

wise, with  $\tilde{d}_0 = \alpha|\tilde{x}_i|$  where  $\alpha$  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  $\tilde{\theta}_i$  are fixed by the statistical distribution of  $\tilde{x}_i$ , 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

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

where  $\vec{r}_i = (r_i(1), \dots, r_i(D))$  measures degrees of local coherence and  $\vec{\psi}_i = (\psi_i(1), \dots, \psi_i(D))$  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  $\alpha$  that guarantee  $Kr_i(n) \geq |x_i(n) - X_g(n)|$  and a sufficient number of neighbors in the neighborhood of  $\tilde{x}_i$ , the phase vectors will come into local synchrony to develop major groups within which the members spontaneously lock to a common frequency vector  $\vec{X}_g = (X_g(1), \dots, X_g(D))$  ( $g = 1, \dots, Q$ ;  $Q$  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  $\sigma$ , may be more convenient:

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

where  $d_{i,j} = |\tilde{\theta}_i - \tilde{\theta}_j|$ . If  $N_i = 0$ ,  $\sigma_i$  is defined to be zero. As perfect synchronization is achieved,  $\sigma \rightarrow 0$ . In the opposite extreme where every  $\tilde{x}_i$  initially has no neighbors

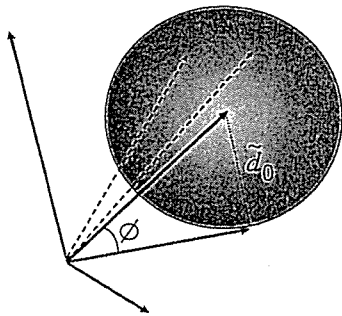


FIG. 1 (color). Neighborhood (gray sphere) of  $\tilde{x}_i$  (black solid arrow). Black dashed arrows are neighbors. Blue arrows are not neighbors. The vertical angle  $\phi$  between  $\tilde{x}_i$  and its outermost neighbor (red arrow) is defined as  $\phi = \sin^{-1}(\tilde{d}_0/|\tilde{x}_i|)$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

within a distance of  $\tilde{d}_0$ , the mean diversity  $\sigma$  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  $\alpha$ .

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  $\tilde{x}_i = (1 + \epsilon, \epsilon, \epsilon)$ ,  $(\epsilon, 1 + \epsilon, \epsilon)$  or  $(\epsilon, \epsilon, 1 + \epsilon)$  with Gaussian random numbers  $\epsilon$  of mean 0 and variance 0.1. The initial values of  $\tilde{\theta}_i$  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  $\sigma$  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  $25 \times 10^6$  cases. We used 12 sets of 2000 samples randomly selected from the database. The samples yielded the data vectors  $\tilde{x}_i$  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

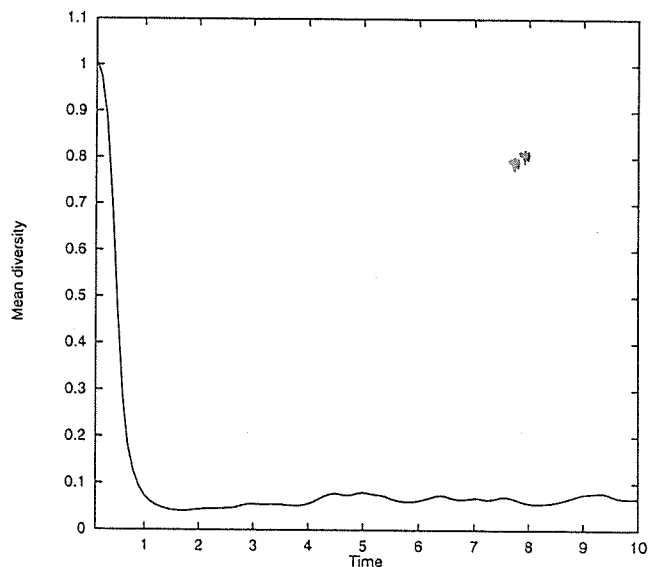


FIG. 2. Mean diversity  $\sigma$  as a function of time.

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  $\sigma$  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. Vicente, 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).

## 集団同期によるデータクラスタリング

宮野 尚哉<sup>†</sup> 筒井 孝子<sup>††</sup>

<sup>†</sup> 立命館大学理工学部 〒 525-8577 滋賀県草津市野路東 1-1-1

<sup>††</sup> 国立保健医療科学院 〒 060-8628 埼玉県和光市南 2-3-6

E-mail: <sup>†</sup>tmiyano@se.ritsumei.ac.jp, <sup>††</sup>tsutsui@niph.go.jp

あらまし 集団同期の蔵本モデルを応用したデータマイニング手法を提案する。位相振動子の自然周波数をベクトル量に拡張し、これらに多変量データを代入する。位相差の関数としての短距離相互作用を振動子間に導入し、ベクトルデータの仮想的な集団同期を実現する。この同期現象はデータ集団における相転移と解釈され、同期周波数を同期集団の代表値と見なすことによってデータクラスタリングが達成される。本手法を要介護保険の認定データに適用し、要介護高齢者の健康状態におけるパターンの抽出を行った。その結果、3つの典型的パターンが存在することが明らかとなった。

キーワード 集団同期, 位相振動子, データマイニング, 医療情報

## Data Clustering by Collective Synchronization in an Ensemble of Multivariate Data

Takaya MIYANO<sup>†</sup> and Takako TSUTSUI<sup>††</sup>

<sup>†</sup> Faculty of Science and Engineering, Ritsumeikan University 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

<sup>††</sup> National Institute of Public Health 2-3-6, Minami, Wako, Saitama 351-0197, Japan

E-mail: <sup>†</sup>tmiyano@se.ritsumei.ac.jp, <sup>††</sup>tsutsui@niph.go.jp

**Abstract** We propose a new method of data mining based on an analogue of the Kuramoto model. In the method, multivariate data are input to the natural frequencies of phase oscillators that form a local network through short-range interactions between phase vectors. Common frequencies of groups of phase oscillators attained by partial phase-locking are interpreted as major templates representing general features of the multivariate data set. This method was applied to the national database of care needs certification for the Japanese public long-term care insurance program. Data synchronization was achieved and generated three major patterns in the ageing process of the frail elderly.

**Key words** Collective Synchronization, Phase Oscillator, Data Mining, Medical Information

### 1. はじめに

今日までの半世紀にわたるコンピュータ科学と情報工学の驚異的な進歩によって、様々な情報がデータベース上にデジタルデータとして保存、蓄積され、ネットワーク化された通信インフラを介して随時利用できるような社会が実現しつつある。インターネットの普及は情報化社会の発展を加速している。最近では、無線通信機能をもつセンサシステムを環境に展開して、環境情報を遠隔地から収集することを可能にするような情報インフラとしてのセンサネットワークの概念が提案されており、ネットワーク化された電子システム上で利用可能なデータ

は、規模と範疇の両者において爆発的に増加しつつある。大規模データベースの重要な用途の一つは、大量のデータから集団の特徴を表す一般的パターンを抽出する、即ち、大量のスパース (sparse) なデータからの汎化された情報を生成するデータマイニングである。このような社会的要求に応えるべく、データマイニングのための様々なアルゴリズムが提案されている。Kohonen による自己組織化写像は有力な手法の一例であり、実際、実用的な手法として広く活用されている [1]。これら従来のアルゴリズムの多くは、データから抽出すべき一般的特徴やパターンに関する何らかの事前情報を必要とする。例えば、Kohonen のアルゴリズムでは、パターン形成の核となるべき

雛形パターンが予め必要とされる。しかしながら、実際のデータマイニングにおいては、データの一般的特徴はデータから獲得すべきものであり、利用可能な事前情報はない場合が多い。

本研究では、何らの事前情報を用いることなく、データ自身にその特徴を生成させるための手法に関する提案を行う。データ自身が一般的特徴を形成する過程は、データが“相転移”を通して秩序構造を形成する過程と見ることもできるかも知れない。筆者らは、このようなアイデアを、集団同期の蔵本理論[2]-[4]を応用して実現できる可能性があることを見出した[5]。この手法は、時系列の近似モデルの良否をデータとモデルの同期を通して評価する手法[6],[7]とはまったく異なる。本報告では、[5]に示した手法と類似の別のダイナミクスを用いて、[5]と同様なデータ分析結果が得られることを示す。[5]では、提案した手法を要介護認定データベースに適用し、要介護高齢者の状態には3つの典型的パターンが存在することを示唆したが、このような仮説は本報告によって支持される。

## 2. 位相振動子ネットワーク上のデータ同期

位相振動子ネットワーク上の集団同期現象を利用したデータマイニング手法[5]の概要は以下のとおりである。今、 $D$ 自由度の多変量データが  $N$  個与えられたとする。これらを  $\bar{x}_i = (x_i(1), \dots, x_i(D))$  ( $i = 1, \dots, N$ ) と表す。位相振動子ネットワークの蔵本モデルのダイナミクスにおいて、自然周波数に多変量データを代入し、かつ、振動子間の相互作用範囲が位相差に応じて限定されたダイナミクスを考える。

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

ただし、 $K > 0$  は結合定数、 $\theta_i(n)$  は位相ベクトル  $\bar{\theta}_i = (\theta_i(1), \dots, \theta_i(D))$  の第  $n$  成分である。位相ベクトルの初期値は乱数によって与えられる。 $\bar{\theta}_i$  の時間微分は、 $\bar{x}_i$  の更新結果を表す。位相振動子の相互作用範囲を決定する関数  $H$  は、データベクトル間距離  $\bar{d}_{i,j} = |\bar{x}_i - \bar{x}_j|$  について  $\bar{d}_{i,j} \leq \bar{d}_0$  ならば

$$H(\bar{d}_{i,j}) = 1$$

$\bar{d}_{i,j} > \bar{d}_0$  ならば

$$H(\bar{d}_{i,j}) = 0$$

と定義される。ただし、 $\bar{d}_0 = \alpha |\bar{x}_i|$ 、 $\alpha > 0$  は定数である。こうして、位相ベクトル  $\bar{x}_i$  が相互作用できる  $N_i$  個の近接ベクトルが決定される。

式(1)で与えられるダイナミクスは、蔵本モデルの優れた特徴である mean field character を局所的に保持するので、データベクトルが無限個与えられた極限におけるデータ同期の統計的性質を表現することができる。これは、局所的な秩序パラメータをベクトル量に拡張することによって確かめられる。

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

$\bar{r}_i = (r_i(1), \dots, r_i(D))$  は局所的な秩序度であり、 $\bar{\psi}_i =$

$(\psi_i(1), \dots, \psi_i(D))$  は局所的な平均場である。これらのパラメータを用いると、式(1)は

$$\frac{d\theta_i(n)}{dt} = x_i(n) + K r_i(n) \sin(\psi_i(n) - \theta_i(n)) \quad (3)$$

と書き直され、local mean field character が明示される。式(3)より、係数  $K$  と  $\alpha$  は、 $K r_i(n) \geq |x_i(n) - X_g(n)|$  を満たし、かつ、 $\bar{x}_i$  の近接ベクトル数が十分に確保されるように与えられればよいことが分かる。こうして、データ集団の部分同期が達成され、データが自発的にグループ化されるであろう。各グループのメンバーが収束する共通の周波数を  $\bar{X}_g = (X_g(1), \dots, X_g(D))$  ( $g = 1, \dots, Q$ ;  $Q$  がグループ数) とすると、これらがもとの多変量データ集団から抽出された一般的特徴としてのテンプレートを表現する。

式(1)に一見類似しているダイナミクスは

$$\frac{d\bar{\theta}_i}{dt} = \bar{x}_i + \frac{K}{N_i} \sum_{j=1}^N (\bar{\theta}_j - \bar{\theta}_i) H(\bar{d}_{i,j}) \sin\left(\frac{\pi \bar{d}_{i,j}}{2\bar{d}_0}\right) \quad (4)$$

$$d_{i,j} = |\bar{\theta}_i - \bar{\theta}_j| \quad (5)$$

である。このダイナミクスは、式(1)とは異なり、mean field character を保持しない。また、振動子間の位相差が減少するにつれて、より速く相互作用が弱まる。

## 3. Case Study

前節で述べた2つのダイナミクス、即ち、式(1)と式(4)を用いて、データクラスタリングの実証実験を行った。実験に用いた多変量データは、2000年よりわが国で施行されている介護保険制度における要介護認定データである[8],[9]。この保険制度では、65歳以上の要介護高齢者は、心身の健康状態に応じて認定された等級に基づいて保険金が地方自治体から支給され、これを経済的原資に用いて様々な介護サービスを購入することができる。要介護認定に当たっては、高齢者は73項目の健康状態に関する調査を受け、調査結果から電子システムによって判定される要介護等級[10]と医師による意見をもとに認定委員会が高齢者の要介護等級を最終決定する。73項目の調査項目と回答に対して与えられるスコアの一覧は、表1および表2の左側2列に記されている。

介護保険制度の施行以来、全国の認定データは、73項目のスコア値と要介護等級からなる多変量データとしてデータベースに保存、蓄積されており、そのデータ規模は2500万症例に達している。この大規模データは、わが国における要介護高齢者の状態像を反映する。このデータから、多数の要介護高齢者が該当するような共通の特徴的パターンが発見されたとなると、その成果は今後の保険制度の改訂において非常に重要な情報となるであろう。しかしながら、このデータベースを構成する73項目の調査事項に対する回答サンプルは、73次元の自由度の多変量データベクトルであり、従来のデータマイニング手法を用いてパターン抽出を行うことは困難である。

本研究では、2500万症例のデータから無作為に選択された2000症例の73次元多変量データからなるサンプルデータ集団を12集団作成し、前節で述べたデータ同期の手法を用いて、

特徴的パターンの抽出実験を行った。  $K = 10$ ,  $\alpha = 0.7$ , および、4 次の Runge-Kutta 法の時間刻み幅 0.1 の条件下で式 (1) によりデータ分析を行ったところ、[5] に記したように、部分的集団同期 (partial synchronization) が達成され、3 つの主なベクトルが得られた。 これらを表 1 に示す。 これら 3 つの共通ベクトルに属するデータメンバーの総数は、全サンプルデータ数のおよそ 75% に相当する。 表 1 と本質的に同様な結果が、  $K = 0.4$ ,  $\alpha = 0.7$ , および、4 次の Runge-Kutta 法の時間刻み幅 0.1 の条件下で式 (4) によって分析実験を行った場合にも得られた。 この結果を表 2 に示す。

いずれのダイナミクスを用いた場合にも、3 つの主な要介護高齢者の状態像が抽出された。 これらを class 1, class 2, class 3 と名付けて分類する。 class 1 の主な特徴は、足に機能的障害があることである。 これは歩行や足を使った運動時の支障となって現れている。 class 3 では、足の機能的障害が進行すると同時に、短期記憶や知的判断の能力低下が顕著に認められる。 class 2 は、class 1 から class 3 への健康状態の悪化の過程の中間地点に該当するようと思われる。 これらのパターンは、高齢化にともなって現れやすい機能低下として、従来より信じられてきたパターンに類似していると言えるかも知れないが、数理的手法によって well-defined のデータから実証的に導かれた情報であるという点で、重要な意義をもつものと言えるであろう。

#### 4. 結 論

本研究では、データ集団に仮想的な集団同期、即ち、データ同期を導入して、データ集団の“相転移”を引き起こすことによって、データクラスタリングを実現した。 このような情報処理の動的プロセスは、生体の脳における神経パルス列の同期と時間間隔による情報コード化に関する仮説を思い起こさせる [11]–[20]。 本手法において得られた部分同期の安定性とデータクラスタリング結果の信頼性との関連、部分同期のパラメータ  $\alpha$  への依存性、あるいは、様々な多変量データへの適用を通じたデータマイニング手法としての有用性の検証等、課題は多いが、今後、これらの解決に向けて研究を続けたい。

#### 謝 辞

本研究を行うにあたり有益な助言と技術的支援をいただいた東野定律博士および谷口仁志氏に感謝の意を表す。 本研究の一部は厚生労働省厚生労働科学研究費補助金による助成のもとで行われたものである。 ここに謝意を表す。

#### 文 献

- [1] T. Kohonen, “Self-Organized Formation of Topologically Correct Feature Maps,” *Biol. Cybern.*, vol.43, pp.59–69, 1982.
- [2] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, Springer, New York, 1984.
- [3] 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.
- [4] R. E. Mirollo and S. H. Strogatz, “The Spectrum of the Locked State for the Kuramoto Model of Coupled Oscillators,” *Physica D*, vol.205, pp.249–266, 2005.
- [5] T. Miyano and T. Tsutsui, “Data Synchronization in a Network of Coupled Phase Oscillators,” *Phys. Rev. Lett.*, vol.98, no.2, pp.024102-1-024102-4, 2007.
- [6] L. A. Aguirre, E. C. Furtado, and L. A. B. Tôrres, “Evaluation of Dynamical Models: Dissipative Synchronization and Other Techniques,” *Phys. Rev. E*, vol.74, pp.066203-1-066203-16, 2006.
- [7] R. Brown, N. F. Rulkov, and E. R. Tracy, “Modeling and Synchronizing Chaotic Systems from Time-Series Data,” *Phys. Rev. E*, vol.49, no.5, pp.3784–3800, 1994.
- [8] T. Tsutsui and N. Muramatsu, “Care-Needs Certification in the Long-Term Care Insurance System of Japan,” *J. Am. Geriatr. Soc.*, vol.53, pp.522–527, 2005.
- [9] J. C. Campbell and N. Ikegami, “Long-Term Care Insurance Comes to Japan,” *Health Aff.*, vol.19, pp.26–39, 2000.
- [10] T. Miyano, T. Tsutsui, Y. Seki, S. Higashino, and H. Taniguchi, “Prediction of Care Class by Local Additive Reference to Prototypical Examples,” *IEEE Trans. Inf. Technol. Biomed.*, vol.9, no.4, pp.502–507, 2005.
- [11] J. A. Acebron, L. L. Bonilla, C. J. P. Vincente, F. Ritort, and R. Spigler, “The Kuramoto Model: A Simple Paradigm for Synchronization Phenomena,” *Rev. Mod. Phys.*, vol.77, pp.137–185, 2005.
- [12] C. M. Gray and W. Singer, “Stimulus-Specific Neuronal Oscillations in Orientation Columns of Cat Visual Cortex,” *Proc. Natl. Acad. Sci. USA*, vol.86, pp.1698–1702, 1989.
- [13] Y. Kuramoto, “Collective Synchronization of Pulse-Coupled Oscillators and Excitable Units,” *Physica D*, vol.50, pp.15–30, 1991.
- [14] H. Sompolinsky, D. Golomb, and D. Kleinfeld, “Cooperative Dynamics in Visual Processing,” *Phys. Rev. A*, vol.43, pp.6990–7011, 1991.
- [15] W. R. Softky and C. Koch, “The Highly Irregular Firing of Cortical Cells is Inconsistent with Temporal Integration of Random EPSPs,” *J. Neurosci.*, vol.13, pp.334–350, 1993.
- [16] J. J. Hopfield, “Pattern Recognition Computation Using Action Potential Timing for Stimulus Representation,” *Nature*, vol.376, pp.33–36, 1995.
- [17] H. Fujii, H. Ito, K. Aihara, N. Ichinose, and M. Tsukada, “Dynamical Cell Assembly Hypothesis - Theoretical Possibility of Spatio-Temporal Coding in the Cortex,” *Neural Networks*, vol.9, pp.1303–1350, 1996.
- [18] P. Seliger, S. C. Young, and L. S. Tsimring, “Plasticity and Learning in a Network of Coupled Phase Oscillators,” *Phys. Rev. E*, vol.65, pp.041906-1–041906-7, 2002.
- [19] H. Haken, “Synchronization and Pattern Recognition in a Pulse-Coupled Neural Net,” *Physica D*, vol.205, pp.1–6, 2005.
- [20] R. Gutig and H. Sompolinsky, “The Tempotron: A Neuron That Learns Spike Timing-Based Decisions,” *Nat. Neurosci.*, vol.9, pp.420–428, 2006.

表 1 要介護高齢者の健康状態における典型的パターン。式 (1) を利用した。表の第 1 列は 73 項目からなる要介護認定質問事項、第 2 列は回答に対する評価スコアを表す。“-1”は健康であること、1~最大値は健康状態に問題があること（整数スコアの増加は悪化）を表す。第 3 列~第 5 列は、データ同期によって得られた代表的 3 パターンである。

調査項目	評価スコア	Class 1	Class 2	Class 3
1. 麻痺 (左一上肢)	-1, 1	-0.9	-0.9	-1
2. 麻痺 (右一上肢)	-1, 1	-0.9	-0.9	-0.9
3. 麻痺 (左一下肢)	-1, 1	-0.2	0	0.2
4. 麻痺 (右一下肢)	-1, 1	-0.2	-0.2	-0.4
5. 麻痺 (その他)	-1, 1	-0.8	-0.8	-0.9
6. 拘縮 (肩関節)	-1, 1	-0.8	-0.8	-0.7
7. 拘縮 (肘関節)	-1, 1	-0.9	-0.9	-0.8
8. 拘縮 (股関節)	-1, 1	-0.8	-0.8	-0.8
9. 拘縮 (膝関節)	-1, 1	-0.3	-0.3	-0.3
10. 拘縮 (足関節)	-1, 1	-0.8	-0.9	-1
11. 拘縮 (その他)	-1, 1	-0.6	-0.6	-0.6
12. 寝返り	-1, 1, 2	-0.5	0.6	0.4
13. 起き上がり	-1, 1, 2	0.1	0.3	0.4
14. 両足での座位	-1, 1, 2, 3	-0.7	0.4	0.6
15. 両足つかない座位	-1, 1, 2, 3	0	0.5	0.5
16. 両足での立位	-1, 1, 2	-0.8	0.3	0.5
17. 歩行	-1, 1, 2	0.6	0.6	0.6
18. 移乗	-1, 1, 2, 3	0.6	0.7	0.7
19. 立ち上がり	-1, 1, 2	-0.6	1.5	1
20. 片足での立位	-1, 1, 2	-0.9	-0.9	-0.8
21. 浴槽の出入り	-1, 1, 2, 3	-0.7	-0.7	-0.6
22. 洗身	-1, 1, 2, 3	-0.8	-0.8	-0.9
23. じょくそう	-1, 1	-0.9	-0.5	0.6
24. 皮膚疾患	-1, 1	-0.9	-0.1	0
25. 片手胸元持ち上げ	-1, 1, 2	-0.9	-0.3	0.3
26. 嚥下	-1, 1, 2	-0.9	-0.2	1.4
27. 尿意	-1, 1, 2	-0.3	1.8	1.3
28. 便意	-1, 1, 2	-0.9	1.3	1.4
29. 排便後の後始末	-1, 1, 2, 3	-0.9	1.7	1.7
30. 排便後の後始末	-1, 1, 2, 3	-0.7	0.1	0.6
31. 食事摂取	-1, 1, 2, 3	-0.5	0.7	0.7
32. 口腔清潔	-1, 1, 2	-0.6	-0.6	-0.7
33. 洗顔	-1, 1, 2	-0.3	-0.3	-0.4
34. 整髪	-1, 1, 2	-0.9	-0.8	-0.2
35. つめ切り	-1, 1, 2	-0.9	-0.9	-1
36. ボタンのかけはずし	-1, 1, 2, 3	-0.9	-0.7	0.1
37. 上衣の着脱	-1, 1, 2, 3	-0.9	-0.9	-0.9
38. スボン等の着脱	-1, 1, 2, 3	-0.9	-0.9	-0.4
39. 靴下の着脱	-1, 1, 2, 3	-0.9	-0.9	-0.8
40. 居室の掃除	-1, 1, 2	-0.9	-0.9	-0.8
41. 薬の内服	-1, 1, 2	-0.9	-0.9	-0.9
42. 金銭の管理	-1, 1, 2	-0.9	-0.9	-0.7
43. ひどい物忘れ	-1, 1, 2	-0.9	-0.9	-0.9
44. 周囲への無関心	-1, 1, 2	-0.9	-0.9	-0.8
45. 視力	-1, 1, 2, 3, 4	-0.9	-0.9	-0.8
46. 聴力	-1, 1, 2, 3, 4	-0.8	-0.8	-0.8
47. 意思の伝達	-1, 1, 2, 3	-0.9	-0.9	-0.8
48. 指示への反応	-1, 1, 2	-0.8	-0.9	-0.8
49. 毎日の日課を理解	-1, 1	-0.9	-0.9	-0.8
50. 生年月日をいう	-1, 1	-0.9	-0.9	-0.4
51. 短期記憶	-1, 1	-0.9	-0.9	-0.8
52. 自分の名前をいう	-1, 1	-0.9	-0.9	-0.8
53. 今の季節を理解	-1, 1	-0.9	-0.9	-0.8
54. 場所の理解	-1, 1	-0.9	-0.9	-0.8
55. 被害的	-1, 1, 2	-0.9	-0.9	-0.8
56. 作話	-1, 1, 2	-0.8	-0.9	-0.9
57. 幻視幻聴	-1, 1, 2	-0.9	-0.9	-0.8
58. 感情が不安定	-1, 1, 2	-0.9	-0.9	-0.8
59. 昼夜逆転	-1, 1, 2	-0.9	-0.9	-0.8
60. 暴言暴行	-1, 1, 2	-0.4	-0.4	0.5
61. 同じ話をする	-1, 1, 2	-0.7	0.1	0.4
62. 大声をだす	-1, 1, 2	0	1.4	0.5
63. 介護に抵抗	-1, 1, 2	-0.5	1.2	1.3
64. 常時の徘徊	-1, 1, 2	-0.9	-0.9	-0.9
65. 落ち着きなし	-1, 1, 2	-0.9	-0.6	0.2
66. 外出して戻れない	-1, 1, 2	-0.9	-0.7	-0.3
67. 一人で出たがる	-1, 1, 2	-0.8	0.6	1.8
68. 収集癖	-1, 1, 2	-0.9	0.5	1.5
69. 火の不始末	-1, 1, 2	-0.9	1.4	1.4
70. 物や衣類を壊す	-1, 1, 2	-0.9	1.5	1.7
71. 不潔行為	-1, 1, 2	0.7	1	1.2
72. 異食行動	-1, 1, 2	-0.9	-0.9	1.4
73. 性的迷惑行為	-1, 1, 2	-0.9	-0.9	-0.9

表 2 要介護高齢者の健康状態における典型的パターン。式 (4) を利用した。表の第 1 列は 73 項目からなる要介護認定質問事項、第 2 列は回答に対するスコアを表す。“-1” は健康であること、1~最大値は健康状態に問題があること（整数スコアの増加は悪化）を表す。第 3 列~第 5 列は、データ同期によって得られた代表的 3 パターンである。

調査項目	評価スコア	Class 1	Class 2	Class 3
1. 麻痺 (左一上肢)	-1, 1	-1	-0.8	-0.6
2. 麻痺 (右一上肢)	-1, 1	-1	-0.9	-0.6
3. 麻痺 (左一下肢)	-1, 1	-0.4	0.1	0.7
4. 麻痺 (右一下肢)	-1, 1	-0.4	0	0.6
5. 麻痺 (その他)	-1, 1	-0.9	-0.9	-0.8
6. 拘縮 (肩関節)	-1, 1	-0.9	-0.8	-0.6
7. 拘縮 (肘関節)	-1, 1	-1	-0.9	-0.8
8. 拘縮 (股関節)	-1, 1	-0.9	-0.9	-0.7
9. 拘縮 (膝関節)	-1, 1	-0.4	-0.3	-0.3
10. 拘縮 (足関節)	-1, 1	-0.9	-0.9	-0.8
11. 拘縮 (その他)	-1, 1	-0.7	-0.7	-0.6
12. 寝返り	-1, 1, 2	-0.7	-0.3	0.5
13. 起き上がり	-1, 1, 2	0	0.4	1
14. 両足での座位	-1, 1, 2, 3	-0.8	-0.4	0.8
15. 両足つかない座位	-1, 1, 2, 3	-0.2	0.4	1.2
16. 両足での立位	-1, 1, 2	-0.9	-0.5	1
17. 歩行	-1, 1, 2	0.5	0.8	1.2
18. 移乗	-1, 1, 2, 3	0.5	0.9	1.5
19. 立ち上がり	-1, 1, 2	-0.7	-0.1	1.4
20. 片足での立位	-1, 1, 2	-1	-1	-0.9
21. 浴槽の出入り	-1, 1, 2, 3	-0.8	-0.7	-0.6
22. 洗身	-1, 1, 2, 3	-0.9	-0.9	-0.5
23. じよくそう	-1, 1	-1	-0.8	0
24. 皮膚疾患	-1, 1	-1	-0.7	0.6
25. 片手胸元持ち上げ	-1, 1, 2	-1	-0.7	0.6
26. 嚥下	-1, 1, 2	-1	-0.8	0.5
27. 尿意	-1, 1, 2	-0.4	0.2	1.6
28. 便意	-1, 1, 2	-0.9	-0.5	1.7
29. 排尿後の後始末	-1, 1, 2, 3	-0.9	-0.4	1.9
30. 排便後の後始末	-1, 1, 2, 3	-0.6	-0.3	1
31. 食事摂取	-1, 1, 2, 3	-0.4	-0.1	1.4
32. 口腔清潔	-1, 1, 2	-0.7	-0.6	-0.3
33. 洗顔	-1, 1, 2	-0.3	-0.2	0.1
34. 整髪	-1, 1, 2	-1	-0.9	-0.1
35. つめ切り	-1, 1, 2	-0.9	-0.9	-0.3
36. ボタンのかけはずし	-1, 1, 2, 3	-0.9	-0.9	0
37. 上衣の着脱	-1, 1, 2, 3	-1	-1	-0.5
38. スボン等の着脱	-1, 1, 2, 3	-0.9	-0.9	-0.1
39. 靴下の着脱	-1, 1, 2, 3	-1	-1	-0.9
40. 居室の掃除	-1, 1, 2	-0.9	-0.9	-0.3
41. 衣の内服	-1, 1, 2	-1	-1	-0.5
42. 金銭の管理	-1, 1, 2	-0.9	-1	-0.8
43. ひどい物忘れ	-1, 1, 2	-1	-1	-0.9
44. 周囲への無関心	-1, 1, 2	-1	-1	-0.8
45. 視力	-1, 1, 2, 3, 4	-0.9	-0.9	-0.7
46. 聴力	-1, 1, 2, 3, 4	-0.9	-0.9	-0.6
47. 意思の伝達	-1, 1, 2, 3	-1	-1	-0.9
48. 指示への反応	-1, 1, 2	-0.8	-0.9	-0.7
49. 毎日の日課を理解	-1, 1	-1	-1	-0.8
50. 生年月日をいう	-1, 1	-1	-1	-0.5
51. 短期記憶	-1, 1	-1	-1	-0.7
52. 自分の名前をいう	-1, 1	-1	-1	-0.9
53. 今の季節を理解	-1, 1	-1	-1	-0.9
54. 場所の理解	-1, 1	-1	-1	-0.9
55. 被害的	-1, 1, 2	-1	-1	-1
56. 作話	-1, 1, 2	-0.9	-0.9	-1
57. 幻視幻聴	-1, 1, 2	-1	-1	-1
58. 感情が不安定	-1, 1, 2	-1	-1	-0.8
59. 昼夜逆転	-1, 1, 2	-1	-1	-0.9
60. 暴言暴行	-1, 1, 2	-0.3	-0.4	0.4
61. 同じ話をする	-1, 1, 2	-0.9	-0.5	0.5
62. 大声をだす	-1, 1, 2	-0.1	0.3	1.2
63. 介護に抵抗	-1, 1, 2	-0.6	0	1.6
64. 常時の徘徊	-1, 1, 2	-1	-1	-1
65. 落ち着きなし	-1, 1, 2	-1	-0.9	-0.1
66. 外出して戻れない	-1, 1, 2	-1	-0.9	-0.2
67. 一人で出たがる	-1, 1, 2	-0.9	-0.5	1.4
68. 収集癖	-1, 1, 2	-1	-0.6	1.3
69. 火の不始末	-1, 1, 2	-1	-0.5	1.5
70. 物や衣類を壊す	-1, 1, 2	-0.9	-0.4	1.8
71. 不潔行為	-1, 1, 2	0.7	1.1	1.9
72. 異食行動	-1, 1, 2	-0.9	-0.9	-0.1
73. 性的迷惑行為	-1, 1, 2	-1	-1	-1