Power Failure Sensitivity Analysis via Grouped $L_{1/2}$ Sparsity Constrained Logistic Regression

Baoshu Li[1,*], Xin Zhou[1], and Ping Dong[1]
[1] Jiangsu Electric Power Information Technology Co., LTD, Nanjing 210000, China
[e-mail: cs_baoshuli@163.com, cs_xinzhou@163.com, cs_pingdong@163.com]
[*]Corresponding author: Baoshu Li

Abstract

To supply precise marketing and differentiated service for the electric power service department, it is very important to predict the customers with high sensitivity of electric power failure. To solve this problem, we propose a novel grouped $l_{1/2}$ sparsity constrained logistic regression method for sensitivity assessment of electric power failure. Different from the l_1 norm and k-support norm, the proposed grouped $l_{1/2}$ sparsity constrained logistic regression method simultaneously imposes the inter-class information and tighter approximation to the nonconvex l_0 sparsity to exploit multiple correlated attributions for prediction. Firstly, the attributes or factors for predicting the customer sensitivity of power failure are selected from customer sheets, such as customer information, electric consuming information, electrical bill, 95598 work sheet, power failure events, etc. Secondly, all these samples with attributes are clustered into several categories, and samples in the same category are assumed to be sharing similar properties. Then, $l_{1/2}$ norm constrained logistic regression model is built to predict the customer's sensitivity of power failure. Alternating direction of multipliers (ADMM) algorithm is finally employed to solve the problem by splitting it into several sub-problems effectively. Experimental results on power electrical dataset with about one million customer data from a province validate that the proposed method has a good prediction accuracy.

1. Introduction

$\mathbf{W}$ith the rapid development of smart grid construction, power companies have accumulated a large amount of business data during the process of production and operation. With the support of data mining and machine learning techniques, such as statistical learning, regression, classification, and other algorithms, hidden information from a large amount of business data can be deeply explored, which highly increases the accuracy of data utilization. Besides, it also provides support for market decision-making and helps power companies provide safe, stable and convenient power services [1].

Although the service quality of power companies has made great progress, the demand for electricity from customers has also been increasing. Some power failure sensitive customers have put forward stricter requirements on the reliability of power supply because the power failure will bring them enormous economic losses and also hurts the reputation of the power supply company [2]. The power failure sensitive customers indicate the customers who pay more attention to power failure through multiple channels during the process of power supply services. As a key medium for the interaction between customers and power companies, the inbound volume of the 95598 service hotline has grown rapidly. Since customers' demands are mainly concentrated in the areas of blackout and repair, the studies of customer failure sensitivity should be launched from the collected data. By analyzing the behavior characteristics of different power failure sensitive customers, it is useful to adopt machine learning, data mining and other techniques to forecast the actual electricity demand of users, and the power services quality can be improved. In this way, customer satisfaction will be improved, and the 95598 workloads will be decreased, as well as the complaints of customers [3].

Aiming to evaluate the customer's power failure sensitivity scientifically, the research of customer power failure sensitivity takes whether the customers call 95598 for power failure issues consultation as the target variable to construct the prediction model. In the field of data mining and machine learning, the models employed for prediction mainly include logistic regression [4][5], decision trees [6], particle swarm optimization (PSO) [7][8], neural networks [9], nearest neighbor classifiers [14], etc. Classic decision tree models developed a tree structure to establish the mapping between sample attributes and sample categories, then iterates from the root node to a certain leaf node to achieve classification prediction. The logistic regression model builds a linear model of the independent variables, performs binary classification prediction on the target variable, and employed the maximum likelihood method to learn the model parameters. Fast calculation and low data quality requirements are the greatest advantages. Currently, it is wildly used in big data analysis, machine learning, etc. Whereas, when the number of independent variables used in the model is too large, the logistic regression model is prone to overfitting. Employing regularization constraints (e.g., l_2 norm [15] and l_1 norm [16][17]) for parameters is an effective way to solve this problem. When the l_1 norm is employed as the regularization constraint, it is a sparse logistic regression model [18], less important factor variables can be selected for prediction. However, subsequent researches have shown that there is a problem intergrowth with the l_1 norm named over-shrink [19], i.e., it is too easy to shrink predictors to zero, only a single factor is selected from multiple strongly correlated predictors for prediction, and the rest are discarded. At the same time, the l_1 norm is not the most compact convexity of the l_0 norm in the bounded area of the unit ball. Later, the k-support norm is presented [20], pointing out that the k-support

norm is the most compact convexification of the l_0 norm in the bounded region of the Euclidean unit sphere. Therefore, a more effective sparseness regularization effect can be formed. At the same time, the bounded constraint in the k-support norm is also helpful to alleviate the over-shrink problem of the l_1 norm. Even k-support norm achieves good performance for the prediction, however, it is only a convex approximation to the l_0 norm, there is still a big room to improve accuracy. To date, the $l_{1/2}$ norm has drawn enormous attention in the field of sparse representation [21][22] because it can achieve a more accurate approximation to the l_0 norm among all l_p ($p < 0 < 1$) norms and it has an analytical solution.

To address these challenges in sensitivity prediction of power failure customers, a novel grouped $l_{1/2}$ sparsity constrained logistic regression model is proposed for solving the problem. As illustrated in **Fig. 1**, possible factors related to customer power failure sensitivity problems are collected to form a sample data set firstly. Then those samples are clustered into several categories, and $l_{1/2}$ sparsity constrained logistic regression model based on failure-sensitive predictions is proposed. The factors in the regression model are selected adaptively, and the most important variable factors for the model prediction are calculated by the dominance analysis method. Eventually, the proposed model is verified and evaluated on the dataset of a provincial power grid company with nearly a million customers. The experimental results show that our proposed method accurately achieves the high sensitivity to power failures customers identification. Customer complaints caused by power failure problems are reduced and overall customer satisfaction is highly improved.
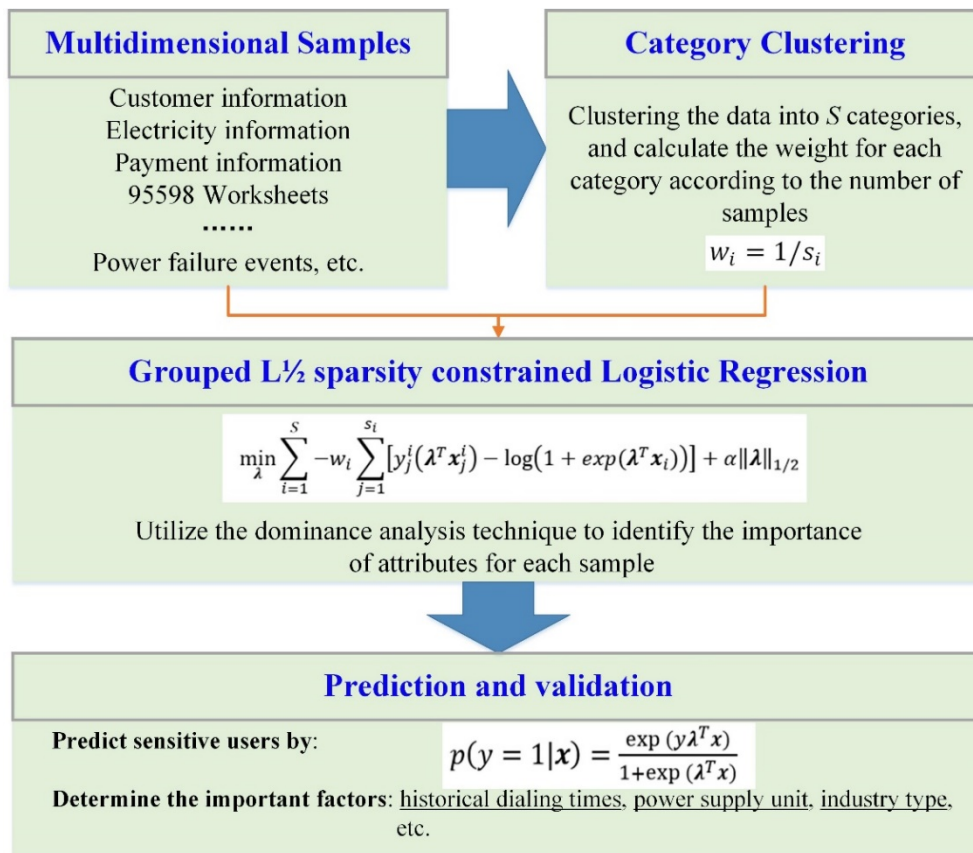

Fig. 1. Flow chart of the proposed method.

Specifically, the contributions of the proposed method can be summarized as follows.
- First, the samples are divided into several categories, and the samples of the same category will be multiplied by a weight in the model, thereby effectively alleviating the problem of sample imbalance during prediction.
- Secondly, the grouped $l_{1/2}$ norm constraint is added to the model to predict attributes related to the test sample, effectively improving the accuracy of the sparse regression prediction model.
- Finally, experiments verify the effectiveness of the proposed method, which has more accurate prediction accuracy than those models with l_1 norm and k-support norm.

The remainder of the paper is organized as follows. Section 2 addresses the related work. The grouped $l_{1/2}$ sparsity constraint logistic regression method is formulated in section 3. Section 4 reports the experimental results and discussion. Finally, section 5 draws the conclusion.

2. Related Work

In this section, we will recall the related works of sparsity constrained logistic regression method for power failure sensitivity analysis. First, we give a brief introduction of sparse logistic regression, then we introduce l_1 norm-based sparse logistic regression method and K-support based sparse logistic regression method.

2.1 Sparse Logistic Regression

To address the issues clearly, we first make some notations. The samples in the data set Ω are divided into a training data set and a test data set, the logistic regression model is trained by the cross-validation strategy. The training data set is denoted as $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots (\boldsymbol{x}_l, y_l)\}$, where l is the number of training samples, where $\boldsymbol{x}_i = [x_1^{(i)}, x_2^{(i)}, \cdots x_d^{(i)}] \in R^d$ is the vector for the i-th sample which includes d elements, y_i is the sample category label corresponding to the i-th sample x_i, when the sample indicates a power failure sensitive customer, the value of y_i is 1; otherwise, the value is 0.

According to the above notations, a regression model is established, and the multivariate linear relationship between the target variable y and a series of factor variables $\boldsymbol{x} = [x_1, x_2, \ldots, x_d] \in R^d$ is formulated as follows:

$$y = \lambda_0, \lambda_1 x_1 + \cdots + \lambda_d x_d \tag{1}$$

where $\boldsymbol{\lambda} = [\lambda_0, \lambda_1, \cdots, \lambda_d] \in R^{d+1}$ is the regression coefficients that need to be solved. According to "logistic regression" model, the probabilistic decision rule can be expressed as follows.

$$p(y|\boldsymbol{x}) = \frac{\exp(y\boldsymbol{\lambda}^T \boldsymbol{x})}{1 + \exp(\boldsymbol{\lambda}^T \boldsymbol{x})} \tag{2}$$

Then the log-likelihood is

$$\mathcal{L}(\boldsymbol{\lambda}) = \log \prod_{i=1}^{l} p(y_i|\boldsymbol{x_i}) = \sum_{i=1}^{l} (y_i \boldsymbol{\lambda}^T \boldsymbol{x_i} - \log(1 + \exp(\boldsymbol{\lambda}^T \boldsymbol{x_i}))) \tag{3}$$

Now, parameter $\boldsymbol{\lambda}$ can be solved by maximizing the log-likelihood function. However, when

the model has too many factor variables, that is the dimension d is larger than a given threshold, the model (1) learned by the maximum likelihood method often cause overfitting. To alleviate this problem, we usually employ sparsity constrained regularization terms, that is,

$$\min_{\lambda} \sum_{i=1}^{l} -\left[y_i(\lambda^T x_i) - \log(1 + exp(\lambda^T x_i))\right] + \alpha\|\lambda\|_0 \tag{4}$$

where $\|\lambda\|_0$ denotes the l_0 norm of parameter λ. By adding the l_0 norm constraint, the solution of the minimization problem (4) will be sparse, that is, the value of most elements is 0, and the attributes of the corresponding sample of non-zero elements are the main factors of power failure. Therefore, by solving the value of lambda, we can find those customers who are sensitive to power failure and the corresponding sensitivity characteristics from the data.

2.2 L_1-norm Based Sparse Logistic Regression

Since the l_0 norm problem is non-convex, it cannot be solved in polynomial time. Scholars have proposed that under given conditions, the l_1 norm problem is equivalent to the l_0 norm problem. The equivalent l_1 norm problem can be formulated as follows.

$$\min_{\lambda} \sum_{i=1}^{l} -\left[y_i(\lambda^T x_i) - \log(1 + exp(\lambda^T x_i))\right] + \alpha\|\lambda\|_1 \tag{5}$$

For problem (5), alternating direction of multipliers can be directly and effectively employed to solve it. Although the l_1 norm problem is easy to solve and can give important factor analysis results, it still has the following disadvantages:

- It is only a convex relaxation of l_0 sparseness. Numerous studies have shown that the result of the l_1 problem cannot get a good solution in practical applications.
- There are still better sparsity regularizations that can replace the l_1 norm to get a more accurate sparse solution.

2.3 K-support based Sparse Logistic Regression

Later, some scholars [3][4][5] introduced a k-support sparsity constraint on the regression coefficient λ of the model. Based on the given training data set, the maximum posterior probability estimation of the model parameters using *k*-support regularization is equivalent to minimize the following objective function:

$$\min_{\lambda} \sum_{i=1}^{l} -\left[y_i(\lambda^T x_i) - \log(1 + exp(\lambda^T x_i))\right] + \alpha\left(\|\lambda\|_k^{sp}\right)^2 \tag{6}$$

where the first item is the likelihood item, and the results predicted by the model will be consistent with the real category of the training sample. α is the regularization parameter, which controls the balance between the likelihood estimate and the regularization constraint. The k-support norm of $\lambda \in R^{d+1}$ is

$$\|\lambda\|_k^{sp} = \left(\sum_{i=1}^{k-r-1}\left(|\lambda|_i^{\downarrow}\right)^2 + \frac{1}{r+1}\left(\sum_{i=1}^{k-r-1}|\lambda|_i^{\downarrow}\right)^2\right)^{1/2} \tag{7}$$

where parameter k is used to force the model to select at most k factor variables for prediction, $|\lambda|_i^{\downarrow}$ is the i-th largest element of vector λ, $r \in \{0, 1, \cdots, k-1\}$ obeys $|\beta|_{k-r-1}^{\downarrow} > \frac{1}{r+1}\left(\sum_{i=k-r}^{d+1}|\beta|_i^{\downarrow}\right) \geq |\beta|_{k-r}^{\downarrow}$. According to (7), the k-support norm is consist of a l_2 norm constraint for the punishment of large coefficients and a l_1 norm constraint for the punishment of small coefficients. Therefore, the k-support norm penalizes the l_2 norm of the selected factor variable group while selecting a small number of factor variable groups for prediction.

As a result, the over-shrinkage problem of l_1 norm regularization can be effectively solved, it is beneficial to select highly relevant factor variables from many factors to predict customers with high sensitivity to the power failure.

Although k-support norm combines the advantages of l_2 norm and l_1 norm, and can obtain a more accurate sparse solution than that of l_1 norm problem, it still has the following disadvantages.
● Model (7) only considers the sparsity of a single sample for factor analysis, ignoring the common characteristics of cluster samples.
● There is still room to improve the accuracy of the sparsity solution of the k-support norm, such as the non-convex l_p $(0 < p < 1)$ norm

2.4 Sensitivity Determination

Based on the obtained model parameter $\boldsymbol{\lambda}$, when the user's factor variable set $\boldsymbol{x}$ is input, the probability of confirming a power failure sensitive customer can be calculated as:

$$p(y = 1|\boldsymbol{x}) = \frac{\exp(\lambda^T x)}{1+\exp(\lambda^T x)} \tag{8}$$

where exp is the base of the natural log, which is approximately 2.71828. By comparing equation (4) with the given threshold ε, we can generate the predicted result $\hat{y}$:

$$\hat{y} = \begin{cases} 1, & p\{y = 1|x\} > \varepsilon \\ 0, & p\{y = 1|x\} \le \varepsilon \end{cases} \tag{9}$$

when $p\{y = 1|x\}$ is bigger than a given threshold ε, a power failure sensitive customer can be confirmed.

3. Proposed Grouped $L_{1/2}$ Sparsity Constrained Logistic Regression Method

3.1 Formulation

Based on the above discussion, in this paper, we utilize a novel grouped $l_{1/2}$ sparsity constrained logistic regression method to predict the customer sensitivity of power failure. First, all samples are clustered into S categories, and those samples in each category are assumed to share similar supporting sparse coefficients. Let samples in category i be denoted as $X_i = \{\boldsymbol{x}_1^i, \boldsymbol{x}_2^i, \dots, \boldsymbol{x}_{s_i}^i\}$ and the corresponding labels be denoted as $Y_i = \{y_1^i, y_2^i, \dots, y_{s_i}^i\}$, here s_i is the number of samples in category i and $\sum_{i=1}^{S} s_i = l$. Therefore, the proposed model can be formulated as follows.

$$\min_{\boldsymbol{\lambda}} \sum_{i=1}^{S} -w_i \sum_{j=1}^{s_i} \left[y_j^i (\boldsymbol{\lambda}^T \boldsymbol{x}_j^i) - \log(1 + exp(\boldsymbol{\lambda}^T \boldsymbol{x}_i)) \right] + \alpha \|\boldsymbol{\lambda}\|_{1/2} \tag{10}$$

where $w_i = 1/s_i$ is the weight for the i-th category and $\|\boldsymbol{\lambda}\|_{1/2} = \left(\sum_{i=1}^{d+1} \lambda_i^{1/2} \right)^2$ is the $l_{1/2}$ norm of weight vector $\boldsymbol{\lambda}$. The proposed model (10) has the following advantages.
● By clustering the samples, and then weighting the objective function according to the number of samples in the category, the problem caused by the imbalance of training

samples can be effectively removed.
- The $l_{1/2}$ norm regularizer in the model (10) can get a solution closer to that of l_0 norm than the l_1 norm and k-support norm, making the result of factor analysis more accurate.

3.2 Optimization Algorithm

Because of the non-convex $l_{1/2}$ norm in model (10), classical gradient descent, Newton iteration and other algorithms cannot directly solve it. Even ADMM algorithm is designed to solve the problem with all convex terms, it has also been proved to be efficient in solving the non-covex problem with $l_{1/2}$ sparsity constrained regularization [24][25].

$$\begin{cases} \lambda^{t+1/2} = \lambda^t + \rho \sum_{i=1}^{S} w_i \sum_{j=1}^{s_i} x_j^i \left(y_j^i - \dfrac{exp(\lambda^T x_j^i)}{1+exp(\lambda^T x_j^i)} \right) \\ \lambda^{t+1} = prox_{\rho\|\lambda\|_{1/2}}(\lambda^{t+1/2}) \end{cases} \quad (11)$$

where t is the number of iterations, ρ is the iteration step size, and the neighborhood operator for the regular term $\rho\|\lambda\|_{1/2}$ is defined as:

$$\lambda^{t+1} = \arg\min_{\lambda}\|\lambda - \lambda^{t+1/2}\|_2^2 + \rho\|\lambda\|_{1/2} \quad (12)$$

This subproblem has an explicit solution. For details, please refer to [20][21]. These two sub-steps are continuously iterated until the algorithm converges. In order to verify the convergence of the numerical optimization algorithm, **Fig. 2** shows the relative iteration deviation on the training data set. The attenuation curve of iteration parameters along with the iteration parameters. It can be drawn that the proposed algorithm only needs a few number of iterations to converge which generated highly calculation effectiveness.
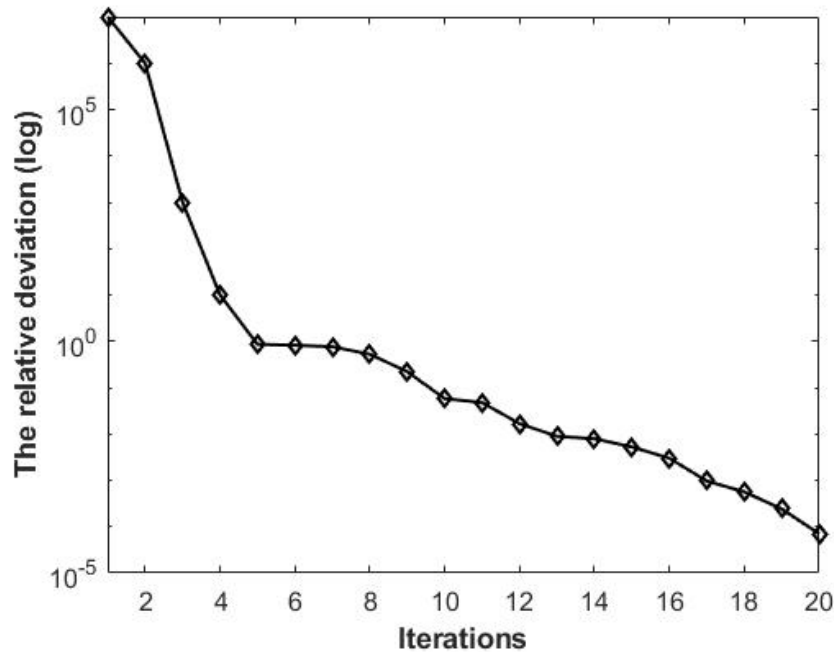

Fig. 2. Iteration deviation decay curve.

3.3 Variable Correlation Analysis

After the regression model is obtained, the importance of each factor variable's influence on the target variable can be further analyzed. The methods for measuring the importance of variables generally include chi-square value, p-value, standardized regression coefficient, partial correlation coefficient, advantage weight, etc. Each method has its own advantages and disadvantages. To obtain more accurate results, we choose the advantage analysis employed in [3]. Advantage analysis decomposes and distributes the total variation of the linear regression model to each independent variable, and has achieved good application results in industry problems.

The grouped $l_{1/2}$ sparse logistic regression model proposed in this paper has d independent variables, and the dominant weight calculation process of the factor variable $\boldsymbol{x}_i$ is as follows.

1)   Calculate R^2 , where $\boldsymbol{x}_i$ is the independent variable, and R^2 is the percentage of the target variable explained by the independent variable in the linear model, that is, the ratio of the sum of the regression squares and the total sum of squares:

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \tag{13}$$

where y is the true value, $\hat{y}$ is the model estimate value, and $\bar{y}$ is the mean of the model estimate. In the sparse logistic regression model, R^2 is defined as:

$$R^2 = 1 - \left[\frac{\mathcal{L}(\lambda_0)}{\mathcal{L}(\lambda)}\right]^{\frac{2}{l}} = 1 - e^{\left\{\frac{-1}{l}[(-2\ln(\mathcal{L}(\lambda_0)))-(-2\ln(\mathcal{L}(\lambda)))]\right\}} \tag{14}$$

where $\mathcal{L}$ is the maximum likelihood function of the model, λ is the model parameter, and l is the number of model observations.

2)   Calculate the incremental contribution ΔR^2 caused when x_i is included in a model with a single independent variable ($x_j, j \neq 1$), and average all ΔR^2 in the group.
3)   Calculate the incremental contribution ΔR^2 caused when x_i is included in a model with double independent variable ($\boldsymbol{x}_j, \boldsymbol{x}_i, i \neq j, i \neq k, j \neq k$), and average all ΔR^2 in the group.
4)   Calculate the mean of the contribution increment ΔR^2 in all the steps above, find the dominant weight of the variable $\boldsymbol{x}_i$. Continue steps 1 to 4 for each factor variable in the model to calculate the dominant weight of each factor.

4. Experimental Results

In this section, we employ one real dataset to validate the effectiveness of the proposed $l_{1/2}$-based sparse regression method.

4.1 Data Set

Due to the low-voltage and high-voltage users are both involved, some differences in the behavior characteristics exist in different types of users. In order to fully describe the characteristics of users, this article integrates multiple dimensions, e.g., basic customer information, power consumption information, payment information, and power failure events to select customer information fields that might be related to the power failure sensitivity, such as power consumption category, power consumption, contract capacity, business type, and

power failure type as follows [3].

1) Basic attributes: the number of users, account opening date, urban and rural category, etc.

2) Power consumption data: user category, industry type, supply voltage, contract capacity, measurement method, load level, etc.

3) Consumption behavior: electricity consumption, electricity bills, electricity bill ladder, electricity bill notification methods, etc.

4) Payment behavior: payment methods, payment channels, payment frequency, etc.

5) 95598 worksheets information: acceptance time, type of business accepted, type of electricity used, urging supervision, etc.

6) Power failure event information: time, duration, power failure type, power failure reason, etc.

Then, the data is preprocessed first to ensure the correctness of the data. It mainly includes the uniqueness of the user number, the integrity of the sample data, the range and value of the variables, missing values, outliers, etc. Furthermore, it is to construct derived variables, i.e., processing the original data to obtain more predictive and explanatory variables. For example, the number of calls to 95598 and the number of reminders.

4.2 Assessment Indexes

After the model is constructed, the prediction accuracy of the model requires to be evaluated. Assume that the confusion matrix formed by the prediction results of the model and the real results are shown in **Table 1**, where TP (true positive) is the number of customers that are correctly predicted to be failure sensitive, FP (false positive) is the number of customers who are incorrectly predicted to be failure sensitive, and TN (true negative) is to predict the correct number of non-sensitive customers, FN (false negative) is the number of customers who are incorrectly predicted to be non-sensitive customers, P = TP+FN is the number of calibrated sensitive customers, N = FP + TN is the calibrated non-sensitive customer.

Table 1. Confusion matrix.

		Experimental Results	
		1	0
Predicted Results	1	TP (true positive)	FP (false positive)
	0	FN (false negative)	TN (true negative)
Total		P	N

The accuracy is the ratio of the number of correct customers to the total number of customers:

$$accuracy = \frac{TP+TN}{P+N} \tag{15}$$

The sensitivity is defined as the ratio of the predicted number of sensitive customers to the total number of sensitive customers:

$$sensitivity = \frac{TP}{P} \tag{16}$$

The specificity is defined as the ratio of correctly predicted nonsensitive customers to all true non-sensitive customers:

$$specificity = \frac{TN}{N} \tag{17}$$

Correspondingly, $1 - specificity = \frac{FP}{N}$ is the ratio of mispredicted non-sensitive customers to all real non-sensitive customers.

Table 2. Comparison of prediction accuracy.

Customer type	Decision tree	l_1 sparsity	k-support	Proposed algorithm
Low voltage residential customers	75.3	84.1	87.5	88.7
Non-resident customers	72.8	82.9	86.1	87.3
High-voltage customers	69.2	78.4	82.3	83.0
All of the customers	74.5	83.2	86.4	87.1

In the experiment, the number of categories and parameter ρ are two important variables in the proposed method. Since the number of categories is determined by the data itself, it is not easy to obtain the value of it automatically. Therefore, we empirically set is to be 11 in the following experiment. For step size ρ, we set it to be 0.001. **Table 2** gives the 5-fold cross-validation accuracy of the proposed grouped $l_{1/2}$ norm constrained logistic regression method as well as the k-support sparse logistic regression model, the classic l_1 sparse logistic regression model, and the classic decision tree model on the test dataset. The l_1 sparse logistic regression model and the discriminant threshold of the algorithm in this experiment are set to 0.70. It can be seen that the prediction performance of the decision tree model is lower than the l_1 sparse logistic regression model. At the same time, the prediction accuracy of the algorithm proposed in this paper is higher than that of the l_1 sparse logistic regression model and k-support sparse logistic regression model for all kinds of customers, indicating that the introduction of grouped $l_{1/2}$ sparsity regularization is conducive to improving the generalization ability of the model.

ROC curve (receiver operating characteristic curve) is further calculated. By changing the discrimination threshold ε in each regression model, the corresponding curves of $sensitivity$ and $1 - specificity$ values under different thresholds are drawn to judge the prediction performance of each model. As shown in **Fig. 3**, the ROC of the proposed model is located

above the l_1 sparse logistic regression model and the k-support sparse logistic regression model, which shows that the proposed grouped $l_{1/2}$ norm constrained logistic regression model has higher accuracy.
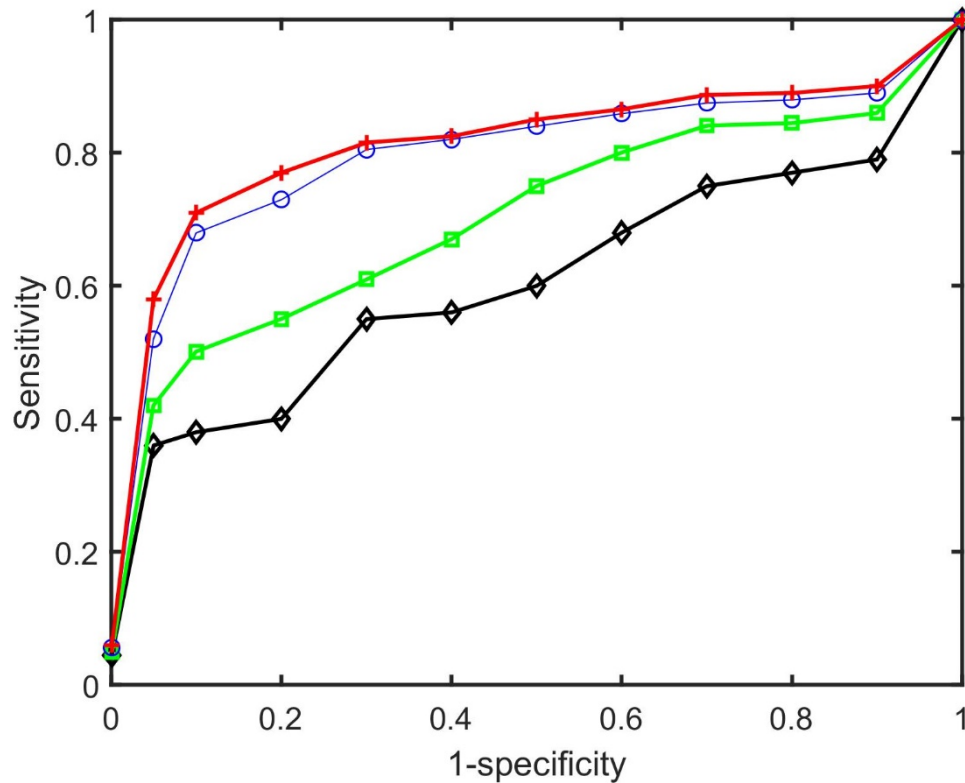

Fig. 3. Comparison analysis of ROC curve.

4.3 Model Verification

Among the 60,000+ high-voltage customers in April in the experiment, customers who generated consultations about power failures, that is, sensitive customers accounted for about 5.6%. When the probability threshold ε of the model is set to 0.7, the prediction accuracy of the proposed model is 82.05%. At the same time, the ROC performance curve of the algorithm is given in **Fig. 4**, and the area under the ROC curve is calculated, that is, the AUC (area under curve) statistics is 0.857. The experimental results show that the model in this paper can predict the failure sensitive customers more accurately.
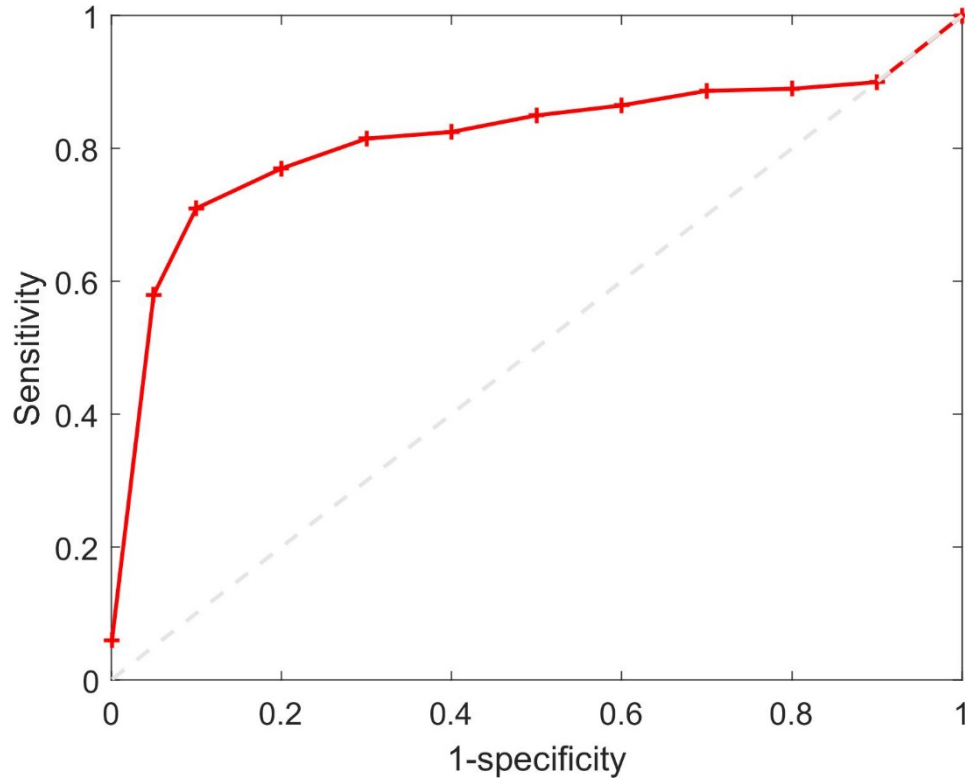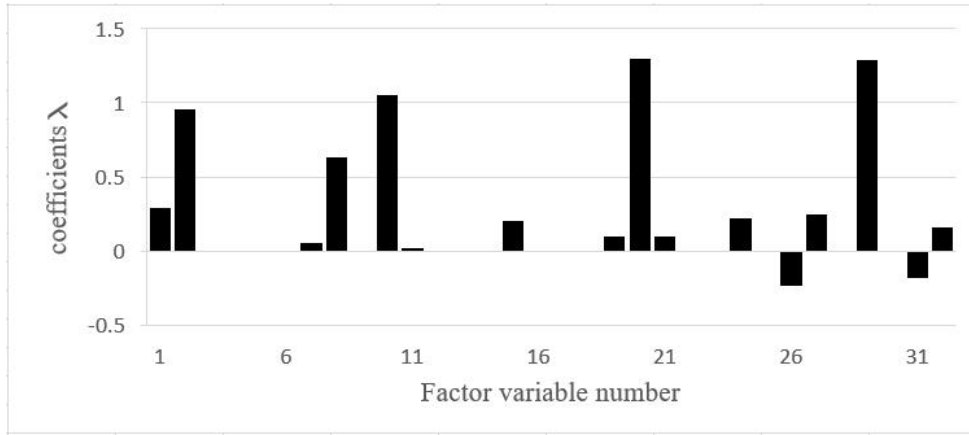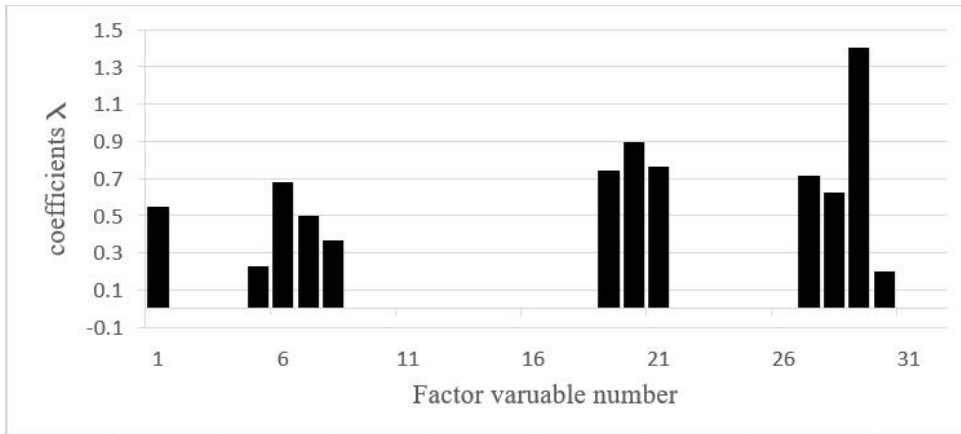
Fig. 4. The ROC curve of the proposed algorithm in April.

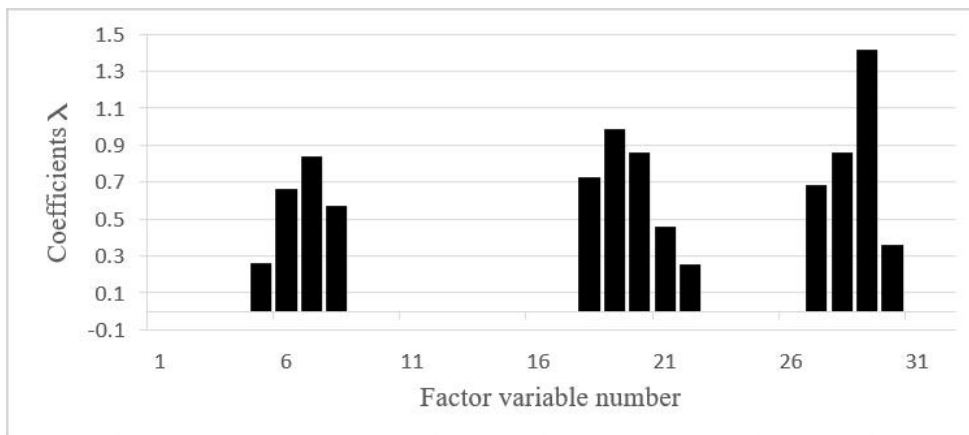## 4.4 Importance Analysis of Independent Variables

Fig. 5 compares the proposed grouped $l_{1/2}$ norm constrained logistic regression method with two sparsity based models after training, i.e., the k-support sparse logistic regression model and the l_1 sparse logistic regression model. The regression coefficient λ is obtained firstly. Only a few factors of the sparse logistic regression model have large amplitude coefficients, and there is a certain degree of excess. Shrinkage problems can easily cause model instability. By applying the grouped $l_{1/2}$ norm constrained and the k-support sparse regular constraint, the model proposed in this paper achieves a good balance between the sparseness and stability and selects the highly relevant factor variables from many factors to predict the failure sensitive customers.

(a) l_1 sparsity

(b) k-support sparsity

(c) Grouped $l_{1/2}$ sparsity

Fig. 5. Comparison analysis of regression coefficient.

The dominance analysis based method is further employed to calculate derives the relative importance of all variables, which indicated in **Table 3**. The number of historical calls to 95598, the type of industry, and other factors have the greatest impact on the sensitivity of customers to power failures. According to the actual work needs, it is possible to select the factor variables with higher influence for prediction, reducing the calculation amount, and improving the efficiency of actual business operations.

Table 3. The relative importance of factor variables.

Independent variable	Variable weight
Load importance	0.0003
Note type	0.0045
Notification Type	0.0409
Number of measuring points	0.1031
Measurement method	0.3332
Contract capacity	0.3731
Average electricity price	0.4013
Urge to supervise	0.4589
Industry Type	0.4816
Power supply unit	0.8012
Call 95598 times	1.6033

5. Conclusion

This paper analyzes the sparse logistic regression algorithm based on the grouped $l_{1/2}$ sparsity and its application in customer power failure sensitivity model evaluation. The experimental evaluations illustrate our proposed model performs well in a real dataset, which indicates that the proposed model can be generalized and applied to practical work. For example, according to the predicted results of power failure sensitive customers, we can continue to perform targeted management optimization measures. When a power failure occurs, priority notifications and key notifications are given to the high-level sensitive customers. Moreover, according to customers' feedback, we can ensure the power line security maintenance, power failure management optimization, and marketing feedback, etc. It should be pointed out that the grouped $l_{1/2}$ sparsity constrained logistic regression model combined with the dominance analysis method determines the key factors. In the face of the actual work requirements, important model parameters should be selected for the progress of prediction, establish the scorecard to reduce the calculation amount, and improve the efficiency of actual business operations.

Acknowledgement

References

[1] A. Bernstein, D. Bienstock, D. Hay D, "Sensitivity analysis of the power grid vulnerability to large-scale cascading failures," *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 3, pp. 33-37, Jan 2012. Article (CrossRef Link)

[2] Y. Yang, T. Nishikawa, A. E. Motter, "Small vulnerable sets determine large network cascades in power grids," *Science*, vol. 358, no. 6365 pp. 886, Nov 2017. Article (CrossRef Link)

[3] J. Geng, X. Zhang, Y. Sun, B. Wu and Q. Zhou, "Power failure sensitivity prediction algorithm using k-support sparse logistic regression," *Computer and Modernization*, vol. 4, pp. 68-73, Apri. 2018. Article (CrossRef Link)

[4] Argyriou, Andreas, and Foygel, Rina, and Srebro, Nathan, "Sparse prediction with the *k*-support norm," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1457-1465, Dec. 2012. Article (CrossRef Link)

[5] S. W. Akhtar et al., "Improving the robustness of neural networks using k-support norm based adversarial training," *IEEE Access*, vol. 4, pp. 9501-9511, Dec. 2016. Article (CrossRef Link)

[6] G. Tso, K. K.Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761-1768, Sep 2007. Article (CrossRef Link)

[7] N. Mohanapriya, B. Kalaavathi, "Adaptive image enhancement using hybrid particle swarm optimization and watershed segmentation," *Intel. Auto. Soft Comput.*, vol. 25, pp. 663–672, Dec 2019.

[8] C. Hung, W. Mao, H. Huang, "Modified PSO algorithm on recurrent fuzzy neural network for system identification," *Intel. Auto. Soft Comput.,* vol. 25, no. 2, pp. 329–341, Jun 2019.

[9] C. Li, Z. Ding, D. Zhao D, "Building energy consumption prediction: An extreme deep learning approach," *Energies*, vol. 10, no. 10, pp. 1525, Oct 2017. Article (CrossRef Link)

[10] F. Xu, X. Zhang, Z. Xin, A. Yang, "Investigation on the Chinese text sentiment analysis based on ConVolutional neural networks in deep learning," *Computer Materials & Continues*, vol. 58, no. 3, pp. 697–709, Mar 2019. Article (CrossRef Link)

[11] Y. Guo, C. Li, Q. Liu, "R^{2}N: a novel deep learning architecture for rain removal from single image," *Computer Materials & Continues*, vol. 58, no. 3, pp. 829–843, Mar 2019. Article (CrossRef Link)

[12] H. Wu, Q. Liu, X. Liu, "A review on deep learning approaches to image classification and object segmentation," *Computer Materials & Continues*, vol. 60, no. 2, pp. 575–597, Oct 2019. Article (CrossRef Link)

[13] X. Zhang, W. Lu, F. Li, X. Peng, R. Zhang, "Deep feature fusion model for sentence semantic matching," *Computer Materials & Continues*, vol. 61, no. 2, pp. 601–616, Nov 2019. Article (CrossRef Link)

[14] N. Stephanie, S. Dustin, R. Todd Constable, "Cluster failure or power failure? Evaluating sensitivity in cluster-level inference," *NeuroImage*, vol. 209, pp. 116468, Apri. 2020. Article (CrossRef Link)

[15] L. Fu, Z. Li, Q. Ye, "Learning robust discriminant subspace based on joint l2,p- and l2,s-norm distance metrics," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-15, 2020. Article (CrossRef Link)

[16] Q. Ye, J. Yang, F. Liu, C. Zhao, N. Ye, T. Yin, "L1-norm distance linear discriminant analysis based on an effective itera-tive algorithm," *IEEE Transactions on Circuits Systems and Video Technology*, vol. 28, no. 1, pp. 114-129, Jan 2018. Article (CrossRef Link)

[17] Q. Ye, H. Zhao, Z. Li, X. Yang, S. Gao, T. Yin, N. Ye, "L1-norm distance minimization based fast robust twin support vector k-plane clustering," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 4494-4503, Sep 2018. Article (CrossRef Link)

[18] S. B. Singh, Beena Nailwal, "Reliability measures and sensitivity analysis of a complex matrix system including power failure," *IJE Transactions A: Basics*, vol. 25, no. 2, pp. 115-130, Jun. 2012.

[19] McDonald, Andrew M., Massimiliano Pontil, and Dimitris Stamos. "Spectral k-support norm regularization," *Advances in neural information processing systems*, vol. 27, pp. 3644-3652, Dec. 2014. Article (CrossRef Link)

[20] J. Lou, and Y. -M. Cheung, "Robust low-rank tensor minimization via a new tensor spectral k-support norm," *IEEE Transactions on Image Processing*, vol. 29, pp. 2314-2327, Oct. 2020. Article (CrossRef Link)

[21] Y. Qian, S. Jia, J. Zhou and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282-4297, Jun. 2011. Article (CrossRef Link)

[22] L. Sun, F. Wu, C. He, T. Zhan, W. Liu, D. Zhang, "Weighted collaborative sparse and $l_{1/2}$ low-rank regularizations with superpixel segmentation for hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, pp. 1-5, 2020. Article (CrossRef Link)

[23] W. Wang and Y. Qian, "Adaptive $l_{1/2}$ sparsity-constrained NMF with half-thresholding algorithm for hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2618-2631, Feb. 2015. Article (CrossRef Link)

[24] S. Kahraman, A. Ertürk and S. Ertürk, "Graph Regularized $l_{1/2}$-sparsity constrained non-negative matrix factorization for hyperspectral and multispectral image fusion," in *Proc. of Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1-4, 2018. Article (CrossRef Link)

[25] M. Guda, S. Gasser, M. S. El-Mahallawy and K. Shehata, "FPGA implementation of $l_{1/2}$ sparsity constrained nonnegative matrix factorization algorithm for remotely sensed hyperspectral image analysis," *IEEE Access*, vol. 8, pp. 12069-12083, Jan. 2020. Article (CrossRef Link).

Baoshu Li received the Bachelor degree in Nanjing University of Science & Technology. He is working in Jiangsu Electric Power Information Technology Co,. Ltd. His current research interests include power grid planning, energy internet.

Xin Zhou received the Bachelor of Engineering degree in Information Engineering from Nanjing University of Information Science & Technology (NUIST). He is working in Jiangsu Electric Power Information Technology Co,. Ltd. His current research interests include Software Engineering.

Ping Dong graduated from Hunan Normal University in 2014. He is R&D Engineer of Jiangsu Electric Power Information Technology Co., China. His current research direction is power grid planning, wisdom health and power information technology.