Journal of Computers Vol. 27, No. 4, 2016, pp. 197-207
doi:10.3966/199115592016122704016

An Integrated Forecasting Model of Complex Uncertainty System Based on Knowledge Discovery

Quan Liang[12*], Ming-xing Nie[12], Kai-jian Liang[3], and Yong-hui Zhang[12]

[1] Fujian Provincial Key Laboratory of Big Data Mining and Applications (Fujian University of Technology), Fuzhou 350108, China
liangquanlq@126.com

[2] School of Information Science and Engineering, Fujian University of Technology, Fuzhou 350108, China
{Nie Ming-xing: 7335303@qq.com, Zhang Yong-hui: 932839251@qq.com}

[3] School of Application Technology, Hunan Institute of Engineering, Xiangtan 421000, China
airmajor@126.com

Received 17 July, 2016; Revised 15 September, 2016; Accepted 3 October, 2016

Abstract. Forecasting for complex uncertainty systems has always been a very difficult problem, how to make accurate forecasting for complex uncertainty systems has also become a focus of the area fore many researchers. Focusing on the properties of complex uncertainty system, by related knowledge discovery theory and methods with prediction heory and methods. The paper presented a model of integrated forecasting model based on knowledge (IFMK) of multiple-targets and multiple-factors, the model utilizes suitable knowledge discovery theory and data mining methods for forecasting. Forecasting method based on IFMK can provide a stable and relatively accurate forecasting results. And analysis of exmaples indicates that the model is valid and practical by forecasting petroleum reservoir, production and demand.

Keywords: complex uncertainty system, forecasting, knowledge discovery

1 Introduction

Forecasting consciousness and simple forecasting methods of human beings have existed for a long time in the activities while human beings acquainted and reformed nature and society [1-3]. In ancient times human beings had to forecast weather for hunting and harvest and had to forecast growth of crops for living. With the development of science, technology and society, new techniques constantly emerge; production competition becomes more and more intense; the forecasting natural phenomenon needs to be more and more accurate, people have to take all factors in consideration, such as politics, military etc. All makes people realize the importance and the urgent of forecasting and forecasting methods, hence forecasting has gradually become an independent discipline. At the same time, applications of forecasting science are increasingly widespread, there has been a variety of forecasting methods and models, especially in recent years, forecasting science has made a considerable progress, brought great convenience to people's lives.

Complex uncertainty systems are difficult to be predicted for a very large number of factors which can't be determined [4-5]. There are many errors, conflicts and contradictions emerging in complex uncertainty system, thus study how to identify errors, eliminate interference and take the right decisions is directly related to the accuracy of the forecasting. Traditional forecasting methods does not solve all the problems, and also have not effective ways to deal with the emerging issues. Therefore, it needs to establish some reliable and effective integrated models to obtain correct forecastings for complex uncertainty system. In this paper, focusing on the properties of complex uncertainty system, proposes the new me-

* Corresponding Author

thod about the researches of integrated forecasting patterns base on knowledge discovery. Its purpose is to research knowledge discovery method suitable for forecasting, combining knowledge discovery and traditional forecasting methods organically, to resolve forecasting problems of complex uncertainty system.

2 Related Works

Based on history and existing information, forecasting, taking advantage of known knowledge and methods, is inferring and judging the future and unknown conditions of research objects in advance. The science which researches on forecasting rules and methods is called forecastology [6-7]. Similar foreign researches are called "future studies," "things to learn," "future researches" and etc. At first, forecasting methods is qualitative analysis. Since Principles of Economics written by Marshall was published in 1980, domestic and overseas economists began to emphasize quantitative analysis. There are lots of forecasting methods, including twenty to thirty forecasting methods used frequently. These methods are based on traditional statistics, probabilism and other theories [8-9]. Forecasting methods of traditional statistics and probabilism, decompose complex uncertainty system into simple and ascertain system to acquaint, and use statistics, probabilism theories and technological methods to process output. Just like Newton Mechanics paradigm, these forecasting methods always build ideal and exquisite forecasting models, emphasize mechanical causality, assuring a constant function relationship among the factors in the system; and thus forecasting is limited to seek the quantitative relation among the factors, and these methods only apply to ideal and simple system [10-12]. Along with the improvement of society and technology, systems become more and more enormous and complicated, requirements about forecasting theory and technique become more and more high, hence scientists devote themselves to research forecasting theories and methods of complex uncertainty systems [13]. So far, there are fuzzy forecasting, grey system forecasting, state equation forecasting, neutral network forecasting and other forecasting methods [1,6,14]. Traditional forecasting methods and existing uncertain forecasting methods both focus on research and apply to one method, overlook the features of complex elements, with mass information, qualitative and quantitative data in complex systems [15-16].

The stock market is a typical complex uncertainty system, forecasting stock price is a hot issue for stock investors, dealers and brokers. However, it is difficult to find out the best time point to buy or sell stock, since many variables will affect the stock market, and stock dataset is time series data. In contrast to other related studies, Li and Kuo proposed a hybrid approach on the basis of the knowledge discovery methodology by integrating K-chart technical analysis for feature representation of stock price movements, discrete wavelet transform for feature extraction to overcome the multi-resolution obstacle, and a novel two-level self-organizing map network for the underlying forecasting model. In particular, a visual trajectory analysis is conducted to reveal the relationship of movements between primary bull and bear markets and help determine appropriate trading strategies for short-term investors and trend followers. The forecasting accuracy and trading profitability of the proposed decision model is validated by performing experiments using the Taiwan Weighted Stock Index (TAIEX) from 1991 to 2002 as the target dataset [17]. For similar issuses, a novel ANFIS (adaptive neuro fuzzy inference system) time series model based on integrated nonlinear feature selection (INFS) method for stock forecasting was proposed [18]. Firstly, it proposed an INFS method to select the important technical indicators objectively. Secondly, it used ANFIS to build time series model and test forecast performance, then utilized adaptive expectation model to strengthen the forecasting performance. The experimental results show that the proposed method outperforms the listing models in accuracy, profit evaluation and statistical test.

In recent years, there are many other new results in associated areas. Fuzzy time series forecasting (FTSF) is a useful tool for forecasting without expert consultation as well as a user-friendly solution for non-expert forecasters. Before selecting the proper forecasting model, analysis of data series is a key step in the implementation of FTSF. Seasonality is one of the change-making dimensions of data series that also include temperature, rainfall, freight rates and vessel traffic. The proposed model is to improve the fuzzy integrated logical forecasting (FILF) model for the seasonal time series by using the bivariate fuzzy time series (FTS) approach [19], which study object is similar to "complex uncertainty system." The model is applied on the volume of vessel traffic on the Istanbul Strait in order to compare the accuracy of the proposed model with benchmark methods, and indicates good performance. Another study is to improve the FILF by utilizing multivariate inference and the partitioning problem for an exponentially distributed time series by using a multiplicative clustering approach. FTS is a growing study field in computer science and its superiority is indicated frequently. Since the conventional time series analysis requires various pre-conditions, the FTS framework is very useful and convenient for many problems in business practice [20]. With the rise of cloud computing, a huge complexity growth of the structure that is the base of the Cloud happens. Thus, to effectively manage applications and resources it is crucial the use of models and tools that create an application profile which is used to apply forecasting models to determine the most suitable amount of resource for each workload. There are models and tools that address the creation of an application profile to later apply some forecasting technique and estimate the amount of resource needed for a workload [21].

3 The Properties of Complex Uncertainty System

3.1 The properties of system

In nature and society, objects exist in the form of system. Any object could be regarded as a system, meanwhile it could also be regarded as a subsystem subordinative to a bigger system. System is a whole part consisted of various factors with specific function and mutual relation [6-7].

So far, the known uncertain information is divided into random information, fuzzy information, grey information and unascertained information. Random information is obtained from random experiment. Fuzzy information is a kind of information with unclear property boundary information, and its concept has the properties of "clear connotation, unclear expitaxy." Grey information is the information partly known and partly unknown. Unascertained information is unclear information [8, 15].

Complex system with uncertain information is complex uncertainty system. There are three fundamental properties: integrity, relativity, environmental suitability.

Integrity. system is a organic integrity consisting of several different elements, not a simple assembly. Only its structure, function, manipulability and operability work as a integrity, can the system work. If any one of it deviates from the integrality, no matter how important it is, it will lose its function.

Relativity. elements in the system are connected organically and interactional, any element can't work without other elements' support; function of any element would support other elements; any element can't work and influence the system integrity independently, so is the reflecting system.

Environmental suitability. any system is in a specific environment, it has to adapt changes of the external environment to establish in objective world and plays its role. In the process of original information inverting into host information, external environment would influence the expression of its essential properties.

In addition to above three properties, the universal objective existing is another one for complex uncertainty system too. In real life, a lot of natural systems are complex uncertainty system. Therefore, study and understanding the rules of complex uncertainty system is of particular importance.

3.2 The models and methods of forecasting

These years, the rise of knowledge discovery brought dawn to research of complex uncertainty system. In traditional statistics, there are many forecasting technology and methods. Its main method is to assign data probability distribution function of a model, draw a conclusion from the form describing probability. However, while forecasting mass information, enormous complex, uncertainty system, it appears helplessness. Existing uncertainty forecasting theories cannot consider all kinds of uncertain elements, and there are a lot of difficulties when forecasting mass data. Research on integrated forecasting model based on knowledge discovery (IFMK), aims at deriving the essence of the existing forecasting theories and methods, overcoming their limitation, to enrich and develop forecasting theory and technology. Its properties are:

Complex model. Relation of variables among the system is not obvious, variables have nonlinear and complex interaction, and knowledge discovery is suitable for solving complex problems.

Large-scale problem. Knowledge discovery always involves mass information, in some way, it adopts complex model to obtain reliable information from mass data. Hence, algorithmic computing complexity and sparsity are especially important in knowledge discovery.

Mass discrete variable. There are continuous and discrete variable data set in complex systems, most of the multivariate analysis in statistics is designed to continuous variable model. Knowledge discovery is suitable for discrete variable. Some regular methods can only use discrete variable, and need to discretize continuous variable.

Description of uncertainty system. Knowledge discovery is engaged in exploring mass and complex uncertainty system, including association rules, classifying rules, clustering rules, timing pattern, similar pattern, chaotic pattern, regression analysis, trend analysis, variance analysis, predictive analysis, etc, data exploring technology consider all kinds of uncertain elements in the system.

Acquisition of deep knowledge. Knowledge discovery aims at extracting patterns that are credible, effective, novel, potential useful, chosen and understood by client ultimately from mass data. However, traditional statistics can only obtain surface layer of data, but cannot gain internal relations and implied information among data attributes, that is, it cannot obtain deep knowledge.

On the other side, there are some typical forecasting methods such as Decision Tree, Artificial Neural Networks, Support Vector Machines, Regularization Method. Other common forecasting methods include Nearest Neighbors, Naive Bayes (belonging to the method of statistical learning). Different methods have advantages of themselves, how to integrate and use various methods is the key content of researching integrated forecasting, and it will bring a great breakthrough for the area of forecasting sciences.

4 Construction of IFMK

The overall structure of IFMK is shown in Fig.1. IFMK applies data mining, based on rough set, concept lattice theory and technology, to sort the data base roughly, to divide mass date into data sets with the same properties and forecast them according to traditional forecasting model suitable for small or medium scale data sets. Meanwhile, taking advantage of self-organizing neural network SOM clustering analysis time data, it excavates frequent timing model. IFMK adopts the four bases synergetic system structure of model base, knowledge base, data base, and method base, to form a new forecasting support system of multiple layer, isomerism, distribution with multiple goals and tasks. It includes:

- Data base. According to practical application, through a series of steps like confirming forecasting target and direction, collecting and analyzing information, abstracting principal factor, data base is ultimately formed data set for forecasting. The data set is the basis for forecasting, which will decide the correctness of the forecasting. Data set is essential for any forecasting methods. Therefore, the establishment of effective data set is a critical part of this integrated forecasting methods, methods of data mining and refining will play a huge role in here.
- Basic knowledge base. It is a kind of knowledge base based on rough set, and it is obtained from equivalence class solution; Classified knowledge base is Hasse Diagram, based on the theory of rough set, utilizing the complete node of concept lattice, every node in the diagram represents a classified knowledge; timing model base is the frequent model excavated from self-organizing neural network SOM clustering, and its expression is a vector quantity composed by a set of segments of the arc tangent.
- Based on excavated knowledge, inference mechanism, through logical deduction, reaches corresponding conclusion. Different knowledge has different inference mechanisms, and we adopt arboreal breadth-first strategy and vector distance scale computing query method. How to select the inference mechanism is also very important, and it can determine the direction and correct of forecastings. Inference mechanism that we adopt can reduce the forecasting error to the greatest extent, it has a major impact on the forecasting results.
- For data base, there are different mining algorithm and model computing method, thereby composing method base. The algorithm uses results of the analysis to define the parameters which will be used to create the optimum mining models. Then, these parameters apply to the entire data set in order to extract a viable model and detailed statistics.
- Classic forecasting model has its own range of application, and different models are suitable for different application backgrounds. In order to make models have broad application fields, forecasting model base management makes users need not to consider technology implementation and process details of model when applying forecasting model.

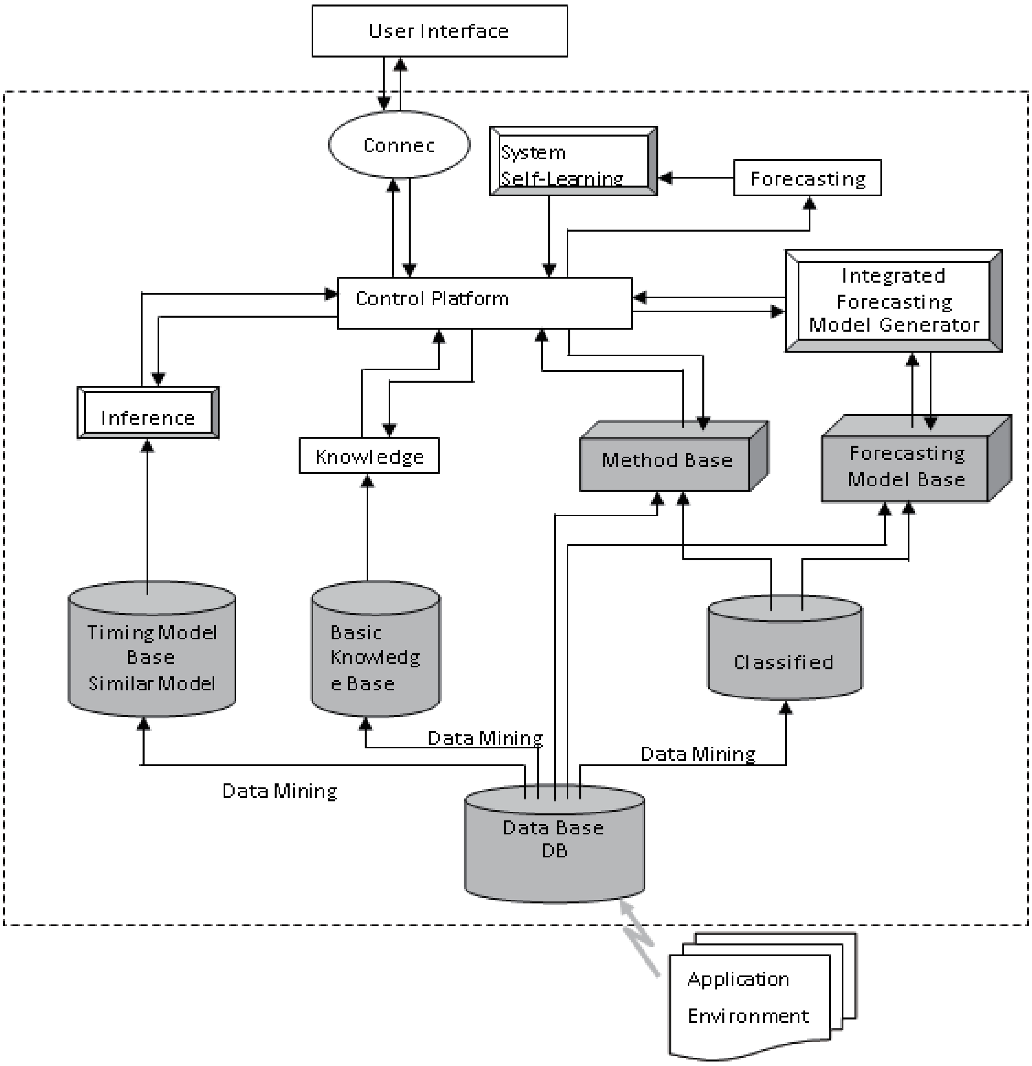

Fig. 1. Overall structure of IFMK

- Integrated forecasting model generator, aiming at different forecasting model, structure and generate selection condition, makes model choices. Usually in machine learning, there are two different models, that is, Discriminative Models and Generating Models. If to construct the model according to the conditional probability, it is the discriminant model, which basic idea is that to establish a discriminant function under conditions of limited samples, not consider the generating models of the samples, and study forecasting models directly. The forecasting model generator will take into account the different situations.
- Based on forecasting outcome, system self-learning adjusts to adaptive systems automatically, in order to find out the suitable forecasting model for application environment ultimately. Self-learning system is a system capable of running their own experience to improve the ability to control algorithm, which is an extension and development of adaptive systems. Self-learning is an important part of the method proposed in the paper. It continues to revise the forecasting results by learning and adjusting, and improve the accuracy of forecasting.
- Control platform is the key link of the system model. On one side, it helps date base, model base, knowledge base and method base cooperate; on the other side, utilizing model generator, it chooses forecasting model and controls system automatically to analyze forecasting outcomes and start learning. Meanwhile it cooperates with users.
- User interface is an indispensable component of IFMK, and it makes interface module and controls platform cooperating with system model. Favorable user interface should be with various properties,such as diversity, effectiveness, convenience, consistency and fault tolerance, etc. Good user in-

terface helps to improve the control and feedback of the system, adjust the forecasting problems and improve the availability and reliability of the forecasting system.

5 Self-Learning Mechanism in IFMK

Complex system has multiple forecasting targets, and every target adopts different forecasting model. Choosing forecasting model needs to meet certain conditions, and some forecasting targets may meet several prerequisite of multi forecasting model. Meanwhile, complex system is a system that constantly changes, and the former forecasting model would become in adaptive since the changes of the system. Therefore, IFMK forecasting system should have the function of self learning, and it not only chooses appropriate forecasting model for forecasting target automatically, but also adjusts forecasting model and adapt practical environment automatically according to the change of the system.

Assume IFMK has k multiple forecasting targets: B_1, B_2, … , B_k, and s forecasting models: M_1, M_2, … , M_s, every forecasting target has a corresponding matrix A, as follows:

$$A = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \cdots & \delta_{1s} \\ \delta_{21} & \delta_{22} & \delta_{23} & \cdots & \delta_{2s} \\ \cdots & \cdots & & & \\ \delta_{n1} & \delta_{n2} & \delta_{n3} & \cdots & \delta_{ns} \end{bmatrix} \tag{1}$$

Among them: δ_{ij} is the relative error of forecasting model *j* in experiment *i*, if the actual value of the target is *y*, forecasting value is $\hat{y}$, then

$$\delta_{ij} = \frac{|y - \hat{y}|}{y} \tag{2}$$

If we define an error threshold value as 0.05, we get a vector quantity:

$$M_i = \left(P_{i1}, P_{i2}, \cdots, P_{is}\right) \tag{3}$$

P_{ij} is the probability value that forecasting error is smaller than error threshold value for forecasting model i. If the number of element value ≤0.05 is u in line i, sum of forecasting times is n, then:

$$P_{ij} = \frac{u}{n} \tag{4}$$

accordingly we obtain matrix B:

$$B = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1S} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2S} \\ \cdots & \cdots & & & \\ P_{k1} & P_{k2} & P_{k3} & \cdots & P_{kS} \end{bmatrix} \tag{5}$$

Matrix k is defined as learning matrix corresponding with k forecasting targets and s forecasting models.

For IFMK system, after n times learning, obtained learning matrix could choose standard for forecasting model of system. That is: for forecasting target i, the forecasting model corresponding element with the highest probability value should be chosen for target’s forecasting model. In system practical application, because of the development of the system, original forecasting model of certain target will not suitable any more. Hence, learning matrix of the system always makes forecasting value and actual value in development as examples for learning matrix. In practical application, it learns and updates constantly, in order to adapt development of system.

Learning quality of learning mechanism mentioned above depends on the choice of probability value and forecasting error threshold value. If forecasting error threshold value is too big, it will not reach the anticipated forecasting accuracy; if too small, probability value of forecasting model is too small, causing difficulty of choosing forecasting model. In application of actual system, the two parameters have to be tested and adjusted for many times. The quality of the two parameters adjusting have a great impact on the forecasting results, it needs to adjust for many times in different application scenarios according to the characteristics of the application itself. After that, the formation of empirical data can be obtained depending on the applications, and on the same conditions, the similar applications can be configured as a unified parameters. This approach can effectively improve the accuracy of forecasting methods.

6 Example Verifications

We will apply research outcomes to forecast petroleum reserves and production as well as demand, in order to test validation and practicability of IFMK. Petroleum industry is complex system, including exploration, exploitation, environmental protection, civil use, generating electricity, chemical industry, industrial fuel, pipeline transportation, price, market and all kinds of problems in many fields [22]. Hence, we have to analyze all the factors influencing reserves and production as well as demand of petroleum, to extract principal factor and relation, and to proceed symbol manipulation.

Through analysis, petroleum reserves have an important relation with geologic structure. According to geologic structure, we obtain the factors related with reserves, including technological level (low, middle, high), geologic structure (Sichuan Basin, Songliao Basin, Bohai Bay Basin, Shangangning Basin, Tarim Basin, Junggar Basin, Tuha Basin, Qaidam Basin, Yitong Basin, Sanshui Basin), tectonic zone or entrapment, ratio of exploitation and reserve, time, proven reserves and so on. Petroleum production generally adopt resource-reverses-production forecasting method, that is, petroleum production has a close relation with the magnitude of resource quantity. If exploration and research work is complete relatively, petroleum resource is fairly precise, it is reliable to forecast petroleum production according petroleum resource. Meanwhile, petroleum production relates with technology, geologic structure, and oleaginousness in oil layer. So far petroleum in China is used mainly for industrial energy resources, chemical materials, lives of urban and rural residents, etc. Factors that influence petroleum demand are development tempo of national economy, growth rate of population, living standard, petroleum consumption and etc. Petroleum Consumption relates with agriculture, industry, construction industry, transportation and telecommunications, commerce, service department, lives of urban and rural residents, growth rate of population and etc. Hence, forecasting petroleum demand in different periods needs to consider the trend of main factors mentioned above. Not because social demand of petroleum quantity is low, industrial development of petroleum is slow and production is stuck, which makes social demand in depression and supply-demand relation tense, actually its demand exceeds its supply.

Fig. 2 is the curve of annual total reserves of petroleum. Obviously, reserves have nonlinearity and instable characteristics. The feature of this non-linear and non-stationary is a typical instance of complex uncertainty systems. The curve shows that the storage capacity of oil and gas is in a high or low level in a period of time, it appears an obvious instability.

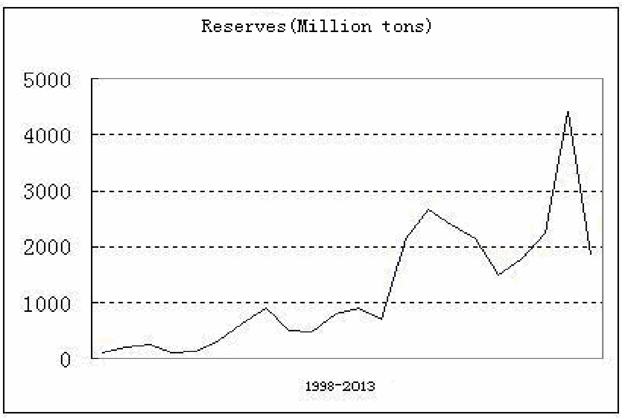

Fig. 2. The curve of annual total reserves of petroleum

Seen from Table 1, time model and neural network are suitable for reserves; trend extrapolation forecasting model is most suitable for production; binary linear regression is most suitable for demand.

Table 1. The forecasting result error of each forecasting model

	Reserves (Million tons)		Yield (Million tons)		Demand (Million tons)	
	Actual value	2140.7	Actual value	304.7	Actual value	173.42
	Predicted value	Relative error (%)	Predicted value	Relative error (%)	Predicted value	Relative error (%)
Unary linear regression	-	100	278.0	8.7	156.3	9.8
Binary linear regression	-	100	287.6	5.6	168.4	2.8
Trend extrapolation model	1173.1	45.2	296.5	2.7	182.6	5.29
Moving average method	3183.2	48.7	324.7	6.8	192.4	10.9
Exponential average method	2919.9	36.4	282.15	7.4	148.3	14.5
Timing model	2307.7	7.8	288.25	5.4	156.4	9.8
Neural network model	2001.5	6.5	319.43	4.8	163.2	5.89

Based on knowledge discovery, IFMK system is a integrated system combining traditional forecasting methods and date mining methods dynamically. Source data from users has to preprocess, in order to apply to knowledge discovery and forecasting. There are many date mining methods of knowledge discovery, and IFMK integrates rough set basic knowledge mining algorithm, classified knowledge mining algorithm, concept lattice construction algorithm, timing model mining algorithm. Forecasting functions chooses forecasting target firstly, using searching mechanism to search corresponding classified knowledge base and timing knowledge base to obtain the corresponding minor data set and knowledge set, and then uses forecasting model to forecast; after getting forecasting result, proceeds analysis, according to the result of analysis, proceeds simulating and self-learning repeatedly, to get forecasting effect ultimately.

Fig. 3 is the forecasting result of neural network forecasting model. It shows the forecasting results has a similar curve of trend with the actual data curve, that is, maintain a consistency of the trend for two curves, forecasting results and actual data. In general, the two have a consistent with a relatively high degree. However, at some point, there is still a high difference, showing the accuracy of forecasting of the defect.

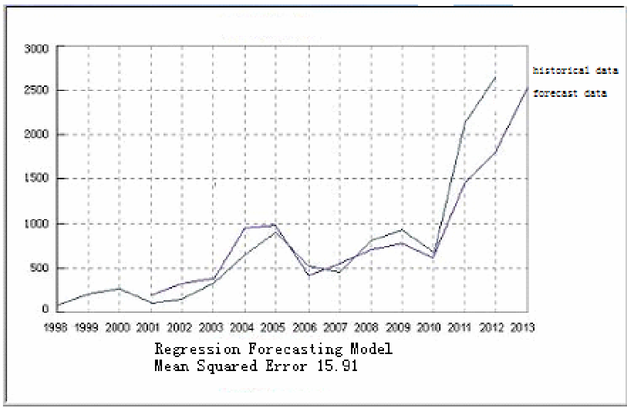

Fig. 3. Neutral network model

Fig. 4 is regression forecasting model. With respect to forecasting of the neural network model, this forecasting method has a higher degree of agreement with the actual data, indicating that the forecasting accuracy is better. The curve of forecasting results fluctuates with the actual data curve, the same general trend is also maintained, which indicates it is an effective method of forecasting too.

Fig. 5 is forecasting model with S function. From forecasting results of IFMK, it conforms to actual data, verifying the validation of IFMK model. In fact, the verification results show that IFMK forecasting method has very high accuracy and stability, especially for complex uncertainty systems, integrated forecasting method provides a higher level of forecasting, it can take advantage of superiorities of different forecasting methods and give a more accurate forecasting result, which is the importance of the method just lies in.

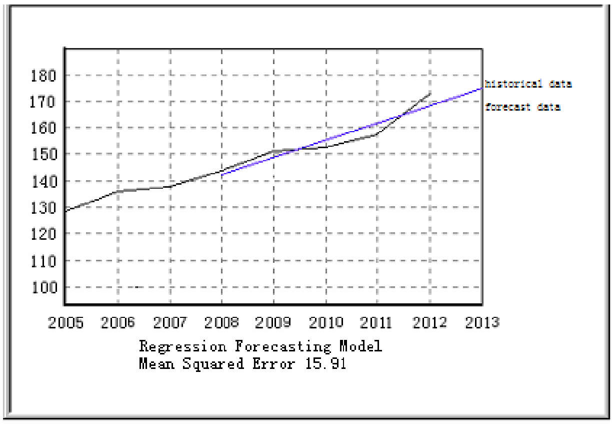

Fig. 4. Regression forecasting model of demand

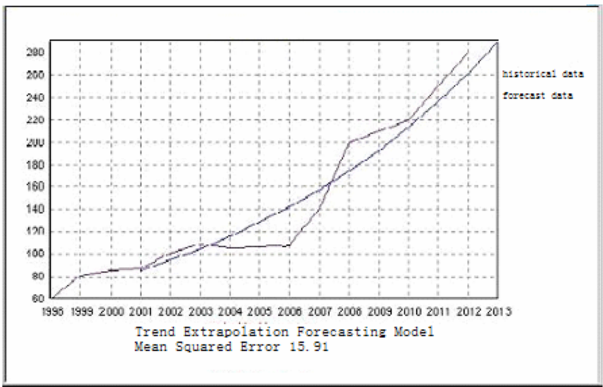

Fig. 5. Annual production trend extrapolation forecasting model

7 Conclusions

Petroleum reverses is related to geological mining structure, whose space-time state is changeable and influence factors are various. While petroleum production and demand have a close relation with national economy development, temporal technology level, industrial and agricultural productivity, living standard and social progress. Therefore, petroleum industry system is the complex system with properties of complexity, nonlinearity, uncertainty, mutability and etc. This article, from specific to general, by extracting properities of complex uncertainty system, analyzes the limitation of forecasting theories and methods in traditional mathematics, which idealized and simplified complex system, and it always builded ideal and exquisite forecasting model, but not suitable for forecasting complex uncertainty system. However, knowledge discovery is a hot issue of present date base and artificial intelligence fields, aiming at the complex formal mass, incomplete, uncertain information source in real world, this article proposes research on IFMK, and offers example verification.

Acknowledgment

This work is supported by Projects of Fujian Provincial Department of Science & Technology (No.2016H0001, No. 2016J01750), Project of Fuzhou Science and Technology Bureau (No. 2015-G-53) and Hunan Province Technology Planning Projects (No. 2014FJ3054). During study the issue, many workmates and friends give us a lot of suggestions which are useful for the work, I should say thanks to them.

References

[1] H.-H. Tsai, Knowledge management vs. data mining: research trend, forecast and citation approach, Expert Systems With Applications 40(8)(2013) 3160-3173.

[2] Q. Liang, J. Zhang, Y.-H. Zhang, J.-M. Liang, The placement method of resources and applications based on request prediction in cloud data center, Information Sciences 279(20)(2014) 735-745.

[3] G.M.P. Wanderley, C.A. Tacla, J.-P. A. Barthès, E.C. Paraiso, Knowledge discovery in task-oriented dialogue, Expert Systems with Applications 42(20)(2015) 6807-6818.

[4] D. J. L. Salmeron, W. Froelich, Dynamic optimization of fuzzy cognitive maps for time series forecasting, Knowledge-Based Systems 105(2016) 29-37.

[5] Q. Liang, Y.-Z. Wang, Y.-H. Zhang, Resource virtualization model using hybrid-graph representation and converging algorithm for cloud computing, International Journal of Automation and Computing 10(6)(2013) 597-606.

[6] O. Addam, A. Chen, W. Hoang, J. Rokne, R. Alhajj, Foreign exchange data crawling and analysis for knowledge discovery leading to informative decision making, Knowledge-Based Systems 102(15)(2016)1-19.

[7] C.-J. Chang, D.-C. Li, Y.-H. Huang, C.-C. Chen, A novel gray forecasting model based on the box plot for small manufacturing data sets, Applied Mathematics and Computation 265(15)(2015) 400-408.

[8] A.I. Saleh, A.H. Rabie, K.M. Abo-Al-Ez, A data mining based load forecasting strategy for smart electrical grids, Advanced Engineering Informatics, 30(3)(2016) 422-448.

[9] S. Barak, M. Modarres, Developing an approach to evaluate stocks by forecasting effective features with data mining methods, Expert Systems with Applications 42(3)(2015) 1325-1339.

[10] Dongxuan? Cai Lijun, Based on The Spatial Interpolation Method for Classification of Unbalanced Data Sets. Computer Simulation. 2012.12.

[11] Z. Yin, Zhanghua, Et Al. Forecasting and Analysis on Consumption and Demand of Petroleum in China, Natural Resource Economics of China, 2009.1.

[12] Gaowei, Hefeng, Parallel Modeling And Description Research Base on Flow Data Set Constraints. Computer Engineering And Design. 2010.6.

[13] Nielei, Wang, Hongqiang, Jiang Weidong, A Method For Constructing a Comprehensive Integration and Demonstration of Integrated System, Journal of National University of Defense Technology, 2012.1

[14] H. Tai, J. Wang, X. Gao, Research on The Determination of Reliability Threshold Value of Decision Data. Ordnance Industry Automation. 2008.5

[15] J. Qin, Thomas Fahringer, Advanced data flow supportfor scientific grid workflow applications, in: SC07, 2007.

[16] J. Xiao, S. Baker, I. Matthews, T. Kanade, Real-time combined 2D+3D active appearance models, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2004.

[17] S.-T. Li, S.-C. Kuo, Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-Based SOM networks, Expert Systems with Applications 34(2)(2008) 935-951.

[18] C.-H. Su, C.-H. Cheng, A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock, Neurocomputing 205(2016) 264-273.

[19] E. Bulut, Modeling Seasonality Using the Fuzzy Integrated Logical Forecasting (FILF) Approach. Expert Systems with

Applications 41(4) (2014) 1806-1812.

[20] O. Duru, A multivariate model of fuzzy integrated logical forecasting method (M-FILF) and multiplicative time series clustering: a model of time-varying volatility for dry cargo freight market, Expert Systems with Applications 39(4)(2012) 4135-4142.

[21] R. Weingärtner, G.B. Bräscher, C.B. Westphall, Cloud resource management: a survey on forecasting and profiling models, Journal of Network and Computer Applications 47(2015) 99-106.

[22] H. Wu, H. Zhu, Review on the concentrated research of data set, Computer Applications and Software 12(6)(2012) 102-115.