7

Tracking and Visualization of Cluster Dynamics by Sequence-based SOM

Ken-ichi Fukui[1], Kazumi Saito[2], Masahiro Kimura[3] and Masayuki Numao[1]
[1]Osaka University, [2]University of Shizuoka, [3]Ryukoku University
Japan

1. Introduction

Since events and physical phenomena change with time, it is important to capture the main transitions and elements of such events and phenomena. Such transitions can be seen to occur in the World Wide Web (Levene & Poulovassilis, 2004), news topics (Allan, 2002), a person's health condition, and the state of an instrument or a plant. Transition, or change, refers to a sequential increase/decrease or generation/extinction of the feature of the object. Visualization of such transitions of dynamic clusters is helpful in understanding such phenomena instinctively and plays a useful role in many application domains, such as fault diagnosis and medial examinations.

Although a number of clustering methods have been proposed (Jain et al., 1999), most conventional clustering methods deal with static data and cannot handle sequential changes of the cluster explicitly. Tracing the trajectory within clusters that have been collectively processed and a sliding window-based method to generate separate clusters can be considered as simple methods. Although the former method cannot trace changes of clusters, it can trace changes in the number of data that belong to each cluster. The latter method can handle changes of clusters to a certain degree. However, there are some problems, such as setting an appropriate window size, the inevitable decrease in the number of data within a window, and the correspondence relationships of clusters between windows.

The present study considers a window-based approach using the temporal neighborhood to address the above-described problems. Kohonen's Self-Organizing Map (SOM) (Kohonen, 2000) is considered to be an appropriate technique for visualizing clusters and their similarity relationships. The SOM is an unsupervised neural network learning technique that produces clusters and subsequently projects them onto a low-dimensional (normally two-dimensional) topology map. The conventional SOM deals with static data. However, we have extended the SOM learning model by introducing the Sequencing Weight Function (SWF), so that the model can visualize the transition of dynamics clusters. This model is referred to herein as the Sequence-based SOM (SbSOM) (Fukui et al., 2008). A SOM-based method was selected because the SOM has a neuron topology in the feature space and that is associated with topology (visualization) space. The introduction of temporal order into the topology is natural. The proposed method mitigates the problems of appropriate

window size and the decrease of the number of data. Moreover, owing to the neighborhood function, the spatio-temporal neighborhood become the topological neighborhood, the correspondence relation problem of clusters can be solved.
Projection methods include conventional Multi-Dimensional Scaling (MDS), using the first and second components of Principal Component Analysis (PCA), and Locally Linear Embedding (LLE) (Roweis & Saul, 2000), and graph-based methods include spring embedding, the Laplacian eigenmap (Belkin & Niyogi, 2002), and the ISOMAP (Tenenbaum et al., 2000). However, these methods project each sample, because clustering is not performed. As such, these are not suitable for large data sets.
A number of studies have been conducted to track cluster changes, but these studies have concentrated on tracking topic changes. For example, TimeMines (Swan & Jensen, 2000) extracts topics and visualizes the periods and intensities of the topics by a χ^2 test from a sequence of words related to the documents. T-Scroll (Ishikawa & Hasegawa, 2007) proposed the introduction of a temporally gradual decrease model to K-means clustering. However, this does not clearly solve the problem of correspondence of clusters between windows. As extension of the static topic model to a dynamic model, Kimura et al. (2005) proposed the multinominal PCA topic model to extract latent topic trends, and Wang and McCallum (2006) proposed the Latent Dirichlet Allocation (LDA) based topic model over time. In addition, in network analysis, Qian et al. (2009) extracted influential research topics and tracked them using a network of research paper citations using the community inference method. An advantage of the proposed approach compared to previous studies is that the SbSOM is a general framework so that it is applicable to any data that can be represented by vectors. In addition, the SbSOM is easy to extend to graph, string, and other structured data by applying the recently developed kernel method (Bishop, 2006). Moreover, the kernelized SOM has been proposed (Lau et al., 2006).
On the other hand, various approaches have been proposed to introduce the concept of time to the SOM model (Barreto & Arajo, 2001). However, these approaches have been introduced as physiological models for short-/long-term memory, or to handle time series data. In contrast, the purpose of the present study is to track cluster changes.
We applied the proposed method to real-world applications, a news articles data set (Fukui et al., 2008), a medical data set (Fukui et al., 2006), and an Acoustic Emission (AE) signal data set (Fukui et al., 2007). Based on the results of these applications, we have confirmed that the map obtained using the SbSOM can interpret as a phenomenon, topic transition from a news articles data set and cluster transition of patients from a medical data set.
The remainder of the present chapter is organized as follows. Section 2 describes the characteristics of cluster dynamics. Section 3 introduces the learning algorithm of the proposed SbSOM. Section 4 presents an application of SbSOM to transition of news topics. Section 5 introduces a visualization method using class labels for sequence data. Section 6 presents an application involving medical data. Finally, conclusions are presented in Section 7.

2. What is Cluster Dynamics?

This section describes which characteristics of cluster change should be tracked. Cluster change can be represented by the following basic three characteristics (see also Fig. 1):

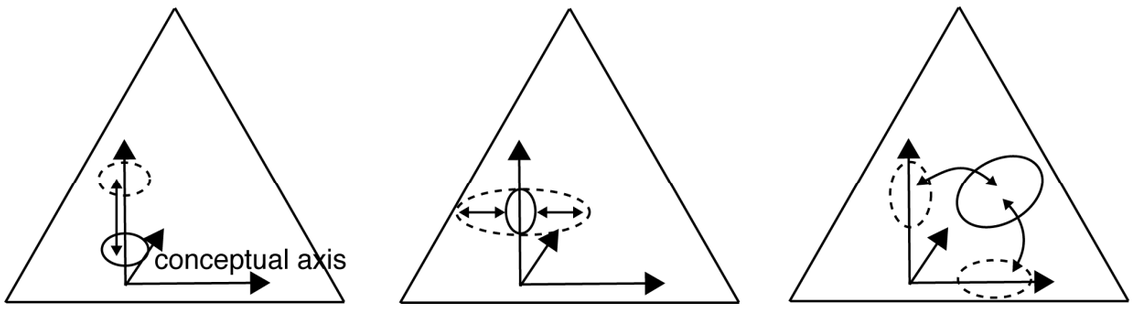

Fig. 1. Characteristics of cluster dynamics. These figures illustrate high-dimensional space based on the concept vector, e.g., a topic represented by a certain direction in the Bag-of-Words vector space.

1) **Movement:** a change in the feature of the cluster. In particular, if a certain direction in the feature space represents a certain concept, e.g., topic, movement and direction indicate the increase/decrease of the degree of the concept, e.g., the level of interest in a topic.
2) **Range:** a change in coverage representing the variety of the cluster. For instance, the divergence/convergence of a topic.
3) **Merger/Separation:** a change wherein more than two clusters merge into one cluster or one cluster separates into multiple clusters. This can be interpreted as a change in topic, i.e., the derivation of a topic. (However, mergers of topics appear to be rare.)

3. Sequence-based Self-Organizing Map

3.1 Algorithm

The Sequence-based SOM (SbSOM) is based on Kohonen's Self-Organizing Map (SOM). Let v-dimensional N inputs be $X_n = (x_{n,1}, \cdots, x_{n,v})$, $(n = 1, \cdots, N)$. Let the position of M neurons in the visualization layer be $r_m = (y_m, z_m)$, $(m = 1, \cdots, M)$, and let the reference vector corresponding to the m^{th} neuron be $W_m = (w_{m,1}, \cdots, w_{m,v})$.

The following is the learning algorithm that uses a batch process and the decreasing strategy of the learning parameter.

Batch sequence-based SOM learning algorithm

Step 1. Initialize the reference vectors $\{W_1, \cdots, W_M\}$ randomly.

Step 2. Search winner neurons for all inputs $\{X_1, \cdots, X_N\}$. (This step is described in the next subsection.)

Step 3. Exit if the winner neurons $\{c(X_1), \cdots, c(X_N)\}$ were not changed.

Step 4. Update the reference vectors $\{W_1, \cdots, W_M\}$ by the following equation:

$$W_m^{new} := W_m + h_{c(X_n),m}[X_n - W_m] \quad (1)$$

where $h_{c(X_n),m}$ is a neighborhood function that defines the effect of the neighborhood of the winner. Typically, a Gaussian function is used:

$$h_{c(X_n),m} = \alpha \exp\left\{-\frac{\left\|r_m - r_{c(X_n)}\right\|}{2\sigma^2}\right\} \tag{2}$$

Step 5. Decrease the learning parameters σ every several iterations. Return to Step 2.

3.2 Sequencing Weight Function

In the conventional SOM, the winner neuron is determined only by spatio-distance. In contrast, in the SbSOM, by introducing the Sequencing Weight Function (SWF), weights are assigned to the neuron topology according to the sequence of the data (Fig. 2). The SWF introduces the concept of time to the topology. Note that this is a loose concept of time because it allows reversal of the data order that appears on the neuron topology by the winner. The winner is determined by spatio-temporal distance using SWF $\phi(n,m)$ as follows:

$$c(X_n) = \arg\min_m \phi(n,m)\left\|X_n - W_m\right\|. \tag{3}$$

Suppose t_n is a time stamp for the n^{th} data, and that the m^{th} neuron is located at η_m / η_M in a certain direction on the topology (in this case, the η-direction). Let the absolute value of the difference in these ratios be $\varepsilon = \left|t_n / t_{end} - \eta_m / \eta_M\right|$. When there is no time stamp associated with the data, but the data is obtained sequentially, ε is given as $\varepsilon = \left|n/N - \eta_m / \eta_M\right|$. Then, the SWF is defined so as to be able to balance the spatio-temporal resolution by allowing reversal of data order:

$$\phi_{\exp}(n,m) = e^{\beta\varepsilon}, \tag{4}$$

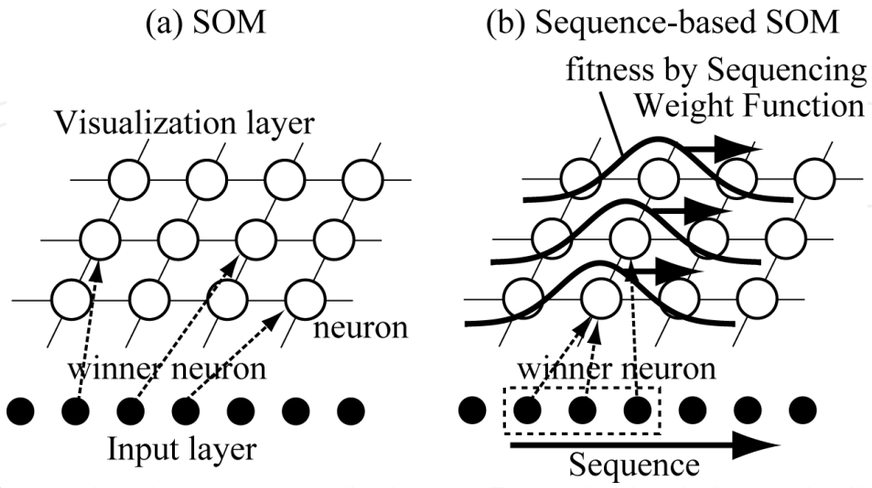

Fig. 2. Difference in winner neuron selection. (a) Determined only by spatio-distance in the SOM. (b) Determined by spatio-distance under the sequencing weight function in the SbSOM.

where $\beta > 0$ is a parameter that controls the degree of influence of the data order, i.e., temporal distance. As β increases, the temporal distance becomes increasingly dominant. Instinctively, by shifting the SWF in a certain direction on the neuron topology, the SbSOM can produce time axis onto the topology (Fig. 2). Note that when $\beta = 0$ (i.e., $\phi(n,m) = 1$), SbSOM will be exactly equivalent to the standard SOM.
In case of style of restricted sliding window, the SWF can be given by a rectangle function:

$$\phi_{rect}(n,m) = \begin{cases} 1 & \varepsilon < 1/2K \\ \infty & otherwise \end{cases} \quad (5)$$

where K is the number of neurons in a certain direction (e.g., the η-direction). When β in is taken to be sufficiently large, $\phi_{\exp}$ is equivalent to ϕ_{rect}. The advantage of the sliding window is reduced computational cost, because the window requires only comparison with reference vectors within a window.

3.3 Neighborhood Function in the SbSOM

The SOM includes the neighborhood function, so that the reference vectors are updated according to the spatio-neighborhood. The neighborhood function in the SbSOM, on the other hand, is a spatio-temporal neighborhood related to SWF (Fig. 3). Therefore, the SbSOM can perform clustering with the help of temporal neighborhood data even when using ϕ_{rect}.

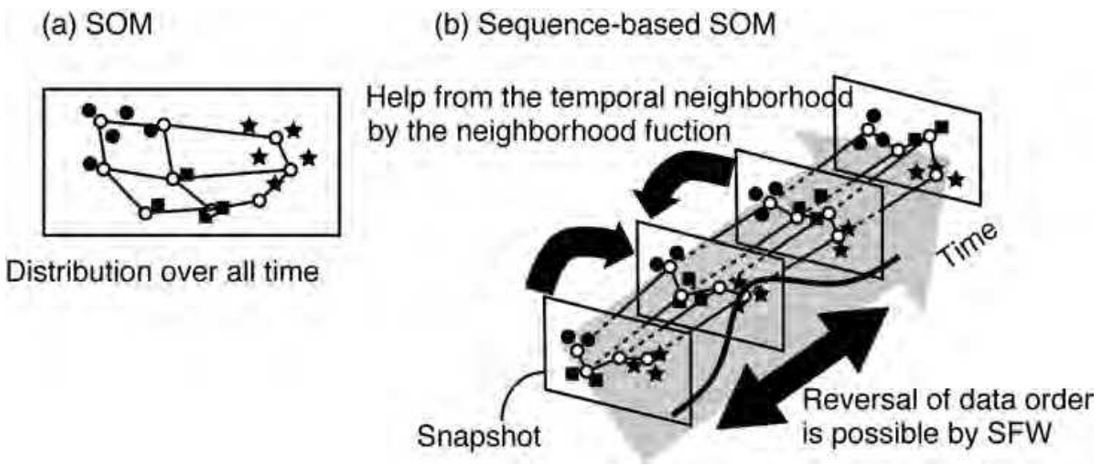

Fig. 3. The spatio-neighborhood becomes the topological neighborhood (a) in the SOM and (b) in the SbSOM. This figure shows a restricted version of the SbSOM.

This property mitigates the problems of appropriate window size and decreased sample data in window-based clustering. Moreover, cluster correspondence can be self-organized because, in the SbSOM, the spatio-temporal neighborhood becomes the topological neighborhood.

4. Experiment 1: Transition of News Topics

4.1 Dataset and Pre-processing

This experiment uses international articles from the 'Mainichi' Japanese newspaper from January to December of 1993[1]. The total number of articles is 5,824, and the total number of unique words is 24,661, after eliminating pre-defined stop words. Each article is represented by a Bag-of-Words model, namely a term frequency - inverse document frequency (tf-idf) model.

Assuming a topic is distributed around certain direction in the vector space of the words, we applied Principal Component Analysis (PCA) to extract topic axes (feature axes). The dimension of the weight vector of the words is reduced by projecting the words into the topic axes space. In addition, a topic class (label) is assigned to each article by the topic axis index with the maximum component. See (Kimura et al., 2005) for a detailed description of topic extraction. Note that the SbSOM is applicable to any topic extraction method as long as the topics are represented by vectors.

In this experiment, the number of PCs was set to 10, which was empirically adjusted. Therefore, the number of extracted topics was 20 as positive and negative direction of each PC indicates different topic, since the origin is the center of the entire articles. The number of articles associated with the topics and the topic titles are listed in Table 2. These topic titles were assigned by two individuals who read articles associated with the topics and merged the topics to obtain consistent titles.

4.2 Results for the News Article Dataset

Visualization results using the conventional SOM and the SbSOM are shown in Fig. 4 and Fig. 5, respectively. The numbers denoted in the figures indicate the representative topic of the cluster as determined by majority decision of topic labels of the articles belonging to the cluster. The neuron topology was set to 24x20 (regular grid), and $\phi_{\exp}$ was used for the SWF with the parameter $\beta = 150.0$. In the SbSOM (Fig. 5), the x direction is the time axis, as indicated by the months of the year.

Clearly, the SOM can perform clustering of related articles (Fig. 4), whereas the SbSOM also provides change of topics. In addition, Fig. 6 shows the intensity distribution of each component that represents a topic, which indicates the level of interest in a topic. For example, the level of interest in the 7th topic (Chinese Situation) is high around March and December, reflecting the election of Chinese president on March 27th and the 100th anniversary of Mao Tse-Tung's birth, respectively.

No.	# of articles	Topic – Subtopics
1	453	Russian situation - Opposing President Yeltsin and the national assembly - Creating a new constitution
2	32	--
3	303	Middle East peace problem - Israel/PLO agreement and trends of involved nations
4	43	--

[1] http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html (in Japanese)

5	471	Cambodian general election
6	148	North Korea matter - Nuclear problem
7	721	Chinese situation - Economics - Change of government
8	452	Bosnian conflict - Bosnian peace conference - Tendency of U.S.A. and the United Nations
9	171	Military and diplomacy in U.S.A. - Foreign policy of Clinton administration
10	290	Bosnian conflict - Muslim influence - Movement of Muslims
11	366	Iraq problem
12	436	North Korea matter - Nuclear problem - North and South Korea conflict
13	345	Japan and foreign countries - Postwar compensation issue - Increasing role in the United Nations
14	230	Middle East situation - General election in Pakistan
15	327	North Korea and China
16	260	EC Integration Summits - including APEC, EC, ASEAN…
17	253	Asian Situation - including China, Taiwan, Korea…
18	128	Russian circumstance - Election of a new national assembly
19	127	Elections in various countries
20	268	Miscellaneous conferences

Table 2. Extracted topics and manually assigned titles.

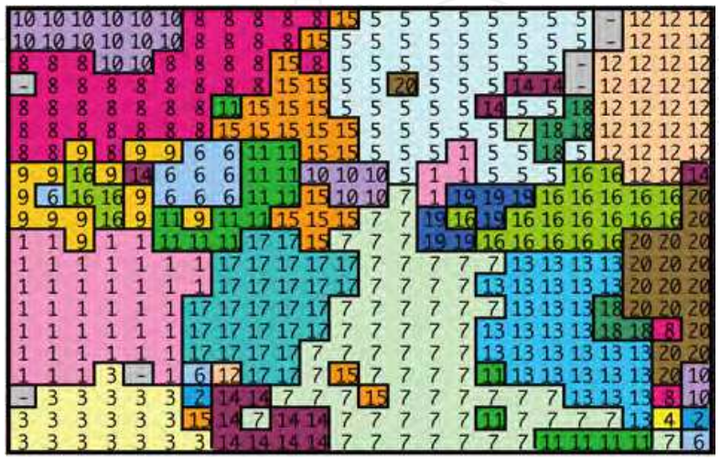

Fig. 4. Mapping result by SOM (represented by majority topic).

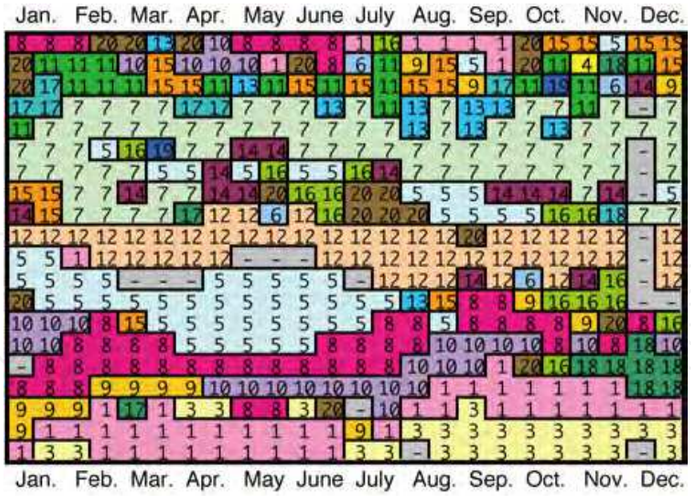

Fig. 5. Mapping result by Sequence-based SOM (represented by majority topic).

Derivation of a Topic

The 9th topic is related to military affairs and diplomacy in the United States. The principal event was the inauguration of President Clinton on January 20th. Therefore, the level of interest in the 9th topic was high around January (Fig. 6). The level of interest in the 9th topic was also high from around March to June, where the majority topic of these clusters is the 5th topic in Fig. 5, which means these two topics are highly related each other. In addition, the levels of interest in the 6th and 11th topics were high around January, where the majority topic is the 9th topic. The 5th topic (Cambodian General Election), the 6th topic (North Korean Nuclear Problem), and the 11th topic (Iraq Problem) are topics that are closely related to President Clinton. That is to say the map shows engagement of President Clinton to these topics.

Furthermore, the 8th and 10th topics are both related to the Bosnian conflict. However, the sub-topics are different for these topics. These topics appear close to each other within the overall map (Fig. 5). In addition, Fig. 6 shows that the 10th topic, which is related to the movement of Muslims, is derived from the 8th topic.

Diversification and Convergence of a Topic

Diversification/convergence of a topic is represented by expansion/contraction in the y direction with the map. Since the y direction in this SbSOM indicates that the neighborhood on the Bag-of-Words vector space is the same as that in the conventional SOM, expansion in y direction represents an increase of the variety of words related to the topic, i.e., diversification of the topic. For example, the variety of words related to topic #5 (Cambodian General Election) increases toward the general election in May and then decreases. Diversification and convergence can also be seen in the component maps (Fig. 6).

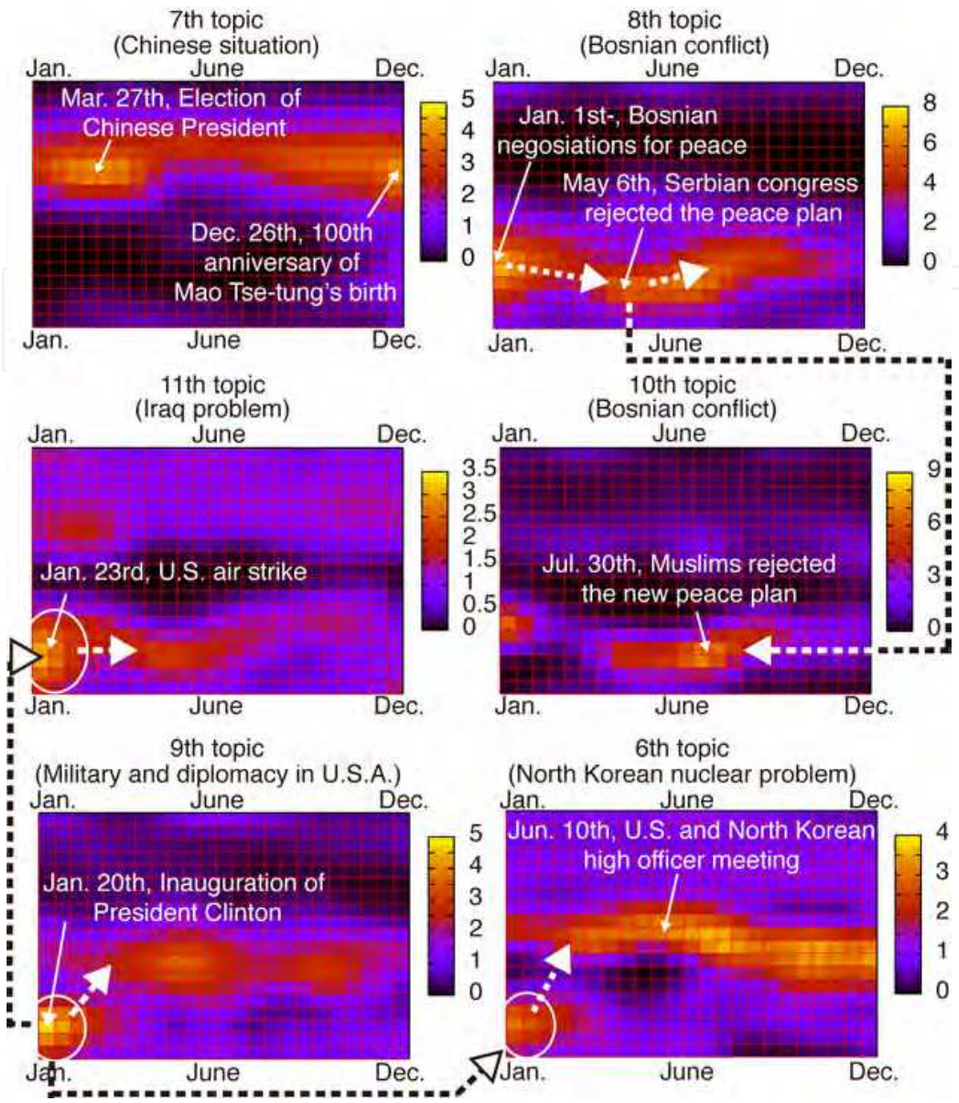

Fig. 6. Maps of level of interest in each topic corresponding to Fig. 5. Overlapped region of high level in different topics indicate relation between the topics. Considering time ordering of high level regions, dotted lines indicate derivation of the topic.

4.3 Effect of the Sequencing Parameter

In this subsection, we investigate the effect of the sequencing parameter in the SWF on visualization. Figure 7 shows one component changing the parameter β. When $\beta = 0$, the SbSOM is exactly the same as the conventional SOM. As β increases, the level of interest gradually spreads in the time direction and stabilizes after $\beta = 100$. This parameter is not sensitive to the result if it is sufficiently large value.

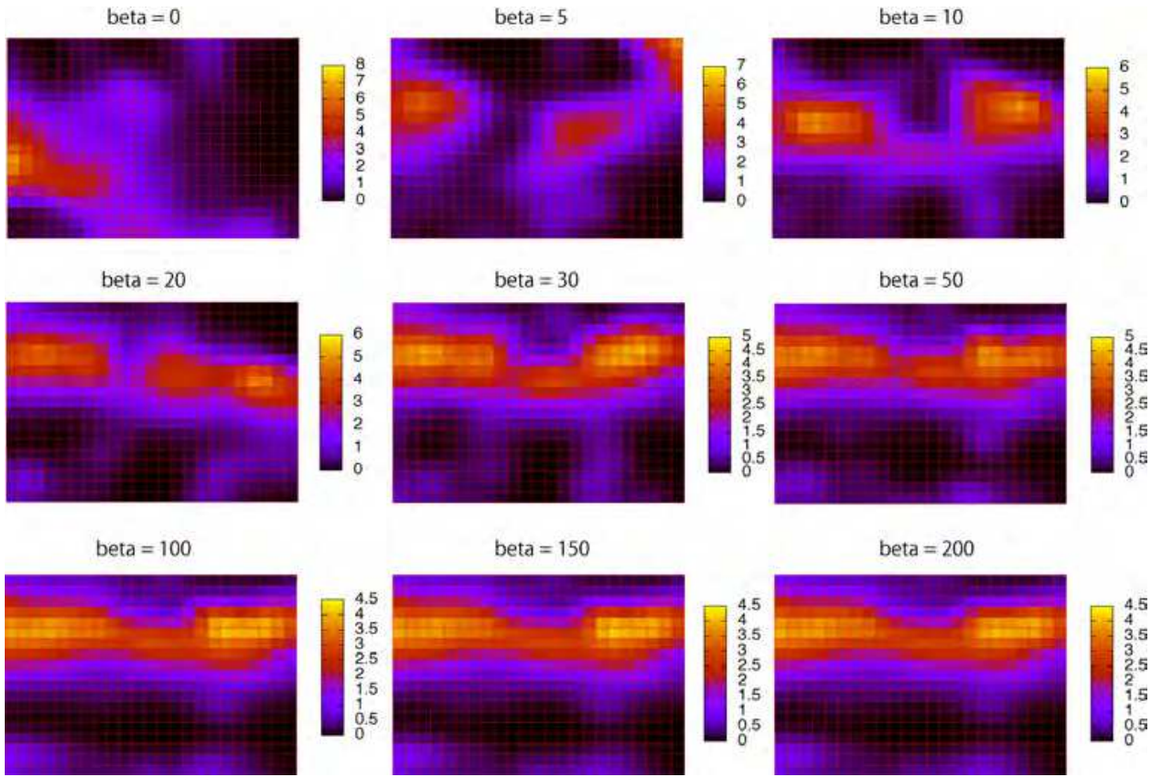

Fig. 7. Effect of the sequencing parameter on visualization.

5. Application to Two Classes of Sequential Data

5.1 Concept

This section presents the visualization architecture for sequential data using the class label. The conventional SOM does not consider the class label while learning because the learning in the SOM is an unsupervised learning. The nature of unsupervised learning is to find latent clusters within data or to find relationships between attributes. However, it is natural to use class label information when available. Although a few studies have extended the SOM to the supervised SOM, e.g., Fritzke, 1994, Hagenbuchner & Tsoi, 2004, these studies did not preserve the nature of the unsupervised learning.

Consequently, preserving the nature of unsupervised learning, we proposed the idea of using the class label to highlight clusters that are related to a specific class (Fig. 8). After SbSOM learning, trajectories of the winner neuron indicating a series of data are obtained. Afterwards, weights assigned to neuron nodes are trained by a single-layer perceptron. The obtained weights are used for visualization, e.g., color gradation. A period that has high correlation with a class is expected to have high weights, and vice versa. The reason for selecting a single-layer perceptron is so that connectivity weights can be paired with neuron nodes.

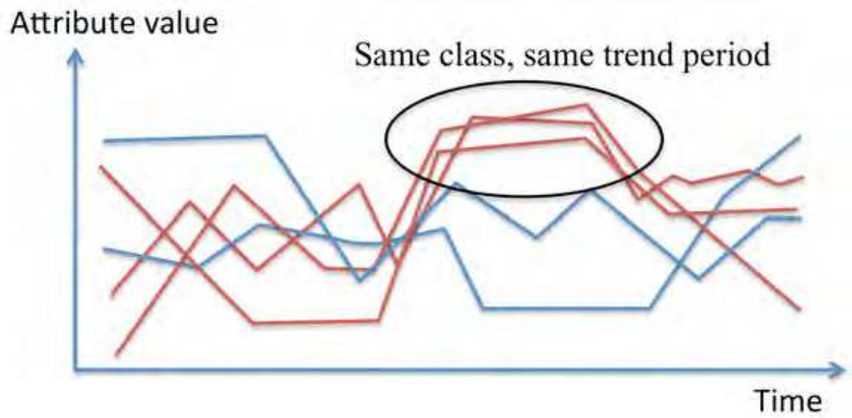

Fig. 8. Concept of using the class label for sequence data. Periods having the same class and trend should be highlighted.

5.2 The algorithm

The proposed visualization architecture is illustrated in Fig. 9. The advantages of this methodology are to derive the interpretation of the discriminant function for the map via the perceptron and to suggest the cluster in a sequence that involves classification through the SOM.

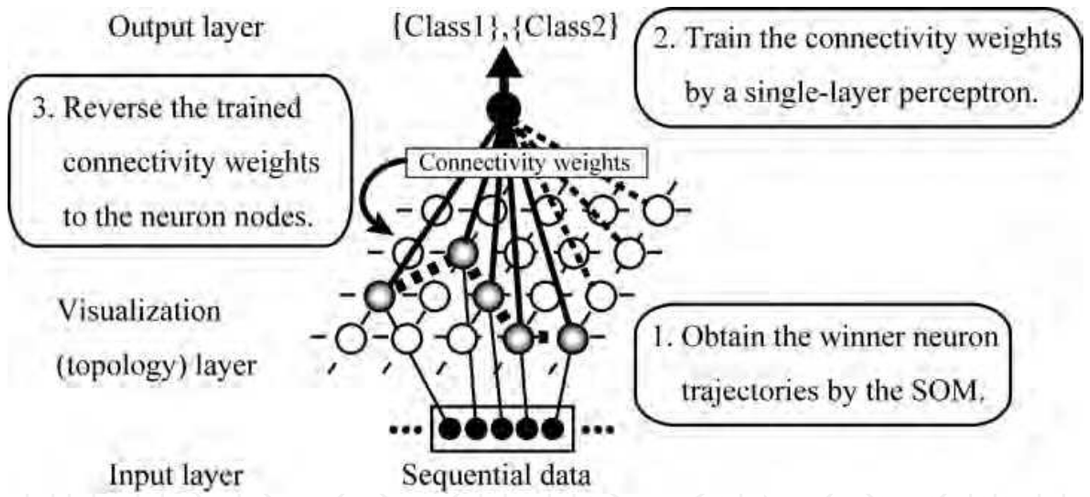

Fig. 9. Visualization architecture for two classes of sequential data using the class label.

Let k be an index of K sequential data, and let n_k be the number of elements in the k^{th} sequential data. The k^{th} sequential data is represented by $\left\{\left(x_r^{(k)}, t_r^{(k)}\right): r = 1, \cdots n_k\right\}$, where $\sum_{k=1}^{K} n_k = N$. The visualization architecture consists of three steps:

Steps of learning weights for visualization

Step 1. Train the reference vectors using the Sequence-based SOM with $\left\{\left(x_r^{(k)}, t_r^{(k)}\right): r = 1, \cdots n_k, k = 1, \cdots, K\right\}$ as input data. For instance, the first sequence of the first datum $\left(x_1^{(1)}, t_1^{(1)}\right)$ is an input. Consequently, the winner neuron trajectories are obtained within the visualization layer.

Step 2. Construct a discriminant function by a single-layer perceptron using the class label as the winner neuron trajectories are input. Concretely, let an input vector be $Y_k = (y_{k,1}, \cdots, y_{k,M})$. Set $y_{k,m}$ corresponding to the m^{th} neuron node for the k^{th} sequence as follows:

$$y_{k,m} = \begin{cases} 1/n_k & if\ \exists r\, X_r^{(k)} \in C_m \\ 0 & otherwise \end{cases} \tag{6}$$

Let the connectivity weights be $W = (w_1, \cdots, w_M)$. The discriminant function is represented by $\hat{z}_k = W \cdot Y_k$. Let $z_k \in \{-1, 1\}$ be class labels. Obtain the optimal weights by minimizing the following objective function:

$$E = \sum_{k=1}^{K} \left(z_k - \hat{z}_k\right)^2. \tag{7}$$

When the objective function is minimized by the gradient decent method, connectivity weights W are updated by the following equation:

$$w_m^* := w_m + \gamma \sum_{k=1}^{K} \left(z_k - \hat{z}_k\right) y_{k,m}, \tag{8}$$

where γ is the learning rate.

Step 3. Finally, the map is visualized by reversing the obtained weights into the neuron nodes and setting the weights as contracting density.

6. Experiment 2: Time Series of a Medical Dataset

6.1 Dataset and Problem Setting

We used time series of blood test data of 137 hepatitis C patients for one year collected from a Japanese hospital (the medical faculty of Chiba University). We used 11 common inspection items indicating hepatic functions, namely, GOT, GPT, TP, ALB, T-BIL, D-BIL, I-BIL, TTT, ZTT, CHE, and T-CHO. The interval of inspection was approximately one month but varied by patient.

In recent years, interferon (IFN) has been used to cure hepatitis. However, IFN is expensive and has strong side effects. Moreover, the average recovery rate of hepatitis C patients on IFN is only approximately 30%. Therefore, it is crucial to predict the effectiveness of IFN based on the inspection results before IFN is administered. This reduces the physical, mental, and cost burdens on the patient. The data is classified a priori into two classes,

namely, 55 positive examples and 82 negative examples in terms of existence of the hepatitis virus through HCV-RNA inspection before and after IFN administration.
The objective of this experiment is to construct a map that suggests periods of patients group which gives the same trends in attributes under the classes of IFN's effectiveness. The results are validated by comparison with the results obtained by a simple display method given by the ratio of positive and negative examples in each cluster.

6.2 Pre-processing

In order to avoid bias among attributes (inspections), the attribute values were standardized based on the index presented by a doctor. All of the attribute values were standardized in continuous value from 1 to 6. The actual value that is in the range of normal state presented by a doctor was set as 3. Also extremely high and low values were cut off. At the same time, since the inspection intervals are different in each patient, the discretized points are normalized by linear interpolation using one-week intervals. Since the inspection intervals are approximately one month, this interpolation interval is sufficient.

6.3 Visualization Results

In this section, two display methods are compared using the results obtained by the Sequence-based SOM (regular grid: 52x15), i.e., using the same winner neuron trajectories. Figure 10 show the visualization obtained by positive-negative ratio based on Sequence-based SOM, which is obtained by:

$$w_i = \frac{NP_i}{\left(NP_i + NN_i\right)}, \tag{9}$$

where NP_i is the number of positive samples associated with node i and NN_i is the number of negative samples. Figure 11 show the visualization obtained by the single-layer perceptron based on the Sequence-based SOM result. In both maps, the horizontal axis indicates the number days before IFN administration, where the right side represents the administration day. This map indicates that if the winner neuron trajectory passes through high-weight nodes, the period may be related to IFN effectiveness. In addition, an expert can refer to the results for other patients in the same cluster in order to investigate the reason for the findings.
In Fig. 10, since some clusters in 30 weeks before the administration have only a few samples, their node weights tend to zero or one, as indicated by the ratio display method. Therefore, there is a bias to clusters with a small number of samples. This may be mitigated by adding confidence to nodes according to the number of samples. The white nodes connected by white lines indicate an example of positive patient trajectory. In contrast, there is no such bias by the connectivity weights display method. The node values appear to be well balanced as they spread across the entire period and show a larger positive region than the map obtained by the positive-negative ratio.

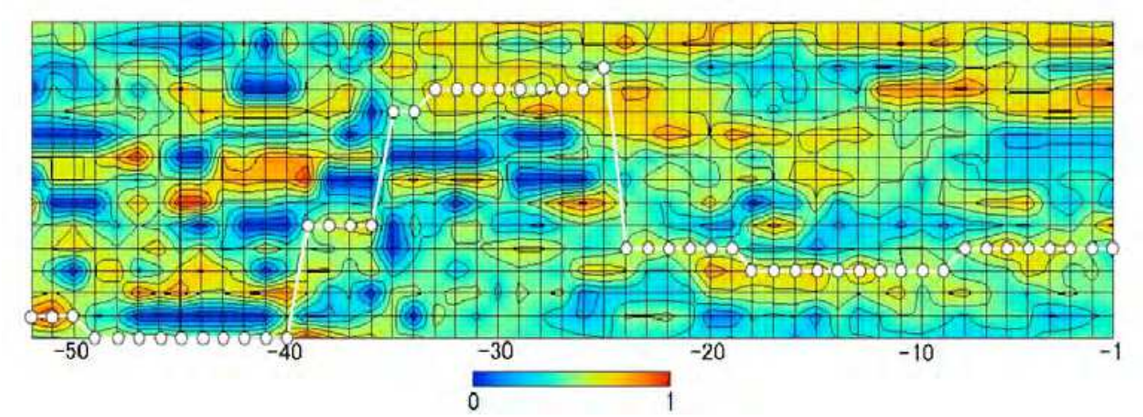

Fig. 10 Visualization by positive-negative ratio based on the Sequence-based SOM result.

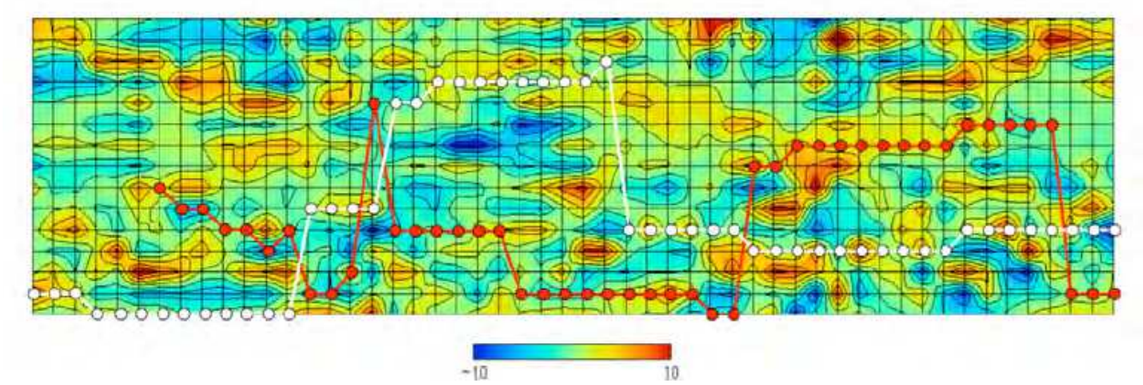

Fig. 11 Visualization by weights obtained by perceptron based on the Sequence-based SOM result.

6.4 Classification Accuracy

We investigated the node values that indicate correct classification, and found that as the weight increased, the correlation with the effectiveness of the IFN also increased. We evaluated the classification performance based on the winner neuron trajectory. In the case of the ratio display method, if the average of the nodes that winner neuron trajectory passes through is greater than 0.5, then the prediction is positive. In the case of the connectivity weights display method, the discriminant function is used as a criterion.

The results of a 10-fold cross validation are shown in Table 3. The average of the maximum and minimum percentages of correct answers obtained by the ratio display method is 47.53% for the test set. The result is equivalent to having a random scheme. Therefore, classification cannot be generalized to predict the effectiveness for an unknown patient. On the other hand, the proposed method provides almost perfect prediction for known data and 64.24% prediction accuracy for unknown patients, albeit with a large variance. The results seem to overfit the training set. Therefore, other methods of estimating appropriate weights, such as maximum entropy, may be considered in the future.

Display method	Positive-negative ratio	Connectivity weights
Training Set	79.78 ± 4.78%	99.60± 0.40%
Test Set	47.53±16.76%	64.29±21.43%

Table 3. Prediction accuracy (10-fold cross validation).

7. Conclusion

In this chapter, we introduced the extension of the SOM to visualize the change of dynamic clusters, movement, range, and merger/separation. The sequence weight function was introduced to the neuron topology in order to introduce temporal order to the topology. Using the proposed Sequence-based SOM, an experiment with news articles revealed transition of topics, level of interest, derivation, and diversification/convergence. In addition, class labels were used to obtain weights in order to display the SbSOM results. We applied this display method to time series of medical data. However, further study is needed about overfitting against the training set.

Acknowledgement

The present study was supported by the Materials Science & Technology Research Center for Industrial Creation (MSTeC) and Kansai Research Foundation for technology promotion (08R010), in Japan.

8. References

Allan, J. (2002). Topic Detection and Tracking: Event-Based Information Organization, Springer, ISBN: 978-0792376644

Andras, P. (2002). Kernel-Kohonen networks, International Journal of Neural Systems, Vol.12, pp.117-135

Barreto, G.A. & Arajo, A.F.R. (2001). Time in Self-Organizing Maps: An Overview of Models, International journal of Computer Research, Special Issue on Neural Networks, Vol.10, No.2. pp.139-179

Belkin, M. & Niyogi, P. (2002). Laplacian eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Processing Systems (NIPS14), pp.585-591

Bishop, C.M. (2006). Pattern Recognition and Machine Learning, Springer-Verlag, ISBN: 978-0387310732

Boulet, R.; Jouve, B.; Rossi, F. & Villa, N. (2008). Batch Kernel SOM and Related Laplacian Methods for Social Network Analysis, Neurocomputing, Vol.71, pp.1257-1273

Fritzke, B. (1994). Growing cell structures: A selforganizing networks for unsupervised and supervised learning, Neural Networks, Vol.7, pp.1441-1460

Fukui, K.; Saito, K.; Kimura, M. & Numao, M. (2006). Visualization Architecture Based on SOM for Two-Class Sequential Data, Lecture Notes in Artificial Intelligence, Vol.4252, pp.929-936, Springer, ISBN: 978-3-540-46537-9

Fukui, K.; Saito, K.; Kimura, M. & Numao, M. (2007). Combining Burst Extraction Method and Sequence-based SOM for Evaluation of Fracture Dynamics in Solid Oxide Fuel

Cell, Proc. of The 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Vol.2, pp.193-196

Fukui, K.; Saito, K.; Kimura, M. & Numao, M. (2008). Sequence-based SOM: Visualizing Transition of Dynamic Clusters, Proc. of IEEE 8th International Conference on Computer & Information Technology (CIT), pp.47-52

Hagenbuchner, M & Tsoi, A.C. (2004). A Supervised Self-Organizing map for Structures, Proceedings IEEE International Joint Conference on Neural Networks (IJCNN), 1923-1928

Ishikawa, Y. & Hasegawa, M. (2007). T-Scroll : Visualizing Trends in a Time-series of Documents for Interactive User Exploration, Lecture Note in Computer Science, Vol.4675, pp.235-246, ISBN: 978-3540748502

Jain, A.K.; Murty, M.N. & Flynn, P.J. (1999). Data Clustering: A Review, ACM Computing Surveys, Vol.31, No.3, pp.264-322

Kimura, M.; Saito, K. & Ueda, N. (2005). Multinominal PCA for extracting major latent topics from document streams, Proceedings of 2005 International Joint Conference on Neural Networks, pp.238-243

Kohonen, T. (2000). Self-Organizing Maps, Springer-Verlag, ISBN: 978-3540679219

Lau, K.W.; Yin H. & Hubbaard, S. (2006). Kernel Self-Organising Maps for Classification, Neurocomputing, Vol.69, pp.2033-2040

Levene, M. & Poulovassilis, A. (2004). Web Dynamics: Adapting To Change In Content, Size, Topology And Use, Springer, ISBN: 978-3540406761

Oian, T.; Srivastava, J.; Peng, Z. & Sheu, P.C.Y. (2009). Simultaneously Finding Fundamental Articles and New Topics Using a Community Tracking Method, PAKDD2009, Lecture Notes in Artificial Intelligence, Vol.5476, pp.796-803

Rossi, F. (2006). Visualization methods for metric studies. Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting

Roweis, S. & Saul, L. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding, Science, pp.2323-2326

Swan, R. & Jensen, D. (2000). TimeMines: Constructing Timelines with Statistical Models of Word Usage, Proc. of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.73-80

Tenenbaum, J.B.; Silva, V. de & Langford, J.C. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, pp.2319-2323

Wang, X. & McCallum, A. (2006). Topics over Time: A Non-Markov Continuous Time Model of Topical Trends, Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Self-Organizing Maps

Edited by George K Matsopoulos

ISBN 978-953-307-074-2
Hard cover, 430 pages
Publisher InTech
Published online 01, April, 2010
Published in print edition April, 2010

The Self-Organizing Map (SOM) is a neural network algorithm, which uses a competitive learning technique to train itself in an unsupervised manner. SOMs are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space and they have been used to create an ordered representation of multi-dimensional data which simplifies complexity and reveals meaningful relationships. Prof. T. Kohonen in the early 1980s first established the relevant theory and explored possible applications of SOMs. Since then, a number of theoretical and practical applications of SOMs have been reported including clustering, prediction, data representation, classification, visualization, etc. This book was prompted by the desire to bring together some of the more recent theoretical and practical developments on SOMs and to provide the background for future developments in promising directions. The book comprises of 25 Chapters which can be categorized into three broad areas: methodology, visualization and practical applications.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen