A TRIDIAGONAL PARSIMONIOUS HIGHER ORDER MULTIVARIATE MARKOV CHAIN MODEL WITH NEW CONVERGENCE CONDITION

Liu Kou[1, a], Chuan-sheng Yang[1,b], Chao Wang[1,c]

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

Zhejiang, 316022, China

[a]1034785223@qq.com, [b] cshyang@zjou.edu.cn,[c]wangchao1321654@163.com

Keywords: Higher-order multivariate Markov chains; Categorical data sequences; Convergence condition.

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

Introduction

Markov chains have been applied in many research areas, such as, speech recognition [1], internet application [2], finance [3], music [4] and so on [5-10] . Developing a better model for a more accurate prediction is an urgent task. The relationships of different categorical data sequences are helpful to accurate prediction.

Different methods for multiple categorical data sequences prediction have been proposed, e.g., the first-order multivariate Markov chain model [11], higher-order multivariate Markov chain model [12] and an improved multivariate Markov chain model [13].

In this paper, we propose a tri-diagonal parsimonious higher-order multivariate Markov chain model with new convergence condition for enhancing the prediction precision and controlling the parameter number of the model.

The organization of this paper is organized as follows. In Section 2, we review several definitions and a model of Markov chain model. In Section 3, a tri-diagonal 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 tri-diagonal parsimonious higher-order multivariate Markov chain model with new convergence condition. Numerical experiments show the effectiveness of our model in Section 5.

Review on the Markov chains

In this section, we briefly introduce several definitions and the first-order Markov chain model. Several definitions of the Markov chain are first introduced [11].

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

$$\Pr ob(x_{t+1} = q_{t+1} \mid x_0 = q_0, x_1 = q_1, \mathbf{K}, x_t = q_t),$$
$$= \Pr ob(x_{t+1} = q_{t+1} \mid x_t = q_t)$$

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

Definition 2 The conditional probability

$$\Pr ob(x_{t+1} = q_{t+1} \mid x_t = q_t)$$

is called as the transition probability of the Markov chain.

Definition 3 Rewriting the transition probability as

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

then the transition probability matrix is

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

Definition 4 Suppose that $X_{t+1} = PX_t$, then $X_t = (x_t^1, x_t^2, \mathbf{K}, x_t^m)^T$ is the state probability distribution and X_0 the initial probability distribution.

Tri-diagonal parsimonious higher-order multivariate Markov chain model with new convergence condition

In this part, a tri-diagonal parsimonious higher-order multivariate Markov chain model with new convergence condition (TPHOMMCM-NCC) is presented.

Tri-diagonal parsimonious higher-order multivariate Markov chain model

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

$$
x_{t+1}^{(j)} = \sum_{h=1, j=k}^{n} I_{j,k}^{(h)} p_h^{(j,k)} x_{t-h+1}^{(k)} + \frac{1}{m-1} \sum_{h=1, j=k}^{n} I_{j,-k}^{(h)} p_h^{(j,k)} (1 - x_{t-h+1}^{(k)})
$$
$$
+ \sum_{|j-k|=1} I_{j,k}^{(1)} p_1^{(j,k)} x_{t-h+1}^{(k)} + \frac{1}{m-1} \sum_{|j-k|=1} I_{j,-k}^{(1)} p_1^{(j,k)} (1 - x_{t-h+1}^{(k)})
$$
(1)

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

$$
\sum_{h=1, j=k}^{n} I_{j,k}^{(h)} + \sum_{h=1, j=k}^{n} I_{j,-k}^{(h)} + \sum_{|j-k|=1} I_{j,k}^{(1)} + \sum_{|j-k|=1} I_{j,-k}^{(1)} = 1,
$$

$$
I_{j,k}^{(h)}, I_{j,-k}^{(h)} \ge 0, \forall j, k \in \{1, 2, \mathbf{K}, s\}, \forall t \in \{n-1, n, \mathbf{K}\},
$$

where $x_{t+1}^{(j)}$ is the state probability distribution at time $t+1$ in the kth sequence and $P_h^{(j,k)}$ the hth step transition probability matrix from the states at time $t-h+1$ in the kth sequence to the states at time $t+1$ in the jth 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 tri-diagonal parsimonious higher-order multivariate Markov chain model in matrix form has

$$
\begin{pmatrix} X_{t+1}^{(j)} \\ X_{t+1}^{(j)} \\ \mathbf{M} \\ X_{t+1}^{(j)} \end{pmatrix} = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \mathbf{K} & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \mathbf{K} & B^{(2,s)} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ B^{(s,1)} & B^{(s,2)} & \mathbf{K} & B^{(s,s)} \end{pmatrix} \begin{pmatrix} X_t^{(j)} \\ X_t^{(j)} \\ \mathbf{M} \\ X_t^{(j)} \end{pmatrix} + \frac{1}{m-1} \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \mathbf{K} & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \mathbf{K} & B^{(2,s)} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ B^{(s,1)} & B^{(s,2)} & \mathbf{K} & B^{(s,s)} \end{pmatrix} \begin{pmatrix} 1 - X_t^{(j)} \\ 1 - X_t^{(j)} \\ \mathbf{M} \\ 1 - X_t^{(j)} \end{pmatrix}
$$
(2)

where if $j = k$

$$
B^{(j,k)} = \begin{pmatrix} I_{j,k}^{(1)} P_1^{(j,k)} & I_{j,k}^{(1)} P_1^{(j,k)} & \mathbf{K} & I_{j,k}^{(1)} P_1^{(j,k)} & I_{j,k}^{(1)} P_1^{(j,k)} \\ I & 0 & \mathbf{K} & 0 & 0 \\ 0 & I & \mathbf{O} & 0 & 0 \\ \mathbf{M} & \mathbf{O} & \mathbf{O} & 0 & 0 \\ 0 & \mathbf{K} & 0 & I & 0 \end{pmatrix} \quad B^{(j,-k)} = \begin{pmatrix} I_{j,k}^{(1)} P_1^{(j,-k)} & I_{j,k}^{(1)} P_1^{(j,-k)} & \mathbf{K} & I_{j,k}^{(1)} P_1^{(j,-k)} & I_{j,k}^{(1)} P_1^{(j,-k)} \\ I & 0 & \mathbf{K} & 0 & 0 \\ 0 & I & \mathbf{O} & 0 & 0 \\ \mathbf{M} & \mathbf{O} & \mathbf{O} & 0 & 0 \\ 0 & \mathbf{K} & 0 & I & 0 \end{pmatrix}
$$

and if $|j - k| = 1,$

$$
B^{(j,k)} = \begin{pmatrix} I_{j,k}^{(1)} P_1^{(j,k)} & 0 & \mathbf{K} & 0 \\ 0 & 0 & \mathbf{K} & 0 \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ 0 & 0 & \mathbf{K} & 0 \end{pmatrix}_{mn \times mn}, \quad B^{(j,-k)} = \begin{pmatrix} I_{j,k}^{(1)} P_1^{(j,-k)} & 0 & \mathbf{K} & 0 \\ 0 & 0 & \mathbf{K} & 0 \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ 0 & 0 & \mathbf{K} & 0 \end{pmatrix}_{mn \times mn}
$$

and

$$B^{+} = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & 0 & K & 0 \\ B^{(2,1)} & B^{(2,2)} & B^{(2,3)} & 0 & 0 \\ M & O & O & O & M \\ 0 & K & 0 & B^{(s,s-1)} & B^{(s,s)} \end{pmatrix}_{ms \times ms}, B^{-} = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & 0 & K & 0 \\ B^{(2,1)} & B^{(2,2)} & B^{(2,3)} & 0 & 0 \\ M & O & O & O & M \\ 0 & K & 0 & B^{(s,s+1)} & B^{(s,s)} \end{pmatrix}_{ms \times ms}$$

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

Convergence condition

After t steps iteration of TPHOMMCM-NCC

$$X_{t+1} = BX_t + \frac{1}{m-1}C(e - X_t)$$

$$= (B - \frac{1}{m-1}C)^{t+1}X_0 + \sum_{k=0}^{t}(B - \frac{1}{m-1}C)^k \frac{1}{m-1}C \cdot e$$

If $r(B - \frac{1}{m-1}C) < 1$, the iteration of TPHOMMCM-NCC 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,

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

According to $\|B\| < 1$,

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

This theorem has been proved.

Parameter estimation

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

Let's first estimate the transition matrix. Probability transition matrix $P_h^{(j,k)}$ is

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

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

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

where

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

One would expect that

$$\left\| B^{+}X + \frac{1}{m-1}B^{-}(1-X) - X \right\| \leq w \tag{3}$$

where $w \geq 0$ and w is as small as possible.

Transform (3) into a minimization problem:

$$
\begin{cases}
\min\limits_{I_{j,k}} \left\| B^{+} X + \dfrac{1}{m-1} B^{-}(1-X) - X \right\| \\[2mm]
\text{subject to } \sum\limits_{h=1,j=k}^{n} I_{j,k}^{(h)} + \sum\limits_{h=1,j=k}^{n} I_{j,-k}^{(h)} + \sum\limits_{|j-k|=1} I_{j,k}^{(1)} + \sum\limits_{|j-k|=1} I_{j,-k}^{(1)} = 1 \\[2mm]
B > \dfrac{1}{m-1} C
\end{cases}
\tag{4}
$$

Application to sales demand prediction with tri-diagonal parsimonious higher-order multivariate Markov chain model

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

The storage space of a soft drink company in Hong Kong is often in the state of overflow or near capacity. 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 hanging over the company's head. Predicting the sales demand with a more precise Markov chain model is beneficial for minimizing the estate cost.

We classifies the sales demand into six 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 states in the same customer group of five important products of the company for a year is given in the Appendix [11].

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

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

n	M1 pn	M1 pe	M2 pn	M2 pe	M3 pn	M3 pe
1	13	314.67	25	259.35	13	269.75
2	18	298.10	50	247.36	18	259.05
3	23	297.98	75	246.31	23	257.68
4	28	298.42	100	246.31	28	257.67
5	33	302.22	125	246.31	33	238.81
6	38	301.40	150	247.16	38	239.28
7	43	301.97	175	246.91	43	239.28
8	48	299.97	200	246.91	48	240.51

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

In Tables, denote that a 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 and $M3$ the tri-diagonal parsimonious higher-order multivariate Markov chain model.

From the results of Figure 1,2,3, we find that the tri-diagonal 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.

Conclusions

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

Acknowledgments

References

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

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

[4] 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).

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

[6] 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).

[7] W. Ching, E. Fung, M. Ng. "A multivariate Markov chain model for categorical data sequences and its applications in demand prediction", *IMA Journal of Management Mathematics*, **13**, pp. 187-199, (2002) .

[8] 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) .

[9] 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).

[10] 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).

[11] 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) .

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

[13] C. Wang, T.Z. Huang. "A new improved parsimonious multivariate Markov chain model", *Journal of Applied Mathematics*, **1**, pp. 1-10, (2013)