Research on the Causes of the Accurately Locating Line Loss Exceptions by Querying the Power Consumption Information Acquisition System Based on the Big Data Platform

Xiaoying Zhu* and Xinsheng Jin*
State Grid Zhejiang Yinzhou District Power Supply Company, Ningbo 315100, China

*Corresponding author e-mail: 106749850@qq.com, *Xsjin@163.com

Abstract. The combination of a power user information collection system with a two-way interactive platform for power business has constructed intelligent power consumption big data for the energy flow, information flow, and business flow of the power grid and customers. Aiming at the problem that the current distribution network transmission line loss anomalies cannot be traced and difficult to locate, based on the data collected by the measurement automation system, through the management of station, line, transformer, and household data, the use of data mining and analysis technology will analyze factors affecting line loss. Through the intelligent diagnosis model of line loss abnormality, it analyses the abnormal causes that affect line loss in real time, promotes the transformation of low-voltage line loss management from result management to process management, improves the power supply operation efficiency and management level, reduces the operating costs of enterprises, and improves the economic benefits of enterprises.

1. Introduction
The informatization needs of smart grids require more refined requirements for grid line loss management, and it is necessary to change the extensive line loss management mode. Grid line loss rate is a comprehensive technical and economic indicator that reflects the production technology and operation and management level of power supply enterprises. Since 2005, the overall line loss rate of the State Grid Corporation of China has been 6.59%, ranking the middle level in the world. Standardizing the line loss management process, accurately calculating the true line loss, and effectively reducing the line loss rate of the power grid are important ways to improve the economic and social benefits of power supply enterprises. The biggest obstacle to traditional line loss statistics is that the power consumption for line loss calculation is obtained by manual meter reading. Due to the reasons such as the meter reading time is not synchronized, the ledger data is incomplete, etc., leading to a large error in the statistical results of line loss. To accurately grasp the true situation of grid line loss, it is necessary to use computer monitoring technology to automatically collect power for supply and sale, to carry out information statistical analysis and management of grid line loss, and to achieve accurate statistics of line loss fundamentally.

It is very important to strengthen the research and innovation on the line loss related theories of the distribution network and find more accurate calculation methods. Using these new methods to improve the efficiency of line loss calculation for power supply companies, and by analyzing the line loss results...
that are consistent with the actual situation, we can find out the weak links in the power grid, which will help the power supply companies to make appropriate adjustments to their network structure for the later power distribution. Network construction or optimization and transformation provides reliable theoretical basis, thereby establishing a more optimized distribution network. Secondly, the calculation of the distribution network line loss is more accurate, which helps to improve the operation and management level of power grid enterprises, and can continuously improve the economic benefits of the enterprise, and achieve a bumper harvest of both management and economic benefits.

Based on this research background, based on big data technology, this paper uses machine learning algorithms, clustering, mining, association relationships and other analysis methods to make full use of the massive data resources generated by the metering system to build an intuitive and effective distribution network line loss analysis and evaluation system. The line loss structure is visualized, the line loss rate changes can be traced to the source, and the loss reduction task can be decomposed. It supports all-round and multi-dimensional management and technical loss reduction.

2. Line loss analysis

Line loss rate is an important economic and technical indicator of power supply enterprises, which comprehensively reflects the level of economic operation and management of power grids and the economic benefits of power supply enterprises. Line loss can be divided into statistical line loss, theoretical line loss and management line loss according to its characteristics. The statistical line loss is the actual line loss of the power grid, and its value is the difference between the power supply and the electricity sales measured by the meter, which reflects the actual loss of the power grid. The theoretical line loss is also called the technical line loss, which is obtained through theoretical calculations based on the parameters of the power supply equipment and the real-time load data of the grid operation. The theoretical line loss reflects the amount of power that the grid should theoretically consume under a specific grid structure and operating mode. The management line loss is the loss of the actual power grid excluding the theoretical line loss, and its value is the difference between the statistical line loss and the theoretical line loss. Through comparative analysis of theoretical line loss and statistical line loss, quickly find the line and station area with abnormal line loss rate, and then further analyse the potential user power abnormal behaviour in the line or station area with abnormal line loss rate.

Power loss refers to the power loss that occurs in various links such as transmission, transformation, distribution, and marketing before the power is produced and used by users. The sum of active and reactive power consumed on the line during transmission is collectively called line loss. General line loss power refers to the total power consumed by all components in the power grid within a certain period. Energy loss refers to the integration of active power over time. The specific expression is as follows:

\[
\Delta H = \int_0^T \Delta P(t) \, dt \times 10^{-3}
\]  

Taiwan area line loss includes statistical line loss, theoretical line loss, management line loss, economic line loss, etc.

2.1. Statistical line loss

The statistical line loss is the difference between power supply and sales. Its line loss calculation formula is:

\[
\lambda = \frac{\Delta A}{A_g} \times 100\% \quad \Delta A = A_g - A_s
\]
Where $\lambda$ is the line loss rate; $\Delta A$ is the line loss power, that is, the difference between the power supply and the sales amount; $A_s$ is the power supply, the amount of electricity recorded in the Taiwan area during the statistical period; $A_i$ is the sales amount, all customers during the period Electricity sales.

2.2. Theoretical line loss

Theoretical line loss refers to factors such as the performance of electrical components, the operating status of the power network, and the layout of the power grid, including fixed energy losses and line losses caused by transformer iron losses, voltmeter coils and core losses, and current meters’ Variable loss caused by loss, etc. Due to factors such as uneven load distribution and complex power network, the calculation of theoretical line loss is generally only simplified. The commonly used formula for calculating typical loss rate is

$$\Delta A = \sum_{i=1}^{n} \Delta A_i \% \sum \Delta A_i$$  \hspace{1cm} (3)

In the formula: $\Delta A\%$ represents the average loss rate of a typical Taiwan area under each capacity, which is obtained through actual measurement; $\Delta A_i$ represents the monthly power supply of the i distribution transformer low-voltage side; $n$ represents the number of groups of distribution transformers divided by capacity.

2.3. Managing Line Loss

The management line loss is the difference between the statistical line loss (actual line loss) and the theoretical line loss [1, 2]. It is the power loss due to management reasons. The management line loss is determined by the difference in the management level of the management enterprise or department of the power grid. The differences are mainly reflected in the production operation management level, enterprise operation management level, equipment maintenance management level, and energy measurement management level. This type of line loss is called Manage line loss, which can be reduced or reduced to near zero by strengthening management.

2.4. Economic line loss

Economic line loss is mainly for lines with fixed equipment conditions, and the theoretical line loss changes with the size of the power supply load, and is not a fixed value. When the line loss rate reaches the minimum, the line loss rate at this time is called economic Line loss, the corresponding current is called economic current, and the formula can be expressed as:

$$\bar{\Delta A}\% = \frac{FAP_{\text{max}}}{f}$$  \hspace{1cm} (4)

In the formula $\Delta A\%$ and $\Delta A$ represent the daily power loss rate and power loss; $f$ represents the load rate, and each unit is determined according to the change in the actual power supply load; $P_{\text{max}}\%$ represents the power loss rate at the maximum load; $F$ represents the loss factor; $A$ represents port power Transformer power supply unit: kWh.

3. Power acquisition system based on big data technology

3.1. Big data storage technology

Data storage technology saves a large amount of multiple types of data at a lower cost and provides multiple access methods to the data. The key technologies involved include: (1) HDFS distributed file system. File data is stored on decentralized low-cost storage media, providing a consistent file access
interface to the outside world, with good fault tolerance and concurrent throughput. (2) HBase columnar storage database. A database based on the column-related storage architecture on the HDFS distributed file system is mainly suitable for batch data processing and instant query. (3) Hive is a data warehouse tool based on Hadoop. It can map structured data files to a database table and provide SQL-like query functions. At the same time, it can convert query statements into Map Reduce tasks to run, and execute simple Data calculation tasks. The grid operation data in this system, such as switch, public and distribution transformer operation data, user electricity information, and other historical / massive real-time data, the amount of data is huge and needs to be stored in HDFS for distributed storage for data processing programs to be distributed Call offline calculations. Structured account data such as equipment ledger and customer files are stored in HBASE for the system to call at any time. Unstructured data such as grid equipment topology is stored in HDFS for use by the system when analyzing the distributed grid equipment topology relationship.

3.2. Big Data Power Acquisition System

Big data analysis accurately locates the cause of line loss anomaly and the design of anti-theft research technology needs to fully consider the characteristics of business data in the power industry, including: (1) support for multiple types of data storage and processing, covering structured data, semi-structured Data and unstructured data, etc.; (2) can store large amounts of data at low cost, and large amounts of data with low value density should not use traditional high-cost storage technologies; (3) the platform can flexibly expand according to user needs to support continuous The increasing amount of data, while ensuring the stability of system performance; (4) The platform should have good generality, while maintaining a certain technological advancement. Based on the above analysis, the design ideas of the open source system Hadoop platform and Spark platform are used in the design of the overall architecture. Based on the distributed file processing technology, HDFS is used as the data storage framework for big data and distributed computing is used. Technology as a data processing framework for big data. The power line loss and power stealing early warning analysis system based on the big data platform is divided into data source layer, data acquisition layer, data storage layer, data processing layer, business model layer, application display layer and typical application scenario construction. The data source layer is the source of system data, including production management PMS system (grid basic ledger and other related data), power grid GIS system (grid topology graphic data), marketing system (user files and transcribed electricity-related data), and electric energy collection system (Switch ledger and automatic collection of power index data), power consumption information collection system (users and distribution transformers automatically collect data).

The user selects specific functions for inquiry through the browser. The related functional services mainly include three parts: line loss synchronization statistics, abnormal user power consumption, and power supply quality. According to different requirements, different data are transferred from the electricity consumption information collection system to an intermediate library for processing. As shown in Figure 1 [3].
3.3. Data Mining Technology

3.3.1. Main Technologies and Business Processes of Data Mining. Data mining can do the following 7 categories: classification, valuation, prediction, correlation grouping or association analysis, clustering, description and visualization, anomaly analysis and mining of complex data types (graphic images, video, audio, etc.). Today, data has become the core of most business behaviours. To be invincible in the fierce competition, we must effectively deal with massive data and find potential information and knowledge that is useful to decision makers. The application process of data mining in business is shown in Figure 2.

1) Business needs. At this stage, you must objectively define the problem, accurately and comprehensively analyse business needs, and recognize the purpose of data mining. 2) Data understanding and preparation. According to the definition of the problem, the data types related to business objects are extracted, their sources are clarified, and the data collection and arrangement are done well. 3) Data extraction. Precisely extract all data stored in the data warehouse and pre-process the data participating in the data mining process. 4) Modelling. Different modelling methods are used for the selected data and actual applications. Generally, multiple technologies may be used, and various parameters are continuously modified until the best results are achieved. 5) Verify the model. At this stage, it is necessary to verify the correctness of the data model. The method is to input the required data into the built model and compare it with the historical data. The results can be used within the tolerance range of the error, otherwise remodelling is required. 6) Data mining. A certain data mining algorithm
is used to mine the extracted data. 7) Evaluation of results. Analyse the mining results, select the optimal model, and determine whether you can achieve your business goals. Consider multiple factors until the business problem is completely resolved [4].

3.3.2. Data pre-processing. Data pre-processing, as a pre-processing process of data mining, generally includes the following four aspects: data extraction, data cleaning, data transformation, and data induction, as shown in Figure 3.

![Figure 3. Data pre-processing.](image)

In terms of data extraction, because the marketing business archive data, periodic settlement electricity data, and energy collection data involve a large number of data tables, complex data, and large amounts of data, writing directly to the line loss intermediate library from the marketing library requires a lot of system resources, so Special programs are used for processing, which requires a large amount of work to be developed. The cycle is relatively long. The quality of the intermediate library writing in the later period also needs to be regularly maintained and checked. The application of the network and system also affects the writing of the intermediate library. Input quality, line loss analysis system also needs to be applied according to the writing of the intermediate library. Due to the practical consideration of the nature of the line loss analysis system, the data selection and monitoring cycles are for research needs. The range of data required and the monitoring cycle time are not fixed. Therefore, it is decided to establish a database link and manage the marketing business system. The database is created by establishing a specified user on the database and granting read-only permissions to a given table or view [5].

3.3.3. K-means data cluster analysis. (1) First, the user needs to specify how many classes (such as K classes) are aggregated. (2) SPSS then determines the initial class centre points of the K classes. SPSS assigns the following sample data points to K as the initial class centre point according to the actual situation of the sample data. (3) Calculate the Euclidean distance from all sample data points to K class centre points. SPSS assigns all samples to the class where each centre point is based on the shortest distance from K class centre points, forming a new K class. Complete an iterative process. The formula for calculating Euclidean distance is:

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

Where K indicates that each sample has K variables; \( x_i \) indicates the value of the first sample on the i variable; \( y_i \) indicates the value of the second sample on the i variable.
4. Application analysis of line loss anomaly data mining

4.1. Platform Construction

This system provides real-time line loss monitoring and detailed analysis functions to accurately, timely and comprehensively reflect the line loss situation. It mainly includes the functions of key unit monitoring, line loss index analysis, line loss abnormal diagnosis, problem tracking and positioning. The interface of the electricity consumption information collection and line loss analysis system is based on the B / S mode and implemented with JavaScript. The main interface is shown in Figure 4:

![Login interface.](image)

This system uses the data mining analysis method of the K-means clustering algorithm to perform the data pre-processing process. The data collected by the electricity information system is filtered to obtain data on abnormal line loss rates of lines, stations, and large users. Judging conditions to determine whether the line loss is abnormal. For example, under normal circumstances, the three-phase voltage exists and the indication is greater than zero, $U_A < 70\% U_N$. There are 5 consecutive points, or $U_B < 70\% U_N$. There are 5 consecutive points, or $U_C < 70\% U_N$. In the case of 5 consecutive points, it is under voltage of the energy meter; a phase current $I$ is under voltage of the energy meter when $2A < I < 50\%$ of the rated current and at least 5 consecutive points. Then calculate the corresponding line loss rate for the screened abnormal data, and calculate and analyse the line loss. Based on the current and accumulated actual line loss rate and the planned index value, select the line with abnormal line loss and the line loss rate for the same period. In comparison, the abnormal fluctuations of the line exceeding the line loss index are classified and screened and dynamically searched, and the classification results are evaluated [6].

4.2. Case Analysis of the Implementation Effect of Data Mining Projects in Comprehensive Management of High-loss Lines

4.2.1. Data collection. After 4 days of real-time monitoring and comprehensive statistical analysis, the western section of the line was located first, and the northern end of the main line was a high-risk suspect area. The suspected users were locked after one-on-one tracking and monitoring of key households in the area. The specific data table and graph are as follows:
Table 1. User metered load (converted to 10 kV side load according to the negative control curve meter current).

| time | A     | B     | C     | A     | B     | C     | A     | B     | C     |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0:00 | 2.87  | -2.84 | 2.84  | 12.74 | 12.74 | 12.42 | 16.7  | 22.9  | 27.4  |
| 1:00 | 0.15  | 0.1   | 0.06  | 8.64  | 8.7   | 8.51  | 9.8   | 16.2  | 14.1  |
| 2:00 | 0.1   | -0.12 | 0.06  | 12.67 | 12.8  | 12.42 | 27.6  | 34.7  | 33.5  |
| 3:00 | 0.12  | -0.14 | 0.08  | 12.67 | 12.8  | 12.42 | 7.9   | 6.8   | 6.6   |
| 4:00 | 2.91  | -2.94 | 2.76  | 8.19  | 8.26  | 7.81  | 7.7   | 7.4   | 7.1   |
| 5:00 | 2.56  | -2.64 | 2.6   | 13.5  | 13.5  | 13.12 | 6.4   | 5.8   | 5.3   |

Figure 5. Metered load curve of power users.

Curve analysis: 0 to 9 points are periods of high line loss. The east section is the actual measured total load of the branch. The total load of the branch is significantly different from the sum of the households that are connected to it. Iron smelting, the main starting time is 23:00, and 8 o’clock, and 2 households in Xie Brick Factory (drawing wire, start all day) are high-risk suspect users. Lock after further one-to-one tracking.

4.2.2. Analysis of economic benefits. A section of intermediate frequency smelting, high supply and low meter, transformer rated capacity of 630 kVA, CT ratio of 800/5, power at start-up is 860 kW, metering load is 1/4 of the actual load, and the amount of electricity under calculation is about 75%. Loss of electricity is about 6057 degrees. The main daily start time is from 11:30 pm to 8:30 the next day, and there are water cooling traces on site transformers. At 00:30 on July 20th, through the data analysis called by the remote electricity monitor, the high-voltage detection unit's primary current: \( I_a \) 53.2, \( I_b \) 55.9, \( I_c \) 54.1; the meter current: \( I_a \) 1.97, \( I_b \) 1.97, \( I_c \) 1.92. The correct current of the meter should be: \( I_a \) 8.31, \( I_b \) 8.06, \( I_c \) 8.22, (the household's power factor is 0.9, taking one-year calculation as an example).

1) The power consumption of over capacity is approximately: \((860 \text{ kVA} - 630 \text{ kVA}) \times 23.3 \text{ yuan} \times 12 \text{ months} = 64308 \text{ yuan}.

2) The cost of stealing electricity is approximately:
   - High-voltage side: \( P = 54.4 \times 10 \text{ kV} \times 1.732 \times 0.9 = 848 \text{ kW} \)
   - Electricity meter: \( P = 220 \times 1.95 \times 3 \times 160 \times 0.9 / 1000 = 186 \text{ kW} \)
   - High-voltage side-electricity meter = 848 - 186 = 662 kW
   - Electricity theft = 662 x 9 hours’ x 30 days’ x December x 0.6011 yuan = 1,289,287 yuan

3) Over-capacity electricity + stealing electricity cost = 64308 yuan + 1289287 yuan = 1353595 yuan
The overcapacity power consumption and transformer loss cause serious power loss and cause great economic losses. The internal sampling part of the meter has been modified. The nature of the device is poor. Other equipment cannot be found, which seriously affects the meter's measurement and loses a large amount of power. Big economic loss [7].

5. Conclusion
In this paper, based on the research on line loss management of power consumption information collection and line loss analysis system, starting from solving the current situation of line loss management, and combining with the construction of power consumption information collection system, a line loss abnormality based on the power consumption information collection system is constructed Data mining software and hardware platform. Real-time analysis of abnormal line loss factors of line loss, station line loss, and large user line loss, screen out lines, stations, or large users with abnormal line loss according to the line loss rate, and analyse these abnormal factors. Due to the inadequacy or inaccuracy of the power information collection system, install data collection terminals on the lines, stations or large users to enable them to collect enough data for analysis, thereby real-time data collection and power consumption Real-time monitoring of information acquisition and line loss analysis system.

Acknowledgments
This work is supported by the State Grid Ningbo Power Supply Company's innovative consulting project “Research on Accurately Locating Line Loss Anomalies and Anti-Theft Based on Big Data Analysis”.

References
[1] Sekizaki, S., Nishizaki, I., & Hayashida, T. Analysis of electricity market model with demand response in distribution network. Ieej Transactions on Electronics Information & Systems, 135(3) (2015) 292-303.
[2] Zhao, Qingzhan, Jin, Guangcai, Zhou, Wenjie, Wang, Chuanjian, & Dai, Jianguo. Information collection system for diseases and pests in cotton field based on mobile gis. Transactions of the Chinese Society of Agricultural Engineering, 31(4) (2015) 183-190.
[3] Pengcheng SONG, Zheng XU, Huanfeng DONG, Hui CAI, & Zhenjian XIE. Security-constrained line loss minimization in distribution systems with high penetration of renewable energy using upfc. Journal of Modern Power Systems & Clean Energy, 5(6) (2017)876-886.
[4] Richard James Crilly. Medical radiation dosimetry—theory of charged particle collision energy loss by brain mparland. Journal of Radiation Oncology, 4(1) (2015)103-103.
[5] Yanlu Huang, Yan Li, Feng Chen, Xu Zheng, & Jing Tang. A modified back/forward sweep method based on the electricity consumption data. Energy & Power Engineering, 09(4) (2017)176-182.
[6] K. Jia, C. Gu, T. Bi, H. Wei, & Q. Yang. Fault characteristics and line protection within the collection system of a large-scale photovoltaic power plant. Diangong Jishu Xuebao/transactions of China Electrotechnical Society, 32(9) (2017)189-198.
[7] Manish Kumar, K.S. Sandhu, & Ashwani Kumar. Wind speed variation impact on transmission loss reduction in electricity market. Procedia Computer Science, 70(1) (2015)526-537.