International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:7, No:1, 2013

A Simplified Higher-Order Markov Chain Model

Chao Wang, Ting-Zhu Huang, Chen Jia

Abstract**—**In this paper, we present a simplified higher-order Markov chain model for multiple categorical data sequences also called as simplified higher-order multivariate Markov chain model. The number of the parameters of the new model is only $O((n+s)sm^2)$ which is less than $O((ns^2m^2)$ in the higher-order multivariate Markov chain model. Numerical experiments illustrate the benefits of our new model.

Keywords**—**Higher-order multivariate Markov chain model, Categorical data sequences, Multivariate Markov chain.

I. Introduction

MARKOV chains are of interest in a wide range of applications, for example, telecommunication systems, manufacturing systems and inventory systems, see for instance [6] and the references therein. In recent years, the prediction of categorical data sequences [4] has become more and more useful in many real world applications such as sales demand predictions [1]-[3], DNA sequencing [5] and credit data modeling [8]-[11]. Improving the models for exploring these relationships among the given categorical data sequences is an important research area in these years.

Different models have been proposed for multiple categorical data sequences predictions. A multivariate Markov chain model has been presented in [1]. They constructed a new matrix by means of the transition probability matrices among different sequences. For effectiveness, an improved multivariate Markov chain model has been exhibited to speed up the convergent rate of computing the stationary or steady state solutions. In the improved model, they incorporated the positive and negative association parts [8]. The number of the parameters is $O((sm^2+s^2)$. A more advanced, namely, higher-order multivariate Markov chain model has been studied in [7]. Moreover, there are some other papers contribute to the multivariate Markov chain model, e.g., [5], [13]-[15], [17].

However, with the development of science technologies and their applications, the data sequences will be longer, the prediction results needed to be more precise. It is inevitable that the computation from a large categorical data sequence group will cause high computational cost. Now, the higher-order multivariate Markov chain model performs the best in the prediction of multivariate discrete-time Markov chains. Thus, it is useful to simplify the higher-order multivariate Markov chain model. For the above purpose, we propose a simplified higher-order multivariate Markov chain model in this paper.

The rest of the paper is organized as follows. In Section II, we review the Markov chain model [3], the multivariate Markov chain model [1], the higher-order multivariate Markov chain model [7] and two lemmas [16], [18]. In Section III, we present a simplified higher-order multivariate Markov chain model for multiple categorical data sequences. Moreover, some properties of the simplified higher-order multivariate Markov chain model are also analyzed. Section IV gives estimation method for the parameters of the simplified higher-order multivariate Markov chain model. Numerical experiments on two examples demonstrate the benefits of our new model in Section V. Finally, concluding remarks are given in Section VI.

II. A Review on the Markov Chain Models

In this section, we briefly introduce some Markov chain models and several lemmas, e.g., the Markov chain model, the multivariate Markov chain model, the higher-order multivariate Markov chain model, and the Perron-Frobenius theorem.

A. The Markov Chain Model

First, we introduce some definitions of the Markov chain from [1], [12]. Consider the state set of the categorical data sequence be $M=\{1,2,\cdots m\}$. The discrete-time Markov chain with m states satisfies the following relationships:

$$\begin{aligned}&\mathrm{Prob}(x_{t+1}=\theta_{t+1}\mid x_0=\theta_0,x_1=\theta_1,\cdots,x_t=\theta_t)\\&=\mathrm{Prob}(x_{t+1}=\theta_{t+1}\mid x_t=\theta_t),\end{aligned}$$

where $\theta_t\in M$, $t\in\{1,2,\cdots\}$. The conditional probabilities $\mathrm{Prob}(x_{t+1}=\theta_{t+1}\mid x_t=\theta_t)$ are called the one-step transition probabilities of the Markov chain. If we rewrite the transition probabilities as

$$p_{ij}=\mathrm{Prob}(x_{t+1}=\theta_{t+1}\mid x_t=\theta_t),\ i,j\in M, \tag{1}$$

the Markov chain model can be represented as follows:

$$X_{t+1}=PX_t,$$

where

$$P=[p_{ij}],0\le p_{ij}\le 1,\sum_{i=1}^{m}p_{ij}=1,\forall i,j\in M, \tag{2}$$

and X_0 is the initial probability distribution, $X_t=(x_t^1,x_t^2,\cdots x_t^m)^T$ is the state probability distribution.

B. The Multivariate Markov Chain Model

A multivariate Markov chain model has been proposed in [1]

Chao Wang and Ting-Zhu Huang are with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, P.R. China (e-mail: wangchao1321654@163.com, tingzhuhuang@126.com).
Chen Jia is with School of Accounting, Jilin University of Finance and Economics, Changchun, Jilin, 611731, P.R. China, (e-mail: chenchen19890707@126.com).

where the number of categorical data sequences is $s>1$ and it has the following form:

$$x_{r+1}^{(j)}=\sum_{k=1}^{s}\lambda_{jk}P^{(jk)}x_r^{(k)},\ \forall j\in\{1,2,\cdots,s\},\ r\in\{0,1,\cdots\} \quad (3)$$

where

$$\lambda_{jk}\geq 0,\sum_{k+1}^{s}\lambda_{jk}=1,\forall j,k\in\{1,2,\cdots,s\} \quad (4)$$

Here, $x_0^{(j)}$ is the initial probability distribution of the *j*th sequence. $x_r^{(k)}$ is the state probability distribution of the *k*th sequence at time *r*. $x_{r+1}^{(j)}$ is the state probability distribution of the *j*th sequence at time $r+1$. Here, $P^{(jk)}$ is the one-step transition probability matrix from the states in the *k*th sequence at time t to the states in the *j*th sequence at time $t+1$. In matrix form, we have

$$\begin{pmatrix} x_{r+1}^{(1)} \\ x_{r+1}^{(2)} \\ \vdots \\ x_{r+1}^{(s)} \end{pmatrix}=\begin{pmatrix} \lambda_{11}P^{(11)} & \lambda_{12}P^{(12)} & \cdots & \lambda_{1s}P^{(1s)} \\ \lambda_{21}P^{(21)} & \lambda_{22}P^{(22)} & \cdots & \lambda_{2s}P^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}P^{(s1)} & \lambda_{s2}P^{(s2)} & \cdots & \lambda_{ss}P^{(ss)} \end{pmatrix}\begin{pmatrix} x_r^{(1)} \\ x_r^{(2)} \\ \vdots \\ x_r^{(s)} \end{pmatrix}, \quad (5)$$

where $P^{(jk)}$ can be obtained directly from the s categorical data sequences and the parameters λ_{jk} can be got by the linear programming, details in [1].

C. The Higher-Order Multivariate Markov Chain Model

Let the number of categorical data sequences be $s>1$, m be the number of states in every sequences. The higher-order multivariate Markov chain model [7] can be presented as follows:

$$x_{r+1}^{(j)}=\sum_{k=1}^{s}\sum_{h=1}^{n}\lambda_{j,k}^{(h)}P_h^{(j,k)}x_{r-h+1}^{(k)},\forall r=n-1,\cdots \quad (6)$$

where $x_0^{(1)},x_0^{(2)},\cdots x_0^{(n)}$ are the initial probability distributions , and $\lambda_{j,k}^{(h)}$ satisfies

$$\sum_{k=1}^{s}\sum_{h=1}^{n}\lambda_{j,k}^{(h)}=1,\lambda_{j,k}^{(h)}\geq 0,\forall 1\leq j,k\leq s,1\leq h\leq n. \quad (7)$$

Here, $x_{r-h+1}^{(k)}$ is the state probability distribution of the *k*th sequence at time $r-h+1$, $P_h^{(j,k)}$ is the *h*th-step transition probability matrix from the states in the *k*th sequence at time $t=r-h+1$ to the states in the *j*th sequence at time $t=j-1$, $x_{r+1}^{(j)}$ is the state probability distribution of the *j*th sequence at time $r+1$. Moreover, it has the following form:

$$X_{r+1}^{(j)}=((x_{r+1}^{(j)})^T,x_r^{(j)})^T,\cdots,x_{r-n+2}^{(j)})^T,)^T\in R^{nm\times 1}. \quad (8)$$

In matrix form, we have

$$\begin{pmatrix} X_{r+1}^{(1)} \\ X_{r+1}^{(2)} \\ \vdots \\ X_{r+1}^{(s)} \end{pmatrix}=\begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \cdots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \cdots & B^{(1,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \cdots & B^{(s,s)} \end{pmatrix}\begin{pmatrix} X_r^{(1)} \\ X_r^{(2)} \\ \vdots \\ X_r^{(s)} \end{pmatrix}, \quad (9)$$

where

$$B=\begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \cdots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \cdots & B^{(1,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \cdots & B^{(s,s)} \end{pmatrix}_{mns\times mns},$$

If $j=k$,

$$B^{(j,k)}=\begin{pmatrix} \lambda_{j,k}^{(1)}P_1^{j,k} & \lambda_{j,k}^{(2)}P_2^{j,k} & \cdots & \lambda_{j,k}^{(n)}P_n^{j,k} \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{pmatrix}_{mn\times mn},$$

else,

$$B^{(j,k)}=\begin{pmatrix} \lambda_{j,k}^{(1)}P_1^{j,k} & \lambda_{j,k}^{(2)}P_2^{j,k} & \cdots & \lambda_{j,k}^{(n)}P_n^{j,k} \\ 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix}_{mn\times mn}$$

Each column sum of the iterative matrix B is not necessary equal to one while each the column sum of transition probability matrices $P_h^{(j,k)}$ is equal to one.

For analyzing the properties of the simplified higher-order multivariate Markov chain model that will be presented in next section, two Lemmas are given at first.

Lemma 1. (Perron-Frobenius Theorem [12]) Let $A\in R^{m\times m}$ be a non-negative and irreducible matrix. Then;

(1). A has a positive real eigenvalue λ equal to its spectral radius, i.e., $\lambda=\max_k|\lambda_k(A)|$ where $\lambda_k(A)$ denotes the k th eigenvalue of A.

(2). To λ there is an eigenvector z of its entries being real and positive, such that $Az=\lambda z$.

(3). λ is a simple eigenvalue of A.

Lemma 2. ([15]) Let B be a iterative matrix of multivariate Markov chain model and X_t be a state distribution at time *t*. If B is irreducible and aperiodic, then there is a unique stationary distribution π satisfying $\pi = P\pi$ and $\lim_{t\to\infty} X_t = \pi.$

III. Simplified Higher-Order Multivariate Markov Chain Model

In this section, for reducing the number of the parameters, we propose a simplified higher-order multivariate Markov chain model for *s* categorical sequences and briefly discuss some properties of the new model.

In this new model, the state probability distribution of the *j*th sequence at time $t+r+1$ depends on the state probability distribution of all the sequences at time $t = r, r-1, \cdots, r-n+1$. For $\forall j,k \in \{1,\cdots,s\}$, $r \in \{n, n-1, \cdots\}$ the simplified higher-order multivariate Markov chain model is given as follows:

$$x_{r+1}^{(j)} = \sum_{h=1, i=j}^{n} \lambda_{j,k}^{(h)} P_n^{(j,k)} x_{r-h+1}^{(k)} + \sum_{h=1, i\neq j}^{s} \lambda_{j,k}^{(h)} P_n^{(j,k)} x_{r-h+1}^{(k)},$$

where $x_0^{(1)}, x_0^{(2)}, \cdots, x_0^{(s)}$ are the initial probability distributions, $x_{r+1}^{(j)}, x_{r-h+1}^{(k)}, P_h^{(j,k)}$ have been defined in Section II.*C* and $\lambda_{j,k}^{(h)}$ satisfies

$$\sum_{h=1, i=j}^{n} \lambda_{j,k}^{(h)} + \sum_{h=1, i\neq j}^{s} \lambda_{j,k}^{(h)} = 1,\ \lambda_{j,k}^{(h)} \geq 0, \tag{10}$$

where $\forall j,k \in \{1,\cdots,s\}$, $h \in \{1,2,\cdots,n\}$.

Assuming that

$$X_{r+1}^{(j)} = ((x_{r+1}^{(j)})^T, (x_r^{(j)})^T, \cdots, (x_{r-n+2}^{(j)})^T)^T \in R^{nm\times 1}, \tag{11}$$

where $\forall j = 1,2,\cdots,s.$ In matrix form, the simplified higher-order multivariate Markov chain model can be presented as

$$\begin{pmatrix} X_{r+1}^{(1)} \\ X_{r+1}^{(2)} \\ \vdots \\ X_{r+1}^{(s)} \end{pmatrix} = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \cdots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \cdots & B^{(1,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \cdots & B^{(s,s)} \end{pmatrix} \begin{pmatrix} X_r^{(1)} \\ X_r^{(2)} \\ \vdots \\ X_r^{(s)} \end{pmatrix}, \tag{12}$$

where

$$B = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \cdots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \cdots & B^{(1,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \cdots & B^{(s,s)} \end{pmatrix}_{mns\times mns}, \tag{13}$$

If $j = k$,

$$B^{(j,k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{j,k} & \lambda_{j,k}^{(2)} P_2^{j,k} & \cdots & \lambda_{j,k}^{(n)} P_n^{j,k} \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{pmatrix}_{mn\times mn} \tag{14}$$

else,

$$B^{(j,k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{j,k} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix}_{mn\times mn}. \tag{15}$$

Each the column sum of transition matrix $P_h^{(j,k)}$ is equal to one. The number of the parameters of the proposed new model is only $O(n+s)sm^2$, which is less than $O(ns^2m^2)$ of the higher-order multivariate Markov chain model.

Next, some properties of the simplified higher-order multivariate Markov chain model are discussed.

Theorem 1: In the simplified higher-order multivariate Markov chain model, if $\lambda_{j,k}^{(h)} \geq 0$ for $\forall j,k \in \{1,\cdots,s\}$, $h \in \{1,2,\cdots,n\}$, then the matrix B has an eigenvalue equal to one and the modulus of all its eigenvalues are less than or equal to one.

Proof: Let

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \cdots & \Lambda_{1s} \\ \Lambda_{21} & \Lambda_{22} & \cdots & \Lambda_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ \Lambda_{s1} & \Lambda_{s2} & \cdots & \Lambda_{ss} \end{pmatrix}, \tag{16}$$

where

$$\Lambda_{i,i} = \begin{pmatrix} \lambda_{i,i}^{(n)} & 1 & \cdots & 0 \\ \lambda_{i,i}^{(n-1)} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & 1 \\ \lambda_{i,i}^{(1)} & 0 & \cdots & 0 \end{pmatrix},\ \forall i \in \{1,\cdots,s\},$$

and if $i \neq j$,

$$\Lambda_{i,j} = \begin{pmatrix} \lambda_{i,j}^{(n)} & 0 & \cdots & 0 \\ \lambda_{i,j}^{(n-1)} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{i,j}^{(1)} & 0 & \cdots & 0 \end{pmatrix}, \ \forall i, j \in \{1, \cdots, s\}.$$

From (10), the matrix Λ is non-negative and its each column sum is equal to one. Using the properties of connection, Λ is irreducible. According to Lemma 2, there exists a positive vector

$$y = [y_1^{(n)}, y_1^{(n-1)}, \cdots, y_1^{(1)}, \cdots, y_s^{(n)}, y_1^{(n-1)}, \cdots y_s^{(1)}]^T,$$

such that

$$\Lambda y = y \text{ and } y^T \Lambda^T = y^T. \tag{17}$$

Since $P_h^{(i,j)}$ is a probability transition matrix, for

$$\forall 1 \le i, j \le s, \ 1 \le h \le n \text{ and } 1_m = [1, \cdots, 1]^T \in R^{1\times m},$$

it has

$$1_m P_h^{(i,j)} = 1_m. \tag{18}$$

It is clear that

$$\begin{aligned} &[y_1^{(n)} 1_m, \cdots, y_1^{(1)} 1_m, \cdots, y_s^{(n)} 1_m, \cdots, y_s^{(1)} 1_m] B \\ &= [y_1^{(n)} 1_m, \cdots, y_1^{(1)} 1_m, \cdots, y_s^{(n)} 1_m, \cdots, y_s^{(1)} 1_m], \end{aligned} \tag{19}$$

such that one is an eigenvalue of B.

Subsequently, we prove the modulus of all the eigenvalues of B are less than or equal to one. Suppose that $D_v = diag(v)$ and $v^T = y \otimes 1_m$. Using (19), it has $v^T B = v^T$. Then $\widehat{B} = D_v B D_v^{-1}$ is similar to B. From (19), it has $\|\widehat{B}\|_1 = 1$, we obtain

$$\rho(B) = \rho(\widehat{B}) \le \|\widehat{B}\|_1 = 1.$$

The conclusions of this theorem have been proved. □

To keep the irreducible of B, we fill the column of $P_h^{(i,j)}$ with $1/m$ when the column sum of B is zero. If $P_h^{(i,j)}$ is irreducible, then B is also irreducible. From Lemma 2, there exists a unique positive vector X such that $BX = X$.

IV. Estimating the Parameters of Simplified Higher-Order Multivariate Markov Chain Model

In this section, we will estimate the parameters of the new multivariate Markov chain model. Let's first estimate the transition matrices $P_h^{(j,k)}$. If the data sequences are given and the state set is $M = \{1, 2, \cdots m\}$, $F_{i_j,i_k}^{(j,k,h)}$ is frequency from the i_k state in the k th sequence at time $t = r - h + 1$ to the i_j state in the j th sequence at time $t = r + 1$ for $\forall 1 \le i_j, i_k \le m$, then the transition frequency matrices $F_h^{(j,k)}$ of the data sequences can be constructed as:

$$F_h^{(j,k)} = \begin{pmatrix} f_{1,1}^{(j,k,h)} & f_{1,2}^{(j,k,h)} & \cdots & f_{1,m}^{(j,k,h)} \\ f_{2,1}^{(j,k,h)} & f_{2,2}^{(j,k,h)} & \cdots & f_{2,m}^{(j,k,h)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m,1}^{(j,k,h)} & f_{m,2}^{(j,k,h)} & \cdots & f_{m,m}^{(j,k,h)} \end{pmatrix}_{m\times m}.$$

Transition probability matrices $P_h^{(j,k)}$ can be obtained by normalizing the frequency transition matrices as follows:

$$P_h^{(j,k)} = \begin{pmatrix} P_{1,1}^{(j,k,h)} & P_{1,2}^{(j,k,h)} & \cdots & P_{1,m}^{(j,k,h)} \\ P_{2,1}^{(j,k,h)} & P_{2,2}^{(j,k,h)} & \cdots & P_{2,m}^{(j,k,h)} \\ \vdots & \vdots & \vdots & \vdots \\ P_{m,1}^{(j,k,h)} & P_{m,2}^{(j,k,h)} & \cdots & P_{m,m}^{(j,k,h)} \end{pmatrix}_{m\times m},$$

where
$$p_{i_j,i_k}^{(j,k,h)} = \begin{cases} \dfrac{f_{i_j,i_k}^{(j,k,h)}}{\sum_{i_j=1}^m f_{i_j,i_k}^{(j,k,h)}}, & if \sum_{i_j=1}^m f_{i_j,i_k}^{(j,k,h)} \ne 0, \\ 1/m, & otherwise. \end{cases}$$

Subsequently, the way of estimating the parameters $\lambda_{j,k}^{(h)}$ will be introduced. Consider X being a joint stationary probability distribution and X can be presented as

$$X = ((X^{(1)})^T, (X^{(2)})^T, \cdots, (X^{(s)})^T)_{mns\times 1}^T,$$

where
$$X^{(j)} = ((x^{(j)})^T, (x^{(j)})^T, \cdots, (x^{(j)})^T)_{mn\times 1}^T,$$

and
$$BX = X.$$

One would expect that

$$\begin{pmatrix} B^{(11)} & B^{(12)} & \cdots & 0 \\ B^{(21)} & B^{(22)} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & B^{(ss)} \end{pmatrix} X \approx X. \tag{20}$$

Certainly, non-equality (20) can be rewritten as:

$$|BX - X| \le \omega, \qquad (21)$$

where $\omega > 0$ and is as small as possible.

One way of estimating $\lambda_{j,k}^{(h)}$ is to transform (21) into a minimization problem as the following form:

$$\begin{cases} \min\limits_{\lambda_{i,j}^{(h)}} \|BX - X\|, \\ subjectto \sum_{k=1}^{s}\sum_{h=1}^{n} \lambda_{j,k}^{(h)} = 1, \\ \lambda_{j,k}^{(h)} \ge 0, \forall h,k,j. \end{cases} \qquad (22)$$

Here, we choose the norm as infinite norm, and thus the above minimization problem (22) can be presented as:

$$\begin{cases} \min\limits_{\lambda_{i,j}^{(h)}} \max_i \left(\left| \left[\sum_{h=1,i=j}^{n} \lambda_{j,k}^{(h)} P_n^{(j,k)} x^{(k)} + \sum_{h=1,i\ne j}^{n} \lambda_{j,k}^{(h)} P_n^{(j,k)} x^{(k)} \right]_i \right| \right), \\ subjectto \sum_{h=1,i=j}^{n} \lambda_{j,k}^{(h)} + \sum_{h=1,i\ne j}^{n} \lambda_{j,k}^{(h)} = 1, \\ \lambda_{j,k}^{(h)} \ge 0, \forall h,k,j. \end{cases} \qquad (23)$$

where $[\cdot]_i$ is the i th entry of the vector. Using the idea of [1], the minimization problem (23) becomes a linear programming problem as follows:

$$\begin{cases} \min\limits_{\lambda_{i,j}^{(h)}} \omega_j \\ subjuctto \\ \begin{pmatrix} \omega_j \\ \omega_j \\ \omega_j \\ \omega_j \end{pmatrix} \ge x^{(j)} - C_j \begin{pmatrix} \tilde{\lambda}_{j,1} \\ \tilde{\lambda}_{j,2} \\ \vdots \\ \tilde{\lambda}_{j,s} \end{pmatrix}, \\ \begin{pmatrix} \omega_j \\ \omega_j \\ \omega_j \\ \omega_j \end{pmatrix} \ge -x^{(j)} + C_j \begin{pmatrix} \tilde{\lambda}_{j,1} \\ \tilde{\lambda}_{j,2} \\ \vdots \\ \tilde{\lambda}_{j,s} \end{pmatrix}, \\ \omega_j \ge 0, \\ \sum_{h=1,i=j}^{n} \lambda_{j,k}^{(h)} + \sum_{h=1,i\ne j}^{n} \lambda_{j,k}^{(h)} = 1, \lambda_{j,k}^{(h)} \ge 0, \forall k,j,h. \end{cases}$$

where

$$C_j = [P_1^{(j,1)} x^{(1)}, \cdots, P_1^{(j,j-1)} x^{(j-1)}, P_1^{(j,j)} x^{(j)}, \cdots P_1^{(j,j)} x^{(j)}, \cdots, P_1^{(j+1,j+1)} x^{(j+1)}, \cdots P_1^{(j,s)} x^{(s)}],$$

and

$$\tilde{\lambda}_{j,k} = \begin{cases} (\lambda_{j,k}^{(1)}, \cdots, \lambda_{j,k}^{(n)})^T, & if \quad j = k, \\ \lambda_{j,k}^{(1)}, & if \quad j \ne k. \end{cases}$$

V. Numerical Experiments

In this section, we report on the numerical results with different higher-order multivariate Markov chain models for two examples obtained with a Matlab 7.0.1 implementation of a Windows XP with 2.93GHz 64-bit processor and 2GB memory.

A Simple Example

There are two categorical data sequences

$$S_1 = \{2,1,3,4,4,3,3,1,3,3,2,3\},$$
$$S_2 = \{2,4,4,4,4,2,3,3,1,4,3,3\}.$$

In this example, we aim at constructing a simplified 2th-order multivariate Markov chain model. Counting the first lag transition frequencies, we have

$$F_1^{(1,1)} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad F_1^{(2,2)} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 3 \end{pmatrix}.$$

With the same process, the second lags transition frequencies are given as follows:

$$F_2^{(1,1)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 0 \\ 1 & 1 & 2 & 1 \end{pmatrix}, \quad F_2^{(2,2)} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Counting the inner first lags transition frequencies, it has

$$F_2^{(2,1)} = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 1 & 0 & 1 & 2 \end{pmatrix}, \quad F_2^{(1,2)} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 2 \\ 2 & 0 & 0 & 1 \end{pmatrix}.$$

Normalizing the transition frequency matrices, we obtain the transition probability matrices:

$$p_1^{(1,1)} = \begin{pmatrix} 0 & 1/2 & 1/5 & 0 \\ 0 & 0 & 1/5 & 1 \\ 1 & 1/2 & 2/5 & 1/2 \\ 0 & 0 & 1/5 & 1/2 \end{pmatrix}, \quad p_1^{(2,2)} = \begin{pmatrix} 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \\ 0 & 1/2 & 2/3 & 0 \\ 1 & 1/2 & 0 & 3/4 \end{pmatrix},$$

$$p_2^{(1,1)} = \begin{pmatrix} 0 & 0 & 1/5 & 0 \\ 0 & 0 & 1/5 & 0 \\ 1/2 & 1 & 2/5 & 1 \\ 1/2 & 0 & 1/5 & 0 \end{pmatrix}, \quad p_2^{(2,2)} = \begin{pmatrix} 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/5 \\ 1 & 1/2 & 0 & 2/5 \\ 0 & 1/2 & 1/2 & 2/5 \end{pmatrix},$$

$$p_1^{(2,1)} = \begin{pmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 3/5 & 0 \\ 1/2 & 1/2 & 2/5 & 1/2 \end{pmatrix}, \quad p_1^{(1,2)} = \begin{pmatrix} 0 & 1/2 & 1/3 & 0 \\ 0 & 0 & 0 & 1/5 \\ 1 & 0 & 2/3 & 2/5 \\ 0 & 1/2 & 0 & 2/5 \end{pmatrix}.$$

The initial state probability distributions can be obtained by computing the proportion of the occurrence of each state in each of the sequence

$$x_0^{(1)} = \left(1/6, 1/6, 1/2, 1/6\right)^T, \; x_0^{(2)} = \left(1/12, 1/6, 1/3, 5/12\right)^T.$$

and $\lambda_{j,k}^{(h)}$ can be calculated by linear programming problem which corresponds to the simplified second-order multivariate Markov chain model of the two categorical data sequences. With the results of $\lambda_{j,k}^{(h)}$ the simplified 2th-order multivariate Markov chain model is presented as follows:

$$\begin{cases} x_{r+1}^{(1)} = 0.6647 P_1^{(1,1)} x_r^{(1)} + 0.3353 P_2^{(1,1)} x_r^{(1)}, \\ x_{r+1}^{(1)} = 0.3747 P_1^{(2,1)} x_r^{(1)} + 0.3199 P_2^{(2,2)} x_r^{(2)} \\ \qquad + 0.3054 P_2^{(2,2)} x_r^{(2)}. \end{cases}$$

B. An Application to Sales Demand Prediction

In this part, the sales demand sequences are presented to show the effectiveness of the simplified higher-order multivariate Markov chain model for multivariate Markov chains. Since the requirement of the market fluctuates heavily, the production planning and the inventory control directly affect the estate cost. Thus, studying the interplay between the storage space requirement and the overall growing sales demand is a pressing issue for the company. Here, our goal is to predict the sales demand of the market for minimizing the estate cost. Consider the products can be classified into six possible states (1, 2, 3, 4, 5, 6), e.g., 1 = no sales volume, 2 = vary low sales volume, 3 = low sales volume, 4 = standard sales volume, 5 = fast sales volume, 6 = vary fast sales volume. The customer's sales demand of five important products of the company for a year has been given in [1].

We choose the simplified 8th-order multivariate Markov chain model to model five categorical data sequences. By computing the proportion of the occurrence of each state in each of the sequence, the initial probability distributions of the five categorical data sequences are

$$\begin{aligned} x_0^1 &= (0.0818, 0.4052, 0.0483, 0.0335, 0.0037, 0.4275)^T, \\ x_0^2 &= (0.3680, 0.1970, 0.0335, 0.0000, 0.0037, 0.3978)^T, \\ x_0^3 &= (0.1450, 0.2045, 0.0186, 0.0000, 0.0037, 0.6283)^T, \\ x_0^4 &= (0.0000, 0.3569, 0.1338, 0.1896, 0.0632, 0.2565)^T, \\ x_0^5 &= (0.0000, 0.3569, 0.1227, 0.2268, 0.0520, 0.2416)^T. \end{aligned}$$

The transition probability matrices $P_h^{(j,k)}$ can be obtained after normalizing the transition frequency matrices. By solving the corresponding linear programming problem, $\lambda_{i,j}^{(h)}$ is obtained and the simplified 2nd-order multivariate Markov chain model can be presented as follows:

$$\begin{cases} x_{r+1}^{(1)} = P_1^{(1,1)} x_r^{(1)}, \\ x_{r+1}^{(2)} = 0.6364 P_1^{(2,2)} x_r^{(2)} + 0.3636 P_3^{(2,2)} x_r^{(2)}, \\ x_{r+1}^{(3)} = P_1^{(3,3)} x_r^{(3)}, \\ x_{r+1}^{(4)} = 0.2809 P_1^{(4,2)} x_r^{(2)} + 0.6198 P_1^{(4,4)} x_r^{(4)} \\ + 0.0993 P_3^{(4,4)} x_r^{(4)}, \\ x_{r+1}^{(5)} = 0.2771 P_1^{(5,2)} x_r^{(2)} + 0.6198 P_1^{(5,5)} x_r^{(5)} \\ + 0.0635 P_1^{(5,5)} x_r^{(5)}. \end{cases} \tag{24}$$

Next we use the simplified 8th-order multivariate Markov chain model to predict the next state of the *k*th sequence $x_t^{(k)}$ at time t, which can be taken as the state with the maximum probability, i.e.,

$$x_t^{(k)} = j, if \left[x_t^{(k)}\right]_i \le \left[x_t^{(k)}\right]_j, \forall 1 \le i \le m.$$

For evaluating the performance and the effectiveness of different multivariate Markov chain models, the prediction accuracy r is proposed in [1] and defined as:

$$r = \frac{1}{T-n} \times \sum_{t=n+1}^{T} \delta_t \times 100\%,$$

where T is the length of the data sequence and

$$\delta_t = \begin{cases} 1, & \text{if } x_t^{(k)} = \theta_t^{(k)} \\ 0, & \text{otherwise} \end{cases}.$$

For comparisons, the numerical results of the 8th-order multivariate Markov chain are also presented in Table I, where we denote "*time*" is the computational time, "*np*" is the number of the parameters in the models, "*H*" is a higher-order multivariate Markov chain model, "*SH*" is a simplified higher-order multivariate Markov chain model and the prediction accuracies of Product A, Product B, Product C, Product D, Product E are denoted as "A", "B", "C", "D", "E" respectively. Observing the data in Table I, the prediction accuracies of our new model are nearly the same as the higher-order multivariate Markov chain model's. In addition, it is obviously that the computational time and the number of the parameters of our new model are much less than those of the higher-order multivariate Markov chain model.

TABLE I
PREDICTION ERRORS OF THE SIMPLIFIED HIGHER-ORDER MULTIVARIATE MARKOV CHAIN MODEL AND HIGHER-ORDER MULTIVARIATE MARKOV CHAIN MODEL IN SALES DEMAND PREDICTIONS

	A	B	C	D	E	*time*	*np*
H	0.4275	0.3978	0.6283	0.3569	0.3569	0.4688	61
SH	0.4275	0.3978	0.6283	0.3569	0.3569	0.6875	201

Subsequently, another prediction criterion for Markov chain models is introduced. Note that "*nA*" is the number of the categorical data in the sequences, ` is a predict probability at time *t*, X_t is a fact value at time *t* and $X_t = [X_t^1, \cdots, X_t^s]^T$. If m_t is the fact state at *t* in *i*th categorical data sequence, $X_t^i = e_{(m_t)} = \{0,\cdots,0,1,0,\cdots,0\}^T \in R^{1\times m}$. We denote the prediction error in the models as "*pe*" which can be estimated by the equation:

$$pe = \sum_{t=9}^{nA} \left\| \overline{X}_t - X_t \right\|_2 .$$

In Table II, we denote that the prediction error of the simplified higher-order multivariate Markov chain model is "pe_1", the prediction error of the higher-order multivariate Markov chain model is "pe_2", the number of the parameters of the simplified higher-order multivariate Markov chain model is """np_1"""", the number of the parameters of the higher-order multivariate Markov chain model is "np_2", the computational time of the simplified higher-order multivariate Markov chain model is "$time_1$", the computational time of the higher-order multivariate Markov chain model is "$time_2$", respectively. Stop criterion of the linear programming problem can be found in *Matlab order linprog*.

Table II provides that the performance of the prediction errors in simplified higher-order multivariate Markov chain model which is comparable with the higher-order multivariate Markov chain model. Table II illustrates the benefits of the simplified higher-order multivariate Markov chain model in terms of the computational time and the number of parameters controlling.

TABLE II
PREDICTION ERRORS OF THE SIMPLIFIED HIGHER-ORDER MULTIVARIATE MARKOV CHAIN MODEL AND HIGHER-ORDER MULTIVARIATE MARKOV CHAIN MODEL IN SALES DEMAND PREDICTIONS

	pe^1	np^1	$time^1$	pe^2	np^2	$time^2$
n=1	372.1810	25	0.0781	372.1810	25	0.0781
n=2	353.3360	50	0.0925	365.2160	30	0.0781
n=3	352.5338	75	0.1250	352.5689	35	0.0781
n=4	353.3630	100	0.1406	353.4019	40	0.0925
n=5	357.6212	125	0.1763	353.8105	45	0.0925
n=6	356.8549	150	0.2031	353.8092	50	0.1094
n=6	357.4291	175	0.2500	353.6237	55	0.1094
n=6	354.7548	200	0.2969	355.7788	60	0.1094

VI. CONCLUSION

In real application, it is inevitable that the computation from a large categorical data sequence group which will cause high computational cost. For saving computational cost, we propose a simplified higher-order multivariate Markov chain model in this paper. The number of the parameters of the new model is only $O((n+s)sm^2)$ which is less than $O((ns^2m^2))$ of the higher-order multivariate Markov chain model. The results of the prediction accuracy and the prediction error in these two model are comparable or nearly the same. Moreover, numerical experiments illustrate the benefits of our new model in time consuming, the number of the parameters controlling and the storage requirements. Certainly, the simplified higher-order multivariate Markov chain model can also be applied in credit risk and other research areas.

REFERENCES

[1] W. Ching, E. Fung and M. Ng, A multivariate Markov chain model for categorical data sequences and its applications in demand prediction, IMA, Journal of Management Mathematics, 13 (2002) 187-199.
[2] W. Ching, E. Fung, and M. Ng, A Higher-order Markov model for the newsboys problem, J. Operational Research Society, 54 (2003) 291-298.
[3] W. Ching, E. Fung and M. Ng, Higher-order Markov chain models for categorical data sequences, Inter. J. Nav. Res. Logist., 51 (2004) 557-574.
[4] W. Ching, M. Ng and E. Fung, On Construction of stochastic genetic networks based on gene expression sequences, International Journal of Neural Systems, 15 (2005) 297-310.
[5] W. Ching and M. Ng, Markov chains: models, algorithms and applications, International Series on Operations Research and Management Science, 2006.
[6] W. Ching, Building higher-order Markov chain models with EXCEL, Intern. J. Mathematical Education in Science and Engineering, 35 (2004) 921-932.
[7] W. Ching, M. Ng, and E. Fung, Higher-order multivariate Markov chains and their applications, Linear Algebra and its Applications, 428 (2008) 492-507.

[8] W. Ching, T. Siu and L. Li, A improved parsimonious multivariate Markov chain model for credit risk, Journal of Credit Risk, 5 (2009) 1-25.
[9] W. Ching, H. Leung, H. Jiang, L. Sun, and T. Siu, A Markovian network model for default risk management, International Journal of Intelligent Engineering Informatics, 1 (2010) 104-124.
[10] V. Chvatal, Linear programming, Freeman, 1983.
[11] M. Davis and V. Lo, Modeling default correlation in bond portfolios, In Mastering Risk, 2 (2001) 141-151.
[12] R. Horn and C. Johnson, Matrix analysis, Cambridge University Press,
[13] M. Kijima, K. Komoribayashi and E. Suzuki, A multivariate Markov model for simulating correlated defaults, Journal of Risk, 4 (2002) 1-32.
[14] A. Raftery, A model for high-order Markov chains, J. R. Statist. Soc. B, 47 (1985) 528-539.
[15] T. Siu, W. Ching, M. Ng and E. Fung, On multivariate credibility approach for portfolio credit risk measurement, Quantitative Finance, 5 (2005) 543-556.
[16] S. Ross, Introduction to probability models, Academic Press, Amsterdam, Boston, 10th ed., 2010.
[17] D. Zhu and W. Ching, A note on the stationary property of highdimensional Markov chain models, International Journal of Pure and Applied Mathematics, 66 (2011) 321-330.
[18] Z.Y. You, L.C. Wang, A concept of nonlinear block diagnal dominance, J. Comput. Appl. Math, 83 (1997) 1-10.

Chao Wang was born in 1983 in Jinlin, China. He received his B.S. (2007) in Applied Mathematics from Bei Hua University and his M.S. (2010) in School of Mathematical Sciences from University of Electronic Science and Technology of China. He is working on his Ph.D. at the University of Electronic Science and Technology in Chengdu, China. His current interests include multivariate Markov Chain model, saddle point problems and linear complementary problem.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in Computational Mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China, in 1986, 1992 and 2000, respectively. During 2005, he was a visiting scholar in Dept. of Computer Science, Loughborough University of UK. He is currently a Full Professor in the School of Mathematical Sciences, UESTC. He is currently an editorial of Advances in Numerical Analysis, Journal of Pure and Applied Mathematics: Advances and Applications, Journal of Electronic Science and Technology of China. He is the author or coauthor of more than 100 research papers. His current research interests include numerical linear algebra with applications, iterative methods of linear systems and saddle point problems, preconditioning technologies, and matrix analysis with applications.

Chen Jia received the B.S. degrees from the School of Accounting, Jilin University of Finance and Economics, Chuangchun, China, in 2008. She is currently a postgraduate in the School of Accounting, Jilin University of Finance and Economics.