An Enhanced HMM-Based for Fuzzy Time Series Forecasting Model

Yi-Chung Cheng[1] Pei-Chih Chen[2] Chih-Chuan Chen[3] Hui-Chi Chuang[4] Sheng-Tun Li[4]

[1]Department of International Business Management, Tainan University of Technology, Taiwan, R.O.C.
[2]Department of Product Design, Tainan University of Technology, Taiwan, R.O.C.
[3]Department of Leisure and Information Management, Taiwan Shoufu University, Taiwan, R.O.C.
[4] Institute of Information Management, National Cheng Kung University, Taiwan, R.O.C.

Abstract

The fast and accurate forecasting method can help makers to make appropriate strategy. Zadeh was given the definition of a fuzzy set in 1965. Song and Chissom proposed the definition and the forecasting framework of fuzzy time series in 1993. Sullivan and Woodall first proposed the forecasting method to handle one factor with probability Markov model in 1994. Li and Cheng proposed a stochastic hidden Markov model which considers two factors in 2010. However, an event can be affected by many factors. In this paper, we present a multi-factor HMM-based forecasting, and utilize more factors to get better forecasting accuracy rate.

Keywords: fuzzy time series, forecasting, hidden Markov model (HMM)

1. Introduction

In the big data era, an efficiency forecasting model is very importance for the maker of enterprises or government. The forecasting problem of time series is an interesting and important research topic. However, traditional time series is complete developing statistic method, but can't deal with vague, vocabulary, or uncertainly data. Zadeh (1965) proposed the fuzzy theory, which closed to human think and can description vague and vocabulary variables.

For time series, the observed uncertain value can be modeled as a fuzzy variable, which is so-called fuzzy time series (FTS) (Möller and Reuter, 2007). The term FTS has been used in different meanings (Möller and Reuter, 2008): (1) time series with uncertain single data (fuzzy data) at each point in time (Hareter, 2004a, 2004b); (2) time series with uzzified real-valued single data at each point in time (Song and Chissom, 1993b); and (3) fuzzy time series based on a set of elementary finite time series and composed of several significant representative courses (Bocklisch and Pässler, 2000).

Based on the fuzzy theory, Song and Chissom (1993) first proposed the fuzzy time series to deal with vague and uncertainly data in time series, and proposed five steps for forecasting. The five steps proposed by Song and Chissom (1993) are: (1) Define and partition the universe of discourse; (2) define fuzzy sets and fuzzify the historical data; (3) construct fuzzy relations; (4)

forecasting and (5) defuzzification. The most fuzzy time series forecasting model based on fuzzy logical relationship (Chen, 1996; Hwang, Chen, and Lee, 1998; Chen, 2002; Yu, 2005; Lee et al., 2006; Li and Cheng, 2007; Lee, Wang, and Chen, 2007; Cheng, Chen, and Wu, 2009; Wang and Chen, 2009; Wong, Bai, and Chu, 2010; Chen and Chen, 2011; Shan, 2012), which is easy to understand but is not applicable for big data.

Sullivan and Woodall (1994) used Markov's matrix based probability statistics method to establish one-factor one-order fuzzy time series forecast model. Li and Cheng (2009) proposed a stochastic hidden Markov model (HMM), which takes into consideration the frequency of relationships thus only solving the two-factor problem. The hidden Markov models have been extensively used in the area like speech recognition, stock (Hassan and Nath, 2005), electrical signal prediction and image processing, etc. But traditional HMM is unable to solve the two-factor problem. In this paper we expand Li and Cheng (2009) probabilistic HMM model to forecasting multiple factors problem. And use one actual datum to verification the differences between one observable variable and muti-observable variables. Final, compare the forecasting accuracy to other model with MAE (Mean Absolute Error), PMAD (Percent Mean Absolute Deviation), MAPE (Mean Absolute Percentage Error), RMSE (Root Mean squared error).

2. Preliminaries and Related Work

2.1. Fuzzy Set and Fuzzy Time Series

Since technological progress and complex, precise value can't description the nature vague, vocabulary and uncertainly information. Zadeh(1965) given a fuzzy set definition to description the vague, vocabulary and uncertainly information, as follows:

Definition 1. A fuzzy set A of universe of discourse U is characterized by a membership function $\mu: U \to (0,1)$ which associates with each element u of U a number $\mu(u)$ in the interval $(0,1)$ which represents the grade of membership of u in A. The fuzzy set A of $U = \bigcup_i u_i$ will be denoted

$$A = \sum_{i=1}^{n} \mu_A(u_i)/u_i = \sum_i \mu_A(u_i) \qquad (1)$$

where Σ stands for union.

And then, Song and Chissom (1993b) proposed the definition of fuzzy time series. The fuzzy time series is a in time sequence, as below:

Definition 2. Let (t) $(t = \cdots, 0, 1, 2, \ldots)$, a subset of R, be the universe of discourse on which fuzzy sets $f(t)$ are defined, and let $F(t)$ be a collection of $f(t)$. Then, $F(t)$ is called a fuzzy time series on $Y(t)$ $(t = \cdots, 0, 1, 2, \ldots)$.

The fuzzy time series is always assumed there is relationship between time t and t-1. The one-order fuzzy relation equation is defined as:

Definition 3. One-order fuzzy relation equation. Let $F(t)$ be a fuzzy time series, the relationship between $F(t)$ and $F(t-1)$ is denoted as below:

$$F(t) = F(t-1) \circ R(t, t-1) \qquad (2)$$

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

$$R(t, t-1) = \bigcup_{ij} R_{ij}(t, t-1) \qquad (3)$$

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

Song and Chissom (1993b) first proposed a complete fuzzy time series forecasting model and divided it into five steps. Nevertheless, there are still many details worthy of further exploration on this architecture. Therefore, many studies follow the framework of Song and Chissom, in order to get better forecasting accuracy. The five steps are: (1) defining universe of discourse and partitioning into several intervals, (2) defining fuzzy sets and fuzzifying historical data, (3) constructing fuzzy relation, (4) forecasting and (5) defuzzification. On step (3), most paper used fuzzy logical relationship to construct fuzzy relation, and attempted to improve the forecasting accuracy. The fuzzy logical relationship is easy to understand but the compute complexity is stubbornly high. On the other hand, the probability forecasting model for fuzzy time series, only Sullivan and Woodall (1994) used Markov matrix and Li and Cheng (2009) used HMM to solve forecasting problem, for one and two factors, represent.

2.2. Hidden Markov Model

The hidden Markov model (HMM) is a statistical model used to deal with symbols or signal sequences. There are always three fundamental questions should be solved in HMM:

(1)Evaluation Problem: Given the model $\lambda = (\pi, A, B)$ and a sequence of observations Y, find $P(Y \mid \lambda)$. The problem is how to determine the likelihood of the observed sequence Y with given model.

(2)Decoding Problem: Given $\lambda = (\pi, A, B)$ and an observation sequence Y, and an optimal state sequence for the underlying Markov process. In other words, we want to uncover the hidden part of the Hidden Markov Model that best explains the observation.

(3)Learning Problem: Given an observation sequence Y and the dimensions N and M, and the model $\lambda = (\pi, A, B)$ that maximizes the probability of Y. This can be viewed as training a model to best fit the observed data. Alternatively, we can view this as a (discrete) hill climb on the parameter space represented by $\pi, A,$ and B. There are many established algorithms to solve the above question. The first problem is determined is the important issue in the paper.

It is very import for an HMM is composed of two states and three probability matrices. Two states are hidden state and observable state. The hidden states set is defined as $H = \{h_1, h_2, \ldots, h_n\}$ and the observable state set is $O = \{o_1^1, o_2^1, \ldots, o_{m_1}^1, o_1^2, o_2^2, \ldots, o_{m_2}^2, \ldots, o_1^k, o_2^k, \ldots, o_{m_k}^k\}$. The hidden states are probabilistically related to the observable states, n and M $(M = m_1 + m_2 + \cdots + m_k)$ are the number of hidden states and observable states, respectively. The three probability matrices are used to describe the relation between hidden and observable states and normally represented as $\lambda = (\pi, A, B)$, where π is initial state vector, A is hidden state transition matrix, and B is confusion matrix with hidden and observable state. The π, A, and B can be defined mathematically as follows:

$$\pi = [\pi_i], \quad \pi_i = P(x_0 = h_i)$$
$$A = [a_{ij}],$$

$$A = \{a_{ij}\} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n-1,1} & \cdots & \cdots & a_{n-1,n-1} \end{bmatrix} \qquad (4)$$

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

$$B = \{b_{ij}\} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1,M-1} \\ b_{21} & b_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ b_{M-1,1} & \cdots & \cdots & b_{M-1,M-1} \end{bmatrix} \qquad (5)$$

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

The probability $P(Y \mid \lambda)$ is the sum of $P(Y, X \mid \lambda)$ with all possible state sequences, so we obtain:

$$Pr(Y \mid \lambda)$$

$$= \sum_X Pr(Y, X \mid \lambda) \qquad (6)$$

$$= \sum_X \mathbf{Pr}(Y \mid X, \lambda) \mathbf{Pr}(X \mid \lambda) \qquad (7)$$

$$= \sum_X \pi_{x_0} b_{x_0,o_0} a_{x_0,x_1} b_{x_1,o_1} a_{x_1,x_2} \dots a_{x_{T-2},x_{T-1}} b_{x_{T-1},o_{T-1}} \qquad (8)$$

3. Model Development

The novel forecasting model is also following Song and Chissom (1993b) forecasting framework. Given fuzzy time series of hidden state F(t), and of k observable variables, G^1 (t), G^2 (t),…, and G^k (t), which there are n and m_i states for F(t) and G^i (t), i=1,2,…,k, respresently.

$$F(t) = \{f_i(t) | t = 1, 2, \dots, T, i = 1, 2, \dots n\}$$
$$G^1(t) = \{g_i^1(t) | t = 1, 2, \dots, T, i = 1, 2, \dots, m_1\}$$
$$G^2(t) = \{g_i^2(t) | t = 1, 2, \dots, T, i = 1, 2, \dots, m_2\} \qquad (9)$$
$$\vdots$$
$$G^k(t) = \{g_i^k(t) | t = 1, 2, \dots, T, i = 1, 2, \dots, m_k\}$$

Step (1). Defining and partition the universe of discourse. For hidden and observable states, the universal of discourse U is easy to make as follows:

$$U^h = \begin{bmatrix} D_{min}^h - D_1^h, & D_{max}^h + D_2^h \end{bmatrix}$$
$$U^{o^1} = \begin{bmatrix} D_{min}^{o^1} - D_1^{o^1}, & D_{max}^{o^1} + D_2^{o^1} \end{bmatrix}$$
$$U^{o^2} = \begin{bmatrix} D_{min}^{o^2} - D_1^{o^2}, & D_{max}^{o^2} + D_2^{o^2} \end{bmatrix} \qquad (10)$$
$$\vdots$$
$$U^{o^k} = \begin{bmatrix} D_{min}^{o^k} - D_1^{o^k}, & D_{max}^{o^k} + D_2^{o^k} \end{bmatrix}$$

Where D_{min}^h, $D_{min}^{o^1}$, $D_{min}^{o^2}$,…, $D_{min}^{o^k}$, D_{max}^h, $D_{max}^{o^1}$, $D_{max}^{o^2}$,…, and $D_{max}^{o^k}$ are the minimum and the maximum in the training date set, D_1^h, $D_1^{o^1}$, $D_1^{o^2}$,…, $D_1^{o^k}$, D_2^h, $D_2^{o^1}$, $D_2^{o^2}$,…, and $D_2^{o^k}$ are the two proper positive integers. And use the popular equal length method to define the interval length as follows:

$$l^h = \frac{1}{n} \left[(D_{max}^h + D_2^h) - (D_{min}^h - D_1^h) \right]$$
$$l^{o^1} = \frac{1}{m_1} \left[(D_{max}^{o^1} + D_2^{o^1}) - (D_{min}^{o^1} - D_1^{o^1}) \right]$$
$$l^{o^2} = \frac{1}{m_2} \left[(D_{max}^{o^2} + D_2^{o^2}) - (D_{min}^{o^2} - D_1^{o^2}) \right] \qquad (11)$$
$$\vdots$$
$$l^{o^k} = \frac{1}{m_k} \left[(D_{max}^{o^k} + D_2^{o^k}) - (D_{min}^{o^k} - D_1^{o^k}) \right]$$

Step (2). Defining the fuzzy sets and fuzzifying the time series. Given a traditional crisp time series, one needs a fuzzification procedure to obtain the corresponding fuzzy time series. For hidden states, n fuzzy sets $h_1, h_2, \dots, h_n$ can be defined on U^h using general membership functions, as expressed below:

$$h_i = \sum_{p=1}^n (\mu_{ip}/u_p) \qquad (12)$$

where μ_{ip} is the membership degree of h_i belonging to u_p.

For k observable states, m_i fuzzy sets $o_1^1, o_2^1, \dots, o_{m_1}^1$, $o_1^2, o_2^2, \dots, o_{m_2}^2, \dots$, and $o_1^k, o_2^k, \dots, o_{m_k}^k$ can be defined on U^{o^i}, $i = 1, 2, \dots, k$, as expressed below:

$$o_{m_i}^i = \sum_{q=1}^{m_i} v_{jq}^i / v_q^i \qquad (13)$$

where v_{jq}^i is the membership degree of $o_{m_i}^i$ belonging to v_q^i.

Then hidden historical datum H_t is fuzzified as h_i where the membership degree in interval u_i is maximal. The observation datum can be fuzzified in the same way as the hidden variables.

Step (3). Constructing an HMM model with $\lambda = (\pi, A, B)$.

First, the initial state vector π is set to be a $1 \times n$ matrix, defined as:

$$\pi_i = Pr((h_i, 1)) = \frac{\#((h_i, 1))}{N_1} \qquad (14)$$

where $\#((h_i, 1))$ is the number of the data whose initial states are h_i and $N_1 = \sum_{i=1}^n \#((h_i, 1))$.

Next, the state transition matrix $A = [a_{ij}]$ is a $n \times n$ matrix defined as below:

$$a_{ij} = Pr((h_j, t) | (h_i, t-1)) = \frac{\#((h_j, t), (h_i, t-1))}{\#((h_i, t-1))} \qquad (15)$$

with $\forall a_{ij} \geq 0$ and $\sum_{j=1}^n a_{ij} = 1$, where $\#((h_j, t), (h_i, t-1))$ denotes that the number of data whose hidden states is h_i at time $t-1$ and h_j at time t.

Finally, the confusion matrix $B = [b_{ij}]$ is a $n \times (m_1 \times m_2 \times \dots \times m_k)$ matrix represented as follows:

$$b_{ij} = Pr((o_j, t) | (h_i, t)) = \frac{\#((o_j, t), (h_i, t))}{\#((h_i, t))} \qquad (16)$$

322

with $\forall b_{ij} \geq 0$ and $\sum_{j=1}^{m_1 \times m_2 \times \ldots \times m_k} a_{ij} = 1$, $1 \leq i \leq n$, where $\#\left((\mathbf{o}_j, t), (h_i, t)\right)$ means that the number of data whose hidden states is h_i at time t and $\mathbf{o}_j$ at time t.

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

$\pi = \{\pi_i\}$, where $\pi_i = P(f(0) = h_i)$

$A = \{a_{ij}\}$, where $a_{ij} = P(f(t) = h_j | f(t-1) = h_i)$

$B = \{b_{ij}\}$, where $b_{ij} = P(g(t) = \mathbf{o}_j | f(t) = h_i)$

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 is the confusion matrix which is the relation between observation v and hidden state.

Step (4) Forecasting. there are a lot of alogrithm that can compute the probability of observation and we can also estmate the next state by getting maximal probability. Here the study just focus on forecasting, so the proposed method just use dynamic programming to calculate maximum likelihood.

Based on dynamic programming method, we construct the following equation:
According to the notation of our study, $y_t = \{y_t^0, y_t^1, \ldots, y_t^{K-1}\}$, we then edit the model as following:

$$
\begin{aligned}
\Pr\left((h_i, t) \big| (\mathbf{o}_j, t)\right) &= \frac{Pr((h_i, t)) \Pr\left((\mathbf{o}_j, t) \big| (h_i, t)\right)}{Pr\left((\mathbf{o}_j, t)\right)} \\
&= \frac{\sum_x Pr((h_i, t) | (h_x, t-1)) Pr((h_x, t)) \Pr\left((\mathbf{o}_{jy}, t) \big| (h_i, t)\right)}{Pr\left((\mathbf{o}_y, t)\right)}
\end{aligned}
\tag{17}
$$

Therefore, a particular HMM can be characterized by $Pr\left((h_i, t) \big| (h_x, t-1)\right)$ and $Pr\left((\mathbf{o}_y, t) \big| (h_i, t)\right)$.

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.

As mentioned before, the probability of (h_i, t) is determined by $Pr\left((h_i, t) \big| (h_x, t-1)\right)$ and $Pr\left((\mathbf{o}_y, t) \big| (h_i, t)\right)$. One has to compute the probabilities of all possible hidden states occurring at time t, $t \geq 2$, by considering the transition influence of the previous hidden state $(h_x, t-1)$ and the observation state $(\mathbf{o}_y, t)$. Such an influential relation can be represented by a function $\alpha_t\left((h_x, t-1), (\mathbf{o}_y, t)\right)$, defined as below:

$$
\alpha_t\left((h_x, t-1), (\mathbf{o}_y, t)\right) = B(:, [y])^T .* A([x], :)
\tag{18}
$$

where $A([x], :)$ is the xth row of state transaction matrix A, and $B(:, [y])$ is the yth column of confusion matrix B. The symbol of operator ' .* ' denotes an element-wise multiplication, a conventionally used nota-

tion in Matlab which multiplies two matrices by multiplying all of the corresponding elements.

However, the sequence we estimated is fuzzy time series. Finally, we need to defuzzy the outcome.
(5) Defuzzification.There are many defuzzification method, we use the most popular one, namely Fuzzy Mean Method, which present as

$$
\mathbf{FM(C)} = \frac{\sum_{i=1}^N \mu_i C_i}{\sum_{i=1}^N \mu_i}
\tag{19}
$$

N is the amount of fuzzy set, μ_i is the ith membership degree, C_i is i^{th} the midpoint of interval corresponding to the ith linguistic value.

4. Experiment and Results Analysis

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 and the calculation is display as MAE (Mean Absolute Error), PMAD (Percent Mean Absolute Deviation), MAPE (Mean Absolute Percentage Error), RMSE (Root Mean squared error), the indices to assess the forecasting abilities of the proposed model, and these are introduced below:

(1) MAE (Mean Absolute Error)

$$
\mathbf{MAE} = \frac{\sum_{i=1}^n |Forecastin\_value_i - Actual\_value_i|}{n}
\tag{20}
$$

(2) PMAD (Percentage Mean Absolute Deviation)

$$
\mathbf{PMAD} = \frac{\sum_{i=1}^n \left| \frac{Forecastin\_value_i - Actual\_value_i}{Actual\_value_i} \right|}{\sum_{i=1}^n Actual\_value_i}
\tag{21}
$$

(3) MAPE (Mean Absolute Percentage Error)

$$
\mathbf{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{Forecastin\_value_i - Actual\_value_i}{Actual\_value_i} \right|
\tag{22}
$$

(4) RMSE (Root Mean Squared Error)

$$
\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Forecastin\_value_i - Actual\_value_i)^2}{n}}
\tag{23}
$$

Here, The monthly weather data of Alishan, from 2004 to 2013, and containing one hidden factor, average temperature, and three observable factors: (1) average relative humidity level, (2) number of rainy days, and (3) total sunshine duration. The first seven years,

from 2004 to 2010, are used for training, from 2011 to 2013, are used for testing. The forecasting result is displayed as following Table 1 and Table 2:

	MAE	PMAD	MAPE	RMSE
All	0.877112	0.092451	0.092451	1.350553
Average humidity level	1.23298	0.124322	0.124322	1.699685
Number of rainy days	0.992742	0.106911	0.106911	1.405663
Total sunshine duration	1.170567	0.117814	0.117814	1.577392

Table 1 The evaluation of Alishan weather with different factors

	MAE	RMSE	PMAD
Proposed Model	1.0496	1.3937	0.0927
Chen (1996)	1.3159	1.6584	0.1145
Hsu et al. (2003)	1.2397	1.5269	0.1079
Li and Cheng (2007)	2.0443	2.7463	0.1779
Chen and Chen (2011)	1.0738	1.4074	0.0923

Table 2 The evaluation of Alishan weather

Obviously, in Table 1, the forecasting result proves that the more factors we considered the more precise result we can forecast. In Table 2, the performs of the proposed model is compare to other forecasting model (Chen(1996), Hsu et al. (2003), Li and Cheng (2007), Chen and Chen (2011)). The above illustrate, we proof two point, one is forecasting with multi-observables is better single observable, another is forecasting with multi-observables HMM is better than traditional fuzzy time series forecasting model (with fuzzy logical relationship to construct fuzzy relation model).

5. Conclusions

The constructing fuzzy relationship in fuzzy time series forecasting model, there are two category, fuzzy logical rule and probability. The fuzzy logical rule is easy to understand, but there is one problem, high compute complexity. The probability forecasting model, the previous study on construct relationship by Markov or HMM, the shortcoming is only deal with one factor or one hidden variable and one observable variable. On the other hand, the drawback of previous probability fuzzy time series forecasting model is that can't forecast with multiple factor data and waste the obtained information. However, the proposed model solves the problem and demonstrates the indication that "predicting with more factors can improve the forecasting result".

There are three points can be focused in the future work. One is, in this model, we assume the relations between the observed factors are independent. However some realistic data cannot satisfy with this limitation.

Therefore, we can to consider the impact of coefficient between observed factors to exclude the factor with collinearity. Another is, the high order HMM fuzzy time series forecasting model is necessary at some time. Third, the compute complexity also is an important issue. In future, the high order muti-factor HMM is must to study, the collinearity is must to exclude, and simple the compute complexity.

References

[1] S. F. Bocklisch and M. Pässler, *Fuzzy time series analysis. in: Hampel*, R., Wagenknecht, M., Chaker, N. (Eds), Advances in Soft Computing - Fuzzy Control, Physica-Verlag, Heidelberg, pp. 331-345, 2000.

[2] S. Chen, Forecasting enrollments based on fuzzy time series, *Fuzzy sets and systems,* 81(3), 311–319, 1996.

[3] S.-M. Chen, Forecasting enrollments based on high-order fuzzy time series, *Cybernetics and Systems,* 33(1), 1–16, 2002.

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

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

[6] D. Hareter, Time series analysis with non-precise data—Part I. presented at the *9th Special Conference Probability Mechanic Structure Reliability*, Sandia Nat. Labs., Albuquerque, NM, 2004.

[7] D. Hareter, Time series analysis with non-precise data—Part II. presented at the *9th Spec. Conference Probability Mechanic Structure Reliability,* Sandia Nat. Labs., Albuquerque, NM, 2004.

[8] M. Hassan and B. Nath, Stock market forecasting using hidden Markov model: a new approach. *Design and Applications,* 2005.

[9] 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), 481–491, 2006.

[10] J.-R. Hwang, S.-M. Chen, and C.-H. Lee, Handling forecasting problems using fuzzy time series, *Fuzzy Sets and Systems,* 100(1), 217–228, 1998.

[11] L.-W. Lee, L.-H. Wang and S.-M. Chen, Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms, *Expert Systems with Applications,* 33(3), 539–550, 2007.

[12] 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), 468–477, 2006.

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

[14] 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), 1255–1266, 2010.

[15] B. Möller and U. Reuter, *Uncertainly Forecasting in Engineering,* Berlin, Germany: Springer-Verlag, 2007.

[16] B. Möller and U. Reuter, Prediction of uncertain structural responses using fuzzy time series, *Compute Structures,* 86, 1123–1139, 2008.

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

[18] Q. Song and B.S. Chissom, Forecasting enrollments with fuzzy time series - Part I, *Fuzzy Sets and Systems,* 54(1), 1-9, 1993a.

[19] Q. Song and B.S. Chissom, Fuzzy time series and its models, *Fuzzy sets and Systems,* 54(3), 269–277, 1993b.

[20] Q. Song and B.S. Chissom, Forecasting enrollments with fuzzy time series - part II, *Fuzzy Sets and Systems,* 62(1), 1-8, 1994.

[21] M. Stamp, A revealing introduction to hidden Markov models, Department of Computer Science San Jose State, 1–20, 2004.

[22] J. Sullivan and W. Woodall, A comparison of fuzzy forecasting and Markov modeling, *Fuzzy Sets and Systems,* 64, 279–293, 1994.

[23] 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), 2143–2154, 2009.

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

[25] 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,* 40(6), 1531–42, 2010.

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