sorted the time-series values of customers' purchasing behavior variables so that customers with similar purchasing behavior, but who made purchases at different points in timing, are comparable. For example, a customer with three-week spending of EUR 10, EUR 20, and EUR 30 will be considered as having the same purchasing pattern as a customer with three-week spending pattern of EUR 20, EUR 30, and EUR 10. The sorted time series were then clustered and the segment profiles were compared. The clustering part of the study is the same as in traditional customer segmentation, only the preprocessing of the data and the representation of the clustering results are specially made to illustrate temporal variations in customers' purchasing behavior.

The time series clustering approach to temporal customer segmentation, however, aims at identifying entities of similar time-series patterns, rather than comparing the cluster structures and profiles across time. In addition to the above mentioned studies that use temporal data directly (e.g., time series clustering), a growing number of studies analyze the changes in cluster composition, assuming that there are temporal changes in cluster structures, for example, shrinking, expanding, emerging and lots of certain clusters. Chakrabarti, et al. (2006) introduced a framework for clustering data over time, which the authors define as a class of problem called *evolutionary clustering*. Evolutionary clustering produces a sequence of clusterings for each time point of temporal data. In order to smooth the clustering sequence, the framework aims to achieve a balance between clustering results being representative of current data and not being too deviant from the previous clustering result. The changes in clustering sequence are quantified by some quality measures that make it difficult for a user to understand the changes in cluster structures. Adomavicius and Bockstedt (2008) introduced a visualization technique for illustrating evolutionary clustering results. The clusters are represented using nodes, which are shown if their sizes are above a predefined threshold. The nodes between two adjacent time points are connected by an edge, which is again controlled by a predefined parameter measuring inter-cluster distances. This approach requires users to experiment with the number of clusters for each point in time and parameters for representing clusters and the connections between clusters at adjacent time points. Denny, et al. (2010) introduced a visualization approach for illustrating temporal changes in cluster structures. Compared to Adomavicius and Bockstedt's (2008) approach, this approach does not require users to experiment with the optimal number of clusters. The authors used distance matrix-based

visualizations and cluster color linking to facilitate comparing different maps and proposed a relative density metric for visualizing the structural changes of clusters. However, with this approach it is difficult to summarize the changes in cluster structures of more than two SOMs. Therefore, it is important to provide a holistic view of multivariate temporal patterns for better understanding of the evolution of and dynamics in a customer base. The method for conducting temporal customer segmentation in this thesis will be discussed in Section 4.3.

2.4.2 Segment migration analysis

As discussed in Section 2.2, customers can evolve over the relationship life cycle and accordingly switch between different segments, that is, a customer ends up in a segment different from the one he or she has been in. Segment migration analysis tracks customers' migration patterns among segments. For a behavioral customer segmentation model, the segment migration analysis can be used to identify customers who actively respond to a sales campaign, those that maintain their normal buying behavior, and those that avoid such events. For a value-based segmentation model, the segment migration analysis can distinguish between customers that increase in value, those that are drifting downwards, and those that are at risk of defection.

Studying customers' migration patterns involves two components: the level of the change and the corresponding probability of the change (Homburg, et al., 2009). Accordingly, the risk in customer relationship refers to the probabilities of migrating among segments of different values to the company. For example, customers that exhibit medium profitability with a high chance of becoming highly profitable deserve different treatment and prioritization than those who are currently highly profitable but have a high risk of defection. These dynamics should be captured proactively and taken into account in a CRM system. Therefore, there is a need for a means to identify future customer dynamics from customers' purchasing behavior and characteristics. The migration patterns in the customer base can be easily tracked and used to calculate the migration probabilities among segments.

These dynamics are commonly represented by a matrix of switching probabilities (c.f. Table 4), which also correspond to a first-order Markov chain, where the estimated switching probability p_{ij} from segment i to segment j is the percentage of customers who were in segment i in period $t-1$ and are in segment j in period t (Homburg, et al., 2009). Bayer (2010) presented a visualization method (c.f. Figure 7) for showing different migration patterns in a migration matrix. The author uses different colors to indicate whether a segment is migrating upwards or downwards along the value deciles.

Table 4. Segment migration probabilities (Source: Bayer (2010))

Value segment time 1	Value segment time 2			Gone (%)
	High (%)	Medium (%)	Low (%)	
High	50	28	14	8
Medium	5	46	19	30
Low	1	23	20	57

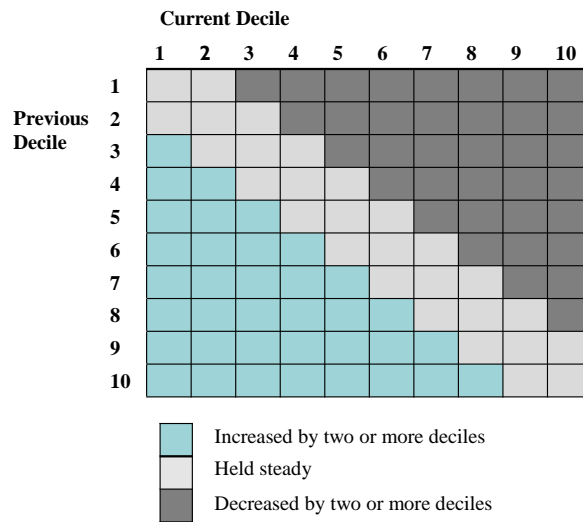


Figure 7. Segmentation by migration patterns (Source: Bayer (2010))

Segment migration analysis can also be conducted by finding changes in segment composition. Lingras, et al. (2004) proposed a method to address the changes in cluster composition and the cluster membership of individual objects. Web usage data are clustered and labeled for the first period, which is repeated for subsequent periods. Then, the changes in cluster size over time are displayed using a set of bar charts. This method does not account for cluster dynamics (e.g., detection of new clusters, lost clusters, and merging clusters) because the method assumes that cluster structures remain the same and the same set of cluster labels can be used across time. Unlike in Figure 7, where customer ranks are used for migration analysis, the assumption of static web usage patterns could not be validated. Ha, et al. (2002) studied the evolution of customer segments as the output of a SOM and a DT model. The managerial implications of the changes in customer behavior were made based on the assumption that segment transition

probabilities are static over time. This assumption, as was stated earlier, is often invalid in today's business environment. In this thesis, migration patterns and static customer segmentation will be integrated cohesively to provide a single view regarding the profiles of customers migrating among segments during campaign periods; that is, a visual clustering approach is applied to model the campaign response among customers in different segments. This approach will be discussed in more detail in Section 4.6.

2.5 Summary

This chapter has presented the concept of CRM. The stages in the relationship between a customer and a company have been discussed, aiming to pinpoint the focus of this thesis in the context of analytical CRM. Considering the importance of customer segmentation in developing a customer relationship, this topic has been addressed from two perspectives, that is, static and dynamic segmentation, where the dynamic segmentation was further divided into temporal customer segmentation and segment migration analysis. The analysis methods for addressing these problems were discussed. In the following chapter, the data mining and the KDD process will be introduced, providing a holistic view of the analytical process for addressing the business problems discussed in this chapter.

3. Data Mining and the Knowledge Discovery Process

This chapter will first introduce the KDD process which provides a general framework of finding knowledge in data. Each step of the process is referred to in corresponding sections of the thesis. Different data mining tasks are discussed for matching the aims of an analytic task to particular data mining algorithms.

3.1 Introduction

The traditional methods of extracting knowledge and useful information from data are heavily dependent upon manual analysis and interpretation, which is often time-consuming, costly, and error-prone. As the amount of data collected and stored in operational transaction systems and databases has been increasing at an unprecedented rate, traditional methods have become impractical for handling such amounts of data, and therefore, more sophisticated and affordable data mining tools need to emerge. Data mining has gained wide attention due to its ability to access, model, and visualize key relationships in data, and it has been applied in a wide variety of areas, for example, science, marketing, investment, fraud detection, manufacturing, and telecommunication (Fayyad, et al., 1996). In CRM, data mining aims at helping companies improve the performance of marketing and sales through a better and deeper understanding of the customers, and for greater accuracy in predicting their behavior.

Data mining is sometimes synonymously referred to as KDD, which has emerged as a field for studying new computational theories and tools to aid humans in extracting useful information and knowledge from large masses of raw data. In this chapter, the process of KDD, an approach that supports the entire process of data mining, is presented.

3.2 The KDD Process

The term KDD was introduced in Piatetsky-Shapiro, et al. (1996), where the authors emphasized that knowledge was the end product of data-driven discovery. Similarly, Fayyad, et al. (1996) stated that “KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.” The authors emphasized that KDD was a *process* involving numerous steps, such as data preprocessing, search for patterns, model evaluation and improvement, in an iterative fashion with many decisions made by the analyst. The process is *nontrivial*, meaning some search or inference based upon the mined result is involved. This implies that unlike the verification-based approach, where the hypothesis is first posed and then verified, the KDD process is a discovery-based approach in that the hypotheses and the assumptions about patterns in data are obtained as a result of the analyses (Piatetsky-Shapiro, et al., 1996). *Identifying patterns* refers to fitting a model to data, or extracting high-level knowledge from voluminous and complex low-level data. This is the ultimate goal of a KDD endeavor. Fayyad, et al. (1996)

considered a pattern to be knowledge based upon some measures of interestingness. The discovered patterns being *valid*, *novel*, *useful*, and *understandable* means that they should be generalizable to new data, be previously unknown, bring utility to decision-makers, and be easily accessible.

The nine-step KDD process proposed by Fayyad, et al. (1996) (c.f. Figure 8) is iterative and interactive, where knowledge is discovered, and feedback from later steps can cause the earlier ones to be redesigned and modified. The outline of this process is introduced below.

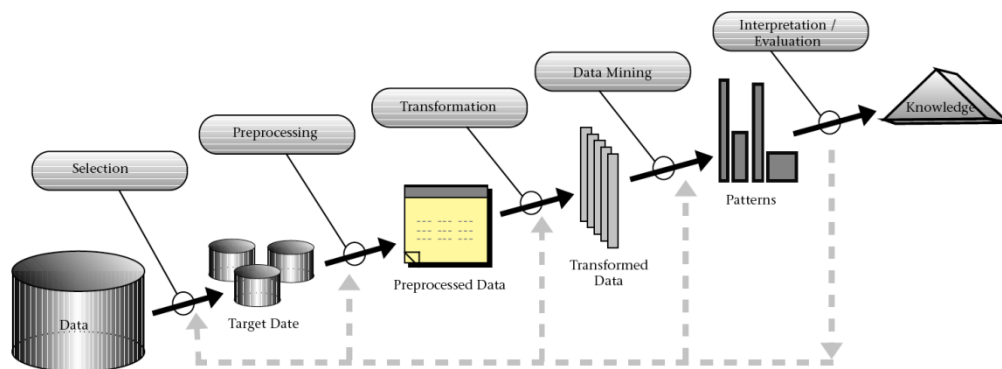


Figure 8. An overview of the KDD process (Source: Fayyad, et al. (1996))

1. Understanding of the application domain

During the initial step, much of the work relates to setting up KDD project goals and obtaining all prior knowledge and relevant resources required in implementing the KDD project. It is important to involve the owners of the business problems in identifying the goal of KDD from their viewpoints. By doing this, the people conducting the KDD project know how to deliver the results to best address the problems. In this thesis, the motivation and aims of the research were discussed in Chapter One.

2. Creating a target dataset

Here, data relevant to the analysis should be selected. First, data need to be integrated from multiple sources, for example, data warehouses, web servers, enterprise resource planning (ERP) systems, marketing databases, and external data providers, into one dataset. Sometimes, efforts should be focused on a subset of variables and data samples. The information regarding the data sources and the aggregation process will be introduced in Sections 5.1.

3. Data cleaning and preprocessing

Unclean data are common in the real world. It has been shown that up to 40% of data collected is unclean in some way (Maimon & Rokach, 2005, p. 22).

Therefore, the data should be preprocessed prior to mining, to avoid the trouble of “garbage in – garbage out.” In this step, the data validity is improved by cleansing, such as imputing missing values, smoothing out noisy data, and removing outliers, to improve the speed and quality of the process.

4. Data reduction and projection

The availability of variables has pros and cons. More variables provide access to information and knowledge piled up in the data. However, too many variables can cause the input data to be sparse and increase the risk of overfitting and correlation among input variables. The third and fourth steps are related to data preparation, which accounts for about 80% of the time spent on a data mining project (Berry & Linoff, 2004, p. 375). In this thesis, the third and fourth steps of the KDD process will be discussed in more detail in Sections 5.2 and 5.3.

5. Matching the goals of the KDD process to a particular data mining method

In step five, the goal of the KDD process is translated into a data mining problem. The knowledge mined from the KDD process should align with the business problems defined in the first step. A classification scheme of data mining methods according to the kinds of knowledge to be mined will be discussed in more detail in the following section.

6. Choosing the data mining methods and tools

In step six, the specific data mining methods are chosen to uncover the underlying patterns. Each specific data mining algorithm has its own advantages and challenges.

As regards to fifth and sixth steps of the KDD process, the goals of the KDD process in this thesis, i.e., customer segmentation and exploring dynamics in a customer segmentation model, are discussed in Chapter Two. In Chapter Four, the motivations for choosing a particular data mining method are discussed.

7. Deploying the chosen data mining method(s)

In step seven, the data mining algorithm is applied to the selected dataset. Different parameters should be experimented with in order to search for patterns and relationships. The terms data mining and KDD are sometimes used interchangeably. Here, Fayyad, et al. (1996) emphasized that the application of data mining algorithms was only an essential step in the whole KDD process.

8. Interpreting the mined patterns

Step eight focuses on reviewing the constructed model with respect to the analytic goals, so that the discovered patterns are valid, novel, useful, and understandable. This step also involves using visualization techniques to show the extracted patterns. As regards to the seventh and eighth steps of the KDD process, the data analysis process and the results will be discussed in Chapter Six.

9. Acting on the discovered knowledge

The purpose of a KDD project is to generalize knowledge from the data collected and to take action based on the discovered knowledge. Effective utilization of the discovered knowledge is of great importance to the success of the entire KDD process. Organizations should ensure that the discovered knowledge can be interpreted into business insights and integrated into business processes. This stage of the KDD process will be discussed in Chapter Seven, i.e., the review of the answers to the research questions.

3.3 Data mining tasks

The nine-step process indicates that KDD is a multidisciplinary activity including data retrieval, preprocessing, analysis, and interpretation of the results. As mentioned above, one essential step of the process is the application of specific data mining algorithms for extracting patterns from data. The patterns and knowledge discovered by data mining are largely concerned with the selected data mining methods. There is no unifying scheme for classifying data mining techniques, but the following are commonly identified in Fayyad, et al. (1996) and Han, et al. (2011, pp. 41-44).

Characterization and discrimination (Summarization)

Data characterization refers to methods that provide a summarization or comparison of subsets of data. Statistical summaries and various visualization tools can be used for presenting the results of data characterization. The widely used data warehousing and BI technologies have become capable of conducting data characterization in the form of multi-dimensional tables, which are also known as OLAP cubes. The information retrieved through interactive operations on the cubes (e.g., roll-up, drill-down, slice and dice, and pivot of data) and visual representation provides more efficient access to the information in data. The methods in this category, corresponding to the *a priori* descriptive methods in Figure 6, can be used for segmentation purpose.

Classification and regression

Classification and regression are also known as supervised learning, in which there are training data and associated target values. Classification and regression attempt to search for patterns in a training dataset so that they can predict, rank, or categorize new data into predefined target variables. In classification models, there are some predefined class labels and pre-classified training data. The model tries to apply the relationships learned from the trained data to the unlabeled data, in order to assign them to predefined classes. Regression models differ from classification in that they deal with continuously valued outcomes. In the CRM context, supervised data mining can be used for response modeling, churn analysis, fraud detection, customer lifetime value estimation, etc. (Linoff & Berry, 2011, pp. 85-86). The methods in this category can also be used for

customer segmentation and correspond to both post-hoc predictive and *a priori* predictive methods in Figure 6.

Clustering

Differing from supervised data mining, clustering models, following the unsupervised learning paradigm, divide data items into a set of clusters without using any predefined class labels. As clustering attempts to maximize intra-cluster similarity and at the same time, minimize inter-cluster similarity. It is therefore suitable to use when class-labeled data does not exist. It can also be used to generate class labels for subsequent analysis using supervised data mining methods. Clustering has been widely used for customer segmentation as discussed in Section 2.3.2. It corresponds to the post-hoc descriptive methods in Figure 6.

Mining frequent patterns, associations, and correlation (Dependency modeling)

This type of techniques aims to identify which items tend to co-occur, such as *frequent itemsets* or *frequent subsequences* (Han, et al., 2011, p. 43). A *frequent itemset* refers to a set of items that appear together frequently in a predefined time frame. This kind of co-occurrence is often expressed in the form of association rules with a set of measures indicating the strength of the correlation. A *frequent subsequence* discovers patterns in sequential or time-stamped data where the subsequence refers to a list of item sets that tend to occur in a predictable order. In the CRM context, association rule mining and sequential mining can be applied to a market basket analysis to identify cross-selling opportunities better.

Change and deviation detection (Outlier analysis)

These types of methods discover “the most significant changes in the data from previously measured or normative values” (Fayyad, et al., 1996). Change detection deals with both monitoring the trends in a dataset over time or the changes of individual objects. As discussed in Chapter Two, change detection can be used to identify the dynamics in a customer base, namely changes in its segments’ structures over time and changes in cluster membership of individual customers. The techniques for detecting the dynamics in a customer base will be discussed in Sections 4.3 and 4.5.

3.4 Summary

In this chapter, the nine-step KDD process has been illustrated. Different kinds of data mining tasks have been discussed, providing a general description and motivation of the appropriate data mining algorithms for extracting patterns from data. In the next chapter, the specific data mining methods used in the thesis will be introduced and discussed.

4. Data mining methods

In this chapter, data mining methods used in this thesis for performing customer segmentation, temporal customer segmentation, and segment migration analysis are presented. The chapter starts with an introduction to visual data mining, which is an important topic of this thesis. Then, VDM techniques such as the SOM and its extensions (i.e., the SOTM and transition probability on the SOM) used for customer segmentation tasks are introduced. It is noted that the SOM and its extensions are introduced in the same section for their technical similarity, even though they are used for different analytical tasks. After that, the classification models used as part of the static customer segmentation model are introduced. The algorithm and measures for performance evaluation for each of data mining methods are documented. Motivations for choosing a particular data mining method will also be discussed.

4.1 Visual data mining

Data have been accumulated at an unprecedented rate, as is noted earlier, resulting in large datasets with a high dimensionality. As discussed in Chapter Three, data mining methods can be used to derive information and insights from massive and dynamic data. It is shown in Section 3.2 that the interactive and iterative KDD process requires the combination of subject-matter knowledge and computational efficiency of data mining algorithms. For this reason, visualization is considered an important tool to facilitate this interaction. The results of a data mining model can be expressed in many forms, and it is commonly held that the visual representation of information has advantages in conveying information to humans compared to the use of tabular or textual reports. Information visualization is a field that uses interactive visual representation of data to amplify cognition (Card, et al., 1999, p. 7). The functions provided by information visualization techniques can be classified into three categories: 1) data exploration: users search for structure, patterns, and trends in data in an attempt to generate relevant hypothesis; 2) hypothesis confirmation: users use visualization techniques to verify or refute a hypothesis; and 3) visualization production: based on the confirmed hypothesis, users produce optimized visual representations. In the context of data mining, information visualization techniques play an important role in visual data exploration and representation of the mined results (Ferreira de Oliveira & Levkowitz, 2003). More recently, the term visual analytics has been introduced to refer to the process of analytical reasoning coupled with interactive visual interfaces (Thomas & Cook, 2005). Visual Analytics methods allow decision makers to combine their expertise and domain knowledge with information embedded in the data to gain insights into complex problems (Keim, et al., 2010). The use of visualization tools can reduce representational complexity and highlight differences and key patterns, which makes it an effective managerial practice (Eppler and Platts, 2009).

One category of visualization techniques capable of visualizing high-dimensional datasets is based on projection techniques, for example, the SOM, principal component analysis (PCA), and Sammon's mapping. The high-dimensional data is mapped onto a lower-dimensional space, usually of two dimensions. The projections on the low-dimensional display enable the exploration of similarities and dissimilarities among data items. In addition, these projection methods can also be used to reduce dimensionality (Kohonen, 2001, p. 34). The dimensionality of the original variables is reduced into a smaller number of linear or nonlinear combinations of original dimensions.

Marghescu (2007) investigated the effectiveness of a number of projection techniques in preserving the structure of original data. The author compared the effectiveness of the projection methods using the Hubert's statistic and its modified version. A strong correlation between the proximity matrices of the original data and the projected data was found, and that the PCA, Sammon's mapping, and the SOM were the most effective projection methods in terms of preserving the distance relationship of the original data. The studies included in this thesis mainly use the SOM and its extensions for customer segmentation. In the following subsections, the motivation of these methods chosen and their working properties will be discussed.

4.2 The SOM

Clustering algorithms have been widely used to approach customer segmentation tasks. The SOM is an unsupervised artificial neural network (ANN) based method for visual clustering. It belongs to a class of projection methods that enables data and dimensionality reduction. The SOM provides nonlinear, ordered, and smooth mapping capabilities for visualizing multivariate data on a lower dimensional display, thereby facilitating the exploration of useful patterns that would be otherwise difficult to identify in the data. The usefulness of the SOM for exploring and clustering financial and business data has been demonstrated in a number of studies covering different areas. For example, the SOM has been used for financial benchmarking (Serrano-Cinca, 1996; Back, et al., 1998; Eklund, et al., 2003, 2008), prediction and exploration of currency crises (Sarlin, 2011; Sarlin & Marghescu, 2011), customer portfolio analysis (Holmbom, et al., 2011), and knowledge management (Segev & Kantola, 2012). The effectiveness of the SOM in customer segmentation over the other methods is compared in a number of studies as listed in Table 5.

A typical SOM consists of two layers: an *input* and an *output layer*, as illustrated in Figure 9. Each input field is connected to the output layer by one node, which is connected with all the nodes in the output layer. The lattice of the grid of the output layer can be either rectangular or hexagonal, in which each node is connected to four or six neighboring nodes, respectively. The hexagonal topology is preferred for its visualization advantage (Vesanto, 1999).

In this thesis, in line with Sarlin (2013a) the motivations for using SOM for customer segmentation are: 1) its simultaneous clustering and projection capabilities, 2) the pre-defined grid structure for linking visualizations, 3) computational efficiency, and 4) flexibility for missing data, which are discussed as follows.

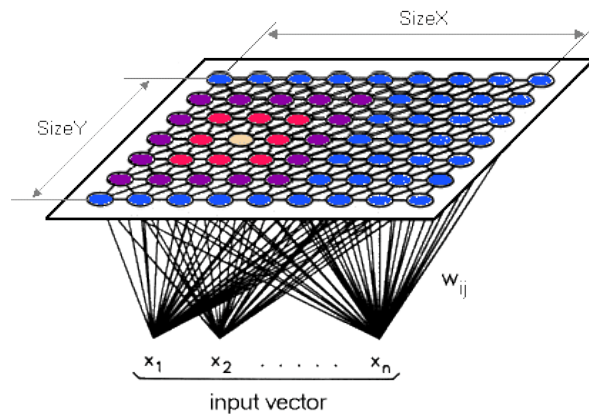


Figure 9. The structure of a typical SOM (Source: Demirhan and Güler (2011)).

Firstly, one of the most important features of the SOM is its ability to preserve neighborhood topology in the projection, which enables simultaneous clustering and projection. As illustrated in Figure 9, the input data are projected to the output layer of the map, which is usually a predefined grid structure. When the map training is complete, each node on the grid is at the center of the data attracted to it and of all its surrounding nodes. This training mechanism ensures that as far as possible, similar input data are mapped onto neighboring nodes on the grid. Venna and Kaski (2001) compared the neighborhood preservation capability of five data projection methods on three datasets. They found that the SOM performs better than traditional multi-dimensional scaling (MDS) based methods, Sammon's mapping, and PCA. More important, unlike PCA, MDS, and Sammon's mapping that aim at preserving local spatial distances in the input data, topology-preserving methods (e.g., the SOM) are more suitable for clustering purposes, where the trustworthiness of the resulting neighborhood and global distance are considered more important. Vesanto and Alhoniemi (2000) also argued that preserving neighborhood relations in the mapping aid in identifying cluster structures that are hidden in the high-dimensional data. Not surprisingly, the SOM's capability of simultaneously conducting clustering tasks and providing a multivariate visual display of the clustering results make it a widely used method for visual clustering.

Secondly, as mentioned in the beginning of this section, one important property of the SOM is its visualization capabilities. In addition to its advantages as a nonlinear projection method introduced above, the predefined grid structure offers the SOM a possibility for linking additional information to the SOM

model for exploratory data analysis, such as component planes (also known as feature planes), U-matrices, density plots, and error plots (Vesanto, 1999). In this thesis, the grid structure is used as a basis for visualizing customers' migration patterns among segments. More details of this approach can be found in Section 4.5.

Table 5. Comparison of related work of using the SOM for segmentation tasks.

Work	Clustering Algorithms	Application Area	Conclusion
Kiang and Kumar (2001)	SOM vs. Factor analysis	Segmentation task	SOM outperforms traditional factor analysis in general and provides more accurate clustering results when the input data are skewed.
Curry, Davies, Phillips, Evans, and Moutinho (2001)	SOM vs. K-means	Strategic groups identification	SOM performs better than K-means. SOM concentrates on similarities and gradual changes in the data. On the contrary, K-means tend to concentrate more on data differences and separating outliers.
Shin and Sohn (2004)	SOM vs. K-means vs. Fuzzy C-means (FCM)	Customer segmentation for stock trading market	Fuzzy K-means cluster analysis is generally more robust and stable than SOM and K-means. SOM performs best in some situations.
Kuo, An, Wang, and Chung (2006)	SOM+genetic K-means vs. K-means vs. SOM+K-means	Market segmentation for freight transport industry	The SOM is more effective in determining the cluster number than K-means. The SOM used in conjunction with genetic K-means outperforms the other two methods.
Budayan, Dikmen, and Birgonul (2009)	SOM vs. FCM vs. K-means vs. Hierarchical clustering	Strategic groups identification	The SOM and FCM can reveal the typology of the strategic groups better than traditional cluster analysis.

Thirdly, the SOM is a relatively computationally efficient method. The computational cost of training a SOM is related to the grid size (Kaski, 1997). This implies that when the number of nodes on the map is much lower than that of the input data, for example, in customer segmentation, the computational complexity of a SOM will be reduced significantly, making the SOM a flexible and scalable method for clustering large-scale data. To improve the computational efficiency further, the batch SOM algorithm is used for our customer segmentation. The sequential training algorithm has been found to be more computationally expensive (Kohonen, 2001, p. 197). Both training algorithms will be introduced in the following section.

Fourthly, in addition to its computational advantage, the SOM is found to be robust to data with missing values (Kohonen, 2001, p. 165), a phenomenon not uncommon in customer data. Unlike traditional projection techniques (e.g., PCA) and distance-based clustering techniques (e.g., K-Means) that completely ignore cases with missing values, the SOM considers only available data in finding best matching units (BMUs) and thus utilizes the available information in training as much as possible.

4.2.1 Training the SOM

There are typically two training algorithms of the SOM: *sequential* and *batch*. In sequential training, input data are presented to the map one by one, and the SOM is updated immediately after every input data. In batch training, all input data are presented to the map together and the map is updated once. In our papers, the batch training was chosen over the sequential one for the following reasons: 1) batch training is much more computationally efficient than the sequential training since it only updates the map once; and 2) the map can be reproduced given the same initialization settings, since all the data are presented to the map simultaneously, thus making maps trained with different combinations of parameters comparable.

The functioning of the batch SOM can be split into two steps: 1) matching data records to their BMUs, and 2) updating each unit towards the attracted data, including those in adjacent locations.

Step 1

The matching is a competitive learning process. All input data x_j (where $j = 1, 2, \dots, N$) are compared with the map's reference vectors m_i (where $i = 1, 2, \dots, M$), and the ones (m_b) whose reference vectors are the closest to the input sample win as the corresponding BMUs. The Euclidean distance can be used to measure the similarity. This process can be formulated as:

$$\|x_j - m_b\| = \min_i \|x_j - m_i\| \quad (1)$$

Step 2

Then, the reference vectors are updated with following batch algorithm:

$$m_i(t+1) = \frac{\sum_{j=1}^N h_{ib(j)}(t)x_j}{\sum_{j=1}^N h_{ib(j)}(t)}, \quad (2)$$

where index j represents the input data belonging to node c , and the neighborhood function is defined as:

$$h_{ic(j)} = \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right), \quad (3)$$

where t is a discrete time coordinate, and $h_{ic(j)}$ a decreasing function of neighborhood radii and time. It is noted that as opposed to sequential training, users do not have to specify the learning rate explicitly, although the time-dependent neighborhood function has to be defined.

4.2.2 Evaluation of the SOM

As previously mentioned, the SOM is capable of conducting simultaneous clustering and projection, and this implies that the SOM has the properties of both vector quantization and projection algorithms. Therefore, most of the quality measures of the SOM concern two aspects of the map: *vector quantization* and *topology preservation*. There is a tradeoff between these two, increasing the quality of vector quantization usually leads to distortion of the map's topology (Pözlbauer, 2004). Commonly used quality measures are *quantization error* (QE), *distortion error* (DE), and *topographic error* (TE). The first two measures aim at evaluating the quality of vector quantization, whereas the third one aims at topology preservation. These three quality measures are introduced as follows.

Quantization error (QE)

QE measures the resolution of the map, which equals the average distance between each input vector and its nearest weight vector. It should be noted that the quantization error decreases monotonically for increasing map sizes. The quantization error can be computed by using the following expression:

$$\varepsilon_{qe} = \frac{1}{N} \sum_{j=1}^N \|x_j - m_{c(j)}\|^2 \quad (4)$$

Distortion error (DE)

DE is very similar to the quantization error except that the distortion error explicitly takes the neighborhood into account. It can be interpreted as an error value for the whole map. According to Vesanto, et al. (2003), the distortion error can be decomposed into three components: quantization error, neighborhood bias, and neighborhood variance. DE can be computed as follows:

$$\varepsilon_{de} = \frac{1}{N} \frac{1}{M} \sum_{j=1}^N \sum_{i=1}^M h_{ic(j)} \|x_j - m_i\|^2 \quad (5)$$

Topographic error (TE)

TE measures the continuity of the mapping or the topology preservation (Kiviluoto, 1996), which is defined as the percentage of all sample vectors for which the best and the second-best matching units are not adjacent.

$$\varepsilon_t = \frac{1}{N} \sum_{k=1}^N u(x_k), \text{ where } u(x_k) \begin{cases} 1, & \text{1st and 2nd BMUs not adjacent} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

4.2.3 Visualizing the SOM

In order for the information from the SOM to be extracted, the map needs to be visualized. Since it is difficult to fit all information into one display, multiple visualizations are commonly used, for example, component planes (Vesanto, 1999). A component plane is constructed for each individual input variable, and it is of the same size as the predefined SOM grid. Different component planes are linked by position. Each node in the same position across the component planes represents same data. A set of component planes, therefore, provides multiple linked views of the same data from the perspectives of different variables. Examples of component planes are given in Figure 10, in which the heat map colors are used to illustrate low (blue colors) to high values (red colors). For instance, Figure 10 displays three component planes, and the first one (c.f. Fig 10(a)) shows high values (elderly customers) for the lower right part of the SOM grid (i.e., segment S1) and low values (young customers) for the mid-right part (i.e., segment S4). Figure 10(b) shows that long-standing customers are in segments S1 and S5, while customers with short tenure are in segment S4. Figure 10(c) shows that male customers are located on the lower-right part of the map.

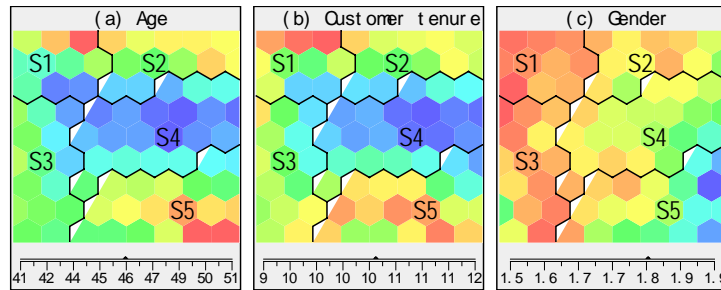


Figure 10. An example of component plane visualization of the SOM model (Source: Yao, et al. (2012a)).

4.3 The SOTM

As discussed in Section 4.2, the SOM is an effective tool for conducting clustering in general and customer segmentation in the context of analytical CRM. However, the SOM provides only a static snapshot of segment structures that can change over time. For this reason, the temporal dimension should be explicitly taken into account in customer segmentation.

Following the discussion in Sarlin (2013b), there are several ways of involving temporal processing in the SOM:

- 1) Embedding time through pre-processing of data.
- 2) Visualizing time on a SOM through post-processing.
- 3) Modification of the activation and learning rule.
- 4) Modification of the network topology.
- 5) Combinatorial approaches of standard SOMs and other visualization techniques.
- 6) Comparing standard SOMs at different points in time.

The first type of methods deal with time by preprocessing time sequence data before training. For example, Chappelier and Grumbach (1996) used a tapped-delay of input data as the input vector fed into the SOM. Although this approach provides a convenient way for introducing time into the SOM and is suitable for tasks such as time series clustering, they cannot be used to cluster entities at different points in time and to compare the cluster structures and profiles across time. Moreover, it is difficult to determine the optimal length of the delay.

The second type of methods first train the SOM without embedding temporal information during the learning process, and the trajectories (Kohonen, 1988) are then used to illustrate the state transition patterns. However, trajectories cannot be used on a large set of data in order not to clutter the display.

The third type of methods process the data in a supervised manner. For example, the Hypermap structure (Kohonen, 1991) decomposes the input vector into a past and a future vector. The past vector is used for matching, and the future part is added when adapting the weights of both parts. These types of methods can be used in time series prediction.

The fourth type of methods deals with time by modifying the network topology by introducing feedback connections. The feedback SOMs (e.g., Temporal Kohonen Map (TKM) (Chappell & Taylor, 1993) and the recurrent SOM (RSOM) (Varsta, et al., 1997)) works by keeping past output values in training. The preservation of past outputs serves as a kind of short-term memory allowing for learning temporal information.

The fifth type of methods combines SOMs and various visualization techniques that have been implemented with the aim of better spatiotemporal visualization (e.g., Guo, et al. (2006)). Yet, they cannot show the temporal dimensions and cross-sectional structures simultaneously.

The last type of methods compare SOMs at different points in time or visualize the changes separately. Back, et al. (1998) applied the SOM for benchmarking the financial performance of pulp and paper companies during 1985–1989. The authors monitored the performance of Finnish companies by comparing several SOM models side by side. However, this task to some extent is hindered by the

inconsistent cluster locations due to the random initialization of reference vectors and random sequence of input data to the network, as was constrained by the functionality of the tool used. Denny and Squire (2005) proposed a SOM-based visualization technique capable of representing the clustering of the one dataset in terms of that of the other with cluster color linking. The orientations of the two SOMs are preserved by using the reference vectors of the first SOM to initialize the second and employing the batch training algorithm, thus facilitating comparisons of two maps. However, this approach is limited by the ability to compare SOMs based on only two points in time. For a more detailed discussion of the methods that involve temporal processing in the SOM, see Sarlin (2013b).

As discussed in Section 2.4.1, evolutionary clustering is a class of problems that produce a sequence of clusterings for each time point of temporal data. An effective evolutionary clustering should achieve a balance between clustering results being representative of each point in time and not being too deviant from the previous clustering result. The methods as discussed above cannot fulfill the requirement of evolutionary clustering and they lack of the capability of objectively representing temporal changes in the cluster structures. The SOTM (Sarlin, 2013b) is a recent adaption of the standard SOM for exploratory temporal structure analysis, i.e., to simultaneously visualize cluster structures, their temporal evolution, and multivariate data. The SOTM is essentially a series of one-dimensional SOMs ordered in consequent time units and represents both time and data topology on a two-dimensional grid, providing a visual means for a abstracted display of temporal multivariate patterns. In contrast to traditional clustering techniques, the SOTM enables exploring complex patterns over time by visualizing the results in a user-friendly way. The special initializations between the neighboring one-dimensional SOMs enables them being comparable. In the following subsections, the architecture and the algorithm of the SOTM are introduced, followed by its evaluation measures and the visualization methods.

4.3.1 Training the SOTM

The training of the SOTM involves a series of training of multiple one-dimensional SOMs at each point in time. Figure 11 presents the functioning of the SOTM, which is detailed as follows.

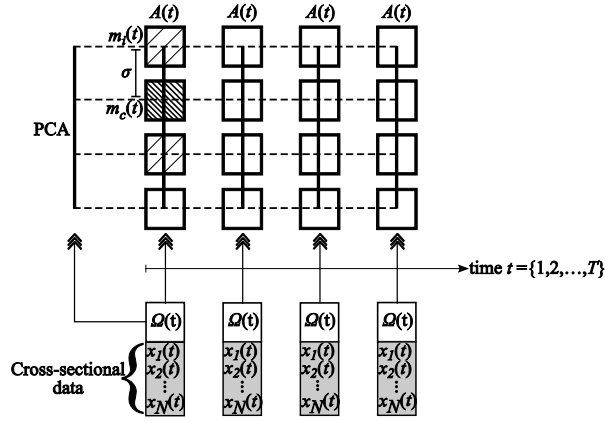


Figure 11. The functioning principles of the SOTM (Source: Sarlin (2013b)).

Step 1

- When $t = 1$, $A(t)$ (i.e., the one-dimensional SOM is composed of nodes $m_i(t)$ at time t) is initialized using PCA on the input data $\Omega(t)$.
- Similarly as the SOM, the SOTM follows the two-step procedure. The batch algorithm documented in Section 4.2.1 is used where the matching is performed by:

$$\|x(t) - m_c(t)\| = \min_i \|x(t) - m_i(t)\| \quad (7)$$

and $A(t)$ is updated using $\Omega(t)$. This batch update process can be formulated as follows:

$$m_i(t) = \frac{\sum_{j=1}^{N(t)} h_{ic(j)}(t) x_j(t)}{\sum_{j=1}^{N(t)} h_{ib(j)}(t)} \quad (8)$$

$$h_{ic(j)}(t) = \exp\left(-\frac{\|r_c(t) - r_i(t)\|^2}{2\sigma^2}\right) \quad (9)$$

where the neighborhood $h_{ic(j)}$ is restricted to vertical relations.

Step 2

- When $t \geq 1$, $A(t)$ is initialized by using the reference vectors of $A(t-1)$.
- Use the batch algorithm documented to update $A(t)$ using $\Omega(t)$. See formulas (7-9).

Step 3

- Order $A(t)$ in ascending order to time t .

After the training is complete, the SOTM composed of a series of one-dimensional SOMs on data ordered in consequent time units represents time and data topology on a two-dimensional grid. For more details of the training algorithm of the SOTM, readers are referred to Sarlin (2013b).

4.3.2 Evaluation of the SOTM

The SOTM can be evaluated as the standard SOM paradigm. Essentially, the quantization error, distortion measure, and topographic error for the SOTM can be calculated as the average of those of the one-dimensional SOMs that form the SOTM (Sarlin, 2013b):

$$\varepsilon_{qe} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \sum_{j=1}^{N(t)} \|x_j(t) - m_{c(j)}(t)\|^2 \quad (10)$$

$$\varepsilon_{dm} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \frac{1}{M(t)} \sum_{j=1}^{N(t)} \sum_{i=1}^{M(t)} h_{ic(j)}(t) \|x(t)_j - m(t)_i\|^2 \quad (11)$$

$$\varepsilon_{te} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \sum_{j=1}^{N(t)} u(x_j(t)), u(x_j(t)) = \begin{cases} 1, & \text{1st and 2nd BMUs not adjacent} \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

4.3.3 Visualizing the SOTM

Similar to the SOM, the multidimensionality of the SOTM is visualized by a set of component planes. Essentially, these component planes show information across clusters (represented with the vertical dimension), time (represented with the horizontal dimension), and multiple variables (represented by the component planes). This provides a holistic view of how the characteristics of clusters, as well as their distribution, evolve over time.

Sammon's mapping (Sammon Jr, 1969), a multidimensional scaling technique that provides a non-linear mapping of data, can also be used for assessing the structural properties of SOTM visually. It tries to preserve the inter-node distances of the SOTM with their original distances in the high-dimensional space, enabling exploration of the changes in cluster structures over time. In a Sammon's mapping of a SOTM, the SOTM nodes are plotted into one dimension first and then rearranged according to Sammon's dimension on the y axis and time on the x axis. The relationships of the horizontally neighboring nodes are shown by connecting lines to have a net-like representation.

A Sammon's mapping of a SOTM can also be visualized using a coloring method by Kaski, et al. (2000), where the well-known uniform color space CIELab (CIELab, 1986) is used so that the perceptual differences of the colors represent the distances in the data space approximated by Sammon's mapping.

Readers are referred to Sarlin (2013b) for more details of visualization of the SOTM.

4.4 Clustering of the SOM and the SOTM

As was mentioned in Section 3.3, clustering is a class of techniques that partition data into groups attempting to minimize intra-cluster distance and maximize inter-cluster distance. The SOM-based clustering methods can be generally divided into two categories: *direct clustering* and *two-level clustering* (Denny & Squire, 2005).

In direct clustering, the individual nodes of a SOM can be treated as separate clusters. Maps with a large number of nodes are always preferred in visualization for their better resolution. However, it is always expected that the number of segments is in a reasonable range (e.g., between five to ten (McDonald & Dunbar, 2004, p. 49)) for a customer segmentation model to be manageable and actionable. Kiang, et al. (2006) applied this approach for customer segmentation. One disadvantage of this approach is that the maximum number of clusters should be predefined.

In two-level clustering, instead, a larger number of units can be further grouped into second-level clusters (Vesanto & Alhoniemi, 2000). The dataset is first projected onto a two-dimensional display using the SOM. Then the resulting SOM is clustered. Since the SOM has reduced the original data, the general problems of computational cost and uncertainty of the clustering result caused by outliers and noisy data are decreased. In the following sub-sections, the motivations for applying second-level clustering on the SOM and SOTM, the agglomerative hierarchical clustering techniques used for the second-level clustering, and the quality measures for evaluating the goodness of the clustering solutions, will be discussed.

4.4.1 Second-level clustering on the SOM and SOTM

Previous studies have shown the two-level SOM to be effective. Vesanto and Alhoniemi (2000) showed on three datasets that the two-level SOM worked equally well as clustering with hierarchical clustering methods and K-means, while being computationally more efficient and capable of visualizing relations in data. Li (2005) shows the superiority of the two-level SOM over some classical clustering algorithms. Lee, et al. (2004), adopting the two-level SOM (using SOM and K-means clustering), conducted a market segmentation of the Asian online game market and find that the two-level SOM is more accurate than K-means clustering or the SOM alone.

Similarly, the motivations of using second-level clustering on the basic SOM are applicable to the SOTM as well (Sarlin & Yao, 2013). While visualization aids may facilitate interpreting the structures of the SOTM, for example, component planes and cluster coloring, a large number of dimensions and units in a SOTM complicate our ability to perceive structures in data, not to mention large

temporal changes. With stationary data, rows of the SOTM would represent similar data at different points in time. However, when the data are nonstationary, horizontally neighboring nodes may represent data of different characteristics. It is, therefore, difficult to judge whether or not nodes in a single row are members of a particular cluster, especially when changes in cluster structures are substantial. For example, a new cluster can be formed by merging existing clusters, and an existing cluster can be split into two sub-clusters. Thus, a second-level clustering can be used for identifying changing, emerging and disappearing clusters on the SOTM. The two-level clustering of the SOTM is conceptually similar to the two-level SOMs and can be used for a rough labeling of units and for identification of changes in cluster structures over time.

In Sarlin and Yao (2013), we defined three types of dynamics in cluster structures as follows: 1) a cluster disappears when its members in time t do not exist in $t + 1$; 2) a cluster emerges when its members exist in $t + 1$ but not in t ; and 3) a cluster changes when the number of its members in time t and $t + 1$ differ.

4.4.2 Agglomerative hierarchical clustering

Agglomerative hierarchical clustering starts by treating each unit of the SOM or SOTM as a separate cluster and merges iteratively clusters with the shortest distance until all units are merged. There are different criteria for measuring the cluster-wise distance, for example, single-linkage, complete-linkage, average linkage, or Ward's method (or minimum variance linkage). In single-linkage, the distance between two clusters is determined by the distance of the closest pair of observations in the two clusters; while in complete-linkage, the distance is based upon the distance of the most distant pair. Unlike single-linkage and complete-linkage that form clusters based on single observation pairs, average-linkage focuses on the inter-cluster distance by averaging the distances of all pairs of the two clusters. While most hierarchical clustering methods use distance as the criterion for merging clusters, Ward's clustering attempts to minimize the variance of the merged cluster. As hierarchical clustering methods are well-known, readers interested in details are referred to Rousseeuw and Kaufman (1990) for further details.

Clusters of single-linkage tend to take the form of long chains and other irregular shapes with little homogeneity, whereas complete-linkage clusters have been shown to be inefficient in separating data (Blashfield, 1976; Hansen & Delattre, 1978). In the seminal literature, Ward's method has already shown to be accurate by outperforming other hierarchical methods (Kuiper & Fisher, 1975; Blashfield, 1976; Mojena, 1977).

In this thesis, agglomerative hierarchical clustering is used to group the SOM and SOTM units. The key motivation for using hierarchical clustering is that it enables us to circumvent a definite choice of number of clusters. In addition, hierarchical clustering provides the means to explore the clustering results with

varying numbers of clusters. For the static customer segmentation task, the SOM-Ward algorithm, taking the ordering of the SOM into account, is used to segment customers. That is, Ward's clustering is limited to merge only adjacent SOM nodes. This modified Ward distance is defined as follows:

$$d_{ij} = \begin{cases} \frac{n_i n_j}{n_i + n_j} * \|c_i - c_j\| & \text{if } i \text{ and } j \text{ are adjacent} \\ \infty & \text{otherwise} \end{cases}, \quad (13)$$

where i and j represent clusters, n_k and n_l the size (i.e., the number of data items) of clusters i and j , and c_i and c_j the squared Euclidean distance between the centroids of clusters i and j .

For temporal customer segmentation with SOTM, the agglomeration process is not limited to only adjacent nodes. Instead, different distance criteria such as single, complete, average-linkage, and Ward's hierarchical clustering are experimented with and the one with the best evaluation measures is selected.

4.4.3 Evaluation of the two-level clustering

There are generally two types of cluster validity measures for evaluating the clustering results and determining the number of clusters: *external* and *internal* measures (for a discussion see, e.g., Theodoridis and Koutroumbas (2008, pp. 866-873)). In the context of CRM, these measures are used to evaluate the degree of homogeneity of each segment and the overall quality of the segmentation model. The external measures (e.g., Rand index (Rand, 1971) and Hubert's statistic (Hubert & Schultz, 1976)) evaluate a clustering model with reference to some external *a priori* information, e.g., given class labels, while the internal measures evaluate (e.g., gap statistic (Tibshirani, et al., 2001), Dunn index (Dunn, 1973) and Silhouette index (Rousseeuw, 1987)) a clustering model in terms of the internal relationships among the data items. The gap statistic evaluates a clustering model based upon the within-cluster dispersion, while the Dunn and the Silhouette indices take into account both cluster compactness and cluster separation.

In our studies, since the structure of the data is not known *a priori*, the clustering validity is evaluated only using internal validity measures, namely the Dunn index and the Silhouette coefficient. Both measures take into account cluster compactness and cluster separation. For both measures, the higher the value is, the better the observations are clustered.

The Dunn index is defined as the ratio of the smallest inter-cluster distance to the largest intra-cluster distance. The Dunn index is computed as:

$$D_K = \min_{1 \leq l \leq K} \left\{ \min_{\substack{1 \leq k \leq K \\ l \neq k}} \left\{ \frac{d(C_l, C_k)}{\max_{1 \leq h \leq K} \{d(C_h)\}} \right\} \right\} \quad (14)$$

where K is the number of clusters, $d(C_l, C_k)$ is the distance between clusters l and k (inter-cluster distance), and $d'(C_h)$ is the maximum distance between observations in cluster h (intra-cluster distance).

For each observation i , its Silhouette coefficient is defined as:

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (15)$$

where a_i is the average distance between i and all other observations in the same cluster, and b_i is the average distance between i and the observations in its nearest cluster. The Silhouette coefficient for a clustering model is simply the average of the Silhouette coefficients of all observations.

4.5 Transition probability on the SOM

While the SOM and the SOTM are ideal tools for clustering and identifying temporal changes in cluster structures respectively, identifying temporal movements in a SOM model under the assumption that the cluster structure stays static over time is not a simple process. Previously, trajectories (Kohonen, 1988) have been a visual means to illustrate changes in cluster membership of individual data records on the SOM grid (Martin-del-Brio & Serrano-Cinca, 1993; Eklund, et al., 2003; Sarlin, 2010). The trajectories, however, can only be applied to a limited set of data in order not to clutter the display. Moreover, as the example shown in Figure 7 provides overall patterns and their strengths of segment migration patterns, similar information cannot be effectively quantified and visualized with trajectories.

Afolabi and Olude (2007) proposed a hybrid approach combining the SOM and a multilayer perceptron (MLP) for forecasting stock prices. Hsu, et al. (2009) also proposed a similar approach for the same task. Although these hybrid approaches facilitate time-series prediction, they involve high computational complexity, and the results cannot be visualized and related to the first stage where data are clustered by the SOM.

Denny, et al. (2012) introduced a methodology and visualization methods for analysis of cluster migration patterns using the SOM. First, the data at different points in time are clustered using two-dimensional SOMs, where the clusters identified are transformed and represented in a set of one-dimensional SOMs in order not to clutter the representation of migration patterns. Then the migration patterns are represented using line thickness. The process of transforming two-dimensional SOMs into their corresponding one-dimensional counterparts, however, is challenging and may distort the original data topology.

As is introduced in Section 4.4, the SOM-based clustering methods can be either direct clustering or two-level clustering. In the first case, as individual nodes in a SOM grid can represent clusters, the temporal migration patterns in a SOM

model can be quantified with the node-to-node transition probabilities in the SOM (Sulkava & Hollmén, 2003; Luysaert, et al., 2004). In the latter case, node-to-cluster transition probabilities have been proposed to be computed on the two-level SOMs and visualized on their own SOM grids (Sarlin, et al., 2012).

The authors proposed a framework for computing, summarizing, and visualizing transition probabilities on the SOM. For each node, the framework calculates its probability of moving to an area on the map in a specified period. Then the transition probabilities of all the nodes on the map to the specified area will be visualized in a separate component plane, thereby linking the transition patterns and the SOM model visually in one view. This framework thus provides a means for visualizing the segment migration patterns on the SOM model. In the following subsections, the methods for computing and visualizing node-to-cluster transition probability are presented.

4.5.1 Computing the transition probability on the SOM

Transition probabilities on the SOM are computed for quantifying the strengths of movements on the two-dimensional SOM grid and can model transitions to a specified region, such as nodes or clusters. Given a SOM model, the strength of segment migrations can be quantified by computing the proportion of the data in a specified region of the map belonging to each segment in the following period. For example, the transition from node i (where $i= 1, 2, \dots, M$) to segment s (where $s = 1, 2, \dots, S$) one period ahead can be computed using $p_{is}(t + 1)$:

$$p_{is}(t + 1) = \frac{n_{is}(t+1)}{\sum_{s=1}^S n_{is}(t+1)} \quad (16)$$

where n_{is} is the number of data moving from node i to cluster s , and t is a time coordinate. That is, the transition probability from node i to cluster s equals the number of data switching from node i to cluster s divided by the sum of data moving from node i to every other cluster.

4.5.2 Visualizing the transition probabilities

As the transition probabilities are associated to each of the nodes of the SOM model, they can be linked to the SOM visualization via positions using component plane representation, which is interpreted in the same way as for the SOM and SOTM. In order to create the component plane representation, a new component plane of the same size as the SOM model is used to represent the transition probabilities of the nodes to a specified area on the SOM model. In this component plane representation, the color code of each node represents the probability of the data items in that node migrating to that particular area. A series of component planes can be created to visualize the migration probabilities of data items to different areas of the SOM model, and therefore, they are termed migration planes.

4.6 Classification techniques

As mentioned in Section 3.3 that in contrast to the SOM that adopts an unsupervised learning paradigm, a classification method is based upon supervised learning where the discrete-valued class labels of the input data are known in advance and they are used for training. The aim of classification is to predict future events by analyzing historical and current data.

Classification can be generally considered as a three-step process. In the first step, a classification model is trained based upon training data with labeled classes, aiming to search for relationships between explanatory variables and the target variable. There is a range of classification algorithms proposed from the fields of machine learning, pattern recognition, and statistics. Three classification algorithms – the DT, the ANN and the SVM – are selected for the follow-on analyses (i.e., classification of high- and low-spending customers) to the static customer segmentation. Since these classification algorithms are well-known and widely used in various data mining applications, they are only briefly introduced in the following sub-sections. In the second step, the model's accuracy will be evaluated. Several evaluation measures are introduced in Section 4.6.4. In the third step, once created and validated, the model could be used for a new dataset that shares the explanatory variables, to predict or classify data for which the class label of the target variable is unknown. In Section 4.6.5, ensemble methods that combine predictions of several models with the aim of improving the overall prediction accuracy will be introduced.

4.6.1 The decision tree (DT)

A DT is a non-parametric, hypothesis-free method for classification. A decision tree (DT) works by dividing data into smaller sets by recursively applying two-way and/or multi-way splits. Compared with other classification techniques, the DT has many advantages. First, as opposed to some data mining methods (e.g., ANN), the DT produces straightforward and intuitive rules for classification and prediction purposes. Second, the DT is computationally efficient and relatively insensitive to outliers and skewed data distributions. Thirdly, the DT is a significant data exploration tool that can be used to unveil the relationship between candidate explanatory and target variables. It can also be used to identify the important variables for predicting or classifying the target variable (Murthy, 1998). In addition, the DT model can be easily translated into a set of if-then rules that can be used to classify the records. In this sense, a DT is both prescriptive and predictive.

A DT follows a top-down approach to split data recursively into smaller mutually exclusive subsets. The learning process involves finding the best variables for the partition at each split. The algorithm selects the best variable according to certain purity measures (e.g., Gini index or information gain). Finally, the data in the terminal nodes tend to be homogeneous with regard to the target variable. This process is repeated until some user-specified criteria are met,

for example, the maximum tree depth, the minimum number of data in a node, or the minimum change of within-node purity. The DT model could include small branches reflecting sampling variation or data noise, which cannot be well generalized. Therefore, the DT should be pruned to remove such branches to improve the classification accuracy. After a DT is built and pruned, it can be applied to classify unseen data, where majority voting is employed to label the data in a node, that is, all data in a node are labeled with the most frequent class within that node.

DT algorithms mainly differ in three aspects: the forms of the splits (e.g., binary split or multiway split), the purity measure for splitting nodes (e.g., Gini index or information gain ratio), and the mechanism for pruning. In this thesis, two well-known DT algorithms are applied to classify high- and low-spending customers: classification and regression tree (CART) (Breiman, 1984) and C5.0 (Quinlan, 2004).

Classification and regression tree (CART)

The CART algorithm constructs a binary classification tree, that is, it applies recursive two-way partitioning to split nodes at each step. The Gini index is used to measure the improvement in purity at each split. The Gini index for a particular dataset D can be calculated as follows:

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2, \quad (17)$$

where p_i is the probability that a record in D belongs to class C_i and is estimated by its corresponding prior probability. The Gini index for a dataset given a two-way partition based upon a variable A is defined as follows:

$$Gini_A(D) = \frac{|D_1|}{|D|} * Gini(D_1) + \frac{|D_2|}{|D|} * Gini(D_2), \quad (18)$$

where D_1 and D_2 denote the subsets of D after the partition, and $|D_1|$ and $|D_2|$ represent the cardinality of the two subsets. Therefore, the gain in purity can be measured by:

$$\Delta Gini(A) = Gini(D) - Gini_A(D), \quad (19)$$

In each step, the variable with the largest gain in purity is selected. The pruning mechanism of the CART involves using a validation set and explicitly adding a penalty to the classification accuracy of the tree in favor of a simple and more generalizable model. The penalty consists of two parts: a complexity parameter ranging from 0 to 1, and the size of the DT. The product of the two parts determines the penalty to the classification accuracy. The sub-tree with the best performance on the validation set is selected as the best pruned tree.

C5.0

In addition to CART, the C5.0 (Quinlan, 2004) (modified and improved after the well-known ID3 and its successor C4.5 (Quinlan, 1993)) with adaptive boosting is used to classify high- and low-value customers. The splitting criteria of the

C5.0 is based upon a concept called information entropy, which is the expected information needed to classify a data record in a dataset D . The entropy of D can be computed as:

$$E(D) = -\sum_{i=1}^m p_i \log_2(p_i), \quad (20)$$

where p_i is the proportion of data belonging to class C_i in a dataset D . Similarly, the entropy of applying a partition based upon an attribute A on the dataset D can be computed as:

$$E_A(D) = -\sum_{j=1}^V \frac{|D_j|}{|D|} * E(D_j), \quad (21)$$

where $|D_j|$ and $|D|$ represent the cardinality of the j_{th} sub-node and that of the parent node, respectively. Therefore, the information gain ratio after applying the partition is determined by:

$$GainRatio(A) = \frac{E(D) - E_A(D)}{E_A(D)}, \quad (22)$$

The C5.0 boosting model works by constructing multiple DT models. The initial model is trained as usual, and in each of the following boosting rounds, all models are built in such a way that they focus on the records that the previous models incorrectly classified. Finally, the records are classified by combining the weighted classification results of individual models according to each model's performance into one overall classification. C5.0 prunes trees in a similar manner to that of the CART, that is, it is based upon the estimated accuracy rate accounting for the complexity of the tree. C5.0 does not, however, use a validation set to estimate the accuracy. Instead, it uses the training data directly and adjusts for the estimated accuracy with a penalty. Then the sub-tree with the highest adjusted accuracy is selected as the best model.

4.6.2 The artificial neural network (ANN)

As its name implies, an ANN is designed to mimic the architecture of the human brain in a simplified way, to process information, and to learn from examples. The Multilayer Perceptron (MLP), a widely used type of ANN, is chosen for our classification tasks in this thesis. As illustrated in Figure 12, a MLP typically consists of at least three layers: an input layer, one or more hidden layers, and an output layer. Each layer consists of a number of nodes. The nodes in the input layer correspond to the input fields and are connected with the nodes in their neighboring hidden layer. Each node in the output layer is connected with all the nodes in its nearest hidden layer, and associated with one output category.

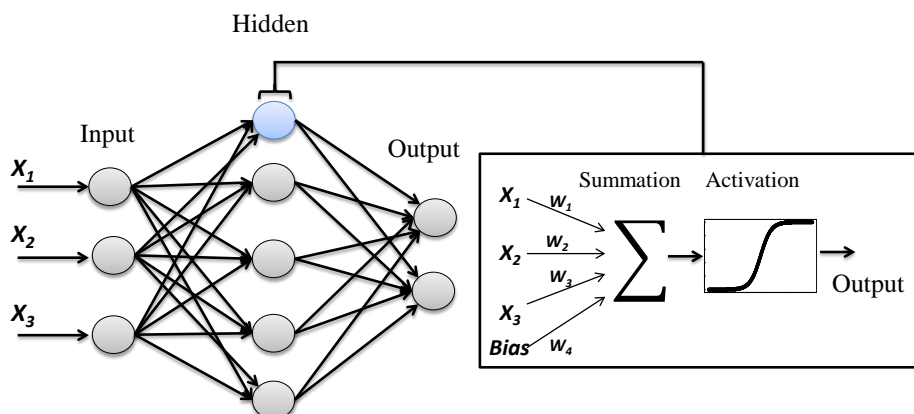


Figure 12. A multilayer perceptron with three layers.

The nodes are the basic units in a MLP. They can be fully or partially interconnected by a weighted connection. The value of the weighted connection defines the strength of the relationship between different nodes. The training of a MLP is based upon back propagation algorithm where the model adjusts the weight among network nodes by constantly reducing the prediction error by propagating the error backward through the whole network. This process of feeding training examples and updating weights can iterate many times until the mean squared error between predicted and actual target values is minimized. Thus, the network gradually becomes capable of understanding the relationship between inputs and outputs. The back propagation learning algorithm is well-known, and readers are referred to Hecht-Nielsen (1992) for more details of the algorithm.

4.6.3 The support vector machine (SVM)

The SVM (Vapnik, 1995) is a kernel-based method capable of conducting classification and regression tasks. The use of kernel transformation effectively overcomes the problem of the “curse of dimensionality”, which enables the application of the SVM to a wide range of datasets. Furthermore, the SVM is based on structural risk minimization (SRM), a principle for model selection based on a trade-off between model complexity and training error (Vapnik & Sterin, 1977). Therefore, it guarantees a global unique optimal solution and can reduce the risk of overfitting. All the advantages mentioned above have made the SVM an extensively applied technique in many industries (Guyon, 2008). In classification tasks, the SVM works by transforming input data using the kernel function into a high-dimensional feature space in which the classes of the data can be separated by a hyperplane. This hyperplane can be used to predict which category the new data belong to. In addition, on each side of the hyperplane, the SVM locates the maximum-margin hyperplanes: two parallel hyperplanes that maximize the distance between the data classes. The larger the distance between the maximum-margin hyperplanes is, the less likely the model will be prone to

overfitting. Readers are referred to Cristianini and Shawe-Taylor (2000) and Vapnik (1995) for technical details of the SVM.

4.6.4 Evaluation

The measures that have been used in the papers are introduced, which is followed by the discussion of different methods for conducting the evaluation.

A confusion matrix is a widely used method for measuring classification performance. As illustrated in Figure 13, a confusion matrix is typically represented by a table where each row represents the levels of an actual class, while each column represents the levels of a predicted class. For a classifier with a binary outcome, the cells in the matrix are commonly labeled as:

- True Positive (TP): Positive (i.e., the main class of interest) cases that are correctly classified.
- True Negative (TN): Negative cases that are correctly classified.
- False Positive (FP): Negative cases that are incorrectly classified as positive.
- False Negative (FN): Positive cases that are incorrectly classified as negative.

Therefore, the accuracy of a classifier can be calculated as follows:

$$\text{Classification accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \quad (23)$$

		Predicted Class		Total
		Yes	No	
Actual Class	Yes	TP	FN	P
	No	FP	TN	N
Total		P'	N'	

Figure 13. An example of confusion matrix.

In addition to accuracy, ROC curves (Swets, 1988) provide a visual means for comparing classifiers. As illustrated in Figure 14, a ROC curve measures the pairs of true positive rate and false positive rate across different cutoff values ranging from 0 to 1, whereas the diagonal line represents a useless model comparable to random guessing. The further the curve deviates from the diagonal line, the better the model is in terms its discriminative power. The area under the curve (AUC) can be used to quantify the performance of a classifier.

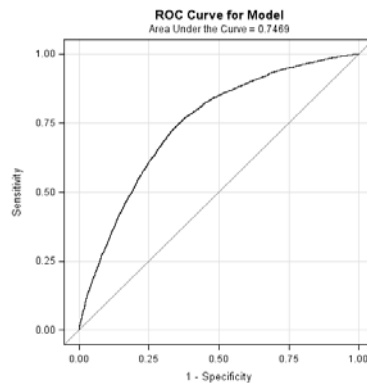


Figure 14. An example of ROC curve.

The performance measures, be it classification accuracy or AUC, can be incorporated into different evaluation approaches, such as hold-out evaluation and k -fold cross-validation. In the hold-out method, the data is partitioned into three mutually exclusive subsets: training set, validation set, and test set. The training set is used to build models. The model that has the best performance in the validation set is selected, and its accuracy is estimated with the test set. The hold-out evaluation cannot make full use of the available data and may suffer from sampling variation due to a lack of data. In k -fold cross-validation, however, the data D are partitioned into ten mutually exclusive subsets of equal size: D_1, D_2, \dots, D_k . The classifier is trained and tested k times. At each iteration i , the training data is $D \setminus D_i$ (i.e., all data excluding D_i), and the testing data is D_i . The accuracy is the ratio of the number of correctly classified data in the k iterations to the total number of data D (Kohavi, 1995).

4.6.5 Ensemble

An ensemble for classification is a composite model made up of a combination of classifiers. The individual classifiers vote, and a class label prediction is returned by the ensemble based on the collection of votes. Ensembles tend to be more accurate than their component classifiers, especially when there is significant diversity among the models (Han, et al., 2011, p. 377). In this thesis, after the performance of the classification models has been evaluated, they will be combined into an ensemble model for more accurate predictive analytics. Three ensemble methods can be used: *voting*, *confidence-weighted voting*, and *highest confidence win*. In voting, the number of times each possible target value appears is summed, and the one with the highest total is chosen as the prediction. Confidence-weighted voting works in a similar way as voting except the confidence of prediction is taken into account and the votes are weighted by the confidence. In highest confidence win, the prediction with the highest confidence is chosen as the prediction of the ensemble model.

4.7 Summary

In this chapter, the SOM and its extensions (i.e., the SOTM and the transition probability on the SOM) have been introduced for processing temporal data. The classification techniques used for predicting customers' development potential have also been discussed. In the next chapter, the data used in the studies and the data preparation process are documented.

5. Data

The raw data accumulated in operational systems are prone to various kinds of errors. They are often noisy, inconsistent, incomplete, or outdated data. These problematic data can prevent a data mining algorithm from using the selected data effectively. Therefore, the data should be preprocessed prior to mining to reduce the mining efforts and to enhance the stability and interpretability of the mined patterns. Data preparation is widely recognized as important and the most time-consuming process in the KDD process (Feelders, et al., 2000; Linoff & Berry, 2011, p. 83).

It has been claimed that data preparation is as much of an art as it is science and that data should take into account the data mining algorithm and the aim of the KDD project at hand (Pyle, 1999, p. 89). Crone, et al. (2006) compared the effects of different data preparation schemes involving variable scaling, sampling, coding of categorical variables, and binning of continuous variables on the performance of different classifiers in the context of direct marketing. The author showed that data preparation exerted more influence on the predictive accuracy of different classifiers than did training parameter choices. Coussement, et al. (2012) examined the influence of inaccurate data on the performance of RFM, DT, and logistic regression for customer segmentation tasks and find that data inaccuracy and its magnitude affected their performance. Arndt and Langbein (2002) discussed data quality issues in the context of customer segmentation and propose that three perspectives, – the marketing perspective, the data quality perspective, and the data mining perspective – should be taken into account when applying clustering techniques to customer segmentation. First, the input data should be chosen in such a way that the derived segments must be interpretable with marketing implications and be stable for a predefined time interval. Therefore, all customer-centric information should be collected and combined to provide a unified view regarding a customer base, and new variables should also be created to enrich the information of the mining file and to facilitate the interpretation of the formed segments. Data collection and integration will be discussed in Section 5.1. Second, since clustering algorithms have difficulties in dealing with problematic data to varying degrees, the issues, such as missing values and outliers, should be addressed prior to the clustering and are covered in Section 5.2. Third, the data preparation should take into account the properties of the clustering algorithm. For example, as most of the clustering algorithms are distance-based, input data should be converted into comparable ranges so that variables with large values will not dominate the results. Similarly, variables with skewed distributions should also be adjusted. These issues can be dealt with using the data transformation techniques introduced in Section 5.3.

5.1 Data collection and integration

Customer-centric data can be from a range of sources, such as billing systems, loyalty card systems, market surveys, or external data obtained from a third party. All these data should be consolidated to provide a single view of customers. On the other hand, the process of choosing data should be aligned with the purpose of their usage in a KDD project. The data used in the thesis are from a national department store chain that belongs to a large, multiservice corporation. They include information about 1.5 million customers for the period of 2007–09. Figure 15 illustrates these source datasets and their relationships.

Through the loyalty card program, demographics and measures of customers' loyalty within the multiservice corporation are collected in two tables: *customers* and *households*. Customers' purchase history is recorded in the file *transaction*, while the file *branches* documents the branches of the department store chain. The file *products* works as a look-up file that provides information about items sold at the department store. Basically, this file provides a definition of product category hierarchy that enables the analysis of customers' purchasing preferences at different levels of granularity.

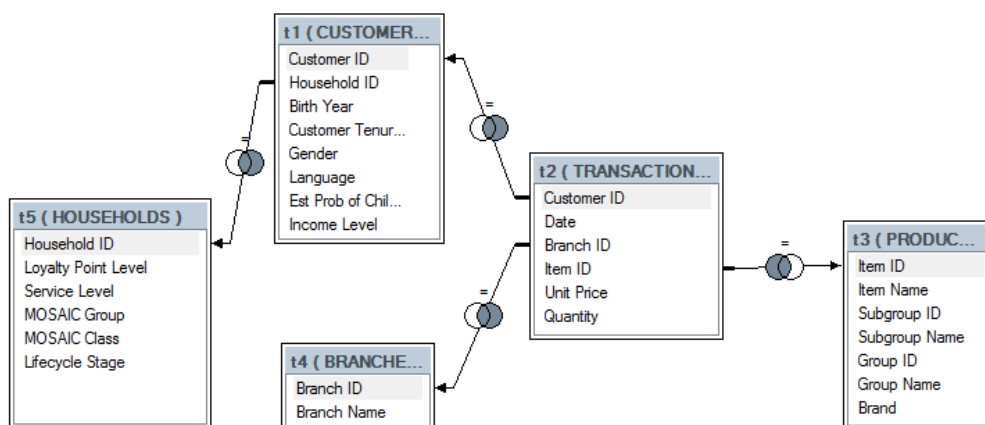


Figure 15. The source files containing customer-related information and their relationships.

As most data mining algorithms require a one-dimensional flat table for modeling, the customer data from multiple sources are integrated to provide a single customer view. In the context of analytical CRM, this one-dimensional flat table should merge all customer-related information to provide a customer signature or marketing customer information file (MCIF) (Tsipstis & Chorianopoulos, 2011, p. 148). In this thesis, the data integration process starts with selecting a time interval or a particular branch for subsequent analysis. Irrelevant variables should also be dropped early to improve data processing efficiency. Data from multiple sources are merged using key fields. The merged

data can be transposed to create multiple variables based on the values of another variable. For example, a series of dummy variables can be created for the variable group ID to indicate whether a customer has purchased the items from a particular product category. The previous step is commonly followed by aggregating data at a certain level, for example, customer or household. After that, new variables should be derived to enrich the content of the customer signature. For example, various ratios and KPIs can be derived to summarize customers' purchasing behavior. Finally, as illustrated in Figure 16, the MCIF or the customer signature file is created with one record for each customer with multiple fields describing customers' demographics, purchasing behavior, and their spending preferences in different product categories.

The Structure of the MCIF		
Descriptive Data	Purchasing Behavior	Product Preference
<ul style="list-style-type: none"> • Age • Gender • Estimated income level • Estimated probability of children • Customer tenure • Mosaic group • Mosaic class • Serive level • Loyalty point level 	<ul style="list-style-type: none"> • Total spending amount (Eur) • Basket size • Average item value (Eur) • Average transaction value (Eur) • Working time transaction (%) • Number of categories • Purchasing frequency 	<ul style="list-style-type: none"> • Recreation (%) • Men (%) • Women (%) • Footwear & suitcase (%) • Beauty (%) • Home (%) • Children (%) • Sports & Outdoors (%)

Figure 16. The variables in the MCIF.

These variables in the MCIF are briefly explained as follows:

- Age in years
- Gender (1 for male, 2 for female).
- Estimated income level: The higher the value, the wealthier the household is considered to be. Possible values are 1, 2, and 3.
- Estimated probability of children: Customers are divided into ten equal sized groups (deciles) according to their probabilities of having children at home. The higher the value of this variable is, the more likely there are children living in the household. The value ranges from 1 to 10.
- Customer tenure: The number of years the customer has been a cardholder.
- Mosaic group (Experian, 2008): The Mosaic group is a socio-economic ranking system that builds upon 250-by-250 meter map grid cells covering all the populated areas of Finland. Each map grid contains an average of seven households. The ranking system combines census data with marketing

research data to classify the whole population of Finland into nine groups: A, B, C, D, E, F, G, H, and I. Each map grid can be assigned to one of the nine groups. The households living in the same map grid can then be described in terms of socio-demographics, such as education, lifestyle, culture, and behavior.

- Mosaic class (Experian, 2008): Based upon the Mosaic group, the Mosaic class divides the nine Mosaic groups further into 33 subclasses.
- Service level: Measures how many service providers in the corporation (the case company is one service provider in the corporation) the customer has used in the last 12 months.
- Loyalty point level: Based on the average spending amount per customer in the corporation, this variable divides customers into five classes: from zero to four.
- Total spending amount: the spending amount of each customer under the analysis period.
- Basket size: Average number of items bought per transaction.
- Average item value: Average value per item purchased.
- Average transaction value: Average value per purchase transaction.
- Working time purchase: The percentage of purchases made during Mon–Fri, 9am–5pm.
- Number of categories: Average total number of distinct product groups in each transaction.
- Purchase frequency: Average number of transactions per day.
- Loyalty point level: Based on the average spending amount per customer in the corporate chain (the case company is one service provider in the corporate chain), this variable divides customers into five classes: zero, one, two, three, and four. A higher value in loyalty point level is an indication of a customer's larger spending amount in the entire corporate chain.
- Service level: Measures how many service providers in the corporate chain the customer has used in the last 12 months.

The variables in the *product preference* section measure the percentage of the spending amount of each customer in each product category of the analysis period.

5.2 Data cleansing

Data cleansing deals with missing values and noisy data. Missing values or incomplete data do not necessarily imply data error. Typical methods for handling missing values includes: 1) case deletion; 2) manual imputation; 3) replace the missing value with a global constant; 4) replace the missing value with a measure of location; 5) replace the missing value with a measure of location of the same class; and 6) use the most probable value to fill in the

missing value. In reality, the task of missing value imputation is task-dependent. The SOM can deal with missing data by using only the available information for matching (Kohonen, 2001, p. 165). For this reason, the SOM itself has been used for missing value imputation (Fessant & Midenet, 2002). The DT deals with missing values using the surrogate variables. Specifically, whenever the value is missing for the field that yields the best split, the next best splitting rule is applied. The DT is also used for imputing missing values by creating a model where the variable having missing data is treated as a target variable and the others are the explanatory variables. Then the missing value is imputed as the value predicted by the model. This method belongs to a class of methods called predictive value imputation (Saar-Tsechansky & Provost, 2007). The motivation for using a DT for imputing the missing data lies in its efficiency in dealing with large datasets and the capability for learning both numeric and categorical data.

Noisy data refers to errors in variables, outliers, and inconsistent data. Here, errors in data and data inconsistency are corrected by using a data dictionary, and outliers can be discovered using basic statistical summaries (e.g., z-score) and data visualization techniques (e.g., Histogram or boxplot).

5.3 Data transformation

Data transformation attempts to transform the data into forms appropriate for data mining. Here, the data transformation techniques are divided into two categories: *data smoothing* and *normalization*.

Data smoothing is used to reduce outlier bias. It is not uncommon in distance-based clustering that a small number of cases with extreme values form a small cluster. This small cluster in the customer segmentation context can be difficult to manage: one of the undesired properties of a segmentation model, as introduced in Section 2.3.3. Therefore, the data need to be transformed to equalize the distribution.

In this thesis, the logarithmic and sigmoid functions are used for transforming data. The logarithmic function (e.g., $y = \log_e^x$) tends to compress data when the value of x is larger than 1, while the transformation has the opposite effect when the value of x is below 1. With this property, when the distribution of a variable is heavily skewed to the right, the logarithmic transformation will constrain the large values and narrow the range of the right tail, producing a more symmetric distribution. This helps clustering algorithms better recognize the patterns in the data. The sigmoid transformation (i.e., $y = \frac{1}{1+e^{-x}}$) produces an “S” shape and transforms any value into the range between 0 and 1. This transformation stretches the data in the middle and compresses the data on the tails of the distribution, and thereby works better when there are significant amounts of outliers.

Normalization is where data are scaled into a predefined small range, so that variables with large values will not dominate. This ensures the pattern captured

is not biased. Two types of normalization techniques are used in the thesis, namely *min-max normalization* (Pyle, 1999, p. 251) and *z-score normalization* (Pyle, 1999, p. 344). In min-max normalization, a value, x_i , of variable A is normalized to x_i' by computing:

$$x_i' = \frac{x_i - \min(A)}{\max(A) - \min(A)} \quad (24)$$

where $\min(A)$ and $\max(A)$ are the minimum and maximum values of variable A , respectively.

The z-score normalization is computed as:

$$x_i' = \frac{x_i - \bar{A}}{\sigma(A)} \quad (25)$$

where \bar{A} and $\sigma(A)$ are the average and standard deviation of the variable A , respectively.

5.4 Summary

This chapter presents the data used in the thesis. The data collection and aggregation process have been documented. The variables in the MCIF have been interpreted. In the last part of this chapter, different data preparation techniques have been documented and explained. The papers included in this thesis will be presented in the following chapter.

6. Results

In this chapter, the data mining methods presented in Chapter Four are applied to address the business tasks discussed in Chapter Two. First, a static customer segmentation is conducted using the two-level SOM, which is followed by a subsequent analysis that identifies potential high-value customers using classification models. Then, the dynamics are addressed in a customer database from two perspectives. On the one hand, temporal customer segmentation is applied to identify the changes in segment structures over time. On the other hand, segment migration analysis is used to illustrate changes in segment membership of individual customers assuming that the segment structure is static over time.

6.1 Customer segmentation

This section presents our findings of the static customer segmentation model based upon Yao, et al. (2010b), *Publication 1*, and Yao, et al. (2010a), *Publication 2*.

As discussed in Section 2.3.2, *a priori* predictive methods can be used for customer segmentation, where a definition of *a priori* segments is first established, followed by the use of predictive models to describe the relationship between segment membership and a set of independent variables. However, this method is limited by the choice of the *a priori* segmentation variable in the first stage of the process (Wedel & Kamakura, 2000, p. 22). In *Publication 1* and *Publication 2*, we employ a two-stage approach that combines *post-hoc descriptive* methods (e.g., clustering) and *post-hoc predictive* methods (e.g., classification) for conducting customer segmentation, aiming to extract more detailed information for focused marketing efforts. The motivation for this methodology lies in the fact that the class labels for the classification model are oftentimes unavailable, which necessitates the first step of using clustering to gain insight into the structure of data and to derive class labels.

In *Publication 1*, we apply a hybrid approach that combines SOM-Ward clustering and DTs to conduct customer segmentation. The SOM-Ward clustering is first used to divide the customer base into distinct groups to differentiate high-spending from low-spending customers. Then, the DT is employed to gain insight into whether there are significant variables for distinguishing between high- and low-spending customers. The trained DT model is both descriptive and predictive in the sense that the model can produce straightforward rules for understanding the relationship of the input and target variables, as well as classifying the records by the splitting rules derived from the DT model. In addition, the capability of training continuous and/or categorical input variables with a DT model makes it possible to capture more information than the standard SOM model that requires its input to be numeric. Finally, the results of the two models are then compared and combined to

perform customer segmentation to identify the high- and low-spending customers, as well as customers with development potential (e.g., customers in the group displaying mid-range spending that have similar characteristics as the high-spending customers).

Publication 2 extends the hybrid approach in *Publication 1* by applying different classifiers for classifying potential customers. Specifically, the SOM-Ward clustering is first used to conduct customer segmentation. Then, three classifiers, the SVM, the ANN, and the DT, are employed to classify high- and low-spending customers. The three classifiers are then compared and combined into an ensemble model to predict potential high-spending customers from the mass customers. It is found that this hybrid approach could provide more detailed and accurate information about the customer base, especially the untapped mass market with potential high revenue contribution, for tailoring actionable marketing strategies.

In sum, the proposed hybrid approach first uses SOM-Ward clustering to conduct exploratory data analysis and to derive a class label regarding customers' value in a data driven fashion. This step is important when an *a priori* class label is impossible or difficult to identify, thereby making the knowledge gained using the SOM potentially very important for subsequent analysis using classification models. It is found that the combined model can provide more detailed information than that provided by either model used alone, thus more actionable information about the customer base for marketing purposes could be retrieved. The experiment process and the results of both publications are summarized in the following sub-sections respectively.

6.1.1 Publication 1

The training dataset is created so that a unique customer signature, including a variable that summarizes customers' total spending amount and a number of descriptive variables of their characteristics, is created for each customer for the period 2006–07. The training of the SOM-Ward model is carried out using Viscovery SOMine 5.0 (<http://www.viscovery.net/>), which is based upon the batch SOM algorithm with Ward's hierarchical clustering for the second-level clustering. All variables included in the training process are preprocessed using a *z*-score transformation to normalize their weight in training and postprocessed in order to have original values when interpreting models. In addition, we assign a higher priority factor to spending amount, aiming to give it more influence in the training process and to enable us to better distinguish between high- and low-spending customers.

Based upon the interpretability and the five criteria for evaluating the desirability of a segmentation model presented in Section 2.3.3, we select the clustering model that consists of seven segments: *exclusive customers*; *high-spending customers*; *customers with high loyalty point level*; *relatively young female customers*; *long-standing customers of the corporate chain*; *relatively old female customers*; and *male customers*. The component planes with heat map coloring

are used to visualize the distribution of each variable over different segments. This enables visual exploration of the characteristics of each segment. Detailed segment profiles can be found in Section 3.2 in *Publication 1*. From the analysis of the SOM-Ward Model, we find that *exclusive customers* and *high-spending customers*, those who spend much in the company, account for 13.7% of the customer base. We arrange the customers in sequence, from the lowest to highest according to their total purchase amounts. The top 13.7% of the customers are labeled high-spending customers, and the bottom 13.7% are labeled low-spending customers. The DT model is then created to classify these two groups of customers. Customers in the middle, whose spending amounts are not clearly high or low, are excluded from the training set. The spending potential of these customers will be predicted once the DT model is created.

The data preprocessing for the DT model is simple since it is a non-parametric classifier that is able to deal with difficult data. In order not to overfit the DT model, the maximum number of levels in the tree is limited to five and the minimum number of records in a node is set to 1000. In this study, the motivation of using a DT is two-fold: its descriptive and predictive capabilities of distinguishing high- and low- spending customers.

First, each of the terminal nodes in the DT is essentially an individual customer segment to which the path from the root node can be interpreted as characteristics that describe high- and low- spending customers. In addition, since the DT is built in a top-down manner, the splitting variables located near the root node are considered better at distinguishing different levels of the target variables. In Section 4.2 of *Publication 1*, we identify important variables (e.g., gender and loyalty point level) in classifying customers in terms of spending amount, as well as the most homogeneous segments in relation to the spending amounts: four terminal nodes with the highest percentage of high-spending customers and two terminal nodes with the highest percentage of low-spending customers.

In Section 5, the created DT model is used to identify customers with spending potential. Customers in Segments one and two, *exclusive customers* and *high-spending customers*, have already been identified as high-spending customers. The DT model then identifies which of the remaining segments have development potential in terms of spending amount.

The confusion matrix in Section 5 shows that the overall accuracy of the model is 72.6%, in which 82.8% of the high-spending customers are correctly identified while 62.4% of the low-spending customers are correctly identified. This accuracy is acceptable since the main focus is on building a model that can identify potential high-spending customers. We score all the cases in Segments Three, Four, Five, Six, and Seven through the created DT model, and then use two criteria to identify segments having development potential: the segments that have the highest number/percentage of customers are identified as high-spending customers, and the segment that has the highest propensity scores.

Based upon the first criterion, Segments Six and Three have a higher proportion of customers identified as high-spending customers; while based upon the latter criterion, Segment Three is identified as the segment with the highest potential segment, as most of the customers with the highest propensity scores are from Segment Three.

6.1.2 Publication 2

The structure of the training dataset and its preprocessing process in *Publication 2* are the same as in *Publication 1*. The data, however, are collected for the period 2007– 08. The SOM-Ward clustering is first used to divide the customer base into seven segments: two segments with high-spending customers and five segments with moderate- and low-spending customers. We use error charts, deviation charts, and component planes to profile segments. In this publication, the binary target variable (i.e., a binary class variable used to label high- and low-spending customers) is created in a different way than in *Publication 1*.

The customers in the shaded area in Figure 4 in *Publication 2* are extracted and labeled as high-spending customers. It is noted that customers with missing values are removed from the training data to have a fair comparison of the three classifiers in the subsequent analysis. Then a RFM analysis is done to locate customers with a low spending amount and a low potentiality of purchasing in the near future. The same number of customers as those labeled as high-spending customers are selected from customers with the lowest RFM scores.

Three classification models, the SVM, the MLP and the DT, were used respectively to classify high- and low-spending customers. The best model for each algorithm is selected based upon a ten-fold cross-validation. The accuracies of the three models based are estimated based on resubstitution error rate, ten-fold cross validation, and the AUC. It is found that the accuracy rates of the three models are high and close to each other. Though the accuracy rate of the boosted DT is higher than that of the other models, it tends to overfit the data whereas the SVM has the lowest risk of overfitting.

The three created models are then combined into an ensemble model to reveal potential mass customers who share similar characteristics with high-spending customers, thus pinpointing the segments with high-value potential. In the publication, we use three ensemble methods, voting, confidence-weighted voting, and highest confidence win, as are introduced in Section 4.6.5. The three methods are used to classify the training data, and the accuracies of the three ensemble methods compared. It is found that the three ensemble methods slightly raise the classification accuracy rate of SVM and ANN on the training data. Although their overall accuracy is slightly lower than that of the boosted DT, the latter is more prone to overfitting data. Therefore, we decide to use the ensemble method with the highest overall accuracy – the confidence-weighted voting – to predict mass customers and locate those customers with development potential. Finally, we create a table that lists the present value rank and the

average propensity score for being high-spending customers for each segment, which can be used as an index for measuring customer potential.

6.2 Temporal customer segmentation

This section presents the findings of the temporal customer segmentation model based upon Yao, et al. (2012b), *Publication 3*, and Yao, et al. (2013), *Publication 4*.

Publication 3 applies the SOTM for visual temporal clustering in identifying demographic customer segments and in tracking customer purchasing behavior over time, which aims to provide a holistic view of multivariate temporal patterns for better understanding of customers. The SOTM is used for 1) performing multivariate clustering of customers over time; 2) visualizing the temporal variation of the multivariate patterns; and 3) detecting and interpreting complex patterns of changes in the customer base and purchasing behavior during three special sales events. The results show that the SOTM enables exploring complex patterns over time by visualizing the results in a user-friendly way.

In *Publication 3*, we use the demographic variables to create a SOTM model in which the behavioral variables are associated with the model to explore customer behavior over time. In *Publication 4*, however, the behavioral variables are directly used in training the model. Since customers' purchasing behavior is much more dynamic compared to their demographics, we apply second-level clustering to the SOTM model to identify the patterns of changing, emerging, and disappearing of customer segment structures.

In sum, the approaches in both publications provide an effective means to facilitate comparisons of a sequence of segmentation models at different points in time, aiming to achieve a balance between a segmentation model being representative of the characteristics of current customers and not being too deviant from the previous segmentation model, as is the aim of evolutionary clustering as discussed in Section 2.4.1. It thus enables one to view changes in customers' shopping behavior and to evaluate the current segmentation model in a broader context. The experiment process and the results of both publications are summarized in the following sub-sections respectively.

6.2.1 Publication 3

The dataset contains weekly aggregated summarized purchasing behavior coupled with their demographics per customer for the department store. The data span 22 weeks from September 2008 to January 2009. For each week, only customers that made at least one purchase are included in the datasets.

We select a SOTM topology with 5×22 nodes, where five nodes represent the data topology for a specific week on the vertical direction and 22 for the time dimension to span the 22 weekly subsets. The number of nodes on the vertical axis is set to five according to the optimal number of customer segments in Yao,

et al. (2012a), where the pooled version of this dataset is used. We test different neighborhood radii ranging from 0.4 to 8. Based upon the three quality measures – quantization error, topographic error and distortion measure – we select the SOTM with the neighborhood radius of 1.2 for a good balance between topology preservation and quantization error.

The results of the SOTM model are visualized with the Sammon’s mapping (c.f. Figure 5 in *Publication 3*) and the component planes (c.f. Figure 6 in *Publication 3*). According to the Sammon’s mapping based upon the training variables (i.e., the demographic variables), the temporal changes in the horizontal direction are gradual, equal, and small, indicating that customers’ demographics in the five segments are stable over time. The multidimensionality of the SOTM model is described using component plane representations, which enable us to summarize the profiles of the five segments.

The component plane representation for the SOTM can be used to identify two types of patterns: a) cross-sectional distributions, and b) temporal changes. Some cross-sectional patterns revealed by the component planes include that male customers are located in the nodes on the lowest part of the map; customers with children at home and high income customers are located in the middle and lower parts of the map; customers with high loyalty point levels and service levels are located in the lower part of the map; and elderly customers are located on the second horizontal row on the map and young in the two rows below. The general temporal patterns are that customers’ demographics stay stable during most of the time, as is already confirmed in the Sammon’s mapping. However, Figure 5 (d, e) in *Publication 3* shows that the pre-Christmas period attracts customers with higher loyalty point and service levels.

In order to illustrate the temporal patterns of customers’ purchasing behavior, we associate behavioral variables to the SOTM model to examine their distributions among the segments in the 5×22 SOTM grid. As shown in Figure 7 in *Publication 3* – the component planes describing the behavioral variables – the temporal changes in customers’ purchasing behavior are nonstationary, especially during the sales events and the pre-Christmas sales. The revenue contribution from each segment, the main responding segments, and customers’ purchasing patterns during different sales events can be captured easily in the component planes and summarized in the tabular analysis. The results show that the purchasing behavior of customers changes during the events, but at the same time, that the sales events differ in the type of shopping behavior that they trigger.

This empirical study demonstrates that the SOTM can serve as a valuable visual exploratory tool for decision makers to see how successful the sales events have been according to different criteria (e.g., responding rate or sales revenue) and to aid in decision making concerning which kind of sales events to have.

6.2.2 Publication 4

Unlike *Publication 3* where behavioral variables do not influence the results of the clustering, they are used for training in *Publication 4*, with the aim of examining the changes in customer behavior from a different perspective. The data collection and preprocessing steps in *Publication 4* are the same as those in *Publication 3*. We also follow the procedure as documented in *Publication 3* in determining the specifications and architectures of the SOTM. We choose a neighborhood radius 2.8 – the smallest radius that achieves the highest quantization accuracies and of no topographic errors – to train a SOTM with 5×22 nodes.

Again, the component plane representation is used to visualize cross-sectional and temporal patterns of the SOTM model. The volatile temporal changes driven by the sales events and Christmas sales can be identified by comparing nodes along the horizontal direction, which also supports and complements the findings in *Publication 3*. With the stationary data of the demographic variables in *Publication 3*, rows of the SOTM roughly represent similar data at different points in time, thereby revealing general cross-sectional structures. However, as the behavioral data are often nonstationary, it is difficult to compare the structures of customer segments over time due to the changing nature of the data. Moreover, as the number of dimensions and nodes in a SOTM grid increases, the ability to perceive the temporal structural changes in data will be hindered. Therefore, a second-level clustering is applied to group the units of the SOTM model to represent temporal changes in cluster structures more objectively. In this study, we experiment with single-linkage, complete-linkage, Ward's method, and average-linkage, and find that average-linkage give more interpretable results. The number of customer segments is varied from three to six to explore the structures in the dataset, as is commonly done with hierarchical clustering methods. This top-down divisive analytical can identify cross-sectional differences at an early stage while more temporal differences can be perceived as the agglomeration process proceeds. In addition, this process enables one to identify different types of dynamics such as disappearing, emerging, and changing segments in segment structure more objectively.

6.3 Segment migration analysis

This section presents our findings of the segment migration analysis based upon Sarlin, et al. (2012), *Publication 5*, and Yao, et al. (2012a), *Publication 6*.

Publication 5 proposes a framework for computing, summarizing, and visualizing transition probabilities on the SOM. For each node, the framework calculates the probability of the objects moving to another node or predefined areas in a specified time period, and links the calculated probabilities to the original SOM grid using separate component planes. Thereby, it is possible to enabling one simultaneously compare the SOM model and the transition patterns (e.g., transition strengths and directions) in one view.

Publication 6 applies the framework proposed in *Publication 5* to customer segmentation and campaign-response analysis. First, we create a customer segmentation model based on customers' shopping behavior. Then, we connect this segmentation model with information about customers' responses to a number of sales campaigns based upon segment migration probabilities to 1) visualize which customer segments react to campaigns, and 2) to identify differences in purchasing behavior during campaign/non-campaign periods. The contribution of this framework is two-fold. First, unlike temporal customer segmentation, which explores structural changes in customer segments in temporal customer segmentation, segment migration analysis provides deeper insights into the dynamics (e.g., the direction and strength of the migration tendency) at the customer level, such as identifying different profiles of customers that respond to a campaign. Second, the method provides a user-friendly visual tool to consolidate the result of customer segmentation and that of segment migration analysis in one view, and thus facilitate convenient exploration of the migration patterns.

6.3.1 Publication 5

The framework proposed in *Publication 5* essentially involves two processes: 1) computing transition probability matrices of node-to-node or node-to-cluster transitions; and 2) linking the computation to the SOM grid using component plane representations. The authors propose three computations and their corresponding visualizations as illustrated in Figure 17. To demonstrate its usefulness in a real-world setting, the framework is then applied in two empirical studies: financial performance comparison of the European banking sector and monitoring indicators of financial crises. The results show that the framework is capable of assessing the evolution of the financial states of companies and in identifying low- and high-risk profiles and cyclicity in data.

Computation

1. TPMs for node-to-cluster switches

$$p_{il}(t+s) = \frac{n_{il}(t+s)}{\sum_{i=1}^c n_{il}(t+s)}$$

where n_{il} is the cardinality of data switching from node i to cluster l , t is a time coordinate and s is the time span for the switch.

2. Maximum transition probability of each node conditional on switching.

$$\max_l(p_{il})$$

3. Compute average transition probabilities for chosen set of s values.

$$p_{il}(t+\{1,2,\dots,S\}) = \frac{\sum_{s=1}^S n_{il}(t+s)}{\sum_{s=1}^S \sum_{i=1}^c n_{il}(t+s)}$$

Visualization

1. Show the probability to transit to each cluster on a separate component plane.

2. Summarize the information in Step 1 to one component plane, where the color code in each node is the probability of the most likely switch and a label represents that cluster.

3. Create the same feature planes as in Step 1, but as an average over a chosen set of s values.

Figure 17. Computations and visualizations of transition probabilities on the SOM.

6.3.2 Publication 6

The framework proposed in *Publication 5* is applied to *Publication 6* to provide a visual tool for identifying the strengths and directions of customers' migration patterns. The segment migration analysis in *Publication 6* consists of two stages. First, we create a behavioral segmentation model using SOM-Ward clustering. Then, we connect patterns of customers' responses to a number of sales campaigns based upon transition probabilities on the created segmentation model, to 1) model which customer segments react to campaigns, and 2) identify differences in purchasing behavior during campaign/non-campaign periods.

The data for training the behavioral segmentation model includes customers with spending amounts of less than EUR 50 in total from the department store chain during 2007–09. The training variables (i.e., 15 variables describing customers' purchasing behavior and purchase preferences) are summarized from the transaction system to quarterly aggregates per customer. A level of aggregation on a quarterly basis was considered appropriate because an aggregation over too short an interval could include noisy patterns, such as irregular and random purchasing behavior, while an aggregation over too long an interval may cancel out possible dynamics. Finally, each customer has eight corresponding records in the training file, one for each quarter during the period 2007–09.

The SOM-Ward clustering is used to conduct behavioral segmentation, and the optimal number of segments is determined to be five based upon the Silhouette measure, as is introduced in Section 4.4.3. The revealed segments are profiled using component plane visualization and the outstanding characteristics of each segment are summarized in Table 1 in *Publication 5*. In order to show customer

dynamics with regard to their responsiveness to campaigns, the framework proposed in *Publication 5* is applied to compute, summarize, and visualize campaign-driven node-to-segment migration patterns. Specifically, we create a response modeling dataset by summarizing customer purchasing behavior and product mix patterns for the periods before and during the campaigns. This dataset is projected onto the created behavioral segmentation model: each of the data records of the response modeling dataset is assigned to their BMUs in the customer segmentation model. The node-to-segment migrations are computed and visualized using component plane representations. Finally, by combining the information from the customer segmentation model and campaign response model, behavioral profiles of the campaign responders are created.

The results show that customers react in different ways to a campaign. The segment that includes the most valuable customers – female customers that spend high amounts and that have a high purchasing frequency – reacts positively to the campaign. On the other hand, the campaigns do not activate well the segment that includes male customers with high income levels and who usually buy expensive items. Both phenomena are interesting and worth investigating by campaign managers.

In sum, this framework combines customer segmentation and campaign-driven segment migration analysis to provide a holistic view of the patterns of customer purchase behavior and the underlying dynamics, in particular, within-segment heterogeneity in terms of campaign response propensity.

6.4 Summary

The chapter summarizes the analytical process and findings of the six papers for addressing the three aspects associated with a customer segmentation task: static customer segmentation, temporal customer segmentation, and segment migration analysis. In the final chapter, the key findings from the empirical study will be stated as well as some implications for future research.

7. Conclusion

This chapter gives a conclusive review of the thesis. The answers to the research questions are given first, followed by a concise description of the business implication of the studies. At the end of this chapter, the recommendations for future research are briefly discussed.

7.1 Review of the answers to the research questions

The main research questions are:

RQ1. How can customer segmentation be conducted and the results visualized?

RQ2. How can the changes of segment structures over time be detected and monitored?

RQ3. How can segment migration analysis be conducted to explore different migration profiles?

RQ1. How can customer segmentation be conducted and the results visualized?

Customer segmentation is an important component of analytical CRM. It divides the customer base into subsets of customers having similar characteristics, which enables companies to target customers in each segment collectively. In this thesis, the two-level SOM and classification models are used to conduct customer segmentation, where a clustering model is first used to explore the data structure and to divide data into groups, aiming to derive a class variable based upon cluster membership for subsequent classification analysis. In a real-world setting, determining a customer value class is not a straightforward task. The customer value class is oftentimes unavailable or difficult to be predefined; and it is important to involve people from business side in the process.

The SOM is a well-known and widely used visual clustering tool. It is chosen for the customer segmentation task for its simultaneous clustering via vector quantization and projection via neighborhood-preservation. The multidimensionality of the segmentation model is illustrated using the component plane representation, which is capable of providing multiple views of the segmentation model from the perspectives of different variables in each two-dimensional display. In addition, the SOM provides a direct means to customer segmentation and to locate interesting “hot spots” for deriving the class variable that distinguishes high- and low-spending customers. Then the three classifiers – the DT, the ANN and the SVM – are applied to classify the high- and low-spending customers. While the SOM can conduct customer segmentation, the three classification models are created for the following reasons: 1) For each customer, the classifiers can calculate the propensity scores of being a high-spending customer. This information can be used to identify segments with

development potential. 2) The classification model is independent of the customer segmentation model based on the SOM, and it can use variables that cannot be utilized by the SOM for the static segmentation model, therefore more information could be retrieved for describing the interesting customer groups. 3) Some classifiers are both descriptive and predictive, the information derived from the classification models can be directly used for targeting new customers. For example, The DT can also be used as a visual means for exploring the relationship between customers' purchase amounts and their demographic and behavioral characteristics. This can also be used to identify the significant features for predicting the high-spending customers.

RQ2. How can the changes of segment structures over time be detected and monitored?

One of the evaluation criteria for a good customer segmentation model is segment stability. However, the problem of segment instability is common in this changing business environment, and a number of authors have identified this phenomenon as discussed in Chapter Two. The study of changes in segment structure assumes that there are temporal changes of segment sizes and composition. To this end, the SOTM was used to conduct temporal customer segmentation to explore the changes of segment structures over time. The SOTM consists of a series of one-dimensional SOMs ordered in consequent time units and represents both time and data topology on a two dimensional grid. The one-dimensional SOM in time t is initialized by the reference vectors of the SOM in time $t - 1$. This short-term memory makes the neighboring one-dimensional SOMs have similar orientations and thus be comparable to each other. This property makes the SOTM a promising means for evolutionary clustering: to produce a sequence of clustering for each time point in time. The short-term memory of the SOTM can deal simultaneously with two conflicting problems: first, the segment model should correctly reflect the current data as much as possible; second, segment models at neighboring points in time should not be deviant from each other too much.

In the thesis, the changes of segment structures are examined from different angles. First, temporal customer segmentation is performed based upon customers' demographic characteristics. As customers' demographic characteristics are relatively stationary over time, rows of the SOTM would represent customers with similar demographic characteristics at different points in time. Then, the behavioral variables are associated to the SOTM model to illustrate the variations in shopping behavior of demographically similar customers.

With stationary data, rows of the SOTM would represent customers with similar characteristics regarding the training variables at different points in time. However, when the data are nonstationary (e.g., customers' purchasing behavior), horizontally neighboring nodes may represent data of different characteristics. It is therefore difficult to judge whether or not nodes in a single row are in the

segment, especially when changes in segment structures are substantial. Therefore a SOTM model is created based on customers' purchasing behavior, then a second-level clustering is applied to the nodes of the SOTM model to represent temporal changes in segment structures more objectively. The two-level SOTM facilitates identification of the patterns of changing, emerging, and disappearing of customer segment structures.

RQ3. How can segment migration analysis be conducted to explore different migration profiles?

In addition to the problem of changing segment structures, the other type of dynamics in customer segmentation is the changes of segment membership of individual customers over time. Segment migration analysis tracks customers' migration patterns among segments and enables identification of within-segment heterogeneity in terms of migration direction and strength. The framework (Sarlin, et al., 2012) for computing, summarizing, and visualizing transition probabilities on the SOM provides a means for conducting segment migration analysis.

Following this approach, the migration probabilities of customers in each node switching to each segment can be calculated and visualized in separate component planes. These combined component planes are able to provide strengths and direction for the segment migration patterns. Unlike the common approach that calculates segment-to-segment migration patterns, the node-to-segment migration probabilities visualized on the migration planes can reveal within-segment heterogeneity in terms of migration patterns.

7.2 Limitations and future work

In this thesis, the SOM and its extensions were used to conduct customer segmentation. Two types of dynamics are addressed in a customer segmentation model: 1) the changes of segment structures over time, and 2) the changes of segment membership of individual customers. The SOTM was used to look at the first type of dynamics, and this provided a series of segmentation models for each point in time. This approach relates to the evolutionary clustering aiming to find a balance between each segmentation model being faithful to their respective data and neighboring segmentation models being comparable to each other. Segment migration analysis based upon calculating, summarizing, and visualizing transition probabilities on the SOM was used to view the second type of dynamics. While these two dynamics have been addressed separately, it is not uncommon for them to happen simultaneously. In other words, customers might switch between evolving segments that are not static over time. Following Hu and Rau (1995), future work should aim at simultaneously addressing changes in segment memberships of individual objects and changes in segment composition. This could, for instance, be performed by combining a visualization of the SOTM with migration probabilities.

References

- ACORN. (2013). *The ACORN user guide*. Retrieved June 13, 2013, from acorn.caci.co.uk/downloads/Acorn-User-guide.pdf.
- Adomavicius, G., & Bockstedt, J. (2008). C-TREND: Temporal cluster graphs for identifying and visualizing trends in multiattribute transactional data. *IEEE Transactions on Knowledge and Data Engineering*, 20(6), 721-735.
- Afolabi, M. O., & Olude, O. (2007). Predicting Stock Prices Using a Hybrid Kohonen Self Organizing Map (SOM). In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, Big Island, Hawaii, January 3-6.
- Ang, L., & Buttle, F. (2006). CRM software applications and business performance. *Journal of Database Marketing & Customer Strategy Management*, 14(1), 4-16.
- Arndt, D., & Langbein, N. (2002). Data quality in the context of customer segmentation. In *Proceedings of the 7th International Conference on Information Quality*, Cambridge, Massachusetts, November 8-10.
- Assael, H., & Roscoe Jr, A. M. (1976). Approaches to market segmentation analysis. *The Journal of Marketing*, 40(4), 67-76.
- Back, B., Sere, K., & Vanharanta, H. (1998). Managing complexity in large data bases using self-organizing maps. *Accounting, Management and Information Technologies*, 8(4), 191-210.
- Bayer, J. (2010). Customer segmentation in the telecommunications industry. *Journal of Database Marketing & Customer Strategy Management*, 17, 247-256.
- Beane, T., & Ennis, D. (1987). Market segmentation: a review. *European Journal of Marketing*, 21(5), 20-42.
- Berry, L. L. (1983). Relationship Marketing. In L.L.Berry, G.L.Shostack & G.D.Upah (Eds.), *Emerging Perspectives on Services Marketing* (pp. 25-28), Chicago: American Marketing Association.
- Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Indianapolis: Wiley Computer Publishing.

- Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83(3), 377.
- Blocker, C. P., & Flint, D. J. (2007). Customer segments as moving targets: Integrating customer value dynamism into segment instability logic. *Industrial Marketing Management*, 36(6), 810-822.
- Boulding, W., Staelin, R., Ehret, M., & Johnston, W. J. (2005). A customer relationship management roadmap: What is known, potential pitfalls, and where to go. *Journal of Marketing*, 69(4), 155-166.
- Brangule-Vlagsma, K., Pieters, R. G. M., & Wedel, M. (2002). The dynamics of value segments: modeling framework and empirical illustration. *International Journal of Research in Marketing*, 19(3), 267-285.
- Breiman, L. (1984). *Classification and regression trees*. New York: Chapman & Hall/CRC.
- Budayan, C., Dikmen, I., & Birgonul, M. T. (2009). Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. *Expert Systems with Applications*, 36(9), 11772-11781.
- Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis: elements of the sociology of corporate life*. London: Heinemann.
- Buttle, F. (2004). *Customer relationship management: Concepts and tools*. Oxford: Elsevier.
- Calantone, R. J., & Sawyer, A. G. (1978). The stability of benefit segments. *Journal of Marketing Research*, 15(3), 395-404.
- Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. London: Morgan Kaufmann.
- Chakrabarti, D., Kumar, R., & Tomkins, A. (2006). Evolutionary clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, Philadelphia, PA, August 20-23.
- Chappelier, J. C., & Grumbach, A. (1996). A Kohonen map for temporal sequences. In *Proceeding of Neural Networks and Their Application*, Marseille, March.

- Chappell, G. J., & Taylor, J. G. (1993). The temporal Kohonen map. *Neural Networks*, 6(3), 441-445.
- Chen, C. (2005). Top 10 Unsolved Information Visualization Problems. *IEEE Computer Graphics and Applications*, 25(4), 12-16.
- Chisnall, P. M. (1975). *Marketing: a behavioural analysis*. Ithaca: McGraw-Hill.
- CIELab, 1986. *Colorimetry*, CIE Publication, No. 15.2.
- Coussement, K., Van den Bossche, F.A.M., & De Bock, K.W. (2012). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research*, In Press, Corrected Proof.
- Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge university press.
- Crone, S. F., Lessmann, S., & Stahlbock, R. (2006). The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, 173(3), 781-800.
- D'Urso, P., & De Giovanni, L. (2008). Temporal self-organizing maps for telecommunications market segmentation. *Neurocomputing*, 71(13), 2880-2892.
- Demirhan, A., & Güler, İ. (2011). Combining stationary wavelet transform and self-organizing maps for brain MR image segmentation. *Engineering Applications of Artificial Intelligence*, 24(2), 358-367.
- Denny, Christen, P., & Williams, G. (2012). Analysis of Cluster Migrations Using Self-Organizing Maps. In L. Cao, J. Huang, J. Bailey, Y. Koh & J. Luo (Eds.), *New Frontiers in Applied Data Mining* (Vol. 7104, pp. 171-182): Springer Berlin Heidelberg.
- Denny, & Squire, D. (2005). Visualization of Cluster Changes by Comparing Self-organizing Maps. In T. Ho, D. Cheung & H. Liu (Eds.), *Advances in Knowledge Discovery and Data Mining* (Vol. 3518, pp. 410-419): Springer Berlin Heidelberg.

- Denny, Williams, G., & Christen, P. (2010). Visualizing temporal cluster changes using Relative Density Self-Organizing Maps. *Knowledge and Information Systems*, 25(2), 281-302.
- Denny, Williams, G. J., & Christen, P. (2008). ReDSOM: relative density visualization of temporal changes in cluster structures using self-organizing maps. In *Proceedings of the Eighth IEEE International Conference on Data Mining*, Pisa, Italy, December 15-19.
- Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3), 32-57.
- Eklund, T., Back, B., Vanharanta, H., & Visa, A. (2003). Using the self-organizing map as a visualization tool in financial benchmarking. *Information Visualization*, 2(3), 171-181.
- Eklund, T., Back, B., Vanharanta, H., & Visa, A. (2008). A face validation of a SOM-based financial benchmarking model. *Journal of Emerging Technologies in Accounting*, 5(1), 109-127.
- Eppler, M. J., & Platts, K. W. (2009). Visual strategizing: The systematic use of visualization in the strategic-planning process. *Long Range Planning*, 42(1), 42-74.
- Experian (2008). *Mosaic osana liiketoimintaa*. Retrieved May 30th 2013, from <http://www.experian.fi/Pages/PALVELUT%20JA%20TUOTTEET/MosaicOsanaLiiketoimintaa.htm>
- Experian. (2010). Optimise the value of your customers and locations, now and in the future. Retrieved May 30th 2013, from <http://www.experian.co.uk/assets/business-strategies/brochures/mosaic-uk-2009-brochure-jun10.pdf>.
- Farley, J. U., Winer, R. S., & Lehmann, D. R. (1987). Stability of membership in market segments identified with a disaggregate consumption model. *Journal of Business Research*, 15(4), 313-328.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-54.
- Feelders, A., Daniels, H., & Holsheimer, M. (2000). Methodological and practical aspects of data mining. *Information & Management*, 37(5), 271-281.

- Ferreira de Oliveira, M. C., & Levkowitz, H. (2003). From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 9(3), 378-394.
- Fessant, F., & Midenet, S. (2002). Self-organising map for data imputation and correction in surveys. *Neural Computing & Applications*, 10(4), 300-310.
- Frank, R. E. (1967). Correlates of buying behavior for grocery products. *The Journal of Marketing*, 31(4), 48-53.
- Guo, D., Chen, J., MacEachren, A. M., & Liao, K. (2006). A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics*, 12(6), 1461-1474.
- Guyon, I. (2008). SVM application list. Retrieved May 30th 2013, from <http://www.clopinet.com/isabelle/Projects/SVM/applist.html>
- Ha, S. H., Bae, S. M., & Park, S. C. (2002). Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case. *Computers & Industrial Engineering*, 43(4), 801-820.
- Haley, R. I. (1968). Benefit segmentation: a decision-oriented research tool. *The Journal of Marketing*, 32(3), 30-35.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques: Concepts and Techniques*. Waltham, MA: Elsevier Science.
- Hansen, P., & Delattre, M. (1978). Complete-link cluster analysis by graph coloring. *Journal of the American Statistical Association*, 73(362), 397-403.
- Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Proceedings of the International Joint Conference on Neural Networks*, San Diego.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1), 75-105.
- Hoek, J., Gendall, P., & Esslemont, D. (1996). Market segmentation: A search for the Holy Grail? *Journal of Marketing Practice: Applied Marketing Science*, 2(1), 25-34.

- Holmbom, A. H., Eklund, T., & Back, B. (2011). Customer portfolio analysis using the SOM. *International Journal of Business Information Systems*, 8(4), 396-412.
- Homburg, C., Steiner, V. V., & Totzek, D. (2009). Managing dynamics in a customer portfolio. *Journal of Marketing*, 73(5), 70-89.
- Hsu, S.H., Hsieh, J.P.A., Chih, T.C., & Hsu, K.C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications*, 36(4), 7947-7951.
- Hu, M.Y., & Rau, P.A. (1995). Stability of Usage Segments, Membership Shifts Across Segments and Implications for Marketing Strategy: An Empirical Examination. *Mid-Atlantic Journal of Business*, 31(2), 161-177.
- Hubert, L., & Schultz, J. (1976). Quadratic assignment as a general data-analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29(2), 190-241.
- Hughes, A. M. (2006). Strategic database marketing. Chicago: McGraw-Hill.
- Iivari, J. (1991). A paradigmatic analysis of contemporary schools of IS development. *European Journal of Information Systems*, 1(4), 249-272.
- Johnson, M. D., & Selnes, F. (2004). Customer portfolio management: Toward a dynamic theory of exchange relationships. *Journal of Marketing*, 68(2), 1-17.
- Kaski, S. (1997). *Data exploration using self-organizing maps*. Helsinki University of Technology, Helsinki.
- Kaski, S., Venna, J., & Kohonen, T. (2000). Coloring that reveals cluster structures in multivariate data. *Australian Journal of Intelligent Information Processing Systems*, 6(2), 82-88.
- Keim, D. A., Mansmann, F., & Thomas, J. (2010). Visual analytics: how much visualization and how much analytics? *ACM SIGKDD Explorations Newsletter*, 11(2), 5-8.
- Kiang, M. Y., Hu, M. Y., & Fisher, D. M. (2006). An extended self-organizing map network for market segmentation—a telecommunication example. *Decision Support Systems*, 42(1), 36-47.

- Kiang, M. Y., & Kumar, A. (2001). An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications. *Information Systems Research*, 12(2), 177-194.
- Kifer, D., Ben-David, S., & Gehrke, J. (2004). Detecting change in data streams. In *Proceedings of the Thirtieth international conference on very large data bases*. Toronto, Canada, 29 August - 3 September.
- Kiviluoto, K. (1996). Topology preservation in self-organizing maps. In *Proceedings of the IEEE International Conference on Neural Networks*. Washington, DC, June 3-6.
- Koch, R. (1999). *The 80/20 principle: the secret to achieving more with less*. New York: Crown Business.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montréal, Canada, August 20-25.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1), 59-69.
- Kohonen, T. (1988). The 'Neural' Phonetic Typewriter. *Computer*, 21(3), 11-22.
- Kohonen, T. (1991). The Hypermap Architecture. In T.Kohonen, K.Mkisara, O.Simula & J.Kangas (Eds.), *Artificial Neural Networks* (Vol. II, pp. 1357-1360). Amsterdam, Netherlands: North-Holland.
- Kohonen, T. (2001). *Self-Organizing Maps*: Springer.
- Kotler, P. (2002). *Marketing management* (Millennium edition ed.): Pearson Custom Publishing.
- Kuiper, F.K., & Fisher, L. (1975). 391: A Monte Carlo comparison of six clustering procedures. *Biometrics*, 31(3), 777-783.
- Lee, S.C., Suh, Y.H., Kim, J.K., & Lee, K.J. (2004). A cross-national market segmentation of online game industry using SOM. *Expert systems with applications*, 27(4), 559-570.
- Lingras, P., Hogo, M., & Snorek, M. (2004). Temporal cluster migration matrices for web usage mining. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, Beijing, China, September 20-24.

- Lingras, P., Hogo, M., Snorek, M., & West, C. (2005). Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Information Sciences*, 172(1–2), 215-240.
- Linoff, G. S., & Berry, M. J. (2011). *Data Mining Techniques: For Marketing, Sales, And Customer Relationship Management (3 edition ed.)*. Indianapolis, Indiana: Wiley Computer Publishing.
- Luyssaert, S., Sulkava, M., Raitio, H., & Hollmén, J. (2004). Evaluation of forest nutrition based on large-scale foliar surveys: are nutrition profiles the way of the future? *Journal of Environmental Monitoring*, 6(2), 160-167.
- Maimon, O., & Rokach, L. (2005). *Data mining and knowledge discovery handbook* : Springer.
- March, S.T., & Smith, G.F. (1995). Design and natural science research on information technology. *Decision support systems*, 15(4), 251-266.
- Marghescu, D. (2007). Evaluation of projection techniques using Hubert's Γ statistics. In *Proceedings of the IADIS 1st European Conference on Data Mining*. Lisbon, Portugal, July 5-7.
- Martin-del-Brio, B., & Serrano-Cinca, C. (1993). Self-organizing neural networks for the analysis and representation of data: Some financial cases. *Neural Computing & Applications*, 1(3), 193-206.
- McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of business research*, 60(6), 656-662.
- McDonald, M., & Dunbar, I. (2004). *Market segmentation: How to do it, how to profit from it*: Butterworth-Heinemann.
- Miglautsch, J. R. (2000). Thoughts on RFM scoring. *The Journal of Database Marketing*, 8(1), 67-72.
- Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. *The Computer Journal*, 20(4), 359-363.
- Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, 2(4), 345-389.

- Ngai, E. W. T., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592-2602.
- Payne, A., & Frow, P. (2005). A strategic framework for customer relationship management. *Journal of Marketing*, 69(4), 167-176.
- Piatetsky-Shapiro, G., Brachman, R., Khabaza, T., Kloesgen, W., & Simoudis, E. (1996). An overview of issues in developing industrial data mining and knowledge discovery applications. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, August 2-4.
- Pözlbauer, G. (2004). Survey and comparison of quality measures for self-organizing maps. In *Proceedings of the Fifth Workshop on Data Analysis*, Tatranska Polianka, Slovak Republic, June 24-27.
- Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: review and suggestions for application. *Journal of marketing research*, 20(2), 134-148.
- Pyle, D. (1999). *Data preparation for data mining*, San Francisco, California: Morgan Kaufmann.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*, San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Quinlan, J. R. (2004). Data mining tools See5 and C5. 0.
- Rand, W. M. (1971). Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336), 846-850.
- Reichheld, F. F., & Teal, T. (2001). *The loyalty effect: The hidden force behind growth, profits, and lasting value*, Boston, Massachusetts: Harvard business press.
- Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of marketing research*, 41(3), 293-305.

- Reinartz, W., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, 64(4), 17-35.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20(1), 53-65.
- Rousseeuw, P. J., & Kaufman, L. (1990). *Finding groups in data: An introduction to cluster analysis*, Hoboken, New Jersey: John Wiley & Sons.
- Ryals, L., & Payne, A. (2001). Customer relationship management in financial services: towards information-enabled relationship marketing. *Journal of Strategic Marketing*, 9(1), 3-27.
- Rygielski, C., Wang, J.-C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24(4), 483-502.
- Saar-Tsechansky, M., & Provost, F. (2007). Handling missing values when applying classification models. *Journal of machine learning research*, 8, 1623-1657.
- Sammon Jr, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 100(5), 401-409.
- Sarlin, P. (2010). Visual monitoring of financial stability with a self-organizing neural network. In *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications* , Cairo, Egypt, November 29 - December 1.
- Sarlin, P. (2011). Evaluating a Self-Organizing Map for Clustering and Visualizing Optimum Currency Area Criteria. *Economics Bulletin*, 31(2), 1483-1495.
- Sarlin, P. (2012). Decomposing the Global Financial Crisis: A Self-Organizing Time Map. In *Proceedings of the 16th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, San Sebastain, Spain, September 10 - 12.
- Sarlin, P. (2013a). Data and dimension reduction for visual financial performance analysis. *Information visualization* (forthcoming).

- Sarlin, P. (2013b). Self-Organizing Time Map: An Abstraction of Temporal Multivariate Patterns. *Neurocomputing*, 99(1), 496-508.
- Sarlin, P., & Marghescu, D. (2011). Visual predictions of currency crises using self-organizing maps. *Intelligent Systems in Accounting, Finance and Management*, 18(1), 15-38.
- Sarlin, P., & Yao, Z. (2013). Clustering of the Self-Organizing Time Map. *Neurocomputing*, 121, 317-327.
- Sarlin, P., Yao, Z., & Eklund, T. (2012). A Framework for State Transitions on the Self-Organizing Map: Some Temporal Financial Applications. *Intelligent Systems in Accounting, Finance and Management*, 19(3), 189-203.
- Saunders, J. (1993). Cluster analysis for market segmentation. *European Journal of Marketing*, 14(7), 422-435.
- Schön, D. A. (1984). *The reflective practitioner: How professionals think in action*: Basic Books.
- Segev, A., & Kantola, J. (2012). Identification of trends from patents using self-organizing maps. *Expert Systems with Applications*, 39(18), 13235-13242.
- Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, 17(3), 227-238.
- Simon, H. A. (1996). *The sciences of the artificial*: MIT press.
- Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21(1), 3-8.
- Steenkamp, J.E.M., & Ter Hofstede, F. (2002). International market segmentation: issues and perspectives. *International Journal of Research in Marketing*, 19(3), 185-213.
- Sulkava, M., & Hollmén, J. (2003). Finding profiles of forest nutrition by clustering of the self-organizing map. In *Proceedings of the Workshop on Self-Organizing Maps*, Kitakyushu, Japan, September 11-14.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285-1293.

- Theodoridis, S., & Koutroumbas, K. (2008). *Pattern Recognition*: Elsevier Science.
- Thomas, J. J., & Cook, K. A. (Eds.). (2005). *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press.
- Thomas, M. (1980). Market segmentation. *Quarterly review of marketing*, 6(1).
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.
- Tseng, M.M., & Piller, F. (2003). *The customer centric enterprise-advances in mass customization and personalization*: Springer.
- Tsiptsis, K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*: Wiley.
- Turban, E., Aronson, J. E., Liang, T. P., & Sharda, R. (2007). *Decision support and business intelligence systems (8 ed.)*: Pearson Education.
- Tynan, A. C., & Drayton, J. (1987). Market segmentation. *Journal of Marketing Management*, 2(3), 301-335.
- Vapnik, V. (1995). *The nature of statistical learning theory*: Springer-Verlag New York, Inc.
- Vapnik, V., & Sterin, A. (1977). On structural risk minimization or overall risk in a problem of pattern recognition. *Automation and Remote Control*, 10(3), 1495-1503.
- Varsta, M., Heikkonen, J., & Millan, J. R. (1997). Context learning with the self-organizing map, In *Proceedings of the Workshop on Self-Organizing Maps*, Espoo, Finland, June 4 - 6.
- Venna, J., & Kaski, S. (2001). Neighborhood preservation in nonlinear projection methods: An experimental study, In *Proceedings of the International Conference on Artificial Neural Networks*, Vienna, Austria, August 21-25.
- Verhoef, P. C., Spring, P. N., Hoekstra, J. C., & Leeflang, P. S. H. (2003). The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. *Decision Support Systems*, 34(4), 471-481.

- Vesanto, J. (1999). SOM-based data visualization methods. *Intelligent data analysis*, 3(2), 111-126.
- Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3), 586-600.
- Vesanto, J., Sulkava, M., & Hollmén, J. (2003). On the decomposition of the self-organizing map distortion measure. In *Proceedings of the workshop on self-organizing maps*, Kitakyushu, Japan, Sept 11-14.
- Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations*: Springer.
- West, C., MacDonald, S., Lingras, P., & Adams, G. (2005). Relationship between Product Based Loyalty and Clustering based on Supermarket Visit and Spending Patterns. *International Journal of Computer Science and Applications*, 2(2), 85-100.
- Williams, G., & Huang, Z. (1997). Mining the knowledge mine. In *Proceedings of the tenth Australian Joint Conference on Artificial Intelligence*, Perth, Australia, November 30 - December 4.
- Wind, Y. (1978). Issues and advances in segmentation research. *Journal of marketing research*, 15(3), 317-337.
- Wong, P. C., Shen, H. W., Johnson, C. R., Chen, C., & Ross, R. B. (2012). The Top 10 Challenges in Extreme-Scale Visual Analytics. *IEEE Computer Graphics and Applications*, 32(4), 63-67.
- Yang, Q., & Wu, X. (2006). 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, 5(04), 597-604.
- Yao, Z., Eklund, T., & Back, B. (2010a). Using SOM-Ward Clustering and Predictive Analytics for Conducting Customer Segmentation. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops*, Sydney, Australia, December 14 - 17.
- Yao, Z., Holmbom, A., Eklund, T., & Back, B. (2010b). Combining unsupervised and supervised data mining techniques for conducting customer portfolio analysis. In *Proceedings of the 10th Industrial conference on data mining*, Berlin, Germany, July 12-14.

Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2012a). Combining visual customer segmentation and response modeling. In *Proceedings of the 20th European Conference on Information Systems*, Barcelona, Spain, June 10-13.

Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2012b). Temporal Customer Segmentation Using the Self-organizing Time Map. In *Proceedings of the 16th International Conference on Information Visualisation*, Montpellier, France, July 10 - 13.

PART TWO: Original Research Papers

Research Paper 1

Yao, Z., Holmbom, A., Eklund, T., & Back, B. (2010). Combining unsupervised and supervised data mining techniques for conducting customer portfolio analysis. In *Proceedings of the 10th Industrial Conference on Data Mining (ICDM 2010)*, Berlin, Germany, 12-14 July, 2010, pp. 292-307.

Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis

Zhiyuan Yao, Annika H. Holmbom, Tomas Eklund, and Barbro Back

Åbo Akademi University and Turku Centre for Computer Science (TUCS),
Joukahainengatan 3-5A, 20520 Turku, Finland
{Zhiyuan.Yao, Annika.H.Holmbom, Tomas.Eklund,
Barbro.Back}@abo.fi

Abstract. Leveraging the power of increasing amounts of data to analyze customer base for attracting and retaining the most valuable customers is a major problem facing companies in this information age. Data mining technologies extract hidden information and knowledge from large data stored in databases or data warehouses, thereby supporting the corporate decision making process. In this study, we apply a two-level approach that combines SOM-Ward clustering and decision trees to conduct customer portfolio analysis for a case company. The created two-level model was then used to identify potential high-value customers from the customer base. It was found that this hybrid approach could provide more detailed and accurate information about the customer base for tailoring actionable marketing strategies.

Keywords: Customer relationship management (CRM), customer portfolio analysis (CPA), Self-organizing maps (SOM), Ward's clustering, decision trees.

1 Introduction

For a long time, the focus of modern companies has been shifting from being product-oriented to customer-centric organizations. In the industry it is commonly held that maintaining existing customers is more cost-effective than attracting new ones, and that 20% of customers create 80% of the profit [1,2]. Reichheld and Teal [3] also point out that a 5% increase in customer retention leads to a 25–95% increase in company profit. Therefore, companies are focusing attention on building relationships with their customers in order to improve satisfaction and retention. This implies that companies must learn much about their customers' needs and demands, their tastes and buying propensities, etc., which is the focus of *Customer Relationship Management (CRM)* [4]. For CRM purposes, data mining techniques can potentially be used to extract hidden information from customer databases or data warehouses.

Data mining techniques can potentially help companies efficiently conduct *Customer Portfolio Analysis (CPA)*, which is the process of analyzing the existing and potential value of customers, thereby allocating limited resources to various customer groups according to the corporate strategy [4,5]. In this study, we propose a hybrid approach that combines the Self-Organizing Map (SOM)-Ward clustering [6,7] and decision trees for conducting CPA, aiming to create a more informative model for

focused marketing efforts, compared to using either method alone. First, we use SOM-Ward clustering to conduct customer segmentation, so that the customer base is divided into distinct groups of customers with similar characteristics and behavior. This will allow us to identify the characteristics that separate high-spending customers from low-spending customers. Then, a decision tree technique will be used to further explore the relationship between customers' spending amounts and their demographic and behavioral characteristics. Finally, the trained decision tree model will be used to identify the segments with development potential, as well as customers in the group displaying mid-range spending that have similar characteristics as the high-spending customers. This group thus represents potential high-value customers if correctly activated. By extension, this type of analysis would allow companies to adjust their marketing efforts in order to better fit their customers' needs and demands, not only helping to enhance their relationship with important customers but also cutting down on advertising costs and improving the profitability of the entire customer base.

Although the SOM-based approach and Decision Trees have been used for market segmentation, classification, and data exploration problems individually, these two approaches have not to our knowledge previously been combined to perform CPA. A dataset of more than one million customers was used to create the models.

The remainder of this paper is organized as follows. Section two introduces the methodology (SOM-Ward and Decision Trees) and the data used in this study. Sections three and four document the training and analysis of the SOM-Ward and Decision Tree models respectively. In Section five, the trained Decision Tree model is used to analyze unclassified customers in order to identify their market potential. Section six presents our conclusions.

2 Methodology

2.1 The SOM and SOM-Ward Clustering

The SOM is a well-known and widely used unsupervised neural network that is able to explore relationships in multidimensional input data and project them onto a two-dimensional map, where similar inputs are self-organized and located together [8]. Additionally, as an unsupervised artificial neural network (ANN), the SOM is a data-driven clustering method. In other words, it works with very little *a priori* information or assumptions concerning the input data. Moreover, the SOM is able to compress the input data while preserving the topological relationships of the underlying data structure [8]. For these reasons, the SOM is considered an important tool for conducting segmentation tasks. The algorithm is well-known and will, therefore, not be further presented in this paper. Readers are referred to Kohonen [8] for details concerning the algorithm.

The SOM has been widely applied as an analytical tool in different business-related areas [9-11], including market segmentation [12-14]. In the above mentioned studies, the SOM is used alone, compared with, or used in conjunction with other clustering techniques for conducting market segmentation tasks. Vesanto and Alhoniemi [15] proposed a two-level approach, e.g., SOM-Ward clustering, for conducting clustering tasks. First, the dataset is projected onto a two-dimensional display using

the SOM. Then, the resulting SOM is divided into groups. Lee et al. [14], adopting the two-level SOM (using SOM and K-means clustering), conducted a market segmentation of the Asian online game market and found that the two-level SOM is more accurate in classification than K-means clustering or the SOM alone. Samarasinghe [16], comparing two clustering methods (SOM-Ward and K-means clustering) drew the conclusion that SOM-Ward clustering resulted in better representations of the top and middle clusters than K-means alone.

As was previously mentioned, SOM-Ward clustering is a two-level clustering approach that combines the SOM and Ward's clustering algorithm. Ward's clustering is an agglomerative (bottom-up) hierarchical clustering method, which starts with a clustering in which each map node is treated as a separate cluster. The two clusters with the minimum distance are merged in each step until there is only one cluster left on the map [7]. SOM-Ward's clustering is a modification of Ward's clustering which limits cluster agglomeration to topologically neighboring nodes.

2.2 The Decision Tree

A Decision Tree is a supervised data mining technique that can be used to partition a large collection of data into smaller sets by recursively applying two-way and/or multi-way splits [17]. Compared to other data mining techniques, the Decision Tree has many advantages. First, as opposed to "black box" data mining techniques, the Decision Tree produces straightforward rules for classification and prediction purposes. Second, it is relatively insensitive to outliers [17] and skewed data distributions [18,19], and some decision tree algorithms are even capable of dealing with both numeric and nominal variables [19]. Thirdly, the decision tree is also a significant data exploration tool that can potentially be used to unveil the relationship between candidate independent and dependent variables. It can also be used to identify the significant variables for predicting the dependent variable [17]. These advantages make the decision tree applicable to a wide variety of business and marketing problems. For example, Fan et al. [20] adopted the decision tree to identify significant determinants of house prices and to predict these. Abrahams et al. [21] used decision trees to create a marketing strategy for a pet insurance company. Sheu et al. [22] adopted it to explore the potential relationship between important influential factors and customer loyalty. The findings of these studies inspire us to adopt the decision tree to explore the relationship between customers' purchase amounts and customers' demographic and behavioral characteristics, with special attention to the characteristics of high- and low-spending customers.

In this study, the CART algorithm [23] is employed to construct a binary classification tree. It is a tree-based classification method that uses recursive two-way partitioning to split the training records into segments where the records tend to fall into a specific class of the target variable. In other words, the records in the terminal nodes tend to be homogeneous with regard to the target variable. The CART algorithm employs an exhaustive search for the best independent variable for classification purposes at each split. First, the algorithm checks all possible independent variables and their potential values at each split. Then, the algorithm chooses an independent variable that maximizes within-node purity to split the node. We use the Gini index of diversity [23] to measure the improvement in purity at each split. For example, if all

the records in a node fall into a specific class of dependent variable, the node is considered pure; however, if the records of each class are proportionate, the node is considered impure. This process is repeated until some user-specified criteria are met, e.g., the maximum tree depth, the minimum number of records in a node or the minimum change in within-node purity improvement [17]. When the tree growing process ends, there is a unique path from the root to each terminal node. These paths can be considered a set of if-then rules that can be used to classify the records.

2.3 The Data

The case company is a national retailer belonging to a large, multiservice Finnish corporation. The corporation uses a common loyalty card system which offers cardholders various discounts and rewards for purchases. The cardholder is required to provide basic personal information in order to register for the loyalty card, and their transactional information is collected and recorded in the system. The dataset, containing 1,480,662 customers, was obtained through the loyalty card system, and contained sales information from several department stores in Finland, for the period 2006-07. The dataset consists of ten variables that fall into two categories: *demographic* and *behavioral* variables.

The demographic variables consist of the following:

- Age
 - Gender: 0 for male, 1 for female.
 - Mosaic group: The Mosaic group is a socio-economic ranking system that builds upon 250-by-250 meter map grid cells covering all the populated areas of Finland. Each map grid contains an average of seven households. The ranking system combines census data with marketing research data to classify the whole population of Finland into nine groups: A, B, C, D, E, F, G, H, and I. Each map grid can be assigned to one of the nine groups. The households living in the same map grid can then be described in terms of socio-demographics, such as education, lifestyle, culture, and behavior.
 - Mosaic class: Based upon the Mosaic group, the Mosaic class divides the nine Mosaic groups further into 33 subclasses.
 - Estimated probability of children: This variable divides households into ten groups of equal size, based upon the probability of them having children living in the same household. A higher value in this variable indicates that the family is more likely to have children living at home. Possible values are from one to ten.
 - Estimated income level: Predicts customers' income level. The higher the value, the wealthier the household is considered. Possible values are one, two and three.
- The behavioral variables consist of the following:
- Loyalty point level: Based on the average spending amount per customer in the corporate chain (the case company is one service provider in the corporate chain), this variable divides customers into five classes: zero, one, two, three, and four. A higher value in loyalty point level is an indication of a customer's larger spending amount in the entire corporate chain.
 - Customer tenure: Number of years since the customer's registration.

- Service level: Measures how many service providers in the corporate chain the customer has used in the last 12 months.
- The spending amount: Records the total spending amount of each customer during the period 2006-2007.

3 The SOM-Ward Model

3.1 Training the SOM-Ward Model

The training of the SOM-Ward Model was carried out using Viscovery SOMine 5.0 (<http://www.viscovery.net/>), which is based upon the batch SOM algorithm [8]. SOMine is a user friendly SOM implementation with a number of analytical tools embedded, including automated two-level clustering using three clustering algorithms, i.e., SOM-Ward, Ward and SOM Single Linkage [24].

To begin with, we preprocessed data to ensure the quality and validity of the clustering result. The Mosaic group is a categorical variable that is not orderly ranked. Since the SOM requires numeric input, we converted the Mosaic group into nine binary variables (either 0 or 1). In addition, the Mosaic class variable was excluded from the SOM-Ward Model as each of the 33 sub-classes of the Mosaic class would have required a dummy variable, which would made visualization extremely difficult.

Assigning a higher priority factor (default is 1) to some variables can be used to give them additional weight and importance in the training process [8], while reducing the priority factor can be used to achieve the opposite. If the priority of a variable is set to zero, the variable has no influence on the training process. We assigned the priority factor of spending amount to 1.4, aiming to give it more influence in the training process. In the pilot tests, it was discovered that the Mosaic group binary variables dominated the segmentation result, leading to clusters exclusively defined by a particular Mosaic group. Therefore, the priority factor of the Mosaic group was set to 0.1. Thus, the Mosaic group data had little influence on the segmentation result, but their distributions in the segments can be investigated when the map has been trained. The priority factors for estimated probability of children and estimated income level were set to 0.5, considering that both variables are based upon estimates and might thus involve some uncertainties. In addition, in order to achieve a more interpretable segmentation result, we slightly adjusted the priority factors of the other variables as well, again based upon the results of the pilot tests. The priority factors of age, gender, service level and customer tenure were adjusted to 1.1, 0.9, 0.9 and 0.8, respectively. Finally, the data were scaled in order to make sure that no variable received undue scale-related bias and to ease the overall training process. Total spending amount and customer tenure were scaled according to range while the rest of the variables were normalized by variance. This step was done automatically by the software. No transformation (e.g., sigmoid or logistic) was applied.

SOMine requires very few parameters for training, mainly because of the batch training process used. The user is only required to provide the *map size*, *map ratio* and *tension* [24]. The default map size is 1,000 nodes. By comparing a set of maps trained in the pilot tests, we chose a map containing 600 nodes to visualize the result. A smaller map is better suited for clustering [25]. The tension parameter is used to

specify the neighborhood interaction. A lower tension will result in a map that adapts more to the data space, resulting in a more detailed map. On the other hand, a higher tension tends to average the data distribution on the map. Based upon the pilot test results, a map generated with the default setting (0.5) was chosen.

3.2 Analysis of the SOM-Ward Model

The characteristics of each segment were identified by examining the variables' component planes (displayed in **Fig. 1**), which show the distributions of each variable across the map. The colors of the nodes in the component planes visualize the value distribution of each variable. Cool colors (blue) indicate low values, while warm ones (red, yellow) indicate high values. Values are indicated by the color scales under the component. For instance, high spending customers were mainly found in Segments One and Two, while long-standing customers are mainly found in Segment Five. In addition to the component planes, two bar charts also illustrate the characteristics of each segment (**Fig. 2** and **Fig. 3**). The height of a bar measures the extent to which the mean value of a variable in a segment deviates from that of the entire data set. The unit of the x-axis is the standard deviation of the entire data set. In this way, both the component planes and the bar charts can visually represent the important characteristics of each segment. A description of each segment follows.

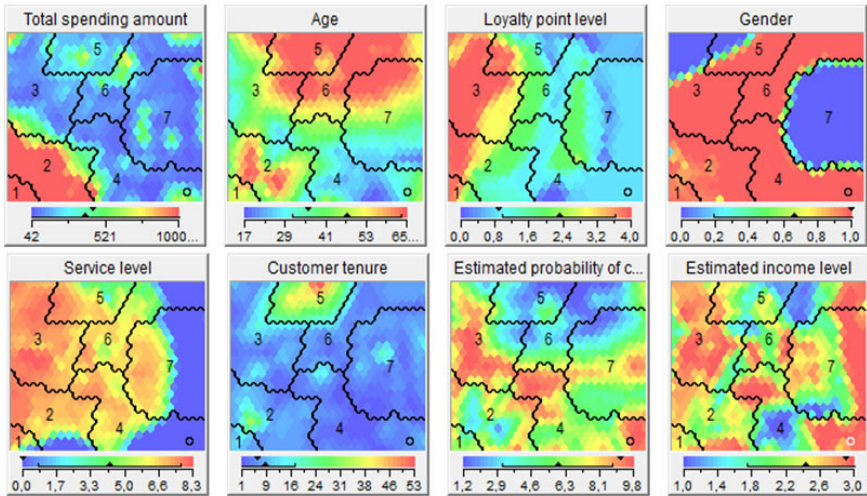


Fig. 1. The component planes of the map

Segment One: Exclusive customers

There are 25,425 customers in Segment One, accounting for 1.7% of the whole customer base. According to **Fig. 1** and **Fig. 2**, the average total purchase amount in this segment is the highest among the seven segments. These customers are mainly female, having a relatively high loyalty point level. **Fig. 3** shows that the customers belonging to this segment are most likely to belong to Mosaic groups A, C, D, and E.

Segment Two: High spending customers

In this segment, there are 177,293 customers, accounting for 12.0% of the customer base. **Fig. 1** and **Fig. 2** show that that most of them, mainly female, are high spending customers. Some of them are around 60 years old, and some have a high loyalty point level. A large percentage of them display a high service level, indicating that they also use many other service providers in the corporate chain. **Fig. 3** reveals that customers belonging to this segment are likely to be from Mosaic groups A, C, D, and E.

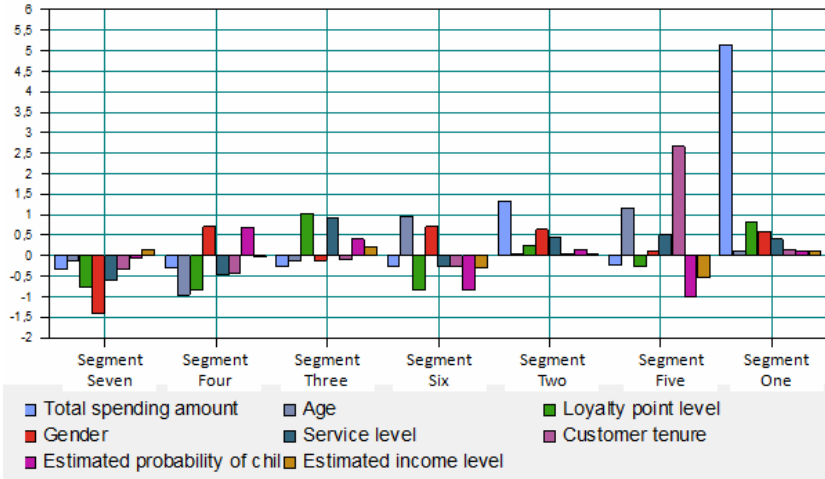


Fig. 2. Bar chart illustrating the characteristics of each segment

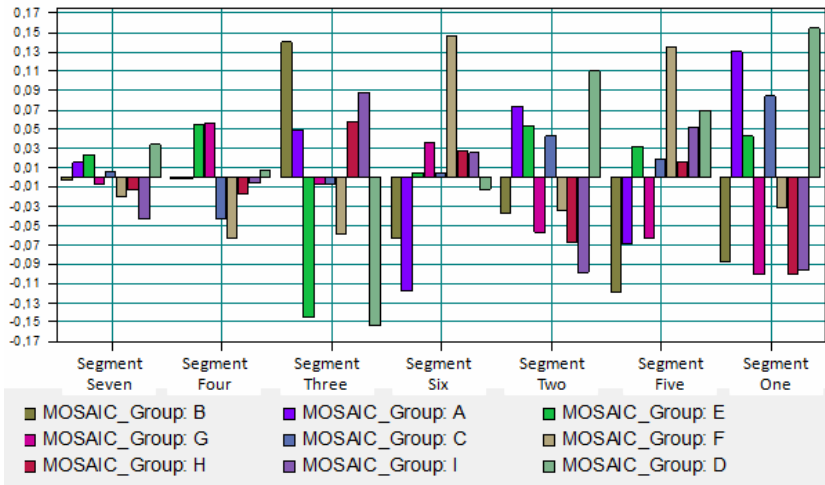


Fig. 3. Bar chart illustrating the proportion of each mosaic group in each segment

Segment Three: Customers with high loyalty point level

There are 283,265 customers in this segment, accounting for 19% of the customer base. **Fig. 1** shows that they have a very high loyalty point level. They use many other service providers in the corporate chain, but their spending amount in the case company is not large. The probability of these customers having children at home is high. **Fig. 3** shows that customers in this segment are likely to be from Mosaic groups B, H, and I.

Segment Four: Relatively young female customers

There are 303,588 customers in this segment, accounting for 20.5% of the customer base. **Fig. 1** and **Fig. 2** show that these customers are mainly females who are much younger than the average of the customer base, and some have a large probability of having children in the same household. However, their spending amount, loyalty point level, service level and customer tenure are below average.

Segment Five: Long-standing customers of the corporate chain

There are 116,537 customers in this segment, accounting for 7.9% of the customer base. **Fig. 1** reveals that these customers have been customers of the corporate chain for a long time. They widely use other service providers in this corporate chain, but their spending amount in the case company is not high. Compared to other segments, these customers are older and their estimated probability of having children living at home is small. However, some of them have a very high loyalty point level. **Fig. 3** indicates that it is likely that these customers belong to Mosaic groups D, F, and I.

Segment Six: Relatively old female customers

Segment Six has 227,194 customers, accounting for 15.3% of the customer base. **Fig. 1** shows that these customers are comparatively senior to those in other segments. Although they are senior in age, they are not long-standing customers of the corporate chain and their spending amount is not large. They are mainly female, and have a low probability of having children living in the same household. **Fig. 3** shows that these customers often belong to Mosaic group F.

Segment Seven: male customers

There are 347,410 customers in Segment Seven, accounting for 23.5% of the customer base. **Fig. 1** shows that these customers are mainly male. They have a low spending amount, and some of them have a high estimated income level.

4 The Decision Tree Model

4.1 Training the Decision Tree Model

The training of the Decision Tree Model is carried out with The PASW Decision Trees module (<http://www.spss.com/statistics/>).

A binary target variable, i.e., the variable of high- and low-spending customers, is created for the Decision Tree Model. From the analysis of the SOM-Ward Model, we found that the customers in Segments One and Two, i.e., those who spend much in the company, account for 13.7% of the customer base. Therefore, we arranged the customers in sequence, from the lowest to highest according to their total purchase

amounts. The top 13.7% of the customers are labeled high-spending customers, and the bottom 13.7% are labeled low-spending customers. Customers in the middle, whose spending amount are not clearly high or low, were excluded from the training set. These customers will be further analyzed in Section 5.

Compared to that of the SOM-Ward Model, the data preprocessing of the Decision Tree Model is much simpler. First, the CART algorithm is able to construct a decision tree model by training continuous and/or categorical predictor variables [26]. Next, the CART algorithm uses surrogates to handle missing values on independent variables [17]. Thus, observations containing independent variables that include missing values are not excluded from the training process. Instead, other independent variables that are highly correlated with the independent variable containing missing values are used for classification. Lastly, the decision tree is relatively insensitive to outliers and skewed data distributions [17]. The above factors reduce the data preprocessing efforts required for training the decision tree. In the Decision Tree Model we used such variables as Mosaic class, which are not easily used in the SOM-Ward Model.

The maximum number of levels in the tree was limited to five and the minimum number of records in a node was set to 1,000, in order to prevent the Decision Tree from becoming very complex. A complex model, from which too many rules are extracted, would not only make the rules hard to generalize into actionable marketing strategies, but would also increase the risk of overfitting. The final model was chosen based upon ten-folds cross validation.

4.2 Analysis of the Decision Tree Model

Appendix 1 shows the created Decision Tree Model. The tree grows from left to right, with the root node (0) located on the left and terminal nodes on the right, with each non-terminal node having two child nodes. The set of two child nodes represents the answers to the decision rules for splitting the records, and these rules are printed on the lines connecting each node to its child nodes. The variable used to split the root node (node 0) is gender. The algorithm compares the classification results produced by all independent variables, and gender is found to lead to the largest improvement of within-node purity. Female customers are in the upper branch of node 0 and male customers are in the lower branch. At node 2, we find that restricting the records to female customers leads the percentage of high-spending customers to increase from 50% to 60.7%. On the other hand, at node 1, restricting the records to male customers leads the percentage of low spending customers to increase from 50% to 77.2%. After the initial split, the decision tree uses loyalty point level to further divide node 1 and node 2 into nodes 3 and 4, and nodes 5 and 6, respectively. The tree shows that a customer with a higher loyalty point level is more likely to be a high-spending customer. For example, when comparing nodes 5 and 6 (the children nodes of node 2), we find that the proportion of high value customers in node 6, i.e., female customers with a loyalty point level above 1, increases compared to that of node 1, while the proportion of high value customers in node 5 decreases compared to that of node 2. The same pattern also appears at the splits of nodes 3, 12, 28 and 30. In addition, the splits at nodes 4, 6, and 26 clearly show that customers belonging to Mosaic groups A, C, D, and E are more likely to spend more. Moreover, the splits at nodes 9 and 13

indicate that customers belonging to Mosaic classes 11, 12, and 17 are more likely to spend more in the company.

As shown in Appendix 1, the process of recursive partitioning does not stop until the tree grows to the terminal nodes. The paths to these terminal nodes describe the rules in the model. The primary focus of this analysis is to identify the characteristics of high- and low-spending customers. Therefore, among all the terminal nodes, we will select four terminal nodes with the highest percentage of high-spending customers, and two terminal nodes with the highest percentage of low-spending customers. The paths from the root node to the six selected terminal nodes are interpreted as characteristics that identify high-spending and low-spending customers.

High-spending customers – Group One: Node 27

This node has 21,526 customers, out of which 92.3% are high-spending customers. The characteristics of the customers in this node are, in order of importance:

1. They are female.
2. Their loyalty point level is larger than 1.
3. They belong to Mosaic classes 11, 12, and 17.

High-spending customers – Group Three: Node 50

This node has 22,172 customers, 83.4% of which are high-spending customers. Their characteristics are:

1. They are female customers.
2. Their loyalty point level is larger than 3. (We combined the rules applied at nodes 2 and 28, because the loyalty point level is used twice.)
3. They belong to Mosaic classes 2, 3, 9, 10, 13, 14, 15, and 16.

High-spending customers – Group Four: Node 19

This node has 1,205 customers, out of which 82.8% are high-spending customers. Characteristics of the customers in this node are:

1. They are male.
2. Their loyalty point level is larger than 3.
3. They belong to Mosaic classes 11, 12, and 17.

Low-spending customers – Group Five: Node 41

This node has 7,144 customers, out of which 98.3% are low-spending customers. Characteristics of the customers in this node are:

1. They are female.
2. Their loyalty point level is less than or equal to 3.
3. They are less than 18.5 years old.

It is also noted that node 23 (the parent node of node 41) also has a very large percentage of low-spending customers. It restricts the age to less than or equal to 20.5.

Low-spending customers – Group Six: Node 31

This node has 7,624 customers, out of which 96.9% are low-spending customers. The characteristics of the customers in this node are:

1. They are male.
2. Their loyalty point level is less than or equal to 1.
3. Their customer tenure with the corporate chain is less than 4.5 years.
4. They are less than 26.5 years old.

It is also noted that node 15 (the parent node of node 31) also has a very large percentage of low-spending customers. However, it has no restriction on age.

5 Customer Portfolio Analysis

Based upon the results of the SOM-Ward and the Decision Tree analyses, we divide the customer base into three groups: high-spending customers, low-spending customers, and customers with development potential. The SOM-Ward model shows that there are seven segments. They are:

1. Exclusive customers
2. High spending customers
3. Customers with high loyalty point level
4. Relatively young, female customers
5. Long-standing customers of the corporate chain
6. Relatively old, female customers
7. Male customers

The map shows that Segments One and Two are high-spending customers. Our purpose is to now identify which of the Segments Three, Four, Five, Six, and Seven have development potential in terms of spending amounts. We will do this by identifying segments that consist of customers displaying similar characteristics as those in Segments One and Two.

We use the decision tree to identify the characteristics that can tell high-spending customers from low-spending ones. The confusion matrix of correct and incorrect classifications in **Table 1** illustrates the accuracy of the decision tree model.

Table 1. The prediction performance of the decision tree

Observed	Predicted		
	Low-spending customers	High-spending customers	Percent Correct
Low-spending customers	125,743	75,685	62,4%
High-spending customers	34,745	166,683	82,8%
Overall Percentage	39.8%	60.2%	72.6%

Table 1 shows that the overall accuracy of the model is 72.6%. 82.8% of the high-spending customers are correctly classified, while only 62.4% of the low spending customers are correctly classified. As our main objective was to build a model that can identify potential high-spending customers, the cost of incorrectly classifying a high-spending customer as a low-spending one is higher than the cost of incorrectly classifying a low-spending customer as a high-spending one. Therefore, this accuracy rate is acceptable.

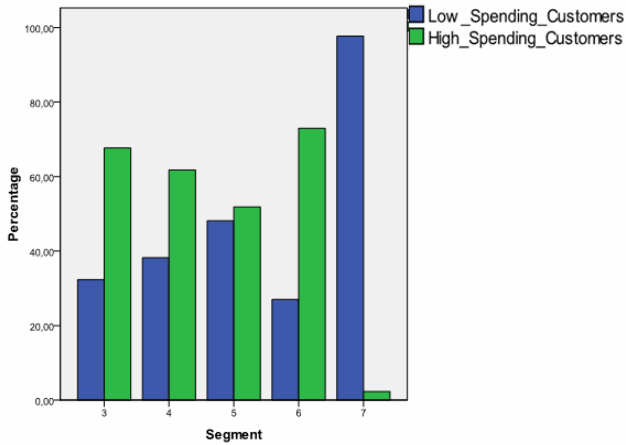


Fig. 4. Percentages of customers in each segment identified as having development potential

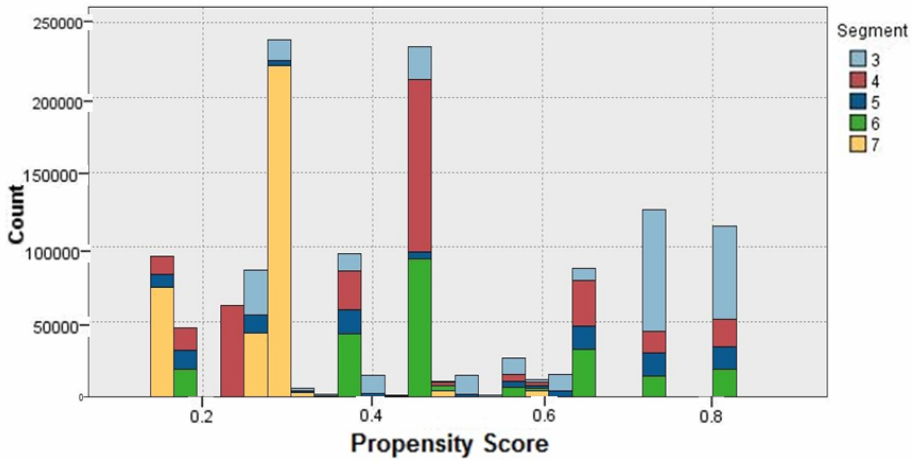


Fig. 5. The propensity scores for the different segments

We then ran all the cases in Segments Three, Four, Five, Six, and Seven through the decision tree model that was created. The ones that we can identify as possessing the same characteristics as the high-spending customers will be our potential group. As the clustered bar chart in Fig. 4 indicates, the customers in Segments Six and Three have more development potential than the customers in the other segments. Each terminal node is a mixture of high-spending customers and low-spending customers. The predicted value is the category with the highest proportion of cases in the terminal node for each case. Therefore, it is possible to use the model to predict these unclassified cases using propensity scores. The propensity score ranks the likelihood of the prediction from 1 (likely to be a high-spending customer) to 0 (not likely to be a high-spending customer). For example, if an unclassified case has the same

characteristics as node 27 in the decision tree model, it will be assigned a propensity score of 0.923, as 92.3% of the cases in node 27 are high-spending customers. The Histogram of the distribution of propensity scores of Segments Three, Four, Five, Six, and Seven is shown in **Fig. 5**.

This figure reveals that the customers in Segment Three are most likely to be potential high value customers, while the customers in Segment Seven are least likely to be potential high value customers. After running all of the data outside of the 27.4% (high and low spending) that were used to train the decision tree, we obtain a list of those customers with their propensity scores or/and predictions appended.

6 Conclusions

A hybrid approach combining unsupervised and supervised data mining techniques has been proposed to conduct customer portfolio analysis. SOM-Ward clustering was first used to conduct customer segmentation. Then, the decision tree was employed to gain insight into whether there are significant determinants for distinguishing between high- and low-spending customers. The results of the two models are then compared and combined to perform customer portfolio analysis, i.e., to identify the high- and low-spending customers, as well as customers with development potential. Each model possesses advantages and disadvantages of its own.

As an unsupervised data mining technique, SOM-Ward clustering is a good tool for exploratory analysis, as is the case when no *a priori* classes have been identified. The SOM is a very visual tool and possesses strong capabilities for dealing with non-linear relationships, missing data, and skewed distributions. However, while the clusters produced using unsupervised methods may be good for gaining an understanding of the customer base, they are not necessarily actionable in terms of marketing strategy as they are not based upon any identified target or aim. In addition, using detailed nominal data (e.g., 33 Mosaic classes) is a problem when using the SOM, as binary variables must be constructed for each potential class. This easily clutters the map and heavily influences training results.

Decision trees, on the other hand, are tailored to a specific purpose by using a supervised learning approach. The decision tree is also a very robust method, easily capable of dealing with difficult data, and requires less data preprocessing and setting of parameters than the SOM. However, the starting point of supervised learning inevitably requires more *a priori* knowledge than unsupervised learning, making the knowledge gained using the SOM potentially very important.

The results of the analysis demonstrate that the combined method of the SOM-Ward clustering and the Decision Tree can potentially be effective in conducting market segmentation. The information provided by the combined model is more detailed and accurate than that provided by either model used alone, thus more actionable information about the customer base for marketing purposes could be retrieved.

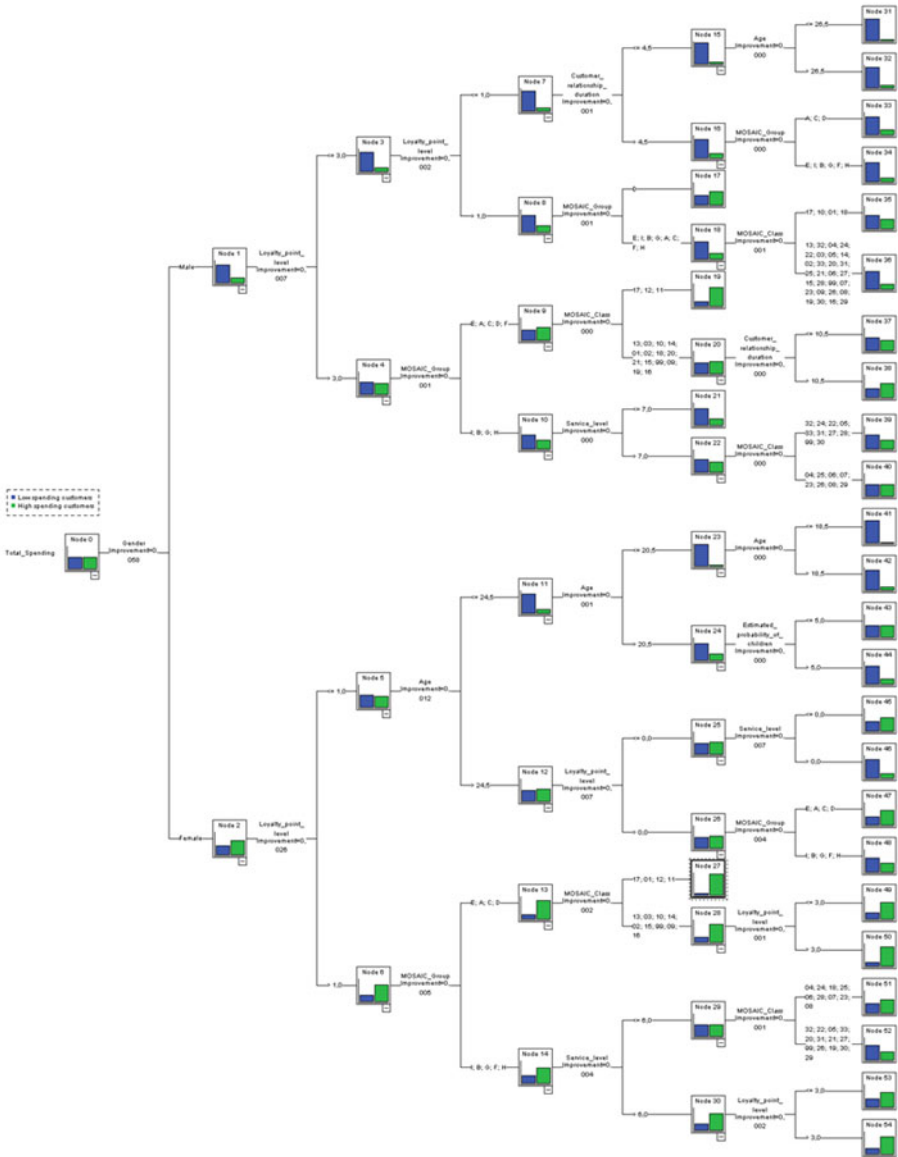
Acknowledgements. The authors gratefully acknowledge the financial support of the National Agency of Technology (Titan, grant no. 40063/08) and the Academy of Finland (grant no. 127656). The case organization's cooperation is also gratefully acknowledged.

References

1. Kim, S., Jung, T., Suh, E., Hwang, H.: Customer Segmentation and Strategy Development Based on Customer Lifetime Value: A Case Study. *Expert Systems with Applications* 31, 101–107 (2006)
2. Park, H., Baik, D.: A Study for Control of Client Value using Cluster Analysis. *Journal of Network and Computer Applications* 29, 262–276 (2006)
3. Reichheld, F.F., Teal, T.: *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*. Harvard Business Press, Boston (2001)
4. Buttle, F.: *Customer Relationship Management Concepts and Tools*. Butterworth-Heinemann, Oxford (2004)
5. Terho, H., Halinen, A.: Customer Portfolio Analysis Practices in Different Exchange Contexts. *Journal of Business Research* 60, 720–730 (2007)
6. Kohonen, T.: *Self-Organization and Associative Memory*. Springer, New York (1988)
7. Ward, J.H.: Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, 236–244 (1963)
8. Kohonen, T.: *Self-Organizing Maps*. Springer, Berlin (2001)
9. Kaski, S., Kangas, J., Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers 1981–1997. *Neural Computing Surveys* 1, 102–350 (1998)
10. Oja, M., Kaski, S., Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers: 1998–2001 Addendum. *Neural Computing Surveys* 3, 1–156 (2003)
11. Deboeck, G.J., Kohonen, T.: *Visual Explorations in Finance with Self-Organizing Maps*. Springer, Berlin (1998)
12. Holmbom, A.H., Eklund, T., Back, B.: Customer Portfolio Analysis using the SOM. In: *Proceedings of 19th Australasian Conference on Information Systems (ACIS 2008)*, pp. 412–422 (2008)
13. D’Urso, P., Giovanni, L.D.: Temporal Self-Organizing Maps for Telecommunications Market Segmentation. *Neurocomputing* 71, 2880–2892 (2008)
14. Lee, S.C., Gu, J.C., Suh, Y.H.: A Comparative Analysis of Clustering Methodology and Application for Market Segmentation: K-Means, SOM and a Two-Level SOM. In: Zhong, N., Ras, Z.W., Tsumoto, S., Suzuki, E. (eds.) *ISMIS 2006*. LNCS (LNAI), vol. 4203, pp. 435–444. Springer, Heidelberg (2006)
15. Vesanto, J., Alhoniemi, E.: Clustering of the Self-Organizing Map. *IEEE Trans. Neural Networks* 11, 586–600 (2000)
16. Samarasinghe, S.: *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*. Auerbach Publications, Boca Raton (2007)
17. Berry, M.J.A., Linoff, G.S.: *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Wiley Publishing Inc., Indianapolis (2004)
18. Murthy, S.K.: Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. *Data Mining and Knowledge Discovery* 2, 345–389 (1998)
19. Rokach, L., Maimon, O.: *Data Mining with Decision Trees: Theory and Applications*. World Scientific Publishing, Singapore (2008)
20. Fan, G.Z., Ong, S.E., Koh, H.C.: Determinants of House Price: A Decision Tree Approach. *Urban Studies* 43, 2301–2315 (2006)
21. Abrahams, A.S., Becker, A.B., Sabido, D., D’Souza, R., Makriyannis, G., Krasnodebski, M.: Inducing a Marketing Strategy for a New Pet Insurance Company using Decision Trees. *Expert Syst. Appl.* 36, 1914–1921 (2009)

22. Sheu, J.J., Su, Y.H., Chu, K.T.: Segmenting Online Game Customers - the Perspective of Experiential Marketing. *Expert Syst. Appl.* 36, 8487–8495 (2009)
23. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: *Classification and Regression Trees*. Wadsworth, Pacific Grove (1984)
24. Deboeck, G.J.: Software Tools for Self-Organizing Maps. In: Deboeck, G.J., Kohonen, T. (eds.) *Visual Explorations in Finance using Self-Organizing Maps*, pp. 179–194. Springer, Berlin (1998)
25. Desmet, P.: Buying Behavior Study with Basket Analysis: Pre-Clustering with a Kohonen Map. *European Journal of Economic and Social Systems* 15, 17–30 (2001)
26. Hill, T., Lewicki, P.: *Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining*. StatSoft, Inc., Tulsa (2006)

Appendix 1. The Decision Tree Model



Research Paper 2

Yao, Z., Eklund, T., & Back, B. (2010), Using SOM-Ward Clustering and Predictive Analytics for Conducting Customer Segmentation, In *Proceedings of the 10th IEEE International Conference on Data Mining Workshops (ICDMW 2010)*, Sydney, Australia, 14-17 December, 2010, pp. 639-646.

Using SOM-Ward clustering and predictive analytics for conducting customer segmentation

Zhiyuan Yao^{1,2}, Tomas Eklund¹, Barbro Back¹

1. Department of Information Technologies, Åbo Akademi University, Turku, Finland
E-mail: {zyao, toeklund, bback}@abo.fi
2. Turku Centre for Computer Science (TUUS), Turku, Finland

Abstract— Continuously increasing amounts of data in data warehouses are providing companies with ample opportunity to conduct analytical customer relationship management (CRM). However, how to utilize the information retrieved from the analysis of these data to retain the most valuable customers, identify customers with additional revenue potential, and achieve cost-effective customer relationship management, continue to pose challenges for companies. This study proposes a two-level approach combining SOM-Ward clustering and predictive analytics to segment the customer base of a case company with 1.5 million customers. First, according to the spending amount, demographic and behavioral characteristics of the customers, we adopt SOM-Ward clustering to segment the customer base into seven segments: exclusive customers, high-spending customers, and five segments of mass customers. Then, three classification models - the support vector machine (SVM), the neural network, and the decision tree, are employed to classify high-spending and low-spending customers. The performance of the three classification models is evaluated and compared. The three models are then combined to predict potential high-spending customers from the mass customers. It is found that this hybrid approach could provide more thorough and detailed information about the customer base, especially the untapped mass market with potential high revenue contribution, for tailoring actionable marketing strategies.

Keywords- customer segmentation; predictive analytics; self-organizing map (SOM); Ward's clustering; support vector machine (SVM); neural network (NN); decision tree

I. INTRODUCTION

Companies have long been diverting their attention from products to customers [1]. The Pareto principle that 20% of the customers create 80% of the profit or revenue [2-4] is commonly agreed upon in the industry. Reichheld and Teal [5] also claim that an increase in customer retention would result in a significant rise in company profit. Hence, companies should try to develop their analytical customer relationship management (CRM) to identify valuable customers, and customers with growth potential, to increase the aggregate value of their customer base. The application of enterprise resource planning (ERP) systems and customer data warehouses present companies with an opportunity to use data mining techniques to convert data and information into knowledge with the aim of improving customer relationships and facilitating decision making. Customer segmentation is an important part of analytical CRM, in

which the customer base is divided into different groups based on similarity [6]. Effective customer segmentation enables companies to interact with customers in each segment collectively, and allocate limited resources to various customer segments according to corporate strategies.

Data mining techniques, cluster analysis in particular, could assist companies in conducting customer segmentation [4, 7]. Cluster analysis is a collection of statistical and machine learning methods capable of dividing heterogeneous data into multiple more homogeneous clusters in a data-driven way, and it is often the first step in customer segmentation [7]. In the present study, we propose a two-step approach to carrying out customer segmentation. We first adopt the SOM-Ward method [8] (i.e., the Self-Organizing Map (SOM) [9] combined with Ward's clustering [10]) to conduct cluster analysis of a customer base. The customer base is divided into distinct groups of customers with similar characteristics and behavior, which allows us to specify the characteristics that distinguish high-spending customers from mass customers. Then, three classification models - the support vector machine (SVM), the artificial neural network (ANN) and the decision tree - will be used respectively to classify high-spending and low-spending customers. Although the three models have been applied in marketing individually, their suitability and performance in marketing have not been fully understood, nor has a comparison been examined in detail. In the present study, we will evaluate and compare the performance of the three methods. Furthermore, the three models combined will be used to reveal potential mass customers who share similar characteristics with high-spending customers, thus pinpointing the segments with high-value potential. By extension, this type of analysis allows companies to take into consideration the current value and potential migration pattern of the customers concurrently while creating marketing strategies. By doing so, companies could not only be more focused in maintaining the relationship with those high-spending customers but could also be more effective in launching cross-selling, up-selling, and deep-selling campaigns among mass customers.

The remainder of this paper is organized as follows. Section II introduces the methodology used in the study, i.e., SOM-Ward clustering and the three classification methods for conducting predictive analytics. Section III presents the data used in this study. Section IV describes the training and analysis of the SOM-Ward model. Section V documents the training process of the three classification models and the classification results; the performance of the three models is

evaluated and compared. In Section VI, the trained classification models are adopted to analyze mass customers to identify their market potential. Finally, in Section VII, a conclusion is drawn and the empirical comparison of the three classification models is discussed.

II. METHODOLOGY

A. The SOM and SOM-Ward clustering

As a widely used unsupervised neural network for clustering tasks [11, 12], the SOM has been applied as an analytical tool in many industry applications [13-15], including market segmentation [12-14]. It is capable of projecting the relationships between high-dimensional data onto a two-dimensional display, where similar input records are located close to each other [11]. By adopting an unsupervised learning paradigm, the SOM conducts clustering tasks in a completely data-driven way [11, 16], i.e., it works with little *a priori* information or assumptions concerning the input data. In addition, the SOM's capability of preserving the topological relationships of the input data and its excellent visualization features motivated the authors to apply it in the present study.

A SOM is typically composed of two layers: an input and an output layer. Each input field is connected to the input layer by exactly one node, which is fully connected with all the nodes in the output layer [7, 17]. When the number of nodes in the output layer is large, the adjacent nodes need to be grouped to conduct clustering tasks. Accordingly, Vesanto and Alhoniemi proposed a two-level approach [18], e.g., the SOM-Ward clustering, to perform clustering tasks. The dataset is first projected onto a two-dimensional display using the SOM, and the resulting SOM is then clustered. Several studies [19-21] have shown the effectiveness of the two-level SOM, especially the superiority of the SOM-Ward over some classical clustering algorithms.

As mentioned previously, the SOM-Ward clustering is a two-level clustering approach that combines local ordering of the SOM and Ward's clustering algorithm to determine the clustering result. Ward's clustering is an agglomerative (bottom-up) hierarchical clustering method [10, 22]. The SOM-Ward starts with a clustering where each node is treated as a separate cluster. The two clusters with the minimum Euclidean distance are merged in each step, until there is only one cluster left on the map. The distance follows the SOM-Ward distance measure, which takes into account not only the Ward distance but also the topological characteristics of the SOM. In other words, the distance between two non-adjacent clusters is considered infinite, which means only adjacent clusters can be merged. A low SOM-Ward distance value represents a more natural clustering for the map, whereas a high value represents a more artificial clustering [8]. In this way, the users can flexibly choose the most appropriate number of clusters for their data mining tasks.

B. Predictive Analytics

Predictive analytics refers to a variety of techniques that deal with the prediction of future events by analyzing

historical and current data [23, 24]. A prediction model starts by applying data mining techniques to historical data, trying to search for relationships between explanatory variables and a response variable. Once created and validated, the model could be used for a new dataset that shares the explanatory variables, to predict the possibility of a response variable. If the response variable is categorical, one should choose classification algorithms to construct a prediction model [23]. Three classification algorithms, i.e., the SVM, the ANN and the decision tree, selected for the present study, are elaborated upon in the following subsections. After the performance of the three models has been evaluated, they will be combined into an ensemble model for more accurate predictive analytics. We used three ensemble methods: voting, confidence-weighted voting and highest confidence win. In voting, the number of times each possible target value appears is summed and the one with the highest total is chosen as the prediction. Confidence-weighted voting works in similar way as voting except the confidence of prediction is taken into account and the votes are weighted by the confidence. In "confidence-weighted voting", the prediction with the highest confidence is chosen as the prediction of the ensemble model.

1) Support Vector Machine

The SVM, introduced by Vapnik [25], is a kernel-based method capable of conducting classification and regression tasks [26]. The use of kernel transformation effectively overcomes the problem of the "curse of dimensionality" [27], which enables the application of the SVM to a wide range of datasets. Furthermore, the SVM is based on structural risk minimization (SRM), i.e., a principle for model selection based on a trade-off between model complexity and training error [28]. Therefore, it guarantees a global unique optimal solution and can reduce the risk of overfitting [26]. All the advantages mentioned above have made the SVM an extensively applied technique in many industries [29].

In classification tasks, the SVM works by transforming input data using the kernel function into a high-dimensional feature space in which the classes of the data can be separated by a hyperplane. This hyperplane can be used to predict which category the new data belong to. In addition, on each side of the hyperplane, the SVM locates the maximum-margin hyperplanes, i.e., two parallel hyperplanes that maximize the distance between the data classes. The larger the distance between the maximum-margin hyperplanes is, the less likely the model will be prone to overfitting [26]. Readers can refer to [27, 30] for technical details and the algorithm of the SVM.

2) Artificial Neural Network

As its name implies, an ANN is designed to mimic the architecture of the human brain in a simplified way, to process information and learn from examples. ANNs have been widely applied in many business areas. In most of those studies, the ANN exhibits a performance as good as, if not better than, that of other methods [31]. The Back-propagation (BP) neural network, one of the currently most widely used neural network algorithms [32], is chosen for

our classification tasks in the present study. As a type of a supervised ANN, the BP neural network functions by adjusting the weight among network nodes by constantly reducing the difference between predicted and actual values during the network training process. This process of feeding training examples and updating weights can iterate many times until the overall error is minimized. Thus, the network gradually becomes capable of understanding how inputs affect outputs. Once the network has been trained, its prediction capability should be evaluated using a test set, after which the network can be applied to predict outcomes for unknown examples. The algorithm of the BP neural network is well-known, so there will be no further explanation of it in this paper. For more details on the algorithm, readers could refer to [32].

3) *Decision Tree*

A decision tree is another classification model that can divide data into smaller and more homogeneous sets by recursively applying two-way and/or multi-way splits [7]. The greatest advantage the decision tree possesses is that it can produce a transparent predictive model, i.e., its output is a set of straightforward and explainable rules describing the relationships between explanatory variables and the response variable. Moreover, the decision tree can identify the significant variables for predicting the response variables; it is relatively insensitive to outliers; and it can deal with missing values [7, 33, 34]. In this study, we use the C5.0 decision tree [35] (modified and improved after the well-known C4.5 [36]) with the boosting algorithm [37] to classify high- and low-spending customers. The C5.0 boosting model works by constructing multiple decision tree models. The initial model is trained as usual, and in each of the following boosting rounds, all models are built in such a way that they focus on the records that were incorrectly classified by the previous models. Finally, records are classified by combining the weighted classification results of individual models, according to each model's performance, into one overall classification.

III. THE DATA

The data used in this study are from a service provider that belongs to a large, multiservice Finnish corporation. Through a loyalty card system, the corporation provides customers with various discounts and rewards based on the bonus points accumulated. Personal information of the cardholders is collected when they apply for the card, and their transactions are recorded in the system. The dataset with a total number of 1,522,701 customers was obtained through the loyalty card system. It contains aggregated sales information from several branches of the service provider in Finland, for the period 2007-08. The dataset consists of ten variables that fall into two bases: demographic and behavioral variables.

The demographic variables are shown as follows.

- Age
- Gender: 0 for male, 1 for female.
- Mosaic group: Mosaic is a household-based socio-economic ranking system developed by Experian

PLC that classifies all Finnish households and neighborhoods (250-by-250 meter map grid cell) into one of nine unique segments. Each segment is described by demographic information, cultural/ethnic composition, lifestyle, purchase habits and so on.

- Mosaic class: Based upon the Mosaic group, the nine Mosaic groups are further divided into 33 subclasses.
- Estimated probability of children: Based upon the probability of their having children at home, this variable divides households into 10 groups of equal size. The higher the value of this variable is, the more likely there are children living in the households. The value ranges from 1 to 10.
- Estimated income level: This variable predicts customers' income level. The higher the value, the wealthier the household is considered to be. Possible values are 1, 2 and 3.

The behavioral variables are shown as follows.

- Loyalty point level: Based on the average purchase amount per customer in the corporate chain (the case company is one service provider in the corporation), this variable divides customers into five classes: 0, 1, 2, 3, and 4. A higher value in loyalty point level indicates a larger spending amount in the entire corporate chain.
- Customer tenure: This indicates how many years the customer has been a cardholder.
- Service level: The variable measures the number of service providers in the corporate chain that the customer has used in the last 12 months.
- The spending amount: The variable shows the total spending amount of each customer during the period 2007-08.

IV. THE SOM-WARD MODEL

A. *Training of the SOM-Ward model*

Viscovery SOMine 5.0, which builds upon the batch SOM algorithm [8], was used to train the SOM-Ward Model. As a user-friendly SOM implementation, SOMine includes three alternative clustering algorithms: SOM-Ward, Ward and SOM Single Linkage [15].

First, we preprocessed the data to ensure the quality and validity of the clustering result. Since the SOM requires a numeric input, we converted the Mosaic group into nine binary dummy variables (either 0 or 1). In addition, the Mosaic class variable was excluded from the SOM-Ward model as each of the 33 sub-classes of the Mosaic class would have required a dummy variable. This would have made visualization extremely difficult.

All the variables included in the training process were scaled to comparable ranges in order to prevent variables with large values from dominating the result. Viscovery SOMine offers two forms of scaling, linear and variance scaling. Linear scaling is simply a linear scaling based upon the range of the variable, and is suggested as default when

the range of the variable is greater than eight times its standard deviation. Otherwise, variance scaling, i.e., the well-known normalization, is used. In this study, range scaling was applied to the variables spending amount, customer tenure and Mosaic groups, while variance scaling (normalization) was applied to the others.

For certain analysis purposes, some variables possess a higher priority than others do. Assigning a higher priority factor to variables gives them additional weight and importance in the training and segmentation processes. By default, the priority factor value is set to 1. The importance of a variable is reduced if its priority factor is set to less than 1, and accordingly, a variable with a priority factor value of 0 has no influence on the training process. The model is intended to explore the relationship between customers' spending amounts and the other characteristics; therefore, we decided to give the variable spending amount more weight in the training process. We assigned its priority value to 1.3 through experimenting with different values. In the pilot tests, it was discovered that the Mosaic group binary variables dominated the cluster formation, leading to clusters exclusively defined by some particular Mosaic group, and therefore, a biased segmentation result. To avoid this, we set the priority factor of the Mosaic group variable to 0.1 to ensure that the Mosaic group data had little impact on the segmentation result but that their distributions in the segments could still be investigated while training the map. Because the variables of the estimated probability of children and estimated income level both involved some estimates, their priority factor values were set to 0.5. In view of the results of the pilot tests, the priority factor values of other variables were somewhat adjusted to obtain a more interpretable segmentation result. The priority factors of age, gender, and customer tenure were set to 1.1, 0.9 and 0.8, respectively.

B. Analysis of the SOM-Ward model

The resulting SOM-Ward model consists of seven clusters. The component planes (Fig. 1) show the distributions of each variable across the map, on which the color scale visualizes the distribution of each variable over different segments. Cold colors indicate low values, while warm ones indicate high values, e.g., high-spending customers were mainly found in Segments Six and Seven, while long-standing customers were mainly found in Segment Five. Apart from the component planes, a series of error bar charts (Fig. 2) demonstrate the characteristics of each variable of the seven clusters. In each error chart there is a horizontal reference line, indicating the mean value of the variable in the whole customer base. The mean value of the variable in each segment is represented by a circular marker, and the 95% confidence interval of a variable's mean for each segment is represented by another two parallel red lines. For Mosaic group, we used a bar chart (Fig. 3) to illustrate its distributions in the seven clusters. The nine Mosaic groups are represented through the bars that are displayed in the same sequence in each segment. The height of a bar measures the extent to which the mean value of a variable in a cluster deviates from that of the entire data set.

The unit of the y-axis is the standard deviation of the entire data set. Specifically, the green and yellow bars (the first and second bar) in Cluster One which are below zero level, show that the proportion of customers belonging to Mosaic groups B and E in Cluster One is less than that of the customer base in general. The key figures and important characteristics of each segment are summarized in Table IV.

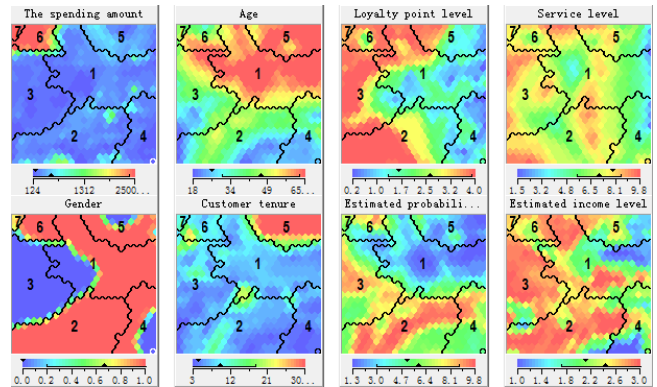


Figure 1. The component planes of the map.

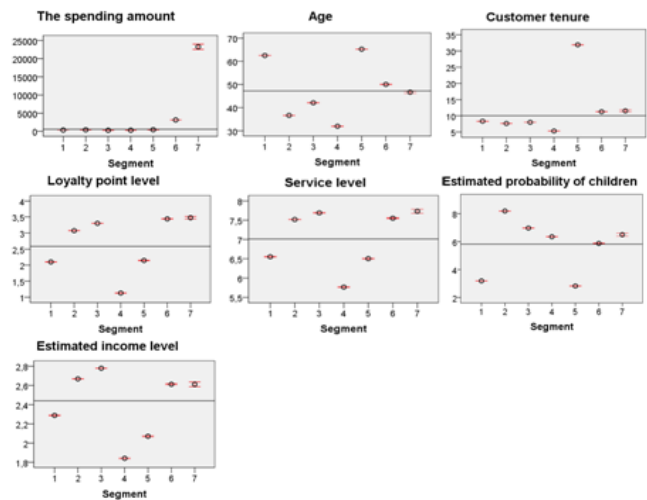


Figure 2. The error bar charts for each variable.

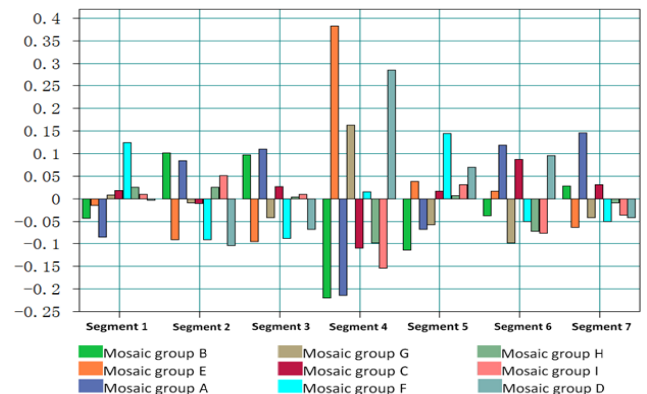


Figure 3. The bar chart illustrating the distributions of Mosaic groups in the seven segments.

V. THE CLASSIFICATION MODELS

A. Creation of the binary target variable

A binary target variable, i.e., the variable of high- and low-spending customers, was created for the classification models. It was revealed from the analysis of the SOM-Ward model that high-spending customers were mainly distributed in Segments Six and Seven, especially in the shaded area in Figure 4, representative of 72,337 customers. After eliminating the records containing missing values, there were altogether 66,535 customers left, whom we labeled as high-spending customers. In SPSS Modeler, SVM only works with non-missing records, records with missing values for any input or output field are excluded from the estimation of the model. In ANN, missing values are handled by substituting neutral values for the missing ones. For this reason, we decided to remove the records with missing values so as to be able to compare all models regardless of the constraints of one modeling technique. Then we conducted an RFM analysis of Segments One to Five composed of mass customers. RFM analysis is a well-known method for analyzing customer purchase behavior and their tendency for buying more products [4]. Here, R (recency) represents the period since the last transaction. The lower the value of R is, the more likely the customers are to purchase again. Therefore, we used reverse recency, i.e. 365×2 -recency to calculate the RFM score. F (frequency) represents the number of times customers purchased within a certain period, and M (monetary value) represents the spending amount of customers within a certain time period. To calculate the RFM score, we needed to assign different weights to the three RFM attributes. We found that the purchase amount of most low-spending customers was no more than 15 Euros, and that they had made only few transactions in two years, but their reverse recency varied from 1 to 730. Therefore, to obtain an unbiased result, we set the weight of recency to 0.01, and frequency and monetary values to 1, respectively. By doing this, we prevented the customers with high reverse recency from dominating the RFM scores. After working out the RFM scores, we selected from each segment the records with the lowest RFM score and with no missing values. From the RFM analysis, we located those customers with low spending amount and low potentiality of purchasing in the near future. The number of customers selected from each segment was proportional to the segment size, and customers with a low RFM score, whom we labeled as low-spending customers, added up to 66,535. We have thus constructed a dataset that contained an equal number of high-spending and low-spending customers. This balanced dataset enabled the model to identify important relationships in the underlying data [24]. In this dataset, the average purchase amount of high-spending customers was 4,649 Euros, while that of low-spending customers was 6 Euros.

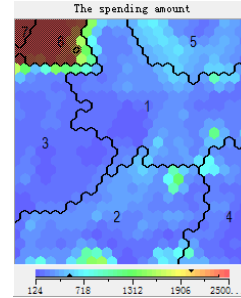


Figure 4. The high-spending customers (shaded area) chosen for the binary target variable.

B. Training of the classification models

The training of the three classification models was implemented using SPSS Modeler 13. In the classification models, we decided to include the Mosaic class in the training process because the three models can handle nominal data well. Since the Mosaic group gives the same, but less detailed, information as Mosaic class, we decided not to use it in this part of analysis. For SVM, SPSS Modeler provides four types of kernel functions: linear function, polynomial function, radial basis function (RBF), and sigmoid function. In this study, we used RBF, which is considered a natural choice when employing SVM [24]. The training of SVM using other kernel types takes an extremely long time. We experimented with different pairs of values for regularization parameter (C) and RBF gamma (γ). The higher the values of these two parameters are, the higher the classification accuracy of the training data is, but at the same time the risk of overfitting is higher. The prediction accuracies of the models were estimated by a ten-fold cross-validation and the model with the highest accuracy rate was chosen.

In employing the ANN, we used a multilayer perceptron with two hidden layers to explore the relationship between the predictors and target variables. The number of hidden layers and the units in each of them are automatically determined by the software. It starts with a large network and prunes the weakest units. By default, the network training process stops if there is no improvement between cycles. Users can also preset the desired accuracy rate and training time as stopping criteria. The default stopping criteria made the training last long because of the large dataset, so we preset training cycles as the stopping criteria. After experimenting with different training cycles, we chose a model with 1,000 training cycles being the stopping criteria according to the accuracy rate of cross-validation.

As mentioned in Section II, we used the boosting method to improve the prediction accuracy of the decision tree. Here we employed 10 boosting rounds, i.e., the method combining 10 base models of C5.0 that complement each other, with each model focusing on the records incorrectly handled by previously constructed ones. We also pruned the trees to prevent overfitting.

The accuracies of the three models based on resubstitution and ten-fold cross validation, and area under the ROC curve (AUC) are listed in Table I. In resubstitution,

all records are used in training and the accuracy is also based on these data. Therefore, the result could be overoptimistic. AUC is a criterion used to measure the accuracy of the model in discriminating between target variable values. If the value of AUC is 0.5, it shows that the model does not perform better than guessing; if the value of AUC is 1, it shows that it is a perfect model. We found that the estimated prediction accuracy rates of the three models were high and close to each other. Though the accuracy rate of the boosted decision tree is higher than that of the other models, it tends to overfit the data whereas the support vector machine has the lowest risk of overfitting.

TABLE I. THE COMPARISON OF THE THREE MODELS.

Classification models	Accuracy (%) (Resubstitution)	Accuracy (%) (Cross-validation)	Area under the ROC curve
Support vector machine	81.25	80.79	0.89
Neural network	80.31	79.60	0.87
Boosted decision tree	82.49	80.58	0.88
Decision tree without boosting	81.73	80	0.86

VI. PREDICTING POTENTIAL HIGH-SPENDING CUSTOMERS

The three models established earlier were then combined into an ensemble model to predict the segments with development potential among those composed primarily of mass customers. SVM cannot generate output probability for records with missing values in SPSS Modeler, which uses the method combining sigmoid and SVM [38], so it cannot be used alone for prediction tasks. We thus decided to incorporate SVM, ANN, and the boosted decision tree to predict the purchasing potential of mass customers. As discussed in Section II, we adopted three ensemble methods: voting, confidence-weighted voting, and highest confidence win. The three methods were used to classify the training data, and the accuracies of the three ensemble methods were compared. The method with the highest accuracy was then used to predict mass customers' potentiality. The confidence here refers to the posterior probability estimate for the predicted target class. As mentioned, the standard SVM does not generate a confidence value, which was estimated using a SVM plus sigmoid method here [38]. The backpropagation ANN uses the sigmoid function and the output activation values range from 0 to 1, with 0.5 being the cutoff value. The confidence was calculated as output activation value, which is twice the absolute difference value of 0.5. In the decision tree, each terminal node was a mixture of high-spending customers and low-spending customers. The predicted value of an unknown record was the category with the highest percentage of cases in the terminal node it belongs to. This percentage value was then used as the confidence value. For the boosted decision tree, the confidence values from each base model were then combined, while each model was

assigned a weight that was proportional to its performance. Table II lists the overall accuracy of the three ensemble methods, and accuracy rates of predicting high-spending and low-spending customers. It was discovered that the three models performed better in classifying the high-spending customers than low-spending customers. Furthermore, the three ensemble methods slightly raise the classification accuracy rate of SVM and ANN on the training data. Although their overall accuracy is slightly lower than that of the boosted decision tree, the latter is more prone to overfitting data. For this reason, we decided to use the ensemble method with the highest overall accuracy, i.e., the confidence-weighted voting, to predict mass customers and locate those customers with development potential.

TABLE II. THE COMPARISON OF THE THREE ENSEMBLE METHODS.

Ensemble method	Training data		
	Overall accuracy (%)	Accuracy of predicting high-spending customers (%)	Accuracy of predicting low-spending customers (%)
Voting	82.37	84.83	79.92
Confidence-weighted voting	82.39	84.22	80.55
Highest confidence win	82.29	83.46	81.11

We then ran all the cases in Segments One to Five through the ensemble model using confidence-weighted voting. As for each customer, every classification model generates a confidence value for the predicted target class. The ensemble model aggregates the confidence values of the predicted classes generated by the classification models. The predicted class of the ensemble model is the one with the highest aggregate confidence value and its output probability is the aggregated confidence value of the class predicted divided by three. Through this output probability, we could obtain the propensity score of a customer being a high-spending customer. If the predicted result of the ensemble model is a high-spending customer, then the propensity score is the same as the output probability of the ensemble model; if the predicted result of the ensemble model is a low-spending customer, then the propensity score is the output probability subtracted by one. Table III lists the present value rank and the average propensity score for being high-spending customers for each segment, which could be used as an index for measuring customer potential. Analyzed together with Table IV, we could find out that customers in Segment Two were most likely to be potential high-spending customers while customers in Segment Four were the least likely. Though customers in Segment Two have smaller purchasing amounts than those in Segment Five, they possess a higher propensity score. Therefore, Segment Two plays no less important role than Segment Five does in the process of devising market strategies. Moreover, this result is visually confirmed in Figure 2, where Segment Two is most similar to Segments Six and Seven based on the variables shown. The average values of customer tenure, loyalty point

level, service level and estimated income level of Segment Two are close to those of Segments Six and Seven.

TABLE III. THE PROPENSITY SCORES FOR SEGMENTS WITH MASS CUSTOMERS

Segment ID	Value rank	Propensity score
7	Exclusive customers	
6	High-spending customers	
2	Mass customers	0.49
5	Mass customers	0.39
1	Mass customers	0.39
3	Mass customers	0.28
4	Mass customers	0.20

VII. CONCLUSION

A two-step approach combining the SOM-Ward clustering and predictive analytics has been proposed to conduct customer segmentation. SOM-Ward clustering is first used to divide customers into exclusive customers, high-spending customers and mass customers, with each segment possessing different demographic and behavioral characteristics. Because mass customers occupy a great percentage of the entire customer base, we continue by using classification models that are capable of distinguishing between high- and low-spending customers to identify those mass customers with development potential.

Based on an unsupervised neural network, the SOM-Ward clustering is a useful tool for exploratory data analysis, as in the case when no *a priori* classes have been identified. The SOM is a visual tool and possesses strong capabilities of dealing with non-linear relationships, missing data, and skewed distributions. However, while the clusters produced using unsupervised methods may warrant a good understanding of the current customer base, they cannot necessarily provide information about the potential value of a segment.

Predictive analytics, on the other hand, are tailored to a specific purpose through a supervised learning approach. The three classification models used in the study can effectively distinguish between high- and low-spending customers and identify mass customers who have similar characteristics as high-spending customers. Companies could thus efficiently manage customer segment migration with the provided actionable information. However, the starting point of predictive analytics inevitably requires more *a priori* knowledge than unsupervised learning does, making the knowledge gained in using the SOM-Ward model potentially significant.

The results of the analysis demonstrate that the combined method of SOM-Ward clustering and predictive analytics can potentially be effective in conducting customer segmentation. The results of market segmentation, incorporated with the results of predictive analytics, are more prospective and predictive. Thus more actionable information about the untapped mass customers for marketing purposes could be retrieved.

ACKNOWLEDGEMENTS.

The authors gratefully acknowledge the financial support of the National Agency of Technology (Titan, grant no. 40063/08) and the Academy of Finland (grant no. 127656). The case organization's cooperation is also gratefully acknowledged.

REFERENCES

- [1] C. Rygielski, J.-C. Wang, and D. C. Yen, "Data mining techniques for customer relationship management," *Technology in Society*, vol. 24, pp. 483-502, 2002.
- [2] S.-Y. Kim, T.-S. Jung, E.-H. Suh, and H.-S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study," *Expert Systems with Applications*, vol. 31, pp. 101-107, 2006.
- [3] H.-S. Park and D.-K. Baik, "A study for control of client value using cluster analysis," *J. Netw. Comput. Appl.*, vol. 29, pp. 262-276, 2006.
- [4] K. Tsipitsis and A. Chorianopoulos, *Data mining techniques in CRM: Inside customer segmentation*: John Wiley and Sons, 2010.
- [5] F. F. Reichheld and T. Teal, *The loyalty effect: The hidden force behind growth, profits, and lasting value*. Harvard Business Press, 2001.
- [6] J. Dyché, *The CRM handbook: A business guide to customer relationship management*. Addison-Wesley, 2002.
- [7] M. J. A. Berry and G. Linoff, *Data mining techniques: For marketing, sales, and customer relationship management*. Wiley, 2004.
- [8] Viscovery Software GmbH, Viscovery SOMine 5.0, <http://www.eudaptics.com/>.
- [9] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, pp. 1464-1480, 1990.
- [10] Jr, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol. 58, pp. 236-244, 1963.
- [11] *Self-organizing maps*: Springer-Verlag New York, Inc., 1997.
- [12] J. Han and M. Kamber, *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*: Morgan Kaufmann, 2000.
- [13] S. Kaski, J. Kangas, and T. Kohonen, "Bibliography of Self-Organizing Map (SOM) Papers 1981-1997," *Neural Computing Surveys* vol. 1, pp. 102-350 1998.
- [14] M. Oja, S. Kaski, and T. Kohonen, "Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum," *Neural Computing Surveys*, vol. 1, pp. 1-176, 2002.
- [15] *Visual Explorations in Finance*: Springer-Verlag New York, Inc., 1998.
- [16] T. Kohonen, J. Hynninen, J. Kangas, and J. Laaksonen, "SOM PAK: The Self-Organizing Map program package," 1996.
- [17] L. Wiskott and T. Sejnowski, "Constrained optimization for neural map formation: A unifying framework for weight growth and normalization (1998)," *Neural Computation*, vol. 10, pp. 671-716, 1998.

[18] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *Neural Networks, IEEE Transactions on*, vol. 11, pp. 586-600, 2000.

[19] S.-C. Lee, J.-C. Gu, and Y.-H. Suh, "A Comparative Analysis of Clustering Methodology and Application for Market Segmentation: K-Means, SOM and a Two-Level SOM," in *Foundations of Intelligent Systems*, 2006, pp. 435-444.

[20] S. Samarasinghe, *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*: CRC Press, 2007.

[21] H. Li, B. Golden, E. Wasil, and P. Zantek, "A Comparison of Software Implementations of SOM Clustering Procedures," in *Intelligent Engineering Systems through Artificial Neural Networks*. vol. 12, A. B. Ca Dagli, Js Ghosh, M Embrechts, Os Ersoy, Ss Kercel, Ed.: ASME Press, 2002, pp. 447-452.

[22] C. Romesburg, *Cluster Analysis for Researchers*: Lulu.com, 2004.

[23] G. J. Myatt and W. P. Johnson, *Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications*: John Wiley and Sons, 2009.

[24] J. F. H. Jr, "Knowledge creation in marketing: the role of predictive analytics," *European Business Review*, vol. 19, pp. 303 - 315, 2007.

[25] V. N. Vapnik, *The nature of statistical learning theory*: Springer-Verlag New York, Inc., 1995.

[26] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*: Springer Publishing Company, Incorporated, 2008.

[27] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines : and other kernel-based learning methods*: Cambridge University Press, 2000.

[28] V. Vapnik and A. Sterin, "On structural risk minimization or overall risk in a problem of pattern recognition," *Automation and Remote Control*, vol. 10, pp. 1495-1503, 1977.

[29] I. Guyon. SVM Application List. <http://www.clopinet.com/isabelle/Projects/SVM/applist.html>.

[30] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998.

[31] A. Vellido, P. J. G. Lisboa, and J. Vaughan, "Neural networks in business: a survey of applications (1992-1998)," *Expert Systems with Applications*, vol. 17, pp. 51-70, 1999.

[32] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural networks for perception (Vol. 2): computation, learning, architectures*: Harcourt Brace & Co., 1992, pp. 65-93.

[33] S. K. Murthy, "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey," *Data Min. Knowl. Discov.*, vol. 2, pp. 345-389, 1998.

[34] L. Rokach and O. Z. Maimon, *Data mining with decision trees: theory and applications*: World Scientific, 2008.

[35] R. R. P. Ltd, "Data Mining Tools See5 and C5.0," 2008.

[36] J. R. Quinlan, *C4.5: programs for machine learning*: Morgan Kaufmann Publishers Inc., 1993.

[37] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proceedings of the Second European Conference on Computational Learning Theory*: Springer-Verlag, 1995.

[38] J. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," in *Advances in Large Margin Classifiers*, 1999, pp. 61-74.

TABLE IV. SEGMENT PROFILE OF THE SOM-WARD MODEL.

ID	Value rank	Average spending amount per customer (€)	Size (%)	Contribution to overall purchase amount (%)	Segment Profile
7	Exclusive customers	23403	0.32	13.6	<ul style="list-style-type: none"> • Highest loyalty point level and service level • Relatively high estimated income level and estimated probability of having children • Mosaic groups A, B, and C
6	High-spending customers	3120	4.90	28.1	<ul style="list-style-type: none"> • Very high loyalty point level and service level • Relatively high estimated income level • Mosaic groups A, C, and D
5	Mass customers	421	8.57	6.6	<ul style="list-style-type: none"> • Highest age and customer tenure • Lowest estimated probability of having children • Relatively low estimated income level • Mosaic groups F, D, E, and I
2	Mass customers	381	28.56	20.0	<ul style="list-style-type: none"> • Relatively young customers • High loyalty point level, service level and estimated income level • Highest estimated probability of having children. • Mosaic groups B, A, I, and H
1	Mass customers	331	25.23	15.4	<ul style="list-style-type: none"> • Senior customers with low customer tenure • Relatively low loyalty point level. • Very low estimated probability of having children • Mosaic group F
4	Mass customers	285	11.68	6.1	<ul style="list-style-type: none"> • Youngest customers with lowest customer tenure • Lowest loyalty point level, service level and estimated income level • Mosaic groups E, D and G
3	Mass customers	270	20.73	10.3	<ul style="list-style-type: none"> • Very high loyalty point level, service level and estimated income level • Mosaic groups A and B

Research Paper 3

Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2012), Temporal customer segmentation using Self-Organizing Time Map. In *Proceedings of the 16th International Conference on Information Visualization (IV 2012)*, Montpellier, France, 10-13 July, 2012, pp. 234-240

Temporal customer segmentation using the Self-Organizing Time Map

Zhiyuan Yao, Peter Sarlin, Tomas Eklund and Barbro Back
Åbo Akademi University, Turku Centre for Computer Science (TUCS), Finland
{Zhiyuan.Yao, Peter.Sarlin, Tomas.Eklund, Barbro.Back}@abo.fi

Abstract

Visual clustering provides effective tools for understanding relationships among clusters in a data space. This paper applies an adaptation of the standard Self-Organizing Map for visual temporal clustering in exploring the customer base and tracking customer behavior of a department store over a 22-week period. In contrast to traditional clustering techniques, which often provide a static snapshot of the customer base and overlook the possible dynamics, the Self-Organizing Time Map enables exploring complex patterns over time by visualizing the results in a user-friendly way. We demonstrate the effectiveness of the application using department store data with more than half a million rows of weekly aggregated customer information.

Keywords--- Visual clustering, Self-Organizing Time Map (SOTM), Temporal customer segmentation.

1. Introduction

Over the past decades, the focus of modern companies has been shifting from being product-oriented to customer-centric. In recent years, this change has been particularly rapid due to the increasing interest in customer relationship management (CRM) [1]. CRM begins with customer identification, i.e., to understand customers' demographic, behavioral and attitudinal characteristics. Customer segmentation is an effective approach to customer understanding and identification. It divides the customer base into distinct and internally homogeneous groups. Effective segmentation enables companies to interact with customers in each segment collectively, such as formulating marketing strategies aimed at different segments.

Various clustering algorithms, e.g., distance- and density- based techniques, have been used to approach the segmentation task. Recently, visualization techniques have gained in popularity for understanding and assessing clustering results [2]. The Self-Organizing Map (SOM) [3] is a well-known and widely used method for visual clustering of high dimensional data. Unlike most clustering algorithms that require post-processing for understanding cluster structures, the SOM is unique for its simultaneous clustering via vector quantization

and projection via neighborhood-preservation. The effectiveness of the SOM in customer segmentation has been demonstrated in a number of studies [4-6].

These segmentation solutions often provide only a static snapshot of the underlying customer base. However, the customer base itself and customers' purchasing behavior are not static but evolve over time, especially during sales events. A segmentation based on a specific timeframe might overlook possible dynamics in the interim. Therefore, providing a holistic view of multivariate temporal patterns for better understanding of customers is the main focus of this paper.

The SOM literature has provided several enhancements for temporal processing. Often, time is introduced implicitly in post-processing (e.g., trajectories [7]). The activation and learning rule (e.g., Hypermap [8]) and network topology (e.g., Temporal SOM [9]) have also been adapted for temporal processing. The standard SOM has also been paired with various visualization techniques for better spatiotemporal visualization (e.g., [10]). Yet, the problems of visualizing changes in inherent data structures over time and disentangling the temporal dimensions and cross-sectional structures still need to be addressed differently. The recently introduced Self-Organizing Time Map (SOTM) [11] provides a visual means for this type of exploratory temporal structure analysis.

In this paper, we use customers' demographic and behavioral characteristics and apply the SOTM for 1) performing multivariate clustering of customers over time; 2) visualizing the temporal variation of the multivariate patterns; and 3) detecting and interpreting complex patterns of changes in the customer base and purchasing behavior during three special sales events.

The remainder of the paper is organized as follows: Section 2 describes the SOTM. Section 3 documents the data used in the study, the experiments with the SOTM and analysis of the results. Finally, we summarize the key findings in Section 4.

2. Methodology

This section presents the SOTM and its related quality measures, followed by a description of the techniques for visualizing the results.

2.1. Self-Organizing Time Map

The SOTM [11] uses the capabilities of the SOM for abstraction of temporal structural changes in data. In principle, the SOTM applies one-dimensional SOMs on data ordered in consequent time units. These one-dimensional SOMs are then set in an ascending order of time to have one axis representing time topology and one representing data topology.

To observe the cross-sectional structures of the dataset for each time unit t (where $t=\{1,2,\dots,T\}$), the SOTM performs a mapping from the input space $\Omega(t)$, with a probability density function $p(x,t)$, onto a one-dimensional array $A(t)$ of output units $m_i(t)$ (where $i=1,2,\dots,M$). To preserve the orientation between consecutive patterns, the SOTM uses short-term memory. In particular, when $t=1$ the first principal component of principal component analysis (PCA) is used for initializing $A(t)$; otherwise, the reference vectors of $A(t-1)$ initialize $A(t)$. Adjustment to temporal changes is achieved by performing a batch update per time t . Thereafter, the timeline is created by arranging $A(t)$ in an ascending order of time t . The topology preservation of the SOTM is hence twofold: the horizontal direction preserves time topology and the vertical preserves data topology. The SOTM follows the two-step procedure, similarly as the SOM. Matching is performed by $\min\|x(t) - m_i(t)\|$ and the batch update by:

$$m_i(t) = \sum_{j=1}^{N(t)} h_{ic}(t)x_j(t) / \sum_{j=1}^{N(t)} h_{ic}(t),$$

where c is a best-matching unit (BMU) and the neighborhood $h_{ic(j)}(t)$ is defined as a Gaussian function restricted to vertical relations. For a comparable timeline, the neighborhood is constant over time. Figure 1 presents the functioning of the SOTM.

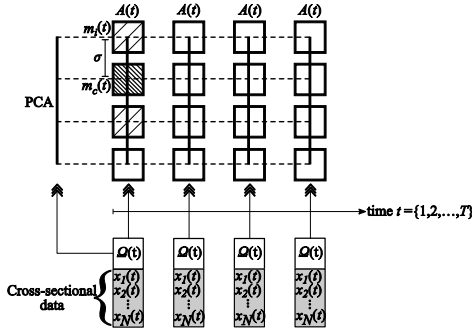


Figure 1: The functioning principles of the SOTM.

2.2. Quality measures of the SOTM

Characteristics of a SOTM can be quantified by a number of quality measures from the standard SOM paradigm: quantization error ε_{qe} , distortion measure ε_{dm} , and topographic error ε_{te} . The measures, while

seeming complex, are nothing more than aggregated qualities of the one-dimensional SOMs (for details see [11]):

$$\begin{aligned} \varepsilon_{qe} &= \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \sum_{j=1}^{N(t)} \|x_j(t) - m_{c(j)}(t)\|, \\ \varepsilon_{dm} &= \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \frac{1}{M(t)} \sum_{j=1}^{N(t)} \sum_{i=1}^{M(t)} h_{ic(j)}(t) \|x(t)_j - m(t)_i\|^2, \\ \varepsilon_{te} &= \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \sum_{j=1}^{N(t)} u(x_j(t)), \end{aligned}$$

where $u(x_j(t))$ is the average proportion of $x_j(t) \in \Omega(t)$ for which first and second BMUs (within $A(t)$) are non-adjacent units.

2.3. Visualizations of the SOTM

The multidimensionality of the SOTM is visualized by feature plane representations. It shows for each variable the temporal evolution of its cross-sectional distribution. As for the standard SOM, the feature planes are different views of the same map, where one unique point represents the same node on all planes. We produce the feature planes using heat map color coding, with cold to warm colors representing low to high values according to their own color scales. As for each feature plane the color scale is common for the entire SOTM ($t=\{1,2,\dots,T\}$), the temporal changes in data distributions are shown by variations in heat.

Sammon's mapping [12], a multidimensional scaling technique, is used for assessing the structural properties of SOTMs. It tries to match the pairwise distances of the SOTM nodes with their original distance in the high-dimensional space, enabling examination of structural properties at time t (vertically) and changes in structures (horizontally). In a Sammon's mapping of a SOTM, we plot all SOTM nodes ($m_i(t)$ where $t=\{1,2,\dots,T\}$) to one dimension. Then, we disentangle time by plotting the SOTM nodes according to Sammon's dimension on the y axis and time on the x axis, and retain neighborhood relations by connecting lines to have a net-like representation. Additionally, a coloring method by Kaski *et al.* [13] is used for visualizing the cluster structure of the SOTM. The well-known uniform color space CIE Lab [14] is used, where the perceptual differences of the colors represent the distances in the data space approximated by the Sammon's mapping. However, as the SOMs of the SOTMs are one-dimensional, we only use one dimension (blue to yellow) of the color space.

3. Temporal customer segmentation

In this section, we describe the data used in the study and the experiment for choosing the model specification. Then, we apply the SOTM to conduct temporal customer segmentation demographically, followed by a temporal analysis of customers' purchasing behavior.

3.1. Data

The data used in this study are from a department store that belongs to a large, multiservice corporation. The dataset contains weekly aggregated sales information per customer for the department store. The data spans 22 weeks from September 2008 to January 2009. For each week, only active customers are included in the data sets. The transaction data for each week were aggregated to a set of customer-level behavioral variables. To this end, we have weekly data on six demographic variables and six behavioral variables. Therefore, each observation in the data set is defined by four components: the customer ID, the week number, a set of demographic variables, and a set of behavioral variables. The distribution of the number of customers across the 22 weeks is illustrated in Figure 2. It is shown that week 40 (Sales event A), week 46 (Sales event B) and pre-Christmas weeks (weeks 50 and 51) attract more customers than average.

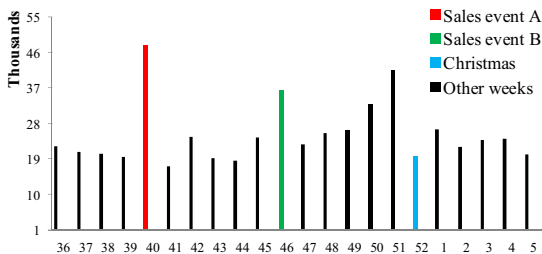


Figure 2. The distribution of the number of customers across the 22 weeks.

The variables used in this study fall into two bases: demographic and purchasing behavior variables. A brief explanation of these variables follows. The demographic variables show background data about the customers.

- Gender: 0 for male and 1 for female.
- Estimated probability of children: The higher the value of this variable is, the more likely there are children living in the household. The values range from 1 to 10.
- Estimated income level: The higher the value, the wealthier the household is considered to be. Possible values are 1, 2 and 3.
- Loyalty point level: Based on the average purchase amount per customer in the corporation, this variable divides customers into five classes: 0, 1, 2, 3, and 4. A higher value in loyalty point level indicates a larger historical spending amount in the entire corporate chain.

- Service level: This variable measures the number of service providers in the corporation that the customer has used in the previous year.
- Age

As mentioned previously, the purchasing behavior variables are summarized from a massive transaction database to weekly aggregated customer data.

- Purchase frequency: Average number of transactions per day.
- Basket size: Average number of items per transaction.
- Average item value: Average value per item purchased.
- Average transaction value: Average value per purchase transaction.
- Working time transaction: The percentage of purchases made from Mon - Fri, 9am – 5pm.
- Number of categories: Average total number of distinct product groups purchased in each transaction.

Only demographic variables are used in training SOTMs. This enables the assessment of not only differences in demographics, but also in their associated purchasing behavior. Each variable is pre-processed by scaling by range for the training as well as post-processed back to their normal distributions for the visualizations.

3.2. Experiments

In this section, we test different specifications and architectures of the SOTM. The SOTM requires the user to set the number of nodes on the y and x axis. While the number of time units sets the x dimension, the number of clusters on the y axis is less straightforward. We set the number of time units to 22 to span the sales events of interest. The number of clusters is set to five according to the optimal number of customer segments in a previous static study on a pooled version of this data set [15]. Hence, we obtain a SOTM with 5×22 nodes, where five nodes represent data topology for a specific week on the vertical direction and 22 the time dimension in terms of weeks on the horizontal direction.

Figure 3 illustrates the quality measures of the SOTM with neighborhood radii ranging from 0.4 to 8. It is shown that the topographic error decreases gradually to zero and stabilizes when the neighborhood radius increases to 1.2; meanwhile, the quantization error starts to increase significantly. Therefore, we chose the SOTM with the neighborhood radius of 1.2 for a good balance between topology preservation and fit to data.

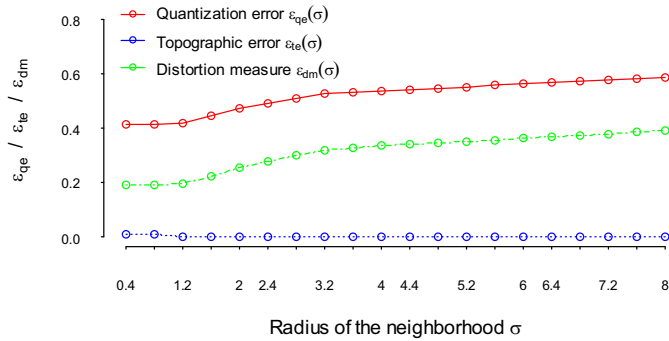
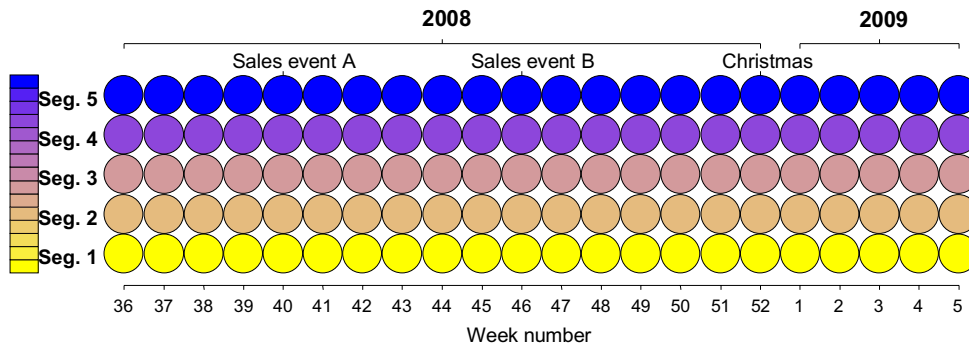


Figure 3. Quality measures over radius of the neighborhood.



Note: The figure represents a SOTM model with 5x22 nodes, where five nodes represent data topology for a specific week on the vertical axis and 22 the time dimension in terms of weeks for a specific segment on the horizontal axis. The perceptual differences of the colors, as defined by a color scale consisting of 15 colors, reflect the differences in the distances between the SOTM nodes. Due to small changes over time on the horizontal axis, each row of nodes forms an own segment, as labeled on the left side (Seg.1 - Seg.5).

Figure 4. The grid structure of a SOTM abstraction of the customer base over time.

3.3. Temporal customer segmentation

The obtained SOTM is shown in Figure 4. Above the figure, we show the events of interest, and on the left side are the five segments. It should be noted that the segment number is an indication of the location of the nodes; however, the data in the horizontally neighboring nodes are still allowed to change overtime. The perceptual differences of the colors in the figure reflect the differences in the distances between the SOTM nodes. Due to the small horizontal changes, as are also illustrated in the Sammon's mapping in Figure 5, the color differences are hardly discernible. In the Sammon's mapping, each point represents the corresponding segment; the vertical solid connections represent data topology and the dashed horizontal connections show time topology on the SOTM. The mapping confirms that the temporal changes are gradual, equal and small, but that the vertical differences between Segments 1 and 2 and Segments 4 and 5 are relatively large. It also reveals that the differences between Segments 2 and 3 increase overtime.

The multidimensionality of Figure 4 can be described using feature plane representations, illustrated in Figure 6. These feature planes are essentially a series of two-dimensional views of the customers, showing

information across segments (represented with the vertical dimension), time (represented with the horizontal dimension), and multiple variables (represented by the feature planes). This provides a holistic view of how the characteristics of customer segments, as well as their distribution, evolve over time.

Using the feature planes, we can observe some general characteristics of the customer segments. For instance, male customers are located in the nodes on the lowest part of the map (cf. Fig. 6(a)), i.e., Segment 1 in Figure 4. Figure 6(b) and Figure 6(c) show that customers with children at home and high income customers are located in the middle and lower parts of the map. Customers with high loyalty point levels and service levels are located in the lower part of the map (cf. Fig. 6(d) and Fig. 6(e)). Elderly customers are located on the second horizontal row on the map and young in the two rows below, as illustrated in Figure 6(f). It should be noted that the lowest average age in the segments is as high as 40 years as indicated in the color scale in Figure 6(f). The demographic characteristics of each segment are summarized as follows:

- **Seg. 1:** Male customers with medium age, medium probability of children, high estimated

income, medium loyalty point level and medium to high service level.

- **Seg. 2:** Female customers with low age, high probability of children, high estimated income, high loyalty point level and high service level.
- **Seg. 3:** Female customers with low age, high probability of children, medium estimated income, low loyalty point level and medium service level.
- **Seg. 4:** Elderly female customers with low probability of children, medium estimated income, low loyalty point level and medium to low service level.
- **Seg. 5:** Female customers with medium age, low probability of children, low estimated income, low loyalty point level and low service level.

By observing changes in the horizontal direction, we can observe temporal changes in the segments. Figure 6 shows that the pre-Christmas period attracts customers with higher loyalty point level (cf. Fig 6(d)) and service

level (cf. Fig 6(e)) in Segment 4. This change is also displayed in the Sammon's mapping in Figure 5 with the small increased vertical differences between Segments 4 and 5 during that period. There is also a slight decrease in the age of shopping customers belonging to Segments 1 and 5 for week 52 (cf. Fig 6(f)). However, while the composition of customers in segments changes over time, the temporal changes in segment characteristics are small, which is also confirmed in the Sammon's mapping in Figure 5.

The frequency map (cf. Fig. 6(g)) gives a visual indication of the number of customers in each segment. It shows that Sales event A (week 40), and the week of Christmas (week 51) attract more customers than Sales event B (week 46), as confirmed by the larger proportion of yellow and red nodes. It is also found that male customers (Seg. 1) and younger high-income female customers with high loyalty points and service levels and a high chance of having children at home (Seg. 2) are more likely to react to the sales events.

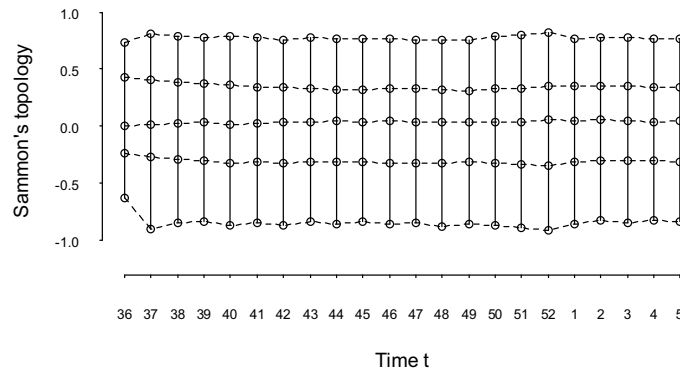
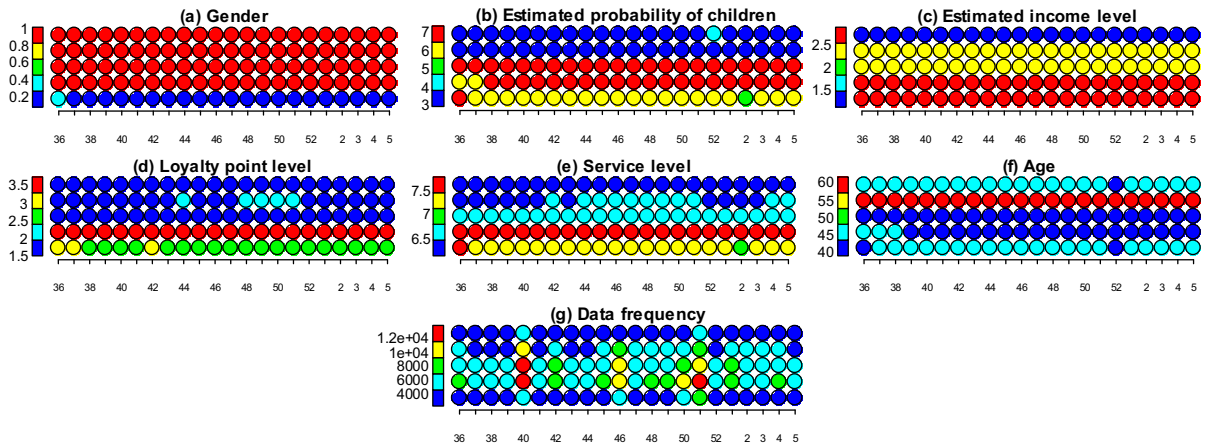


Figure 5. A Sammon's mapping of the SOTM nodes.



Note: Nodes having the same location represent the same data on all planes. The horizontal axis represents data across 22 weeks, while the five nodes on the vertical direction represent data topology at each time t.

Figure 6. Feature planes for the SOTM describing the demographic variables.

3.4. Temporal behavior analysis

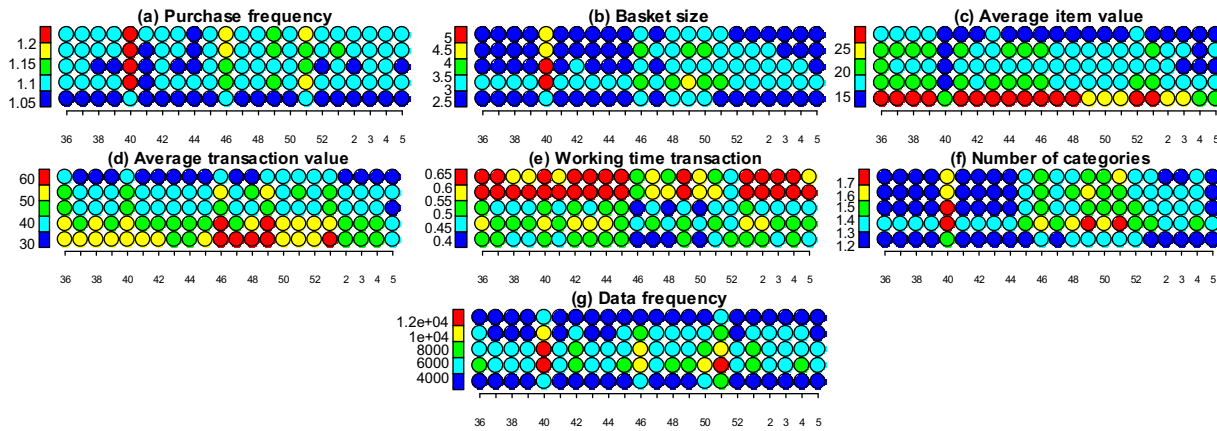
In contrast to the demographic variables used in training and analyzed in section 3.3, the behavioral variables had no impact on forming the segmentation result, but their distributions among the segments in the 5x22 SOTM grid are shown in the associated feature planes in Figure 7. The temporal changes of customers' purchasing behavior are, compared with demographic profiles, large and volatile. Figure 7 can be examined vertically to detect differences among the five segments during a specific week, and horizontally for detecting changes in a certain segment over time. For example, Figure 7(a) shows that the purchase frequency of customers in Seg. 1 is much lower than that of the rest of the segments during Sales event A (week 40). However, the purchase frequency of customers in Seg. 1 did increase during Sales event A, compared to the preceding and succeeding weeks.

The effects of the sales events and pre-Christmas sales are indeed clear. By examining the feature planes horizontally, Figures 7(a, b and f) show that customers went to the department store most frequently, bought more kinds of products, and more of them, during the week-long Sales event A (week 40). However, customers mainly bought low-priced products (cf. Fig. 7(c)) and their average transaction value (cf. Fig. 7(d)) was no higher than those of the other weeks. By examining the figure vertically, it can be seen that customers in Seg. 1 had a lower purchase frequency, basket size, and number of product categories purchased, but more expensive products and higher spending amount than the rest of the

segments. In addition, the frequency map (also shown in Figure 6(g)) is included in Figure 7(g) as well for interpreting the behavioral patterns. Figure 7(g) shows that the number of shopping customers in Segments 2 and 3 is larger than in the other segments, but that the sizes of all segments increased for Sales event A comparing to the preceding weeks, meaning that the customers in all segments actively responded to the event.

Sales event B (week 46) differs from A in that customers are less activated in terms of purchase frequency, number of items purchased and basket size. However, customers tended to buy more expensive products (cf. Fig. 7(c)) than during Sales event A, leading to a higher transaction value (cf. Fig. 7(d)). In addition, the proportion of transactions made during working hours was significantly lower during this week, especially for Segments 1 and 3.

Christmas also exhibits some interesting patterns. The final pre-Christmas week (week 51) attracted many more customers than usual (cf. Fig. 7(g)). Customers tended to buy medium to low priced products (cf. Fig. 7(c)) with small baskets (cf. Fig. 7(b)), thereby leading to the lower spending amount (cf. Fig. 7(d)) than during the preceding weeks. In addition to week 51, pre-Christmas week 49 attracted some large spenders (cf. Fig. 7(d)). During the Christmas and post-Christmas weeks (i.e., weeks 52 and 1), male Customers (Seg. 1) also tended to buy more expensive products (cf. Fig. 7(c)), leading to the high average transaction value (cf. Fig. 7(d)). Table 1 summarizes key information regarding the above-mentioned sales events.



Note: The behavioral variables represented by the feature planes (a) - (f) had no impact on forming the segmentation result. These feature planes display the distribution of the behavioral variables among the segments trained by the demographic variables. Nodes having the same location represent the same data on all planes. The horizontal axis represents data across 22 weeks, while the five nodes on the vertical direction represent data topology at each time t.

Figure 7. Feature planes for the SOTM describing the behavioral variables (associated variables).

Table 1. Summarized information related to the selected weeks.

Events / revenue	Deviation from median revenue	Main responding segments	Important patterns
	-0.5 Millions 2		
Sales event A (Week 40) € 2.77 mil		Segments 1-5	<ul style="list-style-type: none"> The highest purchase frequency among all the weeks. Relatively large basket and cheap products purchased comparing to the other weeks. More working time shopping for elderly women customers (Seg. 5).
Sales event B (Week 46) € 2.23 mil		Segments 1-5	<ul style="list-style-type: none"> Medium-sized basket with more expensive products comparing to event A. High spending amount for segments 1, 2 and 3. More non-working time shopping for segments 1 and 3.
Pre-Christmas (Weeks 49-51) € 1.57 / 1.57 / 2.15 mil		Segments 1-5	<ul style="list-style-type: none"> Medium-sized basket with many kinds of medium-priced products. Increased spending, especially for the first week of pre-Christmas period. More non-working time shopping for segments 1 and 3 in the second week of pre-Christmas period.
Christmas (Week 52) € 0.94 mil		Segments 2 and 3	<ul style="list-style-type: none"> Expensive products bought by male customers responding to the event, though the number of male customers decreased comparing to the pre-Christmas weeks.
Post-Christmas (Week 1) € 1.36 mil		Segments 2, 3 and 4	<ul style="list-style-type: none"> Medium to high spending amount attributed to the expensive products bought by male customers responding to the event, though the number of male customers decreased comparing to the pre-Christmas weeks.

4. Conclusion

In this paper, we have applied SOTM for tracking customers' demographics and behavior for a department store over time. We demonstrate the usefulness of the method on a case company's multivariate customer database with weekly data. The temporal customer segmentation is visualized by a series of feature planes, which provides a holistic view of the customer base and their purchasing behavior over time. The SOTM effectively detects the changes in customer behavior during sales events. The results show that the purchasing behavior of customers changes during the events, but at the same time, that the sales events differ in the type of shopping behavior that they trigger. The results are meaningful and interpretable, and indicate that the SOTM can serve as a valuable visual exploratory tool for decision makers to see how successful the sales events have been as to different criteria and to aid in decision making on which kind of sales events they should have. For business analysts it can also be used to drill down into the interesting patterns revealed for further analysis.

References

[1] G.S. Linoff and M.J. Berry, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, John Wiley & Sons, 2011.
 [2] M.C. Ferreira de Oliveira and H. Levkowitz, From visual data exploration to visual data mining: A survey, *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, 378-394, 2003.
 [3] T. Kohonen, Self-organizing maps, Springer, 2001.

[4] A.H. Holmbom, T. Eklund and B. Back, Customer portfolio analysis using the SOM, *International Journal of Business Information Systems*, vol. 8, 396-412, 2011.
 [5] J. Mo, M.Y. Kiang, P. Zou and Y. Li, A two-stage clustering approach for multi-region segmentation, *Expert Syst. Appl.*, vol. 37, 7120-7131, 2010.
 [6] Z. Yao, A. Holmbom, T. Eklund and B. Back, Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis, In *Advances in Data Mining. Applications and Theoretical Aspects*, P. Perner, Springer Berlin / Heidelberg, 2010, 292-307.
 [7] T. Kohonen, The "neural" phonetic typewriter, *Computer*, vol. 21, 11-22, 1988.
 [8] T. Kohonen, The Hypermap Architecture, *Artificial Neural Networks*, pp. 1357-1360, 1991.
 [9] G.J. Chappell and J.G. Taylor, The temporal Kohonen map, *Neural Networks*, vol. 6, 441-445, 1993.
 [10] D. Guo, J. Chen, A.M. MacEachren and K. Liao, A visualization system for space-time and multivariate patterns (vis-stamp), *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, 1461-1474, 2006.
 [11] P. Sarlin, Self-Organizing Time Map: An Abstraction of Temporal Multivariate Patterns. *TUCS Technical Report*, No. 1022, 2011.
 [12] J.W. Sammon Jr, A nonlinear mapping for data structure analysis, *IEEE Transactions on Computers*, vol. 100, 401-409, 1969.
 [13] S. Kaski, J. Venna and T. Kohonen, Coloring that reveals cluster structures in multivariate data, *Australian Journal of Intelligent Information Processing Systems*, vol. 6, 82-88, 2000.
 [14] Colorimetry, 2nd Ed. CIE Publication No. 15.2, 1986.
 [15] Z.H. Yao, P. Sarlin, T. Eklund and B. Back, Combining visual customer segmentation and response modeling. In *Proceedings of the 20th European Conference on Information Systems*, 2012 (forthcoming).

Research Paper 4

Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2013), Visualizing dynamics in customer behavior with the Self-Organizing Time Map. TUCS Technical Report No. 1085. *Submitted to Business & Information Systems Engineering.*



Visualizing dynamics in customer behavior with the Self-Organizing Time Map

Zhiyuan Yao

Åbo Akademi University, Department of Information Technologies

Peter Sarlin

Åbo Akademi University, Department of Information Technologies

Tomas Eklund

Åbo Akademi University, Department of Information Technologies

Barbro Back

Åbo Akademi University, Department of Information Technologies

Abstract

Visual clustering provides effective tools for understanding relationships among clusters in a data space. This paper applies the Self Organizing Time Map (SOTM) for visual dynamic clustering of the customer base and for tracking customer behavior in a department store over a 22-week period. In addition, in order to objectively represent dynamics in cluster structures, we also apply a second-level clustering to the SOTM model to visualize the temporal changes of segment structures and customers' purchasing behavior. We demonstrate the effectiveness of the application using department store data with more than half a million rows of weekly aggregated customer information.

Keywords: Visual dynamic clustering, Self-Organizing Time Map (SOTM), Temporal customer segmentation.

1 Introduction

Over the past decades, the focus of modern companies has been shifting from being product-oriented to customer-centric (Tseng and Piller, 2003). In recent years, this change has been particularly rapid due to the increasing interest in customer relationship management (CRM). CRM aims to enhance customers' relationships and overall experience through customized communications, cross-selling, and customer segmentation (Payne and Frow, 2005). The concept of segmentation was first introduced in the seminal work of Smith (Smith, 1956), in which the author claimed that "market segmentation involves viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desires of customers for more precise satisfactions of their varying wants". Customer segmentation is an effective approach to customer understanding and identification. It divides the customer base into distinct and internally homogeneous groups. Effective segmentation enables companies to interact with customers in each segment collectively, for example, by formulating marketing strategies aimed at different segments.

One important criterion for evaluating the practicality of a segmentation solution is its stability (Thomas, 1980). It is a commonly ignored fact that the size and characteristics of the identified segments can change over time (Calantone and Sawyer, 1978), especially when markets are becoming increasingly competitive and customers' needs and preferences are evolving constantly. These changes may cause the original targeted marketing strategies to be outdated. Lingras et al. (2005) summarized two types of temporal changes in a segmentation solution: changes in segment membership of individual objects and changes in segment composition. These types of dynamics are well documented in previous studies. While Calantone and Sawyer (1978) and Farley et al. (1987) found that the segments are stable but the individual customers' segment memberships changed significantly, Hoek et al. (1996), Steenkamp and Ter Hofstede (2002), Blocker and Flint (2007) found that segment composition evolves over time. Hoek et al. (1996) stated that "it seems illogical to expect the size, composition, and behavior of market segments defined in these terms to remain constant". Wind (1978) summarized three reasons of segment instability: 1) the basis for segmentation, 2) the volatility of the market place, and 3) consumer characteristics. A more specific and less general segmentation basis can lead to instable segments. The more volatile the market is, the less stable the segments would be. The changes in customers' life cycles can also lead to segment instability. Moreover, Hu and Rau (1995) showed that both changes in segment membership of individual objects and changes in segment composition can happen simultaneously.

Changes in segment membership of individual objects has been addressed using segment migration analysis, which is based upon the assumption that the segment structure stays the same over time. A matrix of switching probabilities among segments is commonly used for segment migration analysis (Homburg et al., 2009). In addition, there are a number of empirical studies (Lingras et al., 2005, Ha et al., 2002, Yao et al., 2012a) addressing the segment migration problems in the context of CRM. However, the managerial implications of these studies were made based upon the assumption that segment composition in terms of size and characteristics are static over time, which is often invalid in today's business environment.

The focus of this study, therefore, is on the changes in segment composition over time. Since markets are becoming more competitive and customers' needs and preferences are evolving constantly, this may lead to changes in the initially identified segments. Changes in segment composition over time has been addressed in (Sarlin and Yao, 2013) where the authors proposed a method for illustrating changes in segment structures over time, e.g., the emergence of new segments, and the changing or disappearance of existing segments. In this study, we apply the Self Organizing Time Map (SOTM) (Sarlin, 2012) for visual dynamic clustering of the customer base, which enables the detection of temporal changes of segment structures and customers' purchasing behavior. This model aims for 1) performing multivariate clustering of customers over time; 2) visualizing the temporal variation of the multivariate patterns; 3) detecting and interpreting complex patterns of changes in the customer base and purchasing behavior during three special sales events; and 4) applying the second-level clustering approach on the SOTM model to identify changing, emerging and disappearing customer segments in an easily interpretable way. The contribution of this study is to provide a holistic view of multivariate temporal patterns for better understanding of the evolution of and dynamics in a customer base.

The remainder of the paper is organized as follows: Section 2 presents related work, while Section 3 describes the methodology for conducting the visual dynamic clustering. Section 4 describes the data used in the study. Section 5 documents the experiments with the SOTM and the analysis of the results. Finally, we summarize the key findings in Section 6.

2 Related studies

Various clustering algorithms have been used to approach segmentation tasks. Likewise, visualization techniques have gained in popularity for understanding and assessing clustering results (Ferreira de Oliveira and Levkowitz, 2003). The Self-Organizing Map (SOM) (Kohonen, 1982) is a well-known and widely used method for

visual clustering of high dimensional data. Unlike most clustering algorithms that require post-processing for understanding cluster structures, the SOM is unique in its simultaneous clustering via vector quantization and projection via neighborhood-preservation. As illustrated in Sarlin (2013) and Vesanto (1999), arguments for using the SOM over alternative techniques are the following: a pre-defined grid structure for linking visualizations, interaction between the two tasks of clustering and projection, flexibility for missing data and outliers, and computational efficiency. The effectiveness of the SOM in customer segmentation has been demonstrated in a number of studies (Holmbom et al., 2011, Mo et al., 2010, Yao et al., 2010, Kiang and Kumar, 2001, Curry et al., 2001). These segmentation solutions often provide a static snapshot of the underlying customer base. However, as was previously noted, the customer base itself and customers' purchasing behavior are not static but evolve over time, especially during sales events. A segmentation based on a specific timeframe might overlook possible dynamics in the interim.

The SOM literature has also provided several enhancements for temporal processing: 1) time is introduced implicitly in post-processing (e.g., trajectories (Kohonen, 1988)), 2) changes of the activation and learning rule (e.g., Hypermap (Kohonen, 1991)), 3) adaptation of network topology (e.g., Temporal SOM (Chappell and Taylor, 1993)), and 4) pairing the standard SOM with various visualization techniques for better spatiotemporal visualization (e.g., (Kuo et al., 2006)). Readers are referred to Sarlin (2012) for more details of the classification of the SOM-based methods for temporal data analysis. However, these methods are not optimal for visualizing changes in data or segment structures over time. In Chakrabarti et al. (2006), a framework was introduced for a type of problem called evolutionary clustering. Evolutionary clustering aims at exploratory temporal structure analysis by producing a sequence of clusterings for each point in time of the underlying temporal data. An effective evolutionary clustering aims to achieve a balance between clustering results being faithful to current data and comparable with its neighboring clustering results. This feature makes evolutionary clustering a promising approach for the detection of temporal changes in segment structures.

Evolutionary clustering in the SOM context can be conducted by comparing standard SOMs at different points in time. Back, et al. (Back et al., 1998) applied the SOM for benchmarking the financial performance of pulp and paper companies during 1985–1989. The authors monitored the performance of Finnish companies by comparing several SOM models side by side. However, this task is to some extent hindered by the inconsistent cluster locations due to the random initialization of reference vectors and random sequence of input data to the network, as was constrained by the functionality of the tool used. These drawbacks can be partly cured by the use of initialization

techniques and the batch SOM algorithm. In Denny and Squire (2005), Denny et al. (2010), the temporal interpretability of the SOM is enhanced by applying pre-defined initializations and additional visualizations. Nevertheless, this approach still has the drawback of an unstable orientation over time and complex comparisons of two-dimensional grids.

The recently introduced SOTM (Sarlin, 2012) provides a visual means for evolutionary clustering. The SOTM is essentially a series of one-dimensional SOMs ordered in consequent time nodes and represents both time and data topology on a two dimensional grid. In Yao et al. (2012b), the SOTM has been used for temporal customer segmentation, in which demographic variables were used to create a SOTM model and the behavioral variables were associated with the model to explore customer behavior over time. With stationary data of the demographic variables, the rows of the SOTM roughly represent similar data at different points in time. However, as the behavioral data are more volatile, it is difficult to identify the structure of the clusters due to the changing nature of the data. In order to objectively represent temporal changes in the multivariate cluster structures, a method based upon second-level clustering of the SOTM has been introduced in Sarlin and Yao (2013). In this paper, we follow the approach in Sarlin and Yao (2013) by applying the SOTM to conduct temporal customer segmentation based upon customer purchasing behaviour and demographic characteristics, followed by a second-level hierarchical clustering of the SOTM.

3 Methodology

3.1 Self-Organizing Time Maps

The SOTM (Sarlin, 2012) uses the capabilities of the SOM for abstraction of temporal structural changes in data. In principle, the SOTM applies one-dimensional SOMs on data ordered in consequent time points. These one-dimensional SOMs are then set in an ascending order of time, in order to have one axis representing the time topology and one representing data topology.

To observe the cross-sectional structures of the dataset for each time point t (where $t=\{1,2,\dots,T\}$), the SOTM performs a mapping from the input space $\mathcal{Q}(t)$, with a probability density function $p(x,t)$, onto a one-dimensional array $A(t)$ of output nodes $m_i(t)$ (where $i=1,2,\dots,M$). To preserve the orientation between consecutive patterns, the SOTM uses short-term memory. In particular, when $t=1$ the first principal component of principal component analysis (PCA) is used for initializing $A(t)$; otherwise, the reference vectors of $A(t-1)$ initialize $A(t)$. Adjustment to temporal changes is achieved by performing a batch update per time t . Thereafter, the timeline is created by arranging

$A(t)$ in an ascending order of time t . The topology preservation of the SOTM is hence twofold: the horizontal direction preserves time topology and the vertical *preserves* data topology. The SOTM follows the two-step procedure, similarly as the SOM. Matching is performed by $\min \|x(t) - m_i(t)\|$ and the batch update by:

$$m_i(t) = \frac{\sum_{j=1}^{N(t)} h_{ic}(t) x_j(t)}{\sum_{j=1}^{N(t)} h_{ic}(t)},$$

where c is a best-matching unit (or node) (BMU) and the neighborhood $h_{ic(j)}(t)$ is defined as a Gaussian function restricted to vertical relations. The radius of the neighborhood function is adjusted with a user-specified neighborhood parameter σ . For a comparable timeline, the neighborhood radius parameter σ is constant over time.

3.2 Quality measures of the SOTM

The characteristics of a SOTM can be quantified by a number of quality measures from the standard SOM paradigm: quantization error ε_{qe} , distortion measure ε_{dm} , and topographic error ε_{te} . These measures essentially show the aggregated qualities of the one-dimensional SOMs (for details see (Sarlin, 2012)). First, the quantization accuracy is measured with the quantization error:

$$\varepsilon_{qe} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \sum_{j=1}^{N(t)} \|x_j(t) - m_{c(j)}(t)\|.$$

Likewise, with the distortion measure, we compute the fit of the map to the shape of the data distribution, but also account for the radius of the neighborhood:

$$\varepsilon_{dm} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \frac{1}{M(t)} \sum_{j=1}^{N(t)} \sum_{i=1}^{M(t)} h_{ic(j)}(t) \|x(t)_j - m(t)_i\|^2.$$

Finally, topographic error measures the quality of the topology preservation:

$$\varepsilon_{te} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N(t)} \sum_{j=1}^{N(t)} u(x_j(t)).$$

where $u(x_j(t))$ is the average proportion of $x_j(t) \in \Omega(t)$ for which first and second BMUs (within $A(t)$) are non-adjacent nodes.

3.3 Second-level clustering of the SOTM

While the SOTM enables visual clustering of temporal and cross-sectional patterns of data, it lacks means for objectively representing temporal changes in cluster structures. With stationary data, the horizontally neighboring nodes of the SOTM would represent similar data. When data are non-stationary, however, the horizontally neighboring nodes may represent data of different characteristics. Moreover, as the number of dimensions and nodes in a SOTM grid increases, the ability to perceive the temporal structural changes in data will be hindered. It is therefore reasonable to apply a second-level clustering to group the output nodes of the SOTM to second-level clusters. This second-level clustering enables one to identify the changes in the cluster structures more objectively. Following Sarlin and Yao (2013), we define three types of dynamics for assessing the changes in cluster structures: 1) a cluster *disappears* when one or more nodes are a member at time t and none is at $t+1$, 2) a cluster *emerges* when no node is a member of it at time t and one or more are at $t+1$, and 3) a cluster *changes* when the positive number of member nodes at time t and $t+1$ differ. A similar two-level clustering approach has been applied to the standard SOM (Vesanto and Alhoniemi, 2000), where the standard SOM, agglomerative hierarchical clustering and partitioning clustering have been used to cluster data based on the SOM.

Clustering is a class of techniques that partition data into groups, while attempting to minimize intra-cluster distance and maximize inter-cluster distance. Since the original data have been reduced by the SOTM, the general problems of computational cost and uncertainty of the clustering result caused by outliers and noisy data are decreased. As suggested in Sarlin and Yao (2013), agglomerative hierarchical clustering is used to group the SOTM nodes. The key motivation for using hierarchical clustering is that this enables us to circumvent the choice of number of clusters K . Instead it provides us means to explore the clustering results with varying K . Agglomerative hierarchical clustering starts by treating each node of the SOTM as a separate cluster ($K=M*T$) and iteratively merges clusters with the shortest distance until all nodes are merged ($K=1$). Merging can be performed using several definitions of distance, e.g., single-linkage, complete-linkage, average linkage or Ward's method. Clusters of single-linkage tend to take the form of long chains and other irregular shapes with little homogeneity, whereas complete-linkage clusters have been shown to be inefficient in separating data (Blashfield, 1976, Hansen and Delattre, 1978). In this study, we experimented with single-linkage, complete-linkage, Ward's method and average-linkage, and found average-linkage gave more interpretable results. Hence, we only report the clustering results based on the average-linkage measure, which is defined as follows (Han et al., 2011):

$$D(X,Y) = \frac{1}{N_X \times N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} d(x_i, y_j)$$

where $d(x_i, y_j)$ is the Euclidean distance between objects x_i and y_j belonging to clusters X and Y respectively, and N_x and N_y are the number of objects in clusters X and Y , respectively.

There are generally two types of cluster validity measures for evaluating the clustering results and determining the number of clusters: external and internal measures (Theodoridis and Koutroumbas, 2008). The external measures (e.g., Rand index (Rand, 1971) and Hubert's statistic (Hubert and Schultz, 1976)) evaluate a clustering solution with reference to some external *a priori* information, e.g., given class labels, while the internal measures (e.g., gap statistic (Tibshirani et al., 2001), Dunn index (Dunn, 1973) and Silhouette index (Rousseeuw, 1987)) evaluate a clustering solution in terms of the internal relationships among the data items. The gap statistic evaluates a clustering solution based upon the within-cluster dispersion, while the Dunn and the Silhouette indices take into account both cluster compactness and cluster separation. Since there are no class labels in our data, we decided to use the Dunn index and the Silhouette coefficient to assess the clustering results. For both measures, the higher the value is, the better the observations are clustered. The Dunn index is defined as the ratio of the smallest inter-cluster distance to the largest intra-cluster distance. The Dunn index is computed as:

$$D_K = \min_{1 \leq l \leq K} \left\{ \min_{\substack{1 \leq k \leq K \\ l \neq k}} \left\{ \frac{d(C_l, C_k)}{\max_{1 \leq h \leq K} \{d'(C_h)\}} \right\} \right\}$$

where K is the number of clusters, $d(C_l, C_k)$ is the distance between clusters l and k (inter-cluster distance) and $d'(C_h)$ is the maximum distance between observations in cluster h (intra-cluster distance).

For each observation i , its Silhouette coefficient is defined as:

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)}$$

where a_i is the average distance between i and all other observations in the same cluster, and b_i is the average distance between i and the observations in its nearest cluster. The Silhouette coefficient for a clustering solution is simply the average of the Silhouette coefficient of all observations.

3.4 Visualization of the SOTM

The multidimensionality of the SOTM is visualized by feature plane representations. It shows for each variable the temporal evolution of its cross-sectional distribution. As for the standard SOM, the feature planes are different views of the same map, where one unique point represents the same node on all planes.

The coloring of the feature planes is here performed using a sequential ColorBrewer's scale (Harrower and Brewer, 2003), where the lightness and saturation of the blue hue varies from light to dark to represent low to high values according to a feature plane-specific scale. As for each feature plane the color scale is common for the entire SOTM ($t=\{1,2,\dots,T\}$), the temporal changes in data distributions are shown by variations in blue hue. Likewise, to explore structures in the high-dimensional space, we use a qualitative scale to represent second-level clusters on the SOTM.

Sammon's mapping (Sammon Jr, 1969), a multidimensional scaling technique, is another approach for assessing the structural properties of SOTMs. It tries to match the pairwise distances of the SOTM nodes with their original distance in the high-dimensional space, enabling examination of structural properties at time t (vertically) and changes in structures (horizontally). In a Sammon's mapping of a SOTM, we plot all SOTM nodes ($m_i(t)$ where $t=\{1,2,\dots,T\}$) to one dimension. Then, we disentangle time by plotting the SOTM nodes according to Sammon's dimension on the y axis and time on the x axis, and retain neighborhood relations by connecting lines to have a net-like representation.

4 Data

The data used in this study are from a department store that belongs to a large, multiservice corporation. The dataset contains weekly aggregated sales information per customer for the department store and spans 22 weeks from September 2008 to January 2009. For each week, only customers that made at least one purchase are included in the data sets. The transaction data for each week were aggregated to a set of customer-level behavioral variables, appended with a number of background variables for profiling the segments. Therefore, each observation in the data set is defined by four components: the customer ID, the week number, a set of behavioral variables, and a set of background variables. The distribution of the number of customers across the 22 weeks is illustrated in Figure 1. The figure shows that week 40 (Sales event A), week 46 (Sales event B) and the pre-Christmas weeks (weeks 49, 50 and 51) attract more customers than average.

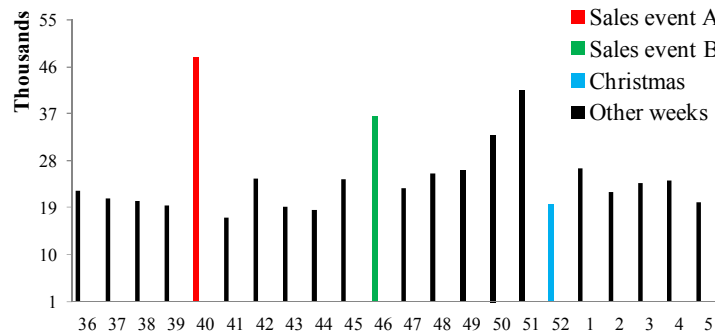


Fig. 1 The distribution of the number of customers across the 22 weeks.

The variables used in this study fall into two bases: purchasing behavior variables and background variables. A brief explanation of these variables follows. The purchasing behavior variables, summarized from a massive transaction database to weekly aggregated customer data, are briefly explained as follows.

- Purchase frequency: Average number of transactions per day.
- Number of items purchased: Average number of items purchased per week.
- Spending amount: Total weekly spending amount.
- Basket size: Average number of items per transaction.
- Average item value: Average value per item purchased.
- Average transaction value: Average value per purchase transaction.
- Working time transaction: The percentage of purchases made from Mon - Fri, 9am - 5pm.
- Number of categories: Average total number of distinct product groups purchased in each transaction.

The background variables, primarily demographic information, show background data about the customers.

- Gender: 0 for male and 1 for female.
- Estimated probability of children: The higher the value of this variable is, the more likely there are children living in the household. The values range from 1 to 10.
- Estimated income level: The higher the value, the wealthier the household is considered to be. Possible values are 1, 2 and 3.

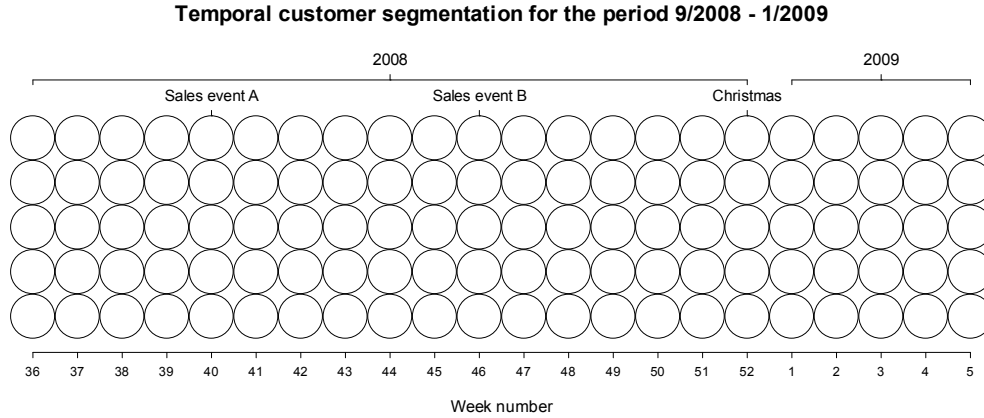
- Loyalty point level: Based on the average purchase amount per customer in the corporation, this variable divides customers into five classes: 0, 1, 2, 3, and 4. A higher value in loyalty point level indicates a larger historical spending amount in the entire corporate chain.
- Service level: This variable measures the number of service providers in the corporation that the customer has used in the previous year.
- Age

5 Visualizing dynamics in customer behavior

In this section, we first describe the experiments for choosing the model specification, and then apply the SOTM to conduct temporal customer segmentation based upon customer purchasing behavior to explore behavior over time. We also associate the background variables as described in Section 4 with the model to profile segments. It is noted that only customer purchase behaviour variables are used in training. The association of background variables enables the assessment of not only the changes in customers' purchase behavior, but also differences in demographics. Since customers' purchasing behavior is not stationary, especially during the sales events and Christmas period, a second-level clustering is applied to the SOTM model to facilitate the identification of cross-sectional and temporal changes in customer segments.

5.1 Experiments

In this section, we test different specifications and architectures of the SOTM. Each variable is pre-processed by scaling by range for training as well as post-processed back to their normal distributions for the visualizations. The SOTM requires the user to set the number of nodes on the y and x axes. While the number of time nodes sets the x dimension, the number of clusters on the y axis is less straightforward. We set the number of time nodes to 22 to span the sales events of interest. The number of clusters is set to five according to the optimal number of customer segments in a previous static study on a pooled version of this data set (Yao et al., 2012a). Hence, we obtain a SOTM with 5×22 nodes, where five nodes represent data topology for a specific week on the vertical direction and 22 the time dimension in terms of weeks on the horizontal direction, as illustrated in Figure 2.



Note: The SOTM model with 5x22 nodes, where five nodes represent data topology for a specific week on the vertical direction and 22 the time dimension in terms of weeks on the horizontal direction.

Fig. 2 The grid structure of a SOTM abstraction of the customer base over time.

Figure 3 illustrates the quality measures of the SOTM with neighborhood radii ranging from 1 to 6. The figure shows that the topographic error decreases gradually to zero and stabilizes when the neighborhood radius increases to 2.8, while the quantization error remains stable. Therefore, we chose a SOTM with a neighborhood radius of 2.8 for minimum quantization error given no topographic errors.

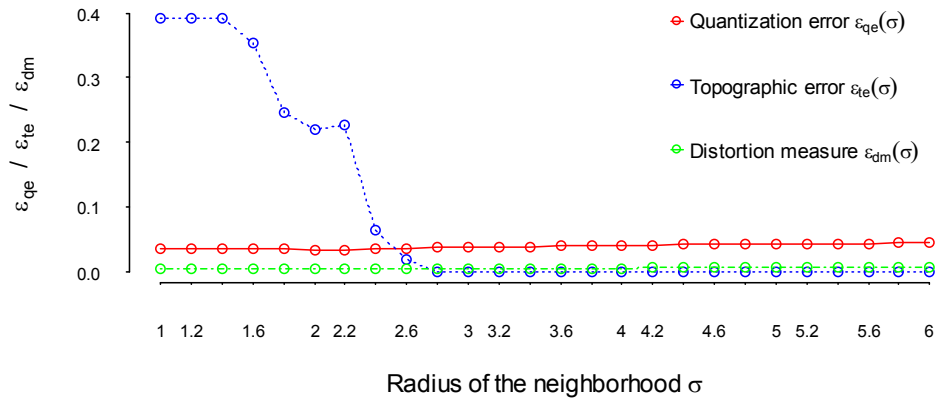


Fig. 3 Quality measures over radius of the neighborhood.

5.2 The SOTM on the behavioral variables

The results of the SOTM model are first illustrated using Sammon's mapping in Figure 4. In the Sammon's mapping, each point represents the corresponding segment; the vertical solid connections represent data topology and the dashed horizontal connections show time topology on the SOTM. The volatile changes in weeks 40, 46 and during

Christmas periods indicate that customers purchasing behavior is different than during the normal weeks. Specifically, customers belonging to the lower part of the map are more reactive to the sales events.

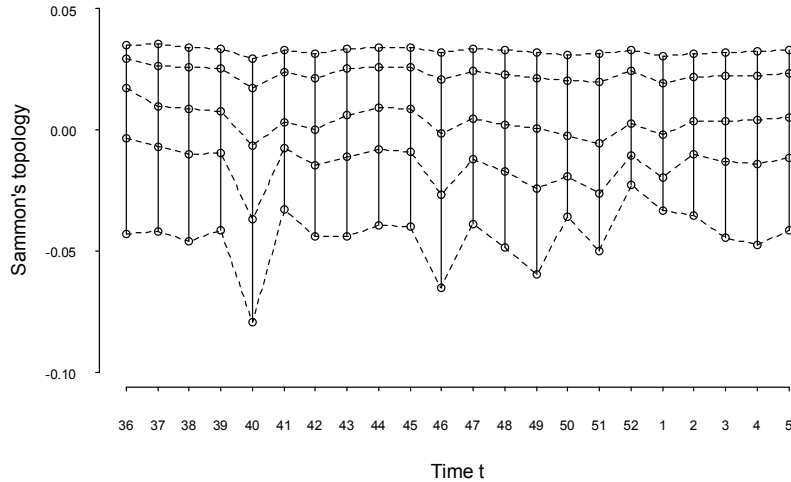
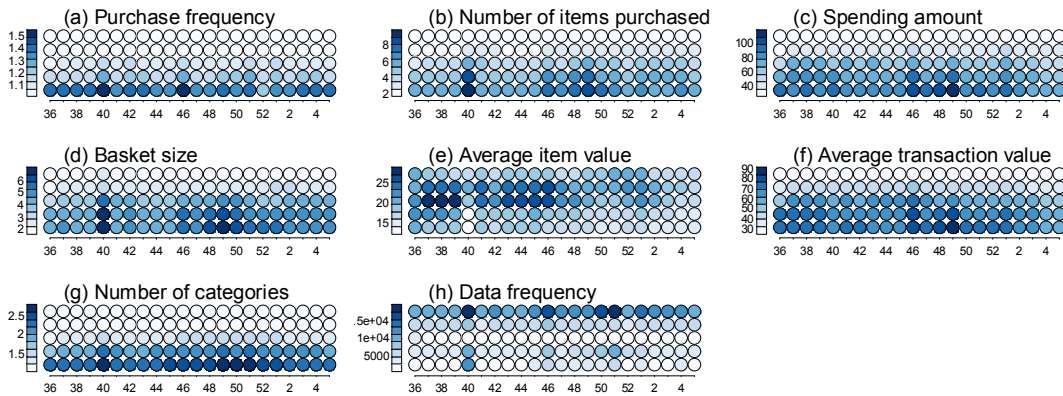


Fig. 4 A Sammon's mapping of the SOTM nodes.

The multidimensionality of Figures 2 and 4 can be described using feature plane representations, as is illustrated in Figure 5. These feature planes are essentially a series of two-dimensional views of the customers, showing information across segments (represented with the vertical dimension), time (represented with the horizontal dimension), and multiple variables (represented by the feature planes). This provides a holistic view of how the characteristics of customer segments evolve over time.



Note: Nodes having the same location represent the same data on all planes. The horizontal axis represents data across 22 weeks, while the five nodes on the vertical direction represent data topology at each time t .

Fig. 5 Feature planes for the SOTM trained by the behavioral variables.

Using the feature planes, we can observe some general characteristics of the customer segments. First, though the data evolve over time, one can roughly identify several groups of customers by comparing the nodes along the vertical direction. The upper part of the map (i.e., the 1st and 2nd rows) represents a group of infrequent and small spending customers with small basket size (cf. Fig. 5 (a, b, c, d)). According to the feature plane showing data frequency, a significant number of customers reside in these nodes. The middle part of the map (i.e., the 3rd row) represents a group of semi-loyal moderate spenders who visited the shop infrequently (cf. Fig. 5 (a)), while purchasing more expensive items (cf. Fig. 5 (e)) and accordingly contributed more to the sales revenue than the customers in the upper part of the map (cf. Fig. 5 (c)). The nodes on the bottom part of the map (i.e., the 4th and 5th rows) represent a group of loyal big spenders who frequently visited the shop (cf. Fig. 5 (a, c)), bought different kinds of products (cf. Fig. 5 (g)) and purchased a large number of products (cf. Fig. 5 (d)).

Since data evolve over time, their temporal changes can be assessed by comparing the nodes along the horizontal direction. Here, we summarize the changes driven by the sales events and Christmas period, respectively. All customer segments, in particular for the nodes on the first row which represents the group of infrequent visitors and small spenders, actively responded to sales event A, as indicated in Fig. 5 (h) by the increased number of customers during week 40. It is also found that customers, especially the loyal big spenders (cf. Fig. 5 (e)), tended to buy lower-priced items during sales event A. On the other hand, these loyal big spenders visited the shop more frequently, bought more items and from more categories. Compared to sales event A, customers responded less actively to sales event B, as indicated by Fig. 5 (h). Customers on the fifth row of the map again displayed increased purchasing frequency. Despite the fact that customers generally purchased fewer numbers of items during sales event B (cf. Fig. 5 (b, d)), the more expensive items they purchased lead to larger average segment-wise spending amounts during sales event B (cf. Fig. 5 (c)). Customers displayed different shopping patterns as Christmas approached. Customers on the fifth row of the map visited the shop more frequently, and bought very large quantities, in the beginning of December (i.e., week 49), followed by a decreasing trend in basket size during the remaining weeks of December. Fig. 5 (g) also indicates that the loyal big spenders started to buy a wide variety of different types of items. In addition, all the segments show a decreasing trend in terms of the prices of the purchased items during the Christmas period.

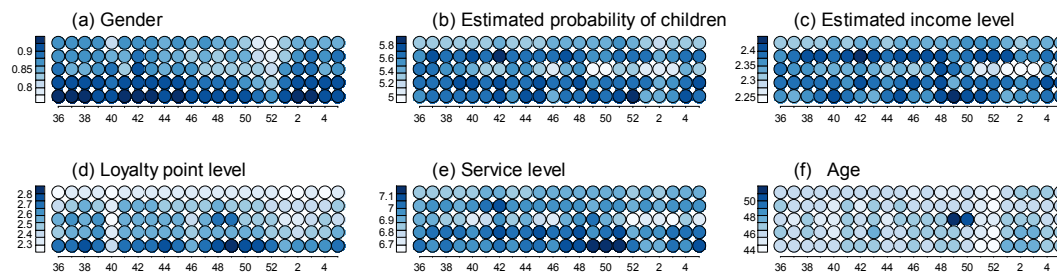
5.3 Analysis based upon the background variables

In contrast to the purchasing behavior variables used in training and analyzed in Section 5.2, the background variables had no impact on forming the segmentation result, but

their distributions among the segments in the 5x22 SOTM grid are shown in the associated feature planes in Figure 6.

By observing the differences in the vertical direction of the feature planes, we can observe some general characteristics of the customer segments. Even though the scales on the left side of the feature planes indicate that the differences among segments are small, some general trends in the cross-sectional differences can still be identified. For instance, female customers are located in the nodes on the lower part of the map (cf. Fig. 6(a)). Customers with high loyalty point (cf. Fig. 6(d)) and service levels (cf. Fig. 6(e)) are also located in the lower parts of the map. This corresponds to the findings in Section 5.2 that high value customers are located in the lower part of the map.

By observing the changes in the horizontal direction, we can observe temporal changes in terms of the background variables. For instance, sales event A attracted customers with lower loyalty point levels (cf. Fig. 6(d)). This to some extent corresponds to the findings in Section 5.2 that sales event A attracted a large number of customers (cf. Fig. 5(h)) who are more price-sensitive (e.g., they bought less expensive items (cf. Fig. 5(e))). The other significant change revealed by the feature plane is that the pre-Christmas period attracted more elderly customers (cf. Fig. 6(f)). Figure 5 shows that these customers have medium purchase behaviour values.



Note: The background variables represented by the feature planes (a) - (f) had no impact on forming the segmentation result. These feature planes display the distribution of the background variables among the segments trained by the purchasing behavior variables. Nodes having the same location represent the same data on all planes. The horizontal axis represents data across 22 weeks, while the five nodes on the vertical direction represent data topology at each time t .

Fig. 6 Feature planes for the SOTM describing the background variables (associated variables).

5.4 Clustering of the SOTM

We apply a second-level clustering on the nodes of the SOTM trained on the behavioral variables in order to better interpret the map. We vary the number of clusters K to explore the structures in the dataset, as is commonly done with hierarchical clustering

methods. In addition, as we do not have a pre-defined number of classes or groups in these data, we use cluster validation measures for evaluating the clustering solutions with different K . Figure 7 shows the Dunn index and Silhouette coefficient for $K = 3, 4, \dots, 10$. While Dunn index indicates that $K = 6, 7, 8, 9$ is optimal and the Silhouette coefficient indicates that $K = 4$ is optimal, a general view of both measures shows only minor differences between different K . The small differences in cluster validation measures motivate us to explore the hierarchical process of agglomerating clusters on the SOTM. The cluster membership planes in Figure 8 illustrate how clusters are agglomerated when increasing K from 3 to 6.

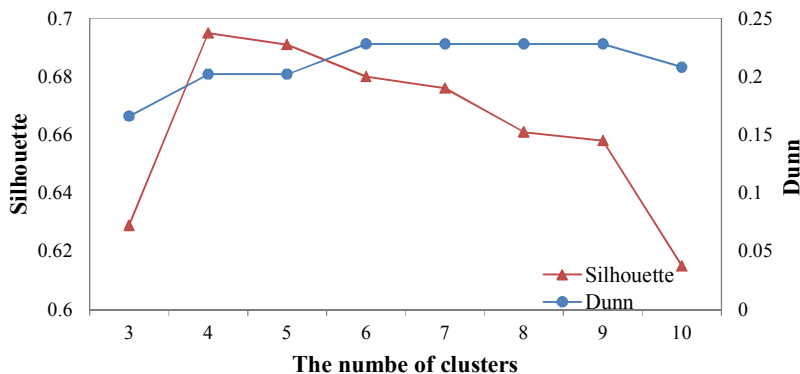


Fig. 7 Cluster validation of the second-level clustering of the SOTM.

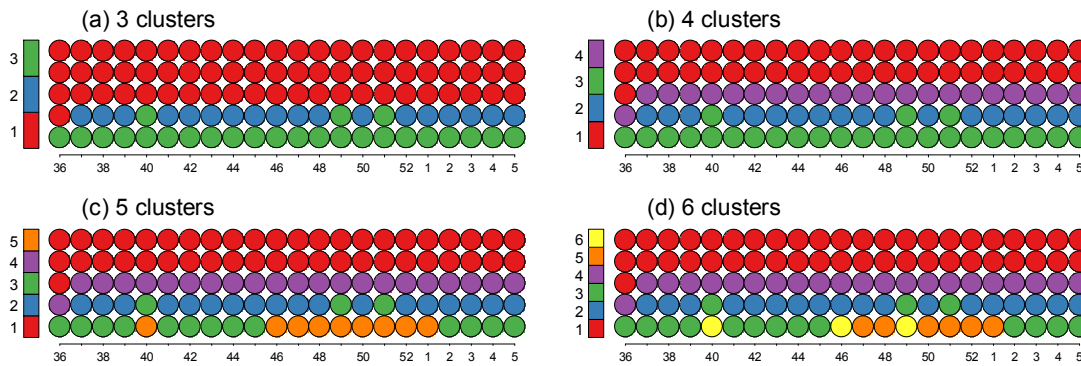


Fig. 8 Cluster membership planes of the SOTM.

While agglomeration proceeds in a bottom-up manner by decreasing K , we summarize the process of the SOTM clustering in a top-down manner for illustrative purposes. First, the 3- and 4-cluster solutions mainly show cross-sectional differences in data, i.e., the lower parts of the map (i.e., the blue and green clusters) represent more loyal and higher-value customers. In the 3-cluster solution, the *changes* during weeks 40, 49 and 51 show that the green cluster moves and merges upwards, indicating the increased

segment size of big spenders during sales event A and the Christmas period. In the 4-cluster solution, the purple cluster emerges to represent medium value customers who exhibit average patterns in most of the behavioral variables. In the 5-cluster solution, the brown cluster first emerges at week 40, and then it disappears and re-emerges starting from sales event B to the end of the Christmas period. The brown cluster essentially represents a group of loyal high spenders who react positively to different sales events. In the 6-cluster solution, the yellow cluster emerges to emphasize the extremely active customers during sales event A, B and in the beginning of December.

6 Conclusion

Dynamics in customer segments is a well documented and broadly acknowledged phenomenon, yet it is seldom explicitly addressed in segmentation approaches. In this paper, we have used the SOTM, in combination with a second-level clustering, for tracking customers' purchasing behaviour and demographics in a department store over time. The model (i) performs multivariate clustering of customers over time; (ii) visualizes the temporal variation of the multivariate patterns; (iii) aids in detecting and interpreting complex patterns of changes in the customer base and purchasing behavior during three special sales events; and (iv) uses a second-level clustering approach to identify changing, emerging and disappearing customer segments in an easily interpretable manner.

We demonstrate the usefulness of the method on a case company's multivariate customer database with weekly data. The SOTM effectively detects the changes in customer behavior during sales events. The results show that the purchasing behavior of customers changes during the events, but at the same time, that the sales events differ in the type of shopping behavior that they trigger. The results are meaningful and interpretable, and indicate that the SOTM can serve as a valuable visual exploratory tool for decision makers to see how successful the sales events have been as to different criteria and to aid in decision making concerning which kind of sales events they should have. For business analysts it can also be used to drill down into the interesting patterns revealed for further analysis.

While this paper provides an approach for visualizing how customer segment compositions change over time, we do not address changes in segment memberships of individual customers, or so-called migration patterns. Following Hu and Rau (1995), future work should aim at simultaneously addressing changes in segment membership of individual objects and changes in segment composition. This could, for instance, be performed by combining a visualization of the SOTM with migration probabilities.

Acknowledgement

The financial support of the Academy of Finland (grant Nos. 127656 and 127592) and the Foundation of Economic Education is gratefully acknowledged.

Reference

- BACK, B., SERE, K. & VANHARANTA, H. 1998. Managing complexity in large data bases using self-organizing maps. *Accounting, Management and Information Technologies*, 8, 191-210.
- BLASHFIELD, R. K. 1976. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83, 377.
- BLOCKER, C. P. & FLINT, D. J. 2007. Customer segments as moving targets: Integrating customer value dynamism into segment instability logic. *Industrial Marketing Management*, 36, 810-822.
- CALANTONE, R. J. & SAWYER, A. G. 1978. The stability of benefit segments. *Journal of Marketing Research*, 15, 395-404.
- CHAKRABARTI, D., KUMAR, R. & TOMKINS, A. 2006. Evolutionary clustering. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. Philadelphia, PA, USA: ACM.
- CHAPPELL, G. J. & TAYLOR, J. G. 1993. The temporal Kohonen map. *Neural Networks*, 6, 441-445.
- CURRY, B., DAVIES, F., PHILLIPS, P., EVANS, M. & MOUTINHO, L. 2001. The Kohonen self - organizing map: an application to the study of strategic groups in the UK hotel industry. *Expert systems*, 18, 19-31.
- DENNY & SQUIRE, D. 2005. Visualization of Cluster Changes by Comparing Self-organizing Maps. In: HO, T., CHEUNG, D. & LIU, H. (eds.) *Advances in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg.
- DENNY, WILLIAMS, G. & CHRISTEN, P. 2010. Visualizing temporal cluster changes using Relative Density Self-Organizing Maps. *Knowledge and Information Systems*, 25, 281-302.
- DUNN, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3, 32-57.
- FARLEY, J. U., WINER, R. S. & LEHMANN, D. R. 1987. Stability of membership in market segments identified with a disaggregate consumption model. *Journal of Business Research*, 15, 313-328.
- FERREIRA DE OLIVEIRA, M. C. & LEVKOWITZ, H. 2003. From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 9, 378-394.
- HA, S. H., BAE, S. M. & PARK, S. C. 2002. Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case. *Computers & Industrial Engineering*, 43, 801-820.
- HAN, J., KAMBER, M. & PEI, J. 2011. *Data Mining: Concepts and Techniques: Concepts and Techniques*, Elsevier Science.

- HANSEN, P. & DELATTRE, M. 1978. Complete-link cluster analysis by graph coloring. *Journal of the American Statistical Association*, 73, 397-403.
- HARROWER, M. & BREWER, C. 2003. Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40, 27-37.
- HOEK, J., GENDALL, P. & ESSLEMONT, D. 1996. Market segmentation: A search for the Holy Grail? *Journal of Marketing Practice: Applied Marketing Science*, 2, 25-34.
- HOLMBOM, A. H., EKLUND, T. & BACK, B. 2011. Customer portfolio analysis using the SOM. *International Journal of Business Information Systems*, 8, 396-412.
- HOMBURG, C., STEINER, V. V. & TOTZEK, D. 2009. Managing dynamics in a customer portfolio. *Journal of Marketing*, 73, 70-89.
- HU, M. Y. & RAU, P. A. 1995. Stability of Usage Segments, Membership Shifts Across Segments and Implications for Marketing Strategy: An Empirical Examination. *Mid-Atlantic Journal of Business*, 31, 161-177.
- HUBERT, L. & SCHULTZ, J. 1976. Quadratic assignment as a general data-analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29, 190-241.
- KIANG, M. Y. & KUMAR, A. 2001. An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications. *Information Systems Research*, 12, 177-194.
- KOHONEN, T. 1982. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43, 59-69.
- KOHONEN, T. 1988. The 'Neural' Phonetic Typewriter. *Computer*, 21, 11-22.
- KOHONEN, T. 1991. The Hypermap Architecture. In: KOHONEN, T., MKISARA, K., SIMULA, O. & KANGAS, J. (eds.) *Artificial Neural Networks*. Amsterdam, Netherlands: North-Holland.
- KUO, R. J., AN, Y. L., WANG, H. S. & CHUNG, W. J. 2006. Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation. *Expert Systems with Applications*, 30, 313-324.
- LINGRAS, P., HOGO, M., SNOREK, M. & WEST, C. 2005. Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Information Sciences*, 172, 215-240.
- MO, J., KIANG, M. Y., ZOU, P. & LI, Y. 2010. A two-stage clustering approach for multi-region segmentation. *Expert Systems with Applications*, 37, 7120-7131.
- PAYNE, A. & FROW, P. 2005. A strategic framework for customer relationship management. *Journal of marketing*, 69, 167-176.
- RAND, W. M. 1971. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66, 846-850.
- ROUSSEEUW, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.
- SAMMON JR, J. W. 1969. A nonlinear mapping for data structure analysis. *Computers, IEEE Transactions on*, 100, 401-409.
- SARLIN, P. 2012. Self-Organizing Time Map: An Abstraction of Temporal Multivariate Patterns. *Neurocomputing*, 99, 496-508.
- SARLIN, P. 2013. Data and dimension reduction for visual financial performance analysis. *Information visualization* (forthcoming).
- SARLIN, P. & YAO, Z. 2013. Clustering of the Self-Organizing Time Map. *Neurocomputing* 121, 317-327.
- SMITH, W. R. 1956. Product differentiation and market segmentation as alternative marketing strategies. *The Journal of Marketing*, 21, 3-8.
- STEENKAMP, J. E. M. & TER HOFSTEDE, F. 2002. International market segmentation: issues and perspectives. *International Journal of Research in Marketing*, 19, 185-213.

- THEODORIDIS, S. & KOUTROUMBAS, K. 2008. *Pattern Recognition*, Elsevier Science.
- THOMAS, M. 1980. Market segmentation. *Quarterly review of marketing*, 6.
- TIBSHIRANI, R., WALTHER, G. & HASTIE, T. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 411-423.
- TSENG, M. M. & PILLER, F. 2003. *The customer centric enterprise-advances in mass customization and personalization*, Springer.
- VESANTO, J. 1999. SOM-based data visualization methods. *Intelligent Data Analysis*, 3, 111-126.
- VESANTO, J. & ALHONIEMI, E. 2000. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11, 586-600.
- WIND, Y. 1978. Issues and advances in segmentation research. *Journal of marketing research*, 15, 317-337.
- YAO, Z., HOLMBOM, A., EKLUND, T. & BACK, B. 2010. Combining unsupervised and supervised data mining techniques for conducting customer portfolio analysis. *10th Industrial conference on data mining*. Germany, Berlin: Springer.
- YAO, Z., SARLIN, P., EKLUND, T. & BACK, B. Combining visual customer segmentation and response modeling. 20th European Conference on Information Systems, 2012a Barcelona, Spain.
- YAO, Z., SARLIN, P., EKLUND, T. & BACK, B. Temporal Customer Segmentation Using the Self-organizing Time Map. 16th International Conference on Information Visualisation (IV), 2012b Montpellier, France. IEEE, 234-240.

Research Paper 5

Sarlin, P., Yao, Z., & Eklund, T. (2012), A Framework for State Transitions on the Self-Organizing Map: Some Temporal Financial Applications. *Intelligent Systems in Accounting, Finance and Management*, Vol.19, No.3, pp. 189-203.

Extends:

Sarlin, P., Yao, Z., & Eklund, T. (2012), Probabilistic Modeling of State Transitions on the Self-Organizing Map: Some Temporal Financial Applications. In *Proceedings of the 45th Annual Hawaii International Conference on Systems Science (HICSS 2012)*, Maui, USA, 4-7 January, 2012, pp. 1128-1137.

A FRAMEWORK FOR STATE TRANSITIONS ON THE SELF-ORGANIZING MAP: SOME TEMPORAL FINANCIAL APPLICATIONS

PETER SARLIN,* ZHIYUAN YAO AND TOMAS EKLUND

Turku Centre for Computer Science, Department of Information Technologies, Åbo Akademi University, Turku, Finland

SUMMARY

Self-organizing maps (SOMs) have commonly been used in temporal applications. This paper enhances the SOM paradigm for temporal data by presenting a framework for computing, summarizing and visualizing transition probabilities on the SOM. The framework includes computing matrices of node-to-node and node-to-cluster transitions and summarizing maximum state transitions. The computations are linked to the SOM grid using transition-plane visualizations. We demonstrate the usefulness of the framework on two SOM models for temporal financial analysis: financial performance comparison of banks and monitoring indicators of currency crises. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: state transitions; self-organizing map (SOM); temporal data; financial institutions; currency crises

1. INTRODUCTION

Today's decision makers are often faced by enormous amounts of financial data available for decision-making purposes. Access to online financial databases, such as Thomson One, Amadeus and Bankscope, can provide nearly endless amounts of multivariate financial time-series data. However, owing to nonlinear relationships, high dimensionality and non-normality often inherent in financial data, utilizing these data can be a significant challenge for traditional statistical tools and spreadsheet programs. Instead, various data-mining and pattern recognition tools have been applied for this purpose.

One potential tool is the self-organizing map (SOM; Kohonen, 1982, 2001), an unsupervised neural network-based projection and clustering method often used for exploratory data analysis. Although most of the early SOM applications have been in the area of medicine and engineering (Oja *et al.*, 2002), the SOM has also been used in a large number of financial applications (Deboeck and Kohonen, 1998), including financial performance comparison (Back *et al.*, 1998; Eklund *et al.*, 2003), bankruptcy prediction (Martín-del-Brío and Serrano-Cinca, 1993; Kiviluoto, 1998), financial crisis monitoring (Sarlin and Marghescu, 2011; Sarlin and Peltonen, 2011), economic welfare analysis (Kaski and Kohonen, 1996), customer churn analysis and segmentation (Lingras *et al.*, 2005; Kiang *et al.*, 2006) and stock price forecasting (Afolabi and Olude, 2007; Hsu *et al.*, 2009), just to name a few.

The general SOM paradigm is an ideal tool for building visualization systems, as it reduces both dimensionality and data; however, manually identifying the positions and patterns in a SOM model is not necessarily a simple process. As the applications above illustrate, financial data typically belong

* Correspondence to: Peter Sarlin, Department of Information Technologies, Åbo Akademi University, Turku Centre for Computer Science, Joukahaisenkatu 3–5, 20520 Turku, Finland. Email: psarlin@abo.fi

to a time series. Variations of the SOM algorithm itself, using delayed or reinforced learning, have been proposed for dealing with temporal data by using leaky integrators or recurrent networks (e.g. Barreto and Araújo, 2001; Guimarães *et al.*, 2003; Hammer *et al.*, 2005). These extensions, however, turn their focus from the entities to the time series. In many fields, such as accounting, finance and economics (e.g. Back *et al.*, 1998; Eklund *et al.*, 2003; Sarlin and Peltonen, 2011) and process monitoring (e.g. Alhoniemi *et al.*, 1999; Fuertes *et al.*, 2010), it is more common than not for data to include both a temporal and a cross-sectional dimension. For instance, when data are cyclical (or scarce), one may want to build a standard SOM model with data on several entities over time to include the temporal and cross-sectional differences. Given a model with this type of panel data, an obvious interest would be the temporal properties of the model.

Sarlin and Eklund (2011a, 2011b) extended the SOM to show membership degrees of each time-series point to each cluster using fuzzy *C*-means clustering. This is suitable for assessing the current state, but says nothing about future transitions. Hence, a method for illustrating state transitions is required. Most often, trajectories have been used in the SOM literature to illustrate these state transitions (e.g. Martín-del-Brío and Serrano-Cinca, 1993; Eklund *et al.*, 2003; Sarlin, 2010). While state transition patterns require a large number of observations for significance, trajectories can only be used on a limited set of data in order not to clutter the display. Thus, they provide no overall information about trends in the dataset. For finding these patterns, be they cyclical or not, the switches should be summarized from transition probabilities, something that is not apparent from studying the elements of the SOM units. These types of transition probabilities generalize the strengths and actual directions of the temporal patterns on the SOM.

For modeling future state transitions, Afolabi and Olude (2007) and Hsu *et al.* (2009) proposed hybrid approaches combining standard SOMs with machine-learning classification techniques like neural networks and support vector machines. While being appropriate for time-series prediction, these suffer from high complexity and computational cost, as well as impaired visualization capabilities. Methods to directly deal with state transitions in SOMs have been introduced in previous applications through transition probabilities. Somervuo (2000) applied competitive hidden Markov models on the SOM for speech recognition. More recently, one-level node-to-node transition probabilities have been used in temporal modeling and clustering, e.g. for finding profiles of forest nutrition (Sulkava and Hollmén, 2003; Luyssaert *et al.*, 2004) and industrial process modeling and fault detection (Fuertes *et al.*, 2010).

In this paper, we propose a framework that further enhances previous attempts to exploit and visualize transition probabilities on the SOM. In particular, two properties are emphasized:

1. computing transition probabilities for a two-level clustering;
2. linking their visualization to the SOM grid.

A two-level clustering refers to using the SOM for clustering data into a large number of nodes and then grouping the nodes of the SOM into clusters, enabling more detailed representations. By computing state transitions for a two-level clustering, we enable discovery of detailed node-to-cluster patterns, while not dealing with enormous transition probability matrices (TPMs). Differences in transition probabilities between nodes within the same clusters reveal additional information, such as stability and cyclicity, since a homogeneous cluster does not necessitate similar state transitions. We also emphasize a user-oriented and easily interpretable visualization of the TPMs. As the transition probabilities are associated with the nodes of an SOM, their visualization can be linked to the regularly shaped SOM grid. We call these low-dimensional representations of transition probabilities ‘transition planes’.

The contribution of this paper is a novel framework for computing, summarizing and visualizing transition probabilities in a user-friendly format, as well as its application to two temporal financial problems. The SOM has been shown to be a particularly feasible tool for building visual monitoring systems for financial performance comparison of companies (Eklund *et al.*, 2008) and for monitoring indicators of country-level financial crises (Sarlin and Peltonen, 2011). We show the added value of the transition-probability framework by applying it to two models in these fields: financial performance comparison of banks (Sarlin and Eklund, 2011) and monitoring indicators of currency crises (Sarlin, 2010). In addition, we also show examples of how the transition information can be used in practice. In financial performance analysis of banks, we illustrate how the evolution of and reaction to transition probabilities over time for an individual entity can be represented using line graphs. In monitoring indicators of currency crises, we use the transition probabilities for profiling by presenting low- and high-risk mean profiles based upon future transitions.

The paper is structured as follows. Section 2 introduces the framework for computing, summarizing and visualizing transition probabilities on the SOM. In Section 3, the framework is applied on financial time series, while Section 4 concludes by presenting our key findings.

2. METHODOLOGY

This section briefly introduces the SOM and transition probabilities, as well as the framework for computing, summarizing and visualizing transition probabilities on the SOM.

2.1. Self-Organizing Maps

The SOM is a method with simultaneous clustering and projection capabilities first developed by Kohonen (1982). As the SOM algorithm is well known and the main emphasis is on transitions on the SOM, we do not present details of it here – for further reference, see Kohonen (2001). The *Viscovery SOMine 5.1* package is used in this study mainly for its excellent visual representation. The training process starts with a linear initialization of the reference vectors. The first step compares all input data vectors x_j (where $j = 1, 2, \dots, N$) with the network's reference vectors m_i (where $i = 1, 2, \dots, M$) to find the best match m_b :

$$\|x_j - m_b\| = \min_i \|x_j - m_i\| \quad (1)$$

Then the second step adjusts each reference vector m_i with the batch updating formula:

$$m_i(t+1) = \frac{\sum_{j=1}^N h_{ib(j)}(t)x_j}{\sum_{j=1}^N h_{ib(j)}(t)} \quad (2)$$

where t is a discrete time coordinate and $h_{ib(j)}$ is a decreasing function of neighbourhood radii and time.

The SOM describes a multidimensional space on a regularly shaped two-dimensional grid of nodes. While all information cannot be visualized in two dimensions, the regular shape of the SOM grid facilitates linking additional information to it. To further enhance the representation of all dimensions, variables are separately shown on their own grids. Each feature plane displays the distribution of that variable on the map, with cold color (blue) indicating low values and warm color (red) indicating high values. As the feature planes are different views of the same map, one unique point represents the same node on all planes. Thereby, the characteristics of the SOM model can be identified by studying the underlying feature planes.

The nodes of the map can further be divided into clusters of similar nodes. We use hierarchical clustering with the following modified Ward's (1963) distance criterion as a basis for merging two candidate clusters:

$$d_{kl} = \begin{cases} \frac{n_k n_l}{n_k + n_l} \cdot \|c_k - c_l\|^2 & \text{if } k \text{ and } l \text{ are adjacent} \\ \infty & \text{otherwise} \end{cases} \quad (3)$$

where k and l represent clusters, n_k and n_l the cardinality of clusters k and l , and $\|c_k - c_l\|^2$ the squared Euclidean distance between the cluster centers of clusters k and l , and the distance between non-adjacent clusters is infinite. When clusters k and l are merged to cluster h , the cardinality is the sum of the cardinalities of k and l and the centroid the mean of c_k and c_l weighted by their cardinalities.

2.2. Transition Probabilities on the SOM

TPMs produce a probabilistic model of the temporal variation in an SOM model. The two-dimensional grid of Section 2.1 is used to compute probabilities of switching from each node to a specified region in a specified time period, where the location per time unit is derived using equation (1). First, we compute for each node m_i the probability of transition to every other node m_u :

$$p_{iu}(t+s) = \frac{n_{iu}(t+s)}{\sum_{u=1}^M n_{iu}(t+s)} \quad (4)$$

where n_{iu} is the cardinality of data switching from m_i to m_u , t is a time coordinate and s is the time span for the switch. In other words, the transition probability $p_{iu}(t+1)$ equals the cardinality of transitions from node m_i to node m_u divided by the sum of transitions from node m_i to $m_{1,2,\dots,M}$. On an SOM grid with four nodes, this could in practice mean that for, say, node m_1 the probability of being in period $t+1$ in $m_{1,2,\dots,4}$ could be 0.5, 0.2, 0.2 and 0.1 respectively. Formally, a TPM corresponds to maximum-likelihood estimates of the switches or a first-order Markov model. It can, however, be computed for different time spans, as appropriate, and summarized to switches between clusters or any other region on the map. For example, node-to-cluster switches are computed using p_{il} , where the transition refers to movements from reference vector i to cluster l (where $l=1, 2, \dots, C$), thus:

$$p_{il}(t+s) = \frac{n_{il}(t+s)}{\sum_{l=1}^C n_{il}(t+s)} \quad (5)$$

For correcting coincidental results due, for example, to lack of data, the TPMs p_{il} (as well as p_{iu}) can be computed as an average of several s values (where $s = 1, 2, \dots, S$):

$$p_{il}(t + \{1, 2, \dots, S\}) = \frac{\sum_{s=1}^S n_{il}(t+s)}{\sum_{s=1}^S \sum_{l=1}^C n_{il}(t+s)} \quad (6)$$

In practice, transition probabilities need to be summarized into easily interpretable formats rather than massive node-to-node TPMs. Thus, we propose the following three computations:

1. TPMs for node-to-cluster switches ($p_{il}(t+s)$ as in equation (5)) for a specified set of s values.
2. Summarize the TPMs from Step 1 by computing to which cluster l an observation in m_i is most likely to switch and with what likelihood; that is, showing maximum transition probabilities $\max_l(p_{il})$ conditional on switching. This combines the direction and strength of all probabilities into a vector.
3. For summarizing the computations in Steps 1 and 2 over time, compute average transition probabilities over a chosen set of s values ($p_{il}(t + \{1, 2, \dots, S\})$) as in equation (6).

Similarly, as with feature planes for individual variables, transition probabilities for nodes can be visualized on their own transition planes, where one unique point represents the same node on the previously presented SOM grid. Thereby, the structure of the transitions on the SOM model can be directly identified by studying the underlying transition planes. The above computations are represented using the following three transition-plane visualizations:

1. Show the probability to transit to a particular cluster for each node on own transition planes, such that the color code of each node represents its probability to transit to that particular cluster.
2. Summarize the information in Step 1 to one transition plane, where the color code in each node is the probability of the most likely switch and a label represents that cluster (non-transiting nodes are left empty).
3. Create the same transition planes as in Steps 1 and 2, but as an average over a chosen set of s values.

While these computations and visualizations effectively summarize transition information, one still needs to pay regard to the representation of transitions when setting the number of nodes and clusters of the SOM. In particular, when the number of clusters approaches the number of nodes (i.e. $C \rightarrow M$) then the node-to-cluster probabilities approach node-to-node probabilities (i.e. $p_{il} \rightarrow p_{iu}$), and this obviously complicates interpretation and decreases the significance of the results. When the number of nodes increases, the significance of probabilities p_{il} and p_{iu} decreases with the decreased data frequency in units. Hence, it is worth noting that the main limitation of state-transition modeling is the requirement of large datasets or small SOMs. To normalize the color scales for different cluster sizes, but still show differences over time, we set the color scales of the transition planes for all s values and sets of s values

as to that of $s = 1$ (i.e. $t + 1$). The temporal dimension of an individual entity can as well be represented by associating each time-series point with the transition probability of its best-matching unit (equation (1)). This enables a line graph representation of the state switch probabilities over time, where clusters are representative financial states and the variation in transition probabilities are indications of future financial performance. The transition probabilities can also be used for profiling by presenting characteristics of low- and high-risk mean profiles based upon future transitions.

3. SOME TEMPORAL FINANCIAL APPLICATIONS

In this section we present two applications of temporal financial analysis: financial performance comparison of the European banking sector and monitoring indicators of financial crises.

3.1. Financial Performance Comparison

Financial performance comparison is a commonly used technique in many industries today and involves the process of comparing the financial performance of one organization against one or more competitors. It can be seen as a subset of performance comparison in general and can be defined as the systematic, broad comparison of performance along several or all subdimensions of organizations' financial measures, not only, for example, the probability of default or bankruptcy.

The temporal nature of financial performance comparison introduces several challenges to an analyst. In part, the multivariate nature of the problem (multiple financial measures for a number of organizations and several years) places requirements on the method used, in particular in terms of visualization capabilities. Also, there are well-known difficult properties of the financial ratio data themselves, with nonlinearity, intercorrelation, outliers, non-normal distributions and so on (e.g. Deakin, 1976; Lev, 1974; Salmi and Martikainen, 1994) that require the use of particularly robust techniques. For these reasons, Back *et al.* (1998) and Eklund *et al.* (2003) proposed the use of the SOM for financial performance comparison, applying the method to the analysis of the international pulp and paper industry. In Eklund *et al.* (2008), the application from Eklund *et al.* (2003) was validated by Finnish managers who found the SOM to be a feasible tool for financial performance comparison.

The previous results in this area motivated us to use the SOM as a financial performance comparison tool for the analysis of the European banking sector. However, as was previously mentioned, the problems related to visual temporal analysis using the SOM will be addressed using the transition probability approach presented in this paper.

An SOM for Financial Performance Comparison of European Banks

The financial performance comparison model was created as a complement to the ongoing EU stress testing of the European banking sector, on account of the fallout from the recent financial crisis. The data – retrieved from the Bankscope financial database and consisting of 24 financial ratios for 855 European banks – covered annual data for the period 1992–2008, resulting in a total of 9,655 rows of data. The SOM model was created by first applying PCA to obtain seven subdimensions of financial performance: *capital ratios* (PC1), *loan ratios* (PC2), *profitability* (PC3), *interest revenue* (PC4), *non-operating items* (PC5), *subordinated debt* (PC6) and *loan loss provisions* (PC7). In general, high values are better, with the exception of loan ratios, which reflect the inverse of capital to net loans and the liquidity of a bank's assets, and loan loss provisions, which measure the ratio of risk to interest rate margins. In addition to these, the ratios Tier 1 and total (Tier 1 + Tier 2) capital were associated with the trained map, as these are the most important ratios from the perspective of stress testing. Moreover,

Ward's (1963) method was used for second-level clustering. The number of nodes and clusters were set according to evaluation criteria based upon map and clustering quality measures. For further details of the model, readers are referred to Sarlin and Eklund (2011b). The SOM model, with data for Deutsche Bank and Banco Santander from 2002–2008, is displayed in Figure 1 and its feature planes are in Figure 2.

By analyzing the spread of individual inputs on the feature planes, we get an overview of the characteristics of the clusters. Using this information, we group the clusters into those with good, average and poor banks. The banks in clusters A, B and C can be seen as good performers, where A can be seen as the best. The banks in cluster D are average performers, as they show average values for all variables. The banks in clusters E, F and G can be seen as poor performers, where F can be seen as the poorest. Since TPMs to seven clusters impair the interpretability of switches, and these clusters can easily be grouped by performance, we consider the best clusters as one (A, B and C), the average cluster as one (D) and the poor clusters as one (E, F and G).

Transition Probabilities on the Financial Performance Comparison Model

We follow the above three-step framework when computing the transition probabilities for the financial performance comparison model. First, TPMs were computed as switches from nodes to clusters ($p_{il}(t+s)$ as in equation (5)). Second, the direction and strength of the switches are summarized by computing maximum transition probabilities $\max(p_{il})$ conditional on switching. Third, we compute the above steps for three different transition time spans ($t+1$, $t+2$ and $t+3$) and an average for $S=3$. A transition-plane visualization of these computations is shown in Figure 3.

The transitions on the SOM reveal several interesting patterns. The poorest group (clusters E, F and G) can be seen as inherently stable, as there are few transitions from the nodes in these clusters, irrespective of the time span. The best group (clusters A, B and C), on the other hand, is less stable, and the probability of transition from the nodes in this group increases with time. The average group (cluster D) is an instable transition cluster. Based on the average probabilities, we can see that banks

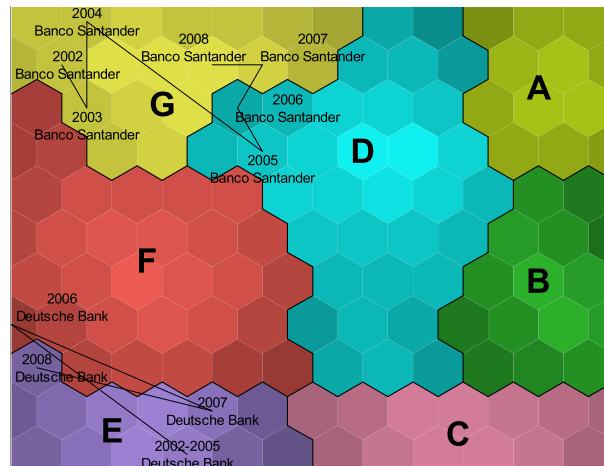


Figure 1. The SOM grid of the bank model.

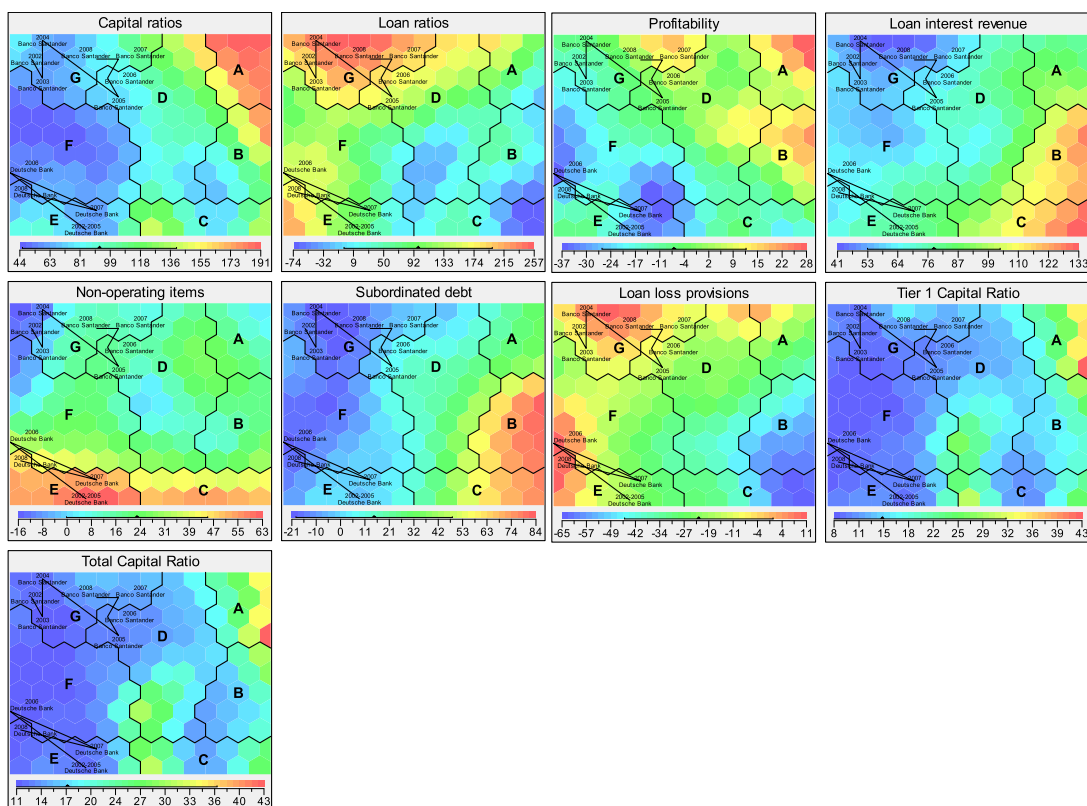


Figure 2. Feature planes of the bank model.

in clusters A, F and G are quite stable, while clusters B, C and E exhibit more transitions, and thus less stability.

The difference in stability between cluster A and clusters B and C might be due to differing business activities, as the feature planes in Figure 2 show that cluster A differs from B and C primarily in capital ratios, loan interest revenue and subordinated debt. These ratios indicate that clusters B and C are higher risk clusters than cluster A, and thus probably more sensitive to changing business conditions, such as interest rates. For cluster B, an interesting strong cluster-to-cluster pattern is the high probabilities of movements to cluster D (average performance).

Another interesting pattern is the difference in stability between cluster E and clusters F and G. While E, F and G are quite similar clusters in terms of performance, a clear difference can be seen in the high ratio of non-operating items of cluster E. Non-operating items are items not related to ongoing, day-to-day operations, such as dividends, financial investments or significant write-downs, which might partially explain the unstable nature of positions in cluster E.

The line graphs in Figure 4 show a practical bank-specific application of the transition-probability framework. The figure shows the state transition probabilities for Deutsche Bank and Banco Santander for 2002–2008 (see Figure 1 for their trajectories). It enables assessing the evolution of probabilities over time, where the clusters centers express the representative financial states while the probability fluctuations represent their variation over time. A label above each time-series point denotes the cluster in which the

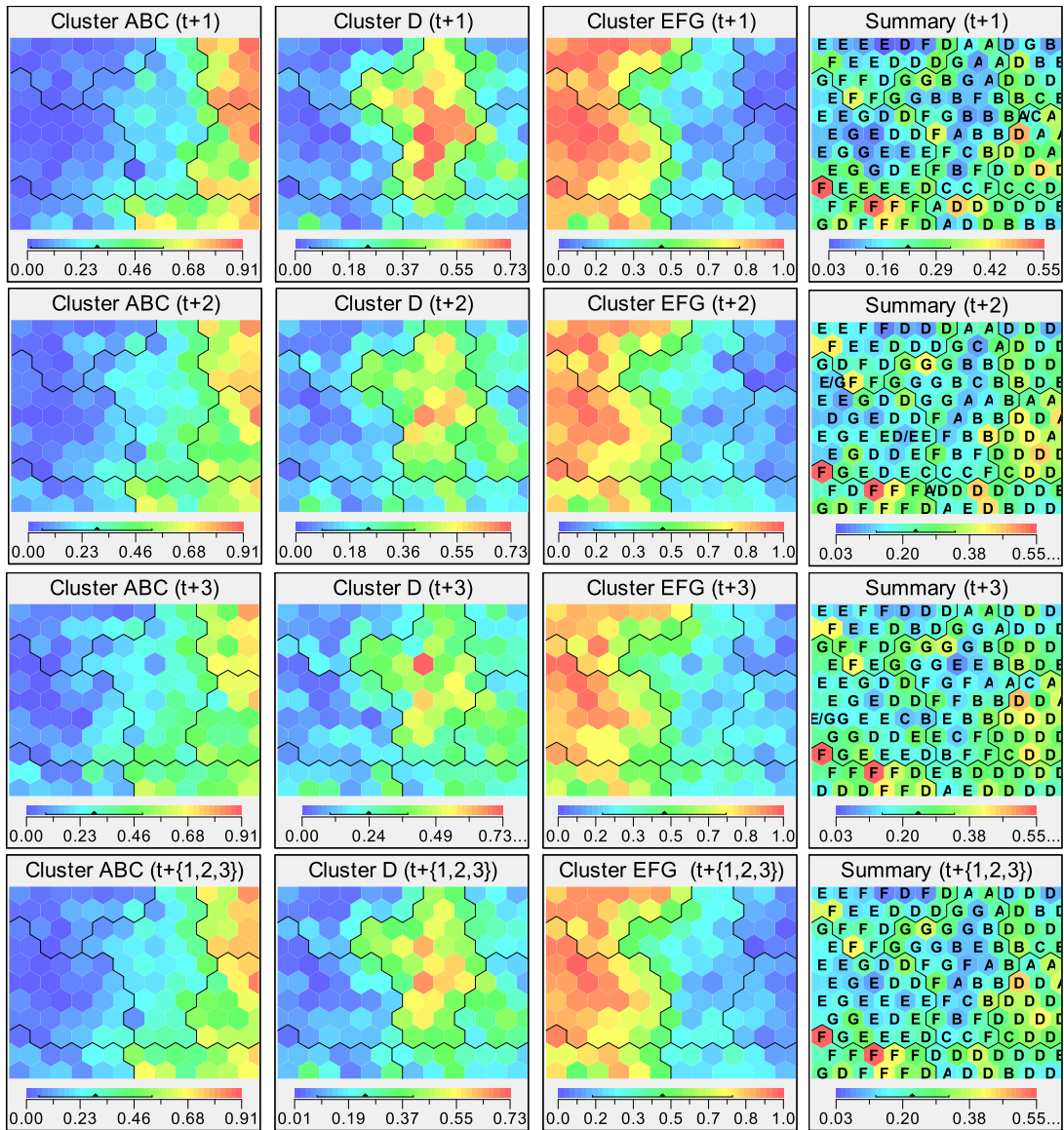


Figure 3. Transition planes of the bank model's transition probabilities. *Notes:* On each row, the first three transition planes represent the probability to switch to the corresponding clusters (or group of clusters). The names of the transition planes refer to the addresses of the transition probabilities. The last transition plane on each row summarizes the direction and strength of the switches by showing maximum transition probabilities, where the greyscale/color coding represents the probability, and the label the address of the switch.

bank is currently located. Since the probability of staying in a cluster is most commonly highest, the interesting patterns are the addresses of the switches and the probability trend of the most likely switch. The transition probabilities for Deutsche Bank are dominated by two clusters, E and F. The

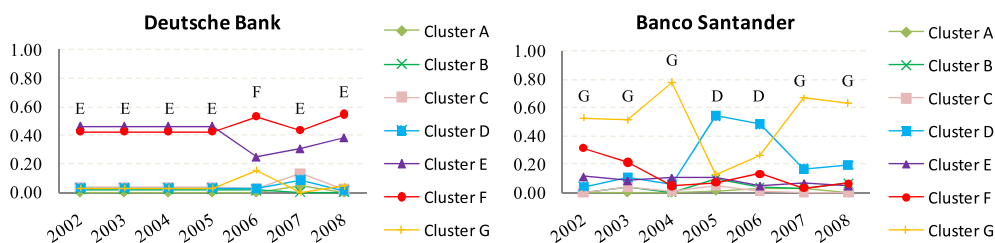


Figure 4. Line graphs of the transition probabilities for Deutsche Bank and Banco Santander.

switches between clusters E and F are thus expected, and the final trend indicates a future movement back to cluster F. For Banco Santander, while the first movement from cluster G to D is unexpected, the movement back to G is accordingly preceded by an increasing trend of the transition probability to that cluster.

3.2. Financial Crisis Monitoring

Recent occurrences of financial crises have demonstrated the importance of understanding the sources of vulnerability. The empirical literature has mainly dealt with early warning systems that rely on conventional statistical modeling methods, such as the univariate signals approach (e.g. Kaminsky *et al.*, 1998) or multivariate logit/probit models (e.g. Berg and Pattillo, 1999), for computing a probability of crisis. However, conventional statistical methods may fail in modeling these complex events driven by multidimensional non-normally distributed and nonlinearly related economic and financial factors, especially in developing models with clear visual capabilities and intuitive interpretability. For addressing these issues, the SOM has been utilized for assessing indicators of different types of financial crises, such as currency crises (Sarlin, 2010; Sarlin and Marghescu, 2011), debt crises (Sarlin, 2011) and systemic financial crises (Sarlin and Peltonen, 2011). Sarlin and Marghescu (2011) and Sarlin and Peltonen (2011) also showed the SOM to outperform a probit/logit model in predicting the events of distress.

From Minsky's (1982) and Kindleberger's (1996) financial fragility view of a credit or asset cycle, it is obvious that these types of data move through stages of a financial stability cycle, such as tranquil, pre-crisis and crisis periods. In the above mentioned applications, the SOM has successfully been combined with trajectory analysis for assessing the performance of individual economies. However, it is difficult to judge the probability of moving to some stage of the cycle from a stand-alone SOM grid and trajectory analysis. Here, we show how assessing strengths and directions of transition patterns on the SOM further enhances the usefulness of an SOM-based visualization system for indicators of currency crises.

An SOM for Crisis Monitoring

The currency crisis model was created for visual monitoring of currency crisis indicators. The model consists of four monthly indicators for 23 emerging market economies from 1971:1 to 1997:12. The indicators included are foreign exchange reserve loss, export loss, real exchange-rate overvaluation relative to trend and current account deficit to GDP, and were chosen and transformed based on a seminal early-warning system created by IMF staff (Berg and Pattillo, 1999). This model, however, is conceptually different from the performance comparison model. Each data point has a class dummy indicating the occurrence of a crisis, pre-crisis or tranquil period. A crisis period is defined to occur

when exchange-rate and reserve volatility exceed a specified threshold, while the pre-crisis periods are defined as 24 months preceding a crisis and the rest of the periods are tranquil. The class labels were associated with the model by only affecting the updating of the reference vectors (equation (2)), not

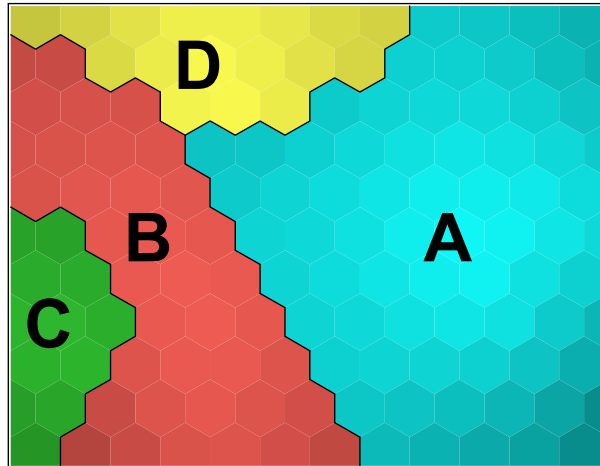


Figure 5. The SOM grid of the crisis model.

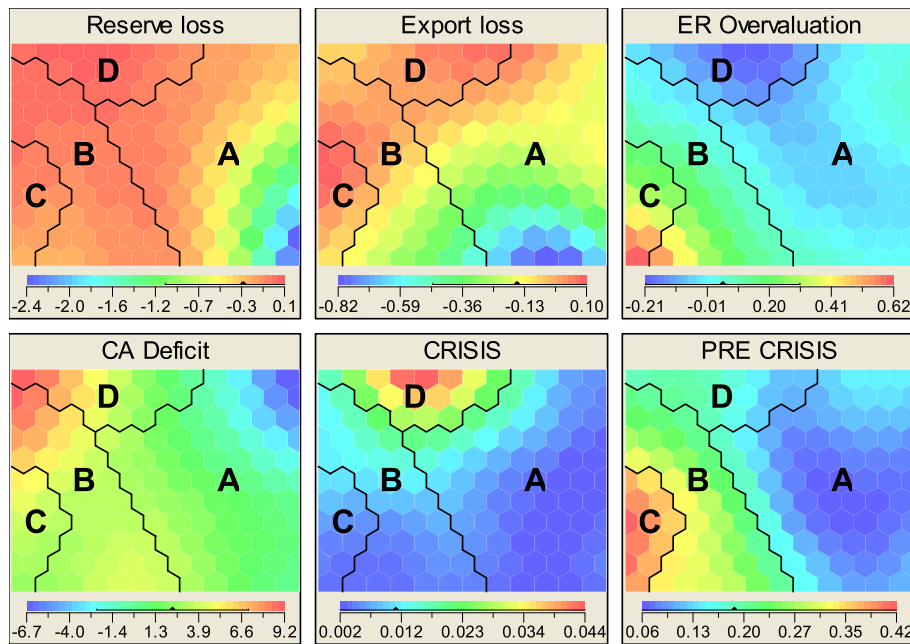


Figure 6. Feature planes of the crisis model.

in finding the best-matching units (equation (1)). The crisis model represents cyclical behaviour resembling a currency crisis or financial stability cycle. Thus, the main purpose of the model is to visualize the evolution of financial indicators to assist the detection of vulnerabilities or threats to financial stability. The model is presented in detail in Sarlin (2010) and a model on the same dataset is evaluated in terms of out-of-sample accuracy in Sarlin and Marghescu (2011).

Each input is standardized using columnwise normalization by range. However, the effects of extremities and outliers are not eliminated, since a crisis episode is *per se* an extreme event. The map consists of 137 nodes ordered on a 13×11 lattice, divided into four crisp clusters representing different stages of the currency crisis cycle. The number of nodes is set so as to allow enough detail while still keeping the map as small as possible, and the number of clusters is derived from the stages of the currency crisis cycle. The units were clustered using Ward's (1963) hierarchical clustering on the associated variables. The map and its feature planes are shown in Figures 5 and 6. The map is roughly divided into a tranquil

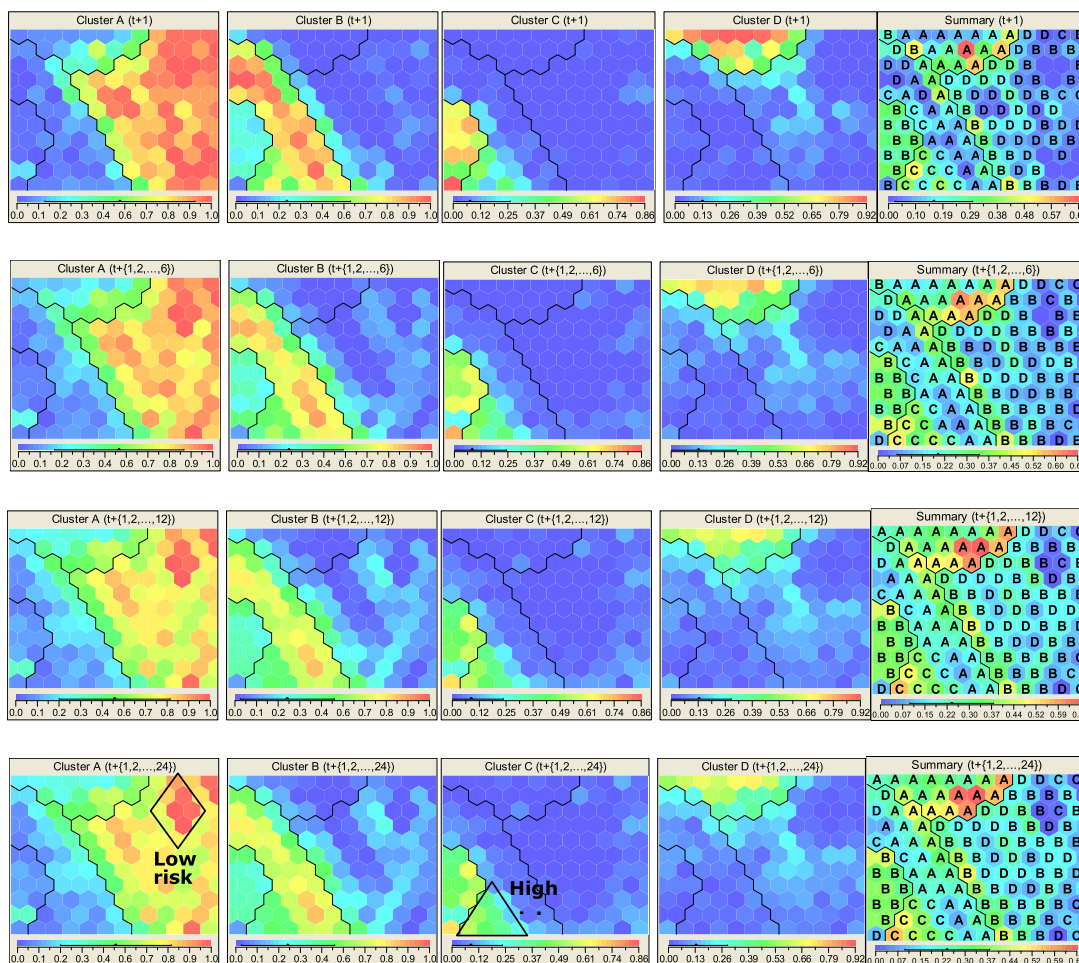


Figure 7. Transition planes of the crisis model's transition probabilities

cluster on the right side of the map (cluster A), a slight early-warning and a pre-crisis cluster in the lower-left part (clusters B and C), and a crisis cluster in the upper-left part (cluster D).

Transition Probabilities on the Financial Crisis Model

We again follow the three-step framework when computing the transition probabilities for the financial crisis model. The first two steps were computed as above, but the third step is computed for one individual time span ($t + 1$) and for three averages for $S = \{6, 12, 24\}$. The transition-plane visualizations of these computations are shown in Figure 7.

Similar to the performance comparison model, the transitions on this SOM reveal several patterns. The cyclical behaviour in the model, however, makes it conceptually different. The transition probabilities in Figure 7 show that the cyclicity follows the four clusters representing stages of the currency crisis cycle. The transition-plane representation also shows that for longer averages the clusters, obviously, become less stable. The summarized transition planes reveal that most of the nodes in the tranquil cluster A switch to the early-warning cluster B, while a group of mid-cluster nodes have a high probability of switching back to the crisis cluster D. Similarly, nodes in cluster B adjacent to cluster A have a high probability of moving to A while those adjacent to C move to C. From the pre-crisis cluster C, the highest probabilities are to switch to cluster B and then further on to the crisis cluster D.

The transition probabilities are used for country profiling by presenting low- and high-risk mean profiles. The low-risk profile is a group of stable nodes with a high probability of staying in the tranquil cluster A, while the high-risk profile is a group of nodes with the high probabilities of moving to the pre-crisis cluster C (see Figure 7). This type of profiling is important, as it allows disentangling the individual sources of risk and vulnerability. Student's t -tests on the difference between high- and low-risk profiles and the average country are shown in Figure 8. For the high-risk profile, the exchange rate overvaluation is significantly larger. For the low-risk profile, the current account deficit is significantly smaller than the average. The rest of the variables, both for high- and low-risk profiles, are close to the average country.

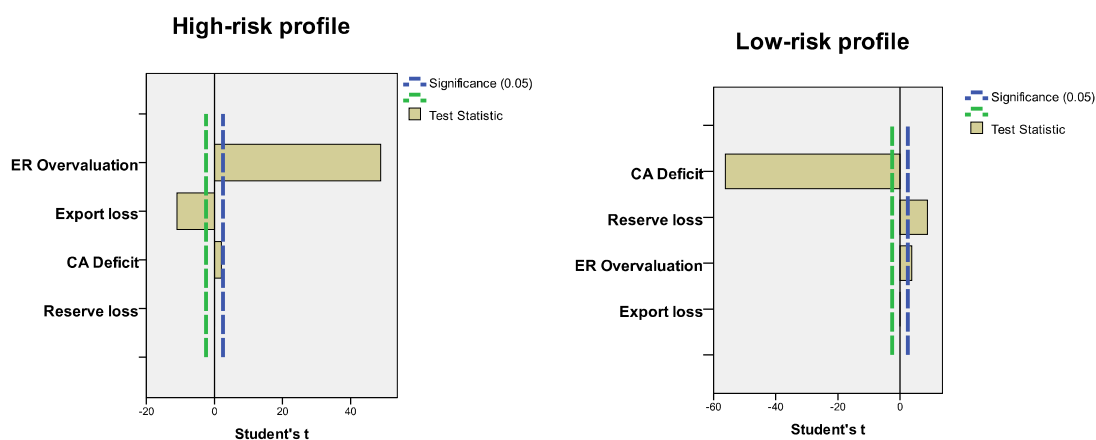


Figure 8. Student's t -tests on profiles of high and low risk and the average country.

4. CONCLUSION

This paper enhances the SOM paradigm for temporal data by presenting a novel framework for computing, summarizing and visualizing transition probabilities on the SOM. The framework includes computing matrices of node-to-node and node-to-cluster transitions and summarizing maximum state transition. The computations are visualized using transition-plane representations. The future state transitions can also be used for finding low- and high-risk profiles as well as for assessing the evolution of probabilities over time, where the cluster center express the representative financial states while the probability fluctuations represent their variation over time.

We demonstrate the usefulness of the framework on two SOM models for temporal financial analysis: financial performance comparison of banks and monitoring indicators of currency crises. In addition to transition patterns assessed for both models, we show an assessment of transition probabilities, and reactions to them, over time for two banks and conduct low- and high-risk country profiling using the financial crisis model. Thus, while the information products of the standard SOM paradigm have been evaluated to bring significant benefits compared with previously used methods (e.g., Eklund *et al.*, 2008), the framework for assessing strengths and directions of transition patterns on the SOM further enhances the usefulness of SOM-based visualization systems. However, it is worth noting that our intention is not to provide an optimal modeling approach for predictive purposes (this aim may be better achieved with some other methods), but to provide an approach for visual analysis and understanding of temporality and cyclicity in data. In addition to financial applications, the transition-probability framework could also be applied to a broad range of other types of temporal problems, such as customer segmentation and response modeling as well as various forms of industrial process modeling and fault detection.

ACKNOWLEDGEMENTS

We gratefully acknowledge three anonymous reviewers and the editor Daniel O’Leary for useful comments and suggestions, as well as the financial support of the Academy of Finland (grant nos. 127656 and 127592). This paper is an extension of work originally presented at the Hawaii International Conference on System Sciences HICSS’45 (2012).

REFERENCES

- Afolabi MO, Olude O. 2007. Predicting stock prices using a hybrid Kohonen self organizing map (SOM). In *Proceedings of the 40th Annual Hawaii International Conference on Systems Sciences (HICSS’07)*, Hawaii, USA, 3–6 January.
- Alhoniemi E, Hollmén J, Simula O, Vesanto J. 1999. Process monitoring and modeling using the self-organizing map. *Integrated Computer Aided Engineering* **6**(1): 3–14.
- Back B, Sere K, Vanharanta H. 1998. Managing complexity in large data bases using self-organizing maps. *Accounting, Management and Information Technologies* **8**(4): 191–210.
- Barreto G, Araújo A. 2001. Time in self-organizing maps: an overview of models. *International Journal of Computer Research* **10**(2): 139–179.
- Berg A, Pattillo C. 1999. What caused the Asian crises: an early warning system approach. *Economic Notes* **28**: 285–334.
- Deakin EB. 1976. Distributions of financial accounting ratios: some empirical evidence. *The Accounting Review* **51**: 90–96.
- Deboeck G, Kohonen T. 1998. *Visual Explorations in Finance with Self-Organizing Maps*. Springer-Verlag: Berlin.

- Eklund T, Back B, Vanharanta H, Visa A. 2003. Using the self-organizing map as a visualization tool in financial benchmarking. *Information Visualization* **2**: 171–181.
- Eklund T, Back B, Vanharanta H, Visa A. 2008. A face validation of a SOM-based financial benchmarking model. *Journal of Emerging Technologies in Accounting* **5**(1): 109–127.
- Fuertes JJ, Domínguez M, Reguera P, Prada MA, Díaz I, Cuadrado AA. 2010. Visual dynamic model based on self-organizing maps for supervision and fault detection in industrial processes. *Engineering Applications of Artificial Intelligence* **23**(1): 8–17.
- Guimarães G, Lobo VS, Moura-Pires F. 2003. A taxonomy of self-organizing maps for temporal sequence processing. *Intelligent Data Analysis* **7**(4): 269–290.
- Hammer B, Micheli A, Neubauer N, Sperduti A, Strickert M. 2005. Self organizing maps for time series. In *Proceedings of the Workshop on Self-Organizing Maps*, Paris, France, 5–8 September.
- Hsu SH, Hsieh J, Chih TC, Hsu KC. 2009. A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications* **36**: 7947–7951.
- Kaminsky G, Lizondo S, Reinhart C. 1998. Leading indicators of currency crises. *IMF Staff Papers* **45**: 1–48.
- Kaski S, Kohonen T. 1996. Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world. In *Neural Networks in Financial Engineering*, Apostolos PNR, Abu-Mostafa Y, Moody J, Weigend A (eds). World Scientific: Singapore; 498–507.
- Kiang M, Hu M, Fisher D. 2006. An extended self-organizing map network for market segmentation - a telecommunication example. *Decision Support Systems* **42**(1): 36–47.
- Kindleberger C. 1996. *Maniacs, Panics, and Crashes*. Cambridge University Press: Cambridge.
- Kiviluoto K. 1998. Predicting bankruptcies with the self-organizing map. *Neurocomputing* **21**(1–3): 191–201.
- Kohonen T. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* **66**: 59–69.
- Kohonen T. 2001. *Self-Organizing Maps*, 3rd edn. Springer-Verlag: Berlin.
- Lev B. 1974. *Financial Statement Analysis: A New Approach*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Lingras P, Hogo M, Snorek M, West C. 2005. Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Information Sciences* **172**(1–2): 215–240.
- Luyssaert S, Sulkava M, Raitio H, Hollmén J. 2004. Evaluation of forest nutrition based on large-scale foliar surveys: are nutrition profiles the way of the future? *Journal of Environmental Monitoring* **6**(2): 160–167.
- Martin-del-Brío B, Serrano-Cinca C. 1993. Self-organizing neural networks for the analysis and representation of data: some financial cases. *Neural Computing and Applications* **1**(2): 193–206.
- Minsky H. 1982. *Can 'it' Happen Again? Essays on Instability and Finance*. M.E. Sharpe: Armonk, NY.
- Oja M, Kaski S, Kohonen S. 2002. Bibliography of self-organizing map (SOM) papers: 1998–2001. *Neural Computing Surveys* **3**: 1–156.
- Salmi T, Martikainen T. 1994. A review of the theoretical and empirical basis of financial ratio analysis. *The Finnish Journal of Business Economics* **4**: 426–448.
- Sarlin P. 2010. Visual monitoring of financial stability with a self-organizing neural network. In *Proceedings of the 10th IEEE International Conference on Intelligent Systems Design and Applications (ISDA'10)*, Cairo, Egypt, 28–30 November.
- Sarlin P. 2011. Sovereign debt monitor: a visual self-organizing maps approach. In *Proceedings of the IEEE Symposium on Computational Intelligence for Financial Engineering & Economics*, Paris, France, 11–15 April.
- Sarlin P, Eklund T. 2011a. Fuzzy clustering of the self-organizing map: some applications on financial time series. In *Advances in Self-Organizing Maps*, Laaksonen J, Honkela T (eds). Springer: Heidelberg; 40–50.
- Sarlin P, Eklund T. 2011b. Financial performance analysis of European banks using a fuzzified self-organizing map. In *Proceedings of the 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2011)*, Kaiserslautern, Germany, 12–14 September.
- Sarlin P, Marghescu D. 2011. Visual predictions of currency crises using self-organizing maps. *Intelligent Systems in Accounting, Finance and Management* **18**(1): 15–38.
- Sarlin P, Peltonen T. 2011. Mapping the state of financial stability. ECB Working Paper, no. 1382.
- Somervuo P. 2000. Competing hidden Markov models on the self-organizing map. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2000)*, Como, Italy, 27 July.
- Sulkava M, Hollmén J. 2003. Finding profiles of forest nutrition by clustering of the self-organizing map. In *Proceedings of the Workshop on Self-Organizing Maps*, Hibikino, Kitakyushu, Japan, 11–14 September.
- Ward J. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**: 236–244.

Research Paper 6

Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2013), Combining Visual Customer Segmentation and Response Modeling. Accepted for publication in *Neural Computing and Applications*.

Extends:

Yao, Z., Sarlin, P., Eklund, T., & Back, B. (2012), Combining Visual Customer Segmentation and Response Modeling. In *Proceedings of the 20th European Conference on Information Systems (ECIS 2012)*, Barcelona, Spain, 10-13 June, 2012, Paper 143.

Combining visual customer segmentation and response modeling

Zhiyuan Yao · Peter Sarlin · Tomas Eklund ·
Barbro Back

Received: 20 November 2012 / Accepted: 1 July 2013
© Springer-Verlag London 2013

Abstract Customer relationship management is a central part of Business Intelligence, and sales campaigns are often used for improving customer relationships. This paper uses advanced analytics to explore customer behavior during sales campaigns. We provide a visual, data-driven and efficient framework for customer-segmentation and campaign-response modeling. First, the customers are grouped by purchasing behavior characteristics using a self-organizing map. To this behavioral segmentation model, we link segment-migration patterns using feature plane representations. This enables visual monitoring of the customer base and tracking customer behavior before and during sales campaigns. In addition to the general segment-migration patterns, this method provides the capability to drill down into each segment to visually explore the dynamics. The framework is applied to a department store chain with more than 1 million customers.

Keywords Business Intelligence · Customer relationship management (CRM) · Visual analytics · Customer segmentation · Campaign-response modeling

1 Introduction

For a long time, the focus of modern companies has been shifting from being product-oriented to customer-centric [1]. In recent years, this change has been particularly rapid due to the increasing interest in Business Intelligence (BI) in general and customer relationship management (CRM) in particular. In industry and the CRM literature, it is commonly held that maintaining and improving existing customer relationships is more cost-effective than attracting new ones [2].

Sales campaigns, one of the most commonly used customer-facing activities, provide a good means for improving customer relationships. Although sales campaigns have been widely used to improve profitability, they have traditionally been more bottom-line focused than customer-focused [3]. One of the reasons for this is a lack of understanding of customer behavior and requirements. Previous research has shown the importance of using customer data for better understanding of customers' needs and behavior, especially in customer-facing activities [4, 5]. Increasingly affordable and powerful computing resources, such as data warehousing and cloud computing, are providing companies with ample opportunity to store data, to analyze data patterns and create knowledge for building better customer relationships [6, 7]. Customer segmentation is an effective approach for evaluating the value of customers and understanding their behavior. Customer segmentation divides the customer base into distinct and internally homogeneous groups. Effective segmentation enables companies to interact with customers in each segment collectively and allocate limited resources to various customer segments according to corporate strategies. A range of data mining techniques have been used for customer segmentation, e.g., decision trees [8],

Z. Yao (✉) · P. Sarlin · T. Eklund · B. Back
Department of Information Technologies, Åbo Akademi
University, Joukahaisenkatu 3–5, 20520 Turku, Finland
e-mail: zyao@abo.fi

P. Sarlin
e-mail: psarlin@abo.fi

T. Eklund
e-mail: toeklund@abo.fi

B. Back
e-mail: bback@abo.fi

self-organizing maps (SOMs) [9–11], *k*-means clustering [12, 13], and combinations of different methods [14, 15]. These studies have successfully demonstrated the usefulness of customer segmentation in a variety of industries. However, the solutions are often stand-alone analytical models, derived based on a specific time frame, and thereby often disregard connections with marketing campaigns. This static snapshot of the customer base might overlook possible real-time dynamics as is pointed out in [16]. For example, customers may exhibit different purchasing behavior before and during campaigns, and accordingly migrate among segments. In order to uncover the changing nature of customers, Ganti et al. [17] provided a framework for measuring changes in data. In addition, several methods [18–20] have been proposed to identify changes in association rules, i.e., the variations in the correlation of purchases of items over time. These methods store static views of the rule set over time, in order to draw conclusions regarding pattern evolution in the data. When applied to customer data, the changes in customer behavior can be summarized and quantified by a set of rules. However, rule sets provide fragmented information regarding customer behavior, and large rule sets are also difficult to manage. Moreover, the retrieved information regarding the changes in customer behavior is difficult to integrate with a customer segmentation model.

In order to provide a holistic view of multivariate temporal patterns in data, time-series clustering has been widely applied to search for similar shapes in trajectories [21]. In the context of CRM, D'Urso and Giovanni [22] proposed a number of dissimilarity measures for comparing time sequences and integrated the dissimilarity measures into clustering techniques to conduct market segmentation. The focus in their study was more on the time series itself, rather than on the entities. Denny [23–25] introduced a visualization approach for illustrating temporal changes in cluster structure. The authors used specific vector initialization to facilitate comparing different maps and proposed a relative density metric for visualizing the structural changes in clusters. The approach focuses more on comparing cluster structure rather than on inter-cluster migration patterns. Yao et al. [26] demonstrated the applicability of the self-organizing time map (SOTM) [27] for analyzing customer data by creating a visual temporal clustering model for exploring the customer base and tracking customer behavior in a department store over time. Although the study provides a holistic overview of temporal changes in customer behavior, it lacks the capability to show the customers' migration patterns among segments, which is common customer behavior, especially during special sales events. Lingras et al. [28] applied the SOM and its extensions to illustrate temporal variations in cluster membership and to identify potential customer attrition. Although

the stability of the clusters is illustrated by the changes in cluster membership, the profiles of the customers migrating among segments are disregarded. Ha et al. [29] studied the evolution of customer segments as the output of a SOM and a decision tree. The managerial implications of the changes in customer behavior were made based on the assumption that the segment transition probabilities are static overtime. This assumption, however, is often invalid in today's business environment. Denny et al. [30] introduced a framework for visual exploration of cluster migration at the entity level. The framework is capable of revealing the path of migration of entities and measuring attribute interestingness of migrating entities. The framework consists of two steps. First, the data at different points in time are clustered using two-dimensional SOMs, where the clusters identified are transformed and represented in a set of one-dimensional SOMs in order not to clutter the representation of migration patterns. Then, the migration patterns are represented using line thickness. However, the process of transforming two-dimensional SOMs into their corresponding one-dimensional counterparts is a challenging task.

In this study, the migration patterns and the static customer segmentation will be cohesively integrated to provide a single view regarding the profiles of customers migrating among segments during campaign periods. Previous research [31, 32] on response modeling focuses on estimation of response rate and campaign profitability, often ignoring the possible customer heterogeneity in terms of response patterns. Recently, more advanced classification techniques have been applied to CRM. Although these techniques have shown improved predictive performance [33], they lack the capability of providing a concise and informative view of customer behavior. In this paper, we apply a visual approach to model campaign response among customers in different segments. First, we create a demographics and purchasing behavior-based customer segmentation model, using a method for visual clustering, the self-organizing map. Then, following the approach in [34], we connect information about customers' responses to a number of sales campaigns based upon segment-migration probabilities, to (1) visualize which customer segments react to campaigns and (2) to identify differences in purchasing behavior during campaign/non-campaign periods. This model is expected to create a more effective analytical CRM system that is able to consolidate the patterns in customers' purchasing behavior and their underlying dynamics in one view and thus provide actionable marketing information.

The remainder of this paper is organized as follows. Section 2 introduces the methodology behind the framework. Section 3 introduces the framework for conducting visual customer segmentation and campaign-driven

segment–migration pattern analysis. Section 4 documents the application of the framework, the analysis of the results and discussion of managerial implications. In Sect. 5, conclusions are drawn by presenting our key findings.

2 Methodology

This section introduces the SOM and transition probabilities on the SOM, as well as describes the data used in this study.

2.1 Self-organizing maps

Clustering algorithms have been widely used to approach customer segmentation tasks. From the turn of the century, visualization techniques have gained in popularity for understanding and assessing clustering results [35]. Visual clustering, in particular, consists of techniques capable of simultaneously conducting clustering tasks and providing a multivariate visual display of the clustering results, thereby facilitating the exploration of useful patterns in the data.

The self-organizing map (SOM) is a well-known and widely used method for visual clustering. Unlike most traditional clustering algorithms that require post-processing for understanding cluster structures, the SOM is a unique technique for data and dimensionality reduction through its simultaneous clustering and projection capabilities. The SOM projects high-dimensional data onto a two-dimensional display, where similar input records are located close to each other. Conceptually, serial or parallel combinations of stand-alone clustering and projection methods come close to what the SOM performs. However, as proposed in [36, 37], motivations for using the SOM over alternative methods are the interaction between the two tasks of clustering and projection, the pre-defined grid structure for linking visualizations, flexibility for missing data and outliers, and computational efficiency. In addition, the SOM enables users to explore the cluster structure on the two-dimensional SOM grid. Therefore, the clustering results can be conveyed to the users in a format that is easy to understand and interpret. The SOM has previously been shown to be an efficient and easy-to-interpret tool for customer segmentation [9, 10, 28, 38, 39]. The functioning of the SOM can be split into two stages: (1) matching data records to their best-matching units and (2) updating each unit toward the attracted data, including those in adjacent locations. The matching is done as follows.

$$\|x_j - m_b\| = \min_i \|x_j - m_i\|, \quad (1)$$

where x_j (where $j = 1, 2, \dots, N$) represents input data vectors, m_i (where $i = 1, 2, \dots, M$) represents network's

reference vectors, and m_b is the best-matching unit. The batch SOM algorithm is used where all input data are presented to the map together and the map is updated at once. Accordingly, the reference vectors are updated as follows.

$$m_i(t+1) = \frac{\sum_{j=1}^N h_{ib(j)}(t)x_j}{\sum_{j=1}^N h_{ib(j)}(t)}, \quad (2)$$

where t is a discrete time coordinate and $h_{ib(j)}$ a decreasing function of neighborhood radii and time. The individual units of a SOM can be treated as separate clusters. When performing visualization, a map with a larger number of units is often preferred for increasing the projection granularity (i.e., detail). However, as the number of units in a SOM increases, the processing of the SOM resembles more data compression into a representative set of units than standard cluster analysis, and thus, ability to perceive the clusters in data is hindered. Therefore, it is reasonable to apply a second-level clustering to group the SOM units into clusters. In a two-level SOM, the dataset is first projected onto a two-dimensional display using the SOM, and the resulting SOM units are then clustered. The two-level SOM [40] has in previous studies shown to be effective; in particular, Li [41] shows the superiority of the combinatorial approach of the SOM and Ward's [42] hierarchical clustering over some classical clustering algorithms. Ward's clustering starts with each unit being treated as a separate cluster. Then, the two clusters (or units) with the minimum distance are merged in each step until there is only one cluster left on the map. In order to take into account the ordering of the SOM, Ward's clustering is limited to agglomerate only adjacent units. This modified Ward's [42] distance is used for merging two candidate clusters:

$$d_{kl} = \begin{cases} \frac{n_k n_l}{n_k + n_l} \cdot \|c_k - c_l\|^2 & \text{if } k \text{ and } l \text{ are adjacent} \\ \infty & \text{otherwise} \end{cases}, \quad (3)$$

where k and l represent clusters, n_k and n_l the cardinality of clusters k and l , and $\|c_k - c_l\|^2$ the squared Euclidean distance between the cluster centers of clusters k and l . In addition, we define the distance between non-adjacent clusters to be infinite. When clusters k and l are merged to cluster h , the cardinality is the sum of the cardinalities of k and l and the centroid the mean of c_k and c_l weighted by their cardinalities.

There are generally two types of cluster validity measures for evaluating the clustering results and determining the number of clusters: external and internal measures (for a discussion see, e.g., [43]). The external measures (e.g., Rand index [44] and Hubert's statistic [45]) evaluate a clustering solution with reference to some external apriori information, e.g., given class labels, while—the internal

measures evaluate (e.g., gap statistic [46], Dunn index [47] and Silhouette index [48]) a clustering solution in terms of the internal relationships among the data items. The gap statistic evaluates a clustering solution based upon the within-cluster dispersion; while the Dunn and the Silhouette indices take into account both cluster compactness and cluster separation. Since the Silhouette index is a widely used internal cluster validity measure, we decided to use it to determine the number of clusters K . The Dunn index provides qualitatively similar results.

For each observation i , its Silhouette coefficient is defined as:

$$S_i = \frac{b_i a_i}{\max(b_i, a_i)} \quad (4)$$

where a_i is the average distance between i and all other observations in the same cluster, and b_i is the average distance between i and the observations in its nearest cluster. The Silhouette index for a clustering solution is simply the average of the Silhouette coefficients of all observations. The higher the value, the better the observations are clustered.

The second-level clustering of the SOM is shown using contour lines, such as those shown in Fig. 4, where we have five segments denoted S_1 – S_5 . While the SOM represents a high-dimensional space on a two-dimensional output space, the multidimensionality can be described using feature planes [36]. A feature plane uses data for each individual variable of the SOM units and is visualized on the SOM grid. On each feature plane, the same position represents the same unit, but visualizes different features (i.e., variables). A set of feature planes, therefore, provide multiple linked views of the same data from a univariate perspective.

2.2 Transition probabilities on the SOM

While the SOM is an ideal tool for data and dimensionality reduction, identifying temporal movements in a SOM model is not a simple process (see [27] for a review of time in SOMs). Previously, trajectories [49] have been a common means to illustrate temporal movements of individual data records on the SOM grid [50, 51]. The use of trajectories suffers, however, from the deficiency that they can only be used on a limited set of data in order not to clutter the display and give no indications of overall patterns and their strengths. For modeling future cluster movement patterns, Afolabi and Olude [52] and Hsu et al. [53] proposed hybrid approaches combining SOMs with classification techniques like neural networks and support vector machines. While being appropriate for time-series prediction, these approaches suffer from high computational complexity as well as impaired visualization capabilities.

A more general probabilistic model of the temporal variations in a SOM model has been introduced through unit-to-unit transition probabilities in stand-alone SOMs [54, 55]. Recently, unit-to-cluster transition probabilities have been proposed to be computed on SOMs with a second-level clustering and visualized on their own SOM grids [34]. Probabilities are computed for movements on the two-dimensional SOM grid and can model transitions to a specified region, such as segments. In CRM terms, this would translate to segment-migration analysis as follows.

Given a SOM model, the location for each data record at each point in time is derived by assigning them to their best-matching unit. Then, we can summarize the segment migrations by computing probabilities of belonging to each segment in the following period, given their current characteristics. Relying on a first-order Markov chain assumption, we compute the probability of migration from unit i (where $i = 1, 2, \dots, M$) to segment s (where $s = 1, 2, \dots, S$) one period ahead using $p_{is}(t + 1)$:

$$p_{is}(t + 1) = \frac{n_{is}(t + 1)}{\sum_{s=1}^S n_{is}(t + 1)} \quad (5)$$

where n_{is} is the number of customers migrating from unit i to segment s and t is a time coordinate. That is, the migration probability from unit i to segment s equals the number of customers switching from unit i to segment s divided by the sum of customer movements from unit i to every other segment. The denominator guarantees that we fulfill the probabilistic constraint $\sum_{s=1}^S p_{is}(t + 1) = 1$. In a SOM model with four segments, this could in practice mean that for, say, unit i , the probability of being in segments $s = 1, 2, \dots, 4$ in period $t + 1$ could be 0.5, 0.2, 0.2 and 0.1, respectively.

As the migration probabilities are associated with each of the units of the SOM model, they can be linked to the SOM visualization. Migration probabilities for units can be visualized on SOM grids, where one unique point represents the same unit on the previously presented SOM. This shows the probability of migration to a particular segment for each unit on its own grid, such that the color code of each unit represents its probability to migrate to that particular segment. We call these visual representations of migration probabilities *migration planes*. To facilitate the interpretation of migration probability strengths to different segments, the color scales have been standardized for all migration planes. Thereby, the structure of the migrations between segments can be directly identified by studying these underlying migration planes.

2.3 Data

The data used in this study are from a national department store chain that belongs to a large, multiservice corporation. Through a loyalty card system, the corporation

provides customers with various discounts and rewards based on the loyalty points accumulated. Personal information about the cardholders is collected when they apply for the card, and their transactions are recorded in the system. The training dataset containing a total of 1,271,433 customers were obtained through the loyalty card system. It contains aggregated sales information from all branches of the department store chain, for the period of 2007–2009. Customers with spending amounts of less than EUR 50 in total from the department store chain during the 2-year period were excluded. The dataset consists of twenty variables that fall into three bases: *demographic variables*, *purchasing behavior variables*, and *product mix variables*.

The *demographic variables* show background data about the customers.

- Age in years
- Gender (1 for male, 2 for female).
- Estimated probability of children: Customers are divided into ten equal-sized groups (deciles) according to their probabilities of having children at home. The values range from 1 to 10. The higher the value of this variable, the more likely is the children living in the household.
- Estimated income level: The higher the value, the wealthier the household is considered to be. Possible values are 1, 2, and 3.
- Customer tenure: The number of years the customer has been a cardholder. Summary statistics of some demographic variables are listed in the “Appendix”.

The *purchasing behavior variables* are summarized from a daily database to quarterly aggregates per customer. A level of aggregation on a quarterly basis was assessed to avoid or mitigate problems related to irregular short-term

behavior and to cancel out fluctuations in behavior over time. Specifically, separate datasets are created to summarize customers’ daily purchasing behavior in each quarter of the 2-year period. These eight datasets are then appended. Hence, in the training data, each customer has eight corresponding records, i.e., one for each quarter during the period 2007–2009.

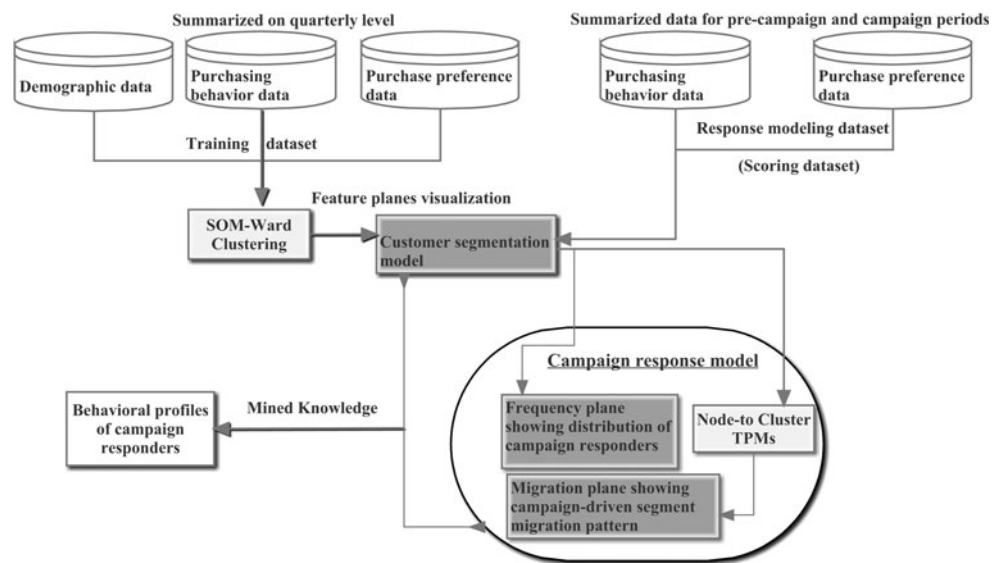
- Basket size: Average number of items bought per transaction.
- Average item value: Average value per item purchased.
- Average transaction value: Average value per purchase transaction.
- Working time purchase: The percentage of purchases made during Monday–Friday, 9 am–5 pm.
- Number of categories: Average total number of distinct product groups in each transaction.
- Spending amount: Average daily spending amount.
- Purchase frequency: Average number of transactions per day.

The *product mix variables* measure the percentage of the spending amount of each customer in each of the eight departments during each quarter, i.e., the mix of products that they tend to buy. This set of variables enables us to identify the quarterly purchasing preferences of each customer.

3 A framework for customer segmentation and response modeling

In this section, we discuss how we apply a visual combined customer-segmentation and response-modeling approach to analyze the loyalty card data. Figure 1 summarizes the

Fig. 1 The customer segmentation and campaign-response modeling approach used in this study



entire process. First, the training data are created by integrating all the customer information, i.e., *the demographic, purchasing behavior, and product mix variables*. Customers are divided into distinct segments using SOM-Ward clustering, and the revealed segments are profiled with the help of feature plane visualization. The response-modeling dataset is created by summarizing customer purchasing behavior and product mix patterns for the periods before and during the campaign. The response-modeling dataset is applied to the customer segmentation model, i.e., each of the data records of the response-modeling dataset is assigned to a best-matching unit on the customer segmentation model. The unit-to-segment migrations are computed and visualized using a series of migration planes. Finally, by combining the information from the customer segmentation model and campaign-response model, behavioral profiles of the campaign responders are created.

3.1 Performing customer segmentation

The Viscovery SOMine [56] package was used for creating the SOM model. Ward's hierarchical clustering was used to conduct second-level clustering on the SOM. All the variables included in the training process were pre-processed using a z -score transformation in order to normalize their weight in training and post-processed in order to have original values when interpreting models. However, for normalizing the influence of the *product mix variables*, their weights in training have been adjusted by dividing their values by the number of product categories. Since no demographic variables were used in building the customer segmentation model, their weights were set to zero. They hence have no influence on training, but enable interpretation of the demographics of segments that are based primarily upon purchasing behavior. In addition to the facilitating heuristics provided by SOMine, the SOM has been parametrized to create a model suitable for both data and dimensionality reduction. In particular, for more detailed visualization, the number of SOM units is chosen to be larger than the expected number of customer segments. The final customer segmentation model has 76 units, which is judged to provide enough detail with possibly low computational cost, as map sizes beyond this would increase cost on these high-volume data.

Then, second-level clustering is performed with Ward's hierarchical clustering, and a suitable number of customer segments is determined based upon the Silhouette index. We compare the Silhouette indices for clustering solutions with different K . For the unit-to-segment-migration patterns, it is necessary to have a K that is low enough to summarize the overwhelming amount of information but still high enough to reveal differences between customers. Therefore, we chose to

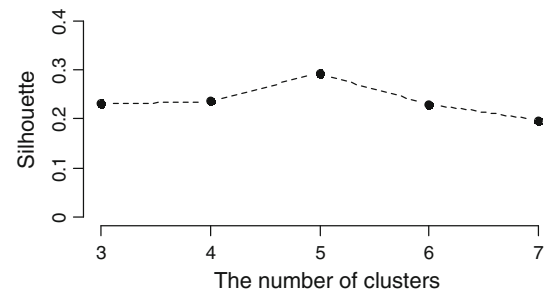


Fig. 2 Cluster validation of the second-level clustering of the SOM model

compare the Silhouette indices for $K = 3, 4, \dots, 7$. Figure 2 shows that the clustering solution with five clusters has the highest Silhouette index, and therefore, we chose to use a five-cluster solution. At the end of Sect. 4.1, we also explore the agglomeration process of the Ward's clustering, and further confirm the interpretability of the 5-cluster solution.

3.2 Performing response modeling

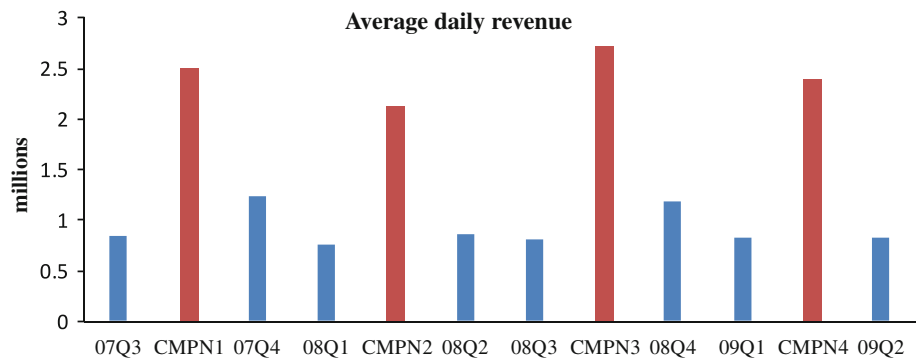
The sales campaign is a special event of the department store chain, organized twice a year and lasting approximately one week at a time. Figure 3 shows the average daily revenue during the four campaign periods (red) during the period 2007–2009, as well as the rest of the pre- and post-campaign periods (blue).

The response-modeling dataset contains the same variables as the ones used for training the SOM model, i.e., the *behavior and product mix variables* introduced in Sect. 2.3. We summarized each customer's *purchasing behavior and product mix variables* for each of the four campaign periods and their corresponding pre-campaign periods (i.e., 2007Q3, 2008Q1, 2008Q3 and 2009Q1). The behavioral variables were again summarized on the daily level in order to make them comparable between the pre-campaign and campaign periods. These data can be mapped to their best-matching units on the SOM model, and thus, migration patterns can be computed. For a customer that made no purchases during a period, we assign 0 for all the behavioral variables except for the variable working time transaction for which we assign a missing value.

4 Results and analysis

In this section, we present the results of the study. First, we present the customer segmentation model created according to the framework in Fig. 1. Then, we apply the response-modeling approach and associate the results with the segmentation model.

Fig. 3 The average daily revenue per quarter and campaign period



4.1 Visual customer segmentation

The resulting customer segmentation model consists of five segments (S_1-S_5). The feature planes (Figs. 4, 5, 6) show the distributions of each variable across the map, on which the gray scale visualizes the distribution of each variable over different segments, i.e., dark shades indicate high values and light shades low values. For example, Fig. 5f shows that high-spending customers can be found in S_3 , while the customers in S_1 include low spending customers. However, customers in S_1 tend to buy products from a number of different categories while customers in S_3 buy from fewer categories (cf. Fig. 5e). The key figures and important characteristics of each segment are summarized in Table 1. The table reveals that customers belonging to S_3 are the most valuable customers, whose spending contributed about 70 % of the total revenue. They are a group of primarily female customers who display high spending amounts, purchase more items, and shop frequently. Customers belonging to S_5 are a group of relatively old, high-income male customers, who purchase expensive products, but shop less frequently than average customers. The customers belonging to S_2 are likely pensioners, who have time to go shopping during working hours. Customers belonging to segments S_1 and S_4 are low-value customers.

While the optimal number of clusters $K = 5$ was determined based upon the Silhouette index, the agglomeration process provides a convenient tool for viewing the hierarchical structure of the segments in a top-down

manner. This also facilitates better understanding of the relative importance of the variables to the formation of the segments. The process started with a clustering solution where K equals the number of units M , and merges clusters until there is only one cluster ($K = 1$) left on the map. We summarize the process of the second-level clustering from low to high K for illustrative purposes.

Figure 7 illustrates how the segment structures evolve when increasing K from 3 to 7. The 3-segment solution consists of three segments which correspond to S_3 (red), S_5 (blue), and a large segment (yellow) combining S_1, S_2 and S_4 , respectively. This is a rough segmentation of customers into three groups representing a group of high-spending customers (cf. Fig. 5f), a group of customers with low spending amounts for single shopping visits and varying shopping frequency (cf. Fig. 5c, g), and another group of customers with high average transaction value but low revenue contribution (cf. Fig. 5c, f). In the 4-segment solution, S_1 separates from the largest segment in the 3-segment solution and becomes an independent segment (yellow). The greater variety of products in the shopping baskets of customers in S_1 (cf. Fig. 5e) makes them well-separated from those in S_2 and S_4 . In the 5-segment solution, segments S_2 (green) and S_4 (black) are split due to their differences in spending amounts and timing of visits to working or non-working hours (cf. Fig. 5d, f). The 6-segment solution divides S_5 into two sub-segments where the right one (purple) differs from the other (blue) by their larger spending in

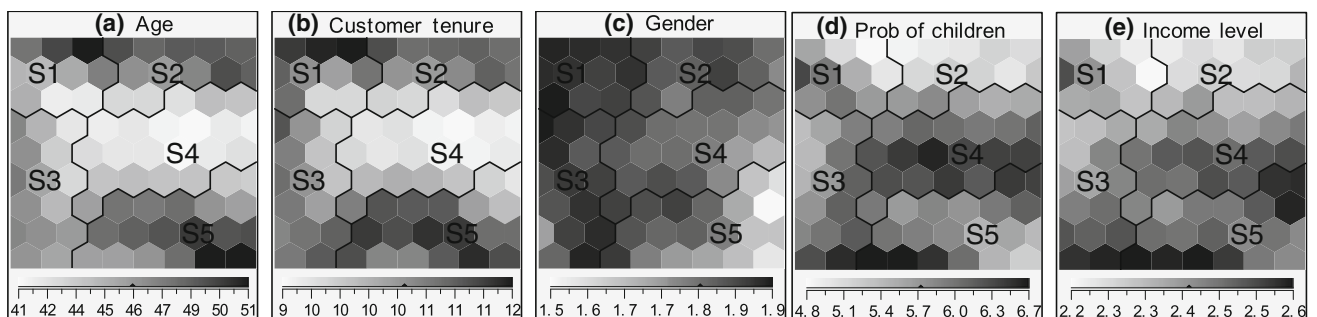


Fig. 4 Demographic profile

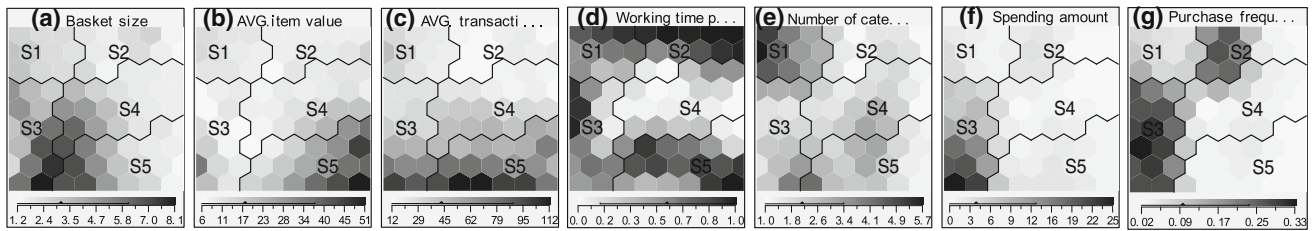


Fig. 5 Purchasing behavior profile

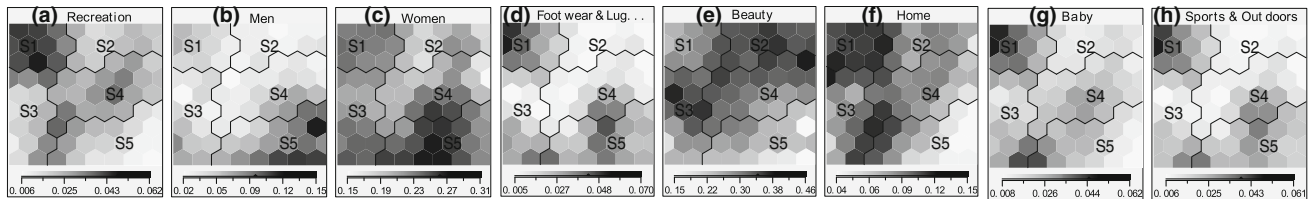


Fig. 6 Product mix profile

Table 1 Segment profiles of the customer segmentation model

Segment Size (%)	CTOR (%)	Behavioral profile	Product mix	Demographics
S ₁ 14	6.6	Purchase from different categories of products	Recreation, footwear and suitcase, home, children, and sports and outdoors	Female customers with a low estimated income level
S ₂ 22	7.8	Most transactions were made during working time	Beauty and home	Low probability of having children, low estimated income level, relatively old customers
S ₃ 16	69.7	High spending amount High purchasing frequency Large basket size	No special product preferences	Female customers with average demographic characteristics
S ₄ 21	4.0	Low-value customers	Recreation and beauty	High probability of having children, young and new customers
S ₅ 27	12.0	Purchase expensive items High-value transactions	Men, women, footwear and suitcase, and sports and outdoors	Relatively old, male customers with high estimated income level

Size refers to the percentage of all customers in the segment. CTOR refers to the contribution to the overall revenue, i.e., the proportion of revenue of the respective segment in the total revenue

expensive items in the male department (cf. Figs. 5b, 6b). Figure 4c also shows that the customers on the right side of the segment are primarily male. In the 7-segment solution, S₂ is split into two sub-segments (orange and green segments) and Fig. 5f shows that the two differs in terms of spending amounts.

4.2 Visual response modeling

The feature plane representation and the Table 1 provide an overview of the customer value, shopping behavior, product preferences and demographics for each segment. However, this segmentation model is based on customers' characteristics during normal periods, and provides little

information regarding their responsiveness to campaigns. In the following section, we will apply the framework introduced in Sect. 3, in order to link the information from the visual customer segmentation model and the campaign-response model.

The response modeling is performed using migration patterns on the SOM. The frequency plane in Fig. 8 was created by projecting the data concerning only the campaign periods (i.e., excluding the pre-campaign data from the response-modeling dataset) to the customer segmentation model. The color shades on the frequency plane represent the number of data records matching each unit. The darker the shade of the unit, the higher is the frequency of matches. Units with no matches are white. The frequency



Fig. 7 Segment membership planes of different segmentations consisting of 3–7 segments

plane provides an indication of the locations of the customers in the segmentation model during the campaigns. However, this gives little information regarding segment-migration patterns driven by the campaigns. To overcome that problem, we compute unit-to-segment patterns and use migration planes for linking the visualization of the migration patterns to the SOM segmentation. Figure 8a–e represents the per unit probability of moving to a particular segment.

The analysis of segment migration on the SOM and the frequency plane reveal several interesting patterns regarding the campaign-driven segment migration. S_5 (i.e., the male customers who under normal campaign periods purchase expensive products) is the segment of customers that is least likely to be activated by the campaign, as is indicated by the warm color of the nodes of S_5 on Fig. 8d. Moreover, almost all customers in S_1 and S_5 changed their purchasing behavior and moved to other segments during the campaign period, as is indicated by the white area in S_1 and S_5 in the frequency plane in Fig. 7. Customers in the left part of S_2 have high probabilities of staying in the same segment (cf. Fig. 8b), i.e., do not change their regular behavior during the campaign. These customers display rather stable purchasing patterns, with low total value but often recurring purchases, which continue even during the campaign. While the migrations on the SOM show the unit-to-segment patterns of the campaign effect, the frequency plane, as we introduced at the beginning of this section, pinpoints the location each customer resides in during the campaign periods. The frequency picture reveals three categories of responses: Area A (left part of S_2), B (S_3), and C (one single unit in S_4). Using the feature plane

visualization of the customer segmentation model and the results of the response model, Table 2 summarizes the three types of campaign responses.

Table 2 shows that customers already in or moving to Area B during the campaign account for 90.8 % of the campaign revenue. Customers belonging to S_3 have high probabilities of remaining in the segment during the campaign. An interesting pattern is that while there is migration from S_3 to other segments during a campaign, very little is to the no response Area C. This indicates that the customers in S_3 are loyal, high-value customers also during a campaign. In addition to these customers, many customers in S_1 also show high probabilities of moving to S_3 during the campaign. Figure 5e indicates that these customers are characterized by the diversity of their market baskets, which can be seen as a sign of customer loyalty and high switching costs (i.e., the costs that a customer bears as a result of becoming the competitor’s customer) [57]. At the same time, significantly changing purchasing patterns as indicated by the shift to S_3 (Area B) indicate a strong campaign response. Most of the customers in Area A are the ones belonging to that area of the segmentation model, i.e., a sub-section of S_2 . These customers would also appear to be quite “loyal”, as very few of them change their behavior during the campaign. This loyalty is, however, of questionable value as they exhibit very low spending amounts. The patterns provided by the model, while not being entirely unexpected, would provide the company with actionable information through a better understanding of profiles of the responders to this campaign, in particular, when combined with a customer-segmentation model.

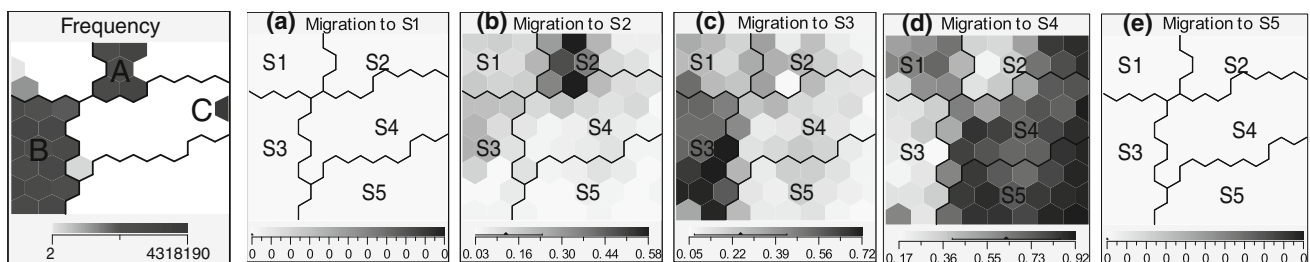


Fig. 8 The frequency plane and the migration plane visualization of the migration patterns and probabilities of migrations

Table 2 Customer profiles of the campaign-response types

Categories	CTSR (%)	Size (%) and sources (no.)	Behavioral profile	Product mix
<i>A Response</i>	9.2	5.5 Stay in: 1,411 Migrate to: 1,417 Migrate out: 282,659	Purchased small items during the campaigns	Beauty and home
<i>B Response</i>	90.8	9.5 Stay in: 6,709 Migrate to: 478,272 Migrate out: 5,994	Made several shopping visits and purchased expensive items during the campaigns	No particular preference of product
<i>C No response</i>	0	85 Stay in: 1,977,075 Migrate to: 2,340,516 Migrate out: 200,152	Made no purchases	None

CTSR refers to contribution to segment revenue and size to the percentage of all response data

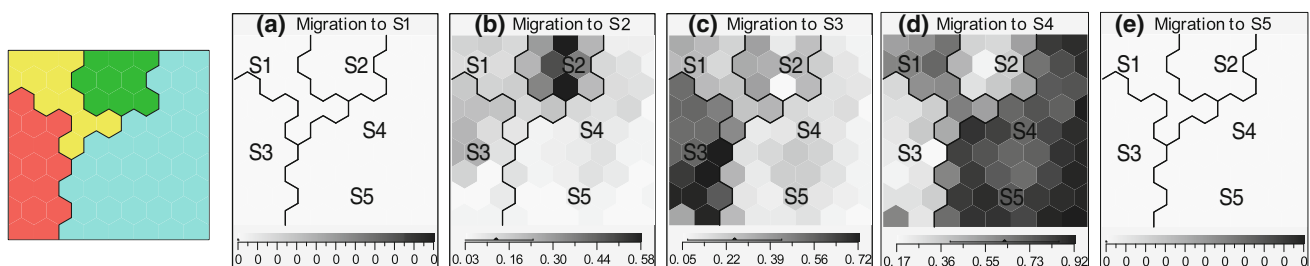


Fig. 9 A second-level clustering of the responder profiles and a series of migration planes showing unit-to-segment-migration probabilities with the second-level clustering applied

The above analysis of the campaign responder profiles is based upon first pinpointing the location each customer resides in during the campaign periods and then determining the types of responders by linking the frequency plane and migration planes to the segmentation model. To more objectively evaluate the responder profiles, we here employ Ward's clustering on the migration patterns. That is, a second-level clustering is applied on the migration probabilities of the SOM units, as is visualized in Fig. 8a–e. The segment membership plane in Fig. 9 indicates four profiles of campaign responders which are visualized on the migration planes (cf. Fig. 9a–e). The analysis based upon Fig. 9 is very similar to the previous result. The yellow segment in Fig. 9 is mainly composed of the units from S_1 in the segmentation model. Figure 9b–d shows that the customers in the yellow segment have high probabilities of moving to S_1 , S_2 and S_3 . The green segment, mainly composed of the units of Area A in the frequency plane of Fig. 8, has higher probabilities of

moving to S_2 (cf. Fig. 9b), or more precisely, to Area A during the campaign. The customers in the red segment, mainly composed of the units from S_3 , have high probabilities of moving to S_3 . The rest of the customers, i.e., the customers in the blue segment (i.e., low-value customers), have high probabilities of moving to S_4 (cf. Fig. 9d), or more precisely, to Area C during the campaign.

5 Conclusion

In this paper, we have developed a framework for detecting changes in customer behavior during a sales campaign. For this purpose, we first clustered the customers based on purchasing behavior characteristics. The revealed segments were profiled using feature plane visualization. We then computed the unit-to-segment migrations by applying the customer segmentation model to the response-modeling dataset. Finally, the segment-

migration patterns were visualized using feature plane representations.

We have demonstrated the usefulness of the framework on a case company's customer dataset, containing more than 1 million customers. In this concrete study, we have been able to show that campaigns have a positive effect on the average daily revenue. Moreover, our study has shown that customers react in different ways to a campaign. The campaign strongly affected the clusters that include the most valuable customers for the company, but in two distinct different ways. The cluster that includes mostly female customers that spend high amounts, have a high purchasing frequency and purchase many items at a time during normal circumstances, are highly activated during a campaign. The cluster that includes male customers with high-income levels and who usually buy expensive items avoid entering the store during a campaign, i.e., they change their behavior from being high-value customers to low-value customers during the campaign. This is an interesting phenomenon and should be investigated by campaign managers.

The result shows that the framework provides a holistic view of the patterns of customer purchase behavior and the underlying dynamics, and it enables the efficient identification of campaign-driven segment-migration patterns and within-segment heterogeneity in terms of campaign-response propensity. Additionally, the integration of customer segmentation, campaign-response and segment-migration modeling provides decision makers with an effective analytical CRM for better campaign management. There are generally two types of dynamics in a customer segmentation model: (1) the changes of segment membership of individual customers and (2) the changes of segment structures over time. This framework for visualizing customers' segment-migration patterns provides an effective means of illustrating the changes of segment membership of individual customers, assuming that the segment structure stay the same over time. In [26], the authors applied the self-organizing time map (SOTM) [27] for exploring the second type of dynamics, i.e., the changes of segment structure over time. Although the two types of dynamics have been addressed separately, the efforts in the future should be focused on dealing with them simultaneously. This would aid in the exploration of customer-migration patterns before, during, and after the campaigns in one view.

Acknowledgments The financial support of the Academy of Finland (Grant Nos. 127656 and 127592) and the Foundation of Economic Education is gratefully acknowledged.

Appendix

See Table 3.

Table 3 The summary statistics of the demographic variables

Variable	Mean	Std Dev	Minimum	Maximum
Age	45.90	15.84	15.00	90.00
Customer tenure	11.00	10.00	0.00	90.00
Prob of children	5.74	2.93	1.00	10.00
Income level	2.41	0.71	1.00	3.00

In addition to the variables, the demographic variables also include gender, which is a categorical variable consisting of two levels: 1 for male (19 %) and 2 for female (81 %)

References

1. Shah D, Rust RT, Parasuraman A, Staelin R, Day GS (2006) The path to customer centricity. *J Serv Res* 9(2):113–124
2. Berry LL (2002) Relationship marketing of services perspectives from 1983 and 2000. *J Relatsh Mark* 1(1):59–77
3. Day GS (1997) Aligning the organization to the market. In: Lehmann DR, Jocz KE (eds) *Reflections on the futures of marketing*. MA: Marketing Science Institute, Cambridge, pp 67–93
4. Jayachandran S, Sharma S, Kaufman P, Raman P (2005) The role of relational information processes and technology use in customer relationship management. *J Mark* 69(4):177–192
5. Kumar V, Venkatesan R, Reinartz W (2008) Performance implications of adopting a customer-focused sales campaign. *J Mark* 72(5):50–68
6. Rygielski C, Wang JC, Yen DC (2002) Data mining techniques for customer relationship management. *Technol Soc* 24(4): 483–502
7. Armbrust M, Fox A, Griffith R, Joseph A, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M (2010) A view of cloud computing. *Commun ACM* 53(4):50–58
8. Hwang H, Jung T, Suh E (2004) An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Syst Appl* 26(2): 181–188
9. Holmbom AH, Eklund T, Back B (2011) Customer portfolio analysis using the SOM. *Int J Bus Inf Syst* 8(4):396–412
10. Mo J, Kiang MY, Zou P, Li Y (2010) A two-stage clustering approach for multi-region segmentation. *Expert Syst Appl* 37(10):7120–7131
11. Yao Z, Holmbom A, Eklund T, Back B (2010) Combining unsupervised and supervised data mining techniques for conducting customer portfolio analysis. In: *Proceedings of the 10th industrial conference on data mining*, Springer, Germany, Berlin
12. Dennis C, Marsland D, Cockett T (2001) Data mining for shopping centres—customer knowledge-management framework. *J Knowl Manag* 5(4):368–374
13. Hosseini SMS, Maleki A, Gholamian MR (2010) Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Syst Appl* 37(7):5259–5264

14. Kuo R, Ho L, Hu C (2002) Integration of self-organizing feature map and K-means algorithm for market segmentation. *Comput Oper Res* 29(11):1475–1493
15. McCarty JA, Hastak M (2007) Segmentation approaches in data-mining: a comparison of RFM, CHAID, and logistic regression. *J Bus Res* 60(6):656–662
16. Grieser L, Wilde KD (2010) Adaptive right-time technologies in customer relationship management. *Bus Inf Syst Eng* 2(1):41–44
17. Ganti V, Gehrke J, Ramakrishnan R (1999) A framework for measuring changes in data characteristics. In: Proceedings of the eighteenth ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems, ACM, Philadelphia, United States
18. Chakrabarti S, Sarawagi S, Dom B (1998) Mining surprising patterns using temporal description length. In: Proceeding of the 24th international conference on very large data bases (VLDB '98), Morgan Kaufmann Publishers Inc., San Francisco, USA
19. Yairi T, Kato Y, Hori K (2001) Fault detection by mining association rules from house-keeping data. In: Proceeding of the international symposium on artificial intelligence, robotics and automation in space, Montreal, Canada
20. Keogh E, Lonardi S, Chiu BY (2002) Finding surprising patterns in a time series database in linear time and space. In: Proceeding of the eighth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, Edmonton, Canada
21. Liao TW (2005) Clustering of time series data—a survey. *Pattern Recogn* 38(11):1857–1874
22. D'Urso P, De Giovanni L (2008) Temporal self-organizing maps for telecommunications market segmentation. *Neurocomputing* 71(13):2880–2892
23. Denny, Squire DM (2005) Visualization of cluster changes by comparing self-organizing maps. In: Proceedings of the 9th Pacific-Asia conference on knowledge discovery and data mining (PAKDD), Springer, Hanoi, Vietnam
24. Denny, Williams GJ, Christen P (2008) ReDSOM: relative density visualization of temporal changes in cluster structures using self-organizing maps. In: Proceedings of the eighth IEEE international conference on data mining (ICDM), IEEE, Pisa, Italy
25. Denny, Williams GJ, Christen P (2010) Visualizing temporal cluster changes using relative density self-organizing maps. *Knowl Inf Syst* 25(2):281–302
26. Yao Z, Sarlin P, Eklund T, Back B (2012) Temporal customer segmentation using the self-organizing time map. In: Proceeding of the 16th international conference on information visualisation (IV), IEEE, Montpellier, France
27. Sarlin P (2013) Self-organizing time map: an abstraction of temporal multivariate patterns. *Neurocomputing* 99(1):496–508
28. Lingras P, Hogo M, Snorek M, West C (2005) Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Inf Sci* 172(1–2):215–240
29. Ha SH, Bae SM, Park SC (2002) Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case. *Comput Ind Eng* 43(4):801–820
30. Denny, Christen P, Williams GJ (2012) Analysis of cluster migrations using self-organizing maps. In: Cao L, Huang J, Bailey J, Koh Y, Luo J (eds) *New frontiers in applied data mining*, vol 7104. Springer, Berlin, pp 171–182
31. Basu AK, Basu A, Batra R (1995) Modeling the response pattern to direct marketing campaigns. *J Mark Res* 32(2):204–212
32. Piatetsky-Shapiro G, Masand B (1999) Estimating campaign benefits and modeling lift. In: Proceeding of the fifth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, San Diego, United States
33. Lessmann S, Voß S (2010) Customer-centric decision support. *Bus Inf Syst Eng* 2(2):79–93
34. Sarlin P, Yao Z, Eklund T (2012) A framework for state transitions on the self-organizing map: some temporal financial applications. *Intell Syst Account Financ Manag* 19(3):189–203
35. Ferreira de Oliveira MC, Levkowitz H (2003) From visual data exploration to visual data mining: a survey. *IEEE Trans Visual Comput Graphics* 9(3):378–394
36. Vesanto J (1999) SOM-based data visualization methods. *Intell Data Anal* 3(2):111–126
37. Sarlin P (1999) Data and dimension reduction for visual financial performance analysis. TUCS Technical Reports. TUCS
38. Lee SC, Suh YH, Kim JK, Lee KJ (2004) A cross-national market segmentation of online game industry using SOM. *Expert Syst Appl* 27(4):559–570
39. Vellido A, Lisboa P, Meehan K (1999) Segmentation of the online shopping market using neural networks. *Expert Syst Appl* 17(4):303–314
40. Vesanto J, Alhoniemi E (2000) Clustering of the self-organizing map. *IEEE Trans Neural Networks* 11(3):586–600
41. Li H (2005) Data visualization of asymmetric data using Sammon mapping and applications of self-organizing maps. University of Maryland, (College Park, Md.)
42. Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58(301):236–244
43. Theodoridis S, Koutroumbas K (2008) *Pattern recognition*. Elsevier Science, Amsterdam
44. Rand WM (1971) Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc* 66(336):846–850
45. Hubert L, Schultz J (1976) Quadratic assignment as a general data-analysis strategy. *Br J Math Stat Psychol* 29(2):190–241
46. Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc Ser B (Stat Methodol)* 63(2):411–423
47. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J Cybern* 3(3):32–57
48. Rousseeuw PJ, Kaufman L (1990) *Finding groups in data: an introduction to cluster analysis*. Wiley-Interscience, Hoboken
49. Kohonen T (1988) The 'neural' phonetic typewriter. *Computer* 21(3):11–22. doi:10.1109/2.28
50. Eklund T, Back B, Vanharanta H, Visa A (2003) Using the self-organizing map as a visualization tool in financial benchmarking. *Inf Vis* 2(3):171–181
51. Sarlin P, Marghescu D (2011) Visual predictions of currency crises using self-organizing maps. *Intell Syst Account Financ Manag* 18(1):15–38
52. Afolabi MO, Olude O (2007) Predicting stock prices using a hybrid Kohonen self organizing map (SOM). In: Proceedings of the 40th annual Hawaii international conference on system sciences, IEEE, Big Island, Hawaii, USA
53. Hsu S-H, Hsieh JP-A, Chih T-C, Hsu K-C (2009) A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Syst Appl* 36(4):7947–7951
54. Sulkava M, Hollmén J (2003) Finding profiles of forest nutrition by clustering of the self-organizing map. In: Proceedings of the workshop on self-organizing maps, Springer, Kitakyushu, Japan
55. Luyssaert S, Sulkava M, Raitio H, Hollmén J (2004) Evaluation of forest nutrition based on large-scale foliar surveys: are nutrition profiles the way of the future? *J Environ Monit* 6(2):160–167
56. SOMine (2009), *Viscovery SOMine user manual*, Austria, Vienna
57. Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. *J Mark* 67(1):77–99

Turku Centre for Computer Science

TUCS Dissertations

1. **Marjo Lipponen**, On Primitive Solutions of the Post Correspondence Problem
2. **Timo Käkölä**, Dual Information Systems in Hyperknowledge Organizations
3. **Ville Leppänen**, Studies on the Realization of PRAM
4. **Cunsheng Ding**, Cryptographic Counter Generators
5. **Sami Viitanen**, Some New Global Optimization Algorithms
6. **Tapio Salakoski**, Representative Classification of Protein Structures
7. **Thomas Långbacka**, An Interactive Environment Supporting the Development of Formally Correct Programs
8. **Thomas Finne**, A Decision Support System for Improving Information Security
9. **Valeria Mihalache**, Cooperation, Communication, Control. Investigations on Grammar Systems.
10. **Marina Waldén**, Formal Reasoning About Distributed Algorithms
11. **Tero Laihonen**, Estimates on the Covering Radius When the Dual Distance is Known
12. **Lucian Ilie**, Decision Problems on Orders of Words
13. **Jukkapekka Hekanaho**, An Evolutionary Approach to Concept Learning
14. **Jouni Järvinen**, Knowledge Representation and Rough Sets
15. **Tomi Pasanen**, In-Place Algorithms for Sorting Problems
16. **Mika Johnsson**, Operational and Tactical Level Optimization in Printed Circuit Board Assembly
17. **Mats Aspñäs**, Multiprocessor Architecture and Programming: The Hathi-2 System
18. **Anna Mikhajlova**, Ensuring Correctness of Object and Component Systems
19. **Vesa Torvinen**, Construction and Evaluation of the Labour Game Method
20. **Jorma Boberg**, Cluster Analysis. A Mathematical Approach with Applications to Protein Structures
21. **Leonid Mikhajlov**, Software Reuse Mechanisms and Techniques: Safety Versus Flexibility
22. **Timo Kaukoranta**, Iterative and Hierarchical Methods for Codebook Generation in Vector Quantization
23. **Gábor Magyar**, On Solution Approaches for Some Industrially Motivated Combinatorial Optimization Problems
24. **Linas Laibinis**, Mechanised Formal Reasoning About Modular Programs
25. **Shuhua Liu**, Improving Executive Support in Strategic Scanning with Software Agent Systems
26. **Jaakko Järvi**, New Techniques in Generic Programming – C++ is more Intentional than Intended
27. **Jan-Christian Lehtinen**, Reproducing Kernel Splines in the Analysis of Medical Data
28. **Martin Büchi**, Safe Language Mechanisms for Modularization and Concurrency
29. **Elena Troubitsyna**, Stepwise Development of Dependable Systems
30. **Janne Näppi**, Computer-Assisted Diagnosis of Breast Calcifications
31. **Jianming Liang**, Dynamic Chest Images Analysis
32. **Tiberiu Seceleanu**, Systematic Design of Synchronous Digital Circuits
33. **Tero Aittokallio**, Characterization and Modelling of the Cardiorespiratory System in Sleep-Disordered Breathing
34. **Ivan Porres**, Modeling and Analyzing Software Behavior in UML
35. **Mauno Rönkkö**, Stepwise Development of Hybrid Systems
36. **Jouni Smed**, Production Planning in Printed Circuit Board Assembly
37. **Vesa Halava**, The Post Correspondence Problem for Market Morphisms
38. **Ion Petre**, Commutation Problems on Sets of Words and Formal Power Series
39. **Vladimir Kvassov**, Information Technology and the Productivity of Managerial Work
40. **Frank Tétard**, Managers, Fragmentation of Working Time, and Information Systems

41. **Jan Manuch**, Defect Theorems and Infinite Words
42. **Kalle Ranto**, Z_4 -Goethals Codes, Decoding and Designs
43. **Arto Lepistö**, On Relations Between Local and Global Periodicity
44. **Mika Hirvensalo**, Studies on Boolean Functions Related to Quantum Computing
45. **Pentti Virtanen**, Measuring and Improving Component-Based Software Development
46. **Adekunle Okunoye**, Knowledge Management and Global Diversity – A Framework to Support Organisations in Developing Countries
47. **Antonina Kloptchenko**, Text Mining Based on the Prototype Matching Method
48. **Juha Kivijärvi**, Optimization Methods for Clustering
49. **Rimvydas Rukšėnas**, Formal Development of Concurrent Components
50. **Dirk Nowotka**, Periodicity and Unbordered Factors of Words
51. **Attila Gyenesei**, Discovering Frequent Fuzzy Patterns in Relations of Quantitative Attributes
52. **Petteri Kaitovaara**, Packaging of IT Services – Conceptual and Empirical Studies
53. **Petri Rosendahl**, Niho Type Cross-Correlation Functions and Related Equations
54. **Péter Majlender**, A Normative Approach to Possibility Theory and Soft Decision Support
55. **Seppo Virtanen**, A Framework for Rapid Design and Evaluation of Protocol Processors
56. **Tomas Eklund**, The Self-Organizing Map in Financial Benchmarking
57. **Mikael Collan**, Giga-Investments: Modelling the Valuation of Very Large Industrial Real Investments
58. **Dag Björklund**, A Kernel Language for Unified Code Synthesis
59. **Shengnan Han**, Understanding User Adoption of Mobile Technology: Focusing on Physicians in Finland
60. **Irina Georgescu**, Rational Choice and Revealed Preference: A Fuzzy Approach
61. **Ping Yan**, Limit Cycles for Generalized Liénard-Type and Lotka-Volterra Systems
62. **Joonas Lehtinen**, Coding of Wavelet-Transformed Images
63. **Tommi Meskanen**, On the NTRU Cryptosystem
64. **Saeed Salehi**, Varieties of Tree Languages
65. **Jukka Arvo**, Efficient Algorithms for Hardware-Accelerated Shadow Computation
66. **Mika Hirvikorpi**, On the Tactical Level Production Planning in Flexible Manufacturing Systems
67. **Adrian Costea**, Computational Intelligence Methods for Quantitative Data Mining
68. **Cristina Seceleanu**, A Methodology for Constructing Correct Reactive Systems
69. **Luigia Petre**, Modeling with Action Systems
70. **Lu Yan**, Systematic Design of Ubiquitous Systems
71. **Mehran Gomari**, On the Generalization Ability of Bayesian Neural Networks
72. **Ville Harkke**, Knowledge Freedom for Medical Professionals – An Evaluation Study of a Mobile Information System for Physicians in Finland
73. **Marius Cosmin Codrea**, Pattern Analysis of Chlorophyll Fluorescence Signals
74. **Aiying Rong**, Cogeneration Planning Under the Deregulated Power Market and Emissions Trading Scheme
75. **Chihab BenMoussa**, Supporting the Sales Force through Mobile Information and Communication Technologies: Focusing on the Pharmaceutical Sales Force
76. **Jussi Salmi**, Improving Data Analysis in Proteomics
77. **Orieta Celiku**, Mechanized Reasoning for Dually-Nondeterministic and Probabilistic Programs
78. **Kaj-Mikael Björk**, Supply Chain Efficiency with Some Forest Industry Improvements
79. **Viorel Preoteasa**, Program Variables – The Core of Mechanical Reasoning about Imperative Programs
80. **Jonne Poikonen**, Absolute Value Extraction and Order Statistic Filtering for a Mixed-Mode Array Image Processor
81. **Luka Milovanov**, Agile Software Development in an Academic Environment
82. **Francisco Augusto Alcaraz Garcia**, Real Options, Default Risk and Soft Applications
83. **Kai K. Kimppa**, Problems with the Justification of Intellectual Property Rights in Relation to Software and Other Digitally Distributable Media
84. **Dragoş Truşcan**, Model Driven Development of Programmable Architectures
85. **Eugen Czeizler**, The Inverse Neighborhood Problem and Applications of Welch Sets in Automata Theory

86. **Sanna Ranto**, Identifying and Locating-Dominating Codes in Binary Hamming Spaces
87. **Tuomas Hakkarainen**, On the Computation of the Class Numbers of Real Abelian Fields
88. **Elena Czeizler**, Intricacies of Word Equations
89. **Marcus Alanen**, A Metamodeling Framework for Software Engineering
90. **Filip Ginter**, Towards Information Extraction in the Biomedical Domain: Methods and Resources
91. **Jarkko Paavola**, Signature Ensembles and Receiver Structures for Oversaturated Synchronous DS-CDMA Systems
92. **Arho Virkki**, The Human Respiratory System: Modelling, Analysis and Control
93. **Olli Luoma**, Efficient Methods for Storing and Querying XML Data with Relational Databases
94. **Dubravka Ilić**, Formal Reasoning about Dependability in Model-Driven Development
95. **Kim Solin**, Abstract Algebra of Program Refinement
96. **Tomi Westerlund**, Time Aware Modelling and Analysis of Systems-on-Chip
97. **Kalle Saari**, On the Frequency and Periodicity of Infinite Words
98. **Tomi Kärki**, Similarity Relations on Words: Relational Codes and Periods
99. **Markus M. Mäkelä**, Essays on Software Product Development: A Strategic Management Viewpoint
100. **Roope Vehkalahti**, Class Field Theoretic Methods in the Design of Lattice Signal Constellations
101. **Anne-Maria Ernvall-Hytönen**, On Short Exponential Sums Involving Fourier Coefficients of Holomorphic Cusp Forms
102. **Chang Li**, Parallelism and Complexity in Gene Assembly
103. **Tapio Pahikkala**, New Kernel Functions and Learning Methods for Text and Data Mining
104. **Denis Shestakov**, Search Interfaces on the Web: Querying and Characterizing
105. **Sampo Pyysalo**, A Dependency Parsing Approach to Biomedical Text Mining
106. **Anna Sell**, Mobile Digital Calendars in Knowledge Work
107. **Dorina Marghescu**, Evaluating Multidimensional Visualization Techniques in Data Mining Tasks
108. **Tero Säntti**, A Co-Processor Approach for Efficient Java Execution in Embedded Systems
109. **Kari Salonen**, Setup Optimization in High-Mix Surface Mount PCB Assembly
110. **Pontus Boström**, Formal Design and Verification of Systems Using Domain-Specific Languages
111. **Camilla J. Hollanti**, Order-Theoretic Methods for Space-Time Coding: Symmetric and Asymmetric Designs
112. **Heidi Himmanen**, On Transmission System Design for Wireless Broadcasting
113. **Sébastien Lafond**, Simulation of Embedded Systems for Energy Consumption Estimation
114. **Evgeni Tsivtsivadze**, Learning Preferences with Kernel-Based Methods
115. **Petri Salmela**, On Commutation and Conjugacy of Rational Languages and the Fixed Point Method
116. **Siamak Taati**, Conservation Laws in Cellular Automata
117. **Vladimir Rogojin**, Gene Assembly in Stichotrichous Ciliates: Elementary Operations, Parallelism and Computation
118. **Alexey Dudkov**, Chip and Signature Interleaving in DS CDMA Systems
119. **Janne Savela**, Role of Selected Spectral Attributes in the Perception of Synthetic Vowels
120. **Kristian Nybom**, Low-Density Parity-Check Codes for Wireless Datacast Networks
121. **Johanna Tuominen**, Formal Power Analysis of Systems-on-Chip
122. **Teijo Lehtonen**, On Fault Tolerance Methods for Networks-on-Chip
123. **Eeva Suvitie**, On Inner Products Involving Holomorphic Cusp Forms and Maass Forms
124. **Linda Mannila**, Teaching Mathematics and Programming – New Approaches with Empirical Evaluation
125. **Hanna Suominen**, Machine Learning and Clinical Text: Supporting Health Information Flow
126. **Tuomo Saarni**, Segmental Durations of Speech
127. **Johannes Eriksson**, Tool-Supported Invariant-Based Programming

128. **Tero Jokela**, Design and Analysis of Forward Error Control Coding and Signaling for Guaranteeing QoS in Wireless Broadcast Systems
129. **Ville Lukkarila**, On Undecidable Dynamical Properties of Reversible One-Dimensional Cellular Automata
130. **Qaisar Ahmad Malik**, Combining Model-Based Testing and Stepwise Formal Development
131. **Mikko-Jussi Laakso**, Promoting Programming Learning: Engagement, Automatic Assessment with Immediate Feedback in Visualizations
132. **Riikka Vuokko**, A Practice Perspective on Organizational Implementation of Information Technology
133. **Jeanette Heidenberg**, Towards Increased Productivity and Quality in Software Development Using Agile, Lean and Collaborative Approaches
134. **Yong Liu**, Solving the Puzzle of Mobile Learning Adoption
135. **Stina Ojala**, Towards an Integrative Information Society: Studies on Individuality in Speech and Sign
136. **Matteo Brunelli**, Some Advances in Mathematical Models for Preference Relations
137. **Ville Junnila**, On Identifying and Locating-Dominating Codes
138. **Andrzej Mizera**, Methods for Construction and Analysis of Computational Models in Systems Biology. Applications to the Modelling of the Heat Shock Response and the Self-Assembly of Intermediate Filaments.
139. **Csaba Ráduly-Baka**, Algorithmic Solutions for Combinatorial Problems in Resource Management of Manufacturing Environments
140. **Jari Kyngäs**, Solving Challenging Real-World Scheduling Problems
141. **Arho Suominen**, Notes on Emerging Technologies
142. **József Mezei**, A Quantitative View on Fuzzy Numbers
143. **Marta Olszewska**, On the Impact of Rigorous Approaches on the Quality of Development
144. **Antti Airola**, Kernel-Based Ranking: Methods for Learning and Performance Estimation
145. **Aleksi Saarela**, Word Equations and Related Topics: Independence, Decidability and Characterizations
146. **Lasse Bergroth**, Kahden merkkijonon pisimmän yhteisen alijonon ongelma ja sen ratkaiseminen
147. **Thomas Canhao Xu**, Hardware/Software Co-Design for Multicore Architectures
148. **Tuomas Mäkilä**, Software Development Process Modeling – Developers Perspective to Contemporary Modeling Techniques
149. **Shahrokh Nikou**, Opening the Black-Box of IT Artifacts: Looking into Mobile Service Characteristics and Individual Perception
150. **Alessandro Buoni**, Fraud Detection in the Banking Sector: A Multi-Agent Approach
151. **Mats Neovius**, Trustworthy Context Dependency in Ubiquitous Systems
152. **Fredrik Degerlund**, Scheduling of Guarded Command Based Models
153. **Amir-Mohammad Rahmani-Sane**, Exploration and Design of Power-Efficient Networked Many-Core Systems
154. **Ville Rantala**, On Dynamic Monitoring Methods for Networks-on-Chip
155. **Mikko Pelto**, On Identifying and Locating-Dominating Codes in the Infinite King Grid
156. **Anton Tarasyuk**, Formal Development and Quantitative Verification of Dependable Systems
157. **Muhammad Mohsin Saleemi**, Towards Combining Interactive Mobile TV and Smart Spaces: Architectures, Tools and Application Development
158. **Tommi J. M. Lehtinen**, Numbers and Languages
159. **Peter Sarlin**, Mapping Financial Stability
160. **Alexander Wei Yin**, On Energy Efficient Computing Platforms
161. **Mikołaj Olszewski**, Scaling Up Stepwise Feature Introduction to Construction of Large Software Systems
162. **Maryam Kamali**, Reusable Formal Architectures for Networked Systems
163. **Zhiyuan Yao**, Visual Customer Segmentation and Behavior Analysis – A SOM-Based Approach

TURKU CENTRE *for* COMPUTER SCIENCE

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www.tucs.fi



University of Turku

Faculty of Mathematics and Natural Sciences

- Department of Information Technology
- Department of Mathematics and Statistics

Turku School of Economics

- Institute of Information Systems Science



Åbo Akademi University

Division for Natural Sciences and Technology

- Department of Information Technologies

ISBN 978-952-12-2950-3

ISSN 1239-1883

Zhiyuan Yao

Zhiyuan Yao

Visual Customer Segmentation and Behavior Analysis

Visual Customer Segmentation and Behavior Analysis – A SOM-Based Approach