ISSN : 0974 - 7435 Volume 10 Issue 24

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(24), 2014 [15435-15442]

Type-2 fuzzy systems theory and its application in the stock trend analysis

Zhao Juan, Jiang Lihui
Department of Mathematics and Physics, Hefei University, Hefei Anhui, 230601, (CHINA)

ABSTRACT

Fuzzy time series model is to solve the classical time series analysis methods can not deal with the fuzzy problem and the birth of. Along with the people to solve complex problems to be increasing, the research is more and more thorough, its application is more and more widely. From the theoretical model, establishing fuzzy relationship matrix, extraction and prediction rules and improve forecasting accuracy has been the focus of the study of fuzzy time series. A fuzzy time series model, C- AFS theory of fuzzy decision tree and based on evidence theory. Based on fuzzy decision tree and evidence theory in the introduction of AFS theory, C-, fuzzy time series model to determine the low prediction rules in the fuzzy relation, fuzzy time series combined with these three methods of forecasting model. The experimental results show that the generalized fuzzy time series model can best sum up the fuzzy sequence contains information which get the forecasting result comparison of similar model better.

KEYWORDS

Time series; Fuzzy relations; Type-2 fuzzy systems; AFS theory; Clustering algorithm.

INTRODUCTION

Time series data sequence is one or more statistical indicators of a chronological phenomenon formed[1]. Although some of the data is not necessarily chronological appearance, if we can lined up a series of information according to their sources, such as the importance of the link or other location has, they can also be seen as a time-series data. Time series analysis methods for this kind of data is, dig out the sequence of the inherent laws provide theoretical support and understanding for people to understand the phenomenon reflected in the sequence, or cow - react sequence system provides basis for decision making[2, 3]. According to their origin, importance or other inner link into a sequence position has information, which can also be seen as the time sequence data[4]. The inner regularity in the sequence provides the theory support for the readers to understand the sequence reflects the phenomenon, or for cattle - sequence of the system to respond to provide basis for decision making.

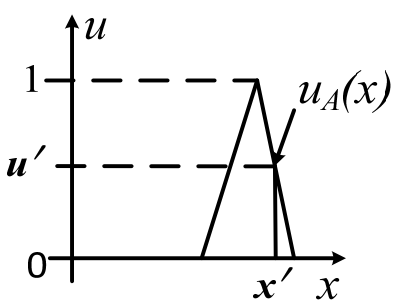
Type-2 fuzzy systems can effectively make up the disadvantages of fuzzy sets to describe a type of fuzzy uncertainty[5-8]. Because the membership function of type-2 fuzzy sets is three-dimensional, provide more freedom, can control multiple fuzzy uncertainty information. However, operation type-2 fuzzy sets with high complexity, making it difficult to be used for real-time operation of the engineering environment[9]. This greatly hindered the development of type-2 fuzzy sets theory. Therefore, in the present after twenty years has not been much attention. According to the type-2 fuzzy set operation complexity, Mendel et al. Further proposed the concept of interval type-2 fuzzy sets, namely the type-2 fuzzy set of two order membership function is defined as the degree of membership for the "1" interval valued fuzzy set[10]. This concept was proposed to simplify operation type-2 fuzzy sets, enhanced the ability of the two type fuzzy system for real-time applications in engineering, to promote application of type-2 fuzzy set theory. Since then, most of the applications of type-2 fuzzy sets theory based on the use of type -2 fuzzy sets and interval[11, 12].

Fuzzy time series model is based on fuzzy theory, using technology and some methods of artificial intelligence and data mining on time series data analysis, find out the information contained in the sequence to provide the basis for the understanding of things and systems. Then dig out information in the sequence, for people's cognition and decision service. This paper is based on the fuzzy time series model is narrow, to three kinds of typical models are summed up in theory, several improving methods are presented, including through adjusting the weights increase model in the prediction process using the observation sample information rate, using the theory of AFS, C- fuzzy decision tree and evidence theory to dig out the fuzzy relation type contains the sample data provided by the prediction rule. In addition, in order to apply the fuzzy time series model is more widely, perfect its theories and methods, this paper also proposed a generalized fuzzy time series model and fuzzy time series forecasting model. The actual effect of model with most enrollment data and the Shanghai Stock Exchange composite trading index is verified.

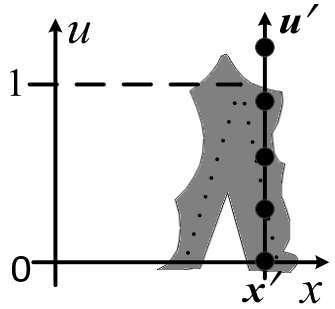
RELATED KNOWLEDGE OF TYPE-2 FUZZY THEORY

Proposed of type-2 fuzzy theory

L.A. Zadeh[13] professor of classics of traditional set theory was extended, the fuzzy set theory is proposed, the classical elements of the set of fuzzy membership values given by, such a set is called a fuzzy set. Professor Zadeh[14] has expanded the concept of a fuzzy set, the type-2 fuzzy set theory is proposed, a fuzzy set expansion to open, the fuzzy membership degree of fuzzy sets. As shown in Figure 1, so as to strengthen the ability of fuzzy set description. In order to describe natural language as an example, a fuzzy set describes the fuzzy language variable itself, while type-2 fuzzy sets extend on fuzzy linguistic variables describing ability more in line with the actual situation. A fuzzy set is not directly described in this example twofold uncertainty. A fuzzy set of uncertain factors and can not directly, completely describe caused by irregular operating environment, system dynamic characteristics of the system or the human language fuzziness[15]. Therefore, a fuzzy set and does not reflect the fuzzy set theory in dealing with fuzzy uncertainty has the advantage of information.



(a) A set of fuzzy membership functions



(b) Fuzzy membership function

Figure 1 : Fuzzy set extended membership function

Architecture of type-2 fuzzy theory

Fuzzy time series model of fuzzy sets of type-2 are included in the frame based on:

1) According to the training set and the fuzzy membership function of fuzzy partition of the domain;
2) Based on the observation of training data has value to construct the fuzzy set;
3) Obtained by the fuzzy relation matrix between all;
4) The observations obtained by fuzzy rules, predictive value.

A fuzzy set of elements in the membership degree is accurate, and the type-2 fuzzy set membership is fuzzy element. Type-2 fuzzy sets in cannot accurately draw fuzzy membership function, still can not determine the information used to describe the variables. Conversely, if the two membership fuzzy set elements to fuzzy, then type-2 fuzzy sets will drop into a fuzzy set. That is to say, two types of fuzzy set is an extension of fuzzy set, including the extension of a type of basic concepts of fuzzy set and the extended set operators.

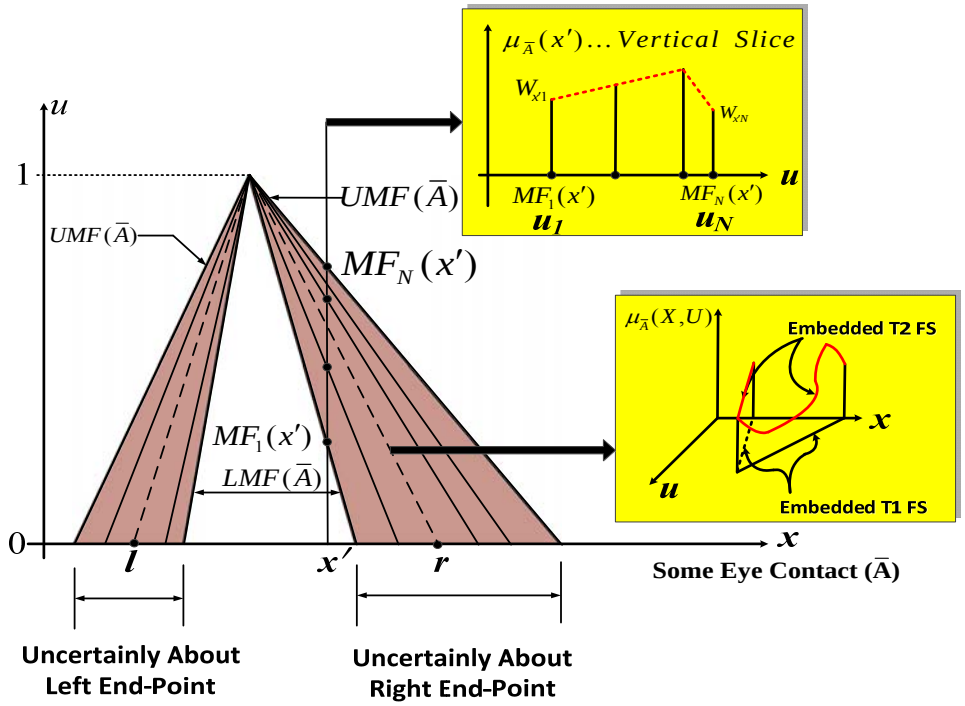


Figure 2 : The type-2 fuzzy triangular membership function

Figure 2 depicts a type-2 fuzzy triangular membership function. Figure insert at the top of section chart depicts the elements of X ′ first-order membership values and the corresponding two order membership values, waveform bottom insert depicts two order a membership function order membership corresponds to an embedded type of membership functions of the fuzzy sets and their elements. From th. From the chart we can see clearly that the two type of fuzzy membership function of three-dimensional structure. Like the other new theory, two types of fuzzy set theory are also defined some new terms are used to describe the uncertainty of the fuzzy model.

Inadequate of existing algorithms

Type-2 fuzzy sets from the known sequence of data to discover the rules, and predicts the future of data, this is the time series analysis method, research content and objectives. In order to improve the prediction accuracy, some machine learning algorithms, such as: the rough set theory, neural network theory and the novel method has also been applied to the fuzzy time series forecasting model research. Some people even put forward some multi factor model, model of high order and high order multi factor model.

However, these new models also exist some problems. Because of the application of these methods in mining data fuzzy relations in the rules of operation process is often a "black box" (such as neural network algorithm process), which weakens the model interpretability, not for people to understand the model results. Some consider only the calculation and prediction accuracy, and ignore the content of interest, such as the prediction result is reliable, which factors impact on the prediction results. Logical operator are selected in the practical application process, often with artificial or accidental factors,

which leads to the membership function and fuzzy logic system is often not objective and rigorous, but also does not have the unification, and made many important mathematical tools can not play a due role, the development of the theory of fuzzy sets and systems theory and practice have a great negative impact.

FUZZY TIME SERIES MODELS AND EVIDENCE-BASED DECISION TREE THEORY

AFS theoretical preparation

According to the classic fuzzy sets the lack of the theory, we investigate the fuzzy sets and fuzzy logic from the more general, more abstract level, proposed the AFS theory. The theory takes the concept of decomposition and synthesis mechanism of abstraction from a mathematical point of view, the fuzzy set theory of mathematics axiom, the mathematical theory and tools people use algebra, topology and close to the study of fuzzy sets and related problems become possible, also makes the building foundation membership functions and fuzzy logic system the objective, rigorous. At present, it has been widely used in machine learning, knowledge representation and discovery research work in many fields.

Know from the general situation of AFS theory, it includes three main mathematical objects: AFS algebra, AFS structure and the cognitive domain. AFS algebra as the basis of this theory, it is the representation of the concept, structure characterization, synthesis, decomposition of mathematics; the AFS structure is on the domain J (complex relationship on mathematical abstraction, it combines with the AFS algebra can be given the structure; fuzzy domain X membership function of the fuzzy logic and AFS the cognitive domain to fuzzy sets into fuzzy sets, it is because of this conversion, the AFS theory is applied to describe the fuzzy concept is intuitive, operation and the operation is very simple.

Let $\boldsymbol{X, M}$ is two sets, $\boldsymbol{2^M}$ is power set of $\boldsymbol{M}$, $\boldsymbol{\tau : X \times X \to 2^M}$. If $\boldsymbol{\tau}$ satisfying the following axioms, $\boldsymbol{(M, \tau, X)}$ is called an ASF structure.

$$\begin{aligned} &\boldsymbol{AX1:\ \forall(x_1,x_2),\ (x_2,x_3)\in X\times X,\ \tau(x_1,x_2)\cap(x_2,x_3)\subseteq\tau(x_1,x_3);}\\ &\boldsymbol{AX2:\ \forall(x_1,x_2)\in X\times X,\ \tau(x_1,x_2)\subseteq\tau(x_1,x_1).} \end{aligned} \tag{1}$$

X is known as the domain, M is called the attribute set, τ is called structure.

Let $\boldsymbol{M}$ be a non-empty set, $\boldsymbol{R}$ is a binary relation on the set of $\boldsymbol{EM^*}$,Defined as follows:

$$\boldsymbol{\sum_{i\in I}\left(\prod_{m\in A_i} m\right), \sum_{j\in J}\left(\prod_{m\in B_i} m\right) \in EM^*\left(\sum_{i\in I}\left(\prod_{m\in A_i} m\right)\right) R \left(\sum_{j\in J}\left(\prod_{m\in B_i} m\right)\right)} \tag{2}$$

Conditions equivalent:

$$\begin{aligned} &\boldsymbol{(1)\ \forall A_i(i\in I), \exists B_h(h\in J), making\ A_i \supseteq B_h;}\\ &\boldsymbol{(2)\ \forall B_j(j\in J), \exists A_k(k\in I), making\ B_j \supseteq A_k.} \end{aligned} \tag{3}$$

Let $\boldsymbol{X, M}$ is two sets, $\boldsymbol{(M, \tau, X)}$ is an ASF structure, $\boldsymbol{(\Omega, F, P)}$ is a probability measure space, $\boldsymbol{M}$ is a collection of simple concepts on $\boldsymbol{\Omega}$, $\boldsymbol{\rho_\gamma}$ is a simple concept $\boldsymbol{\gamma \in M}$ weighting function, $\boldsymbol{X \subseteq \Omega}$ is an observational data sets of probability space $\boldsymbol{(\Omega, F, P)}$,If for any $\boldsymbol{m \in M}$ and $\boldsymbol{x \in \Omega}$,$\boldsymbol{\{m\} \geq (x) \in F}$,you can define the membership functions of fuzzy set $\boldsymbol{\xi = \sum_{i\in I}\left(\prod_{m\in A_i} m\right) \in EM}$ is as follows:

$$\boldsymbol{u_\xi(x) = \sup_{i\in I} Inf_{\gamma\in A_i} \frac{\sum_{u\in A_i \geq (x)} \rho_\gamma(u) N_u}{\sum_{u\in X} \rho_\gamma(u) N_u}, \qquad (\forall x \in X)} \tag{4}$$

Among them, the number of sample U is observed. So the modular concept of arbitrary EM can use the above formula can get it in the cognitive domain membership function and membership point value.

C-fuzzy decision tree

C-fuzzy decision tree is composed of Pedrycz and Sosnowski in 2005, it is the fuzzy C- means clustering algorithm based on the decision tree is constructed based on. After this has been applied to many of its researchers are further studied.

Usually choose (ö as the evaluation parameter. For example, for the k data a tree node, the data can be represented by a φ_a, wherein, $\varphi_k = 1 - c^c \prod_{i=1}^{c} u_{ik}$. And the whole tree threshold can be calculated using the following equation:

$$\varphi = 1 - c^c \left(\frac{1+\alpha}{c}\right)^{\frac{c}{2}} \left(\frac{1-\alpha}{c}\right)^{\frac{c}{2}} \quad (5)$$

Among them, a is a belongs to the interval (0, 1). For example, given a N observation set, the M time series data attributes, domain J set up the first attribute $U_j = \{u_{j,1}, \ \ldots \ , u_{j,n_j}\}$. Then the training set data can be represented as shown in TABLE 1.

TABLE 1 : Results of fuzzified training data

Time	U_1	…	U_i	…	U_M	Fuzzy Sets
	$u^t_{1,1}, \ \ldots \ , u^t_{1,n_1}$	…	$u^t_{i,1}, \ \ldots \ , u^t_{i,n_1}$	…	$u^t_{M,1}, \ \ldots \ , u^t_{M,n_M}$	
S(1)	$u^1_{1,1}, \ \ldots \ , u^1_{1,n_1}$	…	$u^1_{i,1}, \ \ldots \ , u^1_{i,n_1}$	…	$u^1_{M,1}, \ \ldots \ , u^1_{M,n_M}$	$A_{\bar{m}}(1)$
S(2)	$u^2_{1,1}, \ \ldots \ , u^2_{1,n_1}$	…	$u^2_{i,1}, \ \ldots \ , u^2_{i,n_1}$	…	$u^2_{M,1}, \ \ldots \ , u^2_{M,n_M}$	$A_{\bar{m}}(2)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
S(N-1)	$u^{N-1}_{1,1}, \ \ldots \ , u^{N-1}_{1,n_1}$	…	$u^{N-1}_{i,1}, \ \ldots \ , u^{N-1}_{i,n_1}$	…	$u^{N-1}_{M,1}, \ \ldots \ , u^{N-1}_{M,n_M}$	$A_{\bar{m}}(N-1)$
S(N)	$u^N_{1,1}, \ \ldots \ , u^N_{1,n_1}$	…	$u^N_{i,1}, \ \ldots \ , u^N_{i,n_1}$	…	$u^N_{M,1}, \ \ldots \ , u^N_{M,n_M}$	$A_{\bar{m}}(N)$

Fuzzy time series model based on evidence theory

Evidence theory was proposed by Dempster in 1966, since then, his student Shaferd in 1976 to the theory is extended and perfect, therefore, then people will call it the Dempster-shafer theory (D-S theory). It belongs to the research content of artificial intelligence, is a kind of uncertainty reasoning method, first used in expert system, can solve the uncertain or fuzzy nature of the information problem.

As the uncertainty reasoning method, evidence theory can not only meet the Bayesian probability theory than weaker conditions, and can be directly expressed as "uncertainty" and "don't know" the uncertainty of language. It also allows the entire problem and evidence is decomposed into several sub problems, evidence, make corresponding processing in evidence sub problem, using synthetic principle, given the solution of the problem, so the evidence theory is a kind of decision theory, and has gradually developed into an important method in uncertainty reasoning. In addition, it can not only deal with the randomness and fuzziness leads to uncertainty, but can also rely on the accumulation of evidence, narrowing the hypothesis set, so as to obtain the solution of problems.

Integration of different levels of relationship between fuzzy logic

To establish generalized model based on fuzzy logic to the generalized fuzzy logical relationship are classified, get the relation matrix. The method of using fuzzy number matrix between the logical and relationship is statistically identical fuzzy relation in the same type, and take it as the representative of line relation matrix and the corresponding post column elements. Thus, there is a relation matrix for each class of fuzzy relations, altogether can get k relation matrix. According to the definition above, the fuzzy relation very much, and not the same type (or a different level), so you must have an operation in theory, these are not the same level, but also reflects the relationship has effect on decision-making information integrated. Therefore, the following first give a preliminary operation, preparing for the establishment of model.

Let $A^{(j)}_{t_j}$ is the j-th antecedent of main fuzzy logic relations (that fuzzy logic relations *GL (1, j)*), $R^{(j)}(t_j, i)$ is the number of training set $A_{t_j} \to A_i$ according to get fuzzy logic logical relationship relation matrix. Λ_k is defined to calculation is as the integrated relationship matrix of information at all levels:

$$\Lambda_k(A_{t_j}^{(1)},\cdots,A_{t_j}^{(j)},\cdots,A_{t_k}^{(k)}) = (\min_{1\le j\le k} R^{(j)}(t_j,1),\cdots,\min_{1\le j\le k} R^{(j)}(t_j,i),\cdots,\min_{1\le j\le k} R^{(j)}(t_j,n)) \quad (6)$$

Discrete type-2 fuzzy control system design

Accordi ng to the definition of fuzzy relations is in front, every two adjacent observations of three fuzzy sets have relative fuzzy relation. Fuzzy logical relationship close to might as well written as $A_c(t) \to Set_c = \{Ac_1, Ac_2, \square, Ac_c\}$, fuzzy logic relationship between the highest and lowest value index can be respectively denoted as $A_h(t) \to Set_h = \{A_{h_1}, A_{h_2}, \square, A_{h_h}\}$, two operation respectively and Vm and VN in addition, this model and the previous model difference lies in it not by establishing fuzzy logical relationship matrix, and use it as the basis for prediction, but directly use two operations defined above (i.e. and intersection) for different levels of fuzzy relations between the Union and intersection, union and intersection and statistics the number of elements, and then predict the. According to these rules, the predictive model of can be expressed by the following three:

$$Fval_{V_m}(t+1) = \begin{cases} m_c, \ when \ \cup_{S\in\{h,c,l\}} Set_s = \varnothing, \\ \dfrac{\sum_i m_i}{N^c_{V_m}}, when \ A_i \in \cup_{S\in\{h,c,l\}} Set_s, \end{cases}$$

$$Fval_{\Lambda_m}(t+1) = \begin{cases} m_c, \ when \ \cup_{S\in\{h,c,l\}} Set_s = \varnothing, \\ \dfrac{\sum_i m_i}{N^c_{\Lambda_m}}, when \ A_i \in \cup_{S\in\{h,c,l\}} Set_s, \end{cases} \quad (7)$$

RESULTS AND SIMULATION

Experimental results and analysis

This model will be used in the Shanghai stock market in January 1, 2003 to December 31 2012 in the ten years data as test data, in order to facilitate a comparison of fuzzy sets. Fuzzy time series model results, but the price of every stock index as the K attribute, the highest value of every index, the minimum value and closing value as dependency properties, fuzzy interval length is 100, the annual data as a test set, a total of ten test set. Because the model is based on the FCM, there is certain randomness, so that for each test set for the 100 experiment, and the average value as the prediction model.

Figure 3 and Figure 4 compares the variance of fuzzy time series model C- fuzzy decision tree and fuzzy decision tree and fuzzy time series model based on improved C- prediction error based on the consideration of their stability, prediction. Figure 3 is the results of the original model, Figure 4 is the result of the improved model. From the two picture we can see, although the overall variance improved model prediction error will become increasingly large, but the first 8 results, c=3, 4,…,9, is better than the original model results. In fact, in the clustering problem, class number together is generally not great, especially in the use of fuzzy concept for research work, class number, then the fuzzy degree, more not easy to understand, so this method in fuzzy forecasting is feasible. The analysis shows that, the proposed model when the poly class number less than 10, with fast calculation speed to obtain the forecast result more stable, in the following experiments, including process compared with other models, the value of C will be discussed in detail.

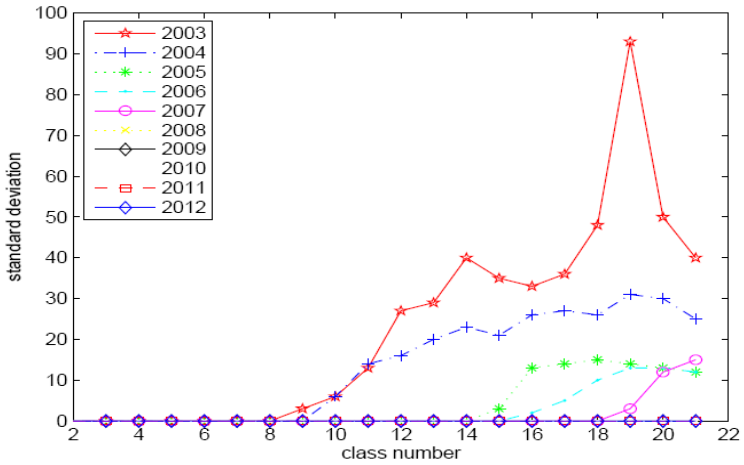


Figure 3 : The results of the original model

Experimental program set

In addition, the model also takes into account the method of information entropy is used to weight the distance, because people from the theory has proved its advantages, the following will it with the original model in 2003, 2004, 2005 and 2006 data as an example, a comparison experiment, the experimental results are listed in TABLE 2. From the table it can be seen that, the improved prediction model based on different values in the C, most of the cases were able to achieve better results.

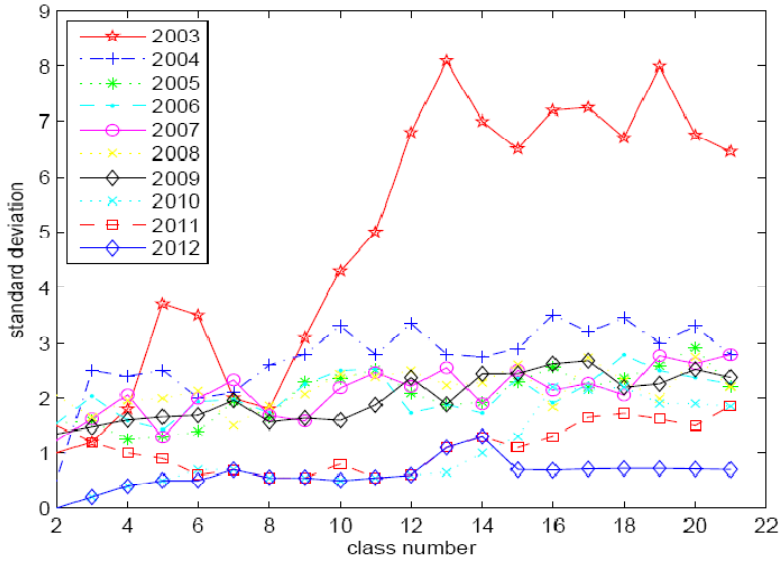


Figure 4 : The result of the improved model

TABLE 2 : Forecasts comparison of conventional models

Year	Model[7]	Model[12]	Model[14]	Model[87]
2003	50.12	38.07	36.57	32.03
2004	44.19	34.51	32.02	24.80
2005	38.38	37.47	35.81	25.23
2006	46.16	42.15	40.33	30.26
2007	42.32	37.23	36.76	27.49
2008	38.80	47.19	35.59	39.90
2009	32.71	36.57	30.98	29.90
2010	38.66	35.98	33.66	26.61

2011	31.02	36.51	33.01	31.71
2012	41.31	40.57	35.42	35.68
Average	40.37	38.63	35.01	30.36

Note: Black font indicates the best results

CONCLUSION

This paper briefly introduces the theory of evidence, it is pointed out that the traditional model cannot deal with fuzzy time series and multi factors of this problem, and then analyzed and improved evidence theory of evidence combination method, presents the evidence theory can solve the multi factor fuzzy time series prediction model in 200 years Shanghai stock index data from 2012 to ten for the model test of practice based on, with experimental results show that the model is feasible, prediction effect is better than that of the reference prediction model. Time series analysis method in the real life and the study has been very mature. It can not only reveal the phenomenon changes from quantity, the future behavior and features also can qualitatively forecast, and to control the risk that may occur in the future, or may encounter in the future to provide decision support. However, the mountain to the reality of the problem is very complicated, different areas have different types of questions, even the same at a different point of view also makes much difference to solve the problem requirements. So, the boundaries may exist in natural science and social science research of fuzzy, or bias problems of uncertainty knowledge.

ACKNOWLEDGEMENT

The work was supported by the Found of National Natural Science Fund Project (71371011), and Anhui Province University Natural Science Fund (KJ2013B234)).

REFERENCES

[1] H.Song, X.M.Gao; Micro Computer Information, **(26)**, 135-136 **(2010)**.
[2] H.F.Han, K.M.Du, Z.F.Sun; Agricultural Engineering, **25**, 158-162 **(2009)**.
[3] B.Zhang, Gaurav S.Sukhatme, Aristides A.G. Requicha; Computer Science Department University of Southern California Los Angeles, **219**, 73-77 **(2010)**.
[4] Q.Q.Gao, Q.D.Zhuang; Sensor Technology, **20**, 30-40 **(2001)**.
[5] R.T.Minamide; Journal of Horticultural Science, **68**, 755-762 **(1993)**.
[6] Q.Chen, X.L.Li, D.Horlacher; Communication in Soil Science and Plant Analysis, **35**, 1725-1740 **(2004)**.
[7] Zheng Zhonglong, Zhao Jianmin, Yang Jie; "NMF with Log Gabor Wavelets for visualization", Lecture Notesin Computer Scienee, **3691(2)**, 26-32, **(2005)**.
[8] Wang Jue, Wang Ren, Miao Duoqian; "Data enriching based on rough set theory", Journal of Environmental Sciences, **29(3)**, 63-69 **(2010)**.
[9] Q.R.Li, L.Y.Wei, S.F.Ma; "The model analysis of vehicles situation and distribution in intersections abased on markov process", IEEE International Conference on Intelligent Transportation Systems., **2(10)**, 1076-1080 **(2003)**.
[10] D.Barash, A.Comaniciu; "A common framework for nonlinear diffusion, Adaptive smoothing bilateral filtering and mean shift ", Image Vision Computing, 73-81 **(2004)**.
[11] C.Sun; "Intelligent surveillance using inductive vehicle sensors", Intellimotion, **8(3)**, 10-100 **(1999)**.
[12] Li Jian Zhong, Li Jin Bao, Shi Sheng Fei; "Concepts, Issues and advance of sensor networks and data management of sensor networks", Journal of Software, **14(10)**, 1717-1727 **(2003)**.
[13] S.Tilak, N.B.Abu-Ghazaleh, W.Heinzelman; "A Taxonomy of Wireless Micro-sensor Network models", Mobile Computing and Commuications Review, **1(2)**, 1-8 **(2002)**.
[14] David Culler, Deborah Estrin, Mani Srivastava; "Overview of sensor networks", IEEE Computer Society Officers, **11(10)**, 15-25 **(2005)**.
[15] T.Matsuo, Y.Kaneko, M.Matano; "Introduction of intelligent vehicle detection sensors", IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, **5(12)**, 709-713 **(1999)**.