Voltage sag severity analysis based on improved FP-Growth algorithm and AHP algorithm

Yan Lin*, Fang Lin, Daoshan Huang, Xiaolin Fang
State Grid Fujian Electric Power Research Institute, Fujian, China

*Corresponding author e-mail: lyepri@163.com

Abstract. In order to accurately evaluate the impact of voltage sag on both utility and customer, a data mining analysis method based on improved FP-growth algorithm and AHP algorithm is proposed. The proposed method selects and classifies voltage sag characteristics and combines voltage sag severity level to construct a data mining analysis framework. The improved FP-growth algorithm is used for data mining on voltage sag events to improve the efficiency of mining. The association rule matching model is constructed by AHP algorithm for evaluating the voltage sag severity level, which improves the accuracy of the evaluation results. Finally, a practical example verifies the practicability of the proposed method.

1. Introduction
Due to the continuous increase of sensitive equipment in the power network, the problem of voltage sag is becoming more and more prominent [1]. To evaluate the effect of voltage sag on equipment and customers is of great significance in many aspects for sensitive customers, especially in economic terms [2,3].

All methods that have been exposed for voltage sag severity evaluation methods are divided into two categories: random estimation method [4,9] and measured statistical method [10,14]. The former includes critical distance method [4,5], fault point method [6,7] and analytical method [8,9]. The critical distance method [4,5] can predict the voltage sag amplitude, but it is only applicable to the symmetric three-phase short circuit situation. The fault point method [6,7] is applicable to all kinds of fault operations, but requires extensive simulation operations, which is not applicable when the network structure is large. Analytical method [8,9] calculates the voltage sag amplitude by constructing impedance matrix, but the model is greatly different from the actual situation.

Based on actual measurement data, the measured statistical method can accurately reflect the voltage sag severity, which can be regarded as three categories: state estimation method [10], comprehensive assessment method [11,12] and data mining analysis method [13,14].

The state estimation method [10] can accurately estimate the voltage sag state of the whole network, but the mathematical model is complex, so it is difficult to find the optimal solution. The comprehensive assessment method [11,12] combines the weights of multiple indicators to evaluate the impact. However, the homogeneity among indicators is ignored and the indicator framework is difficult to consider comprehensively. The voltage sag data mining method [13,14] can evaluate the voltage sag severity by mining the association rules between the characteristic attributes of voltage sag and the voltage sag severity. Reference [13] used AprioriTid algorithm to mine voltage sag data, and the grey target theory was used to match evaluation results of severity of voltage sag. On this basis, reference [14] proposed a data mining and analysis method for voltage sag severity based on mutual
information and improved grey target theory. However, references [13,14] all adopt data mining methods based on traditional Apriori. Apriori algorithm loads more candidate sets before scanning the database, which leads to lower efficiency.

Aiming at this gap, a data mining analysis method in the basis of improved FP-Growth algorithm and AHP algorithm has been proposed in this paper. Firstly, the voltage sag characteristic attributes are selected and classified, and a data mining analysis framework is constructed by combining the severity indexes. Then, the improved FP-Growth algorithm is used to mine the data of voltage sag events, which avoids the problem of invalid rules generated by the traditional FP-Growth algorithm and improves the process during data mining. An association rule matching model is constructed by AHP algorithm to evaluate voltage sag severity and improve the accuracy of the results. Finally, performance and effectiveness of the proposed method are verified through the field case.

2. Data Mining Preparation

2.1. Voltage sag characteristic attribute

The power quality monitoring system collects a lot of data reflecting voltage sag events, and eventually forms a giant database of various information over time. In this database, certain principles must be followed to screen out factors that correspond to the severity of voltage sag, so as to improve mining efficiency and avoid mining "meaningless" association rules. In this paper, six-dimension factors are selected to participate in association rule mining and node voltage sag severity as shown in Table 1. Of course, users can also select other voltage sag characteristic attributes according to their own requirements.

| Characteristic attributes | Attributes description |
|---------------------------|------------------------|
| Observation position      | The geographical location of power quality monitoring device |
| Voltage grade             | 500kV, 220kV, 110kV, 35kV, 10kV, |
| Weather                   | Heavy rain, light rain, overcast, thundershower, cloudy, others |
|                           | Single phase grounding, two phase grounding, |
| Cause of fault            | Two phase short circuit, three phase short circuit, induction motor start, transformer switching |
| Season                    | Spring, summer, autumn, winter |
| Time                      | Early morning, morning, forenoon, noon, afternoon, evening, midnight, late night |

In Table 1, "Observation position" refers to the geographical location of the substation where the concerned bus is located; "Voltage grade" means rated voltage; "Weather" refers to the weather condition at the corresponding substation; "Cause of fault" refers to the cause of voltage sags; "Season" means the season when voltage sags occur; "Time" is the time segment of one day when the voltage sag occurs.

For mining association rules, it is necessary to convert voltage sag characteristic attribute into qualitative quantity. In Table 1, "voltage grade", "weather", "cause of fault" are qualitative descriptive data. "Season" and "time" can be qualitatively divided according to social habits.

"Observation location" adopts k-means clustering method to discrete six qualitative quantities, namely northeast, southeast, central, northwest, west and southwest.

2.2. Voltage sag characteristic attribute

To measure the impact on sensitive consumers caused by voltage sag, it is indispensable to consider the severity of voltage sag and characteristics of sensitive consumers. This kind of severity should consider the voltage tolerance capability on sensitive equipment. A specific definition of voltage sag
severity index $S_e$ is as follows [15]:

$$S_e = \frac{1-U}{1-U_{\text{curve}}(t)}$$  \hspace{1cm} (1)

Where, $U$ is the voltage amplitude per-unit value; $t$ is the duration of voltage sag; $U_{\text{curve}}(t)$ refers to the voltage amplitude on the tolerance curve.

The voltage sag could be divided into 5 grades based on severity: A($0 \leq S_e < 1$), B($1 \leq S_e < 2$), C($2 \leq S_e < 3$), D($3 \leq S_e < 4$), E($S_e \geq 4$).

3. Association Rule Mining

3.1. Overview of Association rules

Let $I = \{i_1, i_2, \cdots, i_m\}$ be a set of $m$ different items. For database $D$, each transaction $T$ is a set of items in $I$, that is, $T \in I$ [16]. An association rule has the form: $X \rightarrow Y$, which $X \in I$, $Y \in I$ and $X \cap Y = \emptyset$.

Two indicators of support $(S(X \rightarrow Y))$ and confidence $(C(X \rightarrow Y))$ are usually used to mine association rules. The calculation formula is:

$$S(X \rightarrow Y) = \frac{\| \{t \in D | X \cup Y \in t\} \|}{|D|}$$  \hspace{1cm} (2)

$$C(X \rightarrow Y) = \frac{P(Y/X)}{P(X)}$$  \hspace{1cm} (3)

Where, $\| \|$ represents the number of sets, and Eq. (2) represents the proportion of transaction $t$ that contains both $X$ and $Y$ in database $D$; Eq. (3) represents the ratio of transactions containing $X$ and $Y$ to transactions containing $X$.

3.2. Improved FP-Growth algorithm

FP-Growth algorithm is an association rule based on data mining algorithm proposed by Jiawei Han in 2000 [17]. This algorithm transforms the itemset of a given database into a Frequent Pattern Tree (FP-Tree), and keeps the correlation information among the itemset in the FP-tree. Because of the tree-like structure of FP-tree, branches with the same prefix can be reused, which can realize data compression and improve data mining efficiency. Compared with the traditional association rule mining Apriori algorithm, this algorithm has the following characteristics: No candidate set is generated in the mining process and only needs to traverse the database twice [18].

The traditional FP-Growth algorithm via mining the association rules that between any attributes and has no restrictions on the conditions and results of the association rules. However, when association rules are used to assessing voltage sag severity, the result of association rules must be the severity index about voltage sag, and the condition is the characteristic attribute of voltage sag. Therefore, a large number of invalid rules will be generated when traditional FP-Growth algorithm is used to mine characteristic attributes and severity indexes. Therefore, this paper proposes an improved FP-Growth algorithm to avoid this situation. The four steps of this algorithm are as follows:

Step A: First, scan the entire database to get and count the frequency for each characteristic attribute item. Then deleting the items that do not meet the minimum support degree, and finally sort the obtained feature attribute items in descending order according to the occurrence frequency to obtain the frequent item sets table. Take the database in Table 2 as an example, where the items a, b, c, d, e and f are characteristic attribute items, and the items g and h are severity index items, all characteristic attribute items were scanned and the occurrence frequency was calculated, among which
item a is appeared 6 times, items c and b are appeared 5 times, item e is appeared 4 times, item d is appeared 3 times, and item f is appeared 1 time. Set the minimum support as 0.25, which means that the item appears at least twice.

Table 2. Tid table of database.

| Tid | Itemsets  |
|-----|-----------|
| 1   | b, c, d, g|
| 2   | e, a, d, g|
| 3   | e, a, f, g|
| 4   | a, d, c, h|
| 5   | b, c, a, h|
| 6   | e, a, d, g|
| 7   | c, b, a, h|
| 8   | c, e, b, g|

Step B: Put the obtained characteristic attribute items that meet the minimum support into the frequent itemsets table as shown in Table 3; Reorder the database in Table 2 from large to small according to the times of each characteristic attribute items appears, and the influence index is ranked to the end. The rearranged database is shown in Table 4.

Table 3. Frequent Itemsets table.

| Tid | Itemsets |
|-----|----------|
| a   | 6        |
| c   | 5        |
| b   | 5        |
| e   | 4        |
| d   | 3        |

Table 4. The restructured Tid table of database.

| Tid | Itemsets  |
|-----|-----------|
| 1   | c, b, d, g|
| 2   | a, e, d, g|
| 3   | a, c, f, g|
| 4   | a, c, b, h|
| 5   | a, c, b, g|
| 6   | a, e, d, g|
| 7   | a, c, b, h|
| 8   | c, b, e, g|

Step C: Scan the database again and create the root node of the FP-Tree with "Null". Every transaction obtained by scanning again is inserted into the FP-Tree according to table 3 to create a path. If the same item appears in the process of building FP-Tree, add one to the node number of the related item. According to the table 4, add 8 item sets to the FP-tree to get the tree diagram as shown in Figure 1.

Step D: Mining FP-tree. For each branch shown in Figure 1, count the number of tail node severity index. If the number of index items meets the minimum support, the path ending with this the severity index item node are output as the candidate ruleset. This rule is supported by the ratio of the number of severity index item to the total number of events, and the confidence of the rule is the ratio of the number of severity indicators to the number of characteristic attributes of the previous node.
For example, the number of the severity index item g of the branch 3 is 2, which meets the minimum support degree, and the confidence degree is 1. Then a, e, d, g is output to the candidate ruleset. The candidate ruleset obtained according to Figure 1 is shown in Table 5.

### Table 5. Candidate rulesets.

| Serial number | Frequent itemsets | Number of items | Support | Confidence |
|---------------|-------------------|-----------------|---------|------------|
| 1             | a, e, d, g        | 2               | 0.25    | 1          |
| 2             | a, c, b, h        | 2               | 0.25    | 0.667      |

### 4. Association Rule Matching Model

If the mined association rules are to be used to guide the actual production, they also need to be matched with the actual fault scenario. By establishing an appropriate matching model, similar results can be output if they are not completely the same as the actual scenes.

In this paper, AHP algorithm is used to construct the matching model of association rules. AHP algorithm [19] is a multi-objective decision making method that combines qualitative and quantitative algorithm. The steps of building an association rule matching model using AHP algorithm are as follows:

**Step1:** According to the actual scenario and association rule matching model, the maximum matching degree is determined as the target layer, the membership degree of each dimension of voltage sag characteristic attribute is determined as the criterion layer, and the mined association rule library is the scheme layer.

**Step2:** Determine the judgment matrix according to the relationship between the indicators of the indicator layer. The judgment matrix $A$ is defined by reference [19]. The definition of the elements in the judgment matrix $A$ is as follows:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$  \hspace{1cm} (4)

Where, $n$ is the number of indicators.

**Step3:** After the maximum eigenvalue $\lambda_{\text{max}}$ of the judgment matrix and its corresponding eigenvector $\xi$ are calculated, normalizing the eigenvector to obtain the weight matrix $W$.

$$W = \begin{bmatrix} w_1 \\ \vdots \\ w_j \\ \vdots \\ w_n \end{bmatrix}$$  \hspace{1cm} (5)

Where, $w_j$ is the weight of the $i$-th index.

**Step4:** Calculate the consistency index and consistency ratio of the judgment matrix, and verify whether it meets the requirements [18]. Otherwise the judgment matrix should be appropriately modified.

**Step5:** Calculate the membership degree of each dimension of the voltage sag characteristic attribute of the association rule, and calculate the matching degree between the association rule and the actual scenario through the weight matrix $W$. The association rules are sorted according to the matching degree.

$$P(k) = \sum_{i=1}^{n} w_i \mu_{ki}$$  \hspace{1cm} (6)

Where $\mu_{ki}$ is the membership degree of the $i$-th characteristic attribute of the $k$-th association rule.
5. Case Analysis

5.1. Validation of Association Rules

Take the power quality monitoring records of a power company as the original data, and extract the characteristic attributes of voltage sags as the historical database for analyzing.

The K-MEANS clustering method is adopted to cluster the latitude and longitude values of the observed position of voltage sags, and six qualitative parts are obtained. The weather information of the voltage sag will be obtained from the weather historical information of the Meteorological Bureau at the time. The fault cause and time information of each sag event are obtained from the monitoring data and classified. Combined with fault event report, monitoring data and power grid topology, the voltage amplitudes of all bus nodes of all events are calculated, duration and sensitive equipment tolerance of known historical sag events, the node voltage sag severity is calculated.

In order to verify the accuracy of association rules, the data from January 2016 to December 2018 are used as training samples to establish a voltage sag severity database; Then the data from January 2019 to December 2019 are used as the test set to match and verify the association rules. When the association rules mined according to the database from 2016 to 2018 still have similar confidence in the sag events in 2019, it shows that the mined association rules can guide users to take corresponding measures to reduce losses.

First of all, data mining is carried out for the training sample set, then selecting 20 association rules from the mined association rules randomly. Referring to Eq. (3), define the confidence of association rules in the test set $C_s(i)$ as shown in Eq.(7), and the define accuracy $\eta$ as shown in Eq.(8):

$$C_s(i) = \frac{n(X_i \rightarrow Y_i)}{m(X_i)}$$  

$$\eta(i) = 1 - \frac{|C(i) - C_s(i)|}{C(i)}$$  

Where, $n(X_i)$ is the number of voltage sag records that meet the $X_i$ in the test scene; $n(X_i \rightarrow Y_i)$ is the number of node voltage sag records that satisfy the association rule $X_i \rightarrow Y_i$, $C(i)$ is the confidence of the association rule itself; $C_s(i)$ is the confidence of the association rule in the test set.

As shown in Figure 2, the confidence level $C$ of the rule itself has the same changing trend as the confidence level $C_s$ of the rule in the test set; The accuracy rate $\eta$ fluctuates around 95%; It shows
that instead of being limited to the mined database, mined association rules and database are universal and can guide users to take corresponding measures to reduce losses.

5.2. Association rule matching model

The VSP of actual scenario is \{southeast, 110.0, winter, morning, overcast, single-phase grounding\} according to Table 1, and match the fault scenario with the mined association rule. When the attribute is the same, the degree of membership Take 1 and take 0 in all other cases. Use the AHP to obtain the weight of the membership degree of each attribute as shown in Table 6.

| Characteristic attributes | Observation location | Voltage grade | Weather | Cause of fault | Season | Time |
|---------------------------|----------------------|---------------|---------|----------------|--------|------|
| Weight                    | 0.106                | 0.415         | 0.208   | 0.135          | 0.084  | 0.052|

The matching degree of the association rule is calculated according to the weight of the membership degree of each attribute in Table 6, and the association rule is output according to the size of the matching degree. The matching results are shown in Table 7.

| Serial number | Association rules                                                                 | Confidence | Matching degree |
|---------------|----------------------------------------------------------------------------------|------------|-----------------|
| 1             | \{southeast, 110.0, winter, morning, overcast, single-phase grounding\} → \{A\}  | 0.906      | 1.0             |
| 2             | \{southeast, 110.0, winter, noon, overcast, single-phase grounding\} → \{A\}       | 0.803      | 0.948           |
| 3             | \{northwest, 110.0, winter, morning, overcast, single-phase grounding\} → \{B\}   | 0.913      | 0.894           |

According to Table 7, the severity of voltage sag in this fault scenario can be obtained. The association rule of serial number 1 has a high degree of confidence and matching. The voltage sag severity level in the association rule shown in serial number 1 could be used to measure the voltage sag severity about this fault scenario. In other word, the voltage sag severity of this fault scenario is "A", indicating that the voltage sag in this fault scenario has little impact on the consumers connected to this node.

6. Conclusion

In order to evaluate the voltage sag severity, this paper proposes a method via data mining analysis that based on FP-growth algorithm and analytic hierarchy process. The proposed method can mine the association rules between the voltage sag characteristic attributes and the voltage sag severity, which can obtain the voltage sag severity in the fault scenario by matching the actual fault scenario with the mined association rules. Moreover, according to the mined association rules, it can also guide sensitive users to choose the right entry point. In the end, the practicality of the proposed method is verified by examples analysis.

7. Acknowledgments

This work was funded by the Science and Technology project of State Grid Fujian Electric Power Co., LTD. Project name: "Research, Development and Application of Voltage Sag Simulation Analysis and Risk Early-warning Decision Aid Platform" (No: 521304190027).

References

[1] A. Luo, Q. Xu, F. Ma and Y. Chen, Overview of power quality analysis and control technology for the smart grid[J]. Journal of Modern Power Systems and Clean Energy, January 2016, vol. 4, no. 1, pp. 1-9.
[2] Xiao X Y , Ma Y Q , Zhang Y , et al. Premium Power Valuation Method Based on Customer Perception of Utility for High Technology Manufacturing Customers[J]. IEEE Transactions on Power Delivery, 2016, 31(4):1-1.
[3] Tan R H G , Ramachandaramurthy V K . Voltage Sag Acceptability Assessment Using Multiple Magnitude-Duration Function[J]. IEEE TRANSACTIONS ON POWER DELIVERY PWRD,
Moschakis M N, Hatzigioryiou N D. Analytical calculation and stochastic assessment of voltage sags[J]. IEEE Transactions on Power Delivery, 2006, 21(3): 1727-1734.

Park C, Jang G. Stochastic Estimation of Voltage Sags in a Large Meshed Network[J]. IEEE Transactions on Power Delivery, 2007, 22(3): 1655-1664.

Zambrano X, Hernandez A, Izzeddine M, et al. Estimation of Voltage Sags From a Limited Set of Monitors in Power Systems[J]. IEEE Transactions on Power Delivery, 2017, 32(2): 656-665.

Chan J Y, Milanovic J V, Delahunty A, et al. Generic Failure-Risk Assessment of Industrial Processes due to Voltage Sags[J]. IEEE Transactions on Power Delivery, 2009, 24(4): 2405-2414.

Liao Y. Fault Location for Single-Circuit Line Based on Bus-Impedance Matrix Utilizing Voltage Measurements[J]. IEEE Transactions on Power Delivery, 2008, 23(2): 609-617.

M. N. Moschakis and N. D. Hatzigioryiou, Analytical calculation and stochastic assessment of voltage sags[J]. in IEEE Transactions on Power Delivery, July 2006, vol. 21, no. 3, pp. 1727-1734.

X. Xiangning, T. Shun, B. Tianshu and X. Yonghai, Study on Distribution Reliability Considering Voltage Sags and Acceptable Indices[J]. in IEEE Transactions on Power Delivery, April 2007, vol. 22, no. 2, pp. 1003-1008.

C. Shen and C. Lu, A Voltage Sag Index Considering Compatibility Between Equipment and Supply[J]. In IEEE Transactions on Power Delivery, April 2007, vol. 22, no. 2, pp. 996-1002.

Zhou X, Wang F, Huang R, et al. Assessment of Voltage Sags in Substations Based on Power System and Equipment Sensitivity Analysis [J]. Proceedings of the CSEE, 2015, 35(08):1940-1946.

Shen X, Yang H, Duan C. Data mining analysis method for voltage sag events based on grey target theory and cloud model [J]. Power grid technology, 2019, 43(02):722-731.

Pu Y, Yang Ho, Ma X. Analysis and evaluation of voltage sag severity based on data mining and improved gray target [J]. Automation of Electric Power Systems, 2020, 44(02): 198-213.

Sabin D D, Bollen M. Overview of IEEE Std 1564-2014 Guide for Voltage Sag Indices [C]. international conference on harmonics and quality of power, 2014: 497-501.

Sheng G, Hou H, Jiang X, et al. A Novel Association Rule Mining Method of Big Data for Power Transformers State Parameters Based on Probabilistic Graph Model [J]. IEEE Transactions on Smart Grid, 2016, 9(2): 695-702.

Han J, Pei J, Yin Y, et al. Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach [J]. Data Mining and Knowledge Discovery, 2004, 8(1): 53-87.

Jiang S, Pan X, Xiao Y, et al. Identification of vulnerable lines in large power grid based on FP-growth algorithm [J]. The Journal of Engineering, 2009, 1(1).

Liang Q, Hu H. Study on Identification of Spurious Public-Private Partnership Projects in China [J]. IEEE Transactions on Engineering Management, 2018: 1-9.