

The Building of Urban Power Grid Information Database and it's Data Mining

Jingmin Wang

e-mail:jingminwang126@126.com

Yinling Li

e-mail: 421785334@qq.com

Abstract: The planning and project decision-making of urban power grid need accurate and comprehensive information resources. Based on the common theme of information needs of power grid, the grid comprehensive information database was built, and a preliminary index system was proposed. On the basis, use data mining technology to get more rules and knowledge of the power grid's information database. Finally, take the association rules algorithm for example, with 'the coordination of the development of power grid and load demand' as the theme, analyze the implied rules, and give it's reference for power grid planning and decision making.

Keywords: Power Grid Planning, Information Database, Data Mining, Association Rules

I. INTRODUCTION

Correct power grid planning and decision-making need timely and comprehensive information. The common themes of information needs about power grid are as follows: the failure analysis of power grid, the reliability analysis of power grid, the transmission lines and equipment status analysis of power grid, load forecasting, intelligentization of power grid, the coordination of the development of power grid and load demand and so on. However, the present information management system of power system cannot meet the information demands. So, it needs to build a comprehensive power grid information database, integrating various information, including the internal operation information and external information. In this paper, the information demand of the comprehensive information database of urban power grid was researched, a preliminary index system and information composition was given, that is to say, the conceptual information database, and specific logical and physical design will not be discussed temporarily.

Assume that after grid integrated information database is constructed, it will accumulate a large amount of complicated data. So, in this paper, we use data mining methods to analyze power grid information database. Data mining technology has efficient analyzing and processing

ability, and it can rapidly and accurately get some hidden rules in the grid information database [1-2]. At home and abroad, the application of data mining technology in electric industry are also going deeper. For example, analyzing solar power plant performance[3], wind power forecasting[4], identification of Power Grid Topological Structure[5], optimization for energy management in a renewable micro-grid system[6], Power grid failure analysis[7], Optimal voltage control[8], Optimal voltage control[9], reliability of power equipment[10], research of state estimation in power network[11]. In practice, the British national grid corporation analyzed real-time wind power using data mining technology, and predicted the future generated electricity; Based on support vector machine(SVM) data mining, Australia's national electricity market predicted the price of electricity. In China, several province and power production enterprises also launched data mining system[12-14].

II. THE INFORMATION COMPOSITION OF URBAN POWER GRID INFORMATION DATABASE

In order to meet the information demand of the themes of power grid planning and decision-making. Urban power grid comprehensive information database need integrate a lot of information, through investigation and research, and communicate with experts, we determines the urban power grid comprehensive information database should include the following preliminary indicators: the equipment of power grid, the power quality, the grid loss, the grid layout, the security and reliability of power grid, the smart grid related indicators, the single substation operation indicators, the single line operation indicators, the electricity customer information and government planning, etc.

The preliminary information composition(index system) of the power grid comprehensive information database is shown in table 1. Due to space limitations, the specific field is not shown in this paper.

Table 1 The preliminary information composition of the urban power grid comprehensive information database

First class indicators	Second class indicators
1 The equipment of power grid	1.1 The ratio of energy-intensive transformer
	1.2 The percent of passing wire cross-sectional area
	1.3 Substation centralized monitoring coverage
	1.4 Single line and single transformer substation of 110kv in proportion
	1.5 The rate of line insulation
2 The power quality,	2.1 Eligibility rate of voltage
	2.2 Eligibility rate of frequency
	2.3 Eligibility rate of Harmonic

3 The grid loss	3.1 The statistical attrition rate of distributing line
	3.2 The theoretical attrition of distribution losses
4 The grid layout	4.1 Power supply radius
	4.2 The power distribution in the load center
	4.3 Throw-cut mode of the capacitor
	4.4 Mode of operation
	4.5 The length of transmission line
5 The security of power grid	5.1 N-1 pass
	5.2 Short-circuit current
	5.3 Eligibility rate of the power factor
6 The reliability of power grid	6.1 Reliability of power distribution reliability
	6.2 Average interruption times of customer
	6.3 Average interruption hours of customer
7 The smart grid related indicators	7.1 The scale basis of smart grid
	7.2 Technology support of smart grid
	7.3 The application effect of smart grid
8 The single substation's operation indicators	8.1 Load rate
	8.2 Transformer overload rate
	8.3 Power supply radius of distribution network lines of Substation
	8.4 The rate of high loss line
	8.5 Substation capacity
9 The single line's operation indicators	9.1 The line overload
	9.2 The line operating year
	9.3 Line status
10 The electricity customer information	10.1 The information of big customer's reporting installation
	10.2 The information of limited supply
	10.3 Information for load growth
	10.4 The information of low voltage complaint
11 Government planning	11.1 The information of population
	11.2 The information of GDP
	11.3 The information of government planning and policy

III. THE ASSOCIATION RULES METHOD OF DATA MINING

Data mining methods, including classification, prediction, clustering, association rules, sequential patterns found, exceptions and trend found, etc. Among them, the association rules method is used to find the information hidden in the data set which is useful to the user. The association rules method are summarized in this paper, because the method has good applicability for power grid information database, and for other methods, this article will not discuss them in this paper.

3.1. Basic concept and steps of association rules method of data mining^[13]

Association rules method has the following basic definitions:

(1) Denote the data sets as D ,

$D = \{t_1, t_2, \dots, t_k, \dots, t_n\}$, Among,
 $t_k = \{i_1, i_2, \dots, i_m, \dots, i_p\}$ ($k = 1, 2, \dots, n$) is called transaction, i_m ($m = 1, 2, \dots, p$) is called item.

(2) Suppose $I = \{i_1, i_2, \dots, i_m\}$ as the set of all the items. Any subset of I is called the Itemset of D . If $|X| = k$, call the set X as k -itemset. Suppose t_k and X as the item and itemset of D , if $X \subseteq t_k$, then call item t_k contained by

itemset X .

(3) The number of items in X contained by D is called the supporting number of item set X , recorded as σ_x . The supporting number of itemset X is recorded as $\text{support}(X)$:

$$\text{support}(X) = \frac{\sigma_x}{|D|} \times 100\%$$

Among, $|D|$ is the item number of set D , if $\text{support}(X)$ is not less than the minimum support of the user specified (min_support). Then, Record X as the frequent item set.

(4) If X, Y is itemset, $X \cap Y \neq \Phi$, implication $X \Rightarrow Y$ is called association rules.

The confidence coefficient of the association rule $X \Rightarrow Y$ is recorded: $\text{confidence}(X \Rightarrow Y)$:

$$\text{confidence}(X \Rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)} \times 100\%$$

The least confidence coefficient is recorded as min_confidence . Support coefficient and confidence coefficient are the two important concepts describing association rules. The former is used to measure statistical significance of the association rules in the whole data set. The latter is used to measure the credible degree of association rules.

3.2. The classical algorithm of association rules- Apriori algorithm

Apriori algorithm is one of the most influential algorithm for association rules. The Algorithm is based on the prior knowledge that all the subsets of frequent itemsets is frequent. The method to finding out the frequent itemsets is iteration step by step. Due to space limitations, this article does not list the program code of Apriori algorithm.

IV. THE CALCULATING EXAMPLE OF THE COORDINATION BETWEEN THE DEVELOPMENT OF POWER GRID AND LOAD DEMAND BASED ON ASSOCIATION RULES ANALYSIS

Among all the themes of power grid planning and decision, the coordination of power grid development and load demand reflect the relationship between the power grid construction and economic development. If the power grid develops too fast, it will cause a waste of resources, and produce unnecessary environmental stress; If the grid develops too slowly, it will become the bottleneck of economic development and threatens the safety of power grid.

This paper take the coordination of urban power grid development and load demand for example, elaborates the process of the data mining of the grid information database, provides an idea for the research of data mining, and other themes can be analyzed by the analogy.

4.1. The properties and record set of the data mining of association rules for 'the coordination of power grid development and load demand'

Take the city of X for example, it will analyze the coordination of power grid development and load demand of each district according to grid partition (area), and analyze which factors influence the coordination of each area.

And the properties of the coordination between power grid's development and load demand include:

- (1) Time dimension: 2001-2013.
- (2) The district's name: 1 zone, 2 zone and 3 zone.
- (3) The type of incongruity:

According to the information of each district in X city from the urban power grid information database (including power supply, power consumption and high load, etc.), and the information of government planning and social and economic development (including regional GDP, GDP growth, population, population growth, the key project, etc.) and power load demand (big customers reporting information, information for load growth, low voltage complaints, information of limited power supply, etc.), integrate the above information, and analyze the regional incongruity type, recorded for level1, level2. The level1 reflects that the power's demand is greater than supply (grid's expansion is lag, it cannot meet the social need, it need to increase the power grid's construction and power grid system's optimizing); Level2 reflect that the power supply exceeds the demand (too fast/advanced grid expansion, more than social need, need a modest slowdown, should not have new grid project).

(4) The factor or phenomenon for incongruity:

A The length of transmission line is not reasonable: integrate the area's economic development and population to calculate the optimal length of transmission line and compare with the actual length of the transmission line. Record it with level1 and level2, and level1 reflect lack of construction and level2 reflect that transmission line is too much.

B Substation equipment's capacity is not reasonable: integrate the following information: substation load rate and transformer overloading rate, power supply radius of distribution network, substation's standby, substation's capacity, name level1 and level2 for lack of construction and overweight.

C Transmission line's condition is bad: integrate circuit overload ratio, circuit operating life and line's state, to judge if the status of transmission is bad.

D Poor reliability and security of power supply: integrate reliability of electric, qualified rate of voltage, centralized monitoring coverage of substation, and the information of power grid's failure, to get if the grid power supply's reliability and safety is poor.

E Other: in addition to the above influencing factors.

On the basis of the establishment of the properties, get the relevant records from 2001 to 2013, such as table 2, due to space limitations, only take 2001 for example.

Table 2 The record of the coordination between development of power grid and load demand for each year

Time	District	Type of incongruity	The factor or phenomenon for incongruity
Year=2001	District=1	Type=level1	Reason= The length of transmission line is not reasonable level1; Reason=Substation equipment's capacity is not reasonable level1; Reason= Transmission line's condition is bad;
Year=2001	District=2	Type=Level1	Reason= The length of transmission line is not reasonable level1; Reason= Transmission line's condition is bad;
Year=2001	District=3	Type=Level2	Reason= The length of transmission line is not reasonable level2; Reason=Substation equipment's capacity is not

reasonable level2;
Reason= Poor reliability and security of power supply::

4.2. The results of numerical example of the data mining for association rules

Use association rules to analyze the grid information database of X city, set the minimum support for 0.1, input the record into the matlab software, and part of the calculated output is shown as table 3:

Table 3 The frequent item set of the association rules analysis of the coordination between development of power grid and load demand

Frequent itemsets of two times
{1 zone, Type=level1}
{1 zone, Reason= The length of transmission line is not reasonable level1}
{1 zone, Reason= Substation equipment's capacity is not reasonable level1}
{Type=level1, Reason= The length of transmission line is not reasonable level1}
{Type=level1, Reason= Substation equipment's capacity is not reasonable level1}
{Type=level1, Reason= Transmission line's condition is bad }
{Reason= The length of transmission line is not reasonable level1, Reason= substation equipment's capacity is not reasonable level1}
{Reason= The length of transmission line is not reasonable level1, Reason= transmission line's condition is bad }
{Reason= Substation equipment's capacity is not reasonable level1, Reason= transmission line's condition is bad }
Frequent itemsets of three times
{1 zone, Type=level1, Reason= The length of transmission line is not reasonable level1}
{1 zone, Type=level1, Reason= substation equipment's capacity is not reasonable level1}
{1 zone, Reason= The length of transmission line is not reasonable level1, Reason= substation equipment's capacity is not reasonable level1}

Association rules can be generated from table 3 which is shown as table 4.

Table 4 The generated association rules

District=1 zone → Type=level1
District=1 zone → Reason =The length of transmission line is not reasonable level1
District=1 zone → Reason= substation equipment's capacity is not reasonable level1
Year=2011, Reason= transmission line's condition is bad level1 → Type=level1
Year=2011, Reason= substation equipment's capacity is not reasonable level1 → Type=level1
Reason= The length of transmission line is not reasonable level1 → Type=level1

4.3. The analysis of the results of association rules' data mining

Through the analysis of output, we found that:

(1) In '1 zone', it always occurs the phenomenon of the power supply exceeding the demand. It shows that power grid's construction is lag and it cannot meet the social development's needs. It need to increase the construction and optimization of power grid system. Also, the transmission line's construction is inadequate.

(2) 'Type =level1', that is electricity demand being greater than supply which always accompanied by the phenomenon that 'substation equipment's capacity is not reasonable level1', 'The length of transmission line is not reasonable' and 'transmission line's condition is bad' .

(3) For the three phenomenon that 'substation equipment's capacity is not reasonable', 'The length of transmission line is not reasonable' and 'transmission line's condition is bad', their correlation appeared considerably. The probability of each two appear at the same time is larger.

(4) Through the analysis of association rules, we can get the knowledge and information that we need. For example, 'Reason= substation equipment's capacity is not reasonable level1 → Type = level1' shows that because of insufficient construction for the substation and too little

investment, it causes that power demand being greater than supply. For similar phenomenon, it will not be stated in detail.

Integrated the above analysis, we can see that 'transmission line's insufficient', 'inadequate substation equipment' and 'insufficient maintenance of line' will cause the power demand being greater than supply. It means that the grid does not have the capacity to provide high quality and reliable electricity, and cannot meet the social development and urban load's growth. It can be concluded that, during the time of power grid's planning and decision-making, due to inaccurate information on the growth of load forecast and the long grid construction's cycle, it lead to insufficient power grid's project construction and inadequate investment which is unable to meet the needs of the economic development.

Thus, on the basis of full data mining of the power grid information database, we can get more information resources and knowledge which can provide support for power grid planning and decision making. According to the change of the power grid's operation state, it can give warnings, and avoid the risk of power grid planning and decision making. Also, it can provide reference and basis for power grid investment decisions, and be aware of problems and avoid the pitfalls, so as to get right decision.

V. CONCLUSION AND PROSPECT

This paper puts forward the formation of urban power grid's comprehensive information database, and propose the preliminary index system, accordingly, provide direct and root information for power grid planning and decision-making; On this basis, introduce the method of data mining based on grid database, took the theme of 'the coordination of power grid's development and load demand'. For example, elaborate the process of the data mining of association rules, and get the relevant knowledge and rules, so as to provide basis for the power grid's project planning and construction. This article only give the preliminary information composition, and the index is not detailed and adequate ,it can't solve all the problems and risk that exist in the power grid's planning and decision-making, and further studies are needed.

ACKNOWLEDGEMENTS

This research is sponsored by:
 The Fundamental Research Funds for the Central Universities (2014XS86).

REFERENCES

- [1] Piatetsky-Shapiro G, Frawley W. Knowledge Discovery in Databases[M].MIT Press, Cambrige,MA,1991:24-27.
- [2] Guo Xintong, Wang Hongzhi, Song Yangqiu. Brief survey of crowd sourcing for data mining[J].DEC,2014:7987-7994.
- [3] Maciejewski H, Valenzuela L, Berenguel M, Fernandez-Reche J, Adamus K, Jarnicki M. Analyzing solar power plant performance through data mining [J].Solar Energy Engineering, v 130,n4,p044503(3 pp.),Nov,2008.
- [4] Mangalova E, Agafonov E. Wind power forecasting using the k-nearest neighbours algorithm[J]. FORECASTING, APR-JUN, 2014:402-406.
- [5] Wei Zhenbo, Liu Junyong, Zhu Mi, Dai Tingling, Wang Jiajia. Identification of Power Grid Topological Structure Based on Network Data Mining[J].Automation of Electric Power Systems,25 Feb,2011,v 35,n 4,p 12-17.
- [6] Parvizimosaed Mehdi, Farmani Farid, Rahimi-Kian Ashkan. A multi-objective optimization for energy management in a renewable micro-grid system: A data mining approach[J].RENEWABLE AND SUSTAINABLE ENERGY,MAR,2014.
- [7] Shuwei Zhai, Qiguang Shi, Yanling Yang, Rong Jin, Huanguang Lei. The monitoring of Power plant's running status and reviewing data mining technology[J]. East China Electric Power,2012,(2).
- [8] Saiz-Marin Elena, Lobato Enrique, Egidio Ignacio. Optimal voltage control by wind farms using data mining techniques[J]. RENEWABLE POWER GENERATION,MAR,2014.
- [9] Chun Zhao Zhou, Yan Yan, Hao Zhang, Chao Wang, Liang Wang. Load Forecasting of Green Power System based on Data Mining[J].Applied Mechanics and Materials, v 340,p 1050-4,2013.
- [10] Arsiry E. A, Antoshchuk S. G, Arsiry V. A, Kravchenko V. I. Data mining for complex simulation to improve reliability of power equipment. Radio Electronics and Computer Systems,2012:89-94,Language:Russian.
- [11] Liang Guoyan, Gao Yang, Zhang Liu, Xu Aoran. Research of State Estimation in Power Network Based on Data Mining Technology. Applied Mechanics and Materials, v 448-453,p 2594-8,2014.
- [12] Zhiqiang Ye, Jiashu Xu. Decision support system Guangdong nuclear power based on knowledge base [J].Beijing, nuclear science and engineering,2006(3),79-81.
- [13] Xu Tan, Study on the algorithm of discovering classification rules based on decision tree [J].Journal of yunnan university,2000(10),415-419.
- [14] Zhengfeng Wang, Xiaofeng Hu. The construction and application of wide-area measurement system of Anhui province's power grid [J].China Power,2008,(7).
- [15] Xun Liang, Data mining algorithms and applications [M].Beijing, Beijing university press,2006.

AUTHOR'S PROFILE



WANG Jing-min

E-mail:jingminwang126@126.com

Brief Introduction: Doctoral tutor, School of Economics and Management, North China Electric Power University, The research direction: Information management and decision analysis, Management science and application, Project assessment and risk management.



LI Yin-ling

E-mail: 421785334@qq.com

Brief Introduction: Master graduate student, School of Economics and Management, North China Electric Power University, The research direction: Management science and application, Project assessment and risk management.