IOP Conf. Series: Materials Science and Engineering **231** (2017) 012011 doi:10.1088/1757-899X/231/1/012011

A simplified parsimonious higher order multivariate Markov chain model with new convergence condition

Chao Wang[12], Chuan-sheng Yang[12]

[1]School of Mathematics, Physics and Information Science, Zhejiang Ocean University

[2]Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province

Abstract. In this paper, we present a simplified parsimonious higher-order multivariate Markov chain model with new convergence condition. (TPHOMMCM-NCC). Moreover, estimation method of the parameters in TPHOMMCM-NCC is give. Numerical experiments illustrate the effectiveness of TPHOMMCM-NCC.

1 Introduction

Markov chains is an important implement in many research areas, such as, internet applications [2] music [3], software testing[4], land cover change [5], energy consumption [6], speech recognition [7], physics, gene expression [9], finance [10-11], DNA[12] and so on. It is helpful to develop a better model for a more accurate prediction.by exploring the relationships of different categorical data sequences is meaningful to accurate prediction.

Different methods for multiple categorical data sequences prediction (which means the relationships of different categorical data sequences are taken into account) has been proposed, e.g., the first-order multivariate Markov chain model, higher-order multivariate Markov chain model and an improved multivariate Markov chain model (to speed up the convergence) [10]. (They add a negative association part which is multiplied a constant for normalizing solutions at the back of the positive association part of the model.).

The organization of this article is organized as follows. In Section 2, we review some basic knowledge of Markov chain model. In Section 3, a simplified parsimonious higher-order multivariate Markov chain model with new convergence condition is proposed for multiple categorical data sequences. In Section 4, we estimate the parameters of the simplified parsimonious higher-order multivariate Markov chain model with new convergence condition. Numerical experiments show the effectiveness of our model in Section 5.

2 A review on the Markov chains

In this section, we briefly introduce several definitions and the first-order Markov chain model.

2.1 The first-order Markov chain

Several definitions of the Markov chain are first introduced [3].

Definition 1 Let the state set of the categorical data sequence with m states be $\mathrm{M} = \{1,2,\ldots,m\}$ and $\theta_k \in \mathrm{M},\ k = \{1,2,\ldots\}$. If the sequence $\{x_0, x_1, x_2, \ldots\}$ with m states satisfies the following relationships:

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

IOP Conf. Series: Materials Science and Engineering **231** (2017) 012011 doi:10.1088/1757-899X/231/1/012011

$$\Pr ob(x_{t+1}=\theta_{t+1}\mid x_0=\theta_0,x_1=\theta_1,\ldots,x_t=\theta_t)$$
$$=\Pr ob(x_{t+1}=\theta_{t+1}\mid x_t=\theta_t)$$

the sequence is called as first-order discrete-time Markov chain.

Definition 2 The conditional probability

$$\Pr ob(x_{t+1}=\theta_{t+1}\mid x_t=\theta_t) \qquad (1)$$

is called as the transition probability of the Markov chain.

Definition 3 Rewriting the transition probability as

$$p_{j,k}=\Pr ob(x_{t+1}=j\mid x_t=k),\forall j,k\in \mathrm{M} \qquad (2)$$

then the transition probability matrix is

$$p=[p_{j,k}],0\le p_{j,k}\le 1,\sum_{j=1}^{m}p_{j,k}=1,\forall j,k\in M.$$

Definition 4 Suppose that

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

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

2.2The simplified parsimonious higher-order multivariate Markov chain model

In this part, a simplified parsimonious higher-order multivariate Markov chain model is introduced.

For $\forall j,k\in\{1,2,\ldots,s\},\forall t\in\{n-1,n,\ldots\}$, TPHOMMCM -NCC is

$$x_{t+1}^{(j)}=\sum_{h=1,j=k}^{n}\lambda_{j,k}^{(h)}p_h^{(j,k)}x_{t-h+1}^{(k)}+\frac{1}{m-1}\sum_{h=1,j=k}^{n}\lambda_{j,-k}^{(h)}p_h^{(j,k)}(1-x_{t-h+1}^{(k)}) \qquad (3)$$
$$+\sum_{j\neq k}\lambda_{j,k}^{(1)}p_1^{(j,k)}x_{t-h+1}^{(k)}+\frac{1}{m-1}\sum_{j\neq k}\lambda_{j,-k}^{(1)}p_1^{(j,k)}(1-x_{t-h+1}^{(k)})$$

where $x_0^{(k)},x_1^{(k)},\ldots x_{n-1}^{(k)}(k=1,2,\ldots,s)$ are the initial probability distributions, the normalization constant $1/m$ keeps $X_{t+1}^{j}=(x_{t+1}^{(1)},x_{t+1}^{(2)},\ldots x_{t+1}^{(j)})^T$ as a probability vector. (3) satisfies

$$\sum_{h=1,j=k}^{n}\lambda_{j,k}^{(h)}+\sum_{h=1,j=k}^{n}\lambda_{j,-k}^{(h)}+\sum_{j\neq k}\lambda_{j,k}^{(1)}+\sum_{j\neq k}\lambda_{j,-k}^{(1)}=1,$$
$$\lambda_{j,k}^{(h)},\lambda_{j,-k}^{(h)}\ge 0,\forall j,k\in\{1,2,\ldots,s\},\forall t\in\{n-1,n,\ldots\},$$

where $x_{t+1}^{(j)}$ is the state probability distribution at time *t+1* in the *k*th sequence and $P_h^{(j,k)}$ the *ht*h step transition probability matrix from the states at time $t-h+1$ in the *k*th sequence to the states at time $t+1$ in the *j*th sequence.

Let $X_{t+1}^{(j)}=((x_{t-n+1}^{(j)})^T,(x_{t-n}^{(j)})^T,(x_t^{(j)})^T)\in IR^{mn\times 1}$, the simplified parsimonious higher-order multivariate Markov chain model in matrix form has

$$\begin{pmatrix} X_{t+1}^{(j)} \\ X_{t+1}^{(j)} \\ \vdots \\ X_{t+1}^{(j)} \end{pmatrix} = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \dots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \dots & B^{(2,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \dots & B^{(s,s)} \end{pmatrix} \begin{pmatrix} X_t^{(j)} \\ X_t^{(j)} \\ \vdots \\ X_t^{(j)} \end{pmatrix} + \frac{1}{m-1} \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \dots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \dots & B^{(2,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \dots & B^{(s,s)} \end{pmatrix} \begin{pmatrix} 1-X_t^{(j)} \\ 1-X_t^{(j)} \\ \vdots \\ 1-X_t^{(j)} \end{pmatrix} \quad (4)$$

where if $j = k$

$$B^{(j,k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,k)} & \lambda_{j,k}^{(1)} P_1^{(j,k)} & \dots & \lambda_{j,k}^{(1)} P_1^{(j,k)} & \lambda_{j,k}^{(1)} P_1^{(j,k)} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & I & 0 \end{pmatrix}, B^{(j,-k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,-k)} & \lambda_{j,k}^{(1)} P_1^{(j,-k)} & \dots & \lambda_{j,k}^{(1)} P_1^{(j,-k)} & \lambda_{j,k}^{(1)} P_1^{(j,-k)} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & I & 0 \end{pmatrix}$$

else if $j \neq k$,

$$B^{(j,k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,k)} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}_{mn \times mn}, B^{(j,-k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,-k)} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}_{mn \times mn}$$

and

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

Here, the column sum of B^{+}, B^{-} are not necessary equal to one.

3 Convergence condition

After *t* steps iteration of TPHOMMCM-NCC

$$\begin{aligned} X_{t+1} &= BX_t + \frac{1}{m-1} C(e - X_t) \\ &= (B - \frac{1}{m-1} C)^{t+1} X_0 + \sum_{k=0}^{t} (B - \frac{1}{m-1} C)^k \frac{1}{m-1} C \cdot e \end{aligned} \quad (5)$$

If $\rho(B - \frac{1}{m-1} C) < 1$, the iteration of TPHOMMCM is convergent.

Theorem 1 Let *X* be the stationary probability distribution of TPHOMMCM. If $B > \frac{1}{m-1} C$, then the iteration of TPHOMMCM is convergent.

Proof. In TPHOMMCM, $B, C > 0, B > \frac{1}{m-1} C,$ we have $0 < B - \frac{1}{m-1} C < B,$ which means

$$\left| B - \frac{1}{m-1} C \right| < B.$$

By Lemma 1,

$$\left| \rho\left(B - \frac{1}{m-1} C \right) \right| < \rho(B).$$

Because $\|B\| < 1$. Therefore,

$$\left|\rho\left(B - \frac{1}{m-1}C\right)\right| < 1.$$

This theorem has been proved.

4 Parameter estimation

In this section, we will estimate the parameters of the simplified parsimonious higher-order multivariate Markov chain model.

Let's first estimate the transition matrices $P_h^{(j,k)}$. Suppose that $M = \{1,2,\ldots,m\}$ is the state set, $F_{i_j,i_k}^{(j,k)}$ is frequency from the i_k state at time $r-h+1$ in the kth sequence to the i_j state at time $r+1$ in the jth sequence for $1 \le i_j, i_k \le m,\ 1 \le h \le n$, then the transition frequency matrices $F_h^{(j,k)}$ of the data sequences is

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

Normalizing the frequency transition matrices, probability transition matrix is

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

where

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

Next, the way of estimating the parameter $\lambda_{j,k}^{(h)}$ will be presented. Let the joint stationary probability distribution be

$$X = ((X^{(1)})^T, (X^{(2)})^T, \ldots, (X^{(2)})^T)^T$$

where

$$X^{(j)} = ((x^{(j)})^T, (x^{(j)})^T, \ldots, (x^{(j)})^T)^T.$$

One would expect that

$$\left\| B^{+}X + \frac{1}{m-1}B^{-}(1-X) - X \right\| \le \omega \qquad (6)$$

where $\omega \ge 0$ and ω is as small as possible.

Transform (6) into a minimization problem:

$$\begin{cases} \min\limits_{\lambda_{j,k}} \left\| B^{+}X + \dfrac{1}{m-1}B^{-}(1-X) - X \right\| \\ \text{subject to } \sum\limits_{h=1,\, j=k}^{n} \lambda_{j,k}^{(h)} + \sum\limits_{h=1,\, j=k}^{n} \lambda_{j,-k}^{(h)} + \sum\limits_{j \ne k} \lambda_{j,k}^{(1)} + \sum\limits_{j \ne k} \lambda_{j,-k}^{(1)} = 1 \\ B > \dfrac{1}{m-1}C \end{cases} \qquad (7)$$

With the same process in [10], TPHOMMCM-NCC can be transformed into following form:

$$\min_{\lambda_{j,k}^{(g)}} \sum_j \omega_j \tag{8}$$

subject to

$$\begin{cases} \begin{pmatrix} \omega_j \\ \vdots \\ \omega_j \\ \omega_j \\ \vdots \\ \omega_j \end{pmatrix} \geq x^{(j)} - C_j \begin{pmatrix} \lambda_{j,1} \\ \vdots \\ \lambda_{j,s} \\ \lambda_{j,-1} \\ \vdots \\ \lambda_{j,-s} \end{pmatrix} \\ \begin{pmatrix} \omega_j \\ \vdots \\ \omega_j \\ \omega_j \\ \vdots \\ \omega_j \end{pmatrix} \geq x^{(j)} - C_j \begin{pmatrix} \lambda_{j,1} \\ \vdots \\ \lambda_{j,s} \\ \lambda_{j,-1} \\ \vdots \\ \lambda_{j,-s} \end{pmatrix} \\ \lambda_{j,k}^{(h)} > \dfrac{1}{m-1} \lambda_{j,-k}^{(h)}, \quad \omega_j \geq 0 \\ \sum_{h=1, j=k}^{n} \lambda_{j,k}^{(h)} + \sum_{h=1, j=k}^{n} \lambda_{j,-k}^{(h)} + \sum_{j \neq k} \lambda_{j,k}^{(1)} + \sum_{j \neq k} \lambda_{j,-k}^{(1)} = 1 \end{cases}$$

where

$$\begin{aligned} C_j = [&P_1^{(j,1)} x^{(1)}, P_1^{(j,2)} x^{(2)}, \dots, P_1^{(j,j)} x^{(j)}, P_2^{(j,j)} x^{(j)}, \dots, P_n^{(j,j)} x^{(j)}, \\ &P_1^{(j,j+1)} x^{(j)}, \dots, P_1^{(j,s)} x^{(j)}, \frac{1}{m-1} P_1^{(j,1)}(1 - x^{(1)}), \frac{1}{m-1} P_1^{(j,2)}(1 - x^{(2)}) \\ &, \dots, \frac{1}{m-1} P_1^{(j,j)}(1 - x^{(j)}), \frac{1}{m-1} P_2^{(j,j)}(1 - x^{(j)}), \dots, \\ &\frac{1}{m-1} P_n^{(j,j)}(1 - x^{(j)}), \frac{1}{m-1} P_1^{(j,j+1)}(1 - x^{(j)}), \dots, \frac{1}{m-1} P_1^{(j,s)}(1 - x^{(j)})] \end{aligned}$$

and

$$\lambda_{j\pm k} = \begin{cases} (\lambda_{j\pm k}^{(1)}, \dots, \lambda_{j\pm k}^{(1)})^T & \text{if } j = k \\ \lambda_{j\pm k}^{(1)} & \text{if } j \neq k \end{cases}$$

5 Application to sales demand prediction with simplified parsimonious higher-order multivariate Markov chain model

In this part, the sales demand categorical data sequences are presented to show the effectiveness of the simplified parsimonious higher-order multivariate Markov chain model.

We classifies the sales demand into six states *(1,2,3,4,5,6)*, The customer's sales demand states in the same customer group of five important products of the company for a year is given in the Appendix [8].

Noting that $\overline{X}_t$ is a predict probability at time t, X_t a fact value at time t and $X_t = [X_t^1, \dots, X_t^s]^T$, nA the data number of each sequence. If m_t is the fact state at t in i th categorical data sequence, $X_t^i = e_{(m_t)} = \{0, \dots, 0, 1, 0, \dots, 0\}^T \in IR^{1\times m}$. We denote the prediction error in the models as pe which can be estimated by the equation:

$$pe = \sum_{t=8}^{nA} \left\| \overline{X}_t - X_t \right\|_2$$

In Tables, denote that α the convergence factor of the convergence condition, n is the order of the model, $M1$ higher-order multivariate Markov chain model, $M2$ parsimonious higher-order multivariate Markov chain model an $M3$ the simplified parsimonious higher-order multivariate Markov chain model.

Table 1: Prediction errs of M1, M2 and M3.

	M1		*M2*		*M3*	
n	*pn*	*pe*	*pn*	*pe*	*pn*	*pe*
1	13	314.67	25	259.35	13	257.49
2	18	298.10	50	247.36	18	244.39
3	23	297.98	75	246.31	23	244.99
4	28	298.42	100	246.31	28	243.76
5	33	302.22	125	246.31	33	243.40
6	38	301.40	150	247.16	38	243.88
7	43	301.97	175	246.91	43	243.87
8	48	299.97	200	246.91	48	243.91

From the results of Figure 1,2,3, we find that the simplified parsimonious higher-order multivariate Markov chain model preforms better than parsimonious higher-order multivariate Markov chain model and the higher-order multivariate Markov chain model in parameter number comparing, time consuming and the prediction precision.

6 Conclusions

We have investigated a simplified parsimonious higher-order multivariate Markov chain model and discussed its convergence condition. Numerical experiments show that the simplified parsimonious higher-order multivariate Markov chain model is efficient. Certainly, this model can be applied in credit risk, gene expression and other research areas.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (Nos. 61573321, 61272021, 61202206, 61173181, 61322211, 41301473 and 41301438), the Zhejiang Provincial Natural Science Foundation of China (Nos. LZ12F03002 and LY14F030001) and Key Fund Project of Sichuan Provincial Department of Education (Nos. 17za0003).

References

[1] A. Markov. “Extension of the limit theorems of probability theory to a sum of variables connected in a chain”, Dynamic Probabilistic Systems, 1, pp. 552-577 (1971) .

[2] B. Nigam, S. Tokekar, S. Jain. “Predicting the next accessed web page using Markov model and pageRank”, International Journal of Data Mining and Emerging Technologies, 3, pp. 73-80, (2013).

[3] D. Herremans, S. Weisser, K. Sorensen, D. Conklin. “Generating structured music for bagana using quality metrics based on Markov models”, Expert Systems with Applications, 42, pp. 7424-7435, (2015).

[4] J.A. Whittaker, M.G. Thomason. “A Markov chain model for statistical software testing”, IEEE Transactions on Software engineering, 20, 812-824, (1994).

[5] J. H. Park, S. H. No, G. S. Lee. “Outlook analysis of future discharge according to land cover change using CA-Markov technique based on GIS”, Journal of the Korean Association of Geographic Information Studies, 16, pp. 25-39, (2013).

[6] J. H. Park, T. H. Hong. “Analysis of South Korea’s economic growth, carbon dioxide emission, and energy consumption using the Markov switching model”, Renewable and Sustainable Energy Reviews, 18, pp. 543-551, (2013).

[7] P. Dighe, A. Asaei, H. Bourlard. “Sparse Hidden Markov Models for Automatic Speech Recognition”, No. EPFL-REPORT-210627 Idiap, (2015).

[8] W. Ching, M. Ng, E. Fung. “Higher-order multivariate Markov chains and their applications”, Linear Algebra and its Applications, 428, pp. 492-507, (2008).

[9] W. Ching, M. Ng, E. Fung. “On Construction of stochastic genetic networks based on gene expression sequences”, International Journal of Neural Systems, 15, pp. 297-310, (2005) .

[10] W. Ching, T. Siu, L. Li. “An improved parsimonious multivariate Markov chain model for credit risk”, Journal of Credit Risk, 5, pp. 1-25, (2009).

[11] R.S. Mamon, R.J. Elliott. “Hidden markov models in finance”, Springer Science and Business Media, 2007.

[12] T.J. Wheeler, J. Clements, S.R. Eddy. “Dfam: a database of repetitive DNA based on profile hidden Markov models”, Nucleic Acids Research, 41, pp. 70-82, (2013).