Research on Lean Analysis Algorithm for Equipment Centralized Monitoring in Big Data Era

Tan Chen23, Wang yong1, Lu Xiaomin1, Wu Jiang23, Zhang Guangcheng23, Bai Ziwei23

1State Grid Jiangsu Electric Power Co.,Ltd., Nanjing 210094, China
2Nari Group Corporation/State Grid Electric Power Research Institute, NanJing 211000, China
3BeiJing KeDong Electric Power Control System Co.,Ltd., Beijing 100192 ,China

Abstract. The equipment monitoring brought by the smart grid big data is difficult to effectively supervise the operation status of the whole network substation, and the typical defects (familial defects) are difficult to classify and locate. This paper proposes the substation operation state evaluation algorithm and typical defect classification algorithm. The operating state evaluation algorithm of the substation is based on different operation data generated by the substation. By normalizing the data of different dimensions, the substation is divided into different operating state levels. The typical defect classification algorithm establishes and maintains the historical experience database, and calculates the conditional probability of each information item to realize the correlation between the signal and the defect, and finally judge whether the signal is from a typical defect. These two algorithms are effective means for equipment monitoring professionals to realize intelligent supervision of substations and equipment in the era of big data.

1. Introduction
As the global energy problem becomes more and more serious, the Smart Grid (SG) has become a hot research topic in various countries. The smart grid integrates modern and advanced technology of the Internet of Things, network, sensor measurement, communication, computing, automation and intelligent control into the current physical grid to form a new smart one. It establishes a panoramic perception network that completely covers the power generation, transmission, substation, power distribution, power consumption and dispatching of the power system. Therefore, the data generated by the grid operation and equipment detection grows exponentially and gradually constitutes big data.

In view of the big data generated by the large-scale operation of the power grid, the national power grid equipment monitoring profession urgently needs effective means to monitor the operation of various substations and equipment. At present, is stuck in the supervision of equipment operation data in the substation, and there is no effective means for supervising the overall operation of the substation.

The classification of typical defects or equipment family defects of equipment can only rely on manual work, and fails to be intelligently and automatically done. Also, signals cannot be used to locate typical defects in the device.

This paper proposes the following two algorithms to evaluate the operation of substation incorporated into centralized monitoring and the classification of typical defects of equipment. This can realize the effective supervision of the provincial power station dispatching to the whole network substation, and solve the problem that the signal and equipment are not related to the typical defects. It
provides an effective means for the equipment monitoring professional to realize the intelligent supervision of the substation and equipment in the era of big data.

2. Whole network substation operation state evaluation algorithm

Take the monitoring signals of each substation as the following input information as an example, the substation operation evaluation is carried out, as shown in the following table:

| NUMBER | SIGNAL                          |
|--------|---------------------------------|
| A₁     | Online Monitoring               |
| A₂     | Number of unclear videos        |
| A₃     | Number of images without video  |
| A₄     | Total number of daily alarm signals |
| A₅     | Total number of weekly and monthly warning signals |
| A₆     | Alarm frequency signal statistics |
| A₇     | Bus voltage over time           |
| A₈     | Main transformer line power over time |
| A₉     | Main transformer line current over time |
| A₁₀    | Equipment missing time statistics |
| A₁₁    | Monitoring authority transfer frequency and time |

The substation with the whole network incorporated into the centralized monitoring is the set \(X[x₁, x₂, \ldots, xₘ]\), where \(x₁, x₂, \ldots, xₘ\) represents the specific \(m\) substations. The samples in \(X\) are represented by 11 description attributes \(A₁, A₂, \cdots, A_{11}\), which are the 11 input signals in Table 1. The data sample is \(xᵢ = (xᵢ₁, xᵢ₂, \cdots, xᵢ₁₁)\), where \(i = 1, 2, \cdots, m\) and \(xᵢ₁, xᵢ₂, \cdots, xᵢ₁₁\) is the specific value of the 11 description attributes \(A₁, A₂, \cdots, A_{11}\) corresponding to the substation \(xᵢ\). The data structure can be viewed as the following matrix:

\[
P = \begin{bmatrix}
  x_{1,1} & x_{1,2} & \cdots & x_{1,11} \\
  x_{2,1} & x_{2,2} & \cdots & x_{2,11} \\
  \vdots   & \vdots   & \ddots & \vdots   \\
  x_{m,1} & x_{m,2} & \cdots & x_{m,11}
\end{bmatrix}
\]  

One of the lines represents the 11 attribute values of a particular substation.

2.1. Data formatting

Since large range of values has a higher impact on the evaluation of the substation than the small range, it is not conducive to reflecting the true dissimilarity, so the attribute value of the range is first normalized. Normalization is to scale each attribute value to the same range to balance the impact on substation evaluation. Each attribute is usually mapped to the [0, 1] interval. The specific method is as follows:

\[
x'_{n,a} = \frac{x_{n,a} - \min(x_{i,a})}{\max(x_{i,a}) - \min(x_{i,a})}
\]  

\(a\) is a fixed value and represents a property of the substation, \(i = 1, 2, \cdots, m\).

The data in the equation (1) is normalized according to the equation (2), and the equation (1) is converted into the following matrix:

\[
P' = \begin{bmatrix}
  x'_{1,1} & x'_{1,1} & \cdots & x'_{1,11} \\
  x'_{2,1} & x'_{2,2} & \cdots & x'_{2,11} \\
  \vdots   & \vdots   & \ddots & \vdots   \\
  x'_{m,1} & x'_{m,2} & \cdots & x'_{m,11}
\end{bmatrix}
\]
2.2. Selecting a typical substation as a grouping center

Historical experience recommends that \( n \) typical substations are selected in the whole network according to the operating conditions of the substation, where \( n > 2 \). The operation state is divided into good, medium and poor, and three typical substations are selected, which are:

\[
K_x(x_{x,1}', x_{x,2}', \cdots, x_{x,11}')
\]

(4)

\[
K_y(x_{y,1}', x_{y,2}', \cdots, x_{y,11}')
\]

(5)

\[
K_z(x_{z,1}', x_{z,2}', \cdots, x_{z,11}')
\]

(6)

2.3. Calculating Euclidean distance

Calculate the Euclidean distance from each substation to these three typical substations as follows:

\[
d_{ij} = \sqrt{\sum_{a=1}^{11}(x'_{a,i} - x'_{a,j})^2}
\]

(7)

Where \( i = 1, 2, \cdots, m \), \( j = x, y, z \), and \( i \neq j \). The substation in the whole network is divided into three groups according to the Euclidean distances of the three typical substations of good, medium and poor, and each of them is divided into the group closest to the Euclidean distance of the three typical substations.

2.4. Adjusting the group center

The 11 attribute values of then \( n_x, n_y \) and \( n_z \) substations clustered in \( K_x, K_y \) and \( K_z \) are respectively averaged to locate a new grouping center, as follows:

\[
K_x'(x_{x,1}'+x_{x,1}'+\cdots+x_{x,y}, x_{x,2}'+x_{x,2}'+\cdots+x_{x,y}, \cdots, x_{x,d}'+x_{x,d}'+\cdots+x_{x,y})
\]

(8)

\[
K_y'(x_{y,1}'+x_{y,1}'+\cdots+x_{y,y}, x_{y,2}'+x_{y,2}'+\cdots+x_{y,y}, \cdots, x_{y,d}'+x_{y,d}'+\cdots+x_{y,y})
\]

(9)

\[
K_z'(x_{z,1}'+x_{z,1}'+\cdots+x_{z,y}, x_{z,2}'+x_{z,2}'+\cdots+x_{z,y}, \cdots, x_{z,d}'+x_{z,d}'+\cdots+x_{z,y})
\]

(10)

Where \( u_y, \cdots v_x \) are \( n_x - 1 \) substations clustered at \( K_x' \), \( u_y, \cdots v_y \) are \( n_y - 1 \) substations clustered at \( K_y' \), \( u_z, \cdots v_z \) are \( n_z - 1 \) aggregated at \( K_z' \) Substation. \( K_x', K_y', K_z' \) are new grouping centers and do not represent specific substations.

2.5. Cycling adjustment group center

Cycle 1.3 steps and 1.4 steps until the substation within the packet is not changing or the group center is no longer changing.

Finally, the whole network substation is divided into three levels: good, medium and poor, which is convenient for the control center to supervise the operation of each substation in the whole network.

3. Equipment typical defect classification algorithm

Typical equipment defects (or familial defects) need to be distinguished by manual experience. It is unable to automatically distinguish whether it is a typical defect, nor can it distinguish whether the device corresponding to the signal is a typical defect by means of a signal. Now through the typical defect classification algorithm of the device, the correlation between the experimental signal and the device defect can be realized, and the function of automatically distinguishing the typical defect can also be realized. The specific algorithm is as follows:

3.1. Creating a historical experience database

The signals when typical defects occur are counted according to experience into the historical experience database, and there are \( N \) kinds of information items. The classification result will be more accurate if the \( N \) gets more. The historical experience database is as follows:

| NUMBER | MESSAGE A | MESSAGE B | MESSAGE N | TYPICAL DEFECTS |
|--------|-----------|-----------|-----------|-----------------|
| 1      | A_1       | B_1       | \( \cdots \) | \( N_1 \)        | Positive/Negative |
3.2. Calculating the conditional probability

Calculate the probability of being a typical defect:

\[ P(\text{positive}) = \frac{\text{the quantity of typical defects}}{n} \]  
\[ P(\text{negative}) = \frac{\text{not the quantity of typical defects}}{n} \]

Calculate the probability that the message item in each column in the history database is a typical defect, that is, the conditional probability \( P(A_1|\text{positive}) \), \( P(A_2|\text{positive}) \), ..., \( P(A_p|\text{positive}) \), \( P(B_1|\text{positive}) \), ..., \( P(B_p|\text{positive}) \), ..., \( P(N_1|\text{positive}) \), \( P(N_2|\text{positive}) \), ..., \( P(N_p|\text{positive}) \).

Calculate the probability that the message item in each column in the history database is not a typical defect, that is, the conditional probability, \( P(A_1|\text{negative}) \), ..., \( P(A_p|\text{negative}) \), \( P(B_1|\text{negative}) \), ..., \( P(B_p|\text{negative}) \), ..., \( P(N_1|\text{negative}) \), \( P(N_2|\text{negative}) \), ..., \( P(N_p|\text{negative}) \).

3.3. Calculating whether the real-time signal corresponds to a typical defect

When a new correlation signal is received, the historical experience database and the calculated conditional probability can be used to determine whether the corresponding defect when the signal occurs is a typical defect. The new signal is \( X \), and the signal content is shown in Table 3:

| MESSAGE A | MESSAGE B | ⋯ | MESSAGE N |
|-----------|-----------|---|-----------|
| \( A_x \) | \( B_x \) | ⋯ | \( N_x \) |

Calculate the probability of the signal \( X \) as a typical defect based on the conditional probability calculated by the historical experience database:

\[ P(\text{positive typical defects}|X) = P(\text{positive})P(A_x|\text{positive})P(B_x|\text{positive}) \cdots P(N_x|\text{positive}) \]  
\[ P(\text{negative typical defects}|X) = P(\text{negative})P(A_x|\text{negative})P(B_x|\text{negative}) \cdots P(N_x|\text{negative}) \]

Compare \( P(\text{positive typical defects}|X) \) and \( P(\text{negative typical defects}|X) \). If \( P(\text{positive typical defects}|X) > P(\text{negative typical defects}|X) \), then the equipment defect corresponding to signal \( X \) is a typical defect. If \( P(\text{positive typical defects}|X) > P(\text{negative typical defects}|X) \), then the equipment defect corresponding to signal \( X \) is not a typical defect.

3.4. Historical experience database maintenance

The messages are maintained in the historical experience database, and the number of signals is changed by \( n+1 \) the next time the conditional probability is calculated.

3.5. Application examples

Taking the typical defects of the switch as an example, a historical experience database is established. Since it has a huge amount of information, this article only lists some information, as shown in Table 4.

Calculate the probability of a typical defect based on the historical experience database:

\[ P(\text{positive}) = \frac{3}{10} = 0.3 \]  
\[ P(\text{negative}) = \frac{7}{10} = 0.7 \]

Calculate the conditional probability of each information item as follows:
The signal X is that Jinshan Power Plant No. 1 changes the main 2201 switch and SF6 air pressure low alarm. Through the signal correlation equipment, the specific information of the signal and the signal generating equipment can be known, as shown in Table 5. Through probability calculation, it can be known that:

\[ P(\text{ABB}|\text{positive}) = 0.33, \quad P(\text{PINGGAO}|\text{positive}) = 0.67, \quad P(\text{frequent}|\text{positive}) = 0.33, \]
\[ P(\text{not frequent}|\text{positive}) = 0.67, \quad P(\text{ABB}|\text{negative}) = 0.51, \quad P(\text{PINGGAO}|\text{negative}) = 0.43, \]
\[ P(\text{frequent}|\text{negative}) = 0.57, \quad P(\text{not frequent}|\text{negative}) = 0.43. \]

The signal X is that Jinshan Power Plant No. 1 changes the main 2201 switch and SF6 air pressure low alarm. Through the signal correlation equipment, the specific information of the signal and the signal generating equipment can be known, as shown in Table 5. Through probability calculation, it can be known that:

\[ P(\text{frequent}|\text{positive}) = 0.33, \quad P(\text{Pinggao}|\text{positive}) = 0.67, \quad P(\text{frequent}|\text{negative}) = 0.57, \]
\[ P(\text{not frequent}|\text{negative}) = 0.57, \quad P(\text{north-east}|\text{positive}) = 0.67, \quad P(\text{north-east}|\text{negative}) = 0.43, \]
\[ P(\text{4}|\text{positive}) = 0.33, \quad P(\text{4}|\text{negative}) = 0.13, \quad P(\text{220kV}|\text{positive}) = 0.67, \]
\[ P(\text{220kV}|\text{negative}) = 0.43, \quad P(\text{hydraulic}|\text{positive}) = 0.67, \quad P(\text{hydraulic}|\text{negative}) = 0.43. \]

Calculate the probability that the signal X is a typical defect:

\[
\text{outdoor}
\]

Calculate the probability that the signal X is not a typical defect:

\[
\text{outdoor}
\]

\[ P(\text{positive}|X) > P(\text{negative}|X), \text{then the defect corresponding to signal X is a typical defect, and this signal is maintained to the historical experience database.} \]

\[ n \text{ becomes 11 in the next calculation.} \]

4. Conclusion

This paper proposes an operating state evaluation algorithm and a typical defect classification algorithm for equipment monitoring of the whole network substation. This helps to adjust the centralized supervision of the whole network substation and can make the substation operating state visually represent. At the same time, it solves the problem that the signal is difficult to be associated with typical defects, and it is convenient for the dispatching center and the inspection unit to distinguish the typical defects or family defects of the equipment. In the era of big data, it is of great significance to the regulation of smart grids.

Table 4. Switch typical defect history experience database.

| Equipment signal number | Manufacturer | Frequent signal | Operation area | Operation time | Voltage | Use environment | Structure type | Operating mechanism form | Extinguishing medium | Typical defect |
|-------------------------|--------------|----------------|----------------|----------------|---------|-----------------|----------------|------------------------|------------------|--------------|
| 1                        | ABB          | Negative       | South-west     | 4              | 15.8kv  | Indoor          | Tank-type      | Liquid spring          | SF6              | Negative |
| 2                        | ABB          | Positive       | North-east     | 1              | 220kv   | Indoor          | Tank-type      | Hydraulic              | Pneumatic         | Positive |
| 3                        | PINGGAO      | Negative       | Central China  | 6              | 10kv    | Outdoor         | Magnetic stanchion | Spring                 | Low-oil           | Positive |
| 4                        | PINGGAO      | Positive       | North-east     | 4              | 10kv    | Outdoor         | Hydraulic      | Tank-type              | SF6              | Positive |
| 5                        | ABB          | Negative       | North China    | 7              | 220kv   | Outdoor         | Magnetic stanchion | Pneumatic              | SF6              | Negative |
| 6                        | ABB          | Positive       | North China    | 8              | 10kv    | Outdoor         | Tank-type      | Liquid spring          | SF6              | Negative |
| 7                        | ABB          | Negative       | North-east     | 3              | 220kv   | Indoor          | Magnetic stanchion | Hydraulic              | Bulk-oil          | Positive |
| 8                        | PINGGAO      | Positive       | South-west     | 8              | 220kv   | Outdoor         | Magnetic stanchion | Pneumatic              | Vacuum            | Negative |
Table 5. Specific information of signal X.

| Manufacturer | Frequent signal | Operation area | Operation time | Voltage Use environment | Structure type | Operating mechanism form | Extinguishing medium |
|--------------|-----------------|----------------|----------------|------------------------|----------------|--------------------------|---------------------|
| PINGGAO     | Positive        | North China    | 220kv         | Outdoor               | Magnetic stanchion | Hydraulic                | Low-oil             |
| PINGGAO     | Negative        | Central China  | 10kv          | Indoor                | Hydraulic        | Bulk-oil                 | Negative            |

Acknowledgment
Scientific Research Program of Jiangsu Electric Power Company, State Grid Corporation of China – Research and Application of Key Regulation Techniques for Intelligent Supervision and Full-dimensional Control of Large Power Grids (J2018065)

References
[1] Amira Mohamed, Shady S. Refaat, Haitham Abu-Rub. A Review on Big Data Management and Decision-Making in Smart Grid[J]. Power Electronics and Drives, 2019, 3(39): 1-13.
[2] Xin Yaozhong, Shi Junjie, Zhou Jingyang, et al. Technology development trends of smart grid dispatching and control systems[J]. Automation of Electric Power Systems, 2015, 39(1): 2-8.
[3] XUE Yusheng, LAI Yening. Integration of macro energy thinking and big data thinking part one big data and power big data[J]. Automation of Electric Power Systems, 2016, 40(01): 1-8.
[4] Leng Xiwu, Chen Guopig, Jiang Yu, Zhang Jiaqi, Xiao Fei. Smart grid monitoring operation big data application model construction method[J]. Automation of Electric Power Systems, 2018, 20(42): 115-122.
[5] Liu Yuquan, Guo Yuanjun, Lu Guojun, et al. Transmission line tripping analysis and correlative factor mining for Guangzhou power grid based on big data[J]. Southern Power System Technology, 2017, 11(4): 38-44.
[6] Wu Guixiong. Hydropower plant equipment maintenance and defect management[J]. Technology Innovation and Application, 2019, 25, 187-188.
[7] WU Jiang, Wang Ning, Cheng Yibo, Liu Qi, Wang Zi, Wang Gang. Analysis of family defect of equipment based on big data[J]. 2016, 13(39): 104-106.
[8] Cheng Peng, Wang Hong, Zuo Jun. Regulation and integration of centralized monitoring information three-level management and control system[J]. Electronics World, 2019, 14(105): 198-190.