Data Synchronization in a Network of Coupled Phase Oscillators

Takaya Miyano*

Department of Micro System Technology, Ritsumeikan University 1-1-1, Noji-higashi, Kusatsu, Shiga 525-8577, Japan

Takako Tsutsui†,‡

National Institute of Public Health 2-3-6, Minami, Wako, Saitama 351-0197, Japan

(Received 29 May 2006; published 11 January 2007)

We devised a new method of data mining for a large-scale database. In the method, a network of locally coupled phase oscillators subject to Kuramoto's model substitutes for given multivariate data to generate major features through phase locking of the oscillators, i.e., phase transition of the data set. We applied the method to the national database of care needs certification for the Japanese public long-term care insurance program, and found three major patterns in the aging process of the frail elderly. This work revealed the latent utility of Kuramoto's model for data processing.

DOI: 10.1103/PhysRevLett.98.024102 PACS numbers: 05.45.Xt, 89.20.−a

In a society of mutual communication, individuals usually look for a consensus, despite their differences, and opinions spontaneously converge into a few representative ideas referred to as public opinion. This common phenomenon residing in democratic societies might reflect synchronization of individual neuronal entities with degrees of freedom. Inspired by this conjecture and recognizing the social needs for using public databases of an explosively growing scale thanks to recent progress in computer science and information technology, we devised a new method of data mining based on spontaneous data clustering for a large-scale database. In the method, a network of locally coupled limit-cycle phase oscillators subject to an analogue of the Kuramoto model [1–4] substitutes for a set of multivariate data by encoding the data vectors into the natural frequencies, yielding instantaneous renewal of the data represented by the time derivatives of the phase vectors. Local phase locking of the network generates a few common frequency vectors that represent major features of the data set. Information is represented and processed by the oscillator's rhythms. This might be reminiscent of a version of a hypothesis for temporal coding in synchronous electrical activity of neurons [5–13], although the present method is not concerned with neuroscience but involves phase transition in a large population of data. Our method requires no initial templates to generate patterns from data, unlike existing methods such as self-organizing mapping [14]. Rather, the method is based on the expectation that prototypical patterns lurk in data themselves. We applied the method to the national database for the Japanese public long-term care insurance program, finding three major classes to categorize aging status of the frail elderly.

Collective synchronization is the phenomenon that a group of events spontaneously comes into occurrence in unison with a common rhythm, despite differences in the individual rhythms of the events, actually emerging in many real-world networked systems. Following the pioneering work by Winfree [15], Kuramoto established a firm foundation of the physics underlying collective synchronization with a network of coupled limit-cycle phase oscillators as a comprehensive and mathematically tractable model to unravel the intriguing machinery [1]. As appreciated by Mirollo and Strogatz [3], the Kuramoto model was a breakthrough in nonlinear science. For instance, the Kuramoto model has been applied to neuroscience to provide new insights into the binding problem in vision, i.e., the linking of sensory input across multiple receptive fields, and information processing based on pulsatile electrical activity of neurons [6,7,11,12]. This paper focused on data clustering in a network of coupled phase oscillators. Nevertheless, the present study might share with previous literature on neuroscience the issue of how information codes are processed in a synchronous dynamic system.

Our aim is expressed as "let data find patterns by themselves without any prior knowledge." Given N multivariate data points with D degrees of freedom, $\vec{x}_i = (x_i(1), \ldots, x_i(D))$ $(i = 1, \ldots, N)$, we devise, as an analogue of the Kuramoto model, a network of coupled phase oscillators to whose natural frequencies the data vectors $\vec{x}_i$ are encoded:

$$\frac{d\theta_i(n)}{dt} = x_i(n) + \frac{K}{N_i}\sum_{j=1}^{N} H(\tilde{d}_{i,j})\sin(\theta_j(n) - \theta_i(n)). \quad (1)$$

Here, K is a positive coupling constant and $\theta_i(n)$ is the nth component of the phase vector $\vec{\theta}_i = (\theta_i(1), \ldots, \theta_i(D))$ whose initial values can be set to random numbers. The derivatives of $\vec{\theta}_i$ with respect to time are instantaneous "frequency" vectors representing renewals of $\vec{x}_i$. We design the nonlinear interaction so as to work only between neighboring phase vectors. The neighborhood of $\vec{\theta}_i$ is defined by a partitioning function H. Let us express the distance as $\tilde{d}_{i,j} = |\vec{x}_i - \vec{x}_j|$. We define the partitioning function as $H(\tilde{d}_{i,j}) = 1$ if $\tilde{d}_{i,j} \le \tilde{d}_0$ and $H(\tilde{d}_{i,j}) = 0$, otherwise, with $\tilde{d}_0 = \alpha|\vec{x}_i|$ where α is an appropriate positive constant. The partitioning function acts as a supervisor that instructs the counterpart to couple, when viewing the time evolution by the governing equations as learning processes. Thus N_i neighboring phase vectors of $\vec{\theta}_i$ are fixed by the statistical distribution of $\vec{x}_i$, as shown in Fig. 1. The interaction occurs within a distance of $\tilde{d}_0$ that might be likened to tolerance of an individual in persuading others having different opinions.

To show that the dynamics inherit the local mean field character, we define a local order parameter as

$$r_i(n)\exp(i\psi_i(n)) = \frac{1}{N_i}\sum_{j=1}^{N} H(\tilde{d}_{i,j})\exp(i\theta_j(n)), \qquad (2)$$

where $\vec{r}_i = (r_i(1), \ldots, r_i(D))$ measures degrees of local coherence and $\vec{\psi}_i = (\psi_i(1), \ldots, \psi_i(D))$ is the local mean field. Then Eq. (1) reads $d\theta_i(n)/dt = x_i(n) + Kr_i(n) \times \sin[\psi_i(n) - \theta_i(n)]$. Under appropriate settings of K and α that guarantee $Kr_i(n) \geq |x_i(n) - X_g(n)|$ and a sufficient number of neighbors in the neighborhood of $\vec{x}_i$, the phase vectors will come into local synchrony to develop major groups within which the members spontaneously lock to a common frequency vector $\vec{X}_g = (X_g(1), \ldots, X_g(D))$ ($g = 1, \ldots, Q$; Q is the number of groups). These processes are coined as data synchronization. We thus obtain template vectors representing major features of the data.

The degree of synchrony in data synchronization should be measured in terms of $r_i(n)$. In practical applications, however, the mean diversity of frequency vectors over synchronized clusters, denoted by σ, may be more convenient:

$$\sigma = \frac{1}{N}\sum_{i=1}^{N}\sigma_i = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N_i}\sum_{j=1}^{N} H(\tilde{d}_{i,j})\frac{d_{i,j}}{\tilde{d}_0}\right], \qquad (3)$$

where $d_{i,j} = |\dot{\vec{\theta}}_i - \dot{\vec{\theta}}_j|$. If $N_i = 0$, σ_i is defined to be zero. As perfect synchronization is achieved, $\sigma \to 0$. In the opposite extreme where every $\vec{x}_i$ initially has no neighbors within a distance of $\tilde{d}_0$, the mean diversity σ will keep taking zero from the beginning of the time evolution of the governing equations. We may be able to avert such perfect desynchronization by increasing α.

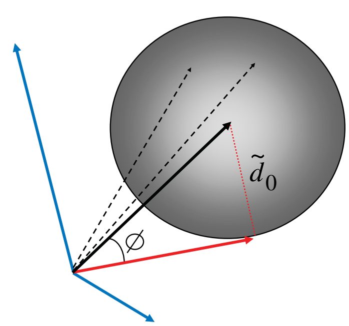

FIG. 1 (color). Neighborhood (gray sphere) of $\vec{x}_i$ (black solid arrow). Black dashed arrows are neighbors. Blue arrows are not neighbors. The vertical angle ϕ between $\vec{x}_i$ and its outermost neighbor (red arrow) is defined as $\phi = \sin^{-1}(\tilde{d}_0/|\vec{x}_i|)$. When $\alpha = 0.7$, $\phi \approx 44.4°$.

We conducted a preliminary numerical experiment for data clustering of multivariate data of 3 degrees of freedom ($D = 3$). In this experiment, we supposed three groups to each of which five data vectors should belong, given as $\vec{x}_i = (1 + \epsilon, \epsilon, \epsilon)$, $(\epsilon,\ 1 + \epsilon,\ \epsilon)$ or $(\epsilon,\ \epsilon,\ 1 + \epsilon)$ with Gaussian random numbers ϵ of mean 0 and variance 0.1. The initial values of $\vec{\theta}_i$ were set to Gaussian random numbers of mean 0 and variance 1. We ran the dynamics of Eq. (1) with $K = 0.4$ and $\alpha = 0.5$ at a time width of 0.05 to achieve perfect synchronization in each group. The member vectors correctly converged to (1, 0, 0), (0, 1, 0), or (0, 0, 1) and σ decreased from 1.01 to 0 in 850 time steps.

We next conducted a case study with a large database of care needs certification in order to examine typical patterns of aging status in the frail elderly. The history of the database is briefly described below. Under the circumstances of rapid demographic aging with more and more frail elderly seeking care, Japan implemented a mandatory social long-term care insurance system in 2000 [16,17]. In this system, a client aged 65 or older who needs nursing care services is given a basic questionnaire of 73 categories to assess his or her health status and quality of life (Table I). Since the start of the insurance system, questionnaire answers (multivariate data) have accumulated in the Japanese national database, which currently consists of about 25×10^6 cases. We used 12 sets of 2000 samples randomly selected from the database. The samples yielded the data vectors $\vec{x}_i$ with 73 degrees of freedom, i.e., $D = 73$. Such immense dimensionality as well as the absence of prior knowledge about the groups into which the multivariate data would be categorized made it difficult for

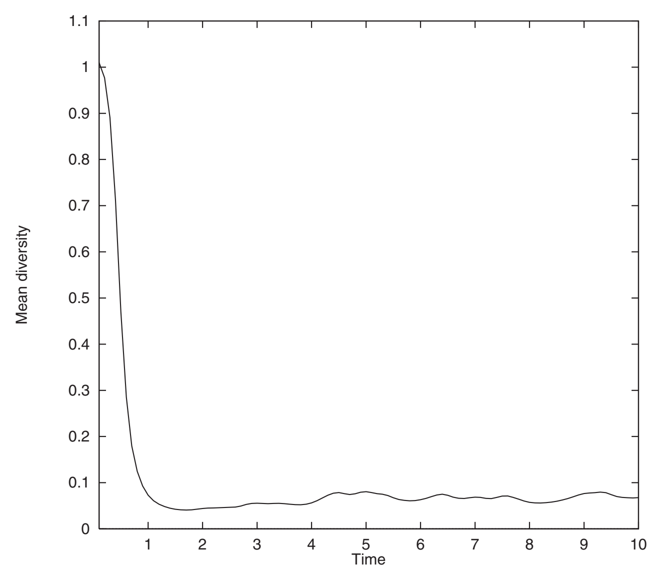

FIG. 2. Mean diversity σ as a function of time.

TABLE I. Three major patterns of aging status in frail elderly. The first column is the basic questionnaire in relation to 73 categories of disability. The second column is integer scores for answers. An integer score of "−1" means "healthy," while integer scores of "1" through the maximum represent increasing degrees of seriousness in each category of disability. The third, fourth, and fifth columns are estimated real scores defining the three major classes (classes 1, 2, and 3).

Questionnaire	Score	Class 1	Class 2	Class 3
1. Paralysis (left arm)	−1, 1	−0.9	−0.9	−1
2. Paralysis (right arm)	−1, 1	−0.9	−0.9	−0.9
3. Paralysis (left leg)	−1, 1	−0.2	0	0.2
4. Paralysis (right leg)	−1, 1	−0.2	−0.2	−0.4
5. Paralysis (other part of body)	−1, 1	−0.8	−0.8	−0.9
6. Contracture (shoulder joints)	−1, 1	−0.8	−0.8	−0.7
7. Contracture (elbow joints)	−1, 1	−0.9	−0.9	−0.8
8. Contracture (hip joints)	−1, 1	−0.8	−0.8	−0.8
9. Contracture (knee joints)	−1, 1	−0.3	−0.3	−0.3
10. Contracture (ankle joints)	−1, 1	−0.8	−0.9	−1
11. Contracture (other part of body)	−1, 1	−0.6	−0.6	−0.6
12. Rolling over in bed	−1, 1, 2	−0.5	0.6	0.4
13. Sitting up in bed	−1, 1, 2	0.1	0.3	0.4
14. Sitting with both feet on floor	−1, 1, 2, 3	−0.7	0.4	0.6
15. Sitting without feet on floor	−1, 1, 2, 3	0	0.5	0.5
16. Standing on both feet	−1, 1, 2	−0.8	0.3	0.5
17. Walking	−1, 1, 2	0.6	0.6	0.6
18. Transferring	−1, 1, 2, 3	0.6	0.7	0.7
19. Standing up from sitting position	−1, 1, 2	−0.6	1.5	1
20. Standing on 1 ft	−1, 1, 2	−0.9	−0.9	−0.8
21. Getting in and out bath	−1, 1, 2, 3	−0.7	−0.7	−0.6
22. Bathing	−1, 1, 2, 3	−0.8	−0.8	−0.9
23. Bedsore (decubitus ulcer)	−1, 1	−0.9	−0.5	0.6
24. Other skin diseases	−1, 1	−0.9	−0.1	0
25. Lifting one arm to the chest	−1, 1, 2	−0.9	−0.3	0.3
26. Swallowing	−1, 1, 2	−0.9	−0.2	1.4
27. Desire to urinate	−1, 1, 2	−0.3	1.8	1.3
28. Desire to defecate	−1, 1, 2	−0.9	1.3	1.4
29. Management after urination	−1, 1, 2, 3	−0.9	1.7	1.7
30. Management after defecation	−1, 1, 2, 3	−0.7	0.1	0.6
31. Taking meals (dietary intake)	−1, 1, 2, 3	−0.5	0.7	0.7
32. Oral hygiene (tooth brushing)	−1, 1, 2	−0.6	−0.6	−0.7
33. Face washing	−1, 1, 2	−0.3	−0.3	−0.4
34. Hair care	−1, 1, 2	−0.9	−0.8	−0.2
35. Nail cutting	−1, 1, 2	−0.9	−0.9	−1
36. Buttoning and unbuttoning clothing	−1, 1, 2, 3	−0.9	−0.7	0.1
37. Putting on and taking off a jacket	−1, 1, 2, 3	−0.9	−0.9	−0.9
38. Putting on and taking off trousers	−1, 1, 2, 3	−0.9	−0.9	−0.4
39. Putting on and taking off socks	−1, 1, 2, 3	−0.9	−0.9	−0.8
40. Cleaning rooms	−1, 1, 2	−0.9	−0.9	−0.8
41. Taking medication	−1, 1, 2	−0.9	−0.9	−0.9
42. Financial management	−1, 1, 2	−0.9	−0.9	−0.7
43. Serious memory loss	−1, 1, 2	−0.9	−0.9	−0.9
44. Loss of interest in circumstances	−1, 1, 2	−0.9	−0.9	−0.8
45. Visual acuity	−1, 1, 2, 3, 4	−0.9	−0.9	−0.8
46. Hearing	−1, 1, 2, 3, 4	−0.8	−0.8	−0.8
47. Mutual communication	−1, 1, 2, 3	−0.9	−0.9	−0.8
48. Response to instructions	−1, 1, 2	−0.8	−0.9	−0.8
49. Understanding a daily schedule	−1, 1	−0.9	−0.9	−0.8
50. Answering date of birth and age	−1, 1	−0.9	−0.9	−0.4
51. Short-term memory	−1, 1	−0.9	−0.9	−0.8
52. Remembering own name	−1, 1	−0.9	−0.9	−0.8
53. Recognition of current season	−1, 1	−0.9	−0.9	−0.8
54. Orientation in place	−1, 1	−0.9	−0.9	−0.8
55. Feeling persecuted	−1, 1, 2	−0.9	−0.9	−0.8
56. Fabricating stories	−1, 1, 2	−0.8	−0.9	−0.9
57. Visual or auditory hallucinations	−1, 1, 2	−0.9	−0.9	−0.8
58. Emotional instability	−1, 1, 2	−0.9	−0.9	−0.8
59. Reversion of day and night	−1, 1, 2	−0.9	−0.9	−0.8
60. Verbal or physical violence	−1, 1, 2	−0.4	−0.4	0.5
61. Repeating the same story	−1, 1, 2	−0.7	0.1	0.4
62. Shouting	−1, 1, 2	0	1.4	0.5
63. Resisting advice or care	−1, 1, 2	−0.5	1.2	1.3
64. Poriomania	−1, 1, 2	−0.9	−0.9	−0.9
65. Restlessness	−1, 1, 2	−0.9	−0.6	0.2
66. Being away from residence	−1, 1, 2	−0.9	−0.7	−0.3
67. Insisting on going out alone	−1, 1, 2	−0.8	0.6	1.8
68. Collecting mania	−1, 1, 2	−0.9	0.5	1.5
69. Inability to manage a fire	−1, 1, 2	−0.9	1.4	1.4
70. Destruction of things or clothes	−1, 1, 2	−0.9	1.5	1.7
71. Unsanitary behavior and living conditions	−1, 1, 2	0.7	1	1.2
72. Pica (consumption of nonnutritive substances)	−1, 1, 2	−0.9	−0.9	1.4
73. Troublesome sexual behavior	−1, 1, 2	−0.9	−0.9	−0.9

existing methods of data clustering like self-organization to work effectively. For each data set, we ran the dynamics of Eq. (1) with $K = 10$ and $\alpha = 0.7$ at a time width of 0.1. Figure 2 shows typical results of the mean diversity σ as a function of time. In 20 time steps, partial phase locking came about to generate three major groups. This tendency was similar in each data set. The results are summarized in Table I. The major groups (classes 1, 2, and 3 in Table I) comprised about 75% of the whole samples. The remaining 25% were out of synchrony or formed minor groups consisting of a few members.

The main features of class 1 are impairment in the legs and resultant functional limitations that affect activities of daily living. This class of disability may be due to long-term accumulation of mechanical load on the legs supporting the body weight. Such disability is likely to develop with aging in legged animals with higher intelligence. For class 3, besides progressive disability affecting the legs, its distinctive features are related to the brain, in particular, bowel or bladder control as well as mental disability concerning basic activities necessary to reside in the community. Interestingly, however, such mental disability, as associated with processing of sensory information such as vision and hearing, does not seem to be very important. Rather, deterioration of intellectual ability such as hypofrontality is noticeable, which is unlikely to come from long-term accumulation of working load in the brain. It is an interesting question as to whether or not such mental disability exists in other primates with higher intellectual abilities. Class 2 is a less progressive version of class 3, which appears to be midway between class 1 to class 3. The present data clustering may capture universal features of the aging process. We conjecture that there may be a major path in aging, starting from physical disability in the legs, represented by class 1 (doddering), through an intermediate status of class 2 (dotage), to the complication of mental disability of class 3 (senility) resulting in serious deterioration of quality of social life. The isolated vectors out of synchrony and the minor groups of vectors could have been classified into other possible major groups that would have come out if many more samples were able to be handled. They may represent the idiosyncratic status of aging to reflect complexity and variety in human aging processes.

In conclusion, we devised dynamics for a network of coupled phase oscillators that substitutes an ensemble of multivariate data. Extracting general features from the data set was performed by phase locking of the oscillators, i.e., phase transition in the ensemble of data. The present Letter suggests that collective synchronization as a physical process occurring in a bounded nonlinear system can play the role of data clustering in a process of learning and generalization from sparse multivariate data. Since the Kuramoto model is an outcome of the perturbation method, referred to as method I in [1], for reaction-diffusion systems, diffusive coupling between physical entities carrying particular information might be necessary to generalize the acquired information in living computational systems.

The authors thank Dr. Sadanori Higashino and Mr. Hitoshi Taniguchi for helpful advice and technical support. This Letter was partially supported by a grant from the Ministry of Health, Labour and Welfare.

*Electronic address: tmiyano@se.ritsumei.ac.jp
†Electronic address: tsutsui@niph.go.jp
‡Also at the Department of Public Health, Nihon University School of Medicine.

[1] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence* (Springer, New York, 1984).
[2] S. H. Strogatz, Physica Amsterdam) **143D**, 1 (2000).
[3] R. E. Mirollo and S. H. Strogatz, Physica (Amsterdam) **205D**, 249 (2005).
[4] J. A. Acebron, L. L. Bonilla, C. J. P. Vincente, F. Ritort, and R. Spigler, Rev. Mod. Phys. **77**, 137 (2005).
[5] C. M. Gray and W. Singer, Proc. Natl. Acad. Sci. U.S.A. **86**, 1698 (1989).
[6] Y. Kuramoto, Physica Amsterdam) **50D**, 15 (1991).
[7] H. Sompolinsky, D. Golomb, and D. Kleinfeld, Phys. Rev. A **43**, 6990 (1991).
[8] W. R. Softky and C. Koch, J. Neurosci. **13**, 334 (1993).
[9] J. J. Hopfield, Nature (London) **376**, 33 (1995).
[10] H. Fujii, H. Ito, K. Aihara, N. Ichinose, and M. Tsukada, Neural Networks **9**, 1303 (1996).
[11] P. Seliger, S. C. Young, and L. S. Tsimring, Phys. Rev. E **65**, 041906 (2002).
[12] H. Haken, Physica Amsterdam) **205D**, 1 (2005).
[13] R. Gutig and H. Sompolinsky, Nat. Neurosci. **9**, 420 (2006).
[14] T. Kohonen, Biol. Cybern. **43**, 59 (1982).
[15] A. T. Winfree, J. Theor. Biol. **16**, 15 (1967).
[16] T. Tsutsui and N. Muramatsu, Journal of the American Geriatrics Society **53**, 522 (2005).
[17] T. Miyano, T. Tsutsui, Y. Seki, S. Higashino, and H. Taniguchi, IEEE Trans. Inf. Technol. Biomed. **9**, 502 (2005).