Zhiyuan Yao | Peter Sarlin | Tomas Eklund | Barbro Back

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

Turku Centre for Computer Science

TUCS Technical Report
No 1085, June 2013

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

TUCS Technical Report
No 1085, June 2013

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.

Data Mining and Knowledge Management Laboratory

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

3.1Self-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 $t=\{1,2,\ldots,T\}$), 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 nodes $m_i(t)$ (where $i=1,2,\ldots,M$). 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:

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

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:

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

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:

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

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:

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

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 ($t=\{1,2,\ldots,T\}$), 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 $t=\{1,2,\ldots,T\}$) 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.

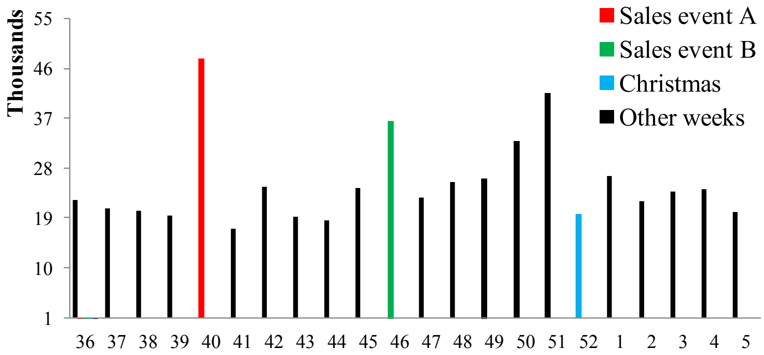


Fig. 1 The distribution of the number of customers across the 22 weeks.

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.

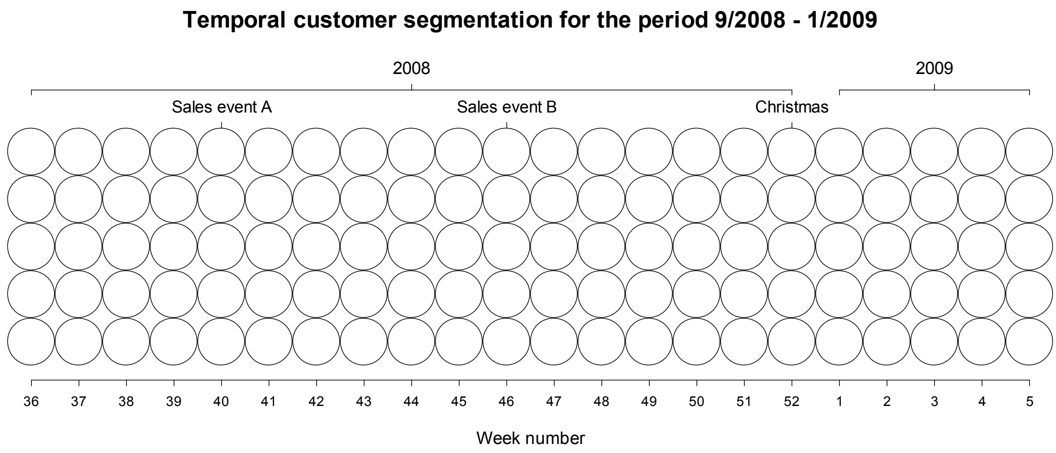
- Loyalty point level: Based on the average purchase amount per customer in the corporation, this variable divides customers into five classes: 0, 1, 2, 3, and 4. A higher value in loyalty point level indicates a larger historical spending amount in the entire corporate chain.
- Service level: This variable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

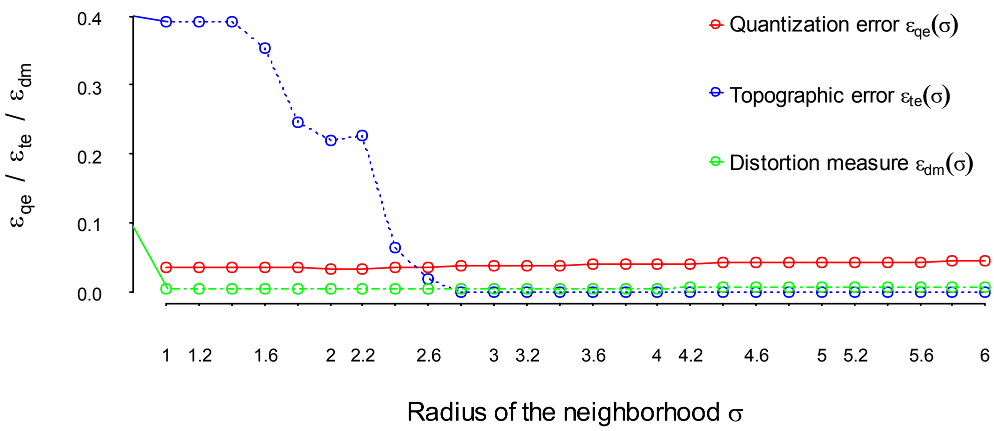


Fig. 3 Quality measures over radius of the neighborhood.

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.

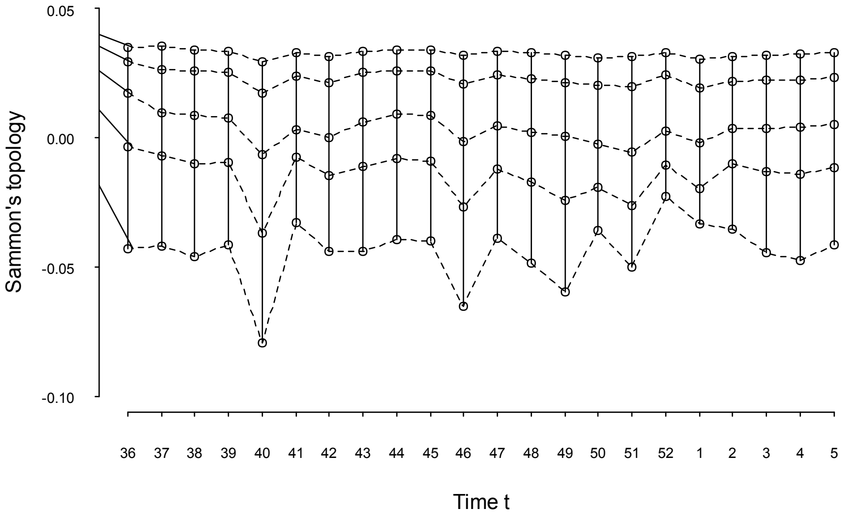
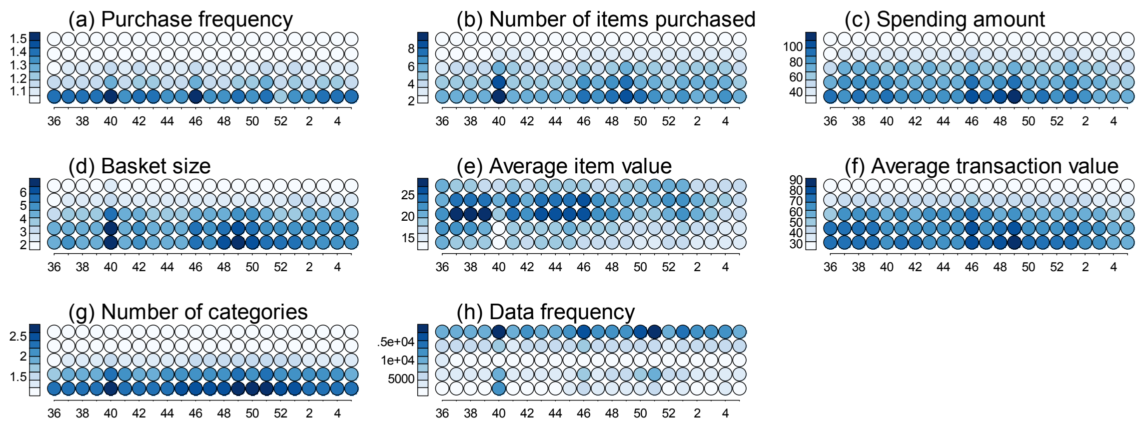


Fig. 4 A Sammon's mapping of the SOTM nodes.

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 1*st* and 2*nd* 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 3*rd* 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 4*th* and 5*th* 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)).

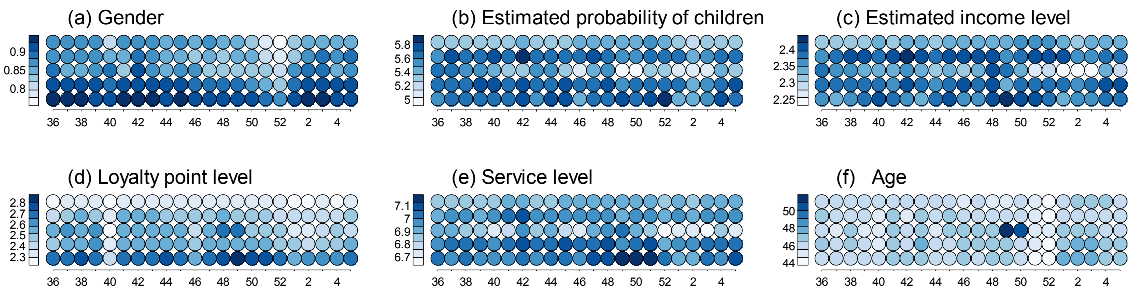
Since data evolve over time, their temporal changes can be assessed by comparing the nodes along the horizontal direction. Here, we summarize the changes driven by the sales events and Christmas period, respectively. All customer segments, in particular for the nodes on the first row which represents 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 $K = 3,4,\ldots,10$. 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.

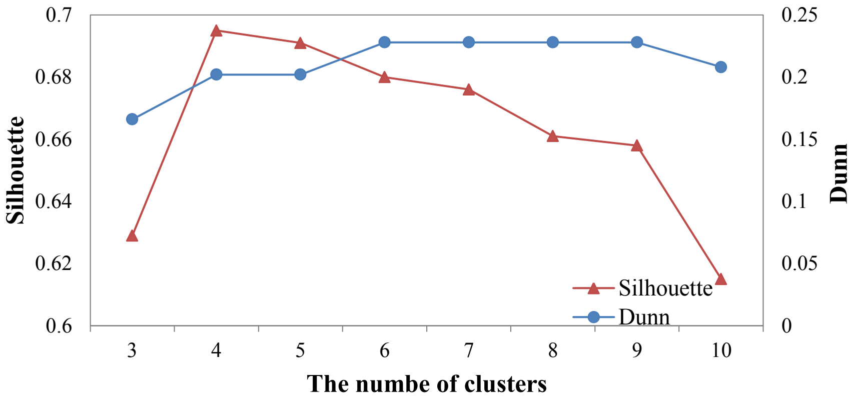


Fig. 7 Cluster validation of the second-level clustering of the SOTM.

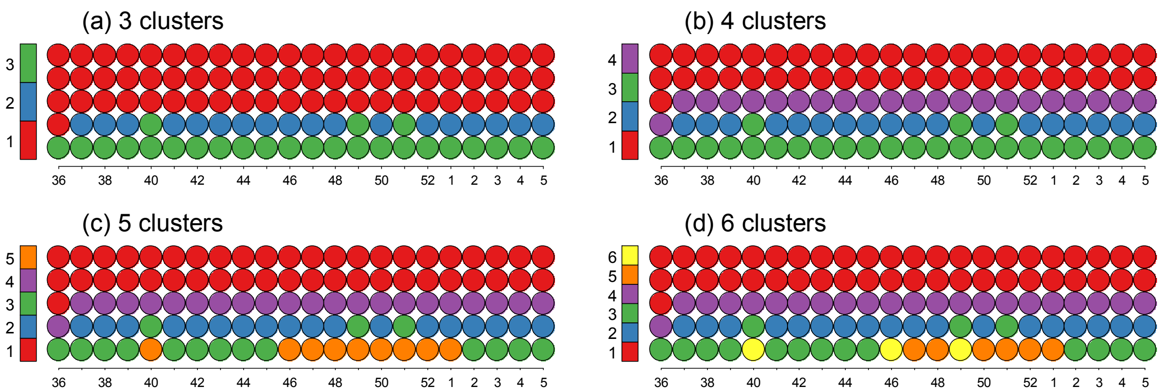


Fig. 8 Cluster membership planes of the SOTM.

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.

TURKU
CENTRE *for*
COMPUTER
SCIENCE

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Information Technologies

Turku School of Economics

- Institute of Information Systems Sciences

ISBN 978-952-12-2923-7
ISSN 1239-1891