A NOVEL STATE ESTIMATION METHOD BASED ON QUALITY TAG FOR DISTRIBUTION NETWORKS

Weiwei XU  Dong LIU  Rong LU, Jidong QIU, Huaming PAN, Zhong WEI
Shanghai Jiaotong University – China  Shanghai Jiaotong University – China  Shanghai Shidong Power Supply Company
E-mail: vvvwei@sjtu.edu.cn  E-mail: dongliu@sjtu.edu.cn  E-mail: lurong@smepc.com

ABSTRACT
A practical SE (state estimation) method based on multiple data source is proposed. With the help of Decision Theory the bad data identification method which combines the historical data and remote data from IDP (Integrated Data Platform) and 6 data rules are proposed. A logical judgment based on the 6 data rules is conducted, and the data quality evaluation method is established accordingly. The data repairing improves the quality of the input data for state estimation. The weighed least-square state (WLS) estimation takes into account the reliability of the measurements, reducing the influence of bad data. Research shows that this method is superior to traditional state estimation methods in that it tackles with bad data more efficiently, resulting in more compatible state information for advanced applications. Experiments in the distribution network of DBC of Shanghai Pudong District prove the method is reliable and efficient.

INTRODUCTION
State estimation \footnote{1} (SE) plays an important role in modern power energy management system. By processing a set of raw measurement data, it provides a real-time system solution which is the basis of the advanced applications for system security monitoring and control. State estimation was first studies in 1970s. Huge effects were devoted to state estimation on the basis of the software and hardware conditions at past decades, and significant advances on transmission networks were made, especially those on state estimation criteria.\footnote{2,3} State estimator has to consider certain issues such as detection and identification of bad data, identification of topology data and system observation. Static state estimator identifies bad data according to the spatial connections between different measurements at constant sampling interval. Dynamic state estimator identifies abnormal events according to the real time measurements of two-dimensional state estimator based on both time and space. However, there are still several issues need to be addressed.

In the past, due to lack of data acquisition and monitoring facilities, state estimation for distribution networks is transformed into a series of power flow matching problems under certain assumptions. Reference \footnote{4–6} provides ideal solutions for power flow in distribution networks. With the implementation of power system automation, new challenges emerge. The conception of “multi-source data” was used to demonstrate the current data status in transmission networks and was subsequently employed in the selection of correct data from different data sources according to their quality signs and priority.\footnote{7,8} Nowadays, the monitoring system in distribution networks has been greatly improved with the development of Distribution Automation (DA) \footnote{9}. As advanced applications are to be utilized in distribution networks, it is necessary and possible to employ SE in distribution networks.

Two types of data are required for distribution network SE: the measurement data and the network data. The measurement data includes remote data and telemetry data. The former is mainly from remote devices of SCADA system for distribution network (DSCADA), the basis for distribution network topology. The later is composited of not only real time measurements from DSCADA, but also historical data from other data acquisition systems and electrical degree data from electricity metering system, which are regarded as redundant data. Since the data volume is huge, and may contain bad data, it is important that a data pre-process is conducted, including bad data detection and identification, and data repairing.

At the same time, innovative technology and theory keep immersing in data processing area, such as expert system, neural network, genetic algorithm, machine learning, et al. which may provide solutions for intelligent algorithms of SE problems. This paper proposed a new methodology for distribution network SE based on the current situation. In order to achieve the object of a practical solution, the work considered practical engineering problems. And the decision tree theory was adopted.

PRESENT SITUATION OF MONITORING AND DATA ACQUISITION SYSTEMS IN DISTRIBUTION NETWORKS

Energy acquisition system provides the database for a set of advanced functions for monitoring and control in distribution networks. The integrity, accuracy and sample frequency of the data has an impact on the efficiency of the advanced functions.

As the construction of distribution network goes in China, all the information systems, SCADA systems in distribution networks keep developing. In some local area
of the distribution networks, multi-source measurements are available.

Fig 1 The network diagram of the test lines with SCADA system

Fig 1 shows the test lines with SCADA system. This central business area in Shanghai was studied as an example. Currently all the 35kV substation in this area are equipped with integrated automation system; all the 10kV costumes are equipped with centralized automatic meter-reading system; the bar transformer, the package transformer, the pole top breaker and other different kinds of distribution substation in 10kV networks are equipped with DTU, TTU, and the FTU, respectively. Therefore, three-phase current, voltage, power consumption, remote switching data and information can be collected, uploaded and stored immediately. All the acquisition systems are integrated to IDP (Integrated Data Platform).

However, in the current stage of measuring system, the quality of measurements in distribution networks is not good. There is significant difference in measuring accuracy between different data sources, and the layout of measuring points are usually unreasonable. So it is very difficult for the application of traditional state estimation methods in distribution networks.

As discussed, the “multi-data source” status currently exists in the distribution networks. Strategies have to be considered to process the raw data, in order to guarantee the quality of input data for advanced applications, and the efficiency of the monitoring and management of distribution networks.

DISTRIBUTION NETWORKS SE BASED ON QUALITY TAGS

Quality tag
Thanks to the wide application of data acquisition systems in distribution networks, it is possible to collect real time data in various ways. This offers a number of advantages, but still suffers from drawbacks such as the difference of accuracy of different systems. So it is essential to judge the reliability of the data acquisition system and select high quality data for subsequent applications. A quality tag is attached to each data source, representing the current data quality.

The quality tag, denoted by \( Q(i) \), is ranging from 0 to 1. Bigger \( Q(i) \) is, more reliable the referring data is. \( Q_V(i) \) and \( Q_P(i) \) represents the quality of node voltage, current and power respectively.

The initiate value of \( Q(i) \) is 1. \( Q(i) \) is obtained from the following equation:

\[
Q(i) = 1 - (1 - \text{flag}) \times W, \quad \text{flag}=0\text{or}1 \quad (1)
\]

‘flag’ is referring to each of the following 6 rules, if the rule is satisfied, then flag=1; else, flag=0; W is the corresponding weight.

Detailed survey and demonstration of the valuation of quality tag are presented in [11]. Considering the huge data volume and the existence of multiple interacting bad data, this paper selects a sample set of data according to a certain proportion. The quality tags of the sample data are valued; a decision tree is derived from the sample data. The initializing of the quality tag of the real time data is conducted by the decision tree. The following sections will discuss the basis of bad data detection and the concept of ID3. The data repairing module process the raw data with initialized quality tags. The repaired data and the corresponding quality tags are used as the input for SE.

Bad data detection basis

For bad data detection in radial distribution networks, the following rules are used:
- Rule 1: the voltage should be in the standard range
- Rule 2: the voltage gradually reduced along the feeder lines[11]
- Rule 3: the KCL rule[11]:
  \[
  I_n > \sum_{m \in M} I_m \quad (2)
  \]
  where M represents the downstream neighbor node set of m.
- Rule 4: based on the particular research object of this paper, the degrees of ammeters and the active power satisfy the following equation:
  \[
  PM_j = [(P_{i=1} + P_i) / 2 \pm \xi] / (3.6 \times 10^8)
  \]
  where \( \omega \) is the average value of the measurements of the latest 7 days at the same sampling interval.
  \[
  \frac{|z_{i-1} - z_i|}{z_i} < \xi \quad (3)
  \]
  - Rule 6: verification with history data:
    \[
    |z_{i-6} - z_i| < \omega \quad (4)
    \]
    where \( z_{i-6} \) is the average value of the measurements of the latest 7 days at the same sampling interval.
The quality tags derived from the bad data detection module are used as input for data repairing module.

**ID3 algorithm in bad data detection**

The most influential algorithm in decision tree learning is ID3, proposed by Quinlan in 1986[10]. ID3 uses the heuristics of minimizing the “entropy” in determining which attribute should be next selected in the decision tree. If attribute A has m values (i.e., A1,A2,...,Am) and the training subset Si having attribute value Ai can be partitioned into \( P_i^+ \) positive training instances and \( P_i^- \) negative training instances, then the entropy E of choosing A as the next attribute can be calculated by the following formula:

\[
E(S_i) = -p_i^+ \log_2 p_i^+ - p_i^- \log_2 p_i^-
\]

Among all the feasible attributes, the one which causes the minimum entropy will be chosen as the next attribute. The same procedure is then repeated until each terminal node in the decision tree contains only training instances of the same class.

Under ideal conditions, if the data satisfy all the six rules, it’s definitely of high quality, vice versa. However, the existence of bad data makes the situation complicated. Since some of the rules are base on comparison of different data, the bad data will have a negative impact on the result. In another word, if some data was not in line with certain rules, it doesn’t necessarily means it’s of poor quality.

**Table 1 Subset of sample data for Decision Tree**

<table>
<thead>
<tr>
<th>No</th>
<th>QT</th>
<th>rule1</th>
<th>rule2</th>
<th>rule3</th>
<th>rule4</th>
<th>rule5</th>
<th>rule6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>U</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>0.6</td>
<td>Y</td>
<td>U</td>
<td>N</td>
<td>U</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>U</td>
<td>U</td>
<td>N</td>
<td>U</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>Y</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>U</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>0.2</td>
<td>U</td>
<td>U</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>0.6</td>
<td>Y</td>
<td>U</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>U</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>0.2</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>0.6</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>0.6</td>
<td>N</td>
<td>N</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>14</td>
<td>0.2</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

*Y=satisfied; N=not satisfied; U=unknown

**Fig 2 Decision tree grow from the sample data**

Table 1 lists a subset of sample data for decision tree. The sample data represents typical data in the prototype. And the quality tags are assigned according to the true value. The quality tags are updated in the process of data repairing.

Table 2 shows that the data repairing process is complete. Set a upper bound \( \chi_{good} \), if \( Q(i) > \chi_{good} \), the data repairing process breaks. The repairing equations are listed as follows:

\[
U_j^r = U_i^{max} \times Qtag + \frac{U_i^{max} + \max_{<i} U_k^{max} \times (1 - Qtag)}{2}
\]

Where j is the father node[11] of k, J is the son node[11] set of j, K is the son node set of k. And \( U_j^r \) is the modified value of \( U_j^{mea} \).

If data K is current, then:

\[
I_j^r = I_i^{max} \times Qtag - \frac{\left[ (I_i^{max} - \sum_{<i} I_j^{max}) + \sum_{<i} I_j^{max} \right]}{2} \times (1 - Qtag)
\]

Where j is the father node[11] of k, J is the son node[11] set of j, K is the son node set of k. And \( I_j^r \) is the modified value of \( I_j^{mea} \).

**THE SE PROCESSING FLOW AND THE TEST RESULTS**

Research shows that this method is superior to traditional state estimation methods in that it tackles with bad data more efficiently, leaving less error to deal with in latter procedures. This method provides conforming state information for posterior advanced application in distribution net works. The application of the proposed method in the distribution network of DBC of Shanghai Pudong District shows that it is reliable and efficient.
Fig 3 illustrates the overall process of quality-tag-based state estimation. Multi-source measurements are integrated into IDP. Bad data detection and identification are based on 6 rules and the decision tree theory. Quality tags are initialized by decision tree and updated in data repairing module. And the modified measurements and their quality tags are used as input for SE. The quality tags are counted as weights in WLS.

This novel method was employed in SE of the test lines. Table 2 compares the results between the traditional WLS and the novel method. Figures in Table 2 show that within the same iteration step, the novel method has higher accuracy. The novel method is convergent with 1.8459e-9, and the traditional one is with 2.17811e-7. Results show that the SE results from the proposed novel method are more compatible.

<table>
<thead>
<tr>
<th>Table 2 Comparison of iteration efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional WLS</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Iteration number</td>
</tr>
<tr>
<td>Difference between</td>
</tr>
<tr>
<td>Assist criteria</td>
</tr>
</tbody>
</table>

CONCLUSION

This paper carried out a pre-process of the measurements for distribution network SE based on the current data acquisition systems and the multi-data source situation. The proposed work minimized the effects of the bad data on the SE program, and overcome the disadvantage of evenly distributed bad data within all the data. This method combined the nature of the measurements in the distribution networks into the theory of decision tree, and accelerated the bad data detection speed and improved the accuracy. Therefore the risks of the misjudging bad data are reduced. In the data repairing module, this work improved the quality of the input of the SE by repairing the bad data according to the redundant measurement data, together with the concept of quality tag. The weighted least square SE based on quality tag was employed in the process of the SE of the testing lines. The results show that this method improves the reliability of the measurements and reduces the effect of bad data, which is promising and practical for distribution automation.

REFERENCES

China.