A Multi-Factor HMM-based Forecasting Model for Fuzzy Time Series

Hui-Chi Chuang, Wen-Shin Chang, Sheng-Tun Li
Department of Industrial and Information Management
Institute of Information Management
National Cheng Kung University
No.1, University Road, Tainan City, Taiwan (R.O.C.)
e-mails:{r78021017, r36024037, stli}@mail.ncku.edu.tw

***Abstract*—In our daily life, people are often using forecasting techniques to predict weather, stock, economy and even some important Key Performance Indicator (KPI), and so forth. Therefore, forecasting methods have recently received increasing attention. In the last years, many researchers used fuzzy time series methods for forecasting because of their capability of dealing with vague data. The followers enhanced their study and proposed a stochastic hidden Markov model, which considers two factors. However, in forecasting problems, an event can be affected by many factors; if we consider more factors for prediction, we usually can get better forecasting results. In this paper, we present a multi-factor HMM-based forecasting model, which is enhanced by a stochastic hidden Markov model, and utilizes more factors to predict the future trend of data and get better forecasting accuracy rate.**

Keywords-fuzzy time series; forecasting; hidden Markov model (HMM).

I. INTRODUCTION

In the information explosion era, forecasting is a useful methodology for enterprises or governments to predict future trends. The more precise the forecasting result, the more appropriate the behavior conducted by managers. In general, many data are present with crisp value, but others are vague and ambiguous instead, such as stock monitoring indicators, signals, and so on. With the purpose of forecasting with vague data, Song and Chissom [9] first proposed the fuzzy time series and suggested seven steps for forecasting. Therefore, numerous fuzzy time series forecasting models have been proposed after them.

The seven steps proposed by Song and Chissom [9] are: (1) define the universe of discourse; (2) partition the universe into several intervals; (3) define fuzzy sets on the universe; (4) fuzzify the historical data; (5) establish fuzzy relations; (6) calculate the forecasted outputs; (7) defuzzification. The literature used Alabama University enrollment data to carry out one factor time relational, invariant fuzzy time series forecast model. Sullivan and Woodall [4] improved the results by using Markov's matrix- based probability statistics method [9] to establish one-factor one-order time invariant forecast model.

The above introduced methods implemented one factor to forecast, but in realistic circumstances, there is more than one factor that affects the forecasted data, such as the temperature and cloud density. Therefore, based on this idea, Lee, Wang, Chen, and Leu [6] proposed a two-factor high-order forecast model, which many people debate problematic in deciding the high-order. Later, Li and Cheng [13] proposed the deterministic fuzzy time series model that uses the concept of state transition diagram to backtrack to the most stable state. Wang and Chen [8] proposed two-factor fuzzy time series model and used automatic clustering techniques to partition intervals. Joshi and Kumar [1] first extended the idea of fuzzy sets into Intuitionistic Fuzzy Sets (IFS) and used it to construction a method for determining the membership degree.

Although fuzzy time series has been developed for decades, there are still some problems. The matrix computing for two-factor fuzzy time series models are too complex. Furthermore, the process of fuzzy time series forecasting is only concerned with whether or not the fuzzy rule is adopted, while ignoring the importance of frequency. Therefore, Sullivan and Woodall [4] used the Markov model to improve the traditional fuzzy time series model. But traditional HMM is unable to solve the two-factor problem. If we can analyze more factors, we can get higher accuracy and fully utilize all the available information. In our study, we enhance the study of Li and Cheng [12] and expand the model for forecasting based on multiple factors.

The fuzzy time series methods are good methodologies, which are suited for data composed of linguistic values, but the fuzzy relationship ignores the importance of the frequency. Hidden Markov models are powerful probability models which use categorical data include linguistic labels, and allow handling of two-factor forecasting problem. But in real world situations, multiple factor data tends to appear, rendering this model ineffective in making accurate assessments.

In this study, we want to combine the benefit of both methodologies, so that we can deal with the problem with many factors and fully utilize the feature of fuzzy theory to get higher accuracy result. There are five sections in this study, which is organized as follows. In Section I, we introduce the background and motivation. In Section Ⅱ, we introduce the basic definition of fuzzy time series and hidden Markov model. Section Ⅲ presents the model development. We propose a fuzzy time series model to deal with the multi-factor forecasting problem and we use HMM to build the fuzzy relationship matrix. In Section Ⅳ, we present the experiment result. We demonstrate the proposed model step by step and make a comparison with other existing methods. In Section Ⅴ, we present the conclusions by discussing the results as explained in Section Ⅳ and proposes future work expecting to improve the proposed method.

 ISBN: 978-1-61208-364-3

II. LITERATURE REVIEW

A. Fuzzy Time Series

Although there are many statistical methods that can be used to solve the problem of time series, it is impossible to resolve this problem when the historical data is linguistic value. It was not until 1993, when Song and Chissom [9] first proposed the fuzzy time series, that this problem has a solution. Because of its easiness to use and comprehend, there are many scholars who have dedicated their time and energy to the field of fuzzy time series making it more complete and accurate. The following is the basic definition of fuzzy time series.

1) Definition 1：Fuzzy Time Series

Let $\{x_t \in R\ ,t=1,2,...n\}$ be a fuzzy time series, U is the universe of discourse. Let $\{L_i(t)\ ,i=1,2,...,l\}$ be the ordered linguistic variables. Each X_t in the universe of discourse can be the denoted as follows:

$$F(t)=\frac{\mu_1(X_t)}{L_1}+\frac{\mu_2(X_t)}{L_2}+...+\frac{\mu_l(X_t)}{L_l} \tag{1}$$

where $\mu_k \in [0,1]$ and $\sum_k^l \mu_k(x_t)=1,\ \forall t=1,2,...,n \cdot \mu_k(X_t)$ is the membership value of the linguistic variable L_k. After transforming, we can say $F(t)$ is the fuzzy time series of X_t..

2) Definition 2：One-order Fuzzy Relation Equation

Let $F(t)$ be a fuzzy time series, the relationship between $F(t)$ and $F(t\text{-}1)$ is denoted as below:

$$F(t)=F(t-1)\circ R(t,t-1) \tag{2}$$

" ∘ " is a composition operator, and where $R(t,t\text{-}1)$ which is composed of R_{ij} is a one-order fuzzy relation. The relationship shows below:

$$R(t,t-1)=\cup_{ij} R_{ij}(t,t-1) \tag{3}$$

where $R_{ij}(t,t-1)$ is the fuzzy relation between $F_j(t)$ and $F_i(t\text{-}1)$.

3) Definition 3：One-order Fuzzy Logic Relationship

In order to reduce the complexity of matrix computation, Chen [11] proposed Fuzzy Logical Relationship (FLR) combined with simple arithmetic operation to replace matrix computation. If $F(t\text{-}1)$ is transformed into linguistic variable A_i, $F(t)$_is A_j, then the relationship between A_i and A_j can be shown below:

$$F(t-1)\rightarrow F(t) \text{ or } A_i \rightarrow A_j \tag{4}$$

The equation on the left side of the arrow is called the Left Hand Side (LHS); the right side is called Right Hand Side (RHS). After fuzzifying the historical data and establishing the fuzzy logical relationship, then we can group the entire data together to establish a logical relationship group. The following shows the fuzzy logical relationship group:

$$\begin{aligned} A_{i1} &\rightarrow A_{j1}, A_{j2},..., A_{jn} \\ A_{i2} &\rightarrow A_{j1}, A_{j2},..., A_{jn} \\ &\vdots \\ A_{ik} &\rightarrow A_{j1}, A_{j2},..., A_{jn} \end{aligned} \tag{5}$$

B. Forecasting Model of Fuzzy Time Series

Song and Chissom [9] first proposed a complete fuzzy time series forecasting model and divided it into seven steps. Later, many scholars have continuously revised this framework in order to get better forecasting accuracy.

1) Define Universe of Discourse U & Partition the universe

Song and Chissom [9] defined the universe of discourse U as follows:

$$U=[D_{\min}-D_1, D_{\max}+D_2] \tag{6}$$

where D_{min} and D_{max} are the minimum and the maximum in the training date set and D_1 and D_2 are the two proper positive integers decided by the analyst. After defining U, we partition the universe into several intervals. One is equal length interval, and another is unequal length interval.

2) Define Fuzzy Set and Linguistic Values

After fuzzyfying historical data we then begin to decide linguistic values like rare, few, plenty and so on. Each interval A_i (i=1,…,n) will have its own membership to get linguistic value u_k (k=1,…,n), which represents the degree of belonging to each interval of each. They used their own experience to formulate the membership degree of each element in each fuzzy set; the example is shown in (7).

$$\begin{aligned} A_1 &= \{u_1/1,\ u_2/0.5,\ u_3/0,\ u_4/0,\ u_5/0,\ u_6/0,\ u_7/0\} \\ A_2 &= \{u_1/0.5,\ u_2/1,\ u_3/0.5,\ u_4/0,\ u_5/0,\ u_6/0,\ u_7/0\} \\ A_3 &= \{u_1/0,\ u_2/0.5,\ u_3/1,\ u_4/0.5,\ u_5/0,\ u_6/0,\ u_7/0\} \\ A_4 &= \{u_1/0,\ u_2/0,\ u_3/0.5,\ u_4/1,\ u_5/0.5,\ u_6/0,\ u_7/0\} \\ A_5 &= \{u_1/0,\ u_2/0,\ u_3/0,\ u_4/0.5,\ u_5/1,\ u_6/0.5,\ u_7/0\} \\ A_6 &= \{u_1/0,\ u_2/0,\ u_3/0,\ u_4/0,\ u_5/0.5,\ u_6/1,\ u_7/0.5\} \\ A_7 &= \{u_1/0,\ u_2/0,\ u_3/0,\ u_4/0,\ u_5/0,\ u_6/0.5,\ u_7/1\} \end{aligned} \tag{7}$$

3) Fuzzify Historical Data

If the historical data is not ambiguous, we need to fuzzify it first, for example hot, cold and linguistic values. The process of fuzzification usually uses triangular fuzzy sets. The triangular fuzzy set of equal length intervals are represented in Figure 1, the unequal length intervals are represented in Figure 2.

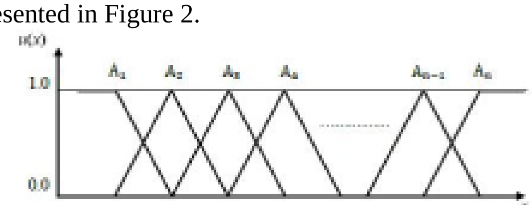

Figure 1. Linguistic values and triangular fuzzy set of equal length intervals.

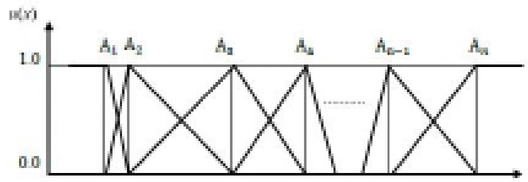

Figure 2. Linguistic values and triangular fuzzy set of unequal length intervals.

4) Establish Fuzzy Relation

Building up fuzzy relations is the most important part in the fuzzy time series forecasting model. Among many scholars, Song and Chissom proposed the time-variant and time-invariant fuzzy relation matrix.

Let us introduce Song and Chissom's method. First, we need to observe the fuzzy relation R_j by historical data. Second, find the relation matrix R which is composed of R_j. By using following equation, we can get the relation matrix R, where R_j comes from first step and N means the total number of fuzzy relations.

$$\mathrm{R} = \bigcup_{j=1}^{N} R_j \tag{8}$$

Finally, we can establish the forecasting model. There are many models for forecasting models and the following formula is one of composition to forecasting data, where "∘"is max-min operator.

$$A_i = A_{i-1} \circ R \tag{9}$$

Then, we can get the forecasting result.

5) Forecasting and Defuzzification

After establishing fuzzy relation matrix or fuzzy logical relationship, we begin to forecast linguistic value and then defuzzify it into precise one. However, there are some scholars who replace the process of forecasting linguistic value with a series of arithmetic operations, directly restoring the value into precise. For example, Wong, Bai, and Chu [14] proposed adaptive time-variant fuzzy time series; they used dynamic analysis window combined with series of heuristic rules to calculate the forecasting result.

The Proposed Defuzzification Method of Song and Chissom is summarized as follows: First, we standardized the calculated membership degree of linguistic value, and then we followed the three criteria which are listed below to defuzzify. (1) If there is only one maximum membership degree, then, we use the corresponding midpoint of the interval to be forecast value. (2) If the membership of an output has two or more consecutive maximums, then select the midpoint of the corresponding conjunct intervals as the forecasted value. (3) Otherwise, standardize the fuzzy output and use the midpoint of each interval to calculate the centroid of fuzzy set as the forecast value.

C. Markov Process

Markov property, named after the Russian mathematician Andrey Markov, refers to the memoryless property of a stochastic process. A stochastic process has the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present values) depends only upon the present state, not on the sequence of events that preceded it.

D. Hidden Markov Model (HMM)

Under realistic situation, there are usually multiple factors that influence the behavior and outcome of an event. For example, when we are trying to predict the temperature today, we may want to know the density of cloud, the amount of rainfall and also the weather yesterday. With the observation today and some previous information, we can predict today's weather. We might obtain the better forecast by combining the information of previous data and the present observation. HMM [15] is exactly the tool that can deal with two factors forecasting problem.

HMM is a statistical model to deal with symbols or signal sequences that are assumed to be a Markov process and have found a lot of applications in many areas like speech processing, weather, economy, population growth, stocks, etc.

The hidden Markov process is based on two important and essential assumptions: (1) The next state is dependent only upon the current state. (2) Each state-transition probability does not vary in time, i.e., it is a time-invariant model.

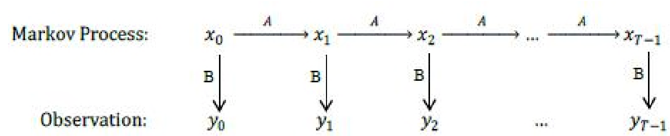

Figure 3. Hidden Markov Model.

The next three matrices of probability describe the common calculations that we would like to be able to perform on a HMM.

(1) Initial state vector $\pi = \{\pi_i\}$

$$\pi_i = \mathrm{P}(x_0 = s_i)$$

(2) State transition matrix A

$$A = \{a_{ij}\} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0,N-1} \\ a_{10} & a_{11} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{N-1,0} & \cdots & \cdots & a_{N-1,N-1} \end{bmatrix} \tag{10}$$

$$a_{ij} = P(state\ j\ at\ t+1 | state\ i\ at\ t)$$

(3) Confusion matrix B

$$B = \{b_{ij}\} = \begin{bmatrix} b_{00} & 01 & \cdots & b_{0,M-1} \\ b_{10} & b_{11} & & \vdots \\ \vdots & & \ddots & \vdots \\ b_{N-1,0} & \cdots & \cdots & a_{N-1,M-1} \end{bmatrix} \quad (11)$$

$$b_{ij} = P(observation\ j\ at\ t \mid state\ i\ at\ t)$$

Therefore, we use $\lambda = (A, B, \pi)$ to present the overall HMM model.

III. MODEL DEVELOPMENT

There are many studies of fuzzy time series, but the complexity of matrix computing still remains in multiple factors forecasting. Even though we can use HMM method to get the result faster and better, traditional HMM just can solve two-factors problem only. If we can analyze more factors, we can get higher accuracy and fully utilize all the information we got.

In this study, we want to preprocess data with fuzzification and use the transition of hidden Markov model to forecast outcome.

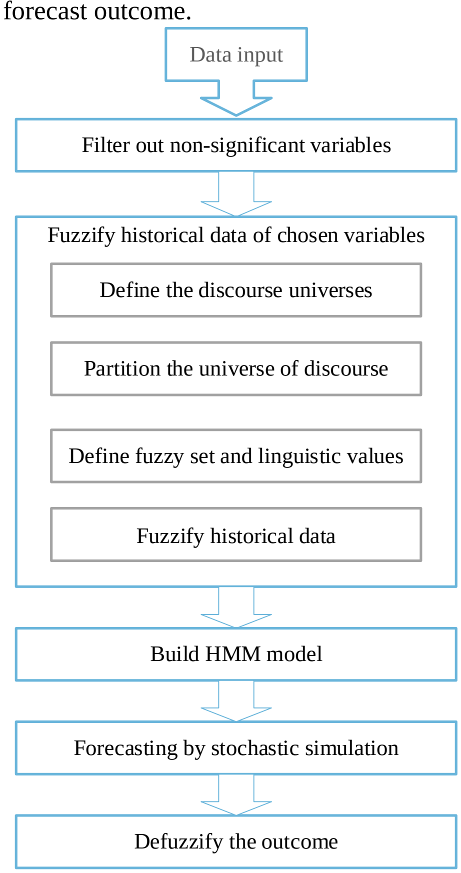

Figure 4. The framework of this study.

A. Fuzzify historical data of chosen variables

We follow the methodology of fuzzy time series forecasting to fuzzify the historical data. In this study, we conduct equal length interval method to partition universe of discourse into *n* intervals.

First, we need to define the discourse universes of historical data. Let $U^s, U^{o_1}, U^{o_2} \ldots U^{o_k}$ be the discourse universes of hidden state and observation variables respectively, and assume that the number of observation variables is *k*.

In general, the discourse universes are defined as follows:

$$\text{Hidden variable:}\quad U^s = [D^s_{min} - D^s_1, D^s_{max} + D^s_2]$$

$$\text{Observation variable 1:}\quad U^{o_1} = [D^{o_1}_{min} - D^{o_1}_1, D^{o_1}_{max} + D^{o_1}_2]$$

$$\vdots$$

$$\text{Observation variable K:}\quad U^{o_k} = [D^{o_k}_{min} - D^{o_k}_1, D^{o_k}_{max} + D^{o_k}_2]$$

where $D^s_{min}, D^s_{max}, D^{o_1}_{min}, D^{o_1}_{max}, D^{o_k}_{min}$ and $D^{o_k}_{max}$ are the respective minimal and maximal values of historical data of hidden and observation variables, and $D^s_1, D^s_2, D^{o_1}_1, D^{o_1}_2, D^{o_k}_1$ and $D^{o_k}_2$ are proper positive values decided by the analyst.

Second, the discourse universes are then partitioned into several equal lengthy intervals. Let us use $u_1, u_2, \ldots, u_n$ for each interval. Therefore, we have fuzzy set $A_1, A_2, \ldots, A_n$. When a value approaches the center of linguistic value A_i, it means the greater degree belongs to the linguistic value A_i, thus, the membership degree is closer to 1; on the contrary, if it is closer to the two bounds of A_i, the membership degree which belongs to A_i will be closer to 0.

In traditional study, each value has two linguistic values and corresponding membership degrees in most case, but the past literatures ignored the smaller one. However, even though our study is smaller, we think that it still has some valuable information. Therefore, we use triangle membership function to fuzzify historical data, and the fuzzy sets $A_1, A_2, \ldots, A_n$ are defined as seen in Figure 5:

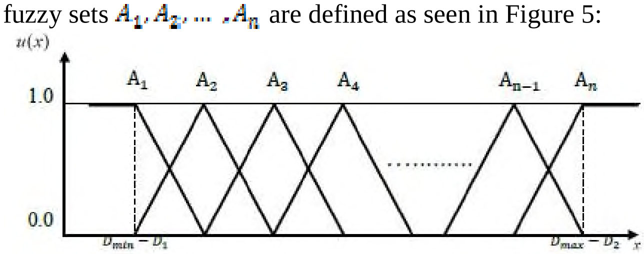

Figure 5. The fuzzy sets.

Finally, let us fuzzify the historical data and find the degree of each data belongs to each A_i $(i = 1, 2, \ldots, n)$. Then, we assume that if the maximum membership of that data is under A_k , then we treat this data as A_k .

B. Build HMM model

The objective of a multi-factor hidden Markov forecasting model is to estimate the probability of hidden state with given observations, so that we can predict the sequence of state that best explains the observed data. The following notations of HMM will be used in this paper:

N = number of state in the model

M_i = number of state for variable i, $i = 0, 1, \ldots, K-1$

K = number of variables
T = lengh of sequence
$S = \{s_0, s_1, \dots, s_{N-1}\}$ = distinct states of the Markov process
$O = \{o^0, o^1, \dots, o^{K-1}\}$ = distinct observation
$o^i = \{o_0^i, o_1^i, \dots, o_{M_i}^i\}$ = distinct observation for variable i, $i = 0,1,\dots,K-1$
$X = \{x_0, x_1, \dots, x_{T-1}\}$ = state sequences with lengh T
$Y = \{y_0, y_1, \dots, y_{T-1}\}$ = observation sequences with lengh T
$y_t = \{y_t^0, y_t^1, \dots, y_t^{K-1}\}$ = all observations for each variable in time t
$t = 0,1,\dots,T-1$

Markov process:	$x_0 \xrightarrow{A}$	$x_1 \xrightarrow{A}$	$x_2 \xrightarrow{A} \dots \xrightarrow{A}$		x_{T-1}
Observations 1 (B^0):	y_0^0	y_1^0	y_2^0	...	y_{T-1}^0
Observations 2 (B^1):	y_0^1	y_1^1	y_2^1	...	y_{T-1}^1
	⋮	⋮	⋮		⋮
Observations K (B^{K-1}):	y_0^{K-1}	y_1^{K-1}	y_2^{K-1}	...	y_{T-1}^{K-1}

Figure 6. A Multi-factor Hidden Markov Model.

In our study, we expand the research to multiple variables HMM and there are three important and essential assumptions of our proposed model: (1) The next state is dependent only upon the current state. (2) Each state-transition probability does not vary in time, i.e., it is a time-invariant model. (3) The observations are independent to each other.

The first and second assumptions come from the assumptions of HMM. The need for a third assumption is because our proposed method deals with multiple factors; in order to simplify the process, we ignore the impact of correlation between observations.

Therefore, the multiple observations HMM can be characterized by the following matrices:

$\pi = \{\pi_i\}$, $where\ \pi_i = P(x_0 = s_i)$
$A = \{a_{ij}\}$, $where\ a_{ij} = P(x_t = s_j | x_{t-1} = s_i)$
$B^0 = \{b_{i,j}^0\}$, $where\ b_{i,j}^1 = P(y_t^0 = o_j^0 | x_t = s_i)$
⋮
$B^{K-1} = \{b_{i,j}^{K-1}\}$, $where\ b_{i,j}^{K-1} = P(y_t^{K-1} = o_j^{K-1} | x_t = s_i)$

π is a vector with the probability of initial state. A is the state-transaction matrix which provides information about the relation of two contiguous hidden states. B^v is the confusion matrix which is the relation between observation *v* and hidden state.

Let us defined those parameters which are estimated by relative frequencies:

First, the initial state $\pi = \{\pi_i\}$ is a $1 \times n$ matrix and is defined as

$$\pi_i = \Pr(x_0 = s_i) = \frac{Count(x_0 = s_i)}{N_1}$$
$$N_1 = \sum_{i=0}^{N-1} Count(x_0 = s_i) \tag{12}$$

where $Count(x_0 = s_i)$ is the number of initial state s_i in the data set, and N_1 is the sum of initial state.

Second, the state transition matrix $A = \{a_{ij}\}$ is a $n \times n$ matrix and is defined as

$$a_{ij} = P(x_t = s_j | x_{t-1} = s_i) = \frac{Count(x_t = s_j, x_{t-1} = s_i)}{Count(x_{t-1} = s_i)}$$
$$\text{where, } \forall a_{ij} \geq 0 \text{ and } \sum_{j=0}^{N-1} a_{ij} = 1, i = 0,1,\dots,N-1 \tag{13}$$

Finally, the confusion matrix $B^v = \{b_{i,j}^v\}$ is a $n \times m$ matrix represented as follows:

$$b_{i,j}^v = P(y_t^v = o_j^v | x_t = s_i) = \frac{Count(y_t^v = o_j^v, x_t = s_i)}{Count(x_t = s_i)}$$
$$\text{where, } \forall b_{ij} \geq 0,\ \sum_{j=1}^{m} b_{ij} = 1, i = 0,1,\dots,N-1, v = 0,1,\dots,K-1 \tag{14}$$

We construct an HMM model $\lambda = (\pi, A, B)$ and use this model to forecast.

C. Forecasting and defuzzificationwihwui

There are many alogrithms that can compute the probability of the observations and we can also estmate the next state by getting maximal probability. Our study only focuses on forecasting, so the proposed method just uses dynamic programming to calculate maximum likelihood.

This study assumes that the variables of observation are independent of each other. Based on the concept of statistical independence, we calculate the probability directly to present the probability of observations.

$$b_{i,j} = \prod_{v=0}^{K-1} b_{i,j}^v \tag{15}$$

Based on the dynamic programming method, we construct the following equation:

$$\begin{aligned} P(Y | \lambda) &= \sum_X P(Y, X | \lambda) \\ &= \sum_X P(Y | X, \lambda) P(X | \lambda) \\ &= \sum_X \pi_{x_0} b_{x_0,y_0} a_{x_0,x_1} b_{x_1,y_1} a_{x_1,x_2} \cdots a_{x_{T-2},x_{T-1}} b_{x_{T-1},y_{T-1}} \\ &= \sum_X \pi_{x_0} b_{x_0,y_0} \prod_{i=0}^{T-2} a_{x_i,x_{i+1}} b_{x_{i+1},y_{i+1}} \end{aligned} \tag{16}$$

According to the notation of our study, $y_t = \{y_t^0, y_t^1, \dots, y_t^{K-1}\}$, we then edit the model as follows:

$$P(Y \mid \lambda) = \sum_{X} \pi_{x_0} b_{x_0, y_0} \prod_{i=0}^{T-2} a_{x_i x_{i+1}} b_{x_{i+1} y_{i+1}}$$

$$b_{i,j} = \prod_{v=0}^{K-1} b_{i,j}^{v} \tag{17}$$

$$P(Y \mid \lambda) = \sum_{X} \pi_{x_0} \prod_{v=0}^{K-1} b_{x_0 y_0}^{v} \left[\prod_{i=0}^{T-2} a_{x_i x_{i+1}} \left(\prod_{v=0}^{K-1} b_{x_{i+1} y_{i+1}}^{v} \right) \right]$$

According to (18), we obtain the probability of hidden state with given observations. Therefore, following the sequence of maximal probability, we can reach the forecasting sequence of hidden state. However, the sequence we estimated is fuzzy time series. Finally, we need to defuzzy the outcome.

There are several defuzzificaton methods that can be chosen. We use the most popular one, namely, centroid method, which standardize the fuzzy output and use the midpoint of each interval to calculate the centroid of fuzzy set as the forecast value. This method is expressed as:

$$\text{Centroid Method} = \frac{\sum_{i=1}^{N} \mu_A(x_i) C_i}{\sum_{i=1}^{N} \mu_A(x_i)} \tag{18}$$

N: the amount of fuzzy set;

$\mu_A(x_i)$: the x_i membership degree belonging to fuzzy se tμ_A;

C_i: the i^{th} midpoint of interval corresponding to the i^{th} linguistic value.

IV. Experiment Results

The present section demonstrates the application of the proposed method and compared the accuracy of its forecasted results with those results obtained by one factor only. In order to evaluate the superiority of proposed model, we use four evaluation indices to evaluate the performance, such as MAE (Mean Absolute Error), PMAD (Percent Mean Absolute Deviation), MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Squared Error).

In this experiment, we conduct the proposed model to forecast the average temperature with other three observations. With the forecasting results, the following Figure 7 shows the respective performances of the Alishan weather forecasting by implementing the proposed model.

This time series data are presented by month and contains four main factors: (1) average temperature (2) average humidity level (3) number of rainy days (4) total sunshine duration. In this experiment, we conduct the proposed model to forecast average temperature with three other factors. In order to confirm the performance of the proposed model, we also exhibit the forecasting result which is estimated by other methods we have mentioned previously. The forecasting result is displayed in Figure 7.

	MAE	**RMSE**	**PMAD**	**MAPE**
Proposed Model	1.0861	1.5100	0.0933	0.1097
Chen(1996)	1.3159	1.6584	0.1145	0.1316
Hsu et al. (2003)	1.2397	1.5269	0.1079	0.1251
Li & Cheng (2007)	2.0443	2.7463	0.1779	0.2078
Chen (2011)	1.0738	1.4074	0.0923	0.1065

Figure 7. The Forcasting result For Weather Data.

All the evaluation indices that are smaller will be favorable. We can discover that the proposed model has better prediction accuracy than most methods except Chen's newest model. Even though Chen's model has better prediction accuracy than the proposed one, the forecasting results are quite similar between two models. In Chen's model, high order information, which requires more computation, was considered. The proposed model uses four factors one order to forecasting, and has the similar performance with the model with one factor high orders. As a result, we can realize the power of multiple factors.

V. Conclusion and Futrue Work

The drawback of traditional forecasting models is that they cannot forecast with multiple factor data and waste the obtained information. However, this proposed model solves the problem and demonstrates the indication that "predicting with more factors can improve the forecasting result".

There is a point can be focused in the future work. In this model, we assume the relations between the observed factors are independent. However some realistic data cannot satisfy with this limitation. Therefore, we need to consider the impact of coefficient between observed factors. But it may make the model to be more complicated in calculation. In the future, we may make our effort on adjusting the model with consideration of coefficient more efficiently.

References

[1] B. P. Joshi and S. Kumar, "Intuitionistic Fuzzy Sets Based Method For Fuzzy Time Series Forecasting." Cybernetics and Systems, 43(1) (2012), pp. 34–47.

[2] C. H. Cheng, Y. S. Chen, and Y. L. Wu, "Forecasting innovation diffusion of products using trend-weighted fuzzy time-series model." Expert Systems with Applications, 36(2) (2009), pp. 1826–1832.

[3] H. K. Yu, "Weighted fuzzy time series models for TAIEX forecasting." Physica A: Statistical Mechanics and its Applications, 349(3) (2005), pp. 609–624.

[4] J. Sullivan and W. Woodall, "A comparison of fuzzy forecasting and Markov modeling." Fuzzy Sets and Systems, 64(1994), pp. 279–293.

[5] K. Huarng and T. H. K. Yu, "The application of neural networks to forecast fuzzy time series." Physica A: Statistical Mechanics and its Applications, 363(2) (2006), pp. 481–491.

[6] L. W. Lee, L. H. Wang, S. M. Chen, and Y. H. Leu, "Handling forecasting problems based on two-factors high-order fuzzy time series." IEEE Transactions on Fuzzy Systems, 14(3) (2006), pp. 468–477.

 ISBN: 978-1-61208-364-3

[7] M. Shah, “Fuzzy based trend mapping and forecasting for time series data.” Expert Systems with Applications, 39(7) (2012), pp. 6351–6358.

[8] N. Y. Wang and S. M. Chen, “Temperature prediction and TAIFEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series.” Expert Systems with Applications, 36(2) (2009), pp. 2143–2154.

[9] Q. Song and B. Chissom, “Fuzzy time series and its models.” Fuzzy sets and Systems, 54(1993), pp. 269–277.

[10] S. M. Chen, “Forecasting Enrollments Based On High-Order Fuzzy Time Series.” Cybernetics and Systems, 33(1) (2002), pp. 1–16.

[11] S. M. Chen and C. D. Chen, “Handling forecasting problems based on high-order fuzzy logical relationships.” Expert Systems with Applications, 38(4) (2011), pp. 3857–3864.

[12] S. T. Li and Y. C. Cheng, “A stochastic HMM-based forecasting model for fuzzy time series. IEEE transactions on systems, man, and cybernetics.” Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society, 40(5) (2010), pp. 1255–1266.

[13] S. T. Li and Y. C. Cheng, “Deterministic fuzzy time series model for forecasting enrollments.” Computers & Mathematics with Applications, 53(12) (2007), pp. 1904–1920.

[14] W. K. Wong, E. Bai, and A. W. C. Chu, “Adaptive time-variant models for fuzzy-time-series forecasting.” IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society, 40(6) (2010), pp. 1531–42.

[15] Y. Wang, X. Hao, X. Zhu, and F. Ye, “An approach of software fault detection based on HMM.” In 2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, 2012, pp. 644–647.