J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 429–447

Journal of Combinatorial Mathematics and
Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Application of Time Series Analysis Model in Monitoring Electricity Consumption Behaviour and Anti-Theft of Electricity

Shuai Yang[1,✉], Qiong Cao[1], Wei Zhang[1], Hao Guo[1]

[1] *State Grid Shanxi Marketing Service Center, Taiyuan, Shanxi, 030000, China*

ABSTRACT

Electricity theft management is closely related to the economy of electric power enterprises. This paper proposes a power theft estimation method based on semi-supervised learning and time series analysis prediction. The electricity consumption data of power theft users are extracted as time series data, and in order to achieve multi-step prediction, MMD is utilized to improve the LSSVR semi-supervised learning algorithm. In addition, a perturbation term is introduced to optimize the convergence effect of the artificial bee colony algorithm, and a time series prediction algorithm based on improved artificial bee colony is established. Bringing in the power theft monitoring process to identify whether the user has power theft behavior, using the real power consumption dataset as the experimental validation data, comparing the identification accuracy of the prediction model. Predict the potential power theft of each user, solve the optimization model with the goal of optimal economic efficiency, and determine the actual ranking order of power theft users. The improved time series prediction algorithm proposed in this paper has a global error of 0.0003 and 0.0027 in dataset 1 and dataset 2, respectively, with the lowest global error and the highest overall accuracy of PSE prediction. And the algorithm predicts the list of users to be scheduled is basically the same as the list of users determined by the real PSE, which can achieve the maximum economic benefits.

Keywords: time series prediction, artificial bee colony algorithm, MMD, LSSVR, electricity consumption behavior monitoring, anti-electricity theft

1. Introduction

For a long time, a small number of lawbreakers are driven by interests and take various means to implement power theft, resulting in the loss of national electricity, damaging the legitimate rights and interests of power supply enterprises, disrupting the normal order of power supply, and also causing security risks to the operation of the power grid [1]. Therefore, there is an urgent need to master a set of effective methods that can realize the online monitoring of abnormal phenomena of electricity consumption, in order to

✉ Corresponding author.
E-mail address: shuaiyang8080@163.com (S. Yang).

Received 24 February 2024; Revised 04 May 2024; Accepted 24 October 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-026

effectively prevent and combat electricity theft and leakage behavior, and to protect the orderly and safe use of electricity.

Common electricity theft techniques are generally used to affect the conventional grounding of the metering equipment, or direct damage to the electrical equipment to achieve the purpose of not measuring or less measurement of electricity [2-3]. In recent years, due to the wide application of intelligent electronic measuring and indicating instruments, the high-tech intelligent power theft phenomenon about electronic power meters is also endless, and the use of new technical means to realize the non-contact power theft phenomenon is also increasingly common [4-7]. This high-tech power theft technology means the use of strong perturbation information to make the measurement device to form a negative deviation, and then realize the power meter to form anomalies or destruction of the purpose, but once the strong perturbation disappeared, then the measurement device to resume work, the covertness is very strong [8-11]. As the time series information contains a large number of explicit or implicit data, by analyzing the law of change and trend of change in the time series information, it is possible to find the data with practical value hidden behind the information, so as to provide a basis for dealing with various situations [12-13]. Therefore, the use of time series prediction algorithms realizes the detection of power theft by predicting the statistical patterns in the historical data of the object so as to fit a functional pattern of its changes over time [14-15].

Zhang, M. et al. use correlation analysis algorithm based on the analysis of common power theft means to establish an anti-power theft diagnostic and analysis model based on massive data information, and lock the suspected users by extracting the abnormal data in order to improve the efficiency and accuracy of prevention and investigation of power theft events [16]. Yan, S. summarizes the mainstream hi-tech power theft measures and establishes an anti-power theft technology learning platform with a remote simulation training base, aiming to improve the business quality of practitioners in the context of high-tech power theft [17]. Li, Y. used big data technology and BP neural network to construct a mathematical model of anti-stolen power, and calculated the suspected coefficient of power theft of a user by analyzing the user's power consumption behavior and assigned the corresponding credit rating [18]. Niu, R. based on the user data information of the power consumption information collection system to Niu, R. based on the user data information of the electricity consumption information collection system, establishes the analysis model of the user's electricity theft behavior, and et al. verified the correctness of the model through software simulation [19]. Chenxi, D. showed that the signals of the user's high-voltage side and low-voltage meter side can be collected to determine whether the user is stealing electricity and put forward the general packet radio service (GPRS) and short message service (SMS) system to realize the wireless long-distance transmission of meter-reading data and alarm information to achieve the purpose of real-time monitoring [20]. Tian, C. studied the power theft detection system based on K-means algorithm and LSTM algorithm, and analyzed the anti-theft role of the system by establishing the decision-making model of power theft prevention under data-driven conditions [21].

This paper clarifies the working mechanism of power consumption behavior monitoring, analyzes the common means of power theft, and establishes a power theft monitoring model. For the massive electricity consumption data, the feature representation of time series is proposed, and the time series anomaly detection based on semi-supervised learning (MMD) is introduced. The artificial bee colony algorithm is improved, and the problem that the artificial bee colony algorithm tends to converge slowly in the early stage and fall into the local optimum in the later stage is optimized by introducing a perturbation term. In addition, the worst food source is replaced to improve the convergence speed of the algorithm. Combining the advantages of the two algorithms, a power theft estimation method with improved time series prediction algorithm is designed and proposed. Real electricity consumption data is selected, and multiple comparison models are used to evaluate the advantages of the algorithm in this paper.

2. Fundamentals of modeling applications

2.1. *Electricity behavior monitoring system work*

The basic workflow of the user electricity behavior monitoring system is shown in Figure 1. Based on the rate principle of Nyquist sampling law of the data acquisition terminal to collect the current data in the circuit, monitor the regional environmental data and process them, it judges whether there is a change in the behavior of the load use in the analyzed circuit. For the current data in the analysis in which there is a change in power usage behavior, differential processing and fast Fourier transform are performed to extract the current eigenvalues. If there is a change in the detected environmental data, it can be formatted directly.

On the premise of obtaining the user's power consumption behavior by relying on the power consumption behavior monitoring system, the analysis of data information can be relied on to summarize the user's power consumption characteristics. Based on the monitoring results, the evaluation system for the user's electricity consumption status is constructed around the indicators of voltage imbalance rate, current imbalance rate, rated voltage deviation, power factor imbalance rate, contracted capacity, power consumption dispersion coefficient, line loss rate, and phase angle, etc. These indicators can reflect the different perspectives of electricity consumption behavior. These indicators can reflect the actual situation of power grid operation from different angles, and thus can be used as the main basis for judging whether there is power theft by users.

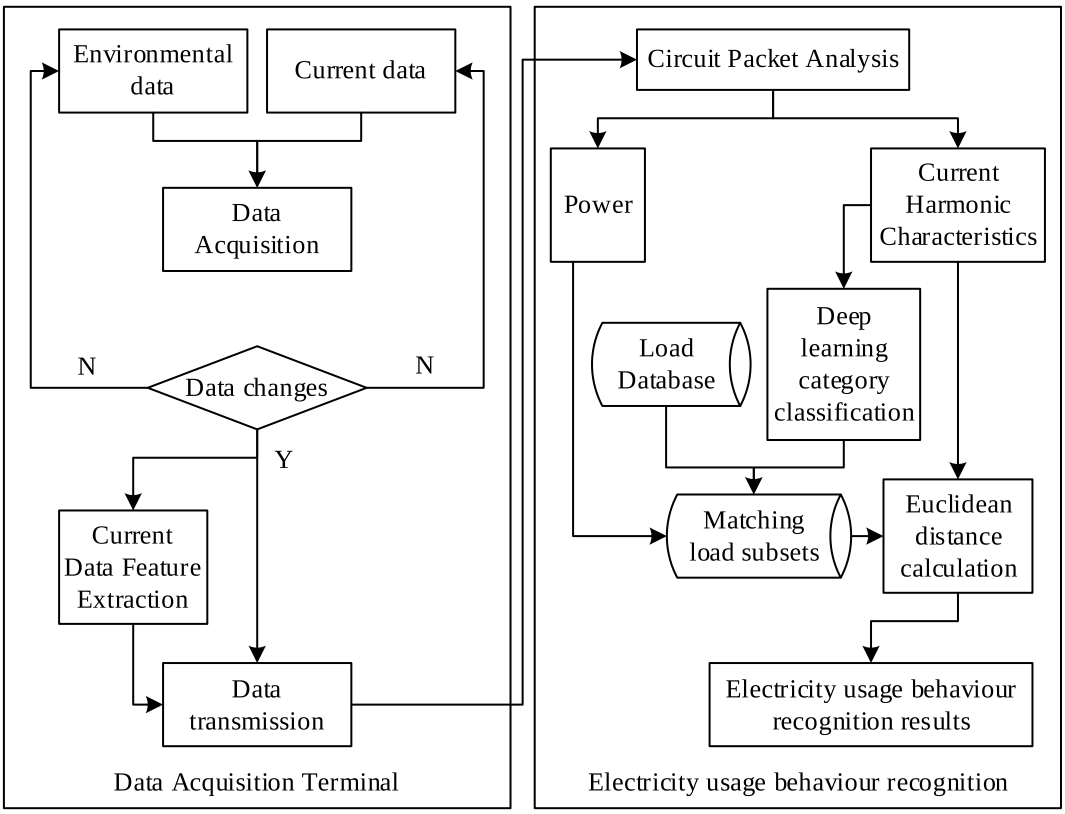


Fig. 1. Monitoring system workflow path for electricity

2.2. *Analysis of electricity theft*

Electricity theft in existing studies refers to the illegal behavior of users who use methods such as tampering with meters, pulling wires privately, and increasing capacity privately to achieve underpayment of electricity bills. The active power calculation formula is:

$$P = UI\cos\varphi \tag{1}$$

where, U and I represent the voltage and current RMS respectively, $\cos\varphi$ is the power factor and P is the average power RMS. The power formula is:

$$W = Pt \tag{2}$$

where, W is the amount of electricity, P and t are the active power and time. According to equation (1) and equation (2), active power is determined by voltage, current and power factor, and active power multiplied by time is equal to the amount of electricity. Therefore, the amount of electricity mainly depends on the voltage, current and power factor, and the vast majority of power theft means are centered around one or more of these three indicators for tampering.

2.2.1. Common means of electricity theft. Existing methods of electricity theft can usually be divided into four categories: undervoltage method, undercurrent method, phase shift method, spreading method.

Under-voltage method of electricity theft refers to the line, junction box or terminal changes to reduce the meter voltage input, so as to achieve the purpose of reducing electricity measurement.

Undercurrent method of electricity theft refers to reduce the current of the meter to achieve the purpose of reducing the measurement of electricity.

Phase shift method of electricity theft refers to tampering with the meter wiring or installation of electricity theft equipment to increase the meter voltage and current phase difference, thereby reducing the measurement of electricity.

Expansion method of electricity theft refers to the destruction of the physical structure of the meter itself, tampering with the parameters, electromagnetic interference and other methods, from the measurement principle of the meter to expand the measurement error, and in general to achieve the purpose of reducing the measurement.

In addition to undervoltage method, undercurrent method, phase shift method, spread the difference between the four means of tampering with the meter to achieve electricity theft, there is a meter has nothing to do with the means of electricity theft, no meter method. The unmetered method of electricity theft refers to the user's private wires, not through the meter directly into the distribution line, resulting in the consumption of electricity can not be measured.

2.2.2. Electricity theft model. For power thieves, regardless of the means of power theft, the purpose is to pay less for electricity, which is ultimately manifested in the reduction of measured power or even zero.

User tampering with the amount of electricity may be continuous or intermittent, may be a fixed percentage of the reduction may be with the change of time and change. Therefore, by analyzing the changes in the time series of the user's electricity in the existing cases of electricity theft, five models can be established based on the changes in the amount of electricity before and after the theft of electricity by the user.

Model 1 assumes that a power thief shrinks the metered power over time by some fixed percentage, mostly undervoltage, undercurrent and stealing power. That is: $X_t = ax_t$, where $0.2 < a < 1$, a is a constant.

Model 2 considers that the power thief reduces the measured power in accordance with a time-varying ratio, which is common in intelligent power theft by tampering with the energy meter. That is: $X_t = ax_t$, where $0.2 < a < 1$, a are random numbers.

Model 3 assumes that in order to minimize the risk of being detected, the power thief sets the metered power to zero by selecting the time period through the remote control switch, which is usually without or with low audit risk. That is: $X_t = f(t)x_t$, where $f(t) = \begin{Bmatrix} 0 & t_1 < t < t_2 \\ 1 & Other \end{Bmatrix}$, t_1 and t_2 are the starting and ending times of the electricity theft period, respectively.

Model 4 assumes that the power thief chooses the time period's to scale down the metered power by the time-varying ratio, thus reducing the risk of being detected and punished. That is: $X_t = f(t)x_t$, where

$f(t)=\begin{cases} 0.2 < a < 1 & t_1 < t < t_2 \\ 1 & Other \end{cases}$. t_1 and t_2 are the starting and ending times of the electricity theft period, respectively.

Model 5 considers that the electricity thief directly disconnects the wiring of the meter or sets the metered power to zero for the whole time period. That is: $X_t = 0$, the whole time period is set to zero.

2.3. *Electricity theft monitoring framework*

The basic architecture of power theft detection used in this paper is shown in Fig. 2. Firstly, the electricity data is preprocessed, and then the abnormal electricity data is simulated and generated according to the electricity theft attack model. Then the power theft detection model is established, the abnormal degree of the detected sample is obtained through model training, and finally the threshold value is used to determine whether the sample has power theft.

The specific process of the deep learning-based power theft detection model is as follows:

Step1: Preprocess the historical electricity consumption data. First of all, the historical electricity consumption data is processed with missing values and abnormal values, and the data of users with a large amount of missing is deleted.

Step2: According to the currently available electricity theft attack model, the normal electricity consumption data obtained from Step1 is used to simulate the generation of electricity theft data in order to verify the generalization ability of the electricity theft detection model. Then the normal sample data and the power theft sample data are fused into a training sample dataset as the input of the power theft detection model and assigned real label values. At the same time, in order to avoid the negative impact on the neural network training caused by the large difference in the values of electricity consumption of different users, the sample data are normalized.

Step3: Establish the power theft detection model, use the training sample data set obtained from Step2 for model training, adjust the model parameters, and calculate the degree of abnormality of the power consumption curve.

Step4: Compare the abnormal degree of power consumption curve output from Step3 with the set threshold of abnormal degree of power consumption curve to get the predicted labeled value, and then compare it with the real labeled value from Step2 to evaluate the performance of the power theft detection model.

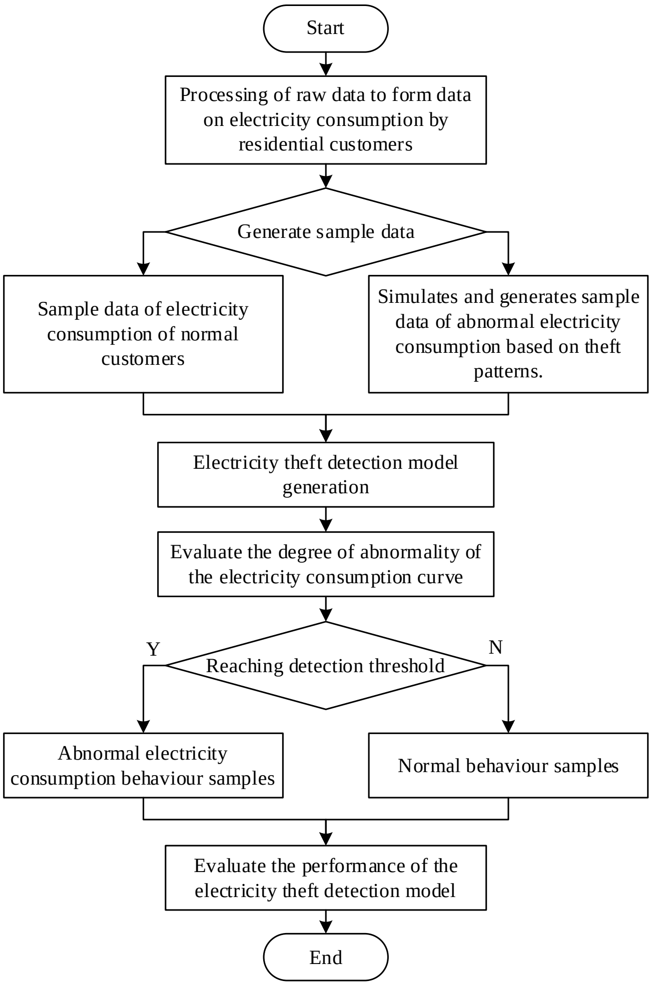


Fig. 2. The basic structure of the journal

3. Electricity theft prediction modeling

3.1. Time series forecasting algorithm

3.1.1. Characterization of time series. Time series data are typically characterized by the high dimensionality of the data, especially multivariate time series data. This characteristic determines that it is difficult to carry out data analysis and data mining work for the original data. In order to improve the computational efficiency of time series data, feature representation methods are usually used [22-23].

Feature representation of time series is a method of transforming time series data into feature vectors for describing and analyzing specific properties and patterns of the time series. These features can include very basic statistical features such as mean, variance, etc. They can also include advanced time and frequency domain features such as autocorrelation, kurtosis, spectrum, etc. By analyzing and processing the feature representation of a time series, the patterns and anomalies of the time series can be mined and identified more effectively, thus realizing tasks such as time series model prediction, anomaly detection, classification, and so on. The main way is to provide a new description of the whole time series data by analyzing the features of the time series data and adopting a new representation to represent its main features. In turn, it replaces the whole time series data with less data to achieve the purpose of improving efficiency and dimension reduction.

3.1.2. Time series anomaly detection. Time series anomaly detection is a method of analyzing time series data to identify abnormal or unusual values, trends, cycles, or patterns in it, as well as to predict possible abnormal or unusual situations in the future, so as to improve efficiency and accuracy in business or engineering applications. Anomalies are usually manifested in time series as values at a point in time or over a period of time that deviate significantly from historical data or patterns.

Semi-supervised anomaly detection uses a small number of existing anomaly samples and a large number of normal samples for anomaly detection. The technique works by modeling normal time series data based on a training dataset and then using that model to distinguish anomalies in the data to be detected. By using unlabeled data, semi-supervised anomaly detection can capture almost all types of anomalies. And because only a very small number of anomaly samples are required, this class of methods is relatively more widely used.

3.2. Electricity theft estimation method based on improved time series prediction algorithm

3.2.1. Least Squares Support Vector Regression Algorithm/LSSVR. For a given set of samples $\{(x,y),i=1,2;\cdot,l\}$, where $x\in R^{n}$, is the input value of the i rd sample. $y_i\in R$, is the corresponding output value.The SVR algorithm maps the inputs into a high-dimensional feature space through some nonlinear mapping $\varphi(\cdot)$. And then the optimal decision function is constructed to solve the problem with the principle of minimizing structural risk. LSSVR, on the other hand, uses a quadratic loss function that transforms the problem into the following form:

$$\min J(w,\xi)=\frac{1}{2}w^{T}w+\frac{c_{l}}{2}\sum_{i=1}^{\xi^{2}}i \tag{3}$$

$$s.t.y_{i}=\varphi(x_{i})w+b+\xi_{i} \tag{4}$$

where $w=[w_1,w_2,\ldots,w_l]^{T}$ is a vector of coefficients to be solved. b is the constant to be solved, c is the regularization parameter, and $\varphi(\cdot)$ is some scalar operator.

The problem is solved using the Lagrangian method:

$$L(w,b,\xi,a)=\frac{1}{2}-w^{T}w+2\sum_{i=1}^{c_{l}}\xi^{2}-\sum_{i=1}^{l}\left\{a_{i}\left[\varphi(x)_{i}\cdot w+\xi_{i}+b-y_{i}\right]\right\} \tag{5}$$

where a_i is the Lagrange multiplier. According to the KKT optimization conditions:

$$\frac{\partial L}{\partial w}=\frac{\partial L}{\partial b}=\frac{\partial L}{\partial \xi_{i}}=\frac{\partial L}{\partial a_{i}}=0 \tag{6}$$

Available:

$$w=\sum_{i=1}^{l}a_{i}\varphi(x_{i}) \tag{7}$$

$$\sum_{i=1}^{1}a_{i}=0 \tag{8}$$

$$\xi_{i}=\frac{a_{i}}{c} \tag{9}$$

$$y_{i}=\varphi(x_{i})\cdot w+\xi_{i}+b \tag{10}$$

Substituting Eq. (7) and Eq. (9) into Eq. (10) gives:

$$y_{i}=\sum_{j=1}^{t}\left(a_{j}\left\langle\varphi(x_{j}),\varphi(x_{i})\right\rangle\right)+b+\frac{a_{i}}{c} \tag{11}$$

If the kernel function is defined to be $k(x_i, x_j) = \langle \varphi(x_j), \varphi(x_i) \rangle$, then:

$$y_i = \sum_{j=1}^{i} \left(a_j k(x_i, x_j) \right) + b + \frac{a_i}{c} \tag{12}$$

Combine Eq. (12) and Eq. (8) into the following system of linear equations:

$$\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & k(x_1, x_1) + \frac{1}{c} & k(x_1, x_2) & \cdots & k(x_1, x_l) \\ 1 & k(x_2, x_1) & k(x_2, x_2) + \frac{1}{c} & \cdots & k(x_2, x_l) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & k(x_l, x_1) & k(x_1, x_2) & \cdots & k(x_l, x_l) + \frac{1}{c} \end{bmatrix} \cdot \begin{bmatrix} b \\ a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ y_2 \\ \vdots \\ y_l \end{bmatrix} \tag{13}$$

With the sample set $\{(x_i, y_i), i = 1, 2, \cdots, l\}$ known, parameter $[b, a_1, a_2, \cdots, a_l]^{\mathrm{T}}$ can be solved according to the system of equations (13). And the decision function used by the LSSVR algorithm to calculate the predicted values is composed of these key parameters:

$$f(x) = \sum_{i=1}^{1} a_i k(x, x_i) + b \tag{14}$$

According to the definition of support vector, the sample corresponding to coefficient $a_i \neq 0$ is the support vector in the calculation. From equation (14), it can be seen that the prediction of the function by the LSSVR algorithm mainly relies on the support vector and the value of coefficient a_i. However, in the actual application of the algorithm test process found that the coefficient a_i are not equal to zero, and there are more coefficients a_i of the absolute value of close to zero, and individual coefficients a_i of the absolute value of a larger. The LSSVR algorithm selects all the training samples as support vectors in the calculation. In order to improve the performance of the algorithm, this paper makes some improvements to the traditional LSSVR and optimizes the support vectors.

3.2.2. Artificial Bee Colony Algorithm/ABC Algorithm:

1) ABC algorithm/ABC algorithm is an optimization algorithm that simulates the process of honey harvesting by bees in nature to find the food source. The quality of the food source corresponds to the fitness of the optimization problem [24]. The formula for updating the position of the food source for the leading and following fronts in the swarm algorithm is:

$$V_{ij} = S_{ij} + r \cdot \left(S_{ij} - S_{ij} \right) \tag{15}$$

where $k \in (1, 2; \cdots L)$, $j \in (1, 2, \cdots, N)$, $k \neq j$. r is a random number between $[-1, 1]$, S_{ij} is the current location of the food source, and S_{kj} is the location of the food source for a randomly selected individual in the field.

The probability that an observer bee chooses the food source selection according to a roulette wheel based on the information shared by the hired bee is:

$$P_i = \frac{f(\delta_i)}{\sum_{i-1}^{L} f(\delta_i)} \tag{16}$$

Depending on whether f_i is greater than 0, the fitness function can be expressed as:

$$f(\delta_i)=\begin{cases}\dfrac{1}{1+f_i} & f_i\geq 0\\ 1+abs(f_i) & f_i<0\end{cases} \tag{17}$$

2) Improvements to the Artificial Bee Swarm Algorithm. The bee swarm algorithm is based on equation (15) to get the location of the new food source. The first term in equation (15) is the current food source information of the individual. The second term is the difference information between the original food source and the neighboring food source. Since the honeybee neighboring individuals were randomly selected, the relationship between the neighboring individual food source and the current food source was not considered. As well as the use of roulette wheel to select the food source is easy to make the algorithm fall into the local optimum. To solve this problem, a perturbation term is added to the artificial bee colony algorithm, and during each iteration, the update of the food source position will be determined by the following equation:

$$\begin{gathered}V_{ij}=S_{ij}+r\cdot\left(S_{ij}-S_{kj}\right)+\left(\alpha-0.5\right)\varphi_j\\ i=1,2,\cdots,N\end{gathered} \tag{18}$$

where α is a random number obeying a uniform distribution varying in the range $[0,1]$. φ_j is the perturbation variable for the j th search for the food source, which takes a positive value. N denotes a N -dimensional space, and the position vector in each dimension denotes the weights of a predictive model. $(\alpha-0.5)\varphi_j$ is an unpredictable random action.

Perturbation variable φ_j can control the strength of the random decision-making behavior of the j nd search for food sources, such as too large to disrupt the following bee's mechanism of selecting food sources, and too small to reduce the population's global search ability, in order to automatically regulate and accelerate the speed of convergence as well as to strengthen the ability of generalization according to the size of the fitness in the process of the bee colony's evolution, this paper defines φ_i as:

$$\begin{aligned}&\varphi_j=\varphi_{\min}\left(\frac{f_{\left(s_j^{1-1-m}\right)}-f_{\left(s_j^{1-1-n}\right)}}{f_{\left(s_j^{1-1}\right)}-f_{\left(s_j^{1-1-m}\right)}}\right)\\ &s.t.\\ &S_j^{1-1}\neq S_j^{1-1-n}\neq S_j^{1-1-m}\\ &m<n,n=\min\{1,2,\cdots,N\},m=\min\{1,2,\cdots,N\}\end{aligned} \tag{19}$$

where, $\varphi_{\min}$ is the perturbation minimum, $f(\cdot)$ is the fitness function, S_j^{1-1}, S_j^{1-1-n}, S_j^{1-1-m}, are the location of the food source searched for the 1-1st, $1-1-n$th, and $1-1-m$th time, respectively, and the value of the perturbation variable φ_j is relatively large at the time of evolution, and the step of change of the j th bee colony is stochastic, and when after a certain amount of evolution, φ_j tends to $\varphi_{\min}$, and the step of change of the colony is stochastic and becomes weaker.

The first part of Eq. (18), which introduces the perturbation term, is the location of the original food source. The second part is the difference from neighboring food sources. The third part is the change step of random irrational behavior. Due to the presence of the perturbation term, even if a local optimum occurs, it is possible to ensure that the new food source of the bee colony is updated to overcome the problem of premature convergence.

During the iteration of the ABC algorithm, it is almost impossible for the worst food source to contribute to the required optimal result, which affects the convergence rate of the algorithm to some extent. The specific way of generating a new food source to replace the worst food source through the OBL strategy is as follows.

In each generation of the cycle, the worst food source is labeled, corresponding to position X_b. The new food source position after invoking the OBL strategy is $\overline{X_b}$, and the new position $\overline{X_{bj}}$ in dimension j is:

$$\overline{X_{bj}} = V_{ij} + S_{jH} + rand \times S_{bj} \tag{20}$$

By substituting the worst food source, not only the convergence speed can be greatly improved, but also to a certain extent, the falling into local optimality can be avoided by replacing the original food source position with the new position if it corresponds to a better food source.

3.2.3. Estimation method of electricity theft based on time series. In order to build a semi-supervised learning model, a new co-learning method is needed, and this subsection proposes to construct a co-training sample matrix by combining current data samples and historical contemporaneous data samples. The co-training sample matrix is utilized to train the LSSVR predictor globally once to obtain the final prediction model. The process of constructing the co-training matrix of current data and historical contemporaneous data needs to consider the difference between the two samples, so the two-sample detection problem is introduced here, which utilizes the distribution distance between the samples to measure the degree of the difference. The MMD method is used to measure the difference between the overall mean of the data in the source domain and the data in the target domain, which is a parameter-less quantity that expresses the difference in the distributions of the source domain and the target domain. In this paper, MMD is used to represent the distribution distance between the current training data samples and the historical contemporaneous data samples.

Definition of MMD If the current data sample $D_S = \{x_{si}\}_{i=1}^{n}$ and the historical contemporaneous data sample $D_T = \{z_{ij}\}_{j=1}^{m}$ satisfy the distributions P and P' distributions, respectively, the MMDs of D_S and D_T in the regenerated kernel Hilbert space (RKHS) are defined as:

$$dist(D_S, D_T) = \left\| E_{x-p}\left[\phi(x)\right] - E_{x-p}\left[\phi(Z)\right] \right\|_H \tag{21}$$

where E denotes the expectation of the distribution. ϕ denotes a nonlinear mapping which maps the samples from the original input space to the RKHS high-dimensional feature space H.

According to the basic theory of RKHS space, Eq. (21) can be rewritten as the following empirical formula, i.e:

$$dist(D_S, D_T) = \left\| \frac{1}{n}\sum_{i=1}^{n}\phi(x_i) - \frac{1}{m}\sum_{j=1}^{m}\phi(z_j) \right\|_H \tag{22}$$

Eqs. (21) and (22) show that the difference in distribution between samples in the RKHS high-dimensional space is equivalent to the difference in means between samples in that space.

In this subsection, the electricity consumption data of power theft users are extracted to form a time series, and an improved semi-supervised learning time series prediction power theft estimation algorithm is established. In order to improve the accuracy of the algorithm, the data need to be preprocessed to filter the noise and form a new time series for the next step of analysis. The algorithm first selects appropriate input training set samples from the preprocessed time series, and combines and transforms the sample data. Secondly, MMD is introduced into the LSSVR algorithm to measure the distance between sample distributions and construct cooperative training samples. And the time series prediction algorithm based on improved artificial bee colony is established by introducing the improved ABC algorithm to optimize the parameters of matrix rows. Finally, the time series to be predicted is input into the trained algorithm to obtain the time series of electricity consumption under normal electricity consumption in the user's electricity theft time zone, and then the size of electricity theft can be estimated. The specific steps are described below:

STEP1: Obtain the electricity consumption data of electricity theft users and select appropriate training sample data. Establish the current time series $D_S = \{x_1, x_2, \cdots x_s\}$ and the time series $D_T = \{x_1, x_2, \cdots x_k\}$ with the length of historical contemporaneous samples as k. Zeroize and normalize D_S and D_T to obtain Eq. (23) and Eq. (24):

$$D_S(t) = Z_1(t) + N_1(t) \tag{23}$$

$$D_T(t) = Z_2(t) + N_2(t) \tag{24}$$

where $Z_i(t)$ represents the normal trend and $N_i(t)$ represents the noise signal.

Fourier transforms $D_S(t)$ and $D_T(t)$ to obtain frequency domains $D_S(Z)$ and $D_T(Z)$, and Fourier inverse transforms to obtain new time series $D'_S(t)$ and $D'_T(t)$ after filtering noise signals, and transforms $D'_S(t)$ to current sample matrix K_T for $l \times n$, and transforms $D'_S(t)$ to historical contemporaneous sample matrix X_T.

STEP2: Using the RBF kernel function K_{RBF}, X_S and X_T are mapped in the kernel space to the kernel matrices K_S and K_T. where K_S has dimension $(n+m)\times n$ and K_T has dimension $(n+m)\times m$. In the high dimensional feature space, the kernel matrices K_S and K_T are the actual inputs to the algorithm.

STEP3: Calculate the value of Λ by $\Lambda^{i,j} = K(\phi(s_i), \phi(s_i))$ and calculate the value of Ω.

STEP4: Set the parameters related to ABC algorithm. Initial number of solutions for initial food source S_N, maximum number of cycles for food source optimization M_{cycle}, maximum number of iterations $N_{\max}$. Set the optimization range for parameters n, ξ, σ and λ.

Initialize the food source according to the following equation of Eq:

$$\begin{gathered} x_y = x_{j_{\min}} + rand(0,1)\left(x_{j\max} - x_{j\min}\right) \\ i = 1, 2, \cdots, S_N \; j = 1, 2, \cdots D \end{gathered} \tag{25}$$

where D denotes the number of parameters to be optimized.

STEP5: According to the characteristics of the LSSVR algorithm, the absolute value of the relative error is chosen as the fitness function, as shown in equation (26), to improve the accuracy of the algorithm. Namely:

$$MAP\left|\frac{\tilde{y}_i - y_i}{y_i}\right| \tag{26}$$

STEP6: Execute the hired bee and observation bee phases.

STEP7: Divide the population according to the fitness function value and population size, individuals with high fitness function value are divided into quality population P, and the rest of the individuals are divided into population Q. Perform mutation operation to update the food source.

STEP8: Perform scout bee phase, discard food sources with no exploitation value and mark them, randomly generate new food sources to continue the search.

STEP9: When the number of iterations exceeds the set number of $N_{\max}$, the search is stopped, and the current optimal food source location is the optimal solution of parameter $(n, \xi, \sigma, \lambda)$.

STEP10: Solve the quadratic programming problem of Eq. β and b. Substitute them to get w. Substitute w and b to get $\tilde{y}(x)=w^T\varphi(x)+b$, and finally inverse normalize $\tilde{y}(x)$ to get the prediction function $\tilde{y}(x)$.

STEP11: The output value Y' of the algorithm is obtained through the training of $\tilde{y}(x)'$. Y' is a row vector, and the performance of the algorithm is analyzed according to the performance index by comparing Y' with the real value. In this paper, the evaluation functions of performance indexes are root mean square error e_{RMSE}, average relative error e_{MAPE} and correlation coefficient R, and the formulas are shown in Eqs. (27), (28) and (29). If e_{RNGE} and e_{MAPE} are smaller and R is closer to 1, it means that the fitting performance of the model is better. i.e:

$$e_{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i-y_i)^2} \tag{27}$$

$$e_{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\tilde{y}_i-y_i}{y_i}\right| \tag{28}$$

$$R=\sqrt{\frac{\left(N\sum_{i=1}^{N}\tilde{y}_iy_i-\sum_{i=1}^{N}\tilde{y}_i\sum_{i=1}^{N}y_i\right)^2}{\left(N\sum_{i=1}^{N}\tilde{y}_i^2-\left(\sum_{i=1}^{N}\tilde{y}_i\right)^2\right)\left(N\sum_{i=1}^{N}y_i^2-\left(\sum_{i=1}^{N}y_i\right)^2\right)}} \tag{29}$$

STEP12: The modeling of power theft estimation method based on semi-supervised learning MMD-ABC-LSSVR time series prediction algorithm can be completed by STEP1~ STEP11. When analyzing the electricity theft in the time segment to be sought, the current time series $D_S^*=\{x_1,x_2,\cdots x_a\}$ (i.e., electricity theft starts from time t_{a+1}) composed of the first a piece of data at the time of electricity theft and the historical contemporaneous time series $D_T^*=\{x_1,x_2,\cdots x_b\}$ of the electricity theft time segment with length b are input into the prediction model, and the time series of electricity consumption under normal electricity consumption in the electricity theft time segment can be obtained, so that the total electricity consumption W of the electricity theft time segment can be estimated, and the estimated value of electricity theft ΔW of the user can be obtained by subtracting the table data W'.

4. Example analysis

4.1. Model parameterization

The dataset for the experiments in this paper is the real electricity consumption dataset released by the National Grid. This dataset contains the electricity consumption data of a total of 68,392 electricity customers within 2068d (from July 1, 2020 to September 31, 2023).

The electricity consumption data of all 5805 electricity theft customers are selected, and then 5805 electricity consumption data of non-stolen electricity customers are randomly selected at 1:1. The training set (8127) and the validation set (3483) were split in a ratio of 7:3 among a total of 11610 electricity consumption data. In addition, 3483 entries were randomly selected from the second file of the dataset as the test set in the same 7:3 ratio of training set to test set. And it is guaranteed that the electricity theft users occupy a certain percentage, where there is no intersection between the test set and the training set. The main fields of each sample include user number, whether the electricity is stolen or not, and daily electricity consumption time series.

4.1.1. Evaluation indicators. Electricity theft detection can be abstracted as a binary classification problem for machine learning. For the electricity theft model it is necessary to choose evaluation metrics to react to the accuracy of the model classification. In this paper, precision rate, recall rate, F1 score and its macro average are selected as evaluation indexes.

The evaluation metrics for each category were synthesized, i.e., they were assessed using the macro-averages of the evaluation metrics: macro-averaged recall R_{macro}, macro-averaged precision, P_{macro}, and macro-averaged F1 scores F_{1macro}. The formulas for their computation were, respectively:

$$R_{macro} = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{30}$$

$$P_{macro} = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{31}$$

$$F_{1macro} = \frac{2\left(P_{macro} \times R_{macro}\right)}{P_{macro} + R_{macro}} \tag{32}$$

4.1.2. Experimental results and analysis. The operating system platform for the experiments in this paper is Windows 11. The hardware environment is CPU Intel(R) Core(TM) i7-7300HQ2.50 GHz. memory 32.00 GB, GPU NVIDIA GeForce GTX 1050 Ti. Video memory 8GB.

The validation set used for the experiments contains electricity consumption data from 1894 non-stealing and 1589 stealing customers, totaling 3843 electricity consumption data. For each model to find its learning rate that can produce the best results. The main experimental parameters are shown in Table 1. The input sequence length of each model is 2068, and the batch size and the number of iterations are the same. In terms of number of layers, the time series prediction model proposed in this paper has 7 layers.

Table 1. Main experimental parameters

Model	Experimental parameter			
	Input sequence length	Bulk size	Iteration number	Layer number
LSSVR	2068	32	300	6
MMD-LSSVR	2068	32	300	6
MMD-ABC-LASSVR	2068	32	300	6
This algorithm	2068	32	300	7

The training of the four models on each interval of learning rate is shown in Table 2. The highest accuracy and corresponding learning rate of each model in each interval are extracted by determining the fitting state by the loss value and accuracy of the training set and validation set in the results of each experiment. The corresponding learning rate of the highest accuracy of the time series prediction model based on the improved artificial bee colony algorithm proposed in this paper can reach five decimal places, i.e., when the highest accuracy is 0.7557, the corresponding learning rate is 0.00005.

Table 2. 4 models of training in each area

Model	Learning interval	Single growth	Maximum accuracy	The highest accuracy rate corresponds to the learning rate
LSSVR	[0.01,0.10]	0.010	0.7506	0.035
	[0.001,0.010]	0.001	0.5012	0.009
MMD-LSSVR	[0.001,0.010]	0.001	0.6361	0.012
	[0.01,0.10]	0.01	0.6894	0.067
MMD-ABC-LASSVR	[0.001,0.010]	0.001	0.6235	0.016
	[0.01,0.10]	0.01	0.6989	0.035
This algorithm	[0.01,0.10]	0.01	0.7125	0.01000
	[0.001,0.010]	0.001	0.7336	0.00125
	[0.0001,0.0010]	0.0001	0.7401	0.00013
	[0.00001,0.00010]	0.00001	0.7557	0.00005

The cases with the respective learning rates that produce the best results are selected separately, and the chosen experimental evaluation metrics are computed after 300 rounds of training: accuracy (Accuracy) reflects the overall correctness of the electricity theft detection and is denoted as A for Accuracy. Macro-averaged Recall R_{macro}, Macro-averaged Precision P_{macro} and Macro-averaged F1 Score F_{1macro}. The results of the models on the validation set at the optimal learning rate are shown in Table 3. The overall accuracy of the four models is above 0.6800. Among them, the accuracy of this paper's algorithm is the highest, and the gap in the performance of the models is mainly reflected in the macro-averaged recall. In the validation set, this paper's algorithm is the model with the highest macro-mean recall, reflecting the model's good detection ability. All the models in this paper have good coverage for the identification of power theft.

Table 3. The results of the model in the optimal learning rate

Model	Learning rate	A	P_{macro}	R_{macro}	F_{1macro}
LSSVR	0.035	0.7506	0.7621	0.7637	0.7654
MMD-LSSVR	0.067	0.6894	0.6991	0.6997	0.7003
MMD-ABC-LASSVR	0.035	0.6989	0.7001	0.7010	0.7015
This algorithm	0.00005	0.7557	0.7978	0.8196	0.7942

4.2. Power consumption anomaly analysis based on power consumption behavior

Based on the theoretical basis that the same type of users have the same characteristics of electricity consumption behavior, the comparison between the user's electricity consumption behavior and the typical load characteristics of the group to which the user belongs is carried out. As a result, users with a large difference in load characteristics are considered abnormal power users.

Based on the time series decomposition, the user load characteristic curve is recognized. The load characteristic curve is then vectorized to establish the load characteristic matrix, and on the basis of the load characteristic matrix, the users are grouped according to their behavioral characteristics of electricity consumption through clustering and subgrouping. After clustering, the load characteristics of users have similarity, based on which the standard load characteristic curve of the user group is established. Finally, the load characteristic correlation calculation is used to identify the abnormal power users, and further correlation analysis is used to lock the suspected power theft users. The load curve based on time series decomposition is shown in Figure 3. Both the standard load profile and the user load profile are in the range of 19h~22h, and the load value reaches the maximum range.

The standard load characteristic curve and the user load characteristic curve are split equidistant in time to form a two-dimensional array, and the correlation coefficient method is applied to each time to calculate its degree of correlation, and then the overall degree of correlation is calculated for all the time segments, and finally the comprehensive correlation coefficient of the two curves is obtained, which is used as the basis for evaluating the consistency of the user's electricity consumption behavior with the standard electricity consumption behavior.

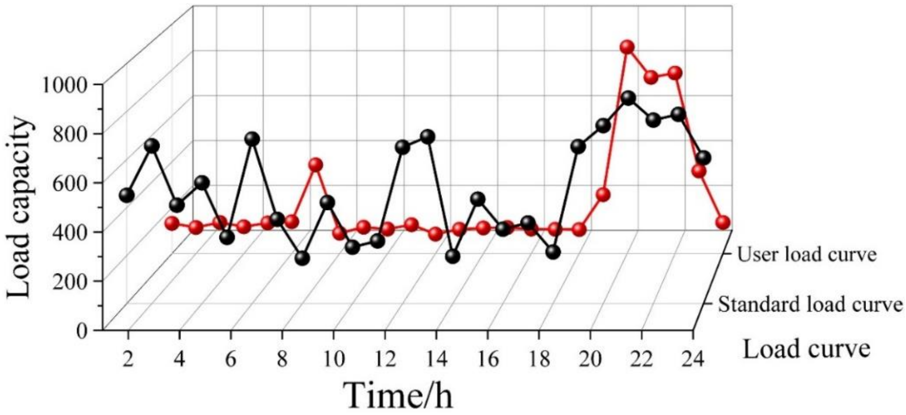


Fig. 3. Load curve based on time series decomposition

4.3. Analysis of the results of the prediction of potential electricity theft

In order to achieve the goal of optimal economic benefits, this paper combines the improved artificial bee colony algorithm with the MMD training strategy, and proposes a time prediction analysis model based on the improved artificial bee colony algorithm, which realizes the accurate prediction of users' potential power theft (PSE). And on the basis of this, the economically optimal user ranking order is determined.

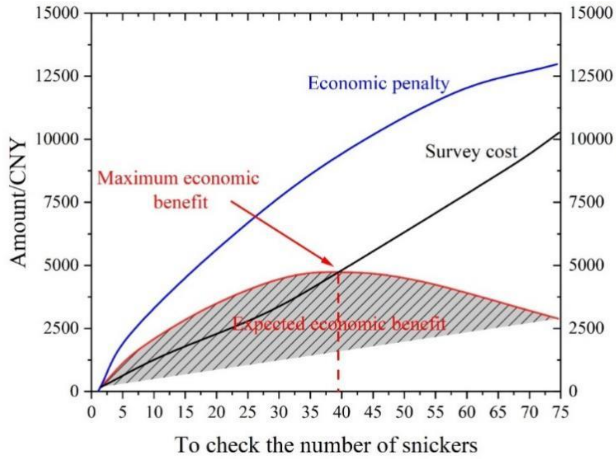
In the PSE prediction of power theft users based on the temporal predictive analysis model of the improved artificial bee colony algorithm, the number of power theft users in each dataset is shown in Table 4.

The users in the auxiliary training set include all historical electricity theft records, and the proportion of the electricity theft users to be predicted is 40% in the test set. The length of time for both electricity theft and normal electricity use is set to 50.

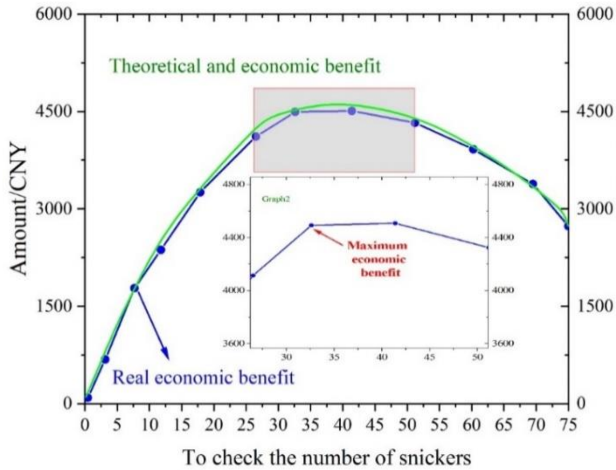
Table 4. The number of snickers in each data center

Data set type	Sample scale	Data set 1		Data set 2	
		Number of users	Time cycle	Number of users	Time cycle
Auxiliary training set	100%	3843	2023/03~2023/9	5805	2020~2023
Target training set	60%	2305	2023/06	3843	2022
Test set	40%	1538	2023/06	1962	2022

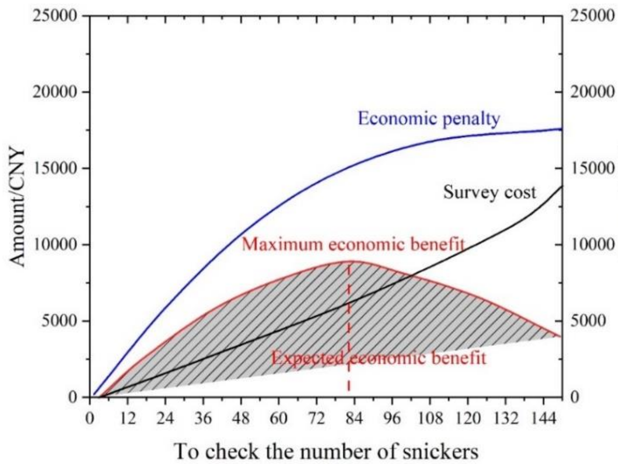
In both cases, the maximum number of iterations, maximum labeling error, and minimum weight were set to 50, 10, and $0.5\alpha^1$, respectively. Parameters related to economic efficiency (φ_{pe}, φ_{pr}, $\overline{N}$, φ_f, φ_c, and φ_e). Are set to 5, 0.15, 50, 500, 300, 100 in dataset 1. 6, 0.10, 300, 300, 100, 50 in dataset 2. The PSE of the power theft users in the test set is predicted by using a temporal predictive analytical model based on the improved artificial bee colony algorithm, and the economic curves of the power theft exclusion are generated. The economic curve of power theft exclusion based on the algorithm of this paper is shown in Fig. 4, and the actual maximum economic benefit obtained by this algorithm in dataset 1 and dataset 2 is basically equal to the theoretical maximum economic benefit. Data set 2 is taken as an example in Fig. (c), and it can be seen that the optimal number of users to be excluded is 83 in the desired economic curve. In Fig. (d), the theoretical and actual economic benefit curves are generated with the user's real PSE as input, and with the theoretical and predicted user order, respectively.



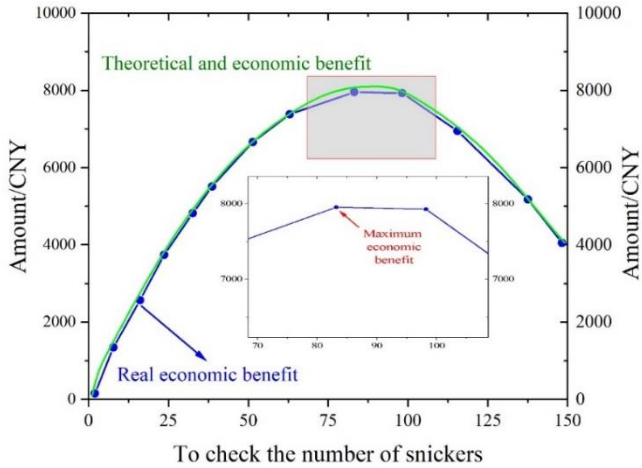
(a) Expected economic curve based on predicting PSE (data set 1)



(b) Theory and practical economic benefit curve (data set 1)



(c) Expected economic curve based on predicting PSE (data set 2)



(d) Theory and practical economic benefit curve (data set 2)

Fig. 4. Based on the algorithm's snicker economic curve

The PSE prediction evaluation metrics of each algorithm are shown in Table 5. It can be seen that the proposed algorithm has the lowest global error, which is 0.0003 and 0.0027 in dataset 1 and dataset 2, respectively. Therefore the overall accuracy of PSE prediction is the highest.

The improved time series based algorithm predicts the highest accuracy of priority order and hence based on its predicted PSE the priority order of most of the users can be determined correctly.

The maximum economic benefit realization rate of the proposed improved artificial bee colony algorithm's time prediction model is 99.45% and 99.76% in dataset 1 and dataset 2, respectively. Because the list of users to be scheduled predicted by the algorithm is basically the same as the list of users determined by the real PSE, the maximum economic benefit can be achieved.

Upon analysis, it is concluded that the reason why the proposed algorithm performs best in PSE prediction and economic efficiency optimization strategies is that, firstly, the algorithm supplements the auxiliary training set as input, which expands the input user samples and makes the training process more effective. In addition, the introduction of MMD training strategy improves the feature learning ability on the target training set. It also automatically weakens the influence of irrelevant auxiliary samples on the training process and enhances the training of the target samples in each iteration.

Table 5. The PSE of the algorithms predicts the evaluation index

Algorithm	Data set 1			Data set 2		
	Global error	Priority order accuracy	Maximum economic efficiency	Global error	Priority order accuracy	Maximum economic efficiency
CART	0.0032	0.9007	0.9785	0.0152	0.7924	0.9133
RF	0.0028	0.7985	0.9689	0.0097	0.5986	0.9214
SVR	0.0069	0.7011	0.9451	0.0784	0.6784	0.8647
LSSVR	0.0025	0.8563	0.9791	0.0061	0.7057	0.9255
MMD-LSSVR	0.0018	0.9124	0.9806	0.0037	0.7699	0.9361
MMD-ABC-LASSVR	0.0012	0.9135	0.9880	0.0029	0.8006	0.9885
This algorithm	0.0003	0.9257	0.9945	0.0027	0.8354	0.9976

5. Conclusion

This paper applies the time series anomaly detection method to the monitoring of electricity consumption behavior and prediction of electricity theft, and proposes a power theft estimation method that combines an artificial bee colony algorithm and a time series analysis model. Based on the power theft detection framework to identify the power theft behavior and predict the potential power theft of power theft users, in order to obtain the maximum economic benefits of power theft investigation.

Using the real dataset as the model validation data, the macro-averaging of each evaluation metric is evaluated to measure the training of each model on each interval of the learning rate, and to compare the results of each model on the validation set at the optimal learning rate. After the same number of training sessions, the overall accuracy of each comparison model is above 0.6800. The macro average recall R_{macro}, macro average precision, P_{macro} and macro average F1 score F_{1macro} of this paper's algorithm are the highest values, which provide better coverage for the identification of electricity theft.

To realize the optimal economic benefit, the improved artificial bee colony algorithm is combined with the MMD training strategy to measure the users' potential power theft (PSE). When the optimal number of users to be discharged is 83, the discharging cost coincides with the expected economic benefit, and the maximum economic benefit point appears. And in the process of increasing the number of users to be excluded, the trend of the actual economic benefit is basically consistent with the theoretical economic benefit. Thus, the user PSE predicted by the time series model based on the artificial bee colony algorithm proposed in this paper can accurately determine the list of users to be checked.

References

[1] Zhai, X., Liu, Y., Yue, W., Shi, H., Sun, Y., He, Y., & Liu, H. (2018, May). Comprehensive evaluation system of stealing electricity. In 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (pp. 75-79). IEEE.

[2] Li, X., Huang, R., Lu, C., Zhang, X., & Ge, K. (2022, October). Research on Anti-Stealing Technology Based on Weather Feature Conditions in Electricity Market. In 2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA) (pp. 436-441). IEEE.

[3] He, W. (2021). Power Marketing Inspection and Anti-stealing Power Based on the Clustering and Correlation Analysis. In 2020 International Conference on Data Processing Techniques and Applications for Cyber-Physical Systems: DPTA 2020 (pp. 225-230). Springer Singapore.

[4] Liu, K., Liu, A., Ma, X., & Jia, X. (2024). Artificial Intelligence-based Real-Time Electricity Metering Data Analysis and its Application to Anti-Theft Actions. International Journal of Advanced Computer Science & Applications, 15(6).

[5] Liu, S., Liang, Y., Wang, J., Jiang, T., Sun, W., & Rui, Y. (2020). Identification of stealing electricity based on big data analysis. Energy Reports, 6, 731-738.

[6] Hu, L., Su, C., Tian, C., & Meng, X. (2022). Evaluation Model of Intelligent Anti-electric Theft Management. In 2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City: Volume 2 (pp. 103-111). Springer Singapore.

[7] Wang, Z., Gong, Y., Chen, Z., Sun, N., Wang, W., Guo, Y., & Zhao, F. (2021). An Electromagnetic Field Detection Module of Preventing Stealing Electricity. In Application of Intelligent Systems in Multi-modal Information Analytics: 2021 International Conference on Multi-modal Information Analytics (MMIA 2021), Volume 2 (pp. 514-521). Springer International Publishing.

[8] Wang, Y., Pan, M., Lan, Z., Wang, L., & Sun, L. (2019). An Electricity-Stealing Identification Method Based on Outlier Algorithm. In Frontier Computing: Theory, Technologies and Applications (FC 2018) 7 (pp. 1383-1388). Springer Singapore.

[9] Hong, H., Su, Y., Zheng, P., Cheng, N., & Zhang, J. (2021). A SVM-based detection method for electricity stealing behavior of charging pile. Procedia Computer Science, 183, 295-302.

[10] Lv, W., Chen, J., Sun, Y., Xu, D., & Liu, X. (2024, January). Application of Data Mining Technology in Anti

Electricity Theft System. In Proceedings of the 2024 International Conference on Power Electronics and Artificial Intelligence (pp. 752-757).

[11] Chen, L., Dong, X., Chen, B., Qiu, X., Wu, Z., Liu, Z., & Lei, Q. (2022). Research on Power-Stealing Behaviors of Large Users Based on Naive Bayes and K-means Algorithm. In Proceedings of the 11th International Conference on Computer Engineering and Networks (pp. 358-367). Springer Singapore.

[12] Zheng, J. (2023, December). Research on Anti-stealing Algorithm of Distributed Photovoltaic Power System Based on Power Big Data. In 2023 International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC) (pp. 776-781). IEEE.

[13] Huang, R., Lu, C., Hu, S., Tang, D., & Pan, B. (2022, February). Research on Characteristics of Electricity Theft Behavior and Risk Early Warning Technology Based on Big Data Information of the Internet of Things. In 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA) (pp. 123-127). IEEE.

[14] Lin, G., Feng, H., Feng, X., Wen, H., Li, Y., Hong, S., & Ni, Z. (2021). Electricity theft detection in power consumption data based on adaptive tuning recurrent neural network. Frontiers in Energy Research, 9, 773805.

[15] Zhang, J., Chen, Q., Zhang, S., & Shao, X. (2019, August). Dedicated Transformer Stealing Identification System Based on Big Data Analysis Technology. In 2019 10th International Conference on Information Technology in Medicine and Education (ITME) (pp. 642-644). IEEE.

[16] Zhang, M., Liu, X., Shang, Y., Kang, L., Wu, Q., & Cheng, W. (2022). Research on comprehensive diagnosis model of anti-stealing electricity based on big data technology. Energy Reports, 8, 916-925.

[17] Yan, S., Liu, X., Jiang, Y., Ma, L., Wang, X., Li, Z., & Lei, J. (2015, November). Analysis of Electricity Stealing and Research of Anti-stealing Measures. In 2015 3rd International Conference on Machinery, Materials and Information Technology Applications (pp. 911-918). Atlantis Press.

[18] Li, Y., Wang, Q., Zhang, D., Sun, X., Shen, Y., & Xu, X. (2016, July). Research and application of electricity anti-stealing system based on neural network. In 2016 3rd International Conference on Information Science and Control Engineering (ICISCE) (pp. 1039-1043). IEEE.

[19] Niu, R., Liu, J., Zhang, X., Guo, W., & Pan, B. (2022, February). Research on Risk Analysis Technology of Electricity Stealing Behavior Characteristics in Smart Grid. In 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA) (pp. 128-131). IEEE.

[20] Chenxi, D., Zhiyong, Z., Mingyang, Y., Wei, Z., & Lei, Z. (2022, September). Research on Intelligent Anti-Stealing System of Computer Big Data Electricity Information Collection System. In 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 813-816). IEEE.

[21] Tian, C., Su, C., Yang, C., & Zheng, Y. (2021). Research on Intelligent Monitoring System for Anti-stealing Electricity. In Big Data Analytics for Cyber-Physical System in Smart City: BDCPS 2020, 28-29 December 2020, Shanghai, China (pp. 714-721). Springer Singapore.

[22] Amna Khalil,Mazhar Javed Awan,Awais Yasin,Tanzeela Kousar,Abdur Rahman & Mohamed Sebaie Youssef.(2024).Hybrid Time Series Model for Advanced Predictive Analysis in COVID-19 Vaccination.Electronics(13),2468-2468.

[23] Xuyi Shi,Jiachen Guang & Liangsu Shao.(2023).Wordle data analysis based on time series analysis model.Academic Journal of Mathematical Sciences(2),

[24] Yan Du,Zhi Cai & Zhiming Ding.(2024).Query Optimization in Distributed Database Based on Improved Artificial Bee Colony Algorithm.Applied Sciences(2).