FORECASTING FUZZY TIME SERIES MODEL ON TWO FACTOR HIGHER ORDER MULTIVARIATE MARKOV CHAIN

D.Jeyabalan Kennedy*

V.Vamitha**

S.Rajaram***

Abstract

During the last decade different models have been designed and developed.Vamithaet.al(2012) proposed a fuzzy time series model to predict thetemperature from correlated categorical data sequence obtained from similar source. In this paper,we propose a new fuzzy time series model called two factors time variant fuzzy time series forecasting based on higher order multivariate Markov chain.Two numerical data sets namely temperature and cloud density of Taipei is selected to illustrate the proposed methodand compare with some of the existing fuzzy time series methods.The proposed method gets higher average forecasting accuracy rate.

Keywords: Time variant,Multivariate Markov chain,Categorical data sequence,Higher orderfuzzy time series.

AMS Mathematics Subject Classification (2010) : 03E72, 62M10,90C40.

* Department of Mathematics; Nazareth Margoschis college; Nazreth; Tamilnadu; India – 628 617

** Research Scholar, Manonmaniam Sundaranar University, Tirunelveli (Department of Mathematics, B.C.M.W.G.Polytehnic, Ettayapuram-628 902) Tamil Nadu, India.

*** Department of Mathematics ; Sri S.R.N.M.College, Sattur-626 203; Tamilnadu; India.

1. Introduction

A model is a simplification of the real world,with only the most important elements included.A good model enables us to workout what is likely to happen if we take certain action,so we can avoid actions which have undesirable consequences.The study of fuzzy models has been of considerable active interest ever since the birth of fuzzy theory.Its progress and developmentare ever growing.The newer areas of application are emerging;there is a large and growing audience interested in the study of fuzzy models.This paper concerns with the time series comprised of imprecise.i.e.,uncertain observed values.In this case of time series the uncertainty of the individual observed values as well as the interpretation of a sequence of uncertain observed values are of interest.The uncertain observed value is thus modeled as a fuzzy variable. In this paper,we propose a new fuzzy time series model based on higher order multivariate Markov chain on categorical data sequences. Data sequences that have a correlation with each other are used.As a result of this,predictions which are very close to the reality can be made,by making use of transition probability matrix. A categorical data sequence of m states can be modeled by m state Markov chain model.The above idea can be extended to model multiple categorical data sequences. The traditional time series forecasting methods cannot be used for forecasting problems in which the historical data are linguistic values. Song and Chissom (1993, 1994) proposed time variant and time invariant fuzzy time series models and fuzzy forecasting to model and forecast processes whose observation are linguistic values. Instead of complicated maximum minimum composition operations Chen (1996) used a simple arithmetic operation for time series forecasting. Thereafter a number of related research works have been reported that follow their framework and aim to improve forecasting accuracy and reduce the computational overhead. These works include enrollments, length of intervals, temperature prediction, weighted method, stock price, hidden Markov model, genetic algorithm, neural – fuzzy system, bulk shipping, seasonal, heuristic models .

2. Fuzzy time series

In the following, we briefly review some basic concepts of fuzzy time series from Song and Chissom (1993, 1994) and its forecasting frame work.

Definition 1: A *fuzzy set* A is defined as an uncertain subset of the fundamental set X.

$$A = \{(x, \mu_A(x)) | x \epsilon X\}$$

The uncertainty is assessed by the membership function $\mu_A(x)$.

Definition 2: Let Y(t) {t = 0,1,2,3,…}, a subset of $\mathbb{R}$, be the universe of discourse on which fuzzy sets $f_i(t)$ (i = 1,2,3…) are defined and let F(t) be the collection of $f_i(t)$. Then F (t) is defined as *fuzzy time series* on Y(t).

From this definition we can see that, (1) F(t) is the function of time.

(2) F(t) can be regarded as a *linguistic variable*, which is a variable whose values are linguistic values represented by fuzzy sets.

(3) $f_i(t)$ (i = 1,2,3…) are possible linguistic values of F(t), where $f_i(t)$ (i = 1,2,3,….) are represented by fuzzy sets.

Song and Chissom employed a fuzzy relational equation to develop their forecasting model under the assumption that the observations at time t are dependent only upon the accumulated results of the observation at previous times, which is defined as follows.

Definition 3: Suppose F(t) is caused only by F(t-1) and is denoted by $F(t-1) \rightarrow F(t)$, then there is a fuzzy relationship between F(t) and F(t-1) and can be expressed as the fuzzy relational equation $F(t) = F(t-1) \circ R(t,t-1)$. Here '∘' is max- min composition operator. The relation R is called first order model of F(t).

Further, if fuzzy relation R(t,t-1) of F(t) is independent of time t, that is to say, for different times t_1 and t_2, $R(t_1,t_1-1) = R(t_2,t_2-1)$, then F(t) is called a *time invariant* fuzzy time series otherwise F(t) is *time variant.*

Definition 4: Suppose $F(t-1) = A_i$ and $F(t) = A_j$ a *fuzzy logical relationship* can be defined as $A_i \rightarrow A_j$ where A_i and A_j are called the left hand side and the right hand side of the fuzzy logical relationship respectively.

Definition 5: If F(t) is caused by more fuzzy sets F(t-n), F(t-n+1), … , F(t-1) the fuzzy relationship is represented by $A_{i1}, A_{i2}, A_{i3},\ldots A_{in} \rightarrow A_j$, where $F(t-n) = A_{i1}$, $F(t-n+1) = A_{i2}, \ldots, F(t-1) = A_{in}$. This relationship is called n^{th} order fuzzy time series model.

3.The Higher order multivariate Markov Chain model

We briefly review some basic concepts of a higher order multivariate Markov chain model to represent the behavior of multiple categorical data sequences generated by similar sources or the same source proposed by Ching et al., (2002). Consider Markov chains having finite number of states $\mathcal{M} = \{1, 2, \ldots, m\}$. In general, a categorical data sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T$ can be logically represented by a sequence of vectors $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_T}$ where T is the length of the sequence, and $\mathbf{x_i} = e_k$ (e_k is the unit vector with the k^{th} entry being one) if it is in state k. A first-order discrete-time Markov chain having m discrete states satisfies the following relationship:

$\Pr(\mathbf{x}_{t+1} = e_{x_{t+1}} | \mathbf{x}_0 = e_{x_0}, \mathbf{x}_1 = e_{x_1}, \ldots, \mathbf{x}_t = e_{x_t}) = \Pr(\mathbf{x}_{t+1} = e_{x_{t+1}} | \mathbf{x}_t = e_{x_t})$

where $x_i \in \mathcal{M}$. The conditional probabilities $\Pr(\mathbf{x}_{n+1} = e_{x_{n+1}} | \mathbf{x}_n = e_{x_n})$ are called the single-step transition probabilities of the Markov chain. They give the conditional probability of making a transition from state i to state j when the time parameter increases from n to n+1. These probabilities are independent of n and are written as

$p_{ij} = \Pr(\mathbf{x}_{n+1} = e_i | \mathbf{x}_n = e_j). \forall\, i, j \in \mathcal{M}.$

The matrix P, formed by placing p_{ij} in row i and column j for all i and j, is called the transition probability matrix. We note that the elements of the matrix P satisfy the following two properties: $0 \le p_{ij} \le 1\ \forall\, i, j \in \mathcal{M}$ and $\sum_{j=1}^{m} p_{ij} = 1,\ \forall\ j \in \mathcal{M}$. We assume that p_{ij} are not all zero for each j.

Here we assume that there are s categorical sequences and each has m possible states in $\mathcal{M}$. Let $\mathbf{x}_r^{(k)}$ be the state vector of the k^{th} sequence at time r. If the k^{th} sequence is in state j at time r, then $\mathbf{x}_r^{(k)} = e_j = (0, \ldots, 0, \underbrace{1}_{jth\,entry}, 0, \ldots, 0)^T$.

In the higher order multivariate Markov chain model, we assume the following relationship:

$X_{r+1}^{(j)} = \sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} P_h^{(jk)} X_{r-h+1}^{(k)}$, for j=1,2,…s., r = n-1,n,…

where $\lambda_{jk}^{(h)} \ge 0$, $1 \le j, k \le s$, $1 \le h \le n$ and

$\sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1$ for j = 1, 2,..,s.

The state probability distribution of the j^{th} sequence $X_{r+1}^{(j)}$ at time r+1 ,depends on the weighted average of $P_h^{(jk)} X_{r-h+1}^{(k)}$.Here $P_h^{(jk)}$ is the h^{th}step transition probability matrix from the states in the k^{th}sequence at time r-h+1 to the states in the j^{th} sequence at time r+1.In the matrix form, we write

$$X_{r+1} \equiv \begin{pmatrix} X_{r+1}^{(1)} \\ X_{r+1}^{(2)} \\ \vdots \\ X_{r+1}^{(s)} \end{pmatrix} = \begin{pmatrix} B^{(11)} & B^{(12)} & \cdots & B^{(1s)} \\ B^{(21)} & B^{(22)} & \cdots & B^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s1)} & B^{(s2)} & \cdots & B^{(ss)} \end{pmatrix} \begin{pmatrix} X_r^{(1)} \\ X_r^{(2)} \\ \vdots \\ X_r^{(s)} \end{pmatrix} \equiv Q X_r \text{ or } X_{r+1} = Q X_r$$

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

and if $i \neq j$ then

$$B^{(ij)} = \begin{pmatrix} \lambda_{ij}^{(1)} P_1^{(ij)} & \lambda_{ij}^{(2)} P_2^{(ij)} & \cdots & \lambda_{ij}^{(n-1)} P_{n-1}^{(ij)} & \lambda_{ij}^{(n)} P_n^{(ij)} \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix}_{mn \times mn} \text{ and}$$

$$Q = \begin{pmatrix} B^{(11)} & B^{(12)} & \cdots & B^{(1s)} \\ B^{(21)} & B^{(22)} & \cdots & B^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s1)} & B^{(s2)} & \cdots & B^{(ss)} \end{pmatrix}.$$

We note that each column sum of Q is not necessarily equal to one but each column sum of $P_h^{(jk)}$ equal to one.

4 . Proposed Model

In this section we introduce a higher order model to forecast the Temperature of Taipei. The historical data of daily average Temperature and the daily cloud density of Taipei from June 1996 to September 1996 are considered. The step-wise procedure of the proposed model of fuzzy time series is detailed as follows:

Step 1:Define the universe of discourse U = [low, up] and V = [low, up] which can cover all observations of Temperature and cloud density respectively in the months of June 1996 to September 1996 of historical data set. Initially partition the universe of discourse into seven linguistic intervals u_i and v_i, i = 1,2,…,7, of equal length.

Step2: Temperatures are categorized into seven possible states. Define fuzzy sets A_1 (very very low), A_2 (very low), A_3 (low), A_4 (normal), A_5 (high), A_6 (very high), A_7 (very very high). Construct the fuzzy sets A_i in accordance with the intervals in step 1. Fuzzify the historical data. For n fuzzy sets, A_1, $A_{2,..}A_n$ can be defined on U as follows:

$A_i = \sum_{j=1}^{n} \frac{\mu_{ij}}{v_j}$ where μ_{ij} is the membership degree of A_i belonging to v_j and is

defined by $\mu_{ij} = \begin{cases} 1 & if \quad j = i \\ 0.5 & if \quad j = i-1 \text{ or } i+1 \\ 0 & if \ otherwise \end{cases}$

Then, for a given historical datum Y_t, its membership degree belonging to interval v_i is determined by the following heuristic rules.

Rule 1: if Y_t is located at v_1, the membership degrees are 1 for v_1, 0.5 for v_2 and 0 otherwise.

Rule 2: if Y_t belongs to v_i, $1 < i < n$, then the degrees are 1, 0.5 and 0.5 for v_i, v_{i-1} and v_{i+1}, respectively and 0 otherwise.

Rule 3: if Y_t is located at v_n, the membership degrees are 1 for v_n, 0.5 for v_{n-1} and 0 otherwise. Then, Y_t is fuzzified as A_j, where the membership degree in interval j is maximal.

Similarly fuzzify the values of cloud density of the corresponding month given in the historical data set.Name the fuzzy sets as D_1, $D_{2...}$,D_n .

Step 3: Represent the subscripts of the fuzzy sets obtained from the four months of temperature data as the members of the categorical data sequences T_1, T_2, T_3, T_4 and from the four months of cloud density data as the members of the categorical data sequences C_1, C_2, C_3, C_4.

Step 4: Given the data sequence,we count the transition frequency $f_{i_j i_k}^{(jk,h)}$ from the state i_k in the sequence $\{X^{(k)}\}$ at time r – h+1 to the state i_j in the sequence $\{X^{(j)}\}$ at time r+1 and therefore we construct the transition frequency matrix for the sequence as follows:

$$F_h^{(jk)} = \begin{pmatrix} f_{11}^{(jk,h)} & \cdots & \cdots & f_{1m}^{(jk,h)} \\ f_{21}^{(jk,h)} & \cdots & \cdots & f_{2m}^{(jk,h)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m1}^{(jk,h)} & \cdots & \cdots & f_{mm}^{(jk,h)} \end{pmatrix}$$

Step 5: After the normalization,the estimates of the transition probability matrices from $F_h^{(jk)}$ can also be obtained as follows:

$$\hat{P}_h^{(jk)} = \begin{pmatrix} \hat{p}_{11}^{(jk,h)} & \dots & \dots & \hat{p}_{1m}^{(jk,h)} \\ \hat{p}_{21}^{(jk,h)} & \dots & \dots & \hat{p}_{2m}^{(jk,h)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{p}_{m1}^{(jk,h)} & \dots & \dots & \hat{p}_{mm}^{(jk,h)} \end{pmatrix}$$

where $\hat{p}_{i_j i_k}^{(jk,h)} = \begin{cases} \dfrac{f_{i_j i_k}^{(jk,h)}}{\sum_{i_j=1}^{m} f_{i_j i_k}^{(jk,h)}} & \text{if } \sum_{i_j=1}^{m} f_{i_j i_k}^{(jk,h)} \neq 0 \\ 0 & \textit{otherwise.} \end{cases}$

Step 6: We need to estimate the parameters $\lambda_{jk}^{(h)}$ .The stationary vector $\hat{\mathbf{x}}$ can be estimated from the sequences by computing the proportion of the occurrence of each state in each of the sequences,and let us denote it by $\hat{\mathbf{x}} = [\hat{\mathbf{x}}^{(1)}, \hat{\mathbf{x}}^{(2)}, \dots, \hat{\mathbf{x}}^{(s)}]^T$ .One would expect

$$\begin{pmatrix} \hat{B}^{(11)} & \hat{B}^{(12)} & \dots & \hat{B}^{(1s)} \\ \hat{B}^{(21)} & \hat{B}^{(22)} & \dots & \hat{B}^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{B}^{(s1)} & \hat{B}^{(s2)} & \dots & \hat{B}^{(ss)} \end{pmatrix} \hat{\mathbf{x}} \approx \hat{\mathbf{x}}$$

From the above equation,it suggests one possible way to estimate the parameters $\lambda = \{\lambda_{jk}^{(h)}\}$ as follows.One may consider solving the following optimization problem:

$$\min_{\lambda_{ij}} \max_i \left\{ \left| \left[\sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} \hat{P}_h^{(jk)} \hat{\mathbf{x}}^{(k)} - \hat{\mathbf{x}}^{(j)} \right]_i \right| \right\}$$

subject to $\sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1$ and $\lambda_{jk}^{(h)} \geq 0, \forall\, h,k.$

Step 7: Formulate s linear programming problems from the above optimization problem as follows : For each j: $\min_{\lambda} w_j$ subject to

$$\begin{cases} \begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geq \hat{\mathbf{x}}^{(j)} - C_j \begin{pmatrix} \bar{\lambda}_{j1} \\ \bar{\lambda}_{j2} \\ \vdots \\ \bar{\lambda}_{js} \end{pmatrix}, \begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geq -\hat{\mathbf{x}}^{(j)} + C_j \begin{pmatrix} \bar{\lambda}_{j1} \\ \bar{\lambda}_{j2} \\ \vdots \\ \bar{\lambda}_{js} \end{pmatrix}, \\ w_j \geq 0, \\ \sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1,\ \lambda_{jk}^{(h)} \geq 0, \forall\, h,j,k. \end{cases}$$

where $C_j = \left[\hat{p}_1^{(j1)} \hat{x}^{(1)} \middle| \dots \middle| \hat{p}_n^{(j1)} \hat{x}^{(1)} \middle| \hat{p}_1^{(j2)} \hat{x}^{(2)} \middle| \dots \middle| \hat{p}_n^{(j2)} \hat{x}^{(2)} \middle| \dots\dots \middle| \hat{p}_1^{(js)} \hat{x}^{(s)} \middle| \dots \middle| \hat{p}_n^{(js)} \hat{x}^{(s)} \right]$

and $\bar{\lambda}_{jh} = (\lambda_{jh}^{(1)} \dots\dots \lambda_{jh}^{(n)})^T$.

Step 8: Construct the models using the values obtained in step 6 and 7 and forecast the vector to the historical data.Compare with the previous forecast vector of temperature and the current forecast vector of cloud density by taking maximum among the vector elements.Calculate the forecasted values as follows:Forecasted value=$\sum_{i=1}^{7} v_i\, m_i$ where v_i are entries of the resulting vector and m_i are the midpoints of the corresponding u_i .

Step 9: Choose β in (0,1)and make an error analysis as follows:

New forecast value=Actual value ∗(1– β) + Forecasted value∗ β.

Compute Root Mean Square Error[RMSE] = $\sqrt{\sum_{i=1}^{n}\frac{(actual_i - forecast_i)^2}{n}}$ for different β values on new forecasted values. Fix β corresponding to minimum RMSE value.The forecasted values with respect to this β are the expected forecasted values.

Step 10: Compute average forecasting error rate (AFER)= $\frac{1}{n}\sum_{i=1}^{n}\frac{|A_i - F_i|}{A_i}$ *100.Compare this model with some of the existing models with respect to AFER values. In the next section,we give an example to demonstrate the construction of a two factor higher order multivariate Markov model using fuzzy time series from four categorical data sequences of Temperature and cloud density.

5.Performance evaluation of the model:

The four categorical data sequences obtained from the temperature data as follows:

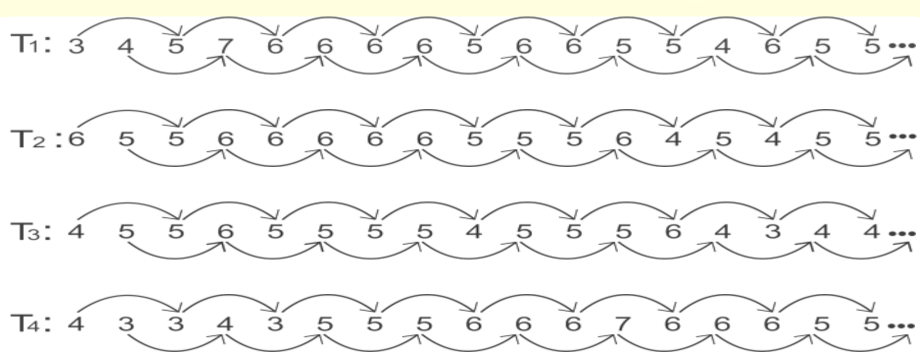
T_1 = {3,4,5,7,6,6,6,6,5,6,6,5,5,4,6,5,5,6,6,7,7,5,4,4,4,4,5,4,5,6}

T_2 = {6,5,5,6,6,6,6,6,5,5,5,6,4,5,4,5,5,6,7,7,7,7,7,7,5,4,5,4,6,4}

T_3 = {4,5,5,6,5,5,5,5,4,5,5,5,6,4,3,4,4,5,5,6,6,6,4,5,5,5,5,4,3,3} and

T_4 = {4,3,3,4,3,5,5,5,6,6,6,7,6,6,6,5,5,5,5,5,3,3,2,4,3,3,2,1,1,1}

By counting the transition frequencies of second order multivariate Markov chain model,



T₁: 3 4 5 7 6 6 6 6 5 6 6 5 5 4 6 5 5 6

T₂ : 6 5 5 6 6 6 6 6 5 5 5 6 4 5 4 5 5 6 ...

T₃: 4 5 5 6 5 5 5 5 4 5 5 5 6 4 3 4 4 5

T₄: 4 3 3 4 3 5 5 5 6 6 6 7 6 6 6 5 5 5 ...

$$F_2^{(11)} = \begin{pmatrix} 0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0\\ 0&0&0&3&2&0&1\\ 0&0&1&2&1&4&1\\ 0&0&0&1&5&3&1\\ 0&0&0&1&0&2&0 \end{pmatrix} \ldots F_2^{(44)} = \begin{pmatrix} 1&1&1&0&0&0&0\\ 0&0&2&0&0&0&0\\ 0&1&1&2&2&0&0\\ 0&0&2&0&0&0&0\\ 0&0&1&1&4&2&0\\ 0&0&0&0&2&3&1\\ 0&0&0&0&0&1&0 \end{pmatrix}$$

After the normalization we have the transition probability matrices

$$\hat{P}_2^{(11)} = \begin{pmatrix} \frac{1}{7}&\frac{1}{7}&0&0&0&0&0\\ \frac{1}{7}&\frac{1}{7}&0&0&0&0&0\\ \frac{1}{7}&\frac{1}{7}&0&0&0&0&0\\ \frac{1}{7}&\frac{1}{7}&0&\frac{3}{7}&\frac{2}{8}&0&\frac{1}{3}\\ \frac{1}{7}&\frac{1}{7}&1&\frac{2}{7}&\frac{1}{8}&\frac{4}{9}&\frac{1}{3}\\ \frac{1}{7}&\frac{1}{7}&0&\frac{1}{7}&\frac{5}{8}&\frac{3}{9}&\frac{1}{3}\\ \frac{1}{7}&\frac{1}{7}&0&\frac{1}{7}&0&\frac{2}{9}&0 \end{pmatrix} \ldots \hat{P}_2^{(44)} = \begin{pmatrix} 1&\frac{1}{2}&\frac{1}{7}&0&0&0&0\\ 0&0&\frac{2}{7}&0&0&0&0\\ 0&\frac{1}{2}&\frac{1}{7}&\frac{2}{3}&\frac{2}{8}&0&0\\ 0&0&\frac{2}{7}&0&0&0&0\\ 0&0&\frac{1}{7}&\frac{1}{3}&\frac{4}{8}&\frac{2}{6}&0\\ 0&0&0&0&\frac{2}{8}&\frac{3}{6}&1\\ 0&0&0&0&0&\frac{1}{6}&0 \end{pmatrix}$$

Moreover,by counting the inter-transition frequencies, we have

$$F_2^{(12)} = \begin{pmatrix} 0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0\\ 0&0&0&1&0&1&4\\ 0&0&0&1&5&2&1\\ 0&0&0&2&4&1&0\\ 0&0&0&0&1&4&1 \end{pmatrix} \ldots F_2^{(43)} = \begin{pmatrix} 0&0&0&1&2&0&0\\ 0&0&0&0&1&1&0\\ 0&0&0&2&3&1&0\\ 0&0&0&0&1&1&0\\ 0&0&1&3&3&1&0\\ 0&0&0&1&4&1&0\\ 0&0&0&0&1&0&0 \end{pmatrix}$$

After the normalization we have the transition probability matrices

$$\hat{P}_2^{(12)} = \begin{pmatrix} \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{4} & 0 & \frac{1}{8} & \frac{4}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{4} & \frac{5}{10} & \frac{2}{8} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{2}{4} & \frac{4}{10} & \frac{4}{8} & 0 \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & \frac{1}{10} & \frac{1}{8} & \frac{1}{6} \end{pmatrix} \ldots \hat{P}_2^{(43)} = \begin{pmatrix} \frac{1}{7} & \frac{1}{7} & 0 & \frac{1}{7} & \frac{2}{15} & 0 & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & 0 & 0 & \frac{1}{15} & \frac{1}{5} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & 0 & \frac{2}{7} & \frac{3}{15} & \frac{1}{5} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & 0 & 0 & \frac{1}{15} & \frac{1}{5} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & 1 & \frac{3}{7} & \frac{3}{15} & \frac{1}{5} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & 0 & \frac{1}{7} & \frac{4}{15} & \frac{1}{5} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & 0 & 0 & \frac{1}{15} & 0 & \frac{1}{7} \end{pmatrix}$$

By solving the corresponding minimization problems, through linear programming

$$w_4 \geq \frac{1}{10} - \frac{1}{10}\lambda_{41}^{(1)} - \frac{5}{48}\lambda_{41}^{(2)} - \frac{13}{120}\lambda_{42}^{(1)} - \frac{37}{240}\lambda_{42}^{(2)} - \frac{7}{60}\lambda_{43}^{(1)} - \frac{1}{10}\lambda_{43}^{(2)} - \frac{2}{15}\lambda_{44}^{(1)} - \frac{1}{6}\lambda_{44}^{(2)}$$

$\vdots$ {seven equations}

$$w_4 \geq \frac{1}{30} - \frac{1}{27}\lambda_{41}^{(1)} - \frac{1}{27}\lambda_{41}^{(2)} - \frac{1}{30}\lambda_{42}^{(1)} - \frac{1}{30}\lambda_{42}^{(2)} - \frac{1}{30}\lambda_{43}^{(1)} - \frac{1}{30}\lambda_{43}^{(2)} - \frac{1}{30}\lambda_{44}^{(1)} - \frac{1}{30}\lambda_{44}^{(2)}$$

and

$$w_4 \geq -\frac{1}{10} + \frac{1}{10}\lambda_{41}^{(1)} + \frac{5}{48}\lambda_{41}^{(2)} + \frac{13}{120}\lambda_{42}^{(1)} + \frac{37}{240}\lambda_{42}^{(2)} + \frac{7}{60}\lambda_{43}^{(1)} + \frac{1}{10}\lambda_{43}^{(2)} + \frac{2}{15}\lambda_{44}^{(1)} + \frac{1}{6}\lambda_{44}^{(2)}$$

$\vdots$ {seven equations}

$$w_4 \geq -\frac{1}{30} + \frac{1}{27}\lambda_{41}^{(1)} + \frac{1}{27}\lambda_{41}^{(2)} + \frac{1}{30}\lambda_{42}^{(1)} + \frac{1}{30}\lambda_{42}^{(2)} + \frac{1}{30}\lambda_{43}^{(1)} + \frac{1}{30}\lambda_{43}^{(2)} + \frac{1}{30}\lambda_{44}^{(1)} + \frac{1}{30}\lambda_{44}^{(2)}$$

By step 7,we obtain the optimal solution as follows:

$$\Lambda = [\lambda_{jk}] = \begin{pmatrix} 0.7930 & 0.0000 & 0.2070 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.7347 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.2653 & 0.0000 \\ 0.4286 & 0.0000 & 0.1905 & 0.0000 & 0.0000 & 0.3809 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix}$$

By step 8, the second order multivariate Markov model for four categorical data sequence obtained from Temperature is as follows:

$$\begin{cases} x_{r+1}^{(1)} = 0.7930\hat{P}_1^{(11)}x_r^{(1)} + 0.2070\hat{P}_1^{(12)}x_r^{(2)} \\ x_{r+1}^{(2)} = 0.7347\hat{P}_2^{(21)}x_{r-1}^{(1)} + 0.2653\hat{P}_1^{(24)}x_r^{(4)} \\ x_{r+1}^{(3)} = 0.4286\hat{P}_1^{(31)}x_r^{(1)} + 0.1905\hat{P}_1^{(32)}x_r^{(2)} + 0.3809\hat{P}_2^{(33)}x_{r-1}^{(3)} \\ x_{r+1}^{4} = 1.0000\hat{P}_2^{(41)}x_{r-1}^{(1)} \end{cases}$$

The four categorical data sequence obtained from the daily cloud density data as follows:

C_1 = {3,2,2,1,1,2,4,3,2,2,3,3,2,2,2,4,4,2,1,4,5,7,5,2,1,2,2,4,2,2}

C_2 = {1,3,2,3,2,2,4,3,1,1,3,1,2,3,2,2,2,2,2,3,2,4,4,3,7,7,7,6,3,4}

C_3 = {7,6,5,3,4,7,5,2,5,3,4,6,4,6,5,6,3,2,2,2,2,4,3,3,2,3,3,1,2,3}

C_4 = {2,4,5,4,4,5,3,6,4,3,3,2,1,4,3,2,2,3,1,1,3,5,5,5,3,2,3,5,6,3}

The second order multivariate Markov model for four categorical data sequence obtained from cloud density is as follows:

$$\begin{cases} x_{r+1}^{(1)} = 0.0700\hat{P}_2^{(11)}x_{r-1}^{(1)} + 0.1500\hat{P}_1^{(12)}x_r^{(2)} + 0.3000\hat{P}_2^{(13)}x_{r-1}^{(3)} + 0.4800\hat{P}_1^{(14)}x_r^{(4)} \\ x_{r+1}^{(2)} = 0.3750\hat{P}_1^{(22)}x_r^{(2)} + 0.6250\hat{P}_2^{(24)}x_{r-1}^{(4)} \\ x_{r+1}^{(3)} = 0.7201\hat{P}_2^{(32)}x_{r-1}^{(2)} + 0.2799\hat{P}_2^{(33)}x_{r-1}^{(3)} \\ x_{r+1}^{(4)} = 0.6522\hat{P}_2^{(41)}x_{r-1}^{(4)} + 0.3478\hat{P}_1^{(44)}x_r^{(4)} \end{cases}$$

By step 8 , for r = 2, the forecast vector of the third position in the second sequence of temperature is$x_3^{(2)} = (0,0,0,0.03,0.81,0.08,0.08)^T$

and the forecast vector of fourth position in the second sequence of the cloud density is

$x_4^{(2)} = (0,0.44, 0.36, 0.20, 0, 0, 0)^T$.

$\text{Max}[x_3^{(2)}, x_4^{(2)}] = (0, 0.44, 0.36, 0.20, 0.81, 0.08, 0.08)^T$

The forecasting value corresponding to fourth position in the second sequence by step 9 is

$$\frac{(0\times m1)+(0.44\times m2)+(0.36\times m3)+(0.20\times m4)+(0.81\times m5)+(0.08\times m6)+(0.08\times m7)}{0.44+0.36+0.20+0.81+0.08+0.08}$$

= 27.5. By step 9, we find β = 0.1 is chosen and the new forecast value is 29.21.

6. Conclusion

In this paper, we have proposed a new forecasting model based on higher order multivariate Markov chain on categorical data sequences for forecasting the daily average temperature of the Taipei, Taiwan. Through the fuzzification of the temperature data we obtained four categorical data sequences and on which higher order multivariate Markov chain on categorical data sequences is applied. In performance evaluation we have computed four orders but for convenience only the second order multivariate Markov chain model is tabulated. From the experimental results the proposed method provides the smallest AFER (see Table 2) and improves on other existing methods using fuzzy times series forecasting.

Table 1: Actual values of temperature, cloud density and forecasted values of temperature(2nd order model)

	JUNE			JULY			AUGUST			SEPTEMBER		
DAY	Actual value temperature	Actual cloud density value	Forecast value --	Actual value temperature	Actual cloud density value	Forecast value --	Actual value temperature	Actual cloud density value	Forecast value --	Actual value temperature	Actual cloud density value	Forecast value --
1	26.1	36	-	29.9	15	-	27.1	100	-	27.5	29	-
2	27.6	23	-	28.4	31	-	28.9	78	-	26.8	53	-
3	29.0	23	-	29.2	26	-	28.9	68	-	26.4	66	-
4	30.5	10	30.14	29.4	34	29.21	29.3	44	29.18	27.5	50	27.43
5	30.0	13	29.71	29.9	24	26.69	28.8	56	28.66	26.6	53	26.58
6	29.5	30	29.28	29.6	28	29.41	28.7	89	28.65	28.2	63	28.13
7	29.7	45	29.49	30.1	50	29.91	29.0	71	28.84	29.2	36	28.96
8	29.4	35	29.25	29.3	34	29.19	28.2	28	28.11	29.0	76	28.87
9	28.8	26	28.66	28.1	15	28.05	27.0	70	27.06	30.3	55	30.06
10	29.4	21	29.21	28.9	8	28.71	28.3	44	28.26	29.9	31	29.71
11	29.3	43	29.11	28.4	36	28.31	28.9	48	28.86	29.9	31	29.68
12	28.5	40	28.42	29.6	13	29.39	28.1	76	28.13	30.5	25	30.18
13	28.7	30	28.58	27.8	26	27.76	29.9	50	29.72	30.2	14	29.95
14	27.5	29	27.50	29.1	44	28.93	27.6	84	27.69	30.3	45	30.04
15	29.5	30	29.25	27.7	25	27.68	26.8	69	26.85	29.5	38	29.29
16	28.8	46	28.65	28.1	24	28.06	27.6	78	27.63	28.3	24	28.20
17	29.0	55	28.85	28.7	26	28.55	27.9	39	27.84	28.6	19	28.34
18	30.3	19	29.99	29.9	25	29.69	29.0	20	28.81	28.1	39	28.06
19	30.2	15	29.92	30.8	21	30.50	29.2	24	29.00	28.4	14	28.30
20	30.9	56	30.62	31.6	35	31.22	29.8	25	29.54	28.3	3	28.21
21	30.8	60	30.53	31.4	29	31.03	29.6	19	29.36	26.4	38	26.54
22	28.7	96	28.63	31.3	48	30.95	29.3	46	29.19	25.7	70	25.90
23	27.8	63	27.81	31.3	53	31.00	28.0	41	27.95	25.0	71	25.19
24	27.4	28	27.39	31.3	44	31.04	28.3	34	28.23	27.0	70	27.04
25	27.7	14	27.61	28.9	100	28.80	28.6	29	28.49	25.8	40	25.99
26	27.1	25	27.10	28.0	100	28.07	28.7	31	28.62	26.4	30	26.38
27	28.4	29	28.29	28.6	91	28.54	29.0	41	28.79	25.6	34	25.65
28	27.8	55	27.74	28.0	84	28.01	27.7	14	27.61	24.2	59	24.40
29	29.0	29	28.82	29.3	38	29.10	26.2	28	26.26	23.3	83	23.61

30	30.2	19	29.86	27.9	46	27.9	26.0	33	26.13	23.5	38	23.91
31	-	-	-	26.9	95	-	27.7	24	-	-	-	-

Table 2: Comparative values of AFER (In percentage)

Month	Lee et.al (2006)	Lee et.al (2007)	Wang and Chen(2009)	Vamitha et.al (2012)	Proposed model			
					I order	II order	III order	IV order
June	1.44	1.24	0.53	0.54	0.54	0.52	0.52	0.46
July	1.59	1.23	0.71	`0.87	0.49	0.50	0.51	0.54
August	1.26	1.09	0.32	0.69	0.31	0.33	0.32	0.32
Septem.	1.89	1.28	0.74	0.83	0.53	0.57	0.57	0.55

References

1. Chen, S. M. Forecasting enrollments based on fuzzy time-series. Fuzzy Sets and Systems, 81,(1996) 311–319.
2. Chen, S.M. Forecasting enrollments based on high-order fuzzy time series. Cybern. Syst., Int. J., 33(1), (2002),1–16.
3. Hsu, Y.Y., Tse, S.M. and Wu, B. A new approach of bivariate fuzzy time series analysis to the forecasting of a stock index. Int. J. Uncertain. Fuzziness Knowl.-Based Syst., 11(6), (2003),671–690.
4. Huarng, T.H.K. Yu, T.H.K. and Hsu, Y.W. A multivariate heuristic model for fuzzy time-series forecasting. IEEE Trans. Syst., Man, Cybern. B, Cybern., 37(4), (2007),836–846.
5. Lee, L.W. Wang, L.H. Chen, S.M. and Leu, Y.H. Handling forecasting problems based on two-factor high-order fuzzy time series. IEEE Trans. Fuzzy Syst., 14(3), (2006),468–477.
6. Lee, L.W. Wang, L.H. and Chen, S.M.Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms. Expert Syst. Appl., 33(3), (2007),539–550.
7. Own, C .M. and Yu, P.T. Forecasting fuzzy time series on a heuristic high order model. Cybernetics and systems: An International Journal, 36,(2005), 705-717.
8. Rajaram,S. & Sakthisree,P. Forecasting enrollments using interval length, weightage factor and fuzzy logical reasoning. International journal of statistics and systems, 5(2).(2010),173-182.
9. Song, Q., & Chissom, B. S. Fuzzy forecasting enrollments with fuzzy timeseries – Part 1. Fuzzy Sets and Systems, 54, (1993a),1–9.

10. Song, Q., & Chissom, B. S. Fuzzy time series and its models. Fuzzy Sets and Systems, 54,(1993b), 269–277.
11. Song, Q., & Chissom, B. S. Fuzzy forecasting enrollments with fuzzy time series – Part 2. Fuzzy Sets and Systems, 62, (1994),1–8.
12. Vamitha.V., Jeyanthi.M.,Rajaram.S.,Revathi.T.Temperature Prediction Using Fuzzy Time Series and Multivariate Markov Chain. International Journal of Fuzzy Mathematics and Systems,2(3),(2012),217-231.
13. Wang,N. Y., & Chen, S. M. Temperature prediction and Taifex forecasting based on automatinc clustering techniques and two factors high order fuzzy time series. Expert systems and applications , 36(2), (2009), 2143-2154.
14. Wai Ki Ching., Eric S., Fung and Michael K. Ng . A Multivariate Markov Chain Model for Categorical Data Sequences and its Applications in Demand Predictions. IMA Journal of Management Mathematics, 13, (2002),187-199.