Classification of short-term loads of enterprises using LightGBM

Meiwei Hao, Yuting Tian, Jing Gao, Yin Wang and Runze Tian

State Grid Tianjin Power Information & Telecommunications Company, Tianjin 300100, China
No.153, Kunwei Road, Hebei District, Tianjin City, P.R. China

Abstract: As the society becomes increasingly energy-conscious these days, enterprises not only make a point of continuous supply for power, but try to improve power utilization efficiency and design scientific power use plans. In this study, the LightGBM algorithm was used to classify the short-term loads of enterprises, and the characteristics of short-term loads were analyzed to realize accurate classification. The research results will be instrumental to the enterprises’ scientific scheduling of power use, thereby improving safety and reliability of the power grid, and providing a strong basis for future grid planning.

1. Background

As demand for electricity increases, electric power plays an increasingly important role in energy consumption. Compared with household power demand, demand for power in enterprises is increasing exponentially. In these years, as the world goes more energy-conscious, enterprises not only pay attention to ensuring continuous power supply, but make it a priority to improve power use efficiency and design scientific power use plans. As the number of electric facilities and orders of an enterprise changes, the power load of the enterprise fluctuates without following a certain pattern, thereby affecting production and maintenance plans or even leading to economic losses.

These years, classification of power load in enterprises is becoming diverse and complicated. In [1], a kernel-based Fuzzy C-means clustering algorithm was designed, and by introducing the idea of layer-by-layer classification, the samples were divided into several sub-spaces which were accurately classified then. In [2], a load classification method based on the load characteristic indicators was proposed. This method classified the load by the characteristic indicators and analyzed the characteristics of load in each class. In [3], a residential power load recognition system was developed by improving the KNN algorithm and the SVM; by optimizing the KNN algorithm, the residential power loads were classified, and to solve the problem of inseparability of linearity of the samples, the SVM was used to displace part of the optimized KNN classifier.

In this paper, a LightGBM-based enterprise short-term load classification method was proposed. The LightGBM algorithm was used to classify the short-term loads and the short-term load characteristics were analyzed to accurately classify the loads, facilitate design of scientific production and power use plans, thereby improving the power efficiency of enterprises.

2. Overall design structure

According to the periods of peaks, bottoms and usual conditions of local power use, proportional calculation was conducted to classify the loads into three types: peak loads, intermediate-flat loads,
bottom loads. The LightGBM algorithm was proposed to train the short-term load data of enterprises, the correlation between the load data and the different power use periods was analyzed to learn and position the load classification and improve classification accuracy. The framework is as follows:

![Diagram](image)

Figure 1. Design plan of the short-term load classification of enterprises

3. Algorithm implementation

3.1. Relevant technologies

LightGBM is a gradient boosting algorithm based on decision trees and has achieved good performance in many machine learning tasks. The LightGBM algorithm has many advantages: fastness, high efficiency, low resource consumption and compatibility with parallel processing. The decision-tree-based LightGBM algorithm was used in this study to analyze short-term load data of enterprises to classify the loads, thereby realizing scientific planning of power supply [4].

3.1.1. Theoretical foundation. Gradient boosting tree is a boosting tree method that optimizes the loss function using fast descent method. Boosting tree is a boosting method that uses the classification tree or regression tree as the basic classifier, and uses the additive model and the forward step algorithm. The boosting tree model can be expressed as the additive model of decision trees [5].

$$f_m(x) = \sum_{n=1}^{m} f(x; \Theta_n)$$

where m is the number of decision trees, $T(x; \Theta_m)$ is the decision tree, and $\Theta_m$ is the parameter of the decision tree.

The forward step algorithm determined the initial boosting tree $f_0(x) = 0$, and the m-th step model was

$$f_m(x) = f_{m-1}(x) + T(x; \Theta_m)$$

If the given model was $f_{m-1}(x)$, the loss function to be solved was:
The $\hat{\Theta}_m$, i.e. the parameter of the m-th tree, could be obtained.

### 3.1.2. LightGBM

LightGBM is a novel boosting model developed by Microsoft in 2015. Though it does not perform as well as XGBoost in terms of the training time, it provides two new techniques: gradient-based one-side sampling (GOSS) and exclusive feature building (EFB) [6]. As for GOSS, the sampling sites of large gradients are found to play an important part in the computation of the information gain. That is to say, sampling sites with large gradients contribute more information gains, so to ensure accurate evaluation of information gains, sampling sites with larger gradients should be selected in down-sampling, while proportional random sampling should be conducted on sampling sites of smaller gradients. EFB binds mutually exclusive features to reduce the number of features. To optimize binding of mutually-exclusive features, a greedy algorithm was proposed to obtain a good approximation ratio. Therefore, the number of features could be reduced without decreasing the site segmentation accuracy.

Moreover, LightGBM uses the histogram-based decision-tree algorithm, and by subtracting the histogram of the parent node of the leaf node and that of the neighboring node in the binary tree, the histogram of the corresponding leaf node is obtained, and the expenditure is reduced via the histogram algorithm. The Lead-wise strategy is used to grow the tree, and the leaf node of the largest loss is grown, and the depth of the tree is limited to avoid overfitting. Regarding parallel optimization, LightGBM provides three ways, i.e. features, data and voting, to optimize the model [7].

### 3.2. Empirical analysis

Empirical research was conducted with load data from an enterprise and the performance of the model was analyzed accordingly.

#### 3.2.1. Data extraction and processing

The dataset contained historic load statistics for about a year. The 96 pieces of load data per day reflected the power use conditions of the enterprise and involved 112 indicators including the holidays, shift system, the business volume. The load data were divided into three groups by characteristics to provide support for classification of short-term loads of the enterprise. The dataset is as follows:

| Table 1. Indicators of short-term load classification of the enterprise. |
|-----------------|-----------------|-----------------|-----------------|
| Column Name     | Meaning         | Type            | Instance        |
| DEVICE_ID       | Device number   | object          | 08089927957     |
| DATA_DATE       | The data of time| date            | 2019-01-03      |
| DATE_TYPE       | The data of type| object          | Weekday         |
| P1              | Load data       | float           | 198.3           |
| P2              | Load data       | float           | 278.2           |

Due to loss and redundancy of the original data, to ensure data availability and reliability, pre-treatment of data was performed. Widely-used data processing methods include normalization, one-hot coding, and sequencing. After data pretreatment, the proportion of the load data was adjusted, and the data were extracted proportionally to form a training dataset and a test set.

#### 3.2.2. Model performance analysis

Through repetitive experiment and grid searching of LightGBM parameters, the optimal model with the best performance was achieved. The optimal parameters of the LightGBM classifier are as shown in Table 2.
Table 2. Parameters of the enterprise short-term load classification model.

| Model Parameter       | Meaning                        | Values |
|-----------------------|--------------------------------|--------|
| BOOSTING_TREE         | The type of base learning      | GBDT   |
| LEARNING_RATE         | Learning rate                  | 0.1    |
| MAX_DEPTH             | The deep of tree               | 10     |
| N_ESTIMATORS          | The number of base learning    | 100    |
| MIN_SPLIT_GAIN        | The minimum gain for performing sharding | 0.0 |
| NUM_LEAVES            | The number of leaves on the tree | 31    |

When building the decision tree, the information gain of the computation features was required to achieve the weight of these features, the larger information gains were selected to split the nodes. The weights of the features were sequenced when the model was trained, and the top 10 features of the model and corresponding scores are as shown in Fig. 2.

![Fig. 2. Sequencing of the feature weights](image)

After training, the model was assessed and the ROC curve was considered. The ROC curve serves two major purposes: 1) to evaluate the effect of one or several indicators on two classifications; 2) to find the optimal threshold values of indicators to obtain the best classification result. It was found that the AUC of the final model was 0.91. The ROC curve is shown in Fig. 3.

![Fig. 3. ROC curve of enterprise short-term load classification](image)
The LightGBM model showed good performance in the training for enterprise short-term load classification. When the test data was introduced, the classification performance remained good. Later, more models will be analyzed to optimize the enterprise short-term load classification model.

4. Conclusions
To summarize, the enterprise load data and the LightGBM algorithm were used to predict the enterprise load classification. By studying the short-term load data, analyzing the indicators including the holidays, seasons, and trading, the proposed model accurately classifies the enterprise short-term loads and realizes rational planning of power use in the enterprise to improve stability and security of the local power supply. Accurate and stable load classification can help enterprise to balance the power use, thereby improving security and reliability of the grid operation and providing better support for the power grid planning [8].

References
[1] Yanhui Xu, Lanyu Zhang, Ge Song. Application of kernel based fuzzy c-means hierarchical clustering algorithm in load classification [J]. Power construction, 2015,36 (04) : 46-51.
[2] Zhimei Que, Qiang Yu, Qihui Xie, Jiyun Chi. A load classification method based on load characteristic index [J]. East China electric power, 2014,42 (11) : 2382-2387.
[3] Ran Liu. Research on residential electrical load identification based on improved nearest neighbor method and support vector machine [D]. Chongqing university, 2014.
[4] Xiaojun Ma, Jinglan Sha, Xueqi Niu. Design and application of P2P project credit rating model based on LightGBM algorithm [J]. Journal of quantitative economy,35 (05) : 144-160.
[5] Boliang Fan, Feng Gao, Peng Kou. Online Boosting regression algorithm and its application in load prediction of energy-intensive enterprises [J]. Information and control, 2014,43 (06) : 750-756.
[6] Yong Xie, Wei Xiang, Mengzhong Ji, Jun Peng, Yihuai Huang. Application analysis of housing monthly rent prediction based on Xgboost and LightGBM algorithm [J]. Computer application and software, 209,36 (09) : 151-155 + 191.
[7] Siyu Wang, Jianping Chen. Research on credit risk assessment model based on LightGBM algorithm [J]. Software guide, 209,18 (10) : 19-22.
[8] Pengjuan Liu. Research on short-term load prediction in the environment of intelligent distribution network [D]. Lanzhou university of technology, 2018.