

# Application Research on Data Mining and Artificial Intelligence Theory in Short-Term Power Load Forecasting

Honghai Wang

School of Electronic and Electrical Engineering, Anhui Sanlian University, Hefei, 230601, China  
[sanlian\\_whh@163.com](mailto:sanlian_whh@163.com)

Data mining technology provides an effective research tool for us to process uncertain, noisy and implicit information. Rough set theory, as a kind of typical data mining algorithm, provides an effective tool for research on analysis and induction of inaccurate data, mining relations among data and discovering potential knowledge. The paper will establish a short-term load prediction model based on rough set theory. It utilizes rough to carry out attribute reduction for various historical data related to load, gets rid of those irrelative attribute to decision information and simplifies input variable so as to shorten the search space of neural network and improve the prediction performance.

## 1. Introduction

Change rule of short-term load is complicated as time changes. It is difficult to describe it with accurate mathematical model and a mass of uncertain factors exist in short-term load prediction, such as various information related to actual load (climate, humidity, precipitation, special event, etc.) (Wu and Niu, 2009) These factors and power load show non-linear relation. In the previous load prediction, these relations are provided by experienced dispatch personnel, but for load prediction, the knowledge is inaccurate and it is difficult to judge the change of various random fluctuation only by experience in current open power market (Chiu and Kao, 1997).

In addition, it still needs many statistic information and priori knowledge in artificial intelligence forecasting method widely applied at present. However, the information may have information deficiency or incompleteness with the change of operation environment of information. Moreover, in short-term power load prediction, various factors affect load prediction and the impact degrees are various. Among these various factors, not all the conditions are necessary. Some of them are relative and some of them are independent, so it only needs some or a few conditions to induce the conclusion. Thus, it needs us to use new intelligence method for analysis and mine various relevant information methods from historical data automatically so as to enhance the precision and efficiency of load prediction (Paarmann and Najjar, 1994).

Data mining technology provides an effective research tool for us to process uncertain, noisy and implicit information. Rough set theory (Jia and Tian, 2008), as a kind of typical data mining algorithm, provides an effective tool for research on analysis and induction of inaccurate data, mining relations among data and discovering potential knowledge. The chapter will establish a short-term load prediction model based on rough set theory. It utilizes rough to carry out attribute reduction for various historical data related to load, gets rid of those irrelative attribute to decision information and simplifies input variable so as to shorten the search space of neural network and improve the prediction performance (Chen and Huang, 2014).

## 2. Basis of rough set theory

Rough set theory was proposed by Polish scholar Z. Pawalk in 1982. It provides a new mathematical tool for processing imprecise and uncompleted information. Rough set theory is established based on classification mechanism. It interprets classification to be equivalence relation in specific space, while relation of equivalence forms the division of space. The theory construes knowledge as data division and each divided set is called concept (Li and Huang, 2014). Main idea of rough set theory is, on the premise of maintaining the

constant information system classification capacity, utilizing the known knowledge base to depict inaccurate or uncertain knowledge with the known knowledge in the knowledge base and import problem decision or classification rule through knowledge supplement and reduction.

The most prominent difference between rough set theory and other theories processing uncertain and inaccurate problems is: rough set theory doesn't need to provide any prior information other than the data set needed to process and the description and processing about problem uncertainty is objective. Moreover, the theory doesn't include the mechanism of processing (Huang et.al, 2002) inaccurate or uncertain original data, so the theory has strong compliment with other theories processing inaccurate or uncertain problems, such as probability theory, fuzzy mathematics, evidence theory, etc.

### 2.1 Information system

Information system is the object of rough set theory. It is a data set, and usually expressed as a data sheet. Each line of the data sheet represents an object and the object can be case, event, etc., while each list of data sheet is attribute of the object and these attributes can be the feature, measurement, etc. of the object.

Information system can be formalized expressed as  $S = \langle U, Q, V, F \rangle$ , including  $U = \{x_1, x_2, \dots, x_n\}$  is finite non-empty object set, usually called domain of discourse,  $Q$  is finite non-empty attribute set. If the attribute set  $Q = A \cup D$ ,  $A, D$  of information system  $S$  is condition attribute set and decision attribute set, such information system can also be called decision system (or decision table),  $V = \cup V_q$  is the union set  $V_q$  of all attributes of  $q$  value domain,  $f$  is the information function specifying the attribute value of each object, namely:  $f: U \times Q \rightarrow V$ .

### 2.2 Relative reduction of attribute

For attribute set  $A, D \in Q$ , matrix unit element of relative discernibility matrix is:

$$m_{ij} = \begin{cases} \{a \in A : f(x_i, a) \neq f(x_j, a)\} \\ \emptyset \end{cases} \quad (1)$$

Relative discernibility function is:

$$f_D^{x_i} = \wedge \{ \vee (m_{ij}) : 1 \leq j \leq n, j \neq i, m_{ij} \neq \emptyset \} \quad (2)$$

When  $a_{i1} \wedge a_{i2} \wedge \dots \wedge a_{ip}$  is prime  $f_D$  of discernibility function, attribute  $\{a_{i1}, \dots, a_{ip}\}$  is relative reduction of  $D$ . Thus, we can seek for relative reduction through seeking for prime  $f_d$ .

### 2.3 Dependence degree of importance of attribute

If all attribute values of attribute set  $D$  totally depend on attribute values of attribute set  $A$ , it can be expressed as  $A$  and  $D$ . The dependence degree of attribute  $D$  to  $A$  is expressed with  $\gamma(A, D)$ :

$$\gamma(A, D) = \text{card}(POS_A(D)) / \text{card}(U) \quad (3)$$

According to the definition of dependence degree, the importance degree of equivalence class of  $U/IND(D)$  is defined as:

$$SGF(a, A, D) = \frac{\gamma(A, D) - \gamma(A - \{a\}, D)}{\gamma(A, D)} \quad (4)$$

## 3. Load forecasting model based on rough set theory

An accurate load prediction model should be able to describe various factors related to load directly. From basic concept of rough set theory, we can see rough set can reduce and get rid of unnecessary information, which is good for information classification and reduction. In order to find directly related condition with load and improve the prediction accuracy and computation speed, the chapter uses rough set theory to reduce attribute of various factors affecting load, seeks for the necessary conditions directly related to power load, considers them as input vector of fuzzy neural network and applies the decision rule established by optimal reduction set to neural network structure design.

As is shown in Diagram 1, the step of neural network load prediction model established based on rough set theory in the paper is as below:

Establish initial information table according to historical load data, relevant information and historical data

Make discretization of original data and establish real numeric type decision tables.

Make attribute reduction to established decision table, get optimal condition attribute set and kernel related to load prediction

Establish neural network model according to decision rule set and determine initial weight of neural network.  
 Provide training to neural network  
 If the fitting error of neural network meets the requirement, then end.

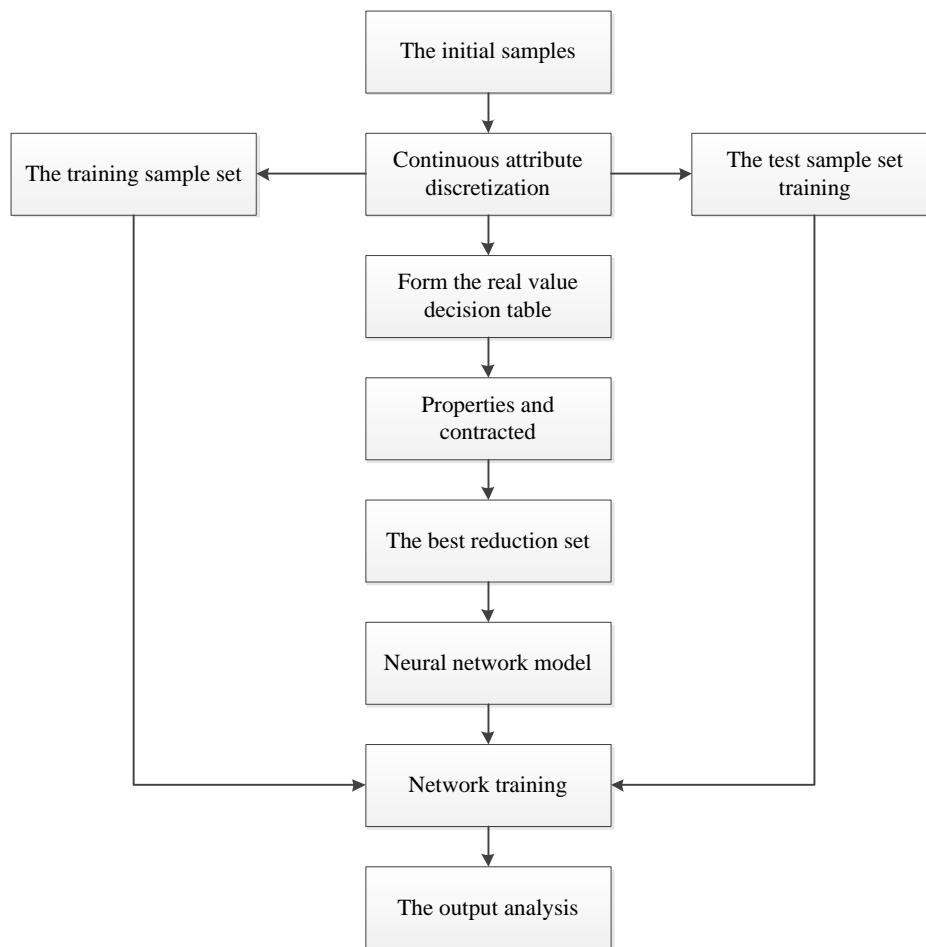


Figure 1: Based on rough set neural network optimization process

### 3.1 Determination of relevant factors

The paper mainly considers the impact of climatic condition change on power load, makes fuzzy processing to climatic condition (including total cloud cover  $cl$ , wind speed meter  $Ws$ , maximum temperature  $T_{max}$ , minimum temperature  $T_{min}$ , humidity  $Hu$ , air pressure and precipitation  $Rf$ ) and according to the features. At this time, the total number of condition attribute is  $N$ . membership function is shown as Diagram 2, the division standard of each climatic condition membership condition predicts daily load LD cluster to be  $\mathcal{L}$  decision categories. In this paper,  $\mathcal{L}=5$ , showing the load is very low (1-VL), low (2-L), ordinary (3-NM), high (4-H) and very high (5-VH).

### 3.2 Determination of fitness function

Discernibility matrix  $M_d(s)$  contains all the complete information differentiating  $x_i, x_j$ , and any reduction set can substitute the whole condition attribute without changing original dependence relation and resolving power, then the reduction set should represent the information in discernibility matrix as much as possible. If attribute B is the reduction set of attribute C, Formula (5) can be expressed as:

$$SGF(B, A, D) = \frac{\gamma(C, D) - \gamma(B, D)}{\gamma(C, D)} = t - \frac{\gamma(B, D)}{\gamma(C, D)} \quad (5)$$

At this time, it can be used to indicate the degree of approximation of attribute set B to C, called approximate error of reduction. If attribute set R is the optimal reduction set C, then double C,  $D = \text{double } R, D$ , namely

SGF(R, C, D)=0. At this time, reduction set cannot make correct decision if it lacks any one element. In addition, a clear and simple decision rule should have less antecedent combination, namely the number of produced reduction set condition attributes should be as less as possible. Therefore, construct fitness function according to the above analysis, as is shown in Formula (6). The first item expresses the number of condition attributed contained by reduction set is lest, the second time expresses the coverage of reduction set in discernibility matrix is largest and the third item expresses the approximation degree of reduction set condition attributes to total condition attributes is greatest.

$$F = a_1 \left( \frac{1}{L(R)} + \sum_{i=1}^M S_i / kl \right) - a_2 sig(R, C, D) \tag{6}$$

Including, L(R) is the number of condition attributes contained by reduction set, s is 0 or 1. When reduction set and some one element of discernibility matrix  $M_D(S)$  have intersection, it is 1, or else 0.  $K_i$  is the number of elements in discernibility matrix: sig is the dependence degree of reduction set R to condition attribute set C,  $a_1$  and  $a_2$  are any non-negative random weight. We hope to get the most simple and most knowledge covered rule, so give greater weight to the first two items of fitting function and give the smaller weigh to the last item.

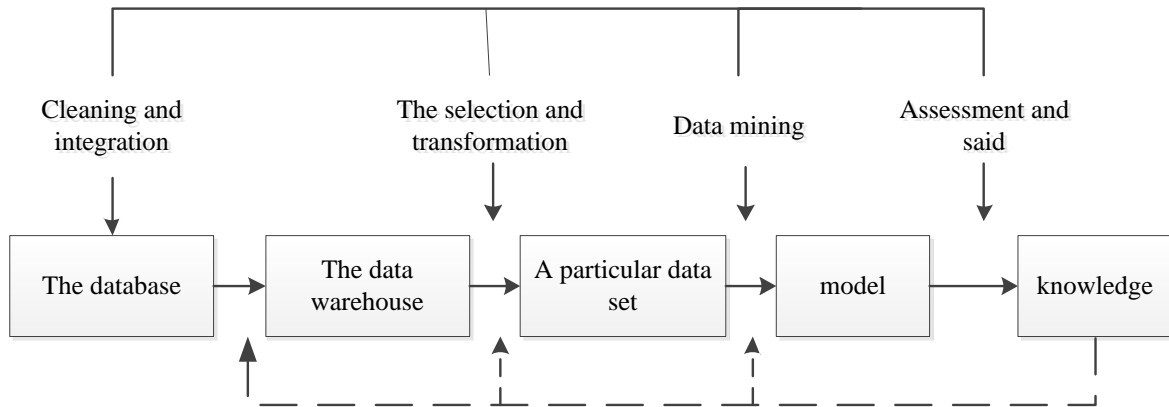


Figure 2: Process schematic knowledge mining

**3.3 Basic process of attribute reduction**

If the attribute quantity of attribute combination in discernibility matrix is 1, it is the kernel attribute of decision table. It indicates that the remaining condition attributes cannot distinguish two different records of decision types in the information table other than attribute, while kernel attribute and the combination of it with any other attributes may constitute reduction set. Therefore, when utilizing discernibility matrix to seek for reduction set, in order to simplify, consider kernel attribute as feature attribute of data set and deem it as good gene to construct new attribute combination in genetic algorithm. However, the remaining useful attribute other than kernel attribute is got from analyse the matrix elements that attribute combination quantity is not 1. The paper considers other attributes other than kernel attributes as the gene pool, produce an attribute at random each time. Join it in the kernel attribute and judge whether the newly produced attribute combination meets the requirement or not. It is good for reducing search space of genetic algorithm and improving computation efficiency.

Computation step of attribute reduction algorithm is as below:

Compute core CORE according to discernibility matrix, get rid of all attributed included by CORE from the original condition attributes and take the remaining attributes as candidate gene.

Select any attributes excluding CORE at random and join them in the CORE, then initial population including n individuals generate. Each attributed set generated is expressed in the form of binary string, including each single bit expresses a condition attribute, 1 expresses the attribute belongs to reduction set, and 0 means it doesn't belong to reduction set.

According to formula (4), compute fitness function of each individual. Produce new population through selection, cross and mutation operation and make fitness function approach maximum gradually. In order to get the simple reduction set, define two mutation probabilities  $p_{m1}$  and  $p_{m2}$ , including  $p_{m1}$  expresses the probability from 1 to 0 and  $p_{m2}$  expresses the probability from 0 to 1.

When the computation reaches the fixed genetic algebra, then finish the computation, or else turn to c). Here, two opposite directions of mutation probabilities  $p_{m1}$  and  $p_{m2}$  are defined, which makes the individual change toward the constant reduction of attribute combination quantity with greater probability  $p_{m1}$ . It is good for finding the minimum attribute combination. During the training, the optimal individual is maintained. These individuals constitute the optimal regulation set  $R$  jointly. It has less quantity of attribute and better resolving power. The reduction set got in the next chapter will be used for optimizing neural network structure so as to get a simple and transparent neural network.

In addition, it still needs many statistic information and priori knowledge in artificial intelligence forecasting method widely applied at present. However, the information may have information deficiency or incompleteness with the change of operation environment of information. Moreover, in short-term power load prediction, various factors affect load prediction and the impact degrees are various. Among these various factors, not all the conditions are necessary. Some of them are relative and some of them are independent, so it only needs some or a few conditions to induce the conclusion. Thus, it needs us to use new intelligence method for analysis and mine various relevant information methods from historical data automatically so as to enhance the precision and efficiency of load prediction.

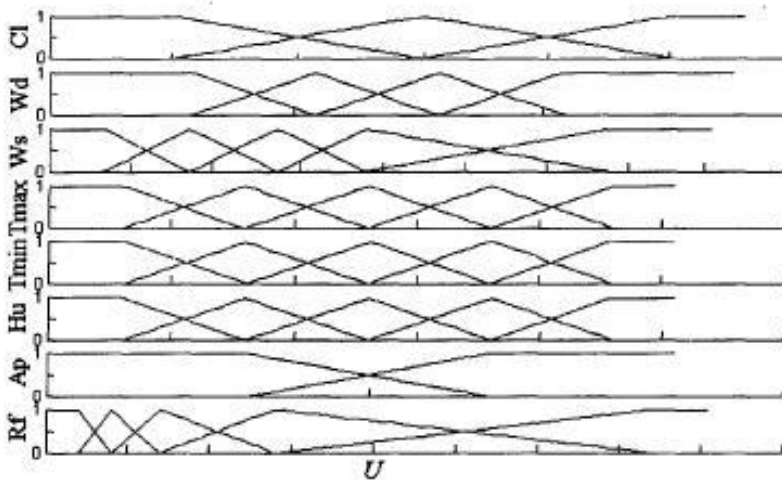


Figure 3: Each feature vector membership function

#### 4. Conclusion

Through effective classification of load mode type, it can provide better load prediction classification model for power load prediction so as to improve fitting performance of neural network. However, a mass of factors affect short-term power load, and excessive input vectors not only increases the scale of neural network model, reduces training efficiency of neural network, but also affects fitting effect of neural network. How to use the information properly and effectively has become the key problem to improve load prediction precision. To this end, the paper studies the factors affecting load prediction and proposes power load prediction algorithm based on rough set theory specific to climatic conditions. It starts with pre-processing of historical data firstly, cuts off redundant attributes and number of neural network input vectors through attribute reduction, then starts with reducing neural network model and improving transparency of neural network model, and applies rough set decision rule to neural network linkage and initial connection weight design. Thus, the paper realizes the purpose of improving prediction performance of model and improving efficiency and precision of load prediction.

Research result indicates the knowledge-based prediction model and load classification model observed through data mining are able to describe the complicated relation in load prediction model and trace load mode change timely so as to improve short-term power load prediction effect. Data mining technology provides an effective research tool for us to process uncertain, noisy and implicit information. Rough set theory, as a kind of typical data mining algorithm, provides an effective tool for research on analysis and induction of inaccurate data, mining relations among data and discovering potential knowledge. The chapter will establish a short-term load prediction model based on rough set theory. It utilizes rough to carry out attribute reduction for various historical data related to load, gets rid of those irrelative attribute to decision information and simplifies input variable so as to shorten the search space of neural network and improve the prediction performance.

### Acknowledgment

This work was supported by the Nature Science Foundation of Anhui Province Education Department (No. KJ2016A251), the Nature Science Foundation of Anhui Sanlian University (No.2014Z020), Quality Project of Anhui Province Education Department (No. 2014jxtd046) and the revitalization plan project of higher education of Anhui Province Grant (No. 2013zytz082).

### Reference

- Chiu C.C., Kao L.J., 1997, Combining a neural network with a rule-based expert system approach for short-term power load forecasting in Taiwan, *Expert Systems with Applications*, 13, 299-30, DOI: 10.1016/S0957-4174(97)00048-1.
- Chen Q.S., Huang W., 2014, Short-term power load forecasting with least squares support vector machines and wavelet transform, *Applied Mechanics and Materials*, 494, 1647-1650, DOI: 10.4028/www.scientific.net/AMM.494-495.1647.
- Huang H.C., Hwang R.C., Hsieh J.G., 2002, Short-term power load forecasting by non-fixed neural network model with fuzzy BP learning algorithm, *International Journal of Power and Energy Systems*, 22, 50-57, DOI: 10.1007/978-3-319-09330-7\_47
- Jia Z.Y. Tian L., 2008, Short-term power load forecasting based on fuzzy-RBF neural network, *Proceedings of International Conference on Risk Management and Engineering Management*, 6, 349-352, DOI: 10.1109/ICRMEM.2008.41.
- Li L.J., Huang W., 2014, A short-term power load forecasting method based on BP neural network, *Applied Mechanics and Materials*, 4945, 1647-1650, DOI: 10.4028/www.scientific.net/AMM.494-495.1647.
- Paarmann L.D., Najjar M.D., 1994, Short-term power load forecasting based on autocorrelation function optimization, *Midwest Symposium on Circuits and Systems*, 2, 1475-1478, DOI: 10.1016/j.epr.2016.04.003.
- Wu J., Niu D.X., 2009, Short-term power load forecasting using Least Squares Support Vector Machines(LS-SVM), *2nd International Workshop on Computer Science and Engineering, WCSE 2009*, 1, 246-250, DOI: 10.1109/WCSE.2009.663.