Index System of the Identification of Electricity Stealing and Leakage Research

Jun TANG¹, Cong ZHONG², Bing-yao Li², Ju-ben XU¹ and Wei-qi GUO¹

¹School of Information and Safety, Zhongnan University of Economics and Law, Hubei, Wuhan, 430073, China
²Shenzhen Power Supply Co., China Southern Power Grid, Shenzhen, Guangdong, 518010, China

Keywords: Electricity stealing and leakage, Index system, Clustering, Fuzzy neural network.

Abstract. With the development of smart grid, data mining makes identification of users of electricity stealing and leakage a reality. This paper systematically studied the data of electricity information for unusualness detection. Firstly, it analyzed the behavior features of those users suspicious of electricity stealing and leakage, constructed a preliminary behavior identification index system. Then it used the discrimination analysis and correlation analysis method for quantitative judgment on the feasibility of screening evaluation index, excluding some indexes, finally got three first level indexes and eight secondary indexes. The paper presented three new indexes, including coefficient of dispersion of A&C phase current’s (voltage’s) difference and the clustering number of monthly electricity load pattern. Calculation method of the index was also described in detail. The index system has shown its feasibility by application of Clustering method and the fuzzy neural network.

Introduction

Stealing and Leakage of Electricity is a serious problem that power supply enterprise commits to solve in China. Stealing and Leakage of Electricity is a kind of situation that customers steal electricity for personal interest or breakdown of metering mechanism leading to counting errors. It not only makes huge economic losses to the state and power supply enterprises, but also exists hidden danger threatening the safety of power grid. So power supply enterprises have been committed to detect abnormal behavior and punish behaviors of electricity stealing.

As related works people widely use classification and clustering technology to identify abnormal power consumption behavior, Deloitte established total analysis for America; IBM offered recognition scheme of abnormal consumption for Canadian Hydro One to identify electricity stealing behavior; American research literature of power distribution big data paid special attention to accurate load adjustment plan, put forward a system solution to predict optimization module based on Hadoop and SAS machine learning. To sum up, foreign literature about identifying electricity stealing and leakage are scattered, lacking comprehensive and systematic introduction of solutions.

This paper systematically studies the electricity information data. On the basis of reference to related literature index, we presented three new indexes, including coefficient of dispersion of A&C phase current’s (voltage’s) difference and the clustering number of monthly electricity load pattern. Finally we constructed the index system of the identification of electricity stealing and leakage, so that it can support other project being carried out.

Analysis of the Abnormal Features of the Electricity

The methods to steal the electricity can be all over the map, it can use reverse meters, phase shift machines, wired remote control, wireless remote control and so forth. Although the methods are various, some are even brilliant and concealed; actually they are all the same. Their main purpose is to make the watt-hour meter measuring less. It will be reflected on the electricity information data. Electricity information acquisition system will collect users’ indicated energy value, voltage, current,
power, power factor data in its each period every 15 minutes. In this paper, we will analyze the data features of electricity stealing and leakage based on the stealing cases and literature.

**Abnormal Line Losses in Circuit**

Calculate the difference between power supply quantity and power sale quantity that line loss watt-hour meter acquired, it is divided into two parts including theoretical line loss and management line loss. Theoretical line loss, is the sum of electric quantity loss of each element in the power grid. However management line loss is caused by poor management, illegal electric power usage, electricity stealing, errors in copying and checking and so on. Theoretical line loss changes slightly within a certain rang, we can assume that the subscriber line circuit and equipment loss is relatively constant, that is to say normal change of the power consumption won’t cause an acute change of feed power loss. When someone use electricity abnormally, the exceedingly increases of management line loss will cause statistical line loss changing a lot. From Figure 1 we can see that, a feeder has electricity stealers, in first 130 days, the statistical line loss remains high (has surpassed 10%). After we found out and punished the stealer, the line loss of the circuit reduced to normal levels immediately. The statistical line loss is of great significance to the auditing work of the power sector. If a feeder has electricity stealers, the line loss rate will show abnormalities, usually higher than the usual one. At present, many power supply enterprises measured the line loss rate to judge which area is severely afflicted area of stealing electricity.

**Abnormal Data of Power Consumption**

The main purpose of power theft is to reduce the reading on the watt-hour meter so that it can pay less electric charge. That’s why the data showed exception when someone stole electricity. The abnormality expressed in two aspects. One is the electricity use reduction. The reduction is compared with the actual consumption but not compared with last month’s consumption. The monthly electricity consumption is not a fixed value; it will fluctuate by the influence of weather, seasons, holidays and so on. Consequently, we need to analyze it combined with the history electricity consumption as well as other customers’ data. The other is the irregularity of the electricity usage. Through data analysis, we found that the majority of users’ monthly consumption data remain relatively stable. Users generally have its own consuming regularities, Electrical equipment and power consumption habit won’t change a lot in a short period of time. Figure 2 is the normal and abnormal users’ standard electricity value comparison, we can clearly see from the picture that the normal users’ electricity consumption is relatively stable, however the abnormal users’ electricity consumption fluctuate strongly.

![Figure 1. Line loss rate of feeder statistical with abnormal user.](image1.png)

![Figure 2. Normal and abnormal users' standard electricity value comparison.](image2.png)

**Abnormal Load Data**

When we use phase shifting, under voltage, undercurrent methods to steal electric-energy, it will cause the abnormality of the voltage and the current, such as unbalance of three phase current, voltage or current lacking a phase, and the sudden changing of voltage or current. When large consumers start using or closing the theft device, the large amount of electricity will cause sudden change of the current. Figure 3 is a discrete coefficient curve of A&C phase current’s (voltage’s) difference of a feed
(x1 represents a power user). In this picture, x6 is the stealer, it can be clearly seen that its discrete coefficient of A&C phase voltage’s difference is much larger than other consumers. However its discrete coefficient of A&C phase current’s difference doesn’t appear to be much different from x1 and x5. More importantly, the abnormality of voltage and current data is sufficient and unnecessary condition. That is to say, even though the consumer’s voltage and current data are normal, we can’t eliminate his/her suspects of electric theft.

Figure 3. Discrete coefficient curve of A&C phase current’s (voltage’s) difference.

Figure 4. Line chart of user’s daily peak, flat and valley period load rate.

**Index System Construction of the Identification of Electricity Stealing and Leakage**

**Construction of Index System**

After we analyzing the cases of electric-theft repeatedly, known the data features of abnormal consumption, we established the index system of the identification of electricity stealing and leakage according to comprehensiveness, sense, multilayer and feasibility principles. The system is divided into two parts, three first-level indicators and thirteen secondary indicators, as shown in table I.

<table>
<thead>
<tr>
<th>First-level</th>
<th>Secondary</th>
<th>Index meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line loss</td>
<td>Median of line loss rate</td>
<td>There are more than 15 days in a month that the line loss rate is more than the median value</td>
</tr>
<tr>
<td></td>
<td>Average of line loss rate</td>
<td>The higher the value is, the more serious the line losses are</td>
</tr>
<tr>
<td>Electric quantity</td>
<td>Matching degree of abnormal dates</td>
<td>Matching degree is the amount of dates that consumers use much greater or less than normal value. Higher Matching degree of abnormal dates indicates most electricity-abusing suspicion.</td>
</tr>
<tr>
<td></td>
<td>Coefficient variance of electricity consuming monthly</td>
<td>Find out consumers whose monthly power supply fluctuate narrowly or dramatically comparing to consumers’ historical dates.</td>
</tr>
<tr>
<td></td>
<td>Users power consumption average ratio monthly</td>
<td>The closer it is from zero, the more consumer’s monthly power supply deviates from industry average.</td>
</tr>
<tr>
<td></td>
<td>Coefficient variance ratio of users and industry power consumption monthly</td>
<td>The Coefficient variance ratios of users and industry power fluctuate in the same range.</td>
</tr>
<tr>
<td></td>
<td>Coefficient variance ratio of users power consumption monthly</td>
<td>The higher this value is, the greater the variation in users’ monthly power consumption. When this variation exceeds threshold value, they are on suspicion of stealing electricity.</td>
</tr>
<tr>
<td></td>
<td>The maximum of load rate monthly</td>
<td>With economic factors taken into consideration, special users choose transformers with load rate at approximately 80%. We may suspect consumers of power abusing, when they prefer with minor load rate.</td>
</tr>
<tr>
<td></td>
<td>Coefficient variance of the difference ratio of daily peak and valley</td>
<td>Coefficient variance of the difference ratio of daily peak and valley. When abnormal variation appears, the discrete coefficient ratio is slightly larger. This indicates suspicion of power stealing.</td>
</tr>
<tr>
<td></td>
<td>Average changing rate of the difference between weekly peak and valley</td>
<td>Under normal circumstances, difference rate between weekly peak and valley should remain constant with tiny variation that close to 0.</td>
</tr>
</tbody>
</table>
Monthly electricity consuming load mode
From the peak period to off-peak period, load rate clustering number is users’ power-consuming model in a month. Higher this value is, the greater suspicion of stealing electricity.

Coefficient of dispersion of A, C phase current’s difference
When uses stealing electricity, they change the current coil of short circuit meters or increase sliding resistor. These changes may cause the disequilibrium of three-phase current.

Coefficient of dispersion of A, C phase voltage’s difference
Voltage instability suggests the suspicion of stealing or leaking electricity. However, we cannot confirm consumers are using power in order only by voltage stability.

Index Selection
Using correlation analysis to secondary indexes belonged to the same first level index. First we do correlation analysis to the two indicators of line loss. We do discrimination analysis to these two indexes, compared with the average, the median of the line loss rate can better reflect the abnormality degree of the line loss rate, it is more stable, so we eliminated the line loss average, keep the median. Similarly, we do the same analysis to the indicators belonged to electric quantity and load, finally we can establish an index system. As shown in table II.

<table>
<thead>
<tr>
<th>First-level</th>
<th>Secondary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line loss</td>
<td>Median of line loss rate</td>
</tr>
<tr>
<td>Electric quantity</td>
<td>Matching degree of abnormal dates</td>
</tr>
<tr>
<td></td>
<td>Coefficient variance t of monthly electricity consuming</td>
</tr>
<tr>
<td></td>
<td>Coefficient variance ratio of users and industry power consumption monthly</td>
</tr>
<tr>
<td>load</td>
<td>Coefficient variance of the difference ratio of daily peak and valley</td>
</tr>
<tr>
<td></td>
<td>Monthly electricity consuming load mode</td>
</tr>
<tr>
<td></td>
<td>Coefficient of dispersion of A, C phase current’s difference</td>
</tr>
<tr>
<td></td>
<td>Coefficient of dispersion of A, C phase voltage’s difference</td>
</tr>
</tbody>
</table>

Index Calculation
Using discrimination analysis, correlation analysis and other quantitative analysis methods to quantitatively determine the indicators’ feasibility, and delete indicators that highly correlated with other indicators or weak in discrimination.

1) Index of line loss rate

\[ L = S - Q \]  (1)

\[ Lr = (L / S) \times 100\% \]  (2)

where \( L \) is the statistical line loss, \( S \) is the quantity of daily power supply, \( Q \) is the quantity of daily power sale, \( Lr \) is the line loss. Median of line loss rate means to calculate the median of the line loss.

2) Index of electric quantity

- Matching degree of abnormal dates

Matching degree of abnormal dates means the matching degree of abnormal power consuming dates and abnormal line loss dates. We can use modified box-plot rules, partial outlier factor and outlier ranking based on cluster to find abnormal values of power consumption and line loss. Modified box-plot rules is a method that decides the value’s outlier degree according to the distance between the value and the central value.

\[ ND(x) = \frac{|x - \bar{U}_x|}{IQR} \]  (3)
where \( x \) is a day’s line loss or power consumption data, \( IQR_x \) is the Interquartile range of the line loss, \( U_i \) is the median of the line loss, \( ND(x) \) is the distance between line loss and standard line loss.

• Coefficient variance of monthly electricity consuming

\[
CV = Q_e / Q_e
\]

where \( CV \) is the Coefficient variance of monthly electricity consuming, \( Q_e \) is the average of monthly power consumption, \( Q_e \) is the standard deviation of monthly power consumption.

• Coefficient variance ratio of users and industry monthly power consumption

\[
R_{cv} = CV_i / CV_2
\]

where \( R_{cv} \) is the Coefficient variance ratio of users and industry monthly power consumption, \( CV_i \) is the Coefficient variance of monthly electricity consuming, \( CV_2 \) is the Coefficient variance of industries’ monthly electricity consuming. And what industry the user belongs to will be decided by the monthly power consumption, and the daily peak.

3) Index of load

• Coefficient variance of the difference ratio of daily peak and valley

\[
R_d = (L_m - L_s) / L_m
\]

\[
CV_d = R_d / R_e
\]

where \( R_d \) is the difference ratio of the peak and the valley, \( L_m \) is the biggest daily load, \( L_s \) is the smallest daily load, \( CV_d \) is the Coefficient variance of the difference ratio of daily peak and valley, \( R_d \) is the standard deviation of daily difference ratio of the peak and the valley, \( R_e \) is the average of daily difference ratio of the peak and the valley.

• Monthly electricity consuming load mode

\[
LR_p = L_p / L_d
\]

\[
LR_f = L_f / L_d
\]

\[
LR_v = L_v / L_d
\]

where \( LR_p \) is the load ratio of the peak, \( LR_f \) is the load ratio of the flat, \( LR_v \) is the load ratio of the valley, \( L_p \) is the average load of the peak, \( L_f \) is the average load of the flat, \( L_v \) is the average load of the valley.

24 hours of a day are divided into three parts according to the electricity consuming habit of each area. They are the peak, the flat and the valley. The three indicators reflected consumers’ electric consuming behavior features. We use clustering analysis to the abnormal consumers’ data every month, the fewer the categories are, the more stable the consumers’ electric consuming behavior is. The group produced by cluster will not be specific. Load rate of the peak, the flat and the valley cluster based on squared Euclidean distance. Squared Euclidean distance of \((x, y)\) is as followed:

\[
\sqrt{\sum_{i=1}^{k} (x_i - y_i)^2}
\]

There we use modified Euclidean distance, the distance between \((x, y)\):

\[
\sqrt{[(x_2 - x_1) - (y_2 - y_1)]^2 + [(x_1 - x_2) - (y_1 - y_2)]^2}
\]
Verification and Conclusion

The identification of electricity stealing and leakage can be realized by classification and predicting model. The fuzzy neural network combined the advantages of artificial neural network and fuzzy comprehensive evaluation method, so it has obvious advantages in dealing with the electric abnormal data which contains uncertain information and noises. This article uses the fuzzy neural network to verify 10350 data of large consumers from a power supply station (from February, 2013 to February, 2014). There are 67 cases of electric theft among those data. We selected 37 users’ data who are stealing the electricity and 60 users’ data who are normal consumers, and put all the data into the fuzzy neural network to train the model. Then we applied the model to users’ data after July, 2013, the results are promising. We use decision accuracy and retrospective accuracy to evaluate the model. According to the formula, if the decision precise and retrospective accuracy are higher, the model is better. From the data in table 4, the retrospective accuracy reached 100%, which means we found all the authenticated electric-energy stealers. The decision accuracy is only 11%, this value is lower than the actual value. For the reason that there must be other electric-energy stealers in the consumers that we have not examined, so the actual matching electric-energy stealer should be higher. The pros and cons of sample size and the coverage will also influence the consequence of the model. Generally speaking, the verification results show that this index system is practical and promising.

Acknowledgment

The paper is sponsored by Shenzhen Power Supply Co. Ltd., which is affiliated by China Southern Power Grid.

References


