ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Application of Internet of Things and Density Clustering Optimization Algorithm in Smart Building Energy Consumption Data Analysis and Platform Development

Jian Pi[1,*], *Ze Xu*[1], *Yaou Lv*[1], *Zihao Zhao*[1] *and Yunfang Wang*[1]

[1] HCIG Xiong'an Construction Development Co., Ltd., Xiongan, Hebei, 070001, China

SUMMARY: *In order to achieve intelligent analysis of building energy consumption using data mining theoretical models and to provide support for energy management solutions for buildings, this study optimizes the DBSCAN clustering algorithm using the improved Tenebrae swarm optimization algorithm, applies the IBSO-DBSCAN algorithm to the data analysis of building energy consumption, and develops an intelligent analysis platform for building energy consumption in the Internet of Things (IoT). The clustering performance of the IBSO-DBSCAN algorithm and the optimization effect of the BSO algorithm are verified through experiments, and the accuracy of the IBSO-DBSCAN algorithm reaches 97.33%, which is 8% higher than that before optimization. The building energy use pattern is clustered into three categories, the energy consumption of category 1 is close to the horizontal state, the energy consumption of category 2 and category 3 shows the "double peak" characteristics, and the energy consumption of the buildings in category 1~3 increases in turn. The DBSCAN clustering algorithm in this paper can be used in the data mining of intelligent analysis platform to provide managers with energy consumption pattern information and decision support.*

KEYWORDS: *bso algorithm; dbscan algorithm; cluster analysis; building energy consumption; intelligent platforms*

1 Introduction

The energy consumption of buildings has become one of the three "big energy consumers" in China, along with industrial energy consumption and transportation energy consumption, and with the continuous rise of the total building volume and the improvement of living comfort, it also shows a sharp upward trend [1-3]. Reduce the energy consumption of public buildings is to realize the top priority of building energy efficiency. Since the beginning of the "Eleventh Five-Year Plan", China has been carrying out the construction of various types of building energy consumption monitoring platforms, and building energy consumption monitoring platforms around the world have entered the operation and maintenance period, and the platform owners are eager to ensure the quality of operation, explore the value of data, and improve energy-saving performance [4-7]. However, these platforms mainly rely on manual completion in the preparation process, and there are problems such as duplication, iteration, low efficiency, etc. With the development and application of the Internet of Things (IoT) and density clustering optimization algorithms, the development of an intelligent building energy consumption data analysis platform is of great significance for the

---

*pijian_1984@126.com
https://doi.org/10.65102/is2026630

optimization of building energy consumption.

The Internet of Things (IoT), i.e., “Internet of Everything”, is an extension and expansion of the network based on the Internet, which combines various information sensing devices with the Internet to form a huge network, realizing the interconnection and interoperability of people, machines, and things at any time and any place [8-11]. And density clustering optimization algorithm is a kind of clustering algorithm commonly used in data mining and machine learning, which is able to discover data points with similar density and classify them into different clusters [12, 13]. In the development of smart building energy consumption data analysis platform, the main task of IoT lies in the acquisition of real-time building data [14, 15], while the density clustering optimization algorithm is able to perform clustering analysis and predictive modeling based on these data, and provide visualization of building energy consumption, anomaly alarms, and optimization suggestions, so as to realize the platform's refined management of building energy consumption [16, 17].

In this paper, an IBSO-DBSCAN building energy consumption data analysis method is proposed through the adaptive inertia weight updating strategy and the normal cloud model updating location, which improves the Tenneco clustering optimization algorithm, and uses the improved BSO algorithm to optimize the DBSCAN algorithm to realize the adaptive selection of parameters Eps and $MinPts$. Subsequently, the clustering algorithm is applied to the IoT building energy consumption intelligent analysis platform to realize the data mining function of the platform and provide support for the energy management scheme of buildings. The evaluation of the clustering performance of building energy consumption data is carried out by using profile coefficients, comparing the classification accuracy of the DBSCAN algorithm before and after optimization based on the Iris dataset, and analyzing the optimization effect of the improved BSO algorithm on the DBSCAN algorithm. Then, the energy consumption data of a large office building is collected and analyzed by clustering, which realizes the organic combination of theoretical method research and practical application and proves the effectiveness and stability of the algorithm proposed in this paper.

2 IBSO-DBSCAN methodology for analyzing building energy consumption data

In order to analyze the energy consumption data of intelligent buildings and mine the building energy use patterns with different characteristics, this paper applies the density clustering algorithm (DBSCAN) and optimizes the DBSCAN algorithm using the Improved Bovine Beefy Cluster Optimization algorithm (IBSO), in order to enhance the global search capability of the DBSCAN algorithm.

2.1 Process of analyzing building energy consumption data

Building energy consumption data analysis is the starting point for energy auditing and energy saving program analysis of the building, which directly reflects the building's energy use status, and by benchmarking with the energy consumption benchmark can detect the abnormal status of the building's operation process in a timely manner. The flow of building energy consumption data analysis is shown in Figure 1, which is divided into three steps: data collection, data processing and energy use pattern clustering. The total building energy consumption data are collected and processed with distorted data, and then the energy consumption data are analyzed by IBSO-DBSCAN clustering to mine the typical daily energy use patterns of intelligent buildings.

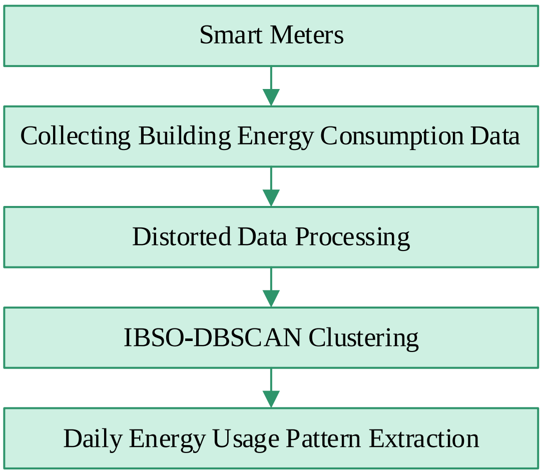


Figure 1: Building energy data analysis process

2.2 DBSCAN clustering algorithm

DBSCAN is a classical density-based spatial clustering algorithm that defines a traditional concept of density to cluster data points located in densely connected regions that need to satisfy two conditions: one is that the distance between them should be less than a given threshold; and two is that they need to be densely packed with each other, i.e., they contain a sufficient number of data points within a certain radius. The given radius is denoted by ε. The number of other data points contained within a spherical neighborhood of radius ε for each data point is defined as the data point density, which is expressed by equation (1):

$$\rho_i = |N_\varepsilon(x_i)| \tag{1}$$

where $N_\varepsilon(x_i) = \{x_j \in X \,|\, d_{i,j} \le \varepsilon\}$, $d_{i,j}$ denotes the distance between x_i and x_j.

DBSCAN categorizes data points into three types: core, boundary and noise points. The algorithm considers data points as points in space and determines the boundary of the cluster based on the density around each point. Core points are data points that contain at least $MinPts$ neighboring points (including themselves) within a radius of ε. $MinPts$ is known as the minimum number of samples. That is, a core point has a sufficient number of neighboring point densities to be considered part of a dense region. Core points are the center points of clusters and they can be connected to form one or more clusters.

A boundary point is a data point that contains fewer than $MinPts$ neighboring points within a radius of ε, but it lies within the ε neighborhood of the core point. In other words, a boundary point does not have a sufficient density of neighboring points to be considered a core point, but it lies within the neighborhood of the core point and is therefore classified as a boundary point. A boundary point may be a point on the boundary of a cluster or a transition point between two different clusters.

Noise points are data points that are neither core nor boundary points. Noise points are located in low-density regions where they do not have a sufficient number of neighboring points and do not belong to any clusters. Noise points are considered as invalid data in the clustering process and are excluded from the clustering results. As shown in equation (2):

$$\begin{array}{ll} \text{If } \rho_i \ge MinPts & \text{Then } x_i \text{ is the core point} \\ \text{If } \rho_i < MinPts & \text{Then } x_i \text{ is the boundary point or noise point} \end{array} \tag{2}$$

When neighborhood radius ε is small, there may be many isolated core points, resulting in too many clusters. Conversely, when the neighborhood radius ε is large, there may be overlapping clusters, resulting in too few clusters. The minimum number of samples $MinPts$ determines the minimum density at which a point is considered to be a core point, and if the value of $MinPts$ is small, there may be more noisy points.

The input parameters of the DBSCAN algorithm include a dataset X, neighborhood radius ε and minimum number of samples in the neighborhood $MinPts$. The steps of the algorithm are as follows:

(1) Randomly select an unvisited point p and mark it as visited.

(2) Find all points in the ε-neighborhood of point p, and if the number of points in this neighborhood is greater than or equal to $MinPts$, take point p as a core point and add all points in its neighborhood to the current cluster. If the number of points in the neighborhood is less than $MinPts$, mark point p as a boundary point.

(3) Select an unvisited point q from the neighborhood of the core point and mark it as visited. If point q is also a core point, add all points in its neighborhood to the current cluster.

(4) Repeat step (3) until all points in the current cluster have been visited.

(5) Repeat steps (1) to (4) until all points have been visited.

(6) Mark all the points that have not been assigned to the cluster as noise points.

The output of the DBSCAN algorithm is a number of clusters, each consisting of a number of core and boundary points.The advantage of the DBSCAN algorithm is that the number of clusters can be identified automatically and clusters of different sizes and shapes can be identified for regions of different densities. The disadvantage is that for high dimensional data, the clustering may not be effective due to dimensional catastrophe.

2.3 Construction of the IBSO-DBSCAN algorithm

2.3.1 BSO algorithm

The Bovine Swarm Optimization (BSO) algorithm is a new swarm intelligent optimization algorithm. In the BSO algorithm, each aspen represents a solution within the solution space, and the number of the population is denoted by N and the dimension is denoted by D. All individual aspens have a randomly initialized position X_i and velocity V_i in the solution space, and the position of the aspens at the tth iteration is denoted by $X_i^t = \left(x_{i,1}^t, x_{i,2}^t, \cdots, x_{i,k}^t\right)$, and the velocity is denoted by $V_i^t = \left(v_{i,1}^t, v_{i,2}^t, \cdots, v_{i,k}^t\right)$, where $1 \le i \le N$, $1 \le k \le D$. The optimal position of the individual is denoted by $pb_i^t = \left(pb_{i,1}^t, pb_{i,2}^t, \cdots, pb_{i,k}^t\right)$ for $pbest_i$, and the optimal position of the group is denoted by $gb^t = \left(gb_1^t, gb_2^t, \cdots, gb_k^t\right)$ for $gbest$.

(1) Position and velocity update of individual aspens. Each tenebrion will search for food source by the concentration of odor detected by the left and right whiskers, and will also move by combining the information of the individual optimal position $pbest$ and the group optimal position $gbest$. The specific position and velocity update equations are as follows:

$$x_{i,k}^{t+1} = x_{i,k}^t + \lambda v_{i,k}^t + \left(1-\lambda\right)\xi_{i,k}^t \tag{3}$$

$$v_{i,k}^{t+1} = \omega v_{i,k}^t + c_1 r_1\left(pbest_{i,k}^t - x_{i,k}^t\right) + c_2 r_2\left(gbest_k^t - x_{i,k}^t\right) \tag{4}$$

where λ is a positive constant to control the effect of velocity and position increments on the positional movement of individual aspens, and λ is set to 0.3 in this paper, ω is the inertia weight, c_1 and c_2 are learning factors, which take the value of 1.2 in this paper, and r_1 and r_2 are random numbers in the range of $\left[0,1\right]$.

(2) Positional growth. Individual tenebrous cattle will use the distance of the left and right whiskers from the food source, together with the sign function and its own speed change, to comprehensively determine the direction of the next search for the food source, and at the same time determine the specific distance to move based on the step update mechanism of the tenebrous cattle whisker search algorithm. The updating formula of position growth is as follows:

$$\xi_{i,k}^{t+1} = \delta^t \cdot v_{i,k}^t \cdot sign\left(f\left(x_{r,k}^t\right) - f\left(x_{l,k}^t\right)\right) \tag{5}$$

$$\delta^t = \theta \cdot \delta^{t-1} \tag{6}$$

where δ^t denotes the step size at the t nd iteration of the tenebrae, $sign$ is the sign function, $x_{l,k}^t$ and $x_{r,k}^t$ denote the positions of the left and right whiskers of the tenebrae in the k th dimension at the t th iteration, respectively, and $f\left(x_{l,k}^t\right)$ and $f\left(x_{r,k}^t\right)$ are the fitness values of the left and right whiskers of the tenebrae in the k th dimension at the t th iteration, respectively, and the tenebrae will be biased to the right at the next iteration if $f\left(x_{r,k}^t\right) > f\left(x_{l,k}^t\right)$, and to the left otherwise. θ controls the degree of attenuation of the individual search step of the aspens.

(3) Position update of the left and right whiskers of the tenrecs. The tench will update the position of the left and right whiskers according to the position and velocity of the left and right whiskers, as well as the distance d between the two whiskers. The specific definitions are as follows:

$$\begin{cases} x_{r,k}^{t+1} = x_{r,k}^t + v_{i,k}^t \cdot \dfrac{d}{2} \\ x_{l,k}^{t+1} = x_{l,k}^t - v_{i,k}^t \cdot \dfrac{d}{2} \end{cases} \tag{7}$$

2.3.2 Adaptive Inertia Weight Update Strategy

In this paper, we propose a new inertia weight updating method to replace the predefined fixed strategy by dynamically adjusting the inertia weight of each tenebrion based on the number of iterations and the quality of individuals at different stages of the algorithm iterations. The new inertia weight updating rule is defined as follows:

$$s_1\left(t\right) = \frac{2t}{T_{\max}} - 1 \tag{8}$$

$$s_2\left(fitness\right) = 1 - \frac{2\left(fitness - fitness_{best}\right)}{fitness_{wrost} - fitness_{best}} \tag{9}$$

$$\omega(t,\ fitness)=\frac{1}{2}\left(\frac{1}{1+e^{s_1(t)/\beta_1}}+\frac{1}{1+e^{s_2(fitness)/\beta_2}}\right) \tag{10}$$

where t is the current number of iterations of the algorithm, β_1 and β_2 are constants. $s_1(t)$ is a function that increases linearly with the number of iterations t, $fitness$ is the fitness value of the individual, and $fitness_{best}$ and $fitness_{wrost}$ are the historical optimal and worst fitness values of the individual, respectively.

2.3.3 Normal cloud model update location

In this paper, a normal cloud model is introduced to improve the position update mechanism of the BSO algorithm to adaptively adjust the global search ability and local exploration ability of the aspen individuals during the operation of the algorithm. The algorithm first calculates the three numerical eigenvalues of the normal cloud model based on Eq. (12), and then uses the forward cloud generator to randomly generate N_{cloud} cloud droplet representing the position of an individual aspen in the solution space $cloud-x$. Finally, it calculates the adaptive degree of an individual aspen and compares them to arrive at the new optimal position of the population and the adaptive degree value. The normal cloud model and the aspen population optimization algorithm system, in this paper, the expectation value of the normal cloud model is set to the current global optimal position $gbest$ to ensure that the worst quality individual in the population can also move to the global optimal position, so as to speed up the convergence of the algorithm and help the algorithm to obtain a more accurate solution. The entropy En and hyperentropy He are defined as functional forms that decrease with the number of iterations to ensure that the individuals can expand the search space and strengthen the global search at the beginning of the algorithm, and the algorithm can carry out a fine search at the end of the algorithm to speed up the convergence speed. The position updating formula of the tenebrous individual is shown in Eq. (11), and the three numerical eigenvalues will be adjusted according to Eq. (12):

$$cloud-x=C\left(gbest,En,He\right) \tag{11}$$

$$\begin{cases} Ex_k=gbest_k \\ En_k=\left(ub_k-lb_k\right)\cdot\left(1-\dfrac{t}{T_{\max}}\right)^{10} \\ He_k=0.1\cdot En_k \end{cases} \tag{12}$$

2.3.4 Adaptation function design

The CS metrics for evaluating the clustering effect are selected as the fitness function of the IBSO-DBSCAN algorithm. The specific fitness function is as follows:

$$fitness=\frac{\sum_{i=1}^{m}\left(\frac{1}{n_i}\sum_{x_j\in C_i}\max_{x_i\in C_i}d\left(x_i,x_j\right)\right)}{\sum_{i=1}^{m}\left(\max_{j\in m,j\neq i}d\left(v_i,v_j\right)\right)} \tag{13}$$

where m is the number of clusters included in the clustering result, n_i is the number of samples in cluster i, $\max_{x_i \in C_i} d\left(x_i, x_j\right)$ is the maximum distance between any two points within class cluster C_i, and $\max_{j \in m, j \neq i} d\left(v_i, v_j\right)$ is the maximum distance between different clusters.

2.3.5 Algorithm flow

The process steps of the IBSO-DBSCAN algorithm are as follows:

Step 1: Initialize the algorithm parameters. Initialize the number of populations N, the maximum number of iterations $T_{\max}$, the initial position of the populations X_i, the speed V_i, and the parameters δ_0, c and θ.

Step 2: Calculate the individual fitness value of aspen group. Based on equation (12), calculate the individual fitness value of aspen and compare the obtained individual extreme value $pbest$ and group extreme value $gbest$.

Step 3: Update the left and right whisker positions and fitness values of the individual aspen group. Calculate the left and right whisker positions of the individual tenebrae group according to equation (6), and calculate the fitness values corresponding to the left and right whiskers of the individual tenebrae group according to equation (13).

Step 4: Update the position increment ξ, velocity V_i, position X_i, and fitness value $fitness$ of the individual aspen herd based on Eq. (3) to Eq. (6).

Step 5: Update the individual position of the aspen herd using the normal cloud model. Based on Eq. (11) update the individual positions $cloud-x$ of the aspen herd and calculate the cloud droplet fitness $cloud-f$ of the individual aspen herd.

Step 6: Calculate and record the optimal fitness value and optimal position of individual and group of aspen bulls. Compare the fitness $fitness$ and cloud droplet fitness $cloud-f$ of individual aspen cows to obtain the optimal fitness value $fitness_{best}$ and the optimal position $gbest$ of the group.

Step 7: Update the step size δ based on equation (7).

Step 8: Judge whether the termination condition of the algorithm is satisfied. If the termination condition is satisfied, the algorithm ends and the output result is the clustering result corresponding to the global optimal parameters Eps and $MinPts$, otherwise return to step 2.

3 Intelligent analysis platform for IoT building energy consumption data

The purpose of the implementation of the IoT building energy consumption intelligent analysis platform is to be able to utilize the researched theoretical model of data mining (IBSO-DBSCAN algorithm) to realize the intelligent analysis of building energy consumption and to provide support for energy-saving solutions for buildings.

3.1 Overall technical architecture

Platform development and operating environment includes two parts: client-side and server-side, client-side development for: technical architecture: Android development. Development environment: Android SDK, the minimum compatible SDK version 2.2, Eclipse integrated development environment. Development language: Java language. Server-side system development environment is: technical architecture: based on .net framework. Development language: C# language. SDK version: support for .net 4.0, asp.net mvc 4.

The overall technical architecture of the IoT building energy consumption intelligent analysis platform is shown in Figure 2. On the basis of dividing the platform logic into four layers, the whole architecture is refined and decomposed one by one as follows:

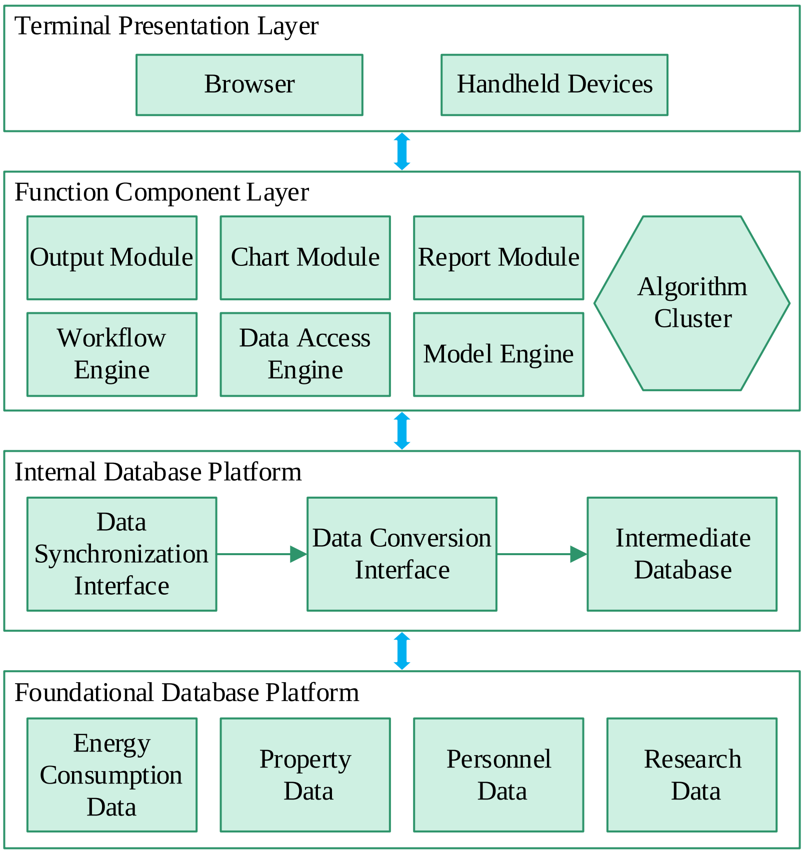


Figure 2: Overall technical architecture

3.1.1 Basic database layer

This is the data object and research object of the data mining platform, in which the energy consumption data are mainly the electric meter, water meter, gas meter, heat meter readings of each node, as well as the related categorized sub-items of energy consumption information. The energy consumption data is characterized by a lot of real-time data, a large amount of data, and fast data update. Real estate data mainly contains some basic information of the building, such as the area of the room, the orientation of the room and the use of the room. Personnel data are the people inside the building in question and their various information. Research data is the research output of each of these people.

3.1.2 Internal database layer

This layer is where the information data is directly operated by the data mining platform, which uses the data synchronization interface to synchronize all the data in the basic database layer, and then uses the data converter to format these data separately and deposit them into the intermediate database.

3.1.3 Functional component layer

This layer mainly realizes various functions and businesses of the platform and manages the database, as well as establishes interface services for the display layer to call. Among them, the workflow engine is responsible for realizing the business process of the platform, mainly for the two parts of data collection and data mining. Data access engine realizes data reading and writing operations and database management tasks, and this part is the interface between the functional component layer and the database platform communication. The model engine is to persist the data taken out and convert it into instance objects to prepare for the subsequent data mining work. The model engine divides the data into memory cache and persistent storage, and selects the appropriate way to store according to the specific data type. The memory cache should not have too much data, and recycle the memory space in time when the data is invalidated, so as not to affect the performance of the cell phone client. The chart report module can display the results from various angles, such as curve graphs, area graphs, scatter plots and so on.

3.1.4 Terminal display layer

This layer is the interface for interaction with end users, and utilizes the service interface provided by the functional component layer to realize the external display of platform functions. The interface requires simple and generous, compatible with current mainstream browsers, easy for users to operate, while outputting clear chart results.

3.2 Functional Module Design

3.2.1 Data management

Manage and maintain the basic energy consumption data and all kinds of external data, organize these data effectively, and carry out certain optimization, so that the platform can quickly provide the corresponding data in the process of data mining. At the same time, the mining results are saved to facilitate multiple calls by users. The business engine takes out the relevant data in the intermediate database and organizes the data through the preprocessing module.

3.2.2 Algorithm management

Maintaining all kinds of data mining algorithms and forming standard input and output interfaces, the platform can quickly map the requests submitted by users to the corresponding algorithms. In addition, the platform supports the expansion of algorithms, and can add and delete various types of algorithms at any time.

3.2.3 Data mining

The data mining algorithm can be a pre-compiled algorithm in the platform, and the embodiment of the present invention integrates the IBSO-DBSCAN clustering algorithm in the platform. In addition, the platform also supports users to upload their own algorithm plug-ins, and user algorithms can be imported into the platform according to certain requirements.

3.2.4 Access management

It is mainly the maintenance of all kinds of external interfaces, which can support other applications to call local algorithms and improve the utilization rate of the platform. The platform display layer mainly adopts a graphical way to provide users with operation interfaces and data analysis results, and mainly focuses on Web page display and cell phone display.

4 Experimental results and analysis

4.1 Sources of data on building energy consumption

The research object of this paper is a large office building in a hot summer and warm winter area, the building is divided into four parts: the main building, conference, east auxiliary building and west auxiliary building, and the central air conditioning system is used for centralized cooling/heating, with a total of one set of cold source system and one set of cooling/heating source system. The energy consumption data (including air conditioning power consumption and lighting power consumption, etc.) of this office building from April 1, 2025 to August 1, 2025 were collected, and the 3σ criterion was used to identify the distorted data of energy consumption.

The 3σ criterion detection results show that a total of 174 distorted data are included, with a distortion rate of 3%. For cases where the continuous missing data for a natural day is greater than 33%, the whole day-by-hour data for that natural day is directly deleted. For cases where the consecutive missing data for a natural day is less than 33%, the mean value method is used to fill in. For cumulative and anomalous mutation values, the mutation data were deleted and filled in using the mean value method.

4.2 Clustering performance evaluation

4.2.1 Comparative analysis of algorithms

Clustering algorithm is a common method for mining daily energy use patterns. In order to test the effectiveness of the proposed clustering method, this paper selects K-means algorithm, traditional K-Medoids method and traditional DBSCAN algorithm to cluster the building energy consumption data, and evaluates the superiority of the clustering effect of the algorithms under the different number of clustering clusters by the evaluation indexes of clustering internal. In this paper, the contour coefficient is selected to evaluate the clustering effect of the algorithm, and the value range of the contour coefficient is [-1,1], the larger the contour coefficient, the better the clustering effect. The comparison of contour coefficients of different algorithms is shown in Figure 3. Compared with other clustering algorithms, the clustering effect of IBSO-DBSCAN is improved by 29.61%~44.62%, which proves that the method can effectively cluster based on the similarity of the time series, and it is suitable for building energy use pattern mining.

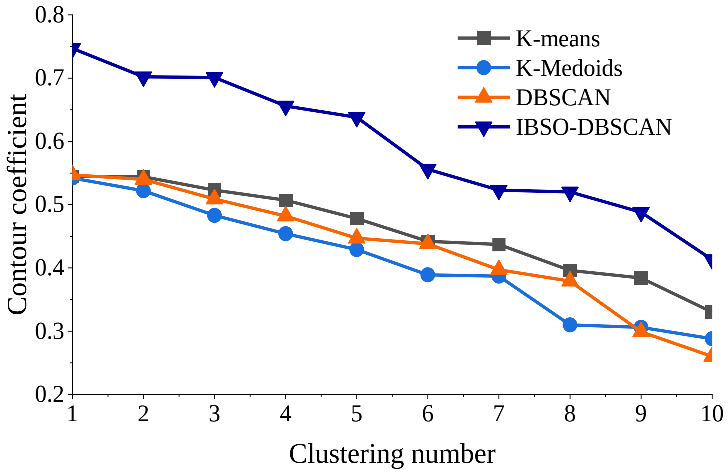


Figure 3: Comparison of contour coefficients of different algorithms

4.2.2 Optimization effect of IBSO

In order to verify the optimization effect of the DBSCAN algorithm for IBSO optimization, the Iris dataset was used to test it and determine its clustering accuracy based on the confusion matrix. The DBSCAN algorithm before and after optimization was used to cluster the Iris dataset, which is a multivariate analysis dataset that is widely used in the field of machine learning, such as cluster analysis and multivariate statistical analysis. The dataset consists of 150 five-dimensional attribute samples forming a 150 × 5 matrix storing three different categories of Iris samples, each with 50 samples each.

The confusion matrix of the clustering results is shown in Table 1. The optimized DBSCAN algorithm has an accuracy of 97.33%, precision of 100%, 97.92%, 94.23%, and recall of 100%, 94%, and 98%, respectively. The accuracy of the unoptimized DBSCAN algorithm is 89.33%, precision rate is 100%, 80.36%, 88.89%, and recall rate is 100%, 90%, and 80%, respectively. It can be seen that the overall accuracy of the optimized DBSCAN clustering algorithm is significantly higher than the unoptimized DBSCAN clustering algorithm.

Table 1: Confusion matrix of clustering results

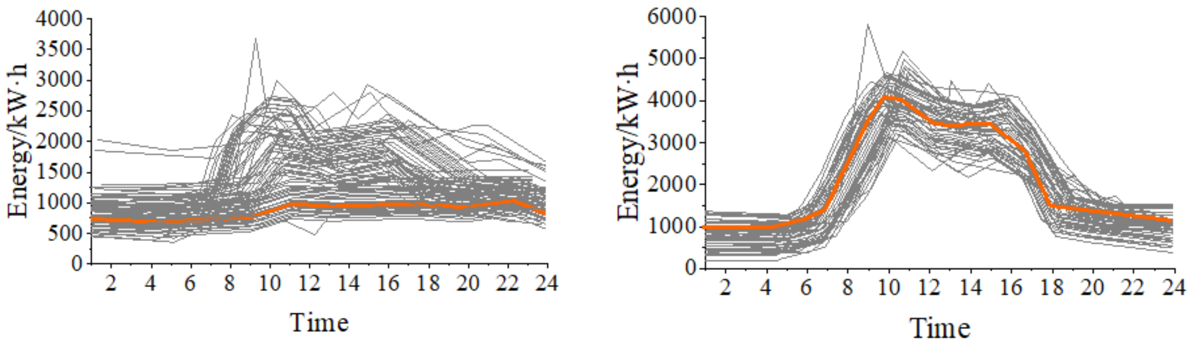
Cluster results of IBSO-DBSCAN					
		Cluster result			Total
		Class 1	Class 2	Class 3	
Real label	Class 1	50	0	0	50
	Class 2	0	47	3	50
	Class 3	0	1	49	50
Total		50	48	52	150
Cluster results of DBSCAN					
		Cluster result			Total
		Class 1	Class 2	Class 3	
Real label	Class 1	49	1	0	50
	Class 2	0	45	5	50
	Class 3	0	10	40	50
Total		49	56	45	150

4.3 Clustering of building energy use patterns

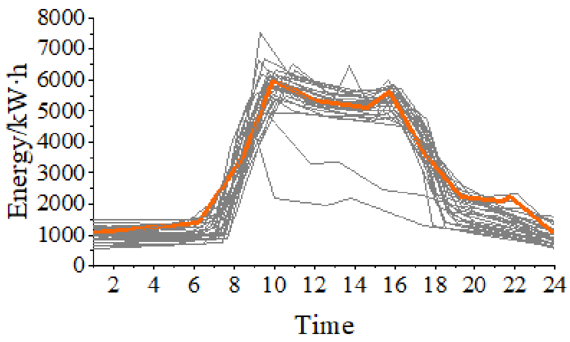
Based on the collected building energy consumption data, this paper selects the IBSO-DBSCAN clustering algorithm to mine building energy use patterns, and the clustering results of building daily energy use patterns are shown in Fig. 4. Each gray line in the figure represents the daily energy use pattern of a building on a certain day, and the orange line indicates the cluster center of each category.

The energy consumption of category 1 clustering center is almost at a level all day, rising at 10:00 a.m. and slightly decreasing at 12:00 p.m., rising at 20:00 p.m. and decreasing at 22:00 p.m. According to the results of the research, there is some overtime work of some people in the building at night time, and at the same time the building will turn on the lighting project. In category 1, there is a part of the energy use pattern that is different from the shape of the center of the cluster, which is similar to the “bimodal” pattern in the center of the cluster of category 2 and 3, but it is classified as category 1 because the intensity of energy use is significantly lower than that of category 2 and 3.

The energy use curves in categories 2 and 3 both show a “double peak” shape of energy consumption: energy consumption starts to rise sharply from 7 a.m. to 9 a.m. and reaches a peak at 10 a.m. The energy consumption of the building then starts to fall. Then the energy consumption of the building starts to decline, and a turning point occurs at 13:00-14:00, and a second peak occurs at 15:00 in the afternoon. However, the two peaks of energy consumption in category 3 reach 5937kW-h and 5718kW-h respectively, which are significantly larger than the peaks of 4172kW-h and 3683kW-h in category 2. The three clustering results represent the three typical daily energy use patterns of the building respectively. The similarity between different energy-use patterns is low, the energy-use intensity varies greatly, and the energy consumption characteristics are distinctly different.



(a)Class 1 (b)Class 2



(c)Class 3

Figure 4: Overall technical architecture

The three daily energy use patterns are visualized based on date attributes as shown in Figure 5. 38.52% of the building energy use from April to July 2025 is accounted for by Mode 1, which mainly occurs on non-working days as well as on working days in the transition season. Mode 2 has a share of 38.52% and mainly occurs during weekdays from April to June. Mode 3 accounted for a minimum of 22.95% and occurred mainly during weekdays in July. The visualization results further illustrate the correlation between total building energy consumption and parameters such as weekday attributes, seasonal attributes, and average outdoor temperature.

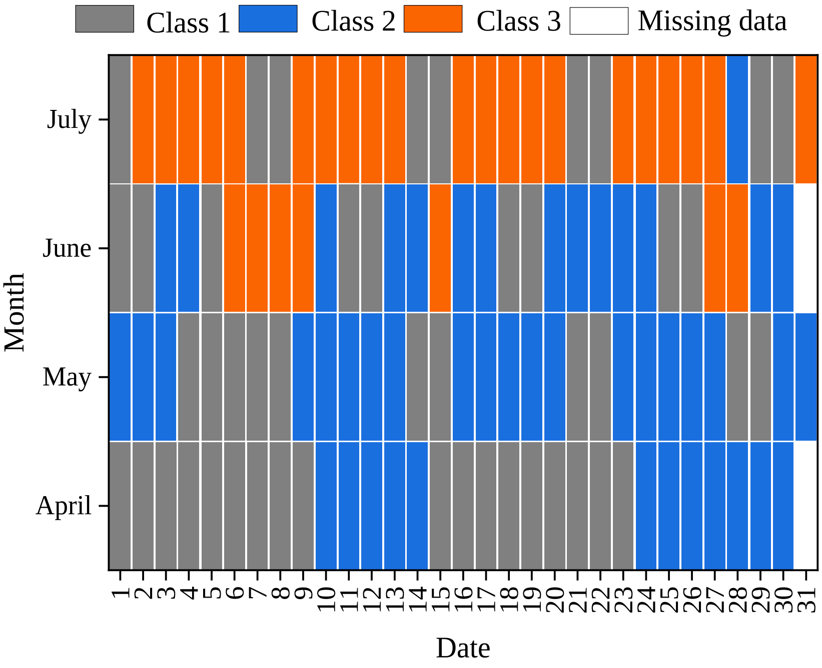


Figure 5: The distribution of daily energy patterns

4.4 Cluster stability analysis

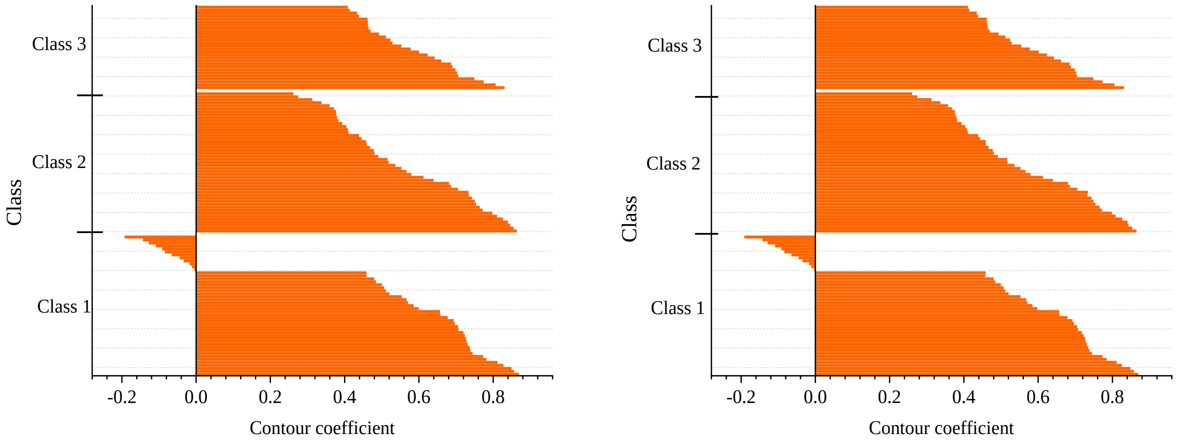
From the above analysis, it can be seen that it is feasible to utilize the IBSO-DBSCAN algorithm for building energy consumption data clustering analysis. Since the stability of the clustering algorithm has a direct impact on the analysis of building energy consumption data, this section utilizes the IBSO-DBSCAN algorithm for several analyses to explore the stability of the clustering results of building energy consumption data.

The collected data was clustered 15 times using the IBSO-DBSCAN algorithm. Table 2 shows the statistical results of its 15 clustering, which shows that a total of three different clustering results occurred. In total, two results 47 and 48 appeared for category 1 data, two results 46 and 47 appeared for category 2 data, and two results 28 and 29 appeared for category 3 data. At the same time, the statistics of the 15 clustering results also found that the error between the average values of the various types of data is less than 4%, indicating that the clustering effect of the DBSCAN algorithm optimized by IBSO on the electricity consumption data has a strong stability.

Table 2: Statistical results of the 15 times clustering

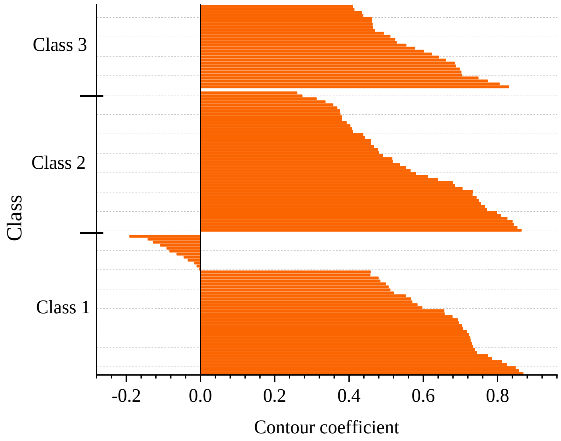
Number	Class 1 number	Mean(kW·h)	Class 2 number	Mean(kW·h)	Class 3 number	Mean(kW·h)
1	48	908.28	46	1693.21	28	2911.81
2	47	908.28	47	1693.21	28	2911.81
3	47	908.28	47	1693.21	28	2911.81
4	47	908.28	47	1693.21	28	2911.81
5	47	908.28	46	1693.21	29	2911.81
6	47	908.28	47	1693.21	28	2911.81
7	48	908.28	46	1693.21	28	2911.81
8	47	908.28	47	1693.21	28	2911.81
9	47	908.28	47	1693.21	28	2911.81
10	47	908.28	47	1693.21	28	2911.81
11	48	908.28	46	1693.21	28	2911.81
12	49	908.28	45	1693.21	28	2911.81
13	47	908.28	47	1693.21	28	2911.81
14	47	908.28	46	1693.21	29	2911.81
15	47	908.28	47	1693.21	28	2911.81

Figure 6 gives an indicator of the profile coefficient for the three clustering results, with values ranging from (-1, 1). The larger the value indicates that the data matches the cluster to which it belongs, and the presence of many lower or negative values indicates that this clustering is not effective. It can be seen that all three clustering results are relatively similar. Negative values are concentrated in the intersection range of category 1 and category 2 data, where the electricity consumption data fluctuates greatly, which can easily lead to misjudgment of the clustering algorithm. Except for a small number of negative values appearing in category 1 data, most of the remaining data have high profile coefficients and can reach 85% accuracy. This proves that the overall clustering results of the DBSCAN algorithm optimized with IBSO are more accurate for electricity consumption data, with the most accurate clustering results for category 3 data. The results of 15 clustering of the data show that the clustering algorithm has strong stability and accuracy.



(a)The contour coefficient of the clustering result 1

(b)The contour coefficient of the clustering result 2



(c)The contour coefficient of the clustering result 3

Figure 6: The contour coefficient of the three kinds of clustering results

5 Conclusion

The study uses the improved BSO algorithm to optimize the density clustering algorithm and proposes a building energy consumption data analysis method based on IBSO-DBSCAN, based on which an intelligent analysis platform for IoT building energy consumption is developed. The collected building energy consumption data are clustered and analyzed to mine the building energy use patterns and evaluate the performance of the IBSO-DBSCAN algorithm. The main research results are as follows:

(1) The IoT building energy consumption intelligent analysis platform developed integrates the IBSO-DBSCAN algorithm with functions of data management, algorithm management, data mining and access management, which can intelligently analyze the building energy consumption data, discover the useful information in the building energy consumption data, and improve the overall level of energy management.

(2) Compared with K-means, K-Medoids and DBSCAN algorithms, the IBSO-DBSCAN algorithm improves the clustering effect by 29.61% to 44.62%. The analysis in the Iris dataset found that the accuracy of the IBSO-DBSCAN algorithm was 97.33%, which was significantly higher than that of the DBSCAN algorithm, which was 89.33%. The clustering effect of IBSO-DBSCAN algorithm and the optimization effect of IBSO are verified.

(3) The building energy consumption pattern is divided into three categories, category 1 has relatively small fluctuations in energy consumption, category 2 and category 3 both show a “double peak” shape of energy consumption, in which the energy consumption of category 3 is significantly higher than that of category 2. The distribution of energy consumption patterns of category 1 and category 2 is comparable during the study period, and category 3 accounts for the smallest proportion, which is 22.95%. The data error of the IBSO-DBSCAN algorithm is less than 4%, which indicates that it has good stability and accuracy in clustering the building energy consumption data.

The introduction of data mining technology in the field of building energy consumption data research has a very broad prospect, and this paper applies some data mining techniques to the study of building energy consumption patterns and obtains certain conclusions. Next, further in-depth research on building energy consumption prediction, energy consumption anomaly detection and other issues can be carried out based on the results of clustering analysis.

About the Author

Jian Pi has successfully obtained a Master's degree in Business Administration. At present, he works as a Political Affairs Expert at Hebei Construction Investment Xiong’an Construction and Development Co., Ltd. His research focuses on Business Administration, particularly strategic management, organizational behavior, and human resource optimization in corporate environments.

Ze Xu, obtained his Bachelor’s degree in Mechanical Design, Manufacturing, and Automation. As a Senior Engineer at Hebei Construction Investment Xiong’an Construction and Development Co., Ltd., his work emphasizes Mechanical Design and Automation, including intelligent manufacturing systems, automation control technologies, and precision mechanical engineering.

Ya’ou Lv earned his Master’s degree specializing in Auditing. He currently serves as an Auditor at Hebei Construction Investment Xiong’an Construction and Development Co., Ltd. His research focuses on Auditing, particularly in areas such as internal audit processes, corporate risk management, compliance auditing frameworks, and advanced techniques in financial data analysis.

Zihao Zhao, holds a Master’s degree in Computer Information Technology. He works as a Lecturer at Hebei Construction Investment Xiong’an Construction and Development Co., Ltd. His research specializes in Computer Information Technology, covering areas such as artificial intelligence applications, big data analytics, and software development methodologies.

Yunfang Wang obtained a Master's degree in the field of Communication Engineering. He is currently a Senior Engineer at Hebei Construction Investment Operation and Management Co., Ltd. His expertise lies in Communication Engineering, with a focus on communication system design, network optimization, and emerging applications of 5G technology.

References

[1] Li, X., Zhou, Y., Yu, S., Jia, G., Li, H., & Li, W. (2019). Urban heat island impacts on building energy consumption: A review of approaches and findings. Energy, 174, 407-419.

[2] Ma, H., Du, N., Yu, S., Lu, W., Zhang, Z., Deng, N., & Li, C. (2017). Analysis of typical public building energy consumption in northern China. Energy and Buildings, 136, 139-150.

[3] Yang, X., Zhang, S., & Xu, W. (2019). Impact of zero energy buildings on medium-to-long term building energy consumption in China. Energy Policy, 129, 574-586.

[4] Li, C., Lu, P., Zhu, W., Zhu, H., & Zhang, X. (2023). Intelligent monitoring platform and application for building energy using information based on digital twin. Energies, 16(19), 6839.

[5] Zhao, T., Xu, J., Zhang, C., & Wang, P. (2021). A monitoring data based bottom-up modeling method and its application for energy consumption prediction of campus building. Journal of Building Engineering, 35, 101962.

[6] Almeida, F., Assunção, M. D., Barbosa, J., Blanco, V., Brandic, I., Da Costa, G., ... & Pierson, J. M. (2018). Energy monitoring as an essential building block towards sustainable ultrascale systems. Sustainable Computing: Informatics and Systems, 17, 27-42.

[7] Ferreira, J. C., Afonso, J. A., Monteiro, V., & Afonso, J. L. (2018). An energy management platform for public buildings. Electronics, 7(11), 294.

[8] Burgess, M. (2018). What is the Internet of Things? WIRED explains. Wired UK, 1357-0978.

[9] Rose, K., Eldridge, S., & Chapin, L. (2015). The internet of things: An overview. The internet society (ISOC), 80(15), 1-53.

[10] Silverio-Fernández, M., Renukappa, S., & Suresh, S. (2018). What is a smart device?-a conceptualisation within the paradigm of the internet of things. Visualization in Engineering, 6(1), 1-10.

[11] Ryan, P. J., & Watson, R. B. (2017). Research challenges for the internet of things: what role can or play?. Systems, 5(1), 24.

[12] Bhattacharjee, P., & Mitra, P. (2021). A survey of density based clustering algorithms. Frontiers of Computer Science, 15(1), 151308.

[13] Campello, R. J., Kröger, P., Sander, J., & Zimek, A. (2020). Density-based clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(2), e1343.

[14] Martín-Garín, A., Millán-García, J. A., Baïri, A., Millán-Medel, J., & Sala-Lizarraga, J. M. (2018). Environmental monitoring system based on an Open Source Platform and the Internet of Things for a building energy retrofit. Automation in Construction, 87, 201-214.

[15] Dutta, P. K., El-kenawy, S. M., Ali, G., & Dhoska, K. (2023). An Energy consumption monitoring and control system in buildings using internet of things. Babylonian Journal of Internet of Things, 2023, 38-47.

[16] Zhang, T., Zhou, M., Guo, X., Qi, L., & Abusorrah, A. (2022). A density-center-based automatic clustering algorithm for IoT data analysis. IEEE Internet of Things Journal, 9(24), 24682-24694.

[17] Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, 1192-1205.