Analysis of Hierarchical and Time-phased Model of Large-scale Power Grid Based on Fp-growth Algorithm

Tianyue Wang¹, Jinxiu Hou² and Zhihong Yu²

¹ North China Electric Power University, Beijing, China
² China Electric Power Research Institute, Beijing, China
E-mail: weety_W@163.com

Abstract. The scale of UHV AC/DC hybrid power grids in China is continuously expanding, and the operational characteristics of power grids have undergone profound changes. With the advancement of measurement technology and communication technology, power data has grown rapidly. Therefore, data mining technology is required to provide safety warnings and risk prevention, and intelligently adjust control strategies to help dispatchers correct the operating status of the power grid. This paper adopts association rules analysis based on fp-growth to construct a hierarchical and time-phased analysis model of power grid, giving association rules for different security and stability levels of the power grid at different time phases. Based on actual historical data of the large-scale power grid and the application of hierarchical and time-phased analysis mode, the association rules of overload conditions of different time periods are extracted.

1. Introduction

At present, long-distance, high-power transmission, extensive access to new energy, and increased interaction between flexible loads and the grid have increased the difficulty of dispatch control. Faced with the complex and ever-changing power grid situation, it is difficult to meet the requirements of high-level security and stability of the power grid, relying on traditional power grid analysis and control technologies. The increasingly developed power grid technology, combined with big data mining analysis and artificial intelligence technology, has become the main trend of future power grid development [1].

Dispatching centres at various levels have accumulated a large amount of dispatching operation data. Based on the Synchronous Vector Measurement Unit (PMU), the sampling rate is 100 times per second, which provides uninterrupted power system synchronization operation data. On the dispatching table, an on-line security and stability analysis is performed every 15 minutes, and the data amount and the result data amount can reach about 1G, and the data amount exceeds 2T in a month, and the data volume is astonishing [2]. Therefore, data mining technology is required to mine knowledge related to the safety and stability of power grids from massive amounts of power data, enhancing the ability of dispatching departments to control the power grid.

The core of big data analysis is data mining algorithms, including decision trees, artificial neural networks, and support vector machines, etc. The analysis of the safety and stability of power systems is mainly through two methods: predicting the system's disturbed trajectory or giving safe and stable identification information [3]-[5].

The method of predicting the disturbed trajectory of the system is based entirely on numerical sequence analysis and data mining, and does not rely on prior knowledge of power system mathematical models and parameters. The literature [6] proposes a method for predicting wide-area...
time and space big data, and discovers the time-space relationship and variation law of power grids from big data, so as to predict the future trajectory and weak links of the power grid, and effectively predict and prevent future stability of the power grid. This kind of method only uses response information, the principle is simple and easy to implement.

Another method is to discriminate the safety and stability of the power grid directly, find the mapping relationship between data characteristics and system security and stability from the power big data sample set, and construct a classifier to predict the safety and stability of the system online without human experience. The literature [7] applied cluster analysis to predict the safety and stability of the system online without human experience. The literature [8] introduces the time series related association rule mining algorithm into the transient stability assessment. The obtained correlation result can provide decision-making basis for the staff. The literature [9] makes use of the characteristics of rapid learning stability of the extreme learning machine to quickly judge the transient stability results. The improved algorithm can continuously update the rolling data to adapt to the dynamic process of the power system.

2. Association rule analysis

2.1. Associative classification principle

Association rule is one of the important machine learning methods. It is mainly used to explore frequent patterns, association dependencies, or causal structures among project collections, and discover valuable association rules based on the frequency of occurrences among data item sets [10]. Let dataset $D$ contain $N$ samples and $Y$ is a collection of categorical variables for the samples in $D$. Each sample $d$ in $D$ is represented by a set of characteristic variables $x$ and a category variable $y$. Let $I$ be the set of all the feature variables in data set $D$. A class association rule is as follows

$$x \Rightarrow y$$

(1)

In the formula, $x$ represents a set of characteristic variables, $x \subseteq I$, $y \in Y$, $x$ and $y$ are disjoint item sets.

There are two important metrics for measuring an association rule - support and confidence. $\text{support}(X \Rightarrow Y)$ is the proportion of transaction item sets containing $[X \cup Y]$ item set in the transaction database $D$, defined as

$$\text{support}(X \Rightarrow Y) = \frac{|X \cup Y|}{|D|}$$

(2)

Support determines the frequency of occurrence of item set $|X \cup Y|$, $|X \cup Y|$ represents the number of transactions contained $|X \cup Y|$ in the data set, $|D|$ indicates the total number of transactions in the data set.

Confidence is the percentage of under the condition that data set $D$ contains item set $X$, it also contains item set $Y$, defined as $\text{conf}(X \Rightarrow Y)$

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

(3)

The role of confidence is to describe the degree of trust of the rules, and support describes how often the rules appear. Mining association rules is to find the association rules that meet the user's minimum support and minimum confidence threshold from all transaction data sets. The basic process of association rule analysis is shown in Figure 1.

It is divided into the following two steps [11]:

Step one: Find the frequent item set in the data set $D$. That is, the support degree is calculated according to the formula (2) and according to the minsupport given by the user, the frequent item sets which support degree is not less than the minimum support degree are selected.
Step 2: Generate association rules using frequent item sets. That is, from the frequent item set obtained in step 1, the rule satisfying the minimum confidence is found according to the formula (3).

![Diagram of association rules process]

**Figure 1.** Basic model of association rules.

### 2.2. FP-Growth algorithm

Mining frequent item sets is an important and time-consuming step in the association analysis process [12]. As one of the most classic algorithms, fp-growth can avoid the generation of a large number of candidate sets and only need to scan the database twice to quickly find frequent item sets. To facilitate efficient processing, we use a new data structure called FP-tree. It can be divided into two steps: FP-tree construction and frequent pattern mining based on FP-tree. The process is shown in the figure 2.

![Diagram of FP-Growth algorithm process]

**Figure 2.** The process of FP-Growth algorithm.

The general idea is to find the frequent single items and then we partition the database based on each such item. Then we recursively grow frequent patterns by doing the above iteratively or recursively for each partitioned database, also called conditional database. The whole mining can be summarized as follows. We recursively construct and mine conditional FP-trees. Until the result FP-tree is empty or until it contains only one path, the single path will generate all the combinations of its sub-paths each
of which is a frequent pattern. Pseudo code of frequent pattern mining algorithm based on FP-tree such as Table 1.

Table 1. Pseudo code of extracting frequent item sets.

| Input: constructed FP-tree |
|---------------------------|
| Output: Full set of frequent patterns |
| Let the set of items in the header table be $I$ |
| for item $\alpha_i$ in $I$ |
| FP-Growth(FP-tree, $\alpha_i$) |
| End |
| FP-Growth(FP-tree, $\alpha_i$) |
| if Tree contains a single Path then |
| for the combination of each node in Path (indicated as $\beta$) |
| generate the mode $\beta \cup \alpha_i$, support_count equals to |
| the minimum support value of each node in $\beta$ |
| else for $\alpha_i$ in the header table of Tree |
| { |
| generate the mode $\beta = \alpha_i \cup \alpha_i$, support_count=$\alpha_i$.support_count |
| construct the conditional pattern base of $\beta$, and conditional FP-Tree$_\beta$ |
| if Tree$_\beta \neq \emptyset$ then |
| FP-Growth(FP-tree, $\alpha_i$) |
| } |

3. Association analysis model

The smart grid dispatching technical support system utilizes a global position system (GPS) and has synchronous timing capabilities, realizing a wide-range space-time unified measurement of large-scale power grids. Therefore, the accumulation of wide-area time-space measurement information in the smart grid scheduling technical support system can accumulate over a long period of time (days, months, and years), which can better reflect the spatial-temporal correlation characteristics of the power grid and have a very large value of excavation and utilization[13]. This paper makes use of the spatio-temporal characteristics of data to build a hierarchical and time--phased model for association analysis. It is aimed at a variety of grid security and stability issues and refines the association rules.

3.1. Hierarchical association analysis

At present, China has formed a hierarchical dispatching operation model of national dispatch, regional dispatch, provincial dispatch, city dispatch, and county dispatch. Due to the division of the management scope and responsibilities of the dispatch, generally according to the geographical location and voltage level, and according to the characteristics of the administrative region and the power system, the scheduling functions and priorities at different levels are different. National dispatch focuses on UHV AC/DC transmission power and emergency support capabilities. The regional dispatch focus on more concerned with the critical liaison sections of the regional power grid, and the provincial dispatch directly controls the output of key power plants in the region and controls the load [14].

At the same time, the security and stability of the power grid also presents typical regional characteristics and spatial correlation. According to different levels of grid security and stability issues, a hierarchical correlation analysis model is established as figure 3.
The hierarchical association analysis model is established through experience and historical data, key factors affecting the stability of different levels of the power grid are obtained, and the combined input feature space is selectively optimized.

3.2. Time-phased association analysis

The power system is a continuous time-varying system, and the power operation data has time correlation, which is reflected in the following two aspects [15]-[17]:

1) The change of power grid operation mode is closely related to time, and the output curve and load curve all show some regularity with time. In the actual operation of the power system, the time-varying curves of a large amount of measurement and calculations contain the inherent operating characteristics of the power system.

2) The large amount of operational data collected by the dispatch centre, as well as the calculated data and the resulting data, all have time scales. These data are the response of various dynamic indicators of the power system at different times and have the characteristics of time series.

When using the association classification method to model the security and stability analysis, we should fully consider the characteristics of its time series. In general, it is difficult to analyse the entire time series in data mining. This paper proposes a time-phased association analysis model, which divides the system state data into local sub-segments, establishes multiple feature input spaces, and obtains association rules at different time periods.

4. Case study and analysis

Based on the online historical data accumulated by a provincial dispatching centre for 4 months, an association analysis model for the heavy load situation of the provincial power grid is established. A total of 6500 samples were collected and organized. The transmission power of important tie lines was extracted from the online data as input features of the model, including two 750kV sections, seven 330kV sections and six 220kV sections. The key sections are usually selected by experts through long-
term operating experience. They are high-voltage grade and large-capacity transmission lines or line sets, which are important security features of the power grid [18].

In this paper, according to the dispatching centre in the calculation of grid operation mode, the line stability limit value is given to evaluate the risk degree of the whole network section. The limit values of the critical sections are generally obtained based on off-line analysis and simulation calculations of the power grid to achieve ‘dimension reduction’ control of complex power systems [19]. According to the ratio of the section power to the stability limit value, determine the risk degree of the safe and stable operation of the grid, and set the evaluation indicators as shown in the Table 2.

**Table 2. Evaluation indicators.**

| Section power/limit value | Stable risk level |
|---------------------------|-------------------|
| 0~50%                     | 0                 |
| 50%~70%                   | 1                 |
| 70%~90%                   | 2                 |
| 90%~110%                  | 3                 |
| 110%~130%                 | 4                 |
| >130%                     | 5                 |

**Table 3. Partial association rules.**

| Time          | Rule 1. Support | Confidence | Rule 2. Support | Confidence |
|---------------|-----------------|------------|-----------------|------------|
| 23:00~1:15    | Support is 20.69% | 87.5%   | Section CS [-1400, -1200) MW → Stable risk level [16,18] |          |
| 1:30~4:45     | Rule 2. Support is 13.44% | 87.5%   | Section DQ [500,600) MW and Section HQ [-400,-200) MW → Stable risk level [16,18] |          |
| 5:00~8:30     | Rule 1. Support is 10.04% | 83.33%  | Section HZ [-400,-200)MW → Stable risk level [16,18] |          |
| 8:45~18:15    | Rule 2. Support is 18.38% | 90.63%  | Section DQ [500,700) MW and Section HQ [-400,-300) MW → Stable risk level [16,18] |          |
| 18:30~22:45   | Rule 1. Support is 16.60% | 81.06%  | Section HZ [-300,300)MW → Stable risk level [16,18] |          |

When the power flow is approaching or even beyond the limits of safe and stable operation, once the disconnection or other serious failure occurs, there may be grid safety and stability problems. According to the statistical analysis of the load situation of the power grid, the entire time series sample is divided into five time phases, and the association rules are mined separately.
In order to ensure the reliability of the extracted association rules, we consider setting the minimum confidence threshold to 80%. In the actual operation of the power grid, severe overload occurs less frequently. If the minimum support threshold is set too high, an effective overload rule cannot be obtained. Therefore, the minimum support threshold is set to 10%. According to the evaluation rules of the table, the overall degree of danger of the grid of each sample is calculated. The degree of danger [16]-[18] is a situation where the grid is overloaded and the safety risk is high. It needs the attention of the dispatcher. Mining partial association rules as shown in Table 3.

From the analysis of the grid structure and the distribution of the power grid in the province, it can be seen that the section CS is an important current collection channel. The new energy generated in the vicinity of this area is all sent through it and then sent through the DC line in a large scale. section CS has long been in heavy-duty operation. From the generated rules, we can also see that in the period of 8:45~18:15, when the new energy output is large, section CS [-1700, -1400) MW → the degree of danger of the power grid [16], [18], confidence degree is 92.8%. Through the above rules, the dispatcher is informed of the sections that require special attention and the dangerous section of the section at different time periods.

It can be seen that the generated rules are in line with the operating rules of the power grid and the operational experience accumulated by the operating personnel for a long time. At the same time, it can also help the dispatching operators to find some unknown and potential operating laws of the power grid and provide auxiliary decision-making.

5. Conclusion

The rapid accumulation of power data and the rapid development of computer and big data technologies have provided new directions for the development of power grids. In this paper, the method of correlation analysis is applied to the analysis of power grid security and stability, and the hierarchical and time-phased correlation analysis model is established. Relying on the actual historical data of large-scale power grids, the correlation analysis between the operating status of key sections of the power grid and the heavy-load overload conditions of the power grid is conducted, and the association rules under different time phases are obtained to demonstrate the feasibility of the model. With the rapid development of the power grid in the future, this method can provide assistant decision-making for dispatchers, to accelerate understanding of the power grid, and increase operational experience.

6. References

[1] LIN Weifang, TANG Yong, The “2·4” blackout in Brazil and its implications for the safe and stable operation of power grids [J]. Automation of Electric Power Systems, 2011, 35(09): 1- 5. [2017-09-12].
[2] Yu Zhihong, Huang Yanhao, et al. Extraction method of stable operation rules based on time series correlation analysis [J/OL]. Proceedings of the Chinese Society of Electrical Engineering, 2015, 35(03):519-526.
[3] Huang Ruoyan. On-Line Power System Transient Stability Online Identification and Algorithm Based on Energy Function Method [D]. Huazhong University of Science and Technology, 2013.
[4] Wu Wei. Research on real-time discrimination and control technology of power system transient stability based on response [D]. China Electric Power Research Institute, 2014.
[5] Zhang Chun, Research on Power System Transient Stability Assessment Method Based on Association Rules [D]. North China Electric Power University, 2017.
[6] Liu Daowei, Zhang Dongxia. Quantitative assessment of the stability situation of large-scale power grids under the space-time big data environment and the construction of adaptive prevention and control system [J]. Chinese Journal of Electrical Engineering, 2015, (02): 268-276.
[7] Jiang Weiying, Research on online comprehensive security early warning and emergency dispatch decision of power grid [D]. Beijing: Tsinghua University, 2008
[8] Kamwa, S. Samantaray and G. Joos, On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors, 2013 IEEE Power & Energy Society General Meeting, Vancouver, BC, 2013, pp. 1-1.
[9] Huang Yanhao, Yu Zhihong. Fast determination strategy for large power grid based on massive online historical data [J]. Proceedings of the CSEE, 2016, 36(03): 596-603.
[10] Huang Tianen, Sun Hongbin. Knowledge management and advanced security early warning based on power grid operation simulation big data [J]. Power System Technology, 2015, 11:3080-3087.
[11] Huang Tianen, Sun Hongbin. Online distributed security feature selection based on power grid operation big data [J]. Automation of Electric Power System, 2016, 04:32-40.
[12] Zhou Yanzhen, Power System Transient Stability Analysis and Preventive Control Based on Data Mining Technology [D]. Beijing: Beijing Jiaotong University, 2017
[13] Yu Z H, Huang Y H, Xie C, et al. Operation rule extracting based on time series association analysis in transient stability study: International Conference on Power System Technology, 2014[C].
[14] I. Kamwa, S. R. Samantaray and G. Joos, Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices, in IEEE Transactions on Smart Grid, vol. 1, no. 2, pp. 144-158, Sept. 2010.
[15] Chi Lining. Classification algorithm mining algorithm based on FP-Growth and its application [D]. Qingdao University, 2012.
[16] He M, Vittal V, Zhang J. Online dynamic security assessment with missing pmu measurements: A data mining approach [J]. IEEE Transactions on Power Systems, 2013, 28 (2):1969-1977.
[17] Sun Hongbin, Wang Kang, Zhang Boming, et al. Transient Stability Rule Extraction Using Linear Decision Tree [J]. Proceedings of the CSEE, 2011(34): 61-67.
[18] Wang B, Fang B, Wang Y, et al. Power System Transient Stability Assessment Based on Big Data and the Core Vector Machine [J]. IEEE Transactions on Smart Grid, 2016:1
[19] Yan Jianfeng, Yu Zhihong, Tian Fang, et al. On-Line Dynamic Safety Assessment and Early Warning System of Power System [J]. Proceedings of the CSEE, 2008(34):87-93.