Peter Sarlin | Zhiyuan Yao

Clustering of the Self-Organizing Time Map

Turku Centre *for* Computer Science

TUCS Technical Report
No 1062, November 2012

Clustering of the Self-Organizing Time Map

Peter Sarlin
Åbo Akademi University, Department of Information Technologies
Zhiyuan Yao
Åbo Akademi University, Department of Information Technologies

TUCS Technical Report
No 1062, November 2012

Abstract

This paper extends the use of the recently introduced Self-Organizing Time Map (SOTM) by pairing it with classical cluster analysis. The SOTM is an adaptation of the Self-Organizing Map for exploratory temporal structure analysis. While enabling visual dynamic clustering of temporal and cross-sectional patterns, the stand-alone SOTM lacks means for objectively representing temporal changes in cluster structures. This paper combines the SOTM with clustering, and illustrates the usefulness of second-level clustering for representing changes in cluster structures in an easily interpretable format. This provides means for identification of changing, emerging and lost clusters over time. Experiments are performed on two toy datasets and two real-world datasets. The first real-world application explores evolution dynamics of European banks before and during the global financial crisis. Not surprisingly, the results indicate a build-up of risks and vulnerabilities throughout the European banking sector prior to the start of the crisis. The second application identifies the cyclicality of currency crises through changes in the most vulnerable clusters.

Keywords: Self-Organizing Time Map; cluster analysis; visual dynamic clustering

TUCS Laboratory
Laboratory for Data Mining and Knowledge Management

1. Introduction

The Self-Organizing Time Map (SOTM) [1] is a recently introduced method for exploratory temporal structure analysis based upon Kohonen's Self-Organizing Map (SOM) [2]. The SOTM is a visual and dynamic approach to clustering by providing a two-dimensional view of (*i*) cross-sectional cluster structures, (*ii*) temporal trends, and (*iii*) multivariate data. This is a common setting in many domains, such as financial tasks with data for several entities, reference periods and explanatory variables. While clustering algorithms, in general, attempt to partition data into natural groups by maximizing inter-cluster distance and/or minimizing intra-cluster distance, the SOTM performs a clustering of a slightly different nature. The SOTM is essentially a series of one-dimensional SOMs on data ordered in consequent time units and represents both time and data topology on a two dimensional grid of prototype vectors. Hence, the topology preservation of the SOTM preserves data topology on the vertical direction and time topology on the horizontal direction. Since the introduction of the approach, the SOTM has been applied to dynamic customer segmentation by Yao *et al*. [3] and to providing an abstraction of the ongoing global financial crisis by Sarlin [4]. These case studies have, however, shown the inherent need for guidance in assessing cluster structures and their temporal evolvement. Previous works [1,3,4] have paired the SOTM with a number of visualization aids. The distance structure of a SOTM has been illustrated using U-matrix [5] visualizations, where a color code between all neighboring units indicates their average distance (see [1]). An adapted version of the cluster coloring by Kaski *et al*. [6] based upon Sammon's mapping [7] has been used to illustrate the distance structure using perceptual differences in colors (see [1,3,4]). The very standard feature planes have been used to represent the spread of values of individual inputs on the SOTM (see [1,3,4]). To quantify changes in cluster structures over time, a measure coined structural change has been introduced (see [1,3,4]). Yet, these approaches lack means for objectively representing temporal changes in the multivariate cluster structures.

With stationary data, rows of the SOTM would represent similar data at different points in time. One ambiguity with the SOTM, and its visualization aids, is, however, that the rows of units, when data are non-stationary, may represent data of different characteristics. Hence, while the stand-alone SOTM illustrates changes, it is difficult to judge whether or not units in a single row are members in a particular cluster and when changes in cluster structures are substantial. The previously provided visualizations of multivariate distances on the SOTM with cluster coloring and the U-matrix do not provide crisp, objective information for identification of changes in cluster structures. In the SOM literature, a common approach to grouping units of a SOM is so-called two-level clustering (e.g. [8,9]), where a SOM, agglomerative hierarchical clustering and partitioning clustering have been used to cluster the SOM units. Similarly, the SOTM could not only be utilized as a stand-alone dynamic clustering technique, but its output could also be used as input for a second stage of two-level clustering. The second-level clustering of the SOTM, while being conceptually similar, differs from two-level SOMs by clustering both units representing cross-sectional data, as well as units representing distinct points in time. Thus, a second-level clustering of the SOTM should be treated as an alternative for a rough labeling of units and for identification of *changing*, *emerging* and *lost* clusters over time.

The novelty of this paper is to provide means to visualize evolving cluster structures. We use four experiments to demonstrate the usefulness of combining the SOTM with a second-level clustering. Due to the lack of a comparable evaluation function, the paper is short of a quantitative evaluation. Instead, we illustrate the functioning, output and usefulness of combining the SOTM with a second-level clustering on an artificial toy dataset with expected patterns. First, we exemplify how stationary data may be labeled by the rows of the SOTM, and confirm it with a second-level clustering. Second, we illustrate how non-stationary data can be processed and show how *changing*, *emerging* and *disappearing* clusters can be identified. The main rationale for the artificial examples is to illustrate how evolving data can be represented with a second-level clustering of the SOTM. Also, this illustrates the interpretation of the SOTM and the second-level clusters, as well as defines the types of dynamics in cluster structures.

The SOTM with a second-level clustering is also applied in two real-world settings: gradual changes of European banks and cyclical behavior of currency-crisis indicators. First, we illustrate how the SOTM

can be used for visual dynamic clustering of financial ratios of European banks, and use a second-level clustering to highlight gradual *changes* in cluster structures over time. While mostly illustrating changes in cluster structures, we can also observe the *emergence* of new clusters and the *disappearance* of old. Second, we use the SOTM to assess cyclical changes in cluster structures of currency crisis indicators. A center of interest is on illustrating the temporal *emergence* of clusters, or *changes* in their size, where we mainly assess states vulnerable to a crisis. This follows from the common notion of reoccurring macro-financial vulnerabilities prior to crises that are then triggered by a wide range of unforeseeable events (see e.g. the work by Reinhart and Rogoff [10,11]).

This paper is structured as follows. Section 2 introduces the SOTM and clustering methods, as well as techniques for evaluating them. In Section 3, we illustrate the rationale behind clustering of the SOTM with two simple toy examples. Section 4 applies the SOTM in two real-world settings by assessing changes in cluster structures in the European banks and indicators of currency crises. Finally, Section 5 concludes by presenting key findings.

2. Methodology

This section presents the SOTM and stand-alone clustering methods, as well as techniques for quantitative evaluation of clustering results.

2.1 Self-Organizing Time Map

The SOM paradigm [2] has its basis in clustering via vector quantization and projection via neighborhood preservation, but provides only a static snapshot of data structures. To introduce time or dynamics, the SOM literature has provided a wide range of adaptations. The most common treatment of time is to introduce it implicitly in post-processing (e.g., trajectories [12]). Another avenue for temporal processing is to modify the activation and learning rule (e.g., Hypermap [13]) and network topology (e.g., Temporal SOM [14]). Combinatorial approaches of standard SOMs and various visualization techniques have been implemented with the aim of better spatiotemporal visualization (e.g., Guo *et al*. [15]). Further, dissimilarity measures for comparing time sequences have been integrated into the SOM paradigm for clustering temporal data [16].

The above extensions, while dealing with temporal transitions and dependencies and mainly focusing on time-series prediction, are not directly applicable for exploratory temporal structure analysis. Assessing changes in cluster characteristics is most commonly performed using the above mentioned post-processing. For instance, Sarlin and Marghescu [17] use a data set with a temporal and cross sectional dimension and explore dynamics with trajectories. Along the same lines, Lingras *et al*. [18] study changes in cluster characteristics with cluster memberships in a standard SOM and a SOM based upon the properties of rough set theory. Until recently, comparing standard two-dimensional SOMs for each time unit, as proposed in [19-21], has been the best available method for exploratory temporal structure analysis. Denny *et al*. **[**20,21**]** enhance temporal interpretability by applying specific initializations and visualizations. Nevertheless, the method has the drawback of an unstable orientation over time and complex comparisons of two-dimensional grids. Visualizing temporal changes in data structures in a manner that disentangles time from cross-sectional structures has thus to be dealt with differently. This particular issue is addressed by the recently introduced SOTM [1].

The SOTM uses the capabilities of the classical SOM for abstraction of changes in data structures over time. To observe the cross-sectional structures of the dataset for each time unit t (where t=1,2,…,T), the SOTM uses one-dimensional SOMs per time unit t. More precisely, the SOTM maps from the input space $x_j(t) \in \Omega(t)$ (where j=1,2,…,$N(t)$), 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,…,M). For t=1 the first principal component of principal component analysis (PCA) is used for initializing $A(t)$. The SOTM uses short-term memory to preserve the orientation between consecutive arrays $A(t)$. The short-term memory refers to reference vectors of $A(t)$ being initialized by those of $A(t$-1). Adjustment to temporal changes is achieved by performing a batch update per time t. The arrays $A(t)$ are then turned into a timeline by arranging them in an ascending order of time t. Hence, one can observe twofold topology preservation:

the vertical direction preserves data topology and the horizontal preserves time topology. 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 (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. The functioning of the SOTM can be summarized as follows.

t = 1
initialize *A*(*t*) using PCA on *Ω*(*t*)
apply the batch update to *A*(*t*) using *Ω*(*t*)
while *t*<*T*
  t = *t* + 1
  initialize *A*(*t*) using $m_i(t)$ of *A*(*t*-1)
  apply the batch update to *A*(*t*) using *Ω*(*t*)
end
order *A*(*t*) in an ascending order of time *t*.

The output of a SOTM is a two-dimensional grid that represents the evolution of a high-dimensional space over time. In this paper, we represent the multidimensionality of the SOTM using feature planes, as is common in the SOM literature. They show the spread of values for each variable on the SOTM grid and enable assessing the temporal evolution of cross-sectional distributions, as the timeline below the figure represents the time dimension in data (see e.g. Fig. 2). The feature planes are produced using a ColorBrewer's [22] scale, where variation of a blue hue occurs in luminance and light to dark represent low to high values. On each feature plane, the differences of the units along the vertical direction shows differences in cross-sections, while differences along the horizontal direction shows differences over time. This allows a twofold interpretation of patterns: *i*) cross-sectional relations and distributions and *ii*) changes in cross-sectional patterns over time.

Characteristics of a SOTM can be quantified by a number of time-restricted quality measures from the SOM paradigm. We focus in this paper on quantization error $\varepsilon_{qe}(t)$ distortion measure $\varepsilon_{dm}(t)$, and topographic error $\varepsilon_{te}(t)$ (for details see [1]):

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

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

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

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 units.

2.2 Cluster analysis

The SOTM, which is essentially a series of one-dimensional SOMs on data ordered in consequent time units, represents time and data topology on a two dimensional grid. While visualization aids may facilitate in interpreting structures of the SOTM, e.g. feature 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. Thus, we turn to a second-level clustering for identifying *changing*, *emerging* and *disappearing* clusters on the SOTM. More precisely, the output units produced by the

SOTM are used as an input for classical cluster analysis. Then, on the SOTM, we illustrate clusters with a qualitative color scheme from ColorBrewer [22], where clusters are differentiated in hue contrast with nearly constant saturation and lightness.

2.2.1 Clustering techniques

Clustering is a class of techniques that partition data into groups attempting to minimize intra-cluster distance and maximize inter-cluster distance. The existing clustering algorithms can be generally classified into the following categories: partitioning clustering, hierarchical clustering, density clustering, grid clustering and model-based clustering. 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. In this study, agglomerative hierarchical clustering will be used to group the SOTM units. The key motivation for using hierarchical clustering, rather than for example *k*-means clustering, is that this enables us to circumvent a definite choice of number of clusters *K*. Instead it provides us the means to explore the clustering results with varying *K*.

Agglomerative hierarchical clustering starts by treating each unit of the SOTM as a separate cluster ($K=M*T$) and merges iteratively clusters with the shortest distance until all units 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. In single-linkage, the distance between two clusters is deternined by the distance of the closest pair of observations in the two clusters. 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. on the contrary, Ward's method attempts to minimize the variance of the merged cluster. As hierarchical clustering methods are well-known, readers interested in details are referred to [23] 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 has been shown to oftentimes be inefficient in separating data [24,25]. Ward's method has already in the seminal literature shown to be accurate by outperforming other hierarchical methods [25-27]. We experimented with single, complete and average-linkage and Ward's hierarchical clustering, but found the results to be qualitatively similar. Hence, supported by the previous literature, we report in this paper only the results with the Ward's method. We use the following Ward's [28] criterion for merging clusters:

$$d_{kl} = \frac{n_k n_l}{n_k + n_l} \cdot \left\| c_k - c_l \right\|^2,$$

where k and l represent two clusters, n_k and n_l the number of reference vectors m_i in the clusters k and l, and $\left\| c_k - c_l \right\|^2$ the squared Euclidean distance between the cluster centers of clusters k and l.

This divides the SOTM units into a number of clusters, but we need precise definitions of the changes to properly assess how cluster structures evolve. Hence, we define the three types of dynamics in cluster structures as follows: *i*) a cluster disappears when one or more units are a member of it in time t and none is in $t+1$, *ii*) a cluster emerges when no unit is a member of it in time t and one or more are in $t+1$, and *iii*) a cluster changes when the (positive) number of units being a member of it in time t and $t+1$ differ.

2.2.2 Validation measures

The clustering validity is assessed by the Dunn index [29] and the Silhouette coefficient [25], both of which take into account cluster compactness and cluster separation [30]. 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:

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

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. Some experiments on toy data

This section introduces second-level clustering of the SOTM with experiments on an artificial toy dataset. The experiments on data with expected patterns illustrate the usefulness of combining clustering techniques with the SOTM.

3.1 Toy data

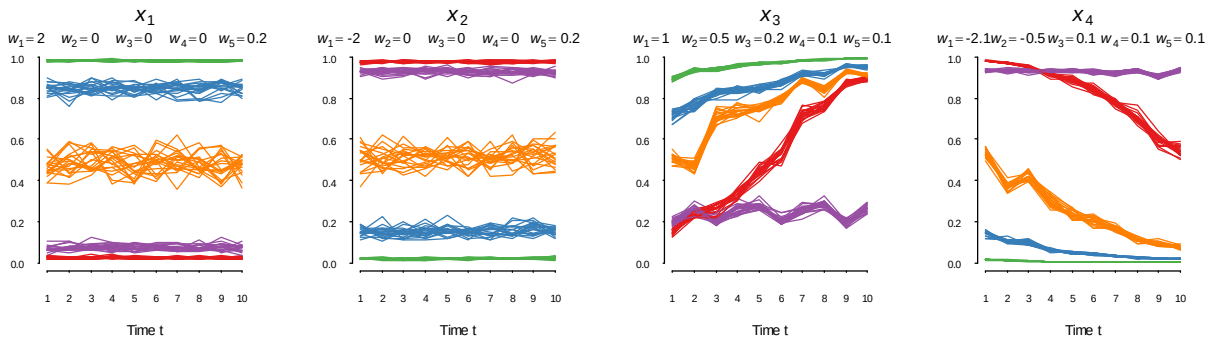
We generate a toy dataset that represents a three-dimensional cube with one dimension for time, one for cross-sectional entities, and one for input variables. We follow the procedure used in [1], but generate data for the purposes of this paper. The expected patterns in data are adjusted by five weights w_{1-5} that steer randomized shocks on four levels: group-specific (g), time-specific (t), variable-specific (r) and common (j) properties. Group-level differences represent clusters in data, time-level properties temporal trends and group-specific and common shocks general noise. The data x are generated as follows:

$$x(r,g,j,t) = \frac{1}{1+\exp\left(-\left[E(r,g,t) + w_4(r,g)e_4(r,t) + w_5(r,g)e_5(r,j,t)\right]\right)}$$

and group-specific trends by

$$E(r,g,t) = w_1(r)e_1(g) + w_2(r)e_2(g)t + w_3(r)e_3(g,t)$$

where $e_{1,3-5} \sim N(0,1)$ are drawn from a normal distribution and $e_2 \sim U(0,1)$ from a uniform distribution, and weights w_{1-5} are specified by the user for adjusting properties in data. The properties in data are as follows: group-specific intercepts (w_1), group-specific slopes over time (w_2), magnitude of group-specific random shocks (w_3), magnitude of time-specific common shocks (w_4) and magnitude of general common shocks (w_5). Fig. 1 reports the used weights w_{1-5} for generating the four variables with five groups of 20 cross-sectional entities over 10 periods, where the color coding illustrates the groups. Particular characteristics of the below four variables are obviously that the time series of x_{1-2} are stationary and those of x_{3-4} are non-stationary. While we may not fulfill all conditions of stationarity, e.g. constant variance and autocorrelation, the aim of these data is to have two variables with a somewhat constant data structure and two with a time-varying and converging structure. That is, the time-varying data are generated such that parts of the groups converge whereas others diverge. Evident changes of x_{3-4} in Fig. 1 are, for instance, that the red group diverges from the purple group and the blue converges to the green group.



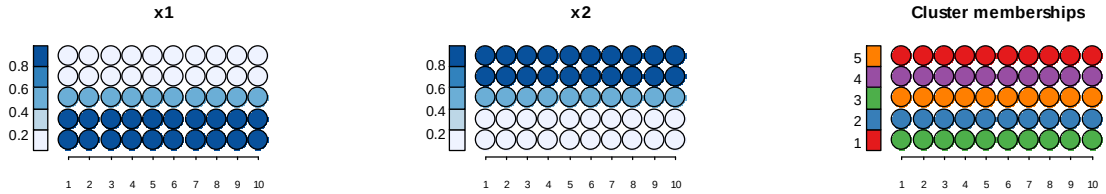
Notes: The figure reports the used weights $w_{1\text{-}5}$ for generating the $x_{1\text{-}4}$ with 100 entities over 10 periods, where the color coding illustrates the 5 groups, *x* axis represents time *t* and *y* axis values.

Fig. 1. Generated artificial toy dataset

3.2 Toy examples with the SOTM

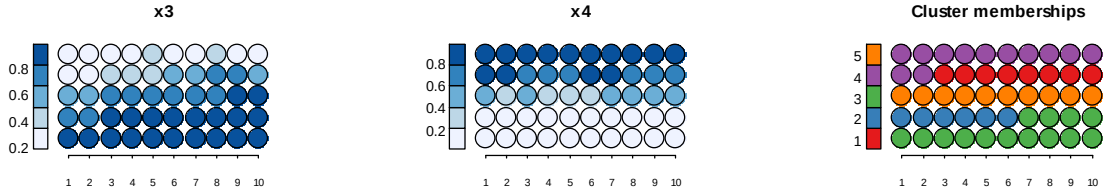
This section presents two toy examples of the SOTM: one with stationary and one with non-stationary data. The key focus is on illustrating the usefulness of second-level clustering of the SOTM. We use first the stationary data x_{1-2} as inputs for a SOTM and then the non-stationary data x_{3-4}. We specify the SOTM to have 5x10 units, where 5 vertical units represent data topology and 10 horizontal units time topology. The number of units (or clusters) on the vertical axis is set to equal the number of generated groups in data, while the number of units on the horizontal axis is fixed by the number of time units *T*. When choosing the final specification of the SOTM, quality measures introduced in Section 2.1 are used. As the main focus ought to be on topographic accuracy, we choose a neighborhood σ that leads to minimum quantization error given no topographic errors. Topographic error is stressed as the interpretation of a SOTM relies heavily on topology preservation, not least the time dimension.

To the trained SOTM, we apply a second-level Ward's clustering to the units of the SOTM. As the number of generated groups is pre-defined, we do not compare the performance of different *K* using clustering validation measures. On the stationary variables x_{1-2}, we obviously set *K* equal to the number of created groups. In the non-stationary case, we use the same *K* to better illustrate the difference to the stationary case. However, exploring different *K* would be useful for illustrating properties of the data. Fig. 2 shows feature planes and cluster memberships for a SOTM on the stationary data x_{1-2}. It is worth noting that the coloring of the clusters follows the coloring of the groups in Fig. 1. Indeed, one can observe that both the two feature planes and the cluster memberships are constant over time. This illustrates that stationary data may be labeled by the rows of the SOTM. Fig. 3, on the other hand, shows feature planes for a SOTM on the non-stationary data x_{3-4}. The feature planes clearly depict the increasing and decreasing trends as well as differences in densities in the structure. This is also reflected in the cluster memberships. The feature of approximating the probability density functions of data $p(x,t)$ lead to a direct interpretation of memberships; the denser some part of the data space $\Omega(t)$, the higher is the number of units in that location. The convergence of the green, blue and orange groups and divergence of the red from the purple group (as shown in Fig. 1) are shown as increases and decreases of vertical units in a second-level cluster. That is, in Fig. 3, we can observe that cluster 2 (blue units) *disappears* in period 7 and that cluster 1 (red units) *emerges* from cluster 4 (purple units) in period 3, as well as cluster 3 (green units) and 4 (purple units) *change* from one vertical unit to two and from two units to one, respectively.



Notes: The two first grids are feature planes and represent the spread of values of $x_{1\text{-}2}$. The final grid shows cluster memberships using color coding. Vertical color scales on the left of feature planes link a cluster number to each color.

Fig. 2. A SOTM applied to toy data with no temporal variation.



Notes: See notes to Fig. 2.

Fig. 3. A SOTM applied to toy data with temporal variation.

4. Applications in real-world settings

This section presents two applications of the SOTM, as well as its second-level clustering, in financial settings. The first application focuses on gradual changes in clusters of European banks and the second on the cyclical changes in indicators of currency crises.

4.1 An application to European banks

We present in this section a full application of the SOTM, as well as its second-level clustering, to a dataset of financial ratios of European banks. The dataset has previously been used in [31,32], in particular for financial benchmarking of high-dimensional data on a low-dimensional display of the SOM. These models do, however, lack means for assessing changes in cluster structures. The dataset consists of a large number of financial ratios of European banks measuring asset and capital quality, as well as operations and liquidity ratios. The data have been retrieved from Bankscope of Bureau van Dijk. The sample comes from all 27 EU countries, including in total 855 banks spanning from 1992–2008, with in total 9,654 bank-year observations.

4.1.1 A SOTM on European banks

The SOTM is used for an abstraction of a dataset of financial ratios of European banks before and during the global financial crisis of 2007–2008. We use a set of 20 financial ratios as training variables, and associate two ratios to the model (Tier 1 and Total Capital Ratio).[1] Associationof variables is performed by not including them in training but only computing averages for the data per unit of the SOTM. These particular variables are associated, due to them being important while having a large number of missing values that would impair training. The model architecture is set to 8x17 units. The 17 horizontal units represent the time dimension and the 8 vertical units represent cross-sectional structures per time point *t*. The units on the time dimension span periods before and during the crisis in 2007–2008 (1992–2008) and equals the number of time units *T*, while the number of units at each point in time is determined based upon its descriptive value. It is important to note that the SOTM, likewise the SOM, is not restricted to treat each unit as an individual cluster. Due to the property of approximating probability density functions $p(x,t)$, only the dense locations in the data at time *t* tend to attract units at time *t*. Quality measures are again used when choosing the final specification of the

[1] Tier 1 capital is a ratio used by regulators in assessing a bank's financial strength and measures a bank's core equity capital to its total risk-weighted assets. The Total Capital Ratio also comprises Tier 2 capital, and provides thus a broader measure of the capital base by also including supplementary capital, such as undisclosed and revaluation reserves, loan-loss provisions and subordinated debt.

SOTM. For a SOTM with 8x17 units, we chose a neighborhood radius $\sigma = 2.0$, as it has the highest quantization accuracies and no topographic errors (see Fig. 4).

As mentioned in Section 2, we represent the multidimensionality of the SOTM using feature planes. On feature planes, differences between units along the vertical direction shows differences in cross-sections and differences along the horizontal direction changes over time. This enables assessing two types of patterns: *i*) cross-sectional relations and distributions and *ii*) changes in patterns over time. Below, we summarize some of the patterns that the feature planes in Fig. 5 reveal. First, while being time-varying, one can roughly divide the vertical dimension into different types of banks: the higher the poorer and the lower the better. The upper part represents banks with somewhat poorer *asset quality*, i.e. loan loss provisions (cf. Fig. 6(a)); significantly poorer *capital ratios*, i.e. a low share of equity and capital funds (cf. Fig. 6(b-g)), a high share of subordinated debt (cf. Fig. 6(h)); poorer *operations ratios*, i.e. low interest and other than operating income (cf. Fig. 6(i-k)), low non-interest expenses (cf. Fig. 6(l)), low pre-tax income (cf. Fig. 6(m)), high non-operating expenses (cf. Fig. 6(n)) and low return on assets and equity (cf. Fig. 6(o-p)), and poorer *liquidity ratios*, i.e. low interbank ratio (cf. Fig. 6(q)) and mixed results for loan ratios (cf. Fig. 6(r-t)). The loan ratios are not only mixed as the upper part is characterized by both the highest and lowest values, but also as there is a wave of decreased liquidity in the latter part of the SOTM. The share of low non-interest expenses to assets (cf. Fig. 6(l)) of the poor banks is somewhat contrary to their general performance, as it can be considered a measure of the operating and overhead expenses of a bank relative to the assets invested. Similarly, then the lower part represents well-performing banks with better *asset quality*, *operations*, and *liquidity ratios*.

Second, to assess temporal changes, one has to observe changes on the horizontal direction. We will divide these into gradual and abrupt changes. Some gradual changes are increases in equity and capital funds ratios (cf. Fig. 6(b-g)) and subordinated debt (cf. Fig. 6(h)), and decreases in interest income (cf. Fig. 6(i-j)) and non-interest expenses (cf. Fig. 6(l)). The abrupt changes include significant decreases of interest income (cf. Fig. 6(i-j)) and the interbank ratio (cf. Fig. 6(q)) in 1994, and increases in loan ratios (cf. Fig. 6(r-t)) for the mid and lower part of the map in 2004. In 2008, we can observe significant changes for several ratios: increases in loan loss provisions (cf. Fig. 6(a)), subordinated debt (cf. Fig. 6(h)), non-interest expenses (cf. Fig. 6(l)), non-operating expenses (cf. Fig. 6(n)) and loan ratios (cf. Fig. 6(r-t)), and decreases in other than operation income (cf. Fig. 6(k)), pre-tax income (cf. Fig. 6(m)), returns (cf. Fig. 6(o-p)), and interbank (cf. Fig. 6(q)), tier 1 and 2 ratios (cf. Fig. 6(u-v)). To sum up, while we can observe direct negative effects of the financial crisis in 2008, several ratios indicate a boom or building up of risks and vulnerabilities during the pre-crisis period. However, we can neither objectively label units nor identify *changing*, *emerging* or *disappearing* clusters over time.

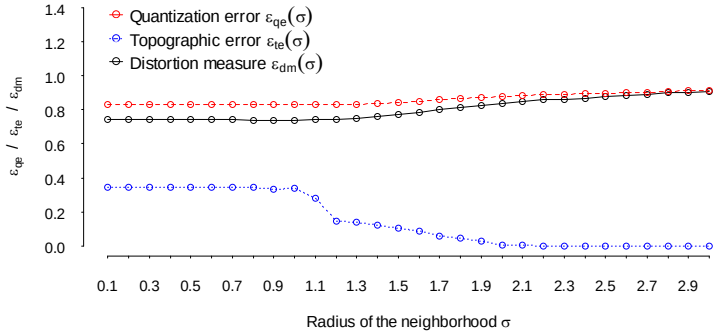


Fig. 4. Quality measures of the SOTM over radius of the neighborhood.

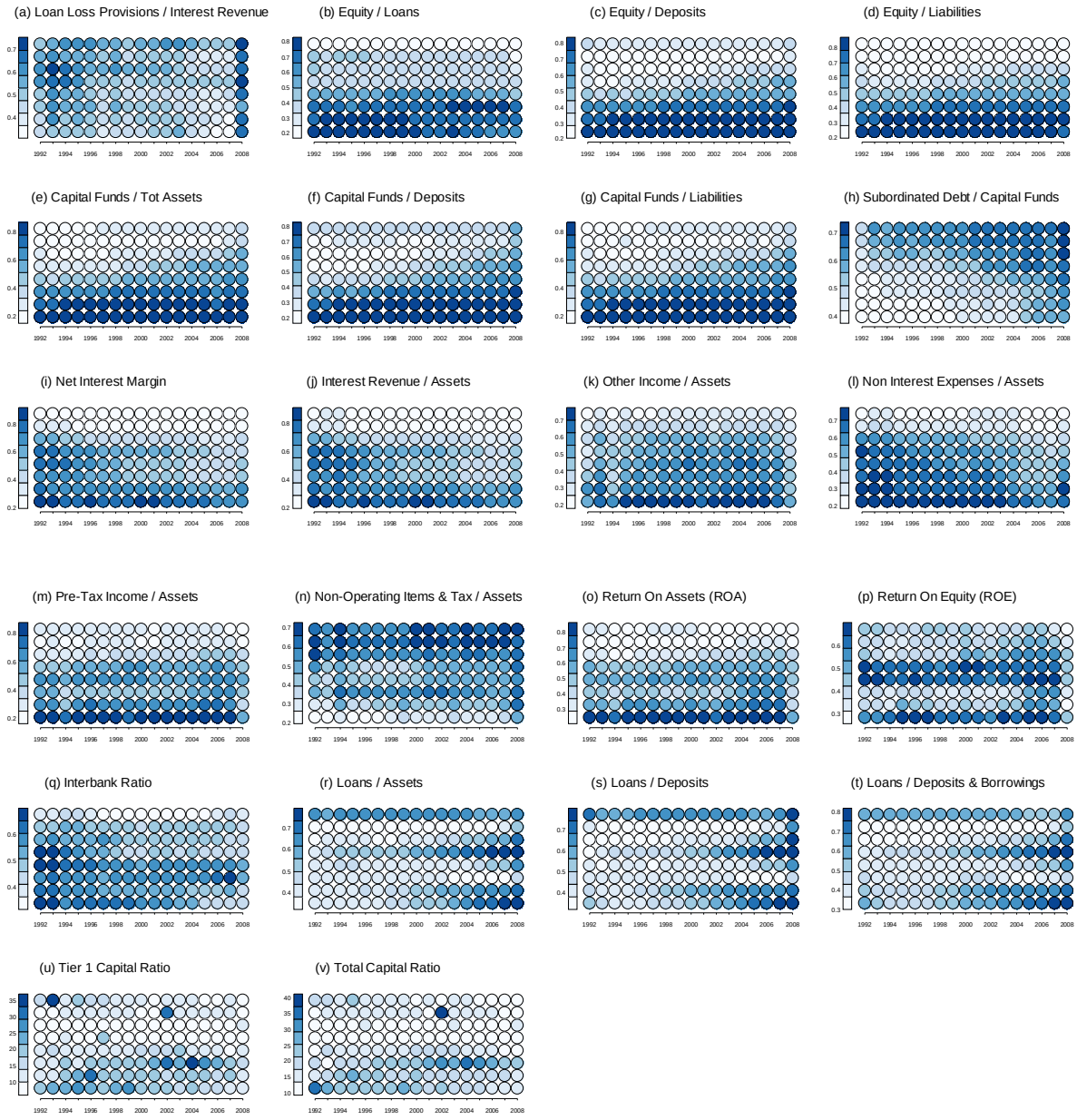


Fig. 5. Feature planes of the bank SOTM.

4.1.2 Clustering of the bank SOTM

We apply a second-level clustering to the units of the trained bank SOTM. As we do not have a pre-defined number of groups in these data, we use cluster validation measures for evaluating the clustering solutions with different K. We do, however, still vary K to explore the structures in the dataset, as is commonly done with hierarchical clustering methods. Fig. 6 shows the Dunn index and Silhouette coefficient for K=3,4,…,10. While Dunn index indicates that K=7,8 is optimal and the Silhouette coefficient indicates that K=9 is optimal, a general view of both measures shows only minor differences between different K. The small differences in cluster validation measures motivate to explore the hierarchical process of agglomerating clusters on the SOTM. The cluster membership planes in Fig. 7 illustrate how clusters are agglomerated when increasing K from 3 to 9. To facilitate the interpretation of the agglomeration process for an increase in K, we set the colors from the ColorBrewer [22] scheme to be constant for all clusters except for the splitted one, where one new color is introduced.

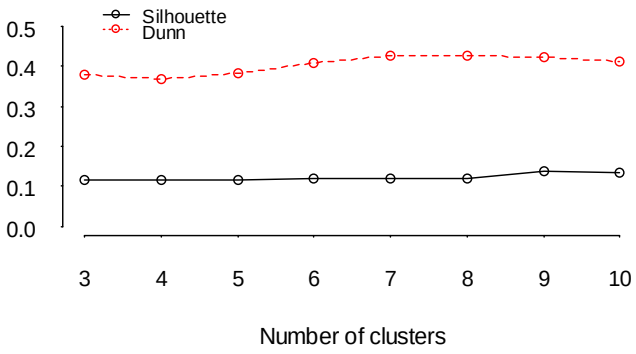
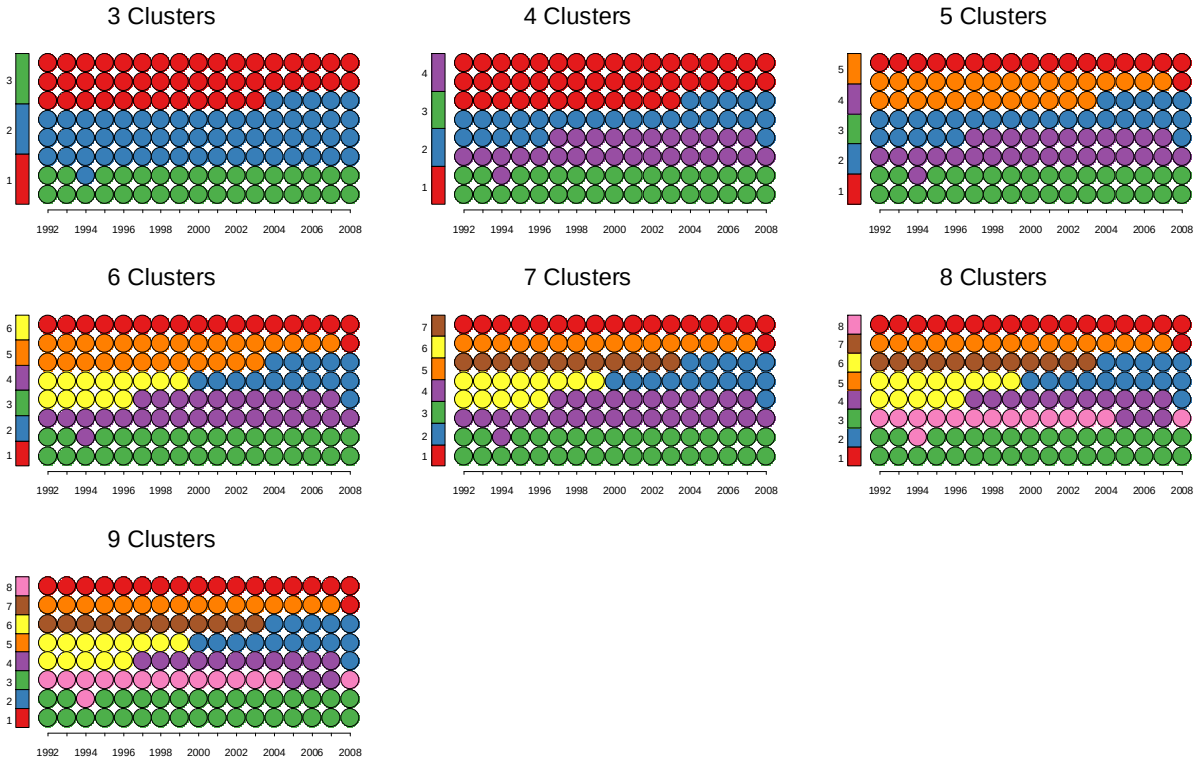


Fig. 6. Cluster validation of the second-level clustering of the bank SOTM.



Notes: Cluster memberships are shown using color coding. The vertical color scale on the left of each grid links a cluster number to each color.

Fig. 7. Cluster membership planes of the bank SOTM.

While agglomeration proceeds by decreasing *K*, we summarize the process of the SOTM clustering from low to high *K* for illustrational purposes. First, the 3 and 4-cluster solutions mainly show cross-sectional differences in data. However, in the 4-cluster solution, gradual *changes* show that clusters 2 and 4 (blue and purple) move and merge upwards. This represents improved boom-like conditions in the entire European banking sector that may have lead to an overheating and build-up of risks and vulnerabilities prior to the peak of the crisis. Second, the 5-cluster solution starts to reveal the abrupt change in the conditions during the peak of the crisis in 2008. The downward movement of cluster memberships of clusters 1 and 2 (red and blue) shows that bank performance decreased significantly as a result of the global financial crisis. Third, the 6-cluster solution illustrates a horizontal agglomeration, or *loss* of one cluster and *emergence* of another, where cluster 2 (blue) in the 5-cluster solution consists of two horizontally separate clusters 2 and 6 (blue andyellow) in the 6-cluster solution. Fourth, the 7, 8 and 9-cluster solutions actually represent a decrease in bank performance prior to the crisis in 2008. The 9-cluster solution reveals that from 2005 onwards clusters 6 and 8 include units on a lower row in addition to only horizontal units. In addition, some clusters *re-emerge* after *disappearing*, while some clusters *emerge* to *disappear* some periods later. As confirmed by the feature planes in Fig. 5(r-t), these

patterns can to a large extent be derived from changes in loan rations. That is, the liquidity of banks starts decreasing prior to the peak of crisis in 2008. The 10-cluster solution, or above, start to include detail to the extent that interpretation is hindered.

4.2 An application to currency crises

This section turns the attention to cyclical changes in indicators of currency crises. We present an application of a SOTM, with its second-level clustering, to indicators commonly used to describe and explain the vulnerabilities to a currency crisis. These indicators were originally introduced to crisis monitoring in a predictive model of the International Monetary Fund by Berg and Pattillo [33]. In [17], a currency crisis map based upon similar data and the SOM was built for visual temporal and cross-sectional surveillance of risks and vulnerabilities. Again, while holding some promise for predicting vulnerabilities and visualizing trajectories of economies, these applications have not involved any assessment of temporal changes in cross sectional cluster structures. The dataset consists of five monthly indicators for 23 emerging market economies from 1986:1–1996:12 with in total 2,916 country-month observations: foreign reserve loss, export loss, real exchange-rate overvaluation relative to trend, current account deficit relative to GDP, and short-term debt to reserves. To control for cross-country differences, we transform each indicator into its country-specific percentile distributions. In addition, we use an exchange market pressure index to date crises, as defined in [33]: a crisis occurs if the sum of a weighted average of monthly percentage depreciation in the currency and monthly percentage declines in reserves exceeds its mean by more than three standard deviations. Using the crisis occurrences, we define an observation to be in a vulnerability state, or pre-crisis period, if it experienced a crisis within 24 months.

4.2.1 A SOTM on currency crises

The SOTM is used for an abstraction of macro-financial indicators prior to the Asian financial crisis in 1997. In addition to a set of training variables, we associate a class variable to the SOTM, namely the occurrence of pre-crisis periods. The rationale for associating the class variables is that it enables us to locate risks and vulnerabilities on the SOTM. Model architecture is set to 6x11 units. The 11 horizontal units represent the time dimension with annual frequency, as monthly data had to be pooled to a more granular solution. Quality measures are again the basis for choosing a final specification of the SOTM; however, for brevity they are not reported here. With the highest quantization accuracies and no topographic errors, we chose a neighborhood radius $\sigma = 2.6$.

In this application we will only briefly discuss the properties of the SOTM, and rather turn the focus to the clustering results. Still, we provide a short discussion on the interpretation of the crisis SOTM. High values of the pre-crisis class variable throughout the lower part of the SOTM illustrates the location of vulnerable states (cf. Fig. 8(f)). In fact, the pre-crisis variable shows that pre-crises, and thus crises, occur in three waves, which is also confirmed by the class frequency: one in 1986, one in 1992 and one in 1996. The indicators are transformed such that a univariate increase leads, on average, to increased risks. Accordingly, reserve loss takes larger values on the bottom part of the map, whereas there is no clear time trend (cf. Fig. 8(a)). While export loss also shows that increases lead to larger vulnerabilities, it illustrates clear increases during the most vulnerable times (cf. Fig. 8(b)). Exchange-rate overvaluation again increases with vulnerabilities, but shows a remarkably increasing trend towards the latter part of the sample (cf. Fig. 8(c)). The current account deficits exhibit similar behavior (cf. Fig. 8(d)). Short-term debt, while decreasing with vulnerabilities, have an increasing relationship with vulnerabilities towards the end of the sample (cf. Fig. 8(e)). However, the short-term debt has, overall, decreased over time.

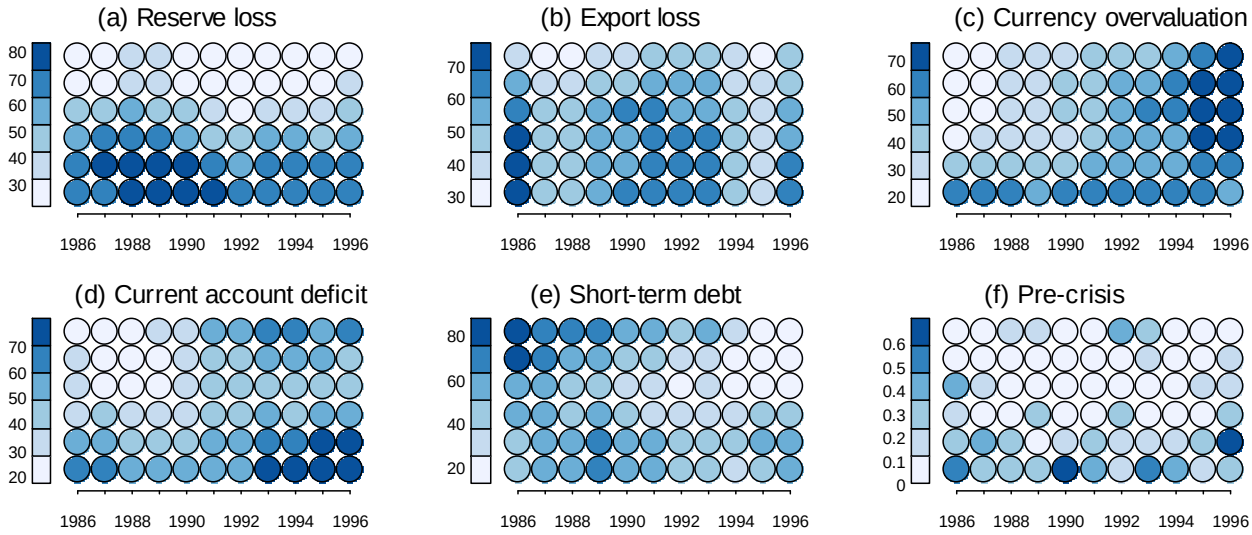


Fig. 8. Feature planes of the crisis SOTM.

4.2.2 Clustering of the crisis SOTM

The second-level clustering of the SOTM better reveals the cyclical nature of the indicators. In the following, rather than discussing all details of cluster structures, the main focus is on changes in the clusters with the highest risks, namely the lower part of the SOTM. Hence, we do not report the cluster validation measures here, but rather turn to directly to the agglomeration process. As above and shown in Fig. 9, we apply Ward's hierarchical clustering, this time ranging from a 3 to a 8-cluster solution due to fewer units. Already the model with 3 clusters illustrates an interesting behavior of cluster 3 (green), the high-risk cluster. Cluster 3 exhibits the following *changes*: it is in the first period a 2-unit solution and decreases then to only including 1 unit, whereas the second wave of vulnerability attracts 2 units and the third wave further increases the solution to include 3 units. While the solutions up to 7 clusters provide information on cluster structures in the less risky parts of the SOTM, e.g., *changing*, *emerging* and *disappearing* clusters, the 8-cluster solution decouples a part from the risky cluster (cluster 3). This part (cluster 8) may be interpreted as even more risky, not the least due to its profile as per the pre-crisis variable (cf. Fig. 8(f)) and the fact that a unit above cluster 8 is a member of cluster 3.

Notes: Cluster memberships are shown using color coding. The vertical color scale on the left of each grid links a cluster number to each color.

Fig. 9. Cluster membership planes of the crisis SOTM.

5. Conclusion

In this paper, we have combined the SOTM with classical cluster analysis. The rationale for clustering the SOTM is twofold: *i*) it provides an objective means for labeling units and *ii*) it enables identification of *changing*, *emerging* and *disappearing* clusters over time. We have used two sorts of experiments for illustrating the rationale behind and usefulness of second-level clustering of the SOTM. First, we used simple toy data for demonstrating the need for clustering of the SOTM. Second, we present applications in two real-world settings: financial ratios of European banks and indicators of currency crisis. Not surprisingly, the results indicate a build-up of risks and vulnerabilities throughout the European banking sector prior to the start of the crisis. The second application identifies the cyclicality of currency crises through increases in the size of the most vulnerable clusters prior to times of crisis. Thus, we have illustrated with four experiments the usefulness of a second-level clustering of the SOTM for better revealing the temporal evolution of cluster structures. In addition, the two real-world applications have further confirmed the usefulness of the SOTM in financial settings. This opens the door for a wide range of applications to illustrating dynamics in customer data, firm data, macroeconomic data, financial data, process monitoring data, etc.

Acknowledgments

We acknowledge Barbro Back for constructive comments and Academy of Finland (grant no. 127592) for financial support.

References

[1] P. Sarlin, Self-Organizing Time Map: An Abstraction of Temporal Multivariate Patterns, Neurocomputing, 99(1) (2012) 496-508.

[2] T. Kohonen, Self-organized formation of topologically correct feature maps, Biological cybernetics, 43 (1982) 59-69.

[3] Z. Yao, P. Sarlin, T. Eklund, B. Back, Temporal Customer Segmentation Using the Self-organizing Time Map, in: Proceedings of the International Conference on Information Visualisation (IV 2012), Montpellier, France, pp. 234-240.

[4] P. Sarlin, Decomposing the Global Financial Crisis: A Self-Organizing Time Map, in: Proceedings of the International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2012), San Sebastian, Spain, 2012.

[5] A. Ultsch, H.P. Siemon, Kohonen's self-organizing feature maps for exploratory data analysis, in: Proceedings of the International Conference on Neural Networks (ICNN 1990), Dordrecht, the Netherlands, 1990, pp. 305–308.

[6] S. Kaski, J. Venna, T. Kohonen, Coloring that reveals cluster structures in multivariate data, Australian Journal of Intelligent Information Processing Systems, 6 (2000) 82-88.

[7] J.W. Sammon Jr, A nonlinear mapping for data structure analysis, IEEE Transactions on Computers 18 (1969) 401–409.

[8] J. Kangas, Time-delayed self-organizing maps, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN 1990), San Diego, USA, 1990, pp. 331–336.

[9] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Transactions on Neural Networks, 11(3) (2000) 586-600.

[10] C.M. Reinhart, K.S. Rogoff, Is the 2007 US sub-prime financial crisis so different? An international historical comparison, American Economic Association, 98(2) (2008) 339-344.

[11] C.M. Reinhart, K.S. Rogoff, The aftermath of financial crises, American Economic Review, 99(2) (2009) 466-472.

[12] T. Kohonen, The "Neural" Phonetic Typewriter, Computer 21(3) (1988) 11–22.

[13] T. Kohonen, The Hypermap Architecture, in: Artificial Neural Networks Vol. II, T. Kohonen, K. Mäkisara, O. Simula, J. Kangas (Eds.), Elsevier, Amsterdam, Netherlands, 1991, pp. 1357–1360.

[14] G.J. Chappell, J.G. Taylor, The temporal Kohonen map. Neural Networks 6 (1993) 441–445.

[15] D. Guo, J. Chen, A.M. MacEachren, K. Liao, A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP), IEEE Transactions on Visualization and Computer Graphics, 12(6) (2006) 1461-1474.

[16] P. D'Urso, L. De Giovanni, Temporal self-organizing maps for telecommunications market segmentation, Neurocomputing, 71(13) (2008) 2880-2892.

[17] P. Sarlin, D. Marghescu, Visual Predictions of Currency Crises using Self-Organizing Maps, Intelligent Systems in Accounting, Finance and Management 18(1) (2011) 15–38.

[18] P. Lingras, M. Hogo, M. Snorek, C. West, Temporal analysis of clusters of supermarket customers: conventional versus interval set approach, Information Sciences, 172(1-2) (2005) 215-240.

[19] B. Back, K. Sere, H. Vanharanta, Managing complexity in large data bases using self-organizing maps, Accounting, Management and Information Technologies, 8(4) (1998) 191-210.

[20] Denny, W. Graham, P. Christen, Visualizing temporal cluster changes using Relative Density Self-Organizing Maps, Knowledge and Information Systems 25(2) (2010) 281–302.

[21] Denny, D.M. Squire, Visualization of cluster changes by comparing Self-Organizing Maps, in: Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2005) Springer, Berlin, 2005, pp. 410–419.

[22] M. Harrower, C. Brewer, Colorbrewer.org: an online tool for selecting colour schemes for maps, The Cartographic Journal, 40(1) (2003) 27-37.

[23] P.J. Rousseeuw, L. Kaufman, Finding groups in data: An introduction to cluster analysis, Wiley-Interscience, 1990.

[24] P. Hansen, M. Delattre, Complete-link cluster analysis by graph coloring, Journal of the American Statistical Association, 73(362) (1978) 397-403.

[25] R.K. Blashfield, Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods, Psychological Bulletin, 83(3) (1976) 377.

[26] F.K. Kuiper, L. Fisher, 391: A Monte Carlo comparison of six clustering procedures, Biometrics, 31(3) (1975) 777-783.

[27] R. Mojena, Hierarchical grouping methods and stopping rules: an evaluation, The Computer Journal, 20(4) (1977) 359-363.

[28] J.H. Ward Jr, Hierarchical grouping to optimize an objective function, Journal of the American statistical association, 58(301) (1963) 236-244.

[29] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, Journal of Cybernetics, 3(3) (1973) 32-57.

[30] R.M. Cormack, A review of classification, Journal of the Royal Statistical Society. Series A (General), 134(3) (1971) 321-367.

[31] P. Sarlin, T. Eklund, 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), Springer Verlag, Kaiserslautern, Germany, 2011, pp. 183–192.

[32] P. Sarlin, Z. Yao, T. Eklund, A Framework for State Transitions on the Self-Organizing Map: Some Temporal Financial Applications, Intelligent Systems in Accounting, Finance and Management, 19(3) (2012) 189-203.

[33]A. Berg, C. Pattillo, What caused the Asian crises: An early warning system approach, Economic Notes, 28(3) (1999) 285-334.

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 978-952-12-2828-5
ISSN 1239-1891