Research On Data Mining Algorithm In Power Marketing Analysis

HaoMiao Wang
State Grid Liaoning Electric Power Co., Ltd
lq@xilai2017.cn, whm0505@139.com

Abstract: In the analysis of power market economic evaluation indicators, it is more common to use technical data mining for data calculation based on technology maturity analysis. This paper is based on a distributed computing platform for data mining, and parallelizes the traditional FP-Growth algorithm, so that the FP-Growth algorithm can be applied to the mining of big data association rules. Experimental results show that the algorithm can greatly improve the efficiency of mining association rules for massive data. The model is trained and tested on a distributed cluster built into the laboratory using customer electricity data, and then the predictions are validated to determine whether the difference between the predicted values and historical data for the same period exceeded the set threshold.

1. Introduction
The electric power industry is the lifeblood of supporting national life and economic development and establishing comprehensive, scientific and executable channel evaluation system. On the basis of rich data support provided by massive big data of electric power marketing, comprehensive evaluation is conducted for each electric power marketing channel, and channels and channels are found according to the evaluation results[1]. Defects in operation management and optimization and improvement are important embodiment of the incorporation of social responsibility into the reform and development work of electric power companies[2].

Data mining is a mature and stable parallel computing framework. The data mining framework adds some unique advantages over all the advantages of the MapReduce framework[3]. For example, all calculations can be stored in memory. Iteration is more efficient; More operators are used, which provides more flexibility in programming.

Based on memory computation, the algorithm studied in this paper uses data mining to store frequent item sets, and introduces the concept of matrix to reduce the number of scanned transaction databases. It also uses local and global pruning methods to reduce the number of candidate frequent item sets.

2. Power marketing analysis based on data mining

2.1 Marketing strategies of power companies
From electric power company itself, its own channels of selling electricity directly by the company itself is responsible for the operation management. Directly selling electricity process is relatively simple, and has a huge customer base, which is easy to carry out information for electricity and subscription notification and a series of intelligent service. It also should devote lots of resources and
take the corresponding operation cost and risk, and limited service network is difficult to meet the
demand of customers universal service[4].

From the perspective of management, the impact of each indicator on the management objective is
not the same. If the subjective experience is completely relied on to determine the index weight, it is
easy to operate but too subjective. Therefore, the analytic hierarchy process (AHP) is used to
determine the index weight. It is based on expert scoring and analytic hierarchy process data processing[5-7].

2.2 Establishment of power marketing indicators
AHP is used to decompose the problem into multiple levels, so as to clarify the complex problems of
the meridian and channel, process the independent and internal relationships of the different evaluation
indicators of the evaluation system, and finally expand layer by layer to separate the indicators at each
level. The importance of is quantified, calculated layer by layer, and the weight of each index at the
same level is obtained by processing and calculating the judgment matrix [8-9]. The specific method of
use is as follows:

(1) All indicators under each level construct a judgment matrix, construct a judgment matrix \( \lambda \),
and mark the fingers of the same layer as \( R_{ij} \)

\[
\lambda = \begin{bmatrix}
1 & R_{12} & \cdots & R_{1n} \\
R_{21} & 1 & \cdots & R_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
R_{n1} & R_{n2} & \cdots & 1
\end{bmatrix}
\]

(2) The subjective judgment of each indicator is quantified, and the relative importance scale
between each two indicators is obtained according to Table 1.

| Scale | Definition                  | Description                                      |
|-------|-----------------------------|--------------------------------------------------|
| 1     | Equally important           | Both factors are equally important               |
| 3     | More important              | Two factors are slightly more important than the other |
| 5     | Obviously important         | Two factors are obviously more important than the other |
| 7     | Much more important         | Two factors are more important than the other     |
| 9     | Very important              | Two factors are more important than the other     |
| 2, 4, 6, 8 |               | Intermediate value above                          |

2.3 Power marketing data calculation based on data mining
The FP-Growth algorithm prediction process of power marketing base data mining platform is shown
in Figure 1.
Figure 1. Parallel FP-Growth algorithm prediction flow chart based on data mining

The weight of each index at the same level is obtained by processing and calculating the judgment matrix.

First, the judgment matrix is calculated, and the maximum characteristics of the matrix can be obtained by normalizing the invisible columns and columns of the judgment matrix. Each element in the vector can be regarded as the relative weight of each index in the index layer\[^{10}\].

Divide the elements of each column of the judgment matrix by the sum of the columns:

\[
\bar{R}_{ij} = \frac{R_{ij}}{\sum_{j=1}^{n} R_{ij}}
\]

Then, the sum of the row vectors of the obtained matrix is obtained:

\[
\bar{\phi} = \sum_{j=1}^{n} \bar{R}_{ij}
\]

Finally, the matrix standardization of the sum of the row vectors is obtained:

\[
\bar{\phi} = \frac{\phi_j}{\sum_{j=1}^{n} \phi_j}
\]

Each element in the resulting vector can then be viewed as the weight of the element at that level.
(3) The reliability of the data cannot be determined after finding the full score of each element, and the reliability of the data needs to be checked. After the results are obtained in the analytic hierarchy process, the data must also be checked for consistency. The steps are as follows: First calculate the maximum feature root $\mu_{\text{max}}$ of the judgment matrix:

$$\mu_{\text{max}} = \frac{\sum_{i=1}^{n} \lambda_i \cdot \phi_i}{n}$$  (4)

Calculate matrix consistency index $\theta$:

$$\theta = \frac{\mu_{\text{max}} - n}{n - 1}$$  (5)

The number of indicators in this layer, that is, according to according to the matrix medium number $N$, find the corresponding average random consistency indicator $\pi$ from Table 2:

| Intermediate n | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 |
|----------------|----|----|----|----|----|----|----|----|----|----|----|
| $\pi$          | 0.56 | 0.89 | 1.06 | 1.17 | 1.28 | 1.36 | 1.49 | 1.57 | 1.63 | 1.69 | 1.78 |

Finally, the consistency index $\psi$ is calculated by the following formula:

$$\psi = \frac{\theta}{\pi}$$  (6)

When $\psi$ is less than 0.1, it can be considered that the judgment matrix satisfies the consistency requirement early enough. When the index number is 1 or 2, the default value of $\pi$ is 0, and the judgment matrix can be regarded as completely consistent. If the results are still unsatisfactory after correction, you can also directly set the weight value determined by the expert. Weights can also be used directly if the selected indicator has sufficient conclusions from previous studies.

3. Experimental Analysis

This study uses a system dynamics model to conduct a comprehensive evaluation and analysis of power marketing channels. Because the system dynamics model is based on state, it can be used for static evaluation of power marketing channels at specific times. Dynamic analysis establishes the weight value of each indicator according to the analytic hierarchy process, and then uses the SD model to calculate the comprehensive evaluation big data of each power marketing channel based on the index value of each power marketing channel obtained from multiple sources (such as power marketing), The higher the score, the higher the level of power marketing.

3.1 Dynamic evaluation of time series of power marketing channels

First of all, the dynamic evaluation and analysis of power marketing channels. Monthly analysis within one year, and trend analysis over many years can be performed. Based on the system dynamics flow chart model constructed in this study, the comprehensive evaluation results of the bank channels can be obtained through the special simulation software for system dynamics, Vensim PLE. This section makes a comprehensive evaluation of power marketing channels. First, the compilation and processing of indicator data is combined with the SD model of comprehensive evaluation of power marketing channels. For comparative analysis, Figure 2 lists the monthly comprehensive evaluation score curve.
As far as the power marketing channel is concerned, the service capability score remains stable throughout the year because it is mainly determined by indicators such as service time and number of service outlets. These values have not changed within a year; the economic benefit score shows a downward trend, because it mainly depends on Changes in power marketing; service quality scores fluctuate less throughout the year, because the two indicators of power marketing success rate and customer satisfaction are lower than in other months, and January and July are relatively low; development potential scores fluctuate throughout the year Yes, this is because the monthly channel marketing volume is obviously different, so the indicators of this layer vary greatly from month to month.

4. Conclusion

Based on the analysis of the marketing process of the power company, based on the dimensions of the index evaluation model, combined with the characteristics of the marketing channel and the operating mode, a comprehensive evaluation index system suitable for the marketing channel is constructed, and the analytic hierarchy process is used to determine all marketing channel evaluation systems. The weight of the indicator and the source of each indicator data are clarified, and a standardized method based on the indicator data is proposed. The evaluation system has strong operability and practical application.

References

[1] Liu Binghan, Fu Zhongguang, Wang Yongzhi, Et al. Big Data Mining Technology based on Parallel Computing and its Application in Power Plant Performance Optimization [J]. Journal of Power Engineering, 2018, 38(6): 431-439.

[2] Hao Yanni, Wu Suping, Tian Weili. Research on data mining algorithm in wine Information Data Analysis System [J]. Computer Science, 2017, 44(1): 491-494.

[3] Yan Lei, Qi Bing. Research on Big Data Mining Technology of Mobile Learning System Based on Android Platform [J]. Modern Electronic Technology, 2017, 40(19): 142-144, 149.

[4] Xing Xue, Yu Dexin, Tian Xiujuan, et al. Analysis of multi-state traffic flow time series characteristics combined with viewability[J]. Acta Phys. Sin, 2017, 66(23): 56-65.

[5] Pan Zhang, Ding Lingyun, Jiang Ning, Et al. Data mining algorithm and application of distribution Automation System based on supportable confidence-lift degree[J]. Electric Measurement and Instrumentation, 2019, 56(10): 62-68.

[6] Li Junwei, Tang Yafang, Hao Zhenghang, et al. Overview of Cluster Analysis and its Application in Power Systems [J]. Modern Electric Power, 2019, 36(3): 1-10.

[7] Li Tianhua, Yuan Yongbo, Zhang Mingyuan. Vulnerability Analysis of Power Grid Nodes based on Variable Fuzzy Clustering [J]. Science and Technology and Engineering, 2018, 18(18): 126-130.
[8] Lin Junhao, Zhang Yan, Zhu Jinzhou, et al. Distribution Network Topology Similarity Analysis Method Based on macro-micro features Hierarchical clustering[J]. Power Service Automation, 2019,43 (13) : 80-88.

[9] Zhao Shuqiang, Zhang Tingting, Li Zhiwei, et al. Error Distribution Model of the Forecast of exporting based on Numerical clustering[J]. Power Service Automation, 2019.43 (13): 155-164.

[10] Zhang Jianglin, Zhang Yachao, Hong Juhua, et al. Load algorithm Based on Discrete K-MODES [J]. Electric Power Automation Equipment, 2019, 39(2): 100-106.