Load Forecasting Method based on Rough Set and Information Fusion Theory

CHEN Lifeng, JIANG Yuechun, HE Fei, ZHANG Jie, YANG Xuqiong
College of Electrical and Information Engineering, Hunan University
Changsha 410082, Hunan Province, China

Abstract*—With the requirements of the power load predicting accuracy becoming higher and higher, and the actual power load is also a complex nonlinear system limited by various uncertain factors. Given only considering the influence of a single model on the load forecasting, the predicting accuracy is absolutely not high under the comprehensive effects of multiple factors. Ensuring to make the electrical model satisfy the requirements, there is access to improving the accuracy of load predicting results through analysis of the relationship between load and multiple factors in power load forecasting. This paper combines the rough set theory with information fusion, firstly a decision table of attributes is established on the basis of historical datum, for the purpose of simplifying the attribute index in data mining to obtain important information. Following step is to rank the importance of the factors influencing the load according to the degree of effect, with grouping these sorted factors into attributes combination and using the same kind of load forecasting method to predict each combination's own results, and finally all groups of the predicting results are integrated by data fusion, where then final predicting result can be obtained. The results of the algorithmic example show that groups of predicting results integrated based on information fusion theory are better than the results got through any casual single attribute combination. And fusion results are a combined action of a variety of related factors, enhancing the favorable trend, where at the same time it has reduced the uncertainty brought by multiple factor comprehensive effect. Based on information fusion the predicting results are helpful to reduce the prediction error, which is feasible and effective.

Key words: Power load prediction; The rough set method; Data mining; Information fusion

I. INTRODUCTION

With the explosive growth on the amount of data in today's information age, the wave of "big data" is coming, fast and furious. The era of big data proposes a new challenge for the development in electricity industry. With the advance of electricity industry, intelligence, automation and interaction will bring many datum casing electrical big data generation. Due to the numerous index of power load forecasting, different goals of all kinds of power load forecasting, even with the same kind of theory, the forecast model to predict no matter from the form, content or effect also vary widely. The load in the medium and long-term prediction of power system is greatly influenced by economy, society and meteorology factors and other uncertainties, etc. How to extract the key factor from multiple factors and reduce the results' uncertainty arising from the comprehensive effect of multiple factors, and improving the precision has been becoming a serious problem to be solved. Aiming at these problems, the paper based on the literature [8] proposing medium-and long-term load forecasting based on partial least squares regression analysis, an more optimized and advanced method is put forward to improve accuracy of prediction. The rough set theory and information fusion are introduced into the model, which respectively advance data pre-processing and optimize the integration of multiple sets of results after prediction, in order to reduce uncertainty caused by varieties factors affecting and improve accuracy of forecasting results.

II. THE ROUGH SET THEORY

In the era of big data, on account of the characteristics on grid data volume, which is large amount, variety type, and more speed, the traditional attribute simplified methods have been unable to complete the pre-processing on electrical big data. To this end the paper proposes a method on power big data simplified processing based on cloud computing technology. The rough set theory which offers a new way for data mining, is a new mathematical tool to deal with uncertain and imprecise problems. In this paper, rough sets method is used in power load forecasting to analyze the dependence of load on each factor and the importance of various properties in condition attribute set.

In the decision system R={U, A}, the rough set divides A(attribute) into two categories C and D, where C is condition attribute, D is decision attribute, and U is domain. The ternary collection R={U,C,D} is known as a decision-making system. An important application of rough set in the data analysis is to find the dependence among attributes from a known decision system R={U,C,D}. The dependency is defined as follows.

If D depends on C in extent of the degree k ($0 \le k \le 1$), expressing that $C \Rightarrow_k D$, if there is.

$$k = \gamma(C, D) = \frac{card(POS_C(D))}{card(U)} \qquad (1)$$

The coefficient k represents a certain proportion in all elements of domain U, namely on the basis of the attribute set C, the element can be fully or partially divided inside of area in partition U/D, k is called dependency.

In practical application, the form of expression on decision-making system is a two-dimensional table. The decision attribute D and conditions attribute C correspond to the properties of the columns in the table, while all of the objects are contained in the domain corresponding to each row in the table, and all the records of tables consist of the

domain U. The two-dimensional table, namely decision table, is composed of condition attributes and the decision attribute in the domain. Specific steps of attribute reduction based on rough set theory are as follows.

1) By using certain discrete algorithm of rough set, historical data of the condition attributes and decision attribute variables in decision table are made discretization. A new decision table is to go on attribute reduction after attributes discretization and then decision rules after reduction can be obtained. Calculate the degree of dependence of decision attribute D on condition attribute set C.

$$H(R_D / R_C) = -\sum_{[x]\subset U/R_C} p[x] \sum_{[y]\subset U/R_D} p([y]/[x]) \ln(p([y]/[x])) \quad (2)$$

The larger $H(R_D / R_C)$ is, the greater decision attribute D is depending on condition attributes C, whereas lower.

2) Calculate the weights of attributes, and sort the degree of importance of each factor in the attribute set. Formula (3), (4) is used to calculate the importance of properties in the attribute set C.

$$\omega(\mathrm{c}_i, C, D) = H(R_D / R_{c_i}) - H(R_D / R_C) \quad (3)$$

$$H(R_D / R_{c_i}) = -\sum_{[x]\subset U/R_{c_i}} p[x] \sum_{[y]\subset U/R_D} p([y]/[x]) \ln(p[y]/[x])), j = 1,2,\cdots,n \quad (4)$$

The greater the result of formula (3) is, it can be indicated that the higher importance of attribute in C is, whereas lower.

III. INFORMATION FUSION THEORY

Information Fusion, also known as data fusion, is multi-sensor information resource using time and space. In order to reduce effect on making decision brought by environmental uncertainty, using data from multiple sources can improve the reliability and robustness of the outcome, and produce more accurate, more complete, more reliable estimation and interpretation than by a single data. This article discusses the application of D-S evidence theory in load forecasting.

D-S evidence theory is a promotion on the Bayesian statistical inference theory. Evidence theory can provide support for one or more of the proposition, and the level of support on each proposition the evidence refers to is called the reliability, which can use set functions to express, namely Basic Probability Assignment. If identification framework Θ in the set of functions is on the meet:

$$m(\phi) = 0, \quad \sum_{A\subset\Theta} m(A) = 1 \quad (5)$$

m is called basic probability assignment function on framework Θ, m(A) called basic credible number. If $m(A) > 0$, then A is called the focus element. Combination rule of D-S evidence theory is a law reflecting the combined effects of evidence. The function defined by the following formula is new basic probability assignment, thus reliability's synthetic rule can be drawn, namely

$$m(A) = \frac{\sum_{A_i \cap B_j = A} \mathrm{m}_1(\mathrm{A_i})\, \mathrm{m}_2(\mathrm{B_j})}{1 - \sum_{A_i \cap B_j = \phi} \mathrm{m}_1(\mathrm{A_i})\, \mathrm{m}_2(\mathrm{B_j})} \quad (6)$$

IV. LOAD FORECASTING METHOD

The process of load prediction in power system typically includes many uncertainties, such as the observation error of original data, selection on feature and attribute, selection of predicting algorithm, and so on.Based on information fusion methods, combining with the existing load predicting method form a unified framework for load forecasting. Using this unified approach no matter in different environments, or taking different forms, it also can help to reduce uncertainty caused by varieties of factors and to improve the predicting accuracy.

To reduce the effect on selecting properties of load prediction, the rough set method can be adopted to calculate the influence degree of each attribute. And following steps are going on attributes' sorting and grouping, and then unified partial least squares regression analysis is applied on for each group to respectively predict each group's results. Finally D-S evidence theory is applied to fuse the results together as a final prediction. It can clearly be seen that results after fusion combine comprehensive influence of multiple factors with attribute's information to enhance the positive trend, which achieve in reducing the uncertainty brought by different methods. Based on the same lines, a variety of predicting methods can be used to predict the same set of results, then the results are going on integration, and a better prediction can be got than by a single method.

Figure 1 shows the overall framework of load forecasting method based on information fusion. The paper adopts rough set method to do with history data preprocessing, sorting attribute according to the degree of importance and grouping it. and D-S evidence theory is applied to integration.

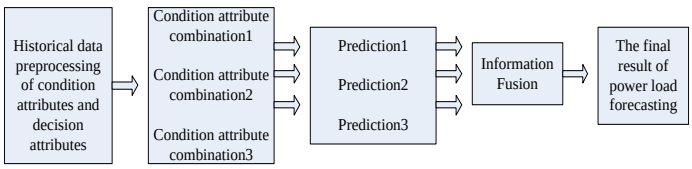

Fig. 1. Flow diagram of load forecasting method based on information fusion

A. *Mathematical model of integrating weight in Evidence Theory*

When using Dempster combination rule, firstly weights must be extracted before going on integrating weights. Following the mathematical model of extracting weights is to be introduced. The predicting error of the three models respectively were e_1, e_2 and e_3, taking ω_1, ω_2 and ω_3 as the corresponding weights,the total predicting error and its variance can be expressed as

$$E = \omega_1 e_1 + \omega_2 e_2 + \omega_3 e_3 \quad (7)$$

$$D(E) = \omega_1^2 D(e_1) + \omega_2^2 D(e_2) + \omega_3^2 D(e_3) + 2\omega_1\omega_2 \mathrm{cov}(e_1, e_2) + 2\omega_1\omega_3 \mathrm{cov}(e_1, e_3) + 2\omega_2\omega_3 \mathrm{cov}(e_2, e_3) \quad (8)$$

Where, $\mathrm{cov}(e_i, e_j)(i, j = 1,2,3)$ represents the covariance of the e_i , e_j .The following step is solving the problems on ω_1 , ω_2 and ω_3 to find the minimum value of $D(E)$, giving

$$\frac{\partial D(E)}{\partial \omega_1} = 1, \frac{\partial D(E)}{\partial \omega_2} = 1, \frac{\partial D(E)}{\partial \omega_3} = 1 \quad (9)$$

B. *Mathematical model of integrating weights*

When the weights are going on integration, the higher accuracy of the predicting values to greater weight, and whereas less weight is given. In the integration model of load forecasting, it can be an analogy thinking that the weight of the model is similar to the value of the basic reliability on evidence theory when using Dempster combination rule on evidence theory for integration on the multiple sets of weights. Formula (6) is applied to go on fusion by reliability combination rule, the value of basic reliability fused can be made as a predictive model's incorporates weights.

V. CASE STUDY

In order to further illustrate the advantage of the improved and advanced PLSR method based on rough set, examples adopt the original data of literature [8] to do simulation calculation. On the basis of its medium and long-term load forecasting based on partial least squares regression analysis, this paper proposes optimizing methods to improve prediction, and introduces the rough set theory and information fusion method to data preprocessing in front of predicting and optimization on the basement of fusing multiple sets of results after predicting. The implement process of load forecasting is shown in Figure 2.

Firstly the importance of the condition attribute x_i among condition attribute set will be calculated, following it being sorted and grouped. According to the rough set method's steps above mentioned, above all data must be going on preprocessing and charactering. Then according to the formula (3), (4), the importance of each condition attribute in the condition attribute set can be calculated, the results namely ω_i are shown in Table 1. Rough set method is applied to analyze the relative importance of conditions.The top six factors ranking in the degree of importance of condition attribute set are selected to set up group 1 in examples simulation, which is $x_1, x_2, x_5, x_6, x_8, x_9$ .The top six factors and load D build up a decision-making system as combination 1. The factors raking 7 and 8 that is x_4, x_7 are filled into the group 1 adding load D as a new decision-making system that is combination 2, and finally all 10 factors and load D consist in decision-making system as a combination 3. Based on indicators of social economic factors and annual electric load of Hengyang from 2002 to 2005,historical data have been given in literature [8], and by using unified partial least squares regression analysis the total electricity consumption prediction from 2002 to 2005 in the combination of three groups decision-making system can be respectively calculated, and its specific results can see in table 2.

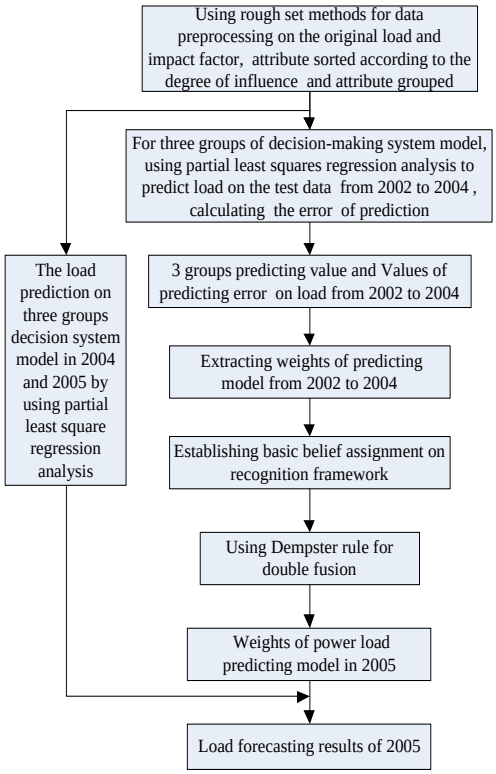

Fig. 2. The process of load forecasting

TABLE 1. Importance of each condition attribute in set of condition attributes

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0.0227	0.0243	0.103	0.0137	0.0374	0.0543	0.0163	0.0241	0.0238	0.0074

TABLE 2. Predicting value of each model

Years	Actual consumption /TWh	Combination 1			Combination 2			Combination 3		
		Predictive value /TWh	Absolute error /TWh	Relative error %	Predictive value /TWh	Absolute error /TWh	Relative error %	Predictive value /TWh	Absolute error /TWh	Relative error %
2002	4.15165	3.90972	0.24193	5.820522	3.89854	0.25311	6.096612	3.89309	0.25856	6.227885
2003	4.81841	4.34932	0.46909	9.735369	4.27047	0.54801	11.373254	4.19916	0.61925	12.85175
2004	5.44120	4.88678	0.55442	10.189671	4.79323	0.64797	11.908586	4.70931	0.73189	13.45097
2005	6.03652	5.30735	0.72917	12.079311	5.18431	0.85221	14.117571	5.07392	0.96260	15.94622

A. The extraction of weight

Table 2 shows the predicting load results of three groups decision system from 2002 to 2004,mathematical model of integrating weight in evidence theory aforementioned and introducing information fusion theory are applied to predict the final value of load in 2005 to go on fusion of multiple sets of predicting results for destination on obtaining a final prediction. Based on this method the load forecasting in the follow-up year can also be carried out in this way.

First of all, the weight of the three groups' decision-making systems in 2002-2004 were respectively extracted, and then it can be multiple fused through the Dempster rule, and then the fusing results are made as the weight of each predicting model in 2005. The values of weight corresponding to each group's predicting model in the right year from 2002 to 2004 were calculated on extracting weight model,as shown in table 3.

B. The mathematical model of integrating weight

By combining D-S evidence theory concerning the basic concept of reliability distribution, assigning with the mathematical model aforementioned, the weight in table 3 can be converted into the corresponding values of basic reliability, and then through the Dempster combination rule of reliability function, the corresponding belief function is conducted multiple fusion, in the predicting model the right weight of each group to be predicted can be got. Specific results of integrating weight as shown in table 4.

Table 3. The value of weights corresponding to each group predicting model in 2002-2004

Years	Weights of Group 1	Weights of Group 2	Weights of Group 3
2002	0.359656	0.327282	0.313062
2003	0.438943	0.316277	0.24478
2004	0.439849	0.315821	0.24433

TABLE 4. Specific results of integrating weight

Integration step	Group 1	Group 2	Group 3
Re-integration	0.463358	0.333816	0.224921
Double Fusion	0.526763	0.304304	0.186341

TABLE 5. Load forecasting results D-S evidence theory fusion

Years	Actual consumption /TWh	Predictive value /TWh	Absolute error /TWh	Relative error %
2004	5.44120	4.923608	0.517592	9.512457
2005	6.03652	5.318801	0.717719	11.889612

C. Forecasting results analysis

By the above the weights of the fusion results can be calculated in the fusion of predictive results of forecasting years Y, namely its initial value of each group's predicting model multiplied by the corresponding value of weighting fusion can be concluded as final predicting results integrated of forecasting year, the result in table 5.

Comparing the results before fusion with the results after fusion, it can be concluded that D-S information fusion theory conducting integration in multiple sets of predictions were better than the results obtained by the single combination of attributes, which synthesize the effects of many factors and strengthen the positive trend, thereby to reduce the uncertainty caused by the different methods and errors of the prediction.

VI. CONCLUSION

This paper presents a power load forecasting method based on rough set theory and information fusion, moreover gives predicting model and its flow chart. This method can combine effects of various factors on the power load and reduce uncertainties caused by attribute selection, method selection and data observation errors. The partial least squares regression analysis was chosen as a predictive calculation method in this paper, and the ideas in combination of the rough set method with the information fusion are applied on analysis and predicting on the actual data, D-S information fusion theory conducting integration in multiple sets of predictions were superior to predictions got by separate property portfolio then a better predictive value is obtained than by a single method, which proves the effectiveness of the method. It can be concluded that the idea of combining rough set method with information fusion method can reduce the uncertainty brought by multiple factors comprehensive effecting to improve the accuracy of load prediction.

References

[1] Zhang Yuanyuan, Sun Guoqiang, Qi Xiaoming. A Survey on Rough Set Theory[J]. Metrology and Mesurement Technology, 2008:1-4.

[2] Gao Shuang, Dong Lei. Mid-long Term Wind Speed Prediction based on Rough Set Theory[J]. Proceeding of the CSEE,2012,32(1):32-37.

[3] Li Ping,Song Kun. Medium and long-term load forecasting using associated clustering based on rough sets[J].Relay, 2008,36(1):43-49.

[4] Xu Zhongwei, Li Longshu. Algorithm study of data mining based on rough set theory[J].Microcomputer Development, 2001 (01):136-142.

[5] Wang Guoyin, Hu Feng. Rough Set Extensions in incomplete Information Systems [J]. Frontiers of Electrical and Electronic Engineering in China, 2008, 35(04):68-79.

[6] Gu Xihua,Niu Dongxiao. Power Load Forecasting of Multiple Factors influencing the Gray Neural Network Combination[J].Electric Power of East China, 2006, 34(07):548-551.

[7] Liu Yaonian,Li Jian. Medium long-term load prediction based on rough set theory and least squares support vector machine[J]. Chinese power, 2007,40(10):42-44.

[8] Mao Lifan, Jiang Yuechun, Medium- and Long-Term Load Forecasting Based on Partial Least Squares Regression Analysis[J]. Power System Technology, 2008, 32(19):71-78.

[9] Zhang Qinbao, Cheng Haozhong. Short- term load forecasting based on attribute reduction algorithm of rough sets and support vector machine[J]. Power System Technology, 2006,30(8):56-61.

[10] Xiao Zhi, Ye Shijie. Rough set method for short term load forecasting[J]. Journal of systems engineering, 2009,24(2) : 143-150.

[11] Ni Ming,Shan Yuanda. Evidence Theory and its Application[J]. Automation of Electric Power System 1996 ,20(03):76-80.

[12] Wu Jingqiu, Sun Qi.Short-term load forecasting model fusion based on D-Sevidence theory[J]. Electric Power Automation Equipment, 2009,29(04):66-70.

[13] Jin Long. Prediction Method based on Support Vector Machine and Information Fusion[J]. Oil and Gas Geophysical Seismic.2006 ,41(01).

[14] Sun Quan,Ye Xiuqing. A new combination rules of evidence theory[J]. ACTA Electronic SCIENCE. 2000(08):117-119.

[15] AI-Hamadi HM, Soliman SA. Long-term and mid -term Electric Load Forecasting based on Short-term Cor-relation and Annual Growth. Electric Power -Systems Research. 2005,32(04)118-130.

[16] Sun Qi, Yang Wei.Research on load forecasting of fusion between D-S evidential theory improved BP network[J]. Relay,2007,35(07):61-65.