Electric Larceny Detection Based on Support Vector Machine

Li Songnong
[1.] State Grid Chongqing Electric Power Research Institute, Chongqing, 401123, China
[2.] Chongqing Key Laboratory of Energy Internet Advanced Metrology and Measurement Technology Chongqing, 401123, China
E-mail: lxpecolicee@163.com

Ye Jun
[1.] State Grid Chongqing Electric Power Research Institute, Chongqing, 401123, China
[2.] Chongqing Key Laboratory of Energy Internet Advanced Metrology and Measurement Technology Chongqing, 401123, China
E-mail: gtovictor@163.com

Zeng Yan
[1.] State Grid Chongqing Electric Power Research Institute, Chongqing, 401123, China [2.] Chongqing Key Laboratory of Energy Internet Advanced Metrology and Measurement Technology, Chongqing, 401123, China
E-mail: cqepzy@163.com

Sun Hongliang
[1.] State Grid Chongqing Electric Power Research Institute, Chongqing, 401123, China
[2.] Chongqing Key Laboratory of Energy Internet Advanced Metrology and Measurement Technology, Chongqing, 401123, China
E-mail: cqepsunhl@163.com

***Abstract*—The design and application of power system line loss calculation and lean management system have important guiding significance in guiding loss reduction and energy saving and promoting line loss management. In recent years, the electric energy data acquire system, as a tool that can effectively meet the power enterprise's demand for power consumption information, has also accumulated a large amount of user power consumption data while meeting the power supply marketing automation needs. These power consumption data contain huge user power usage information. Therefore, the user data collected by the power electric energy data acquire system can be analyzed and processed to identify users with high suspicion of power severance, so as to reduce the management line loss. To this end, this paper studies a small-volume user anomaly power detection scheme based on Support Vector Machine (SVM), which can effectively identify the abnormal power consumption mode by tracking and screening the load data of the user for a period of time. An unbalanced sample synthesis processing model based on SMOTE+Bagging is constructed. The differential evolution algorithm is used to optimize the SVM parameters, which solves the problem that SVM classification performance is more affected by parameters. At the same time, the operational efficiency of the SVM-based Bagging integrated classification model is guaranteed.**

Keywords-Component; Lean Management, Management Line Loss, Support Vector Machine, SMOTE+Bagging, Unbalanced Sample

I. INTRODUCTION

Line loss is a comprehensive reflection of the economic operation management level of power grids and the economic benefits of power supply enterprises. The effective containment of electric larceny behavior has great significance for the development of power companies and the whole society. this paper studies a small-volume abnormal power detection scheme based on Support Vector Machine (SVM). By tracking and screening the load data of the user for a period of time, the identification of the abnormal power mode is effectively completed. The specific implementation of the scheme will be introduced below.

II. UNBALANCED SAMPLE SYNTHESIS PROCESSING MODEL BASED ON SMOTE+BAGGING

A. *Analysis of the Influence of Unbalanced Samples on SVM Detection Results*

The detection of abnormal electrical behavior is also a typical classification problem under unbalanced samples. This paper uses SMOTE over-sampling and Bagging combination classification technology to construct the SVM integrated classification model. The model can effectively deal with the unbalanced sample problem in abnormal power consumption detection, thereby achieving the purpose of improving the overall detection accuracy of the electric larceny detection.

B. *SMOTE oversampling technology*

Oversampling refers to the purpose of balancing two types of sample data by repeated sampling of a small number of sample data. Chawla proposed a synthetic oversampling SMOTE (Synthetic Minority Over-sampling Technique) algorithm. The schematic diagram of the Smote algorithm is shown in Figure II-1. The SMOTE algorithm randomly inserts newly synthesized minority samples by connecting a small number of samples and their surrounding samples. It effectively avoids the problem that the classification decision interval is small due to the excessive sample repetition rate, which leads to the over-fitting phenomenon of the classifier. The specific SMOTE algorithm is detailed in the reference [1].

C. Bagging integrated classification

The Bagging algorithm is a combined classification method proposed by Breiman. By combining the classification results of multiple classifiers, the purpose of improving the classification effect of the overall classifier is achieved. The specific Bagging algorithm is detailed in the reference [2]

D. Unbalanced sample comprehensive detection model based on SMOTE+Bagging

Combining the advantages of the above two methods, a comprehensive abnormal energy usage sample processing model based on SMOTE algorithm and Bagging integrated classification method is constructed. The SVM+SMOTE+Bagging model demonstrates the powerful integration benefits of the Bagging algorithm and the excellent processing power of the SVM algorithm for balanced data. This model greatly improves the processing ability of unbalanced data, and provides powerful support for the abnormal recognition of unbalanced user power usage data.

III. SVM PARAMETER OPTIMIZATION

A. SVM parameter impact analysis

The Abnormal use of electricity detection model constructed in this paper is based on SVM as the basis for detection. The superiority of SVM detection performance is related to the performance of the entire detection system. The classification effect of the SVM classifier depends largely on the reasonable selection of the penalty factor C and the parameters in the kernel function [3].

The penalty factor C can be regarded as a degree of punishment for dealing with deviations from the sample. The kernel parameter g reflects the degree of correlation between the support vectors [4]. Therefore, the reasonable selection of parameters has great significance for the improvement of SVM overall classification performance.

As an emerging group intelligence global search algorithm, differential evolution algorithm has attracted more and more attention due to its low control parameters and high optimization efficiency [5, 6]. Therefore, this paper considers the introduction of differential evolution algorithm to reasonably select the parameters of SVM to improve the classification accuracy of SVM detection model and ensure the operation efficiency of SVM-based comprehensive detection model.

B. Differential Evolution Algorithm

Differential Evolution (DE) is an intelligent optimization algorithm that relies on group random search [7]. The method mainly guides the optimization strategy by the group intelligence generated by the cooperation and competition between individuals within the group, so that each iteration can produce a better solution than the original. The algorithm is similar to the idea of genetic algorithm. For each generation of samples, the three processes of mutation, crossover and selection are used to generate better solutions, and then through continuous iteration, the global optimal solution or approximate optimal solution is finally found. . For details of the DE algorithm, see reference [7].

C. Parameter Optimization of SVM Based on Differential Evolution Algorithm

The steps of using the differential evolution algorithm to find the optimal penalty factor C and the kernel function parameter g are as follows:

Step 1: Set the search range of the SVM penalty parameter and the kernel function parameter, and set the relevant parameters of the differential evolution algorithm, including the population size N, the population dimension D, the maximum evolution algebra T, the scaling factor F, and the crossover probability CR;

Step 2: randomly generate an initial population within the specified search scope;

Step 3: For the t-th generation population, calculate the fitness of each individual in the population and determine whether the set accuracy is met or whether t is equal to the maximum evolution algebra T. If yes, go to step 6, otherwise t=t+1 for the next generation of evolution. The fitness function set here is the classification correct rate of the support vector machine k-weight cross-validation. The classification correctness rate R is defined as follows:

$$R = \frac{\text{Number of samples correctly classified}}{\text{Total number of samples}} \tag{1}$$

Step 4: Select individuals from the current generation population, and perform operations such as mutation, intersection, and selection;

Step 5: Calculate the new c, g parameters and go to step 3;

Step 6: Take the obtained optimal parameters c and g as parameters of the SVM, and construct the subsequent SVM detection model.

The flow chart of the above steps is as follows:

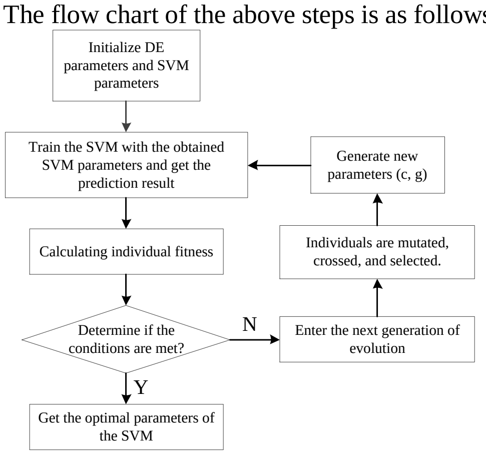

Figure 1. SVM parameter optimization flow chart based on differential evolution

IV. ABNORMAL ELECTRICITY IDENTIFICATION BASED ON SVM INTEGRATED PROCESSING MODEL

A. *User behavior analysis*

Figure IV-1 shows the average daily load curve information for two typical normal users over the whole year after standardization. The average daily load here is a two-week time period. Since the original data fluctuates less, it will show more severe jitter as shown in Figure IV-2 after standardization.

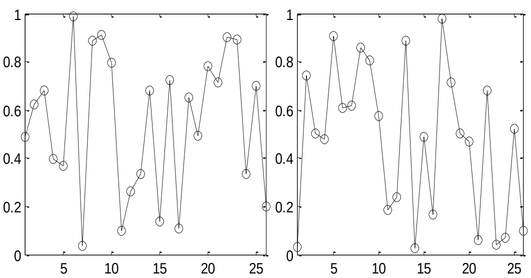

Figure 2. Figure IV-1. Normal user power load characteristic curve

If the user implements the electric larceny behavior, the power load characteristic mode will have a significant deviation from the normal power consumption. Figure IV-2 shows the average daily load curve information for two typical tamper users.

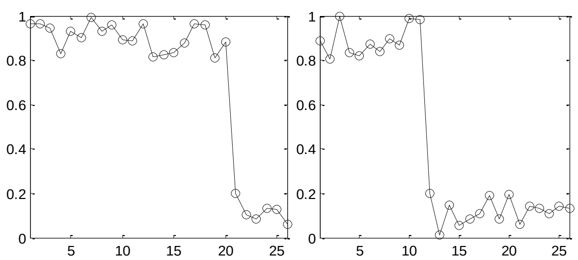

Figure 3. Typical abnormal electric load characteristic curve

B. *Overall inspection design*

The electric larceny detection scheme constructed based on the optimized SVM comprehensive detection model is shown in the figure IV-3. The whole detection system is divided into four parts: data acquisition and preprocessing, power feature pattern extraction, construction of abnormal electricity detection model and detection system decision analysis.

V. RESULTS VERIFICATION ANALYSIS

Dataset introduction: The dataset comes from an electric power company's anti-stealing inspection. The data sample contains 904 household users' electricity consumption data for a whole year. Firstly, the dataset is cleaned and preprocessed, and then 893 samples are obtained. Among them, there are 870 normal power samples and 23 abnormal power samples. In order to reflect the influence of different algorithms and processing methods on the final detection result of electric larceny, 500 abnormal power samples are added from the system library to the data set to form the final verification data set.

A. *Result verification analysis*

Then, through different proportions of normal and abnormal power samples, the performance detection of the abnormal power consumption detection model based on SMOTE+Bagging is completed. Finally, in order to reflect the influence of the method on the detection results more vividly, a certain amount of electricity data is selected to form a training set and a test set to complete the detection performance comparison of SVM, DE-SVM and SMOTE+Bagging based on SVM model.

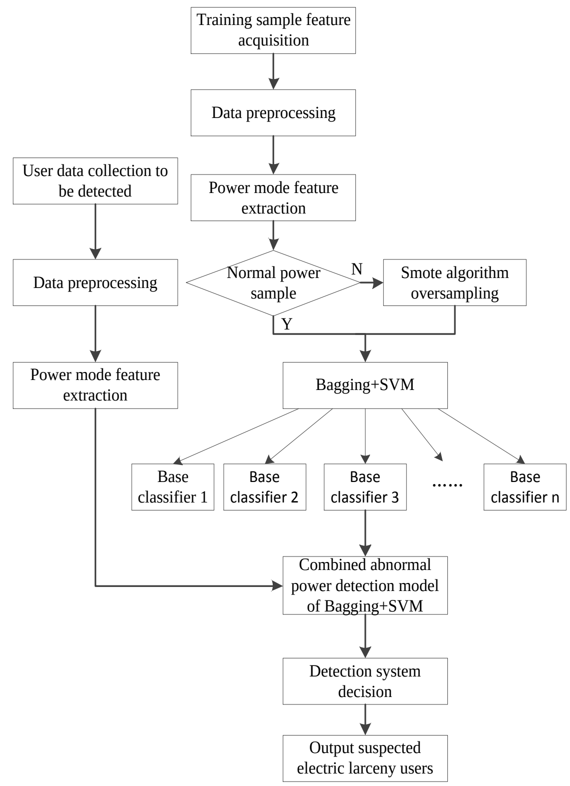

Figure 4. SVM-based abnormal power inspection system flow chart

B. *SVM parameter optimization*

The experimental design is as follows: Select a fixed number (520) of normal power samples from the total power sample set, and select abnormal power samples according to different ratios of 1:1, 3:1, 5:1, 10:1, and use To form a validation sample set. The algorithm results are the average of 30 runs. The parameters of several group intelligent optimization algorithms are set as follows: population size N=40, maximum evolution algebra T=200, termination threshold is set to 0.0001, SVM parameter C, g ranges from [2^{-10}, 2^{10}]. Specific tests are shown in the table 1, table 2.

TABLE I. PARAMETER OPTIMIZATION CONSUMPTION TIME UNDER DIFFERENT SAMPLE RATIOS (UNIT: S)

Sample ratio	GA-SVM	PSO-SVM	DE-SVM
1:1	821.11	918.04	467.69
3:1	446.24	495.95	246.81
5:1	308.75	383.26	173.71
10:1	256.65	289.37	117.80

TABLE II. F VALUE AT DIFFERENT SAMPLE RATIOS

Sample ratio	GA-SVM	PSO-SVM	DE-SVM
1:1	0.7566	0.7550	0.7603
3:1	0.7202	0.7003	0.7131
5:1	0.6782	0.6802	0.6817
10:1	0.4615	0.4706	0.4919

C. Comprehensive model detection based on SMOTE and Bagging

SVM, SVM+Bagging, SVM+SMOTE are selected to perform the similar processing on the same sample set to objectively present the effect of the detection model. The specific results are shown in Table Tables 3.

TABLE III. F VALUES OF VARIOUS METHODS UNDER DIFFERENT SAMPLE RATIOS

Sample ratio	SVM	SVM+Bagging	SVM+SMOTE	SVM+SMOTE+Bagging
1:1	0.7541	0.7723	0.7489	0.7692
3:1	0.7123	0.7208	0.7396	0.7603
5:1	0.6809	0.6911	0.7304	0.7535
10:1	0.4946	0.4939	0.7098	0.7267

D. Comprphensive verification analysis

SVM, DE-SVM and SMOTE+Bagging+SVM models are used for abnormal power consumption. The detection effect is verified, and the verification result is shown in Figure 4.8-4.10.

In summary, the differential evolution algorithm used in this paper can effectively improve the overall recognition accuracy of the SVM-based anomaly power detection model and reduce the SVM parameter optimization time. At the same time, the SMOTE+Bagging integrated processing model constructed for the unbalanced sample problem can significantly improve the detection accuracy of a few abnormal users, which is of great value for improving the SVM-based abnormal power detection effect.

VI. SUMMARY

In this paper, based on the analysis of user's electricity consumption characteristics, a supervised abnormal electricity detection scheme based on SVM is constructed. In addition, the verification of the measured data proves the effectiveness of the method and achieves more accurate tamper detection for small-scale users.

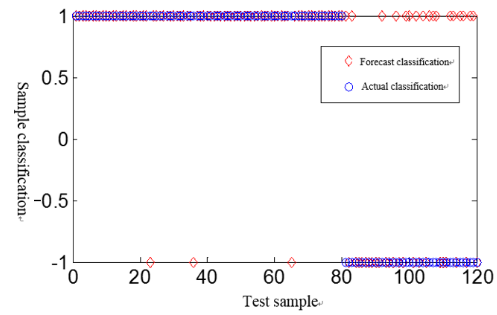

Figure 5. SVM test set accuracy

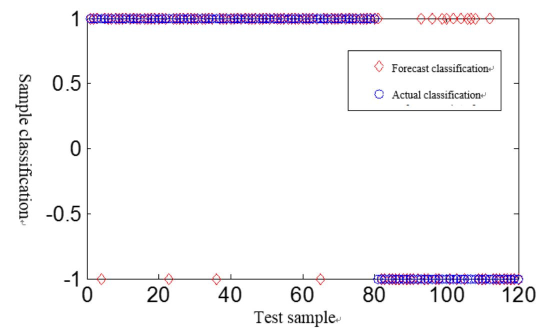

Figure 6. DE-SVM test set accuracy

ACKNOWLEDGMENT

This work is supported by State Grid science and technology project(1100-201919158A-0-0-00)

REFERENCES

[1] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1):321-357.

[2] Breiman L. Bagging predictors [J]. Machine Learning, 1996, 24(2): 23-140.

[3] Phan A V, Nguyen M L, Bui L T. Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems[J]. Applied Intelligence, 2016:1-15.

[4] Wan P, Wen J, Wu C. A discriminating method of driving anger based on sample entropy of EEG and BVP[C]//Transportation Information and Safety (ICTIS), 2015 International Conference on. IEEE, New York: IEEE Press ,2015,32(4): 156-161.

[5] LIU Hongzhi, GAO Liqun, KONG Xiangyong, et al. Application of improved differential evolution algorithm in reliability redundancy allocation problem[J]. Control and Decision, 2015, 30(5): 917-922.

[6] Zeng Xueqiang, Liu Zhigang, Fu Weijie, et al. Application of Improved Differential Evolution Algorithm in Reactive Power Optimization of Power Systems[J]. Power System Technology, 2012, 36(2): 121-125.

[7] Storn R, Price K. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces[J]. Journal of Global Optimization, 1997, 11(4):341-359.