Power User Sensitivity Analysis and Power Outage Complaint Prediction

Jiahan Ding¹*, Yunchao Shi¹, Ruiqian Zhu¹, Xiaoxiong Wei¹, Bo Chen¹, Jianfeng Yu²
¹State Grid Zhejiang Marketing Service Center, 310000, Hangzhou Zhejiang, People R China.
²State Grid Zhejiang Electric Power Co., LTD, 310000, Hangzhou Zhejiang, People R China.
*Corresponding author e-mail: 17622811135@sgzjm.org

Abstract. The sensitivity of power users is an important basis for improving the quality of power customer service and refining customer service content. In order to improve the accuracy of power user sensitivity classification, this paper optimizes and improves the decision tree algorithm based on the ant algorithm, builds a power user sensitivity analysis model, and verifies its effectiveness with simulation experiments, which provides a more powerful data reference for improving the quality of power user services and other tasks. Based on the analysis of the sensitivity of power users, combined with the analysis of related characteristic data of power outages, this paper predicts the probability of power user outage complaints, hoping to provide data reference for improving power user satisfaction and reducing power outage complaints.

Keywords: Sensitivity analysis; complaint prediction; decision tree; ant colony algorithm

1. Introduction
In recent years, with the continuous development and advancement of my country's power system reform, the requirements for the service level of power customers have also increased. Power user sensitivity is an important data support for improving the service level of power customers. By mastering user sensitivity, it is possible to formulate distribution network architecture optimization and other related plans more rationally and targeted, and to refine the service content of power users.

The current research on the sensitivity of power users is mainly based on building a calculation model for analysis. The calculation and analysis methods used include logistic regression, entropy and so on. In order to improve the accuracy of power user sensitivity classification, this paper optimizes and improves the decision tree algorithm based on the ant algorithm, thereby constructing a power user sensitivity analysis model [1-3].

2. Data preprocessing
2.1. Data screening
The feedback information of power-related services in my country currently mainly comes from the customer service call system, but the number of power users in my country is very large, and the power customer service call data information of each user is quite different, and the distribution of most user-related data is uneven. Therefore, considering the specific situation of the characteristics of power users in my country, this article selects the data of a power supply bureau's marketing management system from 2018 to 2019 for analysis in specific research. In addition, the collected high-dimensional original business data contains many attributes that have no or little relevance to the sensitivity of power users, and there may also be some duplicate data, incomplete data, invalid data, etc. Therefore, it is necessary to filter the original data to ensure the efficiency of analysis and prediction [4-6].

2.2. Data standardization
The filtered data still has differences in dimensions and magnitudes, which need to be standardized to avoid the occurrence of excessively large variables due to irregular data. The standardization methods mainly include maximum-minimum standardization, Z-score standardization, and decimal standardization. Considering the advantages of Z-score standardization in multi-dimensional data processing, this paper chooses to use the Z-score standardization method based on the mean and standard deviation of the original data. The calculation formula is as follows:

$$x' = \frac{x - \bar{x}}{s}$$  \hspace{1cm} (1)

After data standardization processing, a new combined data set including several variables is generated, including 1 label variable. The label variables are 0 and 1, respectively indicating insensitive and sensitive.

3. Power user sensitivity analysis based on improved decision tree

3.1. Principle of decision tree
The tree model is a widely used machine learning model, and the decision tree is its core classifier. It has two basic structures: a binary tree and a polytree, as shown in Figure 1.

![Trinomial tree structure](image1.png)  \hspace{1cm} ![Polymetric tree structure](image2.png)

(a) Trinomial tree structure  \hspace{1cm} (b) Polymetric tree structure

**Figure 1.** Basic structure of decision tree

The decision tree represents the mapping relationship between object attributes and object values. In data mining, the decision tree is an analysis and prediction model.

Decision tree classification starts from the root node. Each node represents an object. During classification, it means that a variable in the data is compared with a feature node; each branched branch represents a possible attribute value. During classification, it represents a variable and the result of feature node comparison; each leaf node represents the value of the object corresponding to
the path from the root node to this node, and the final result is the leaf node that represents a certain type of distribution during classification [7-9].

The whole decision tree classification process consists of two parts: building regression tree and pruning. The establishment of classification regression tree is as follows:

For a data set D containing m attributes, the attribution category of the data is C, and the sample belongs to category C (1) is P, then:

$$Gini(A) = 1 - \sum_{i=1}^{m} P_i^2$$  \hspace{1cm} (2)

After attribute A is decomposed, the formula for Gini is:

$$Gini_{,A}(D) = \frac{|D_1|}{D} Gini(D_1) + \frac{|D_2|}{D} Gini(D_2)$$ \hspace{1cm} (3)

The model needs to have low bias and variance, usually a compromise between bias and variance, so in order to improve the classification effect, it needs to be pruned:

$$\hat{\mathcal{C}} = \frac{R(t) - R(T_i)}{|N_{T_i}|-1}$$ \hspace{1cm} (4)

In the formula, $R(T_i)$ is the error cost of $T_i$, $R(t)$ is the error cost of pruning $t$, and $|N_{T_i}|$ is the number of leaf nodes. When there are multiple non-leaf nodes, delete the node with the largest value of $|N_{T_i}|$.

3.2. Ant colony algorithm

The ant colony algorithm is an algorithm that simulates the process of ants looking for food to find an optimal path. Its mathematical model is as follows:

(1) Before starting, parameters such as task array and node array need to be initialized. Suppose the number of ants $Y(0)$ included in the target ant colony is M, the maximum number of iterations is maxgen, and the number of nodes is n. $fit_i$ is the fitness of the i-th ant. Find the fitness function:

$$fit(y) = [y_1, y_2, \ldots, y_M]^T$$  \hspace{1cm} (5)

$$y_j = \frac{1}{m} \sum_{j=1}^{m} (O_j - T_j)^2$$  \hspace{1cm} (6)

$O_j$ is the predicted output value of the j-th sample, and $T_j$ is the predicted output value of the j-th sample.

(2) $\tau_{ij}$ is the pheromone between i and j, $\tau_{ij}(0)$ is initially C; $\eta_{ij}$ is the heuristic information between i and j, $\eta_{ij}(0)$ is initially C; Q is the normal quantity, $\Delta_{\tau_{ij}}$ is the pheromone of the kth ant on (i,j). Find pheromone:

$$\tau_{ij}(t+n) = (1-p)\tau(t) + \Delta_{\tau_{ij}}$$  \hspace{1cm} (7)

$$\Delta_{\tau_{ij}} = \sum_{k=1}^{m} \Delta \tau_{ij}$$  \hspace{1cm} (8)
(3) The ants move according to the pheromone and the selection function, $P_{ij}^k(t)$ is the probability of the kth ant moving from i to j:

$$P_{ij}^k = \frac{[\tau_{ij}(t)]^\alpha[\eta_{ij}(t)]^\beta}{\sum_{s \in J_k(i)}[\tau_{is}(t)]^\alpha[\eta_{is}(t)]^\beta}, j \in J_k(i); P_{ij}^k = 0, j \not\in J_k(i)$$

(10)

$\alpha$ is the weight of the pheromone, $\beta$ is the weight of the heuristic factor, and $J_k(i)$ is the node selectable by ant k.

Record the location and path of the ants, and select the optimal path.

When the maximum number of iterations is reached or the error threshold is reached, the iteration is stopped and the result is output, otherwise the loop continues.

3.3. Improved decision tree based on ant colony algorithm

In order to improve the accuracy of power user sensitivity classification, this paper uses ant algorithm to optimize and improve the decision tree algorithm, which specifically includes the discretization of attributes, heuristic information, and pheromone.

(1) Discretization of attributes:

Take the minimum value MIN of the attribute and the maximum value MAX of the attribute, and then divide them into N equal parts:

$$A_i = MIN + (MAX - MIN) / N \ast i, i = 1, 2, ..., N.$$  (11)

Take the largest point of $A_k$ and divide the attributes into $[MIN, A_k]$ and $[A_k, MAX]$. Repeat segmentation until the optimal global threshold is obtained.

(2) Combining pheromone and information gain as heuristic information, the information gain is expressed as:

$$GainRatio(\varphi) = \frac{Gain(\varphi)}{SplitInfo(\varphi)}$$  (12)

The pheromone is $[edge_j, level, x_i]$, the value range is $[\tau_{\text{min}}^\text{edge}, \tau_{\text{max}}^\text{edge}]$, and the pheromone update method is as follows:

$$\tau(E, L, x_i) = \begin{cases} \rho \tau(E, L, x_i), & \text{if } (E, L, x_i) \not\in \text{tree}_{ib}; \\ \rho \tau(E, L, x_i) + \Delta \tau(\text{tree}_{ib}), & \text{if } (E, L, x_i) \in \text{tree}_{ib} \end{cases}$$  (13)

$\rho$ is the volatilization parameter and $\text{tree}_{ib}$ is the i-th local optimal solution.

Combine pheromone and information gain as heuristic information to generate a decision tree. Ants select attribute nodes based on heuristic information and pheromone, and the probability $P_i$ of selecting nodes $x_i$ is: $\forall$
\[ \rho_i = \frac{\tau(E, L, x_i)\eta_i}{\sum\tau(E, L, x_i)\eta_i}, \quad \forall i \in F \]  

(14)

3.4. Simulation experiment

3.4.1 Algorithm verification

In order to verify the effectiveness of the improved decision tree model proposed in this paper, some data in the marketing system of a power supply bureau were selected for statistical and simulation comparison experiments. The experimental results are shown in Figure 2.

![Figure 2. Simulation and comparison experiment results](image)

It can be seen from the results of simulation and comparison experiments that compared with support vector machines (SVM) and traditional decision tree algorithms, the improved decision tree algorithm based on ant colony algorithm proposed in this paper has good classification accuracy, which can prove the method’s effectiveness.

3.4.2 Sensitivity analysis of power users

In this paper, the resident users in a certain area are taken as the object, the corresponding power supply data is selected as the sample data, and the improved decision tree method based on the ant colony algorithm is used to conduct a comparative experiment on the sensitivity analysis of power users. The experimental evaluation criteria include judgment accuracy \( P \), judgment coverage rate \( R \), harmonic mean \( F \), false positive rate \( FPR \) and false negative rate \( FNR \). The experimental results are shown in Table 1.

|                  | ACO-C4.5 | SVM  | Traditional decision tree |
|------------------|----------|------|---------------------------|
| \( P \)          | 94.29    | 1.38 | 55.28                     |
| \( R \)          | 93.21    | 68.25| 1.36                      |
| \( F \)          | 93.75    | 26.82| 1.72                      |
| \( FPR \)        | 0.1      | 99.3 | 0.01                      |
| \( FNR \)        | 6.73     | 31.28| 99.20                     |

Table 1. Comparison results of classification effect

According to the data shown in Table 1, it can be seen that compared with support vector machines (SVM) and traditional decision tree algorithms, the improved decision tree method based on ant colony algorithm for power user sensitivity analysis not only has a high recall rate, but also both the false positive rate and the false negative rate are relatively low.
4. Forecast of power outage complaints by power users

After statistically analyzing the data of power user complaints in a certain area, this article found that complaints related to power outages, such as frequent planned outages, early and delayed power transmission, accounted for about 45% of the total complaints. In addition to the complaints generated during the power outage, there are also some complaints generated during the period of time after the power is restored, including complaints about customer service attitudes, complaints about dissatisfaction with outage explanations, etc., the specific statistical results are shown in Figure 3.

![Figure 3. Complaint interval distribution after power recovery](image)

As can be seen from Figure 3, about 80% of the related complaints that occurred after the power was restored in this area were concentrated in the first 10 hours after the power was restored. Based on this situation, this paper screened out the non-standard data set, selected the power outage data in the area in the past two years as the training set, and determined whether there were related complaints in the first 10 hours after the power was restored as a label, and determined other relevant feature values. Forecast the power outage complaints of residents in this area. The prediction results are shown in Table 2.

| No. | Improved decision tree | Support vector machine | Decision tree |
|-----|------------------------|------------------------|--------------|
|     | Recall     | F1            | Recall     | F1            | Recall     | F1            |
| 1   | 93.26%     | 95.40%        | 84.27%     | 87.21%        | 87.56%     | 84.21%        |
| 2   | 87.23%     | 89.13%        | 75.53%     | 84.02%        | 80.34%     | 82.49%        |
| 3   | 91.21%     | 89.25%        | 90.11%     | 90.61%        | 80.78%     | 79.57%        |
| 4   | 91.30%     | 92.31%        | 85.87%     | 86.34%        | 87.67%     | 82.49%        |
| 5   | 93.62%     | 95.65%        | 85.11%     | 88.40%        | 88.67%     | 83.61%        |
| **Average** | **91.62%** | **92.06%** | **83.31%** | **86.20%** | **84.27%** | **81.46%** |

According to the results shown in Table 2, it can be seen that the improved decision tree algorithm proposed in this paper shows good performance in the prediction of power outage complaints and
meets the requirements for prediction of power outage complaints.

5. Conclusions
This paper studies the classification model of power customer sensitivity. Starting from the need to improve the performance of the algorithm, a new classification model and calculation method are proposed, and the validity is confirmed through algorithm verification experiments [10].

In addition, based on the analysis of the data, this article finds that a period of time after power is restored is a period of high incidence of power outage-related complaints, and based on this situation, the characteristics of the power outage complaint forecast data are refined. The comparison of experimental results proves that the improved decision tree proposed in this paper can meet the requirements of power outage complaint prediction and has better performance.

References
[1] Yang Xiaxia, Su Feng, Huang Xuxia. Research on Imbalanced Data Classification Method Based on Improved Random Forest Algorithm. Network Security Technology and Application, 2020(10): 70-71.
[2] Chen Ziyang, Liu Chen, Chen Xiaofeng. Research on Sensitivity of Electric Power Customer Complaints Based on Big Data Mining. China New Telecommunications, 2019, 21(19): 139-140.
[3] Shi Pu, Zhang Gangshuai, Zhao Yaomin, Huang Hui, Li Jinyu. Analysis of the Sensitivity of Power Users and Prediction of Power Outage Complaints. Power Supply, 2019, 36(10): 67-72.
[4] Lv Pengpeng, Tao Xiaofeng, Xu Zhiguang, Miao Ping, Xiong Xia, Bi Shanyu. Power Customer Outage Sensitivity Prediction Based on Bayesian Network. Electric Power Engineering Technology, 2019, 38(05): 49-54.
[5] Wang Li. Research on the Countermeasures to Improve the Service Level of Guangdong Power Supply. Electrical Technology, 2019, 20(07): 74-79.
[6] Xie Shengxiang, Zhang Huishi. Research and Application of Power Customer Attribute Mining Based on Big Data Technology. Electric Power and Energy, 2018, 39(06): 849-852+860.
[7] Zou Yuelin, Zhang Longjun, Liu Kun. Application Research of Big Data Analysis in Key Business Monitoring of Marketing and Power Purchase. Electronic World, 2018(24): 79-80.
[8] Lu Qiang. Differentiated Services for Power Customer Complaint Behavior Analysis. Communication World, 2018, 25(12): 151-152.
[9] Zhao Shaodong, Wang Chengsi. Research on Comprehensive Evaluation Model of Power Customer Sensitivity Based on Entropy Method. Electrical Engineering Technology, 2018(11): 43-45+48.
[10] Zhu Yingchun. Analysis of Emergencies and Countermeasures in Power Supply Service. Communication World, 2014(16): 112-113.