2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

On the Link of Data Synchronization to Self-Organizing Map Algorithm

Takaya Miyano† and Takako Tsutsui‡

†Department of Micro System Technology, Ritsumeikan University,
1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, JAPAN
‡Department of Health and Social Services, National Institute of Public Health,
2-3-6 Minami, Wako, Saitama 351-0197, JAPAN
Email: tmiyano@se.ritsumei.ac.jp, tsutsui@niph.go.jp

Abstract—We have recently devised a method for feature extraction from multivariate data using an analogue of Kuramoto's dynamics for modeling collective synchronization in a network of coupled phase oscillators. In our method, which we call data synchronization, the phase oscillators carrying information in their natural and updated rhythms achieve partial synchronizations of the oscillators whose common rhythms are interpreted as the template vectors representing the general features of the data set. In this study, we discuss the link of data synchronization to the self-organizing map algorithm as a popular method for data mining and show through numerical experiments how data synchronization can fix the problem of the self-organizing map algorithm on the initial setting of reference vectors.

1. Introduction

Finding patterns from data is one of the interesting and important applications of large-scale databases that grow to become indispensable information infrastructures of the society. To meet such social needs, a variety of mathematical methods for data mining have been developed. Among others, the self-organizing map (SOM) algorithm devised by Kohonen, which is currently one of the three major prototypes of artificial neural networks, is a powerful and popular method applicable to multivariate data [1, 2]. Unfortunately, the SOM algorithm has a bottleneck that reference vectors as candidates for general feature patterns to be extracted from given data must be provided at the initial stage of learning to initiate the learning process. Even if the reference vectors are generated by random selection from data, as is often performed in actual applications, the appropriate number of reference vectors has to be determined by some rules. In many practical cases, no such rules are available.

Recently, we have developed a method for spontaneous data clustering to perform feature extraction from multivariate data [3]–[5], on the basis of Kuramoto's model for collective synchronization in a network of coupled phase oscillators [6]–[8]. In our method, which we call data synchronization, phase oscillators carry given multivariate data in their natural rhythms and update their rhythms through nonlinear coupling between oscillators to achieve partial synchronizations of the oscillators. The common rhythms of the self-organized synchronous groups are interpreted as the template vectors that represent the general features of the data set. Data synchronization does not require reference vectors such that the SOM algorithm requires to initiate the learning process. In the previous work [4], we found that the equations governing data synchronization become equivalent to the competitive learning rule for SOM when they are linearized about partial synchronous solutions. This implies that the reference vectors for the SOM algorithm are spontaneously generated during the nonlinear regime of the dynamics governing data synchronization. Our findings might suggest that data synchronization can overcome the bottleneck of the SOM algorithm.

In this paper, we briefly summarize our recent findings about the link of data synchronization to the SOM algorithm and show through numerical experiments of data clustering how the bottleneck of the SOM algorithm can be overcome in data synchronization.

2. Self-organizing Map Algorithm

We briefly describe the main points of the competitive learning rule for the SOM algorithm. Given N sample vectors $\{\boldsymbol{x}_i = (x_{i1}, \cdots, x_{iD})\}_{i=1}^{N}$ with D degrees of freedom, we generate a set of M reference vectors with D degrees of freedom, $\{\boldsymbol{m}_i\}_{i=1}^{M}$, by the method mentioned later in this section. To update the reference vectors, we use a continuous batch learning rule of the form

$$\dot{\boldsymbol{m}}_i = \frac{1}{N_i}\sum_{j=1}^{N} h(\rho_{i,j})\,(\boldsymbol{x}_j - \boldsymbol{m}_i)\ , \quad (1)$$

$$= \frac{1}{N_i}\sum_{k=1}^{N_i} \kappa(t)\,\left(\boldsymbol{x}_{j(k)} - \boldsymbol{m}_i\right)\ , \quad (2)$$

$$\rho_{i,j} = | \boldsymbol{x}_j - \boldsymbol{m}_i | ,$$

where the overdot denotes differentiation with respect to dimensionless time, N_i is the number of nearest sample vectors $\boldsymbol{x}_{j(k)}$ $(k = 1, \cdots, N_i)$ to $\boldsymbol{m}_i$, and $h(\rho_{i,j})$ is the partitioning function defined as $h(\rho_{i,j}) = \kappa(t)$ if

$$\rho_{i,j} = \min_k(| \boldsymbol{x}_j - \boldsymbol{m}_k |)$$

with the adaptation gain $\kappa(t)$ $(0 < \kappa(t) < 1)$, and $h(\rho_{i,j}) = 0$ otherwise. The nearest reference vector to $\boldsymbol{x}_{j(k)}$ exclusively utilizes $\boldsymbol{x}_{j(k)}$ at a time to update itself. In the limit $N \to \infty$,

$$\frac{1}{N_i} \sum_{k=1}^{N_i} \kappa(t) \boldsymbol{x}_{j(k)} \to \kappa(t) \boldsymbol{X}_i , \quad (3)$$

where $\boldsymbol{X}_i$ is the mean vector over all sample vectors in the neighborhood of $\boldsymbol{m}_i$. Consequently,

$$\dot{\boldsymbol{m}}_i = \kappa(t) (\boldsymbol{X}_i - \boldsymbol{m}_i) . \quad (4)$$

Hence, this learning rule inherits the local mean-field characteristic of the data space.

The reference vectors can be set at the initial stage of the learning process using prior knowledge about the features to be extracted. If there is no such knowledge, random selection from the sample vectors suffices. In either case, however, there is no rule for determining the appropriate number of reference vectors. In this sense, the initial settings of the reference vectors are determined haphazardly. Equation (4) leads to the convergence of the reference vectors to certain template vectors representing the general features of the population ensemble of the learning data. As will be shown in the next section, it turns out that the competitive learning rule for SOM can be derived from the equations governing data synchronization as their linearized version about partial synchronous solutions.

3. Link of Data Synchronization to SOM Algorithm

In data synchronization, we use dynamics as an analogue of Kuramoto's dynamics for modeling collective synchronization in a network of coupled phase oscillators. For details, see [3, 4]. In our method, we assign the multivariate data, $\{\boldsymbol{x}_i\}_{i=1}^N$, to the natural frequencies of the phase oscillators subject to the following equations with $n = 1, \cdots, D$:

$$\dot{\theta}_{in} = x_{in} + \frac{K}{N_i} \sum_{j=1}^{N} H\left(\tilde{d}_{i,j}\right) \sin(\theta_{jn} - \theta_{in}) , \quad (5)$$

$$\tilde{d}_{i,j} = | \boldsymbol{x}_i - \boldsymbol{x}_j | ,$$

where K is a positive coupling constant, N_i is the number of neighboring vectors to $\boldsymbol{x}_i$ and θ_{in} is the nth entry of the phase vectors $\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{iD})$, whose initial values are given as random numbers. $\dot{\boldsymbol{\theta}}_i = \boldsymbol{\omega}_i$ represents the updated value of $\boldsymbol{x}_i$ at each instant in the time evolution. The partitioning function H determines the range of mutual interaction between phase oscillators. It is defined as $H\left(\tilde{d}_{i,j}\right) = 1$ if $\tilde{d}_{i,j} \leq \tilde{d}_0$ and $H\left(\tilde{d}_{i,j}\right) = 0$ otherwise. Here, $\tilde{d}_0 = \alpha \mid \boldsymbol{x}_i \mid$ with a positive constant α, which determines N_i neighboring vectors with which the phase vector $\boldsymbol{\theta}_i$ can interact. Thus, each oscillator conveys the original and updated data through its natural and adaptive rhythms, respectively. This enables the self-organization of data vectors.

The dynamics of Eq. (5) have a local mean-field characteristic that is associated with the statistical properties of the self-organized data in the limit $N \to \infty$. Let us define the local-order parameters and local mean-field phase vectors, $\boldsymbol{r}_i = (r_{i1}, \cdots, r_{iD})$ and $\boldsymbol{\psi}_i = (\psi_{i1}, \cdots, \psi_{iD})$, respectively, as

$$r_{in} \exp\left(\tilde{i}\psi_{in}\right) = \frac{1}{N_i} \sum_{j=1}^{N} H\left(\tilde{d}_{i,j}\right) \exp\left(\tilde{i}\theta_{jn}\right) , \quad (6)$$

where $\tilde{i} = \sqrt{-1}$. Thus, we can rewrite Eq. (5) as

$$\dot{\theta}_{in} = x_{in} + K r_{in} \sin(\psi_{in} - \theta_{in}) . \quad (7)$$

Let us consider the phase vectors belonging to the gth group, denoted as Γ_g, of G partially synchronized groups and denote each entry of a phase vector belonging to the gth group as $\theta_{i(g)n} = \psi_{gn} + \Delta\theta_{i(g)n}$ and its common frequency as $X_{gn} = \dot{\psi}_{gn}$, where $\boldsymbol{X}_g = (X_{g1}, \cdots, X_{gD})$ with g from 1 to G. Then, Eq. (7) becomes

$$\Delta\dot{\theta}_{i(g)n} = x_{i(g)n} - X_{gn} - K r_{gn} \sin\left(\Delta\theta_{i(g)n}\right) . \quad (8)$$

Accordingly, the frequency vectors will converge to the true mean frequency vector in the partially synchronized group in the limit $N \to \infty$. In this way, the multivariate data undergo spontaneous grouping. The common frequency vectors $\boldsymbol{X}_g$ are interpreted as the template vectors that represent the general features of the learning data.

From the necessary condition for Eq. (8) to have fixed points, the coupling constant K is set to a sufficiently large value. The coarse-graining parameter α that determines the resolution for discriminating one partially synchronized group from another is set so as to yield a sufficient number of neighbors in the neighborhood of $\boldsymbol{x}_i$. In fact, the vertical angle ϕ between $\boldsymbol{x}_i$ and its outermost neighbor is defined as $\phi = \sin^{-1}\left(\tilde{d}_0 / \mid \boldsymbol{x}_i \mid\right) = \sin^{-1}\alpha$.

To establish the link between data synchronization and the SOM algorithm, we rewrite Eq. (7) as

$$\omega_{in} = x_{in} + K r_{gn} \sin(\psi_{gn} - \theta_{in}) , \quad (9)$$

where $i \in \Gamma_g$. Equation (9) is linearized about $\psi_{gn} - \theta_{in} = 0$, i.e., the partially synchronized state:

$$\omega_{in} \approx x_{in} + Kr_{gn}(\psi_{gn} - \theta_{in}) \ . \tag{10}$$

Taking the derivative of both sides of Eq. (10) with respect to time, we obtain

$$\dot{\omega}_{in} = Kr_{gn} \left(\Omega_{gn} - \omega_{in}\right) \ . \tag{11}$$

Here, $\boldsymbol{r}_g$ is assumed to change slowly with time and be approximately constant near $\psi_{gn} - \theta_{in} = 0$ and $\boldsymbol{\Omega}_g = \dot{\boldsymbol{\psi}}_g$. By the substitutions of

$$\begin{aligned} \boldsymbol{\omega}_i &= \boldsymbol{m}_i \ , \\ Kr_{gn} &= \kappa(t) = \kappa = \text{constant} \ , \\ \boldsymbol{\Omega}_g &= \boldsymbol{X}_g \ , \end{aligned}$$

we can reproduce Eq. (2) from Eq. (11). Thus, the competitive learning rule for SOM is shown to be a linearized version of the dynamics governing data synchronization about partially synchronized states.

4. Numerical Experiments

We conducted numerical experiments of data clustering for multivariate data with three degrees of freedom ($D = 3$). In these experiments, we supposed three groups to each of which five data vectors should belong, given as $\boldsymbol{x}_i = (1+\epsilon, \epsilon, \epsilon)$, $(\epsilon, 1+\epsilon, \epsilon)$ or $(\epsilon, \epsilon, 1+\epsilon)$ with Gaussian random numbers ϵ of mean 0 and variance 0.1. These groups were labeled as group 1, group 2 and group 3 for convenience, represented by the template vectors $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, respectively. The magnitude of the variance in ϵ, i.e., the diversity in the data vectors of each group implies that particular settings of the coupling constant $K = 0.5$ and the adaptation gain $\kappa(t) = 0.5$ are sufficient for data synchronization as well as the SOM algorithm to achieve data clustering.

In a first experiment, we ran the dynamics of Eq. (2) at a time width of 0.05 with $\kappa(t) = 0.5$ and observed how data clustering was performed by the SOM algorithm. In this experiment, each of three reference vectors was selected from groups 1, 2 and 3, respectively. This represents an appropriate initial setting of the reference vectors. Results are shown in Fig. 1. Owing to the appropriate selection of the reference vectors, the template vector of each group was correctly generated.

In a second experiment, we again ran the dynamics of Eq. (2) at a time width of 0.05 with $\kappa(t) = 0.5$. In this experiment, two reference vectors were selected from group 1 and one vector from group 3. This represents an inappropriate setting of the reference vectors. Results are shown in Fig. 2. Although the template vector of group 3 was correctly generated, those of groups 1 and 2 were not correctly generated due to the inappropriate selection of the reference vectors.

In a third experiment, we ran the dynamics of Eq. (5) at a time width of 0.05 with $K = 0.5$ and $\alpha = 0.3$, 0.5, 0.7 and 1, and observed how data clustering was performed by data synchronization. This experiment was free from the initial setting of the reference vectors, since data synchronization requires no reference vectors to initiate the learning process. Instead, the initial values of $\boldsymbol{\theta}_i$ were set to be Gaussian random numbers of mean 0 and variance 1. Results for $\alpha = 0.5$ are shown in Fig. 3. We obtained similar results for $\alpha = 0.3$, 0.7 and 1 to those shown in Fig. 3. Despite different settings of α, the template vector of each group was correctly generated by data synchronization as was performed using the SOM algorithm with the appropriate setting of the reference vectors.

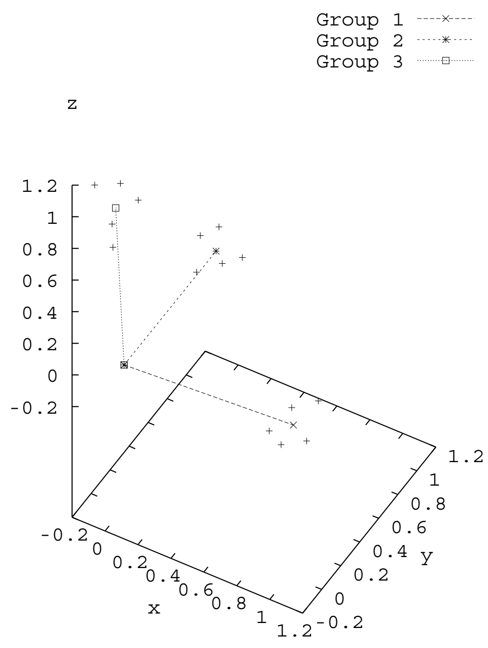

Figure 1: Data clustering of three-dimensional data vectors by SOM algorithm with appropriate selection of reference vectors. Learning data are shown by + and extracted feature vectors by three dotted lines.

5. Discussion and Conclusion

The present numerical experiments demonstrate the pathology of the SOM algorithm that feature extraction strongly depends on the initial setting of reference vectors. As has been shown in Fig. 2, an inappropriate setting of reference vectors leads to the generation

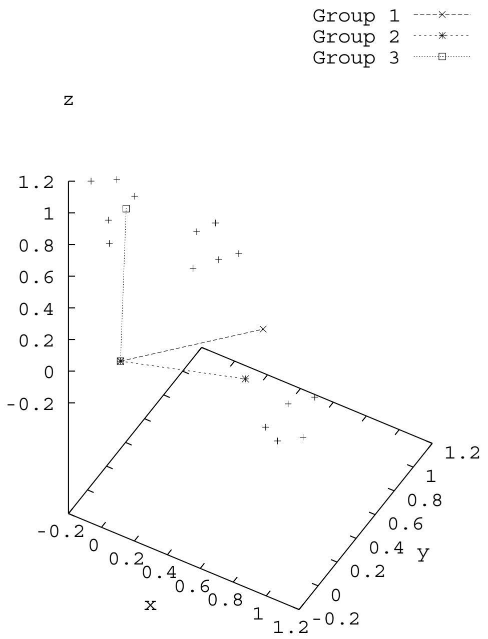

Figure 2: Data clustering of three-dimensional data vectors by SOM algorithm with inappropriate selection of reference vectors. Learning data are shown by + and extracted feature vectors by three dotted lines.

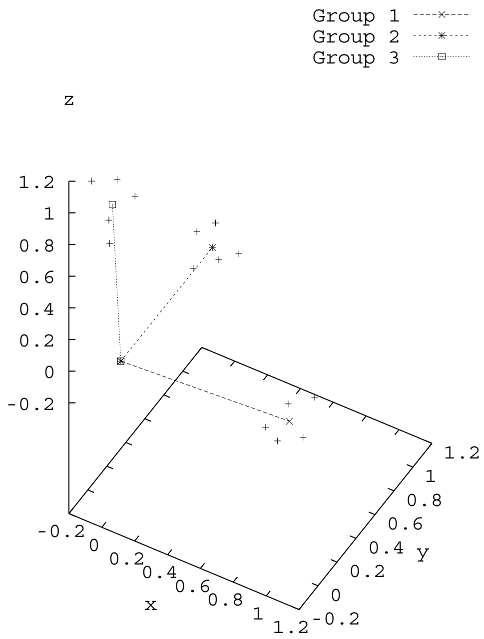

Figure 3: Data clustering of three-dimensional data vectors by data synchronization with $\alpha = 0.5$. Learning data are shown by + and extracted feature vectors by three dotted lines.

of incorrect template vectors, despite an appropriate setting of the number of the reference vectors. Such a situation is likely to occur in practical applications of the SOM algorithm when no prior information about the feature patterns to be extracted is available. The disadvantages of the SOM algorithm can be overcome by data synchronization. Although data synchronization requires appropriate settings of the coupling constant K and coarse-graining level α, it can be insensitive to α due to particular statistical distribution of data and thus can generate correct template vectors regardless of the setting of α, as has been shown in Fig. 3. In general, the setting of α is crucial in data synchronization, since it determines the resolution for discriminating one synchronous cluster of data vectors from another. Nevertheless, we have no general and systematic method for optimizing α. Developing such a method is an issue of interest that is worth investigating in a future work.

References

[1] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol.43, pp.59–69, 1982.

[2] T. Kohonen, The self-organizing map, *Proc. IEEE*, vol.78, pp.1464–1480, 1990.

[3] T. Miyano and T. Tsutsui, "Data synchronization in a network of coupled phase oscillators," *Phys. Rev. Lett.*, vol.98, pp.024102-1–024102-4, 2007.

[4] T. Miyano and T. Tsutsui, "Collective synchronization as a method of learning and generalization from sparse data," *Phys. Rev.* E, vol.77, pp.026112-1–026112-11, 2008.

[5] T. Miyano and T. Tsutsui, "Finding major patterns of aging process by data synchronization," *IEICE Trans. Fundamentals*, in print.

[6] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, Springer, New York, 1984.

[7] S. H. Strogatz, "From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators," *Physica* D, vol.143, pp.1–20, 2000.

[8] J. A. Acebrón, L. L. Bonilla, C. J. P. Vincente, F. Rotort, and R. Spigler, "The Kuramoto model: A simple paradigm for synchronization phenomena," *Rev. Mod. Phys.*, vol.77, pp.137–185, 2005.