ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Physica D 198 (2004) 51–73

PHYSICA D

www.elsevier.com/locate/physd

Modeling share dynamics by extracting competition structure

Masahiro Kimura*, Kazumi Saito, Naonori Ueda

NTT Communication Science Laboratories, 2-4 Hikaridai, Seika-cho, Kyoto 619-0237, Japan

Received 29 January 2004; accepted 9 August 2004

Communicated by Y. Kuramoto

Abstract

We propose a new method for analyzing multivariate time-series data governed by competitive dynamics such as fluctuations in the number of visitors to Web sites that form a market. To achieve this aim, we construct a probabilistic dynamical model using a replicator equation and derive its learning algorithm. This method is implemented for both categorizing the sites into groups of competitors and predicting the future shares of the sites based on the observed time-series data. We confirmed experimentally, using synthetic data, that the method successfully identifies the true model structure, and exhibits better prediction performance than conventional methods that leave competitive dynamics out of consideration. We also experimentally demonstrated, using real data of visitors to 20 Web sites offering streaming video contents, that the method suggested a reasonable competition structure that conventional methods failed to find and that it outperformed them in terms of predictive performance.

PACS: 89.20.Hh; 07.05.Mh; 02.50.Tt; 05.45.Tp

Keywords: Web dynamics; Multivariate time-series modeling; Learning algorithm; Clustering; Prediction

1. Introduction

The World Wide Web provides a vast information space and is growing as a novel important medium of communication, and scientific and technological investigations of the Web are becoming important and challenging research issues [17,7,16,20,28,4,11,26,19]. From the viewpoints of sociology and economics, the Web can be regarded as a global market in which site maintainers offer information goods to users, and the number of visitors to a site during a period can become a proxy for that site's success [2,25]. Namely, we can consider that Web sites offering similar services compete to increase visitors. It is important, therefore, to model fluctuations in the

[*] Corresponding author. Tel.: +81 774 93 5131; fax: +81 774 93 5155.
E-mail address: kimura@cslab.kecl.ntt.co.jp (M. Kimura).

0167-2789/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2004.08.022

number of visitors to Web sites that form a market in terms of competitive dynamics, and investigating such Web dynamics is attracting the attention of the research community involved in complex systems and nonlinear dynamics [25].

In general, for any set of Web sites, not all the sites compete for the same users. For example, if we consider "asahi.com", "NIKKEI NET", "YOMIURI ON-LINE", "The New York Times on the Web" and "washingtonpost.com" as a set of Web sites offering the latest news, we may consider that the first three sites principally compete for Japanese users, and the last two sites principally compete for American users; thus, there is little competition between these two groups. That is, given a set of Web sites forming a market, the sites can generally be categorized into several competitive groups such that if sites belong to the same competitive group, then they strongly compete for the same population of users, otherwise they hardly compete at all. Incorporating competition structure enables us to model this sort of detailed competitive process. Also, it is important to extract the competition structure of the sites from observed data on fluctuations in numbers of visitors to the sites, since such knowledge can help site maintainers plan their business strategies. Therefore, it can be necessary to incorporate competition structure to model the competitive dynamics of sites.

Recently, researchers have discovered several statistical regularities in the access patterns of Web surfers, such as that the distribution of visitors to sites follows a power law [17,2,22]. To qualitatively explain these regularities, they have proposed statistical theories [17,2] and agent models [22] for Web user behavior. Maurer and Huberman [25] have proposed a model of usage dynamics for Web sites offering similar services and competing for the same population of users, in order to qualitatively explore the effects of competition among the sites on the nature of the market. However, the models of these previous works are not equipped with any learning mechanisms since they were principally aimed at finding qualitative explanations of observed phenomena, and not at use as predictive models (quantitative models) for the observed data. For example, Maurer and Huberman's [25] model is described as a nonlinear differential equation with as many parameters as the square of the number of sites, and the parameter fitting can generally be difficult. Therefore, it is not appropriate to use those previous models to predict tomorrow's share of each site in a certain market based on those observations.

On the other hand, when attempting to predict the number of visitors to each site in a market in the near future using the observed time-series data, we might be able to apply black-box models such as AR models, artificial neural networks and nonlinear time-series analysis models [5,18,6]. However, these models do not explicitly represent the underlying structures and the mechanisms of phenomena such as the competition structure of the market and the competitive dynamics of sites.

In this paper, we consider the problem of modeling short-term fluctuations in the number of visitors to sites that form a market. Given time-series data on the number of visitors to the sites, we aim to construct such a predictive model of this fluctuation process that has the following properties:

- It can represent the structure and the mechanism by which the sites interact in terms of competition structure and competitive dynamics.
- It can quantitatively simulate the observed time-series data on the number of visitors to the sites and be used to predict the near-future population shares of the sites.

For this purpose, we propose a new probabilistic dynamical model based on competition structure and competitive dynamics, and present its learning algorithm. In particular, we incorporate a *replicator equation* in evolutionary game theory [15] to model the *share dynamics* of the sites competing for the same population of users. Furthermore, we propose adopting as our predictive model of the actual process the model that has learned the observed time-series data. Using this learned model, we can categorize the sites into groups of competitors and predict the one-step future shares of the sites. In particular, the proposed method enables us to extract the competition structure of the sites based on the time-series data on the number of visitors to the sites.

In Section 2, we mathematically formulate the problem discussed in the paper and outline the proposed method, then in Section 3 we describe the proposed model in detail. In Section 4, we give a learning algorithm for the proposed model. In Section 5, we define how to evaluate the proposed method, and introduce the conventional methods with which it should be compared. In Section 6, we experimentally demonstrate that the proposed method works well and can also outperform conventional methods, in both the task of predicting the one-step future shares of the sites and the task of categorizing the sites into groups of competitors.

2. Overview

We begin by mathematically formulating the problem of modeling fluctuations in the number of visitors to sites that form a market (see, Fig. 1). Let $S = \{s_1, \ldots, s_N\}$ be a set of Web sites that form a certain market, including a set of sites offering similar services. For each i, let $m_i(t)$ denote the number of visitors to site s_i at time-step t. Here, we appropriately set the unit of time (for example, one time-step is 1 day, 2 days, or 1 week), and we interpret $m_i(t)$ as the number of visitors to s_i during the period $(t-1, t]$. We call

$$\boldsymbol{m}(t) = (m_1(t), \ldots, m_N(t))$$

the *visit vector to market S at time-step t*. In this paper, we deal with the problem of modeling the evolution of $\boldsymbol{m}(t)$.

Let $M(t)$ denote the total number of visitors to S at time-step t, that is,

$$M(t) = \sum_{i=1}^{N} m_i(t).$$

For example, according to Adamic and Huberman [2], we may be able to model the evolution of $M(t)$ by the stochastic dynamics

$$M(t+1) - M(t) = \nu(t)M(t) \quad (t \geq 1),$$

where each $\nu(t)$ is independently generated from some Gaussian distribution, and the fraction is omitted to obtain the total number $M(t)$ of visitors to S at each time-step t. However, we believe that the dynamics of $M(t)$ should in advance be modeled by investigating and observing a more extensive social system including S. Therefore, we assume that $M(t)$ is given, and focus on the problem of predicting the *share* of each site $s_i \in S$,

$$x_i(t) = \frac{m_i(t)}{M(t)}.$$

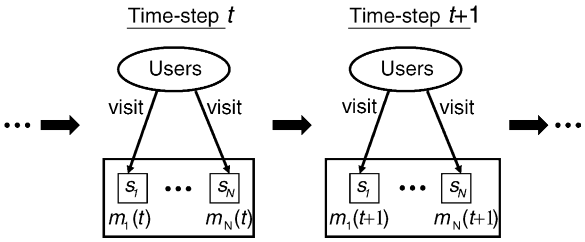

Fig. 1. A conceptual illustration of the usage dynamics for the sites $s_1, \ldots, s_N \in S$.

We call

$$\boldsymbol{x}(t) = (x_1(t), \dots, x_N(t))$$

the *share vector of the market S at time-step t*. Note that since $\boldsymbol{x}(t)$ is an element of the canonical $(N-1)$ dimensional simplex

$$\Delta^{N-1} = \left\{ (y_1, \dots, y_N); 0 \le y_1, \dots, y_N \le 1, \sum_{i=1}^{N} y_i = 1 \right\},$$

it is essentially necessary to model a stochastic dynamics on Δ^{N-1}.

Now, let us outline the proposed method for modeling the dynamics of visit vector $\boldsymbol{m}(t)$ to S. Based on the above discussion, we hypothesize that the process of fluctuations in the number of visitors to the sites of S is described as a stochastic process of the form $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1))$ $(t \ge 1)$. Namely, we assume that for any positive integer t, given visit vector $\boldsymbol{m}(t)$ to S at time-step t and total number $M(t+1)$ of visitors to S at time-step $t+1$, visit vector $\boldsymbol{m}(t+1)$ to S at time-step $t+1$ is generated according to the probability $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1))$. We first construct such a probabilistic dynamical model $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \boldsymbol{\Gamma})$ $(t \ge 1)$ of a generic process for the fluctuation phenomenon that represents the competition structure of S and the competitive dynamics of the sites, where $\boldsymbol{\Gamma}$ denotes the model parameter vector. Next, by learning from the observed time-series data of visit vectors to S, we acquire an optimal predictive model $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \hat{\boldsymbol{\Gamma}})$ $(t \ge 1)$ that approximates the actual process. We adopt the acquired model as the objective predictive model of the actual process.

3. Proposed model

In this section, we propose a probabilistic dynamical model $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \boldsymbol{\Gamma})$ $(t \ge 1)$ that can explain fluctuations in the number of visitors to the sites of S in terms of the competition structure and the competitive dynamics. We begin by describing the basic ideas of the proposed model.

3.1. Basic ideas

We assume that S can be divided into K competitive groups. Namely, we hypothesize that if sites belong to the same competitive group, they strongly compete for the same population of users, otherwise, they do not have the same population of users and hardly compete at all. Let

$$S = \bigcup_{k=1}^{K} C_k \quad \text{(disjoint union)}; \qquad C_k = \{s_i; i \in I_k\} \quad (k = 1, \dots, K) \tag{1}$$

denote the competition structure of S, where each C_k is a competitive group of S. Note that in the proposed model, the competition structure (1) of S is an adjustable parameter. In particular, the number K of competitive groups is also a parameter.

We interpret the behavior exhibited when a general user of market S visits a site $s_i \in C_k$ as follows: A general user of S first chooses a group C_k among groups $C_1, \dots, C_K$, and chooses a site s_i among sites $\{s_j; j \in I_k\}$ (see Fig. 2). We assume that during the period of our concern, the ratio of the users of a group C_k to the users of S is independent of time. Namely, we hypothesize that the probability that a general user of S chooses a group C_k at

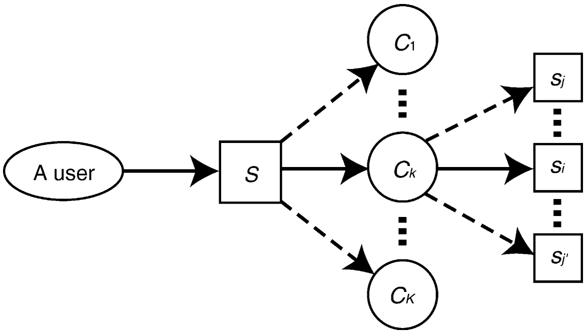

Fig. 2. Schematic diagram of the behavior of a general user that visits a site of S. A general user first chooses a group C_k among groups $C_1, \dots, C_K$. Next, the user who visits C_k chooses a site s_i among sites $\{s_j;\ j \in I_k\}$.

time-step t is a constant θ_k for any positive integer t (see Fig. 2). Here, $\theta_1, \dots, \theta_K$ are parameters such that

$$0 < \theta_1, \dots, \theta_K < 1; \qquad \sum_{k=1}^{K} \theta_k = 1.$$

Next, we consider the probability that a general user of C_k chooses a site $s_i \in C_k$ at time-step $t+1$ for any $t \geq 1$ (see Fig. 2). Note that this probability can be regarded as the average share $\overline{x_{k,i}(t+1)}$ of site s_i within group C_k at time-step $t+1$. Let $\overline{\boldsymbol{x}_k(t+1)}$ be the average share vector of group C_k constructed by arranging $\{\overline{x_{k,j}(t)};\ j \in I_k\}$. Also, let $x_{k,i}(t)$ be the actual share of site s_i within group C_k at time-step t, that is,

$$x_{k,i}(t) = \frac{m_i(t)}{\sum_{j \in I_k} m_j(t)} \quad (i \in I_k)$$

and let $\boldsymbol{x}_k(t)$ be the share vector of group C_k constructed by arranging $\{x_{k,j}(t);\ j \in I_k\}$. We hypothesize that the average share vector $\overline{\boldsymbol{x}_k(t+1)}$ of group C_k is basically determined by a *replicator equation* that is frequently adopted in ecosystem modeling [15]; that is, we assume that the share dynamics within group C_k is governed by

$$\overline{\boldsymbol{x}_k(t+1)} = \boldsymbol{f}_k(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k) \quad (t \geq 1); \qquad \overline{x_{k,i}(t+1)} = f_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k) \quad (i \in I_k, t \geq 1),$$

where $\boldsymbol{f}_k$ is some dynamics based on a replicator equation, and $\boldsymbol{\Phi}_k$ is a parameter vector.

Moreover, we hypothesize based on our user model (see Fig. 2) that for any $t \geq 1$, the probability $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \boldsymbol{\Gamma})$ is given by the following multinomial distribution:

$$P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \boldsymbol{\Gamma}) \propto \prod_{i=1}^{N} \{\theta_{\lambda_i} f_{\lambda_i,i}(\boldsymbol{x}_{\lambda_i}(t); \boldsymbol{\Phi}_{\lambda_i})\}^{m_i(t+1)}, \tag{2}$$

where each λ_i denotes the label of the group to which site s_i belongs; this means that site s_i belongs to group C_{λ_i}. Hence, we have obtained a probabilistic dynamical model of fluctuations in the number of visitors to the sites of S to which site s_i belongs; this means that site s_i belongs to group C_{λ_i}. Hence, we have obtained a probabilistic dynamical model of fluctuations in the number of visitors to the sites of S.

Our model represents not only the competitive dynamics of the sites in terms of replicator equations in evolutionary game theory [15] but also the competition structure of S. Moreover, it represents the probabilistic nature of fluctuations in the number of visitors to the sites in terms of multinomial distributions. In the following subsections, the proposed model is described in detail. First, a method to represent competition structures is introduced. Next, the share dynamics $\boldsymbol{f}_k$ within group C_k is rigorously defined. Finally, the proposed model is explicitly described as a probabilistic generative model, and its properties are also mathematically stated.

3.2. Competition structure

We represent a categorization of S into K competitive groups by an N-dimensional vector called a K division vector of S. Here, an N-dimensional vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ is called a *K division vector of S* if the following conditions are satisfied:

- $\lambda_1, \ldots, \lambda_N \in \{1, \ldots, K\}$.
- For $\forall k \in \{1, \ldots, K\}$, there exists some $i \in \{1, \ldots, N\}$ such that $\lambda_i = k$.

Then, categorization (1) of S is represented by the K division vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ of S such that

$$s_i \in C_k \Leftrightarrow \lambda_i = k \quad (k = 1, \ldots, K)$$

for $i = 1, \ldots, N$.

3.3. Replicator equations

Suppose that a categorization (1) of S into K competitive groups is given. For each k, we define the share dynamics $\boldsymbol{f}_k$ within group C_k based on a replicator equation. Let N_k denote the number of elements in C_k.

We first hypothesize that $\boldsymbol{f}_k$ is governed by a mixture of replicator dynamics and uniform dynamics,

$$f_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k) = (1 - \xi_k) g_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \eta_k) + \frac{\xi_k}{N_k} \quad (i \in I_k), \tag{3}$$

where ξ_k is a mixing parameter with $0 < \xi_k < 1$, and $g_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \eta_k)$ $(i \in I_k)$ represents replicator dynamics. Here, $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ and η_k are parameters. Namely, we hypothesize that the share dynamics within group C_k is not only subject to a replicator equation but also driven by the uniform external force.

Next, we define replicator equation $\overline{x_{k,i}(t+1)} = g_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \eta_k)$ $(i \in I_k)$. In terms of evolutionary game theory, we consider each site in group C_k as a type of strategy in a certain population, and regard the share of the site as the frequency of the type. Recall that a replicator equation is defined by the following principle [15]:

The rate of increase of type i is a measure of its success. This success is expressed as the difference between the fitness of i and the average fitness of the population.

In our problem, we may identify the fitness of type i with the attractiveness of site s_i. Thus, a replicator equation is defined by

$$\frac{\overline{x_{k,i}(t+1)} - x_{k,i}(t)}{x_{k,i}(t)} = A_{k,i}(\boldsymbol{x}_k(t)) - \bar{A}_k(\boldsymbol{x}_k(t)) \quad (i \in I_k), \tag{4}$$

where $A_{k,i}(\boldsymbol{x}_k(t))$ denotes the attractiveness of site s_i for state $\boldsymbol{x}_k(t)$ and $\bar{A}_k(\boldsymbol{x}_k(t))$ denotes the average attractiveness of group C_k for state $\boldsymbol{x}_k(t)$,

$$\bar{A}_k(\boldsymbol{x}_k(t)) = \sum_{i \in I_k} x_{k,i}(t) A_{k,i}(\boldsymbol{x}_k(t)). \tag{5}$$

We hypothesize that the attractiveness $A_{k,i}(\boldsymbol{x}_k(t))$ is defined by a mixture of the *general attractiveness* of s_i within C_k and the *rarity attractiveness* of s_i within C_k. Here, the general attractiveness of s_i at time-step t means some quantity proportional to the current share $x_{k,i}(t)$, and the rarity attractiveness of s_i at time-step t means some quantity proportional to the current opposite share $(1 - x_{k,i}(t))$. Therefore, the attractiveness $A_{k,i}(\boldsymbol{x}_k(t))$ is defined by

$$A_{k,i}(\boldsymbol{x}_k(t)) = (1 - \eta_k)\alpha_{k,i}x_{k,i}(t) + \eta_k\beta_{k,i}(1 - x_{k,i}(t)), \tag{6}$$

where η_k is a mixing parameter with $0 < \eta_k < 1$, and $\alpha_{k,i}$ and $\beta_{k,i}$ are, respectively, parameters that represent the potential values for the general attractiveness and the rarity attractiveness of s_i within C_k such that

$$0 < \alpha_{k,i}, \beta_{k,i} < 1 \quad (i \in I_k); \qquad \sum_{i \in I_k} \alpha_{k,i} = \sum_{i \in I_k} \beta_{k,i} = 1.$$

Let $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ denote the parameter vectors constructed by arranging $\{\alpha_{k,i}; i \in I_k\}$ and $\{\beta_{k,i}; i \in I_k\}$, respectively. Hence, from Eqs. (4), (5) and (6), we have

$$g_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \eta_k) = x_{k,i}(t)\left\{1 + A_{k,i}(\boldsymbol{x}_k(t)) - \sum_{j \in I_k} x_{k,j}(t)A_{k,j}(\boldsymbol{x}_k(t))\right\} \tag{7}$$

for any $i \in I_k$.

Now, we can explicitly describe the share dynamics within group C_k, $\overline{\boldsymbol{x}_k(t+1)} = \boldsymbol{f}_k(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k)$. From Eqs. (3) and (7), $\boldsymbol{f}_k(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k)$ is given by

$$f_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k) = \sum_{j \in I_k} a_{k,i,j}(t)(1 - \xi_k)(1 - \eta_k)\alpha_{k,j} + \sum_{j \in I_k} b_{k,i,j}(t)(1 - \xi_k)\eta_k\beta_{k,j} + \frac{\xi_k}{N_k} \tag{8}$$

for any $i \in I_k$ and any $t \geq 1$, where

$$a_{k,i,i}(t) = x_{k,i}(t)\{1 + x_{k,i}(t)(1 - x_{k,i}(t))\} \quad (i \in I_k, t \geq 1),$$
$$a_{k,i,j}(t) = x_{k,i}(t)(1 - x_{k,j}(t)^2) \quad (i, j \in I_k; j \neq i,\ t \geq 1),$$
$$b_{k,i,i}(t) = x_{k,i}(t)\{1 + (1 - x_{k,i}(t))^2\} \quad (i \in I_k,\ t \geq 1),$$
$$b_{k,i,j}(t) = x_{k,i}(t)(1 - x_{k,j}(t) + x_{k,j}(t)^2) \quad (i, j \in I_k; j \neq i, t \geq 1)$$

and $\boldsymbol{\Phi}_k$ is the parameter vector constructed by arranging $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$, η_k and ξ_k.

Note that both an actual share vector $\boldsymbol{x}_k(t)$ and an average share vector $\overline{\boldsymbol{x}_k(t+1)}$ are elements of the canonical $(N_k - 1)$ dimensional simplex. The following proposition shows that the definition of the share dynamics $\boldsymbol{f}_k$ within group C_k is well-defined.

Proposition 1. *Let t be a positive integer. For an arbitrary share vector* $\boldsymbol{x}_k(t)$ *of group* C_k *at time-step t,* $\boldsymbol{f}_k(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k)$ *is an element of the canonical* $(N_k - 1)$*-dimensional simplex.*

Proof. From Eq. (7) and $\sum_{i \in I_k} x_{k,i}(t) = 1$, it is easily seen that

$$\sum_{i \in I_k} g_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \eta_k) = 1.$$

Thus, from Eq. (3), we obtain

$$\sum_{i \in I_k} f_{k,i}(\boldsymbol{x}_k(t); \boldsymbol{\Phi}_k) = 1. \tag{9}$$

Notice that in Eq. (8), each $a_{k,i,j}(t)$ and $b_{k,i,j}(t)$ are nonnegative. Hence, we have $f_{k,i}(\boldsymbol{x}_k(t)) \geq 0$ for any $i \in I_k$. By this fact and Eq. (9), we can easily prove the proposition. □

3.4. Probabilistic generative model

We can now explicitly describe the proposed model.

We begin by specifying the model parameters. The model parameter vector $\mathbf{\Gamma}$ consists of the number K of competitive groups, the K division vector $\boldsymbol{\lambda}$, and the parameter vector $\mathbf{\Theta}$ constructed by arranging parameters $\theta_1, \ldots, \theta_K$ and parameter vectors $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K$. Note that parameter vector $\boldsymbol{\lambda}$ depends on parameter K, and parameter vector $\mathbf{\Theta}$ depends on parameter vector $\boldsymbol{\lambda}$.

Next, we describe the proposed model as a probabilistic generative model. Suppose that the values of model parameter vector $\mathbf{\Gamma}$ are given, and the corresponding competition structure of S is Eq. (1). Suppose also that time-series data $\{M(t); 1 \leq t \leq T+1\}$ of the total number of visitors to S are given. Then, given an initial visit vector $\boldsymbol{m}(1)$ to S such that $\sum_{i=1}^{N} m_i(1) = M(1)$, the model generates the time-series $\{\boldsymbol{m}(2), \ldots, \boldsymbol{m}(T+1)\}$ of visit vectors to S. Let us explain this generative process in detail. For an arbitrary integer t with $1 \leq t \leq T$, the visit vector $\boldsymbol{m}(t+1)$ at time-step $t+1$ is generated in the following way: Consider $M(t+1)$ general visitors to S at time-step $t+1$. First, a general visitor chooses a group according to the multinomial distribution with parameters $\theta_1, \ldots, \theta_K$. Suppose that a group C_k is chosen. Then, the visitor chooses a site of group C_k according to the multinomial distribution with parameters $\boldsymbol{f}_k(\boldsymbol{x}_k(t); \mathbf{\Phi}_k)$. This trial is performed $M(t+1)$ times, and results in the visit vector $\boldsymbol{m}(t+1)$ at time-step $t+1$. Namely, $\boldsymbol{m}(t+1)$ is generated according to the probability $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \mathbf{\Gamma})$. Let $n_k(t+1)$ denote the total number of visitors to C_k at time-step $t+1$, that is,

$$n_k(t+1) = \sum_{i \in I_k} m_i(t+1) \quad (k = 1, \ldots, K). \tag{10}$$

From Eq. (2), the probability $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \mathbf{\Gamma})$ is explicitly described by:

$$P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \mathbf{\Theta}, \boldsymbol{\lambda}) = \frac{M(t+1)!}{\prod_{k=1}^{K} n_k(t+1)!} \prod_{k=1}^{K} {\theta_k}^{n_k(t+1)} \frac{n_k(t+1)!}{\prod_{i \in I_k} m_i(t+1)!} \prod_{i \in I_k} f_{k,i}(\boldsymbol{x}_k(t); \mathbf{\Phi}_k)^{m_i(t+1)}. \tag{11}$$

Finally, let us state some properties of the proposed model. In the proposed model, we consider the mean and the variance for the share $n_k(t+1)/M(t+1)$ of group C_k at time-step $t+1$, and the mean and the variance for the share $m_i(t+1)/n_k(t+1)$ of site $s_i \in C_k$ within group C_k at time-step $t+1$. The next proposition presents the calculations of these values, and is easily proved by using the properties of multinomial distributions (see Appendix A).

Proposition 2. *For the conditional means and variances in the proposed probabilistic model, the following relations hold*: *For* $\forall k \in \{1, \ldots, K\}$, $\forall i \in I_k$ *and* $\forall t \geq 1$,

$$\left\langle \frac{n_k(t+1)}{M(t+1)} \middle| \mathbf{\Gamma} \right\rangle = \theta_k, \qquad \left\langle \frac{m_i(t+1)}{n_k(t+1)} \middle| \boldsymbol{m}(t), \mathbf{\Gamma} \right\rangle = f_{k,i}(\boldsymbol{x}_k(t); \mathbf{\Phi}_k),$$

$$\left\langle \left\{ \frac{n_k(t+1)}{M(t+1)} - \theta_k \right\}^2 \middle| \mathbf{\Gamma} \right\rangle = \frac{\theta_k(1-\theta_k)}{M(t+1)},$$

$$\left\langle \left\{ \frac{m_i(t+1)}{n_k(t+1)} - f_{k,i}(\boldsymbol{x}_k(t); \mathbf{\Phi}_k) \right\}^2 \middle| \mathbf{\Gamma} \right\rangle = \frac{f_{k,i}(\boldsymbol{x}_k(t); \mathbf{\Phi}_k)\{1 - f_{k,i}(\boldsymbol{x}_k(t); \mathbf{\Phi}_k)\}}{n_k(t+1)},$$

where $\langle X|Y \rangle$ *denotes the conditional expectation of* X *given* Y.

This proposition implies that many visitors to S reduce the noise levels (the variances) and make the probabilistic model behave like a deterministic one.

4. Learning algorithm

Let $\{\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T+1)\}$ be the observed time-series data of visit vectors to S, where T is an integer with $T \geq 2$. Based on this data set, we would like to construct the probabilistic dynamical model $P(\boldsymbol{m}(t+1)|\boldsymbol{m}(t), M(t+1), \hat{\boldsymbol{\Gamma}})$ $(t \geq 1)$ of the actual fluctuation process. Namely, for this data set, we estimate the optimal number $\hat{K}$ of competitive groups, the optimal $\hat{K}$ division vector $\hat{\boldsymbol{\lambda}}$ and the optimal parameter vector $\hat{\boldsymbol{\Theta}}$.

Given the optimal number $\hat{K}$ of competitive groups, we estimate $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\lambda}}$ based on the maximum likelihood method, that is, by maximizing the function

$$P(\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T+1)|\boldsymbol{\Theta}, \boldsymbol{\lambda}, \hat{K})$$

with respect to $\boldsymbol{\Theta}$ and $\boldsymbol{\lambda}$. Here, $P(\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T+1)|\boldsymbol{\Theta}, \boldsymbol{\lambda}, \hat{K})$ denotes the joint probability of the observed data $\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T+1)$ given parameters $\boldsymbol{\Theta}$, $\boldsymbol{\lambda}$ and $\hat{K}$.

On the other hand, we determine the optimal number $\hat{K}$ based on the criterion of maximizing prediction performance. Let us appropriately fix an integer ΔT with $1 \leq \Delta T \leq T$, and consider the problem of predicting the share vector $\boldsymbol{x}(T_0+1)$ at time-step T_0+1 from the time-series data $\{\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T_0)\}$ for any integer T_0 with $T - \Delta T + 1 \leq T_0 \leq T$. When the number K of competitive groups in S is given, we estimate the optimal values of both parameter vector $\boldsymbol{\Theta}$ and K division vector $\boldsymbol{\lambda}$ from the data set $\{\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T_0)\}$ by using the maximum likelihood method, and predict the expected share vector at time-step T_0+1 by using the estimated model. Let $\hat{\boldsymbol{x}}(T_0+1) = (\hat{x}_1(T_0+1), \ldots, \hat{x}_N(T_0+1))$ be the share vector at time-step T_0+1 predicted by the estimated model, that is,

$$\hat{x}_i(T_0+1) = \theta_{\lambda_i} f_{\lambda_i,i}(\boldsymbol{x}_{\lambda_i}(T_0); \boldsymbol{\Phi}_{\lambda_i}) \quad (i = 1, \ldots, N).$$

We evaluate the prediction performance of the estimated model by measuring the prediction error $E(T_0+1; K)$ for the actual share vector $\boldsymbol{x}(T_0+1)$,

$$E(T_0+1; K) = \frac{1}{2} \sum_{i=1}^{N} |\hat{x}_i(T_0+1) - x_i(T_0+1)|.$$

Hence, we estimate the optimal number $\hat{K}$ of competitive groups by minimizing the average prediction error $\bar{E}(K)$,

$$\bar{E}(K) = \frac{1}{\Delta T} \sum_{T_0=T-\Delta T+1}^{T} E(T_0+1; K)$$

with respect to K.

In Sections 4.1 and 4.2, we describe in detail how to estimate the optimal values of both parameter vector $\boldsymbol{\Theta}$ and K division vector $\boldsymbol{\lambda}$ from the data set $\{\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T_0)\}$ by using the maximum likelihood method, when the number K of competitive groups of S is given.

4.1. Estimation of $\boldsymbol{\Theta}$

First, we describe how to estimate the optimal values of parameter vector $\boldsymbol{\Theta}$ from the data set $\{\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T_0)\}$ when the number K of competitive groups in S and the K division $\boldsymbol{\lambda}$ of S are given. Suppose that categorization (1) of S into K competitive groups corresponds to $\boldsymbol{\lambda}$. For this task, we basically use the maximum likelihood method, where an optimal value of $\boldsymbol{\Theta}$ is basically estimated by minimizing the function

$$\mathcal{L}_K(\boldsymbol{\Theta}; \boldsymbol{\lambda}, T_0) = -\log\ P(\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T_0)|\boldsymbol{\Theta}, \boldsymbol{\lambda}, K)$$

with respect to $\mathbf{\Theta}$. By Eq. (11), we have

$$\log\ P(\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T_0)|\mathbf{\Theta}, \boldsymbol{\lambda}, K) = \sum_{k=1}^{K}\sum_{t=1}^{T_0-1} n_k(t+1)\log\ \theta_k + \sum_{k=1}^{K}\sum_{t=1}^{T_0-1}\sum_{i\in I_k} m_i(t+1)\log\ f_{k,i}(\boldsymbol{x}_k(t); \mathbf{\Phi}_k). \tag{12}$$

From Eqs. (8) and (12), we consider setting

$$\phi_{k,i} = (1-\xi_k)(1-\eta_k)\alpha_{k,i} \quad (i \in I_k), \qquad \psi_{k,i} = (1-\xi_k)\eta_k\beta_{k,i} \quad (i \in I_k) \tag{13}$$

for each k. Then, it is easily seen that each $\mathbf{\Phi}_k$ can be identified with the parameter vector constructed by arranging $\phi_{k,i}$'s, $\psi_{k,i}$'s and ξ_k such that

$$0 < \phi_{k,i}, \psi_{k,i}, \xi_k < 1 \quad (i \in I_k); \qquad \sum_{i\in I_k}(\phi_{k,i}+\psi_{k,i}) + \xi_k = 1.$$

Namely, $\mathbf{\Phi}_k$ is identified with an interior point of the $2N_k$ dimensional simplex Δ^{2N_k}. Since $\mathbf{\Theta}$ is represented by

$$\mathbf{\Theta} = ((\theta_1, \ldots, \theta_K), \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K),$$

the domain Ω of parameter vector $\mathbf{\Theta}$ is the interior of the product set $\Delta^{K-1} \times \Delta^{2N_1} \times \cdots \times \Delta^{2N_K}$. Therefore, to effectively estimate the optimal values of parameter vector $\mathbf{\Theta}$, we consider introducing the regularizer, called the Laplace smoothing, into the objective function. This means we estimate the optimal values of $\mathbf{\Theta}$ by minimizing the function

$$\mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0) = -\log\ P(\boldsymbol{m}(1), \ldots, \boldsymbol{m}(T_0)|\mathbf{\Theta}, \boldsymbol{\lambda}, K) - \sum_{k=1}^{K}\left(\log\ \theta_k + \sum_{i\in I_k}(\log\ \phi_{k,i} + \log\ \psi_{k,i}) + \log\ \xi_k\right).$$

Hence, from Eq. (12), the function $\mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0)$ becomes

$$\mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0) = \mathcal{L}_K^0(\theta_1, \ldots, \theta_K; \boldsymbol{\lambda}, T_0) + \mathcal{L}_K^1(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K; \boldsymbol{\lambda}, T_0), \tag{14}$$

where

$$\mathcal{L}_K^0(\theta_1, \ldots, \theta_K; \boldsymbol{\lambda}, T_0) = -\sum_{k=1}^{K}\left(1 + \sum_{t=1}^{T_0-1} n_k(t+1)\right)\log\ \theta_k \tag{15}$$

and

$$\begin{aligned}&\mathcal{L}_K^1(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K; \boldsymbol{\lambda}, T_0)\\ &\quad = -\sum_{k=1}^{K}\left\{\sum_{t=1}^{T_0-1}\sum_{i\in I_k} m_i(t+1)\log\ f_{k,i}\,(\boldsymbol{x}_k(t); \mathbf{\Phi}_k) + \sum_{j\in I_k}(\log\ \phi_{k,j} + \log\ \psi_{k,j}) + \log\ \xi_k\right\}.\end{aligned} \tag{16}$$

Then, we have the following proposition.

Proposition 3. *Given the number K of competitive groups, the K division vector $\boldsymbol{\lambda}$ and the length T_0 of training time-series data, the function $\Omega \ni \mathbf{\Theta} \mapsto \mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0) \in \mathbb{R}$ has locally minimal points, and these minimal points are also the global minimum points of this function.*

Proof. First, note that the domain Ω is the interior of $\Delta^{K-1} \times \Delta^{2N_1} \times \Delta^{2N_K}$, and its closure $\bar{\Omega}$ is a compact convex subset of the Euclidean space $\mathbb{R}^{2(K+N_1+\cdots+N_K)}$. The differentiable function $\exp(-\mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0))$ on $\bar{\Omega}$ has a maximum point in Ω since it is nonnegative on $\bar{\Omega}$ and identically zero on the boundary $\partial\Omega$ of Ω. This implies that the function $\mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0)$ has a minimum point in Ω. On the other hand, it is easily shown from Eqs. (8), (13)–(16) that the function $\Omega \ni \mathbf{\Theta} \mapsto \mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0) \in \mathbb{R}$ is convex. Hence, a locally minimal point of the function $\mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0)$ always becomes its global minimum point. We have thus proved the proposition. □

Let us compute a minimal point of the function $\Omega \ni \mathbf{\Theta} \mapsto \mathcal{L}_K(\mathbf{\Theta}; \boldsymbol{\lambda}, T_0) \in \mathbb{R}$. By Eq. (14), our task reduces to calculate $(\theta_1, \ldots, \theta_K)$ and $(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K)$ which minimize $\mathcal{L}_K^0(\theta_1, \ldots, \theta_K; \boldsymbol{\lambda}, T_0)$ and $\mathcal{L}_K^1(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K; \boldsymbol{\lambda}, T_0)$, respectively.

We first consider minimizing $\mathcal{L}_K^0(\theta_1, \ldots, \theta_K; \boldsymbol{\lambda}, T_0)$ with respect to $(\theta_1, \ldots, \theta_K)$. From Eq. (15), it is easily seen that the minimal point $(\theta_1^*, \ldots, \theta_K^*)$ of this function is obtained by

$$\theta_k^* = \frac{1 + \sum_{t=2}^{T_0} n_k(t)}{K + \sum_{t=2}^{T_0} M(t)} \quad (k = 1, \ldots, K). \tag{17}$$

Next, we consider minimizing $\mathcal{L}_K^1(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K; \boldsymbol{\lambda}, T_0)$ with respect to $(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K)$. This can be efficiently performed by using an iterative algorithm based on the EM algorithm [9]. Let $\tilde{\mathbf{\Phi}}_k$ be the current estimate of $\mathbf{\Phi}_k$ for each k. Here, we define the Q-function $Q_k(\mathbf{\Phi}_k|\tilde{\mathbf{\Phi}}_k)$ by

$$\begin{aligned} Q_k(\mathbf{\Phi}_k|\tilde{\mathbf{\Phi}}_k) = {} & \sum_{j \in I_k} \left\{ 1 + \sum_{t=1}^{T_0-1} \sum_{i \in I_k} \frac{m_i(t+1) a_{k,i,j}(t) \tilde{\phi}_{k,j}}{f_{k,i}(\boldsymbol{x}_k(t); \tilde{\mathbf{\Phi}}_k)} \right\} \log \phi_{k,j} \\ & + \sum_{j \in I_k} \left\{ 1 + \sum_{t=1}^{T_0-1} \sum_{i \in I_k} \frac{m_i(t+1) b_{k,i,j}(t) \tilde{\psi}_{k,j}}{f_{k,i}(\boldsymbol{x}_k(t); \tilde{\mathbf{\Phi}}_k)} \right\} \log \psi_{k,j} \\ & + \left\{ 1 + \sum_{t=1}^{T_0-1} \sum_{i \in I_k} \frac{m_i(t+1) \tilde{\xi}_k}{N_k f_{k,i}(\boldsymbol{x}_k(t); \tilde{\mathbf{\Phi}}_k)} \right\} \log \xi_k \quad (k = 1, \ldots, K). \end{aligned}$$

Then, by Eqs. (8), (13), (16), and Jensen's inequality, we have

$$\mathcal{L}_K^1(\tilde{\mathbf{\Phi}}_1, \ldots, \tilde{\mathbf{\Phi}}_K; \boldsymbol{\lambda}, T_0) - \mathcal{L}_K^1(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K; \boldsymbol{\lambda}, T_0) \geq \sum_{k=1}^{K} (Q_k(\mathbf{\Phi}_k|\tilde{\mathbf{\Phi}}_k) - Q_k(\tilde{\mathbf{\Phi}}_k|\tilde{\mathbf{\Phi}}_k)).$$

Thus, for each k, the update formula of $\mathbf{\Phi}_k$ is obtained by maximizing $Q_k(\mathbf{\Phi}_k|\tilde{\mathbf{\Phi}}_k)$. It is easily shown that the update formula of $\mathbf{\Phi}_k$ is given by

$$\begin{aligned} \phi_{k,j} &= \frac{1}{\sum_{t=2}^{T_0} n_k(t) + 2N_k + 1} \left\{ \sum_{t=1}^{T_0-1} \sum_{i \in I_k} \frac{m_i(t+1) a_{k,i,j}(t) \tilde{\phi}_{k,j}}{f_{k,i}(\boldsymbol{x}_k(t); \tilde{\mathbf{\Phi}}_k)} + 1 \right\}, \\ \psi_{k,j} &= \frac{1}{\sum_{t=2}^{T_0} n_k(t) + 2N_k + 1} \left\{ \sum_{t=1}^{T_0-1} \sum_{i \in I_k} \frac{m_i(t+1) b_{k,i,j}(t) \tilde{\psi}_{k,j}}{f_{k,i}(\boldsymbol{x}_k(t); \tilde{\mathbf{\Phi}}_k)} + 1 \right\}, \\ \xi_k &= \frac{1}{\sum_{t=2}^{T_0} n_k(t) + 2N_k + 1} \left\{ \sum_{t=1}^{T_0-1} \sum_{i \in I_k} \frac{m_i(t+1) \tilde{\xi}_k}{N_k f_{k,i}(\boldsymbol{x}_k(t); \tilde{\mathbf{\Phi}}_k)} + 1 \right\} \end{aligned} \tag{18}$$

for any $j \in I_k$ and $k = 1, \dots, K$. Hence, we have obtained the algorithm to find the minimal point $(\mathbf{\Phi}_1^*, \dots, \mathbf{\Phi}_K^*)$ of the function $\mathcal{L}_K^1(\mathbf{\Phi}_1, \dots, \mathbf{\Phi}_K; \boldsymbol{\lambda}, T_0)$.

4.2. Estimation of $\boldsymbol{\lambda}$

Next, we describe how to estimate the optimal K division vector $\boldsymbol{\lambda}(K)$ from the data set $\{\boldsymbol{m}(1), \dots, \boldsymbol{m}(T_0)\}$ when the number K of competitive groups in S is given. Basically, we use the following framework:

- For a K division vector $\boldsymbol{\lambda}$, we compute the optimal values $\mathbf{\Theta}(\boldsymbol{\lambda})$ of parameter vector $\mathbf{\Theta}$ by using the method described in Section 4.1.
- Next, we find the optimal K division vector $\boldsymbol{\lambda}(K)$ by minimizing the function $\mathcal{L}_K(\mathbf{\Theta}(\boldsymbol{\lambda}); \boldsymbol{\lambda}, T_0)$ with respect to $\boldsymbol{\lambda}$.

To search the optimal K division vector $\boldsymbol{\lambda}(K)$ in this framework, we can apply the exhaustive search method, the simulated annealing method, and so forth. However, due to the problem of computational quantity, we use such a method that modifies K division vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ to a locally optimal direction by changing each λ_i from 1 to K. Specifically, we estimate the optimal K division vector $\boldsymbol{\lambda}(K)$ by the following algorithm:

```
Step 1. Generate at random a $K$ division vector $\boldsymbol{\lambda}^0$ as the initial division of $S$.
Step 2. Substitute $\boldsymbol{\lambda}^0$ into $\boldsymbol{\lambda}$.
Step 3. For the $K$ division vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$, construct inductively the $K$ division vector $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_N)$
        as follows:
        for $i = 1 : N$
            for $k = 1 : K$
                Set
                    $\boldsymbol{\lambda}_{i,k} = \begin{cases} (k, \lambda_2, \dots, \lambda_N) & (i = 1), \\ (\tilde{\lambda}_1, \dots, \tilde{\lambda}_{i-1}, k, \lambda_{i+1}, \dots, \lambda_N) & (i > 1), \end{cases}$
                if $\boldsymbol{\lambda}_{i,k}$ is a $K$ division vector of $S$.
                    Compute the optimal values $\mathbf{\Theta}(\boldsymbol{\lambda}_{i,k})$ of parameter vector $\mathbf{\Theta}$ for the $K$ division vector $\boldsymbol{\lambda}_{i,k}$ by
                    using the method described in Section 4.1.
                end if
            end for
            Let $\tilde{\lambda}_i$ be the minimal point of the function $\mathcal{L}_K(\mathbf{\Theta}(\boldsymbol{\lambda}_{i,k}); \boldsymbol{\lambda}_{i,k}, T_0)$ with respect to $k$.
        end for
Step 4. Substitute $\tilde{\boldsymbol{\lambda}}$ into $\boldsymbol{\lambda}$.
Step 5. Iterate Steps 3 and 4 until $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}$.
Step 6. Substitute $\tilde{\boldsymbol{\lambda}}$ into $\boldsymbol{\mu}$.
Step 7. For the $K$ division vector $\boldsymbol{\mu}$ of $S$, construct the $K$ division vector $\tilde{\boldsymbol{\mu}}$ of $S$ as follows:
        7.1. For $\forall i, j \in \{1, \dots, N\}$ with $i \neq j$ and $\forall k, \ell \in \{1, \dots, K\}$, let $\boldsymbol{\mu}_{i,k;j,\ell}$ denote the $N$-dimensional vector
             such that the $i$th element is $k$, the $j$th element is $\ell$, and the other elements are equal to those of $\boldsymbol{\mu}$.
        7.2. If $\boldsymbol{\mu}_{i,k;j,\ell}$ is a $K$ division vector of $S$, compute the optimal values $\mathbf{\Theta}(\boldsymbol{\mu}_{i,k;j,\ell})$ of parameter vector $\mathbf{\Theta}$ for
             the $K$ division vector $\boldsymbol{\mu}_{i,k;j,\ell}$ by using the method described in Section 4.1.
        7.3. Let $(i^*, k^*; j^*, \ell^*)$ be the minimal point of the function $\mathcal{L}_K(\mathbf{\Theta}(\boldsymbol{\mu}_{i,k;j,\ell}); \boldsymbol{\mu}_{i,k;j,\ell}, T_0)$ with respect to $(i, j;$
             $k, \ell)$, and set $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_{i^*,k^*;j^*,\ell^*}$.
Step 8. If $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}$, go to Step 9, otherwise return to Step 3.
Step 9. Define the optimal $K$ division vector $\boldsymbol{\lambda}(K)$ of $S$ by setting $\boldsymbol{\lambda}(K) = \tilde{\boldsymbol{\mu}}$.
```

5. Evaluation methods

In this section, we describe how to evaluate the proposed method.

This is done by using both the performance for predicting the one-step future shares of the sites of S and the performance for identifying the competition structure of S. Below, we define how to measure the prediction accuracy of a method and the categorization (clustering) accuracy of a method. Moreover, we introduce conventional methods with which the proposed method should be compared for the performance evaluations.

5.1. Prediction accuracy

We evaluate the performance of a prediction method by exploiting the average prediction error defined in Section 4 to determine the optimal number of competitive groups.

Let $\{\boldsymbol{m}(1), \dots, \boldsymbol{m}(T+1)\}$ be the observed time-series data of visit vectors to S. We would like to evaluate the performance of a method to predict the one-step future shares for this data set. Let us fix an integer ΔT with $1 \leq \Delta T \leq T$, and for each integer T_0 with $T - \Delta T + 1 \leq T_0 \leq T$, consider the task of predicting the share vector $\boldsymbol{m}(T_0+1)/M(T_0+1)$ from the data set $\{\boldsymbol{m}(1), \dots, \boldsymbol{m}(T_0)\}$. To predict the share vector at time-step T_0+1, we define the prediction error $\mathcal{E}(T_0+1)$ of a method by

$$\mathcal{E}(T_0+1) = \frac{1}{2}\sum_{i=1}^{N}\left|\hat{x}_i(T_0+1) - \frac{m_i(T_0+1)}{M(T_0+1)}\right|,$$

where each $\hat{x}_i(T_0+1)$ is the share of site s_i at time-step T_0+1 that the method predicts. Then, we measure the prediction accuracy of the method by the average prediction error $\bar{\mathcal{E}}$,

$$\bar{\mathcal{E}} = \frac{1}{\Delta T}\sum_{T_0=T-\Delta T+1}^{T}\mathcal{E}(T_0+1),$$

that is, we evaluate the performance of the prediction method by $\bar{\mathcal{E}}$. Here, note that $0 \leq \mathcal{E}(T_0+1) \leq 1$ and $0 \leq \bar{\mathcal{E}} \leq 1$.

5.2. Categorization accuracy

To measure the categorization accuracy of a method, we basically exploit the performance measure frequently used to evaluate clustering results. Suppose that the number K of competitive groups in S is known.

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ be the true K division vector of S and let $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_N)$ be the K division vector of S estimated by a method from the time-series data $\{\boldsymbol{m}(1), \dots, \boldsymbol{m}(T_0)\}$. Also, let $S = \cup_{k=1}^{K} C_k$ and $S = \cup_{\ell=1}^{K} \hat{C}_\ell$, respectively, denote the categorizations of S into K competitive groups that correspond to $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}$. Then, we evaluate the performance of the categorization method by the micro-averaged precision $\mathcal{A}P_{T_0}(\hat{\boldsymbol{\lambda}}; \boldsymbol{\lambda})$ defined below.

We define the *micro-averaged precision* $\mathcal{A}P_{T_0}(\hat{\boldsymbol{\lambda}}; \boldsymbol{\lambda})$ of the estimated division $\hat{\boldsymbol{\lambda}}$ to the true division $\boldsymbol{\lambda}$ in the following way: For each $\hat{C}_\ell$, let $C_{k[\ell]}$ denote the set C_k that maximizes the number of visitors to the set $C_k \cap \hat{C}_\ell$ of sites during the period $[1, T_0]$. Furthermore, we understand that the group $\hat{C}_\ell$ for the division $\hat{\boldsymbol{\lambda}}$ corresponds to the group $C_{k[\ell]}$ for the division $\boldsymbol{\lambda}$. Then, the micro-averaged precision $\mathcal{A}P_{T_0}(\hat{\boldsymbol{\lambda}}; \boldsymbol{\lambda})$ is the ratio of the number of visitors to the set $\hat{C}_\ell \cap C_{k[\ell]}$ to the number of visitors to S during the period $[1, T_0]$, that is,

$$\mathcal{A}P_{T_0}(\hat{\boldsymbol{\lambda}}; \boldsymbol{\lambda}) = 100\,\frac{\sum_{\ell=1}^{K}\max_{k=1,\dots,K}\left\{\sum_{t=1}^{T_0}\sum_{i;\lambda_i=k,\hat{\lambda}_i=\ell} m_i(t)\right\}}{\sum_{t=1}^{T_0} M(t)}.$$

Notice that this definition of micro-averaged precision is somewhat different from its usual definition for document clustering [27]. Under this definition, the micro-averaged precision deteriorates if the sites having many visitors are misclassified, though it does not deteriorate by much if the sites having few visitors are misclassified.

5.3. Conventional methods

We introduce the conventional methods with which the proposed method should be compared. First, we describe the conventional methods for predicting the one-step future shares of the sites of S. Second, we describe the conventional method for identifying the competition structure of S.

5.3.1. Prediction task

The prediction task requires predicting the actual share vector $\boldsymbol{x}(T_0+1)$ at time-step T_0+1 from the time-series data $\{\boldsymbol{m}(1), \dots, \boldsymbol{m}(T_0)\}$ of visit vectors to S. For this task, it is not appropriate to apply stochastic process models based on Gaussian noise since they do not necessarily map the simplex Δ^{N-1} onto the simplex Δ^{N-1}; such a model can make the share of a site negative. This implies that straightforward applications of AR models and artificial neural networks will generally not work well. Therefore, as the naive methods with which the proposed method should be compared for this task, we adopt the *naive multinomial method* (NMM) and the *parallel displacement method* (PDM).

First, the NMM is the following: It is a method that supposes each $m_i(t)$ is generated according to some multinomial distribution that is independent of time t. In addition, it estimates the multinomial distribution by the maximal likelihood method from the observed time-series data. It then predicts the share vector at time-step T_0+1 by the mean vector $(\hat{x}_1, \dots, \hat{x}_N)$ of the estimated multinomial distribution.

Next, the PDM is described as follows: This method predicts the share vector $(\hat{x}_1(T_0+1), \dots, \hat{x}_N(T_0+1))$ at time-step T_0+1 by the observed share vector $(m_1(T_0)/M(T_0), \dots, m_N(T_0)/M(T_0))$ at time-step T_0. Here, we remark that the PDM is a promising approach for the prediction task on highly fluctuating time-series data.

5.3.2. Categorization task

It is often necessary to categorize a set of Web sites into subsets of related sites. Several methods have been proposed to extract topically related sites and communities from the data for texts and hyperlinks of the sites [7,21,8,12–14]. However, in our categorization task, we must extract groups of competitors based on the dynamical data for users. More specifically, the categorization task requires dividing S into K competitive groups from the observed time-series data $\{\boldsymbol{m}(1), \dots, \boldsymbol{m}(T_0)\}$ of visit vectors to S when the number K of divisions is given. For this task, it is not appropriate to apply those previous methods that exploit static data for the sites without using user data.

In the field of Econophysics [24], Mantegna [23] successfully extracted groups of stocks belonging to the same types of industry from the time-series data of stock prices in a financial market. In general, the stocks belonging to the same type of industry can be considered to compete. Also, the number of visitors to a site can be regarded as a proxy for that site's success in the Web market [2,25]. Therefore, by relating the time-series of the number of visitors to the sites of S with the time-series of stock prices in a financial market, we consider applying Mantegna's method [23] to our categorization task.

Mantegna's method operates as follows: The method identifies each site s_i with the point $\boldsymbol{w}_i = (w_i(1), \dots, w_i(T_0-1))$ on the (T_0-2)-dimensional sphere of radius $\sqrt{T_0-1}$ with center of the origin in the Euclidean space $\mathbb{R}^{T_0-1}$, and constructs the dendrogram of the points $\boldsymbol{w}_1, \dots, \boldsymbol{w}_N$ based on the Euclidean distance of $\mathbb{R}^{T_0-1}$. Here, for $i=1, \dots, N$ and $t=1, \dots, T_0-1$,

$$w_i(t) = \frac{y_i(t)-\mu_i}{\sigma_i}, \qquad y_i(t) = \log m_i(t+1) - \log m_i(t),$$

where each μ_i and σ_i^2 are, respectively, the mean and variance of the data $\{y_i(1), \dots, y_i(T_0-1)\}$.

Since our categorization task requires dividing S into K sets, we adopt the following method as the naive method with which the proposed method should be compared: The method first identifies each site $s_i \in S$ with the points $\boldsymbol{w}_i$ on the sphere according to Mantegna's method. Next, it divides the points $\boldsymbol{w}_1, \dots, \boldsymbol{w}_N$ into K sets by the spherical K means method that is frequently used for clustering of text data sets [10].

6. Experimental evaluation

By conducting experiments using synthetic and real data, we examine the effectiveness of the proposed method.

6.1. *Evaluation for synthetic data*

By carrying out experiments using synthetic data, we confirmed that the proposed method works well, and below, we present one of those experimental results.

6.1.1. *Synthetic data*

We consider identifying the following model:

- The number of sites in market S is 10, that is, $N = 10$.
- The number of competitive groups in S is 3, that is, $K = 3$.
- The three division vector $\boldsymbol{\lambda}$ of S is randomly generated.
- The parameter vector $\boldsymbol{\Theta}$ is randomly generated.

Based on [2], we assume that the total number $M(t)$ of visitors to S at time-step t is determined through the omission of fractions from the stochastic dynamics $M(t+1) - M(t) = \nu(t)M(t)$ for $\forall t \geq 0$, where each $\nu(t)$ is independently generated from the Gaussian distribution with mean ν_0 and variance σ^2. In our experiment, we used $\nu_0 = 0$, $\sigma = 0.1$ and $M(0) = 500$.

We generated the time-series data $\{\boldsymbol{m}(1), \dots, \boldsymbol{m}(51)\}$ of visit vectors to S from this model by randomly choosing the initial visit vector $\boldsymbol{m}(1) = (m_1(1), \dots, m_{10}(1))$ such that $\sum_{i=1}^{10} m_i(1) = M(1)$. Note that $T = 50$ in this experiment. Fig. 3 displays the 10 time-series $m_1(t), \dots, m_{10}(t)$ $(t = 1, \dots, 51)$, where a line-style indicates a competitive group. Also, Fig. 4 displays the time-series $M(t)$ $(t = 1, \dots, 51)$.

6.1.2. *Performance evaluation*

Using this synthetic data, we evaluated the effectiveness of the proposed method for both the prediction task and the categorization task.

First, we investigated the prediction performance for the proposed method and the conventional methods. Table 1 displays average prediction error $\bar{\mathcal{E}}$ for the proposed method, the NMM and the PDM, where $\Delta T = 10$ was used. We applied the proposed method for $K = 1, 2, 3, 4$ and 5. We observe from Table 1 that the proposed method selected the model with $K = 3$, that is, it found that the number of competitive groups in S is three. This result and Table 1 show that the proposed method could predict the one time-step future shares of the sites with an accuracy of about 96% on average, and outperform the conventional methods.

Next, once $K = 3$ was known, we investigated the categorization performance for the proposed method and the conventional method. In Fig. 5, for both methods, we plot the micro-averaged precision $\mathcal{A}P_{T_0}(\hat{\boldsymbol{\lambda}}; \boldsymbol{\lambda})$ of the estimated division $\hat{\boldsymbol{\lambda}}$ to the true division $\boldsymbol{\lambda}$ with respect to the final time-step T_0 of training time-series data. Fig 5 shows that the proposed method could be more stable and accurate than the conventional method. Moreover, since the proposed method selected $K = 3$, it is seen from Fig. 5 that the proposed method could perfectly identify the competition structure of the true model.

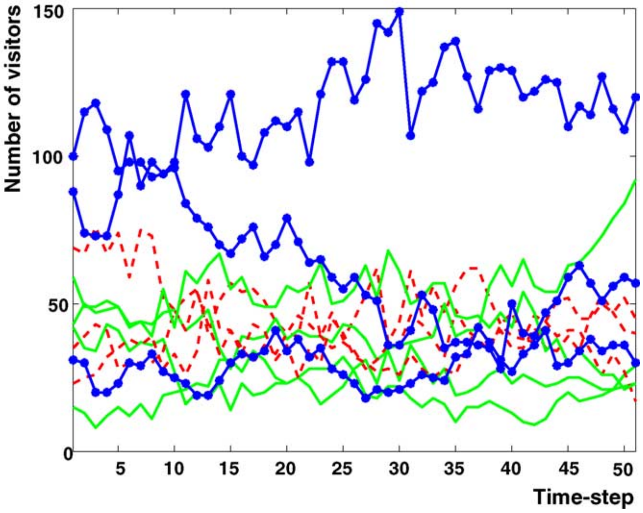

Fig. 3. Fluctuations in the numbers of visitors to the sites of *S* for synthetic data.

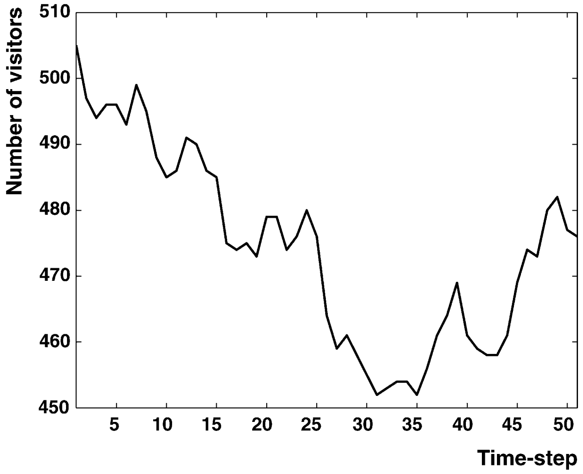

Fig. 4. Fluctuations in the total number of visitors to the sites of *S* for synthetic data.

Table 1
Prediction performance for synthetic data

	Prediction error
$K = 1$	0.049
$K = 2$	0.046
$K = 3$	0.044
$K = 4$	0.050
$K = 5$	0.051
NMM	0.117
PDM	0.048

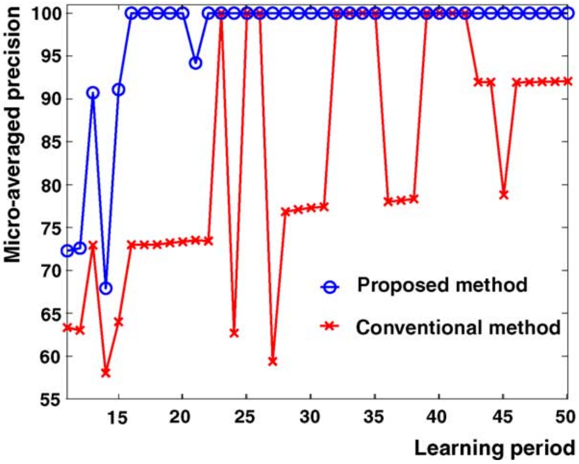

Fig. 5. Categorization performance for synthetic data.

Hence, we observed that the proposed method can successfully identify the true competition structure of S, and fairly predict the one-step future shares of the sites. In particular, we observed that it can outperform the conventional methods for both the prediction task and the categorization task.

6.2. Evaluation for real Web data

We evaluate the performance of the proposed method using real Web data.

6.2.1. Real web data

Fig. 6 displays the time-series data $m_1(t), \ldots, m_{20}(t)\,(t = 1, \ldots, 51)$, obtained from the usage logs for 20 Japanese Web sites that offer streaming video contents.[1] Here, one time-step equals 2 days. Note that $N = 20$ and $T = 50$. Fig. 7 displays the time-series $M(t)\,(t = 1, \ldots, 51)$. From Fig. 7, we observe that the number of total visitors to this market fluctuate heavily.

6.2.2. Performance evaluation

First, we investigated the prediction performance for the proposed method and the conventional methods. Table 2 displays average prediction error $\bar{\mathcal{E}}$ for the proposed method, the NMM and the PDM, where $\Delta T = 10$ was used. We applied the proposed method to models for $K = 1, 2, 3, 4$ and 5. We observe from Table 2 that the proposed method selected the model with $K = 3$; that is, it concluded that the number of competitive groups in S is three. This result and Table 2 show that the proposed method could predict the one-step future shares of the sites with an accuracy of about 80% on average, and outperform the conventional methods.

Next, when $K = 3$ was specified, we investigated the categorization performance for the proposed method and the conventional method. In the case of real Web data, we cannot know the true competition structure. Thus, for each method, we used as its true division, the division $\boldsymbol{\lambda}$ estimated from the whole time-series data $\{\boldsymbol{m}(1), \ldots, \boldsymbol{m}(51)\}$. We compared the proposed method and the conventional method for the stability of division. When we extract the competition structure of S from time-series data, we consider that the extracted structure should be stable for the

[1] This data was obtained from the public trial service of “Broadband Contents Guide” that Nippon Telegraph and Telephone Corporation (NTT) performed in 2002 in cooperation with Tokyo News Service, Ltd. [1].

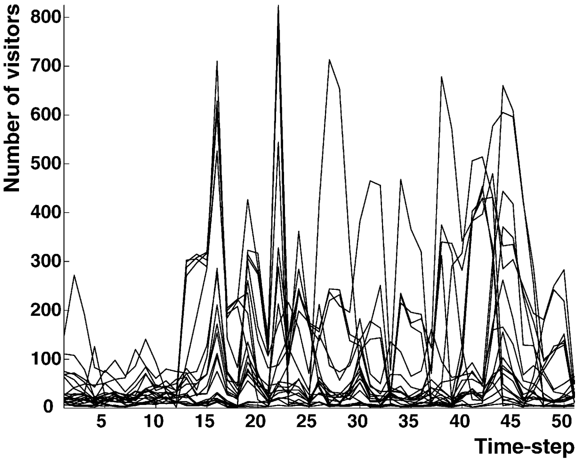

Fig. 6. Fluctuations in the numbers of visitors to the sites of *S* for real data.

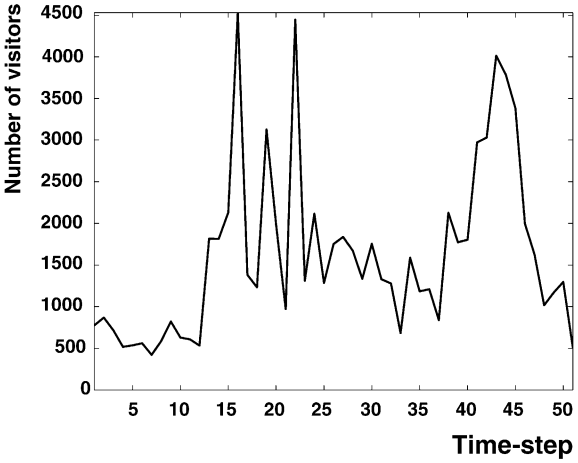

Fig. 7. Fluctuations in the total number of visitors to the sites of *S* for real data.

Table 2
Prediction performance for real data

	Prediction error
$K = 1$	0.217
$K = 2$	0.223
$K = 3$	0.206
$K = 4$	0.208
$K = 5$	0.208
NMM	0.299
PDM	0.217

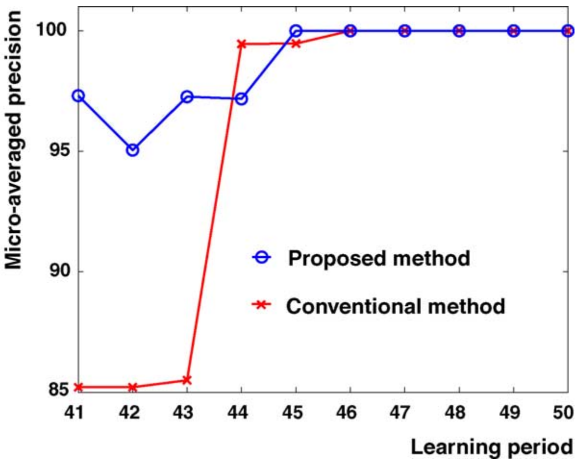

Fig. 8. Categorization performance for real data.

time-axis. In Fig. 8, for each method, the micro-averaged precision $\mathcal{AP}_{T_0}(\hat{\boldsymbol{\lambda}}; \boldsymbol{\lambda})$ of the estimated division $\hat{\boldsymbol{\lambda}}$ to its true division $\boldsymbol{\lambda}$ is plotted with respect to the final time-step T_0 of training time-series data. Fig. 8 shows that the proposed method could be more stable than the conventional method.

Hence, we observed that the proposed method works reasonably well and can also outperform the conventional methods for both the prediction task and the categorization task.

Table 3
The categorization result by the proposed method

ID	Web site description
Group 1	
1	A site for popular music-videos
2	An Internet-TV-broadcasting site managed by a satellite TV-broadcasting company
3	An information site for lifestyle and shopping in Shibuya, Tokyo
4	A site for event information
5	A site for sports entertainments
6	An information site for table tennis
7	A site for interactive dramas
Group 2	
8	An information site for next-generation sports and classical culture, managed by a video-production company
9	An information site for ecological goods
10	An information site for music
11	An internet-broadcasting site managed by a radio-broadcasting company
12	An internet-broadcasting site for a local community
13	A site for artistic dramas, managed by an old movie company
14	A site for mystery dramas
Group 3	
15	A general information site managed by a newspaper publishing company
16	A general information site managed by an IT-news company
17	A general information site managed by a communication-service company
18	A general information site managed by a local IT company
19	An education site for children
20	An information site for hair restoration

Table 4
The categorization result by the conventional method

ID	Web site description
Group 1	
1	A site for popular music-videos
2	An Internet-TV-broadcasting site managed by a satellite TV-broadcasting company
3	An information site for lifestyle and shopping in Shibuya, Tokyo
4	A site for event information
5	A site for sports entertainments
14	An internet-broadcasting site for a local community
16	A general information site managed by an IT-news company
17	A general information site managed by a communication-service company
18	A general information site managed by a local IT company
19	An education site for children
20	An information site for hair restoration
Group 2	
6	An information site for table tennis
8	An information site for next-generation sports and classical culture, managed by a video-production company
9	An information site for ecological goods
10	An information site for music
11	An internet-broadcasting site managed by a radio-broadcasting company
Group 3	
7	A site for interactive dramas
13	A site for artistic dramas, managed by an old movie company
14	A site for mystery dramas
15	A general information site managed by a newspaper publishing company

6.2.3. *Qualitative evaluation*

Now, we qualitatively compared the proposed method and conventional method for the categorization task when $K = 3$ was specified. Tables 3 and 4, respectively, indicate the categorization results by the proposed method and conventional method from all the observed time-series data. For the categorization result of the proposed method, we can observe from Table 3 that Group 1 is a group of popular entertainment sites, Group 2 is a group of artistic hobby sites, and Group 3 is a group of knowledge-providing sites. Since these groups can be considered to have distinct user classes, we observe that the proposed method suggested a reasonable competition structure. On the other hand, Table 4 shows that the conventional method mixed together the group of popular entertainment sites and the group of knowledge-providing sites in Group 1.

7. Concluding remarks

The Web is a complex system that changes over time, and highly expected to understand its inherent structures. In this paper, we explored the problem of modeling fluctuations in the number of visitors to Web sites that form a market in terms of competitive dynamics. We proposed a probabilistic mixture model of multinomial distributions, each of whose parameter values are estimated using a replicator equation. We constructed an effective algorithm for both identifying groups of competitive sites and estimating the underlying replicator equation for each group from observed fluctuations of user populations for the sites. We implemented this method to both categorize the sites into groups of competitors and to predict the one-step future population shares of the sites based on these observations. We experimentally showed that the proposed method can outperform the conventional methods for both the prediction task and the categorization task. Using synthetic data, we experimentally confirmed that the proposed method successfully identifies the true model structure, and fairly predicts the one-step future shares of the sites. Furthermore, using real data from the usage logs of 20 Japanese Web sites that offer streaming video contents, we experimentally demonstrated that the proposed method could suggest a reasonable competition structure and predict the one-step future shares of the sites with an accuracy of about 80% on average. Namely, we showed that the proposed method can construct an effective predictive model of short-term fluctuations in the number of visitors to the sites based on observations.

Our research aims to model the usage dynamics of Web sites from game-theoretic points of view. Applying the framework of dynamical systems game [3] to this problem will be one promising direction for future research. Furthermore, extensive verification of the proposed method with various real Web data remains an important task. However, we have already made substantial progress, and we are encouraged by the initial results of our efforts to model the Web dynamics.

Acknowledgements

The authors would like to thank the members of Intelligent Media Project, NTT Cyber Solutions Laboratories, for providing the data obtained in the public trial service of “Broadband Contents Guide” [1].

Appendix A. Proof of Proposition 2

Let $Z = \{z_j; j \in J\}$ be a finite set. We fix a positive integer V. We consider sampling V elements from set Z with replacement according to the multinomial distribution with parameters $\{q_j; j \in J\}$, where $0 \leq q_j \leq 1$ $(j \in J)$. Then, the probability $P((v_j)_{j\in J})$ that each z_j is sampled v_j times when V elements are sampled from Z with replacement is obtained by

$$P((v_j)_{j\in J}) = \frac{V!}{\prod_{j\in J} v_j!} \prod_{j\in j} {q_j}^{v_j}.$$

Let $\langle X\rangle_V$ denote the expectation of X when trials are run V times under this multinomial distribution. Then, we have the next lemma.

Lemma A.1. *Suppose that V elements are sampled from set Z according to the multinomial distribution* $\{q_j; j \in J\}$. *For each* $j \in J$, let v_j *denote the number that* z_j *is sampled. Then the following equations hold:*

$$\left\langle \frac{v_j}{V} \right\rangle_V = q_j \quad (j \in J) \tag{A.1}$$

$$\left\langle \left(\frac{v_j}{V} - \left\langle \frac{v_j}{V} \right\rangle \right)^2 \right\rangle_V = \frac{q_j(1-q_j)}{M} \quad (j \in J). \tag{A.2}$$

Proof. First, let us prove Eq. (A.1). We have

$$\left\langle \frac{v_j}{V} \right\rangle_V = \frac{1}{V} \sum_{\ell=0}^{V} \ell P(v_j = \ell) = \frac{1}{V} \sum_{\ell=1}^{V} \ell \sum_{\substack{v_i \geq 0,\, (i\neq j);\\ \ell + \sum_{i\neq j} v_i = V}} \frac{V!}{\ell! \prod_{i\neq j} v_i!} {q_j}^{\ell} \prod_{i\neq j} {q_i}^{v_i}$$

$$= \frac{1}{V} \sum_{\substack{w_i \geq 0\, (\forall i);\\ \sum_i w_i = V-1}} \frac{V(V-1)!}{\prod_i w_i!} q_j \prod_i {q_i}^{w_i} = q_j \left(\sum_{j\in J} q_j \right)^{V-1} = q_j.$$

Hence, Eq. (A.1) holds.

Next, we prove Eq. (A.2). If $V = 1$, it is trivial. Let us suppose that $V > 1$. Then, we have

$$\left\langle \left(\frac{v_j}{V}\right)^2 \right\rangle_V = \frac{1}{V^2}\sum_{\ell=0}^{V} \ell^2 P(v_j=\ell) = \frac{1}{V^2} \sum_{\substack{w_i \geq 0\,(\forall i);\\ \sum_i w_i = V-1}} \frac{(w_j+1)V(V-1)!}{\prod_i w_i!} q_j \prod_i {q_i}^{w_i}$$

$$= \frac{{q_j}^2}{V} \sum_{\substack{w_i \geq 0\,(\forall i);\\ \sum_i w_i = V-2}} \frac{(V-1)(V-2)!}{\prod_i w_i!} \prod_i {q_i}^{w_i} + \frac{q_j}{V} \sum_{\substack{w_i \geq 0\,(\forall i);\\ \sum_i w_i = V-1}} \frac{(V-1)!}{\prod_i w_i!} \prod_i {q_i}^{w_i}$$

$$= \frac{(V-1){q_j}^2}{V} \left(\sum_{j \in J} q_j\right)^{V-2} + \frac{q_j}{V} \left(\sum_{j \in J} q_j\right)^{V-1} = \frac{(V-1){q_j}^2 + q_j}{V}.$$

Hence, we can easily obtain Eq. (A.2) from $(\langle v_j/V - \langle v_j/V \rangle_V \rangle_V)^2 = \langle (v_j/V)^2 \rangle_V - {\langle v_j/V \rangle_V}^2$ and Eq. (A.1).

We have completed the proof of Lemma A.1. □

Proposition 2 follows immediately from Lemma A.1.

References

[1] S. Abe, S. Miyahara, Y. Hayashi, Y. Tonomura, Map-like content guide system "AssociaGuide": The Report of the Public Trial Service "Broadband Content Guide", ITE Technical Report 26, No. 81, 2002, pp. 1–4 (in Japanese).

[2] L.A. Adamic, B.A. Huberman, The nature of markets in the World Wide Web, Quart. J. Electr. Commerce 1 (2000) 5–12.

[3] E. Akiyama, K. Kaneko, Dynamical systems game theory and dynamics of games, Physica D 147 (2000) 221–258.

[4] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.

[5] C. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Oxford, 1995.

[6] L. Cao, A. Mees, K. Judd, Dynamics from multivariate time series, Physica D 121 (1998) 75–88.

[7] S. Chakrabarti, B. Dom, R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, J. Kleinberg, Mining the Web's link structure, IEEE Comput. 32 (8) (1999) 60–67.

[8] D. Cohn, T. Hofmann, The missing link: a probabilistic model of document content and hypertext connectivity, Adv. Neural Inform. Process. Syst. 13 (2001) 430–436.

[9] A.P. Dempster, N.M. Larid, D.B. Rubin, Maximum likelihood from incomplete data via EM algorithm, J. Roy. Statist. Soc. B 39 (1977) 1–38.

[10] I.S. Dhillon, D.S. Modha, Concept decompositions for large sparse text data using clustering, Mach. Learn. 42 (2001) 143–175.

[11] S.N. Dorogovtsev, J.F.F. Mendes, Evolution of networks, Adv. Phys. 51 (2002) 1079–1187.

[12] J.-P. Eckmann, E. Moses, Curvature of co-links uncovers hidden thematic layers in the World Wide Web, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 5825–5829.

[13] G.W. Flake, S. Lawrence, C.L. Giles, F.M. Coetzee, Self-organization and identification of Web communities, IEEE Comput. 35 (3) (2002) 66–71.

[14] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 7821–7826.

[15] J. Hofbauer, K. Sigmund, Evolutionary Games and Population Dynamics, Cambridge University Press, Cambridge, 1998.

[16] B.A. Huberman, L.A. Adamic, Growth dynamics of the World-Wide Web, Nature 401 (1999) 131.

[17] B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow, R.M. Lukose, Strong regularities in World Wide Web surfing, Science 280 (1998) 95–97.

[18] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, 1997.

[19] M. Kimura, K. Saito, N. Ueda, Modeling of growing networks with directional attachment and communities, Neural Networks 17 (2004) 975–988.

[20] J. Kleinberg, S. Lawrence, The structure of the Web, Science 294 (2001) 1849–1850.

[21] R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, Extracting large-scale knowledge bases from the web, in: Proceedings of the 25th International Conference on Very Large Data Bases, 1999, pp. 639–650.

[22] J. Liu, S. Zhang, J. Yang, Characterizing Web usage regularities with information foraging agents, IEEE Trans. Knowledge Data Eng. 16 (2004) 566–584.

[23] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1999) 193–197.

[24] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, 2000.

[25] S.M. Maurer, B.A. Huberman, Competitive dynamics of Web sites, J. Econ. Dynam. Contr. 27 (2003) 2195–2206.

[26] S.K. Pal, V. Talwar, P. Mitra, Web mining in soft computing framework: relevance, state of the art and future directions, IEEE Trans. Neural Netw. 13 (2002) 1163–1177.

[27] N. Slonim, N. Friedman, N. Tishby, Unsupervised document classification using sequential information maximization, in: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2002, pp. 129–136.

[28] S.H. Strogatz, Exploring complex networks, Nature 410 (2001) 268–276.