International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 4267-4287

https://doi.org/10.70517/ijhsa464276

Smart Grid Oriented Power User Behavior Analysis with Intelligent Clustering Methods Based on Knowledge Graph and Hybrid Models

Fei Lu[1,*]
[1] State Grid Dandong Power Supply Company, Dandong, Liaoning, 118000, China
Corresponding authors: (e-mail: lufeiLF123@126.com).

Abstract With the continuous expansion of the power grid scale, how to analyze the behavior of power users and intelligent clustering methods has become an urgent problem to be solved by power grid companies. In order to solve the problems described above, the traditional DTW algorithm is first optimized and improved with the help of similarity algorithm, so that it meets the requirements of power user behavior analysis. After that, the knowledge graph is used to preprocess the power user behavior, store it in the form of dataset, and realize the intelligent clustering of power user behavior through the clustering analysis of Gaussian mixture model. Build the experimental environment, set the comparison algorithm and evaluation indexes, and use the data analysis software to verify and analyze the intelligent clustering scheme of power user behavior based on KGEG algorithm. In the data set A~J, the data of the three indicators of this paper's algorithm is much better than the other three comparison algorithms, and the distribution of the data of the three indicators is in the range of 0.6~0.9, which confirms the application effectiveness of this paper's user behavioral intelligent subgrouping research program, so as to improve the level of development of the smart grid.

Index Terms DTW algorithm, similarity algorithm, knowledge graph, Gaussian mixture model

I. Introduction

With the progress of science and technology and the improvement of people's living standards, the demand for electricity is growing, and in order to cope with the increasing demand for electricity, smart grid comes into being [1], [2]. Smart grid is a kind of power system that uses advanced technology and information and communication technology to optimize the power supply and power consumption behavior, and in smart grid, power consumption behavior analysis is the foundation and key to realize smart grid [3]-[6].

Electricity consumption behavior analysis refers to the collection, analysis and mining of electricity consumption data to understand users' electricity consumption patterns, characteristics and needs [7], [8]. Electricity consumption behavior analysis can help electric power companies and energy suppliers, to better understand the users' electricity consumption pattern and demand, so as to reasonably allocate and configure electricity resources, through the analysis of the users' historical electricity consumption data, it can find out the electricity consumption pattern and characteristics of different users [9]-[12]. For example, some users may use more electricity at night and less during the day, some users may often use TV, air conditioner and washing machine at the same time, while others may only use electric devices at night. These electricity usage patterns and characteristics are very important for electricity suppliers to help them plan electricity resources rationally and make adjustments and responses in advance [13]-[15]. Electricity consumption behavior analysis can also help users understand their own electricity consumption habits and consumption behaviors, so as to improve electricity consumption efficiency and save electricity, and through the analysis of their own electricity consumption data, users can find their own peak, trough, and peak-to-valley differences [16]-[19]. For example, avoiding the simultaneous use of high-power electrical appliances during the peak period of electricity consumption can avoid overloading and excessive electrical load and improve the efficiency of electricity consumption. In the low period of electricity consumption, one can choose to use electric devices for charging or use the period of lower electricity price for laundry and cooking, etc., so as to save the cost of electricity [20]-[23]. In addition, the analysis of electricity consumption behavior can also help users find potential energy waste and hidden dangers of electricity consumption, and make timely improvements and optimization [24], [25].

Aiming at the limitations of traditional dynamic time planning algorithm on top of user behavior analysis, this paper proposes to adopt similarity algorithm to optimize and improve the traditional dynamic time planning algorithm, in order to achieve the purpose of accurate analysis of user behavior. Through the theoretical analysis of knowledge graph, expectation maximization, and Gaussian mixture model, the intelligent clustering research scheme of user behavior based on GKEG algorithm is designed. With the help of TensorFlow software to build the experimental environment, to determine the data set of this research, Purity, ACC, NMI as the evaluation index of the effect of intelligent clustering, in order to highlight the priority of this paper's intelligent clustering scheme, but also set up three control algorithms, respectively, k-means, GMM, DAE-k. Finally, according to the results of the algorithmic comparison and analysis, to verify the research of this paper's effectiveness and feasibility.

II. Research on Electricity User Behavior under Smart Grid Framework

Algorithm + model applications mainly realize in-depth insights into power grid business and operational development laws, data value-added applications use data as a new production factor to incubate new business forms, and real-time intelligence applications combine big data and artificial intelligence technologies to provide companies with real-time intelligent analysis capabilities. Specifically, in the direction of power user behavior analysis under power big data analytics, the business tends to focus on the change of power consumption over time, and can measure the similarity between users based on this change, classify users into different categories, and also analyze the causes of this change to improve the service system. In this chapter, the user similarity algorithm with improved DTW is mainly used to study the behavior of electricity users in the framework of smart grid. The overall structure of this chapter is as follows:

II. A. Dynamic time planning algorithm and similarity algorithm

II. A. 1) Dynamic time planning algorithm (DTW)

The DTW algorithm is a method to compute the similarity between two sequences by finding the optimal mapping dynamically [26]. If there are two different one-dimensional time series for a single variable X and Y, $X = x(i)$, $i = 1, 2, \cdots, m$. $Y = y(j)$, $j = 1, 2, \cdots, n$. where m, n are the lengths of the sequences X and Y, respectively. The distance between the points $x(i)$, $y(j)$ is defined as the matrix element (i, j), and the corresponding matrix D is constructed with the size of m rows and n columns, and the local distances d of matrix elements are denoted as:

$$d(i, j) = (x(i) - y(j))^2 \tag{1}$$

In the planning process of the DTW algorithm, the same time points of two sequences do not necessarily correspond to each other two by two, but rather, the points in the two sequences with similar states are found to correspond to each other through dynamic planning. The mapping of the two time series is represented by an integer path W. Namely:

$$W = \begin{pmatrix} w_x(k) \\ w_y(k) \end{pmatrix}, k = 1, 2, \cdots, p \tag{2}$$

where $w_x(k)$ denotes the $w_x(k)$th element in the sequence X, $w_y(k)$ represents the $w_y(k)$th element in the sequence Y, p denotes the length of W, and $\begin{pmatrix} w_x(k) \\ w_y(k) \end{pmatrix}$ denotes two elements mapped to each other. Regularized paths must satisfy boundary conditions, continuity and monotonicity constraints, and $DTW(x, y)$ is used to denote the shortest distance between tensors X and Y:

$$r(i, j) = d(i, j) + \varepsilon_{ij} \tag{3}$$

$$DTW(x, y) = \min\{r(m, n)\} \tag{4}$$

where $r(i, j)$ denotes the sum of local distances from $(0,0)$ to $r(i, j)$ paths in the distance matrix D. According to the constraints, the point with the smallest sum of distances from $(i-1, j-1)$, $(i-1, j)$, or $(i, j-1)$ is chosen as the starting point and designated as ε_{ij}. After finding the optimal planning path W, the original time series X and Y can be redefined by $\bar{x}(k)$ and $\bar{y}(k)$:

$$\begin{cases} \bar{x}(k) = x\left(w_x(k)\right) \\ \bar{y}(k) = y\left(w_y(k)\right) \end{cases}, k = 1, 2, \cdots, p \tag{5}$$

Thus the DTW distance between two unitary tensors can be expressed as the Euclidean distance between the new sequences $\bar{x}(k)$ and $\bar{y}(k)$:

$$DTW(x, y) = \sum_{k=1}^{P} (\bar{x}(k) - \bar{y}(k))^2 \tag{6}$$

II. A. 2) Similarity Algorithm

Similarity is to compare the similarity of two things, generally by calculating the distance between the features of things, if the distance between two objects is small, then their similarity is large, and vice versa [27]. Currently, the common distance or similarity measures mainly include Pearson's correlation coefficient, Euclidean distance, cosine distance, etc. Here we take two vectors X and Y as an example of similarity, each of which has n-dimensional features, and try to differentiate between the various measures of distance.

The Euclidean distance is formulated as:

$$d_{x,y} = \sqrt{\sum_{i=1}^{n} \left(X_i - Y_i\right)^2} \tag{7}$$

The cosine similarity is calculated as:

$$\cos_{x,y} = \frac{\sum_{i=1}^{n} X_i * Y_i}{\sqrt{\sum_{i=1}^{n} \left(X_i\right)^2} \cdot \sqrt{\sum_{i=1}^{n} \left(Y_i\right)^2}} \tag{8}$$

The Pearson's correlation coefficient is calculated as:

$$\rho_{x,y} = \frac{E\left(\left(x - \mu_x\right)\left(Y - \mu_y\right)\right)}{\sigma_x \sigma_y} \tag{9}$$

The Pearson correlation coefficient can remain translation invariant, i.e., if we move X to $a + bX$, where a and b are constant and do not change the correlation coefficients of the two variables, this solves the problem of cosine similarity being affected by the translation of the vectors.

II. B.Research on user similarity analysis method based on improved DTW

User similarity is the basis for realizing power user-side writing, and appropriate similarity calculation can accurately measure user behavior and provide a basis for further clustering and prediction. The user similarity analysis scenario for smart grid is designed to measure the similarity of power users and classify them accordingly, which can provide strong support for the following specific services such as peak tariff design, network-wide load balancing, and enterprise customized services. In this project, a user similarity analysis method (IDTW) based on improved DTW is proposed, which introduces range constraints and translation constraints to measure the similarity between power users, and has certain novelty. First, for the defects of high time complexity and pathological matching in DTW, this method introduces search range constraints, which reduces the search complexity from $O(N^2)$ to the average O(N) level while restricting unreasonable point pair matching. Secondly, the curve finite translation is introduced to improve the accuracy of sequence matching for non-stretch alignment, and improve the metric results for electricity users with slightly shifted time axis. Finally, the accuracy of similarity calculation is verified by a clustering algorithm that divides electricity users based on the calculated similarity.

II. B. 1) Description of the problem

Most of the existing similarity calculations are realized based on Euclidean and cosine distances, which are not applicable to power load data, and the subject needs to find a suitable comparison algorithm for time series data. Dynamic Time Warping (DTW) algorithm is a common method to analyze time series data, but its disadvantages are quite obvious. Figure 1 shows the schematic diagram of the dynamic time regularization algorithm search, and Figure 2 shows the schematic diagram of scaling matching. If DTW calculates the matching relationship between two time series, it will traverse the whole M*N matrix to find the optimal path, and the time complexity is too high, at the same time, the DTW algorithm has the phenomenon of pathological matching. The curve in Fig. 2 will be matched by DTW through pathological scaling, which is not compatible with the power data where frequency and period information are crucial. For this reason, this topic needs to propose improvements on the basis of the original dynamic time regularization algorithm to avoid the unnecessary pathological matching problem, and strive to reduce the computational complexity, while adapting the algorithm to the temporal nature of the power load data as well as the characteristics of the cycle frequency limitation.

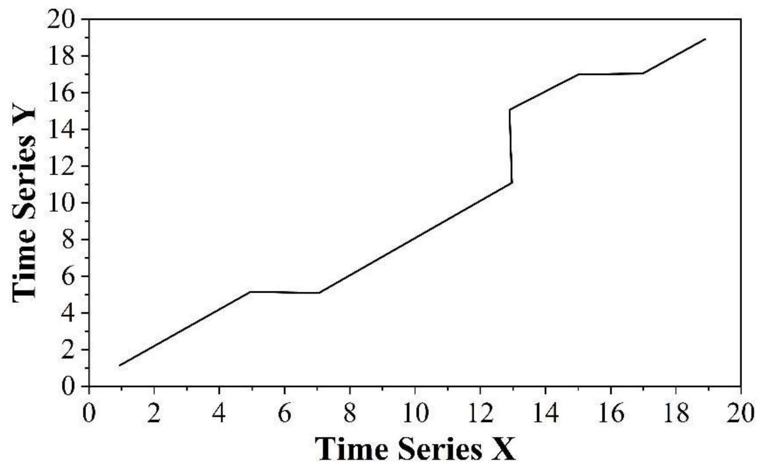


Figure 1: Dynamic time warping algorithm search diagram

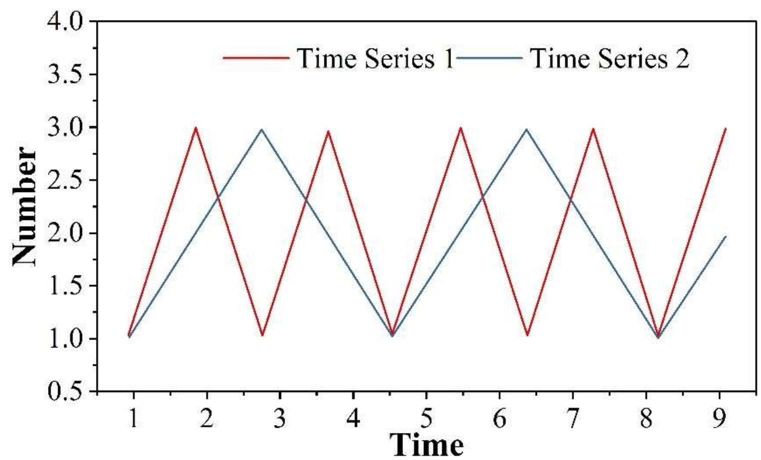


Figure 2: Scaling matching diagram

II. B. 2) Additional constraints

The variables that appear in the algorithm are defined below:

Time series X: $X = x_1, x_2, \cdots, x_m$.

Time series Y: $Y = y_1, y_2, \cdots, y_n$.

Distance matrix D: $D[m,n]$, where any element $D[i,j]$ denotes the distance between time point x_i and y_j.

Matching path P: $P = p_1, p_2, \cdots, p_k$, where $p_i = (x_a, y_b)$ and $\max(m,n) \le k < m+n-1$.

Cumulative distance R: $R(i,j)$ denotes the minimum value of the cumulative distance from the starting point to the point $D[i,j]$ in the distance matrix.

With the definition of variables as above, and based on the original idea of dynamic time regularization, the search range restriction, and the introduction of finite curve translation, we can reduce the algorithm idea to the following five constraints:

Constraint 1: Boundarity constraint. Arbitrary time sequences may be translated or stretched, but their sequential order will not change, and the selected matching path must be from the start to the end of the sequence pair, i.e., from the lower-left corner of the distance matrix to the upper-right corner of the extension. I.e:

$$P_1 = (x_1, y_1) \ And \ P_k = (x_m, y_n) \tag{10}$$

Constraint 2: Continuity constraint. Each point cannot be compared across a point, but can only be aligned with its own neighboring points, ensuring that each point in the sequence X and Y appears in the matching path P. I.e:

$$P_{k-1} = (x'_a, y'_b) \ And \ P_k = (x_2, y_b), \ Have \ a - a' \le 1 \ And \ b - b' \le 1 \tag{11}$$

Constraint 3: Monotonicity constraint. The pairs of points on the matching path P must evolve monotonically over time, and the matching path itself is not allowed to cross or overlap. I.e:

$$P_{k-1} = (x'_a, y'_b) \ And \ P_k = (x_a, y_b), \ Have \ a - a' \ge 0 \ And \ b - b' \ge 0 \tag{12}$$

With the above three constraints, it can be determined that there can only be three directions for generating matching paths in the distance matrix. That is, if the point (i, j) is known, then the next selected matrix position must be $(i+1, j)$, $(i, j+1)$, $(i+1, j+1)$ one of them.

Constraint 4: Search range constraint. The search range of the matrix represents the degree of stretching of the curves; the more similar the curves match the paths the closer they are to the diagonal. This method restricts the search range of the path to the Search_Limit range near the diagonal of the matrix, i.e., the maximum stretching degree of the curve is 2*Search_Limit. That is:

$$P_i = (x_a, y_b), Have \left| x_a - y_b \right| \le Search\_Limit \tag{13}$$

Constraint V: Finite translation constraint. For time series X and Y with obvious translation similarity, limited translation of the curves is allowed, and the maximum length of translation is Trans_Limit. The distance matrix and matching paths are re-computed after translation, and for the start and end points of the curves that are moved out of the comparison range, their optimal matching relationship is maintained and counted in the initial value of the distance matrix $D[0,0]$. Let the initial curve be:

$$X = x_i, x_{i+1}, \ldots\ldots, x_j \ And \ Y = y_a, y_{a+1}, \ldots\ldots, y_b \tag{14}$$

Then the curve is transformed after translation:

$$X' = x_{i+1}, x_{i+2}, \ldots\ldots, x_j \ And \ Y' = y_a, y_{a+1}, \ldots\ldots, y_{b-1} \tag{15}$$

Calculate the initial value of the new distance matrix as:

$$D'[0,0] = \sum D[i, \ all \ ] + \sum D[all, b] \tag{16}$$

Based on the previous constraints, the computation of the cumulative distance R can be realized by dynamic programming:

$$R(i.j) = D[i.j] + \min\left(R(i-1.j) \cdot R(i.j-1) \cdot R(i-1, j-1)\right) \tag{17}$$

where (i, all), (all, b) are all pairs of points involved in x_i and y_b in the matching path P. Then the goal of the whole improved DTW algorithm is to find $\min R(m,n)$ while generating the waylens of this minimum distance which is the best matching path.

II. B. 3) Algorithm implementation

The whole IDTW algorithm flow is shown in Fig. 3. First initialize all parameters, calculate the distance between pairs of points in the time series, and generate the distance matrix. Secondly, carry out dynamic planning on the matrix, calculate the optimal matching path under the current search scope limitations: then according to the current matching results to determine whether the results can be improved by panning, if not, then the current results that is the optimal choice, the direct output. If not, the current result is the optimal choice and output directly. On the contrary, it enters into the step of translation optimization, determines the direction of translation, modifies the time series and calculates the initial value of the distance matrix, and carries out dynamic planning again. Repeat the above operations until the optimal match is found or the optimization cannot be continued, and the final output of the cumulative distance is the similarity between the two sequences.

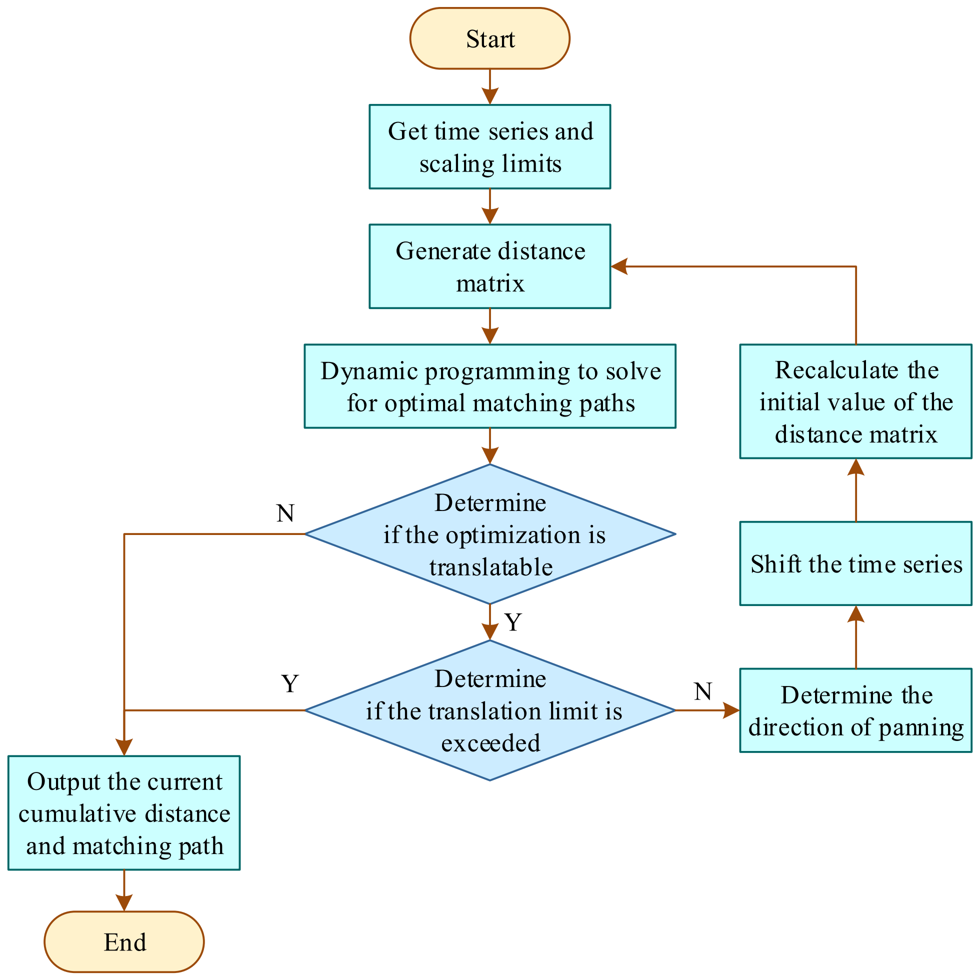


Figure 3: Improved dynamic time warping algorithm flow chart

II. B. 4) Clustering algorithm selection

In order to verify the accuracy of similarity analysis, the experiment chooses to analyze the accuracy of the metric by clustering based on the calculated similarity. There are various clustering algorithms, according to different clustering basis can be divided into: distance-based clustering methods, such as K-Means algorithm, K-Medoids algorithm; clustering algorithms based on hierarchy, such as the BIRCH algorithm, CURE algorithm; clustering methods based on the density of clustering methods, such as the DBSCAN algorithm, OPTICS algorithm and so on. In order to cluster the customer's electricity load, we have measured the similarity between electricity consumption curves by IDTW algorithm. On this basis, we have chosen K-Medoids algorithm with better interpretability for clustering the users.

III. Research on Intelligent Clustering Based on Knowledge Graph and Hybrid Modeling

On the basis of the power user behavior analysis in the previous paper, in order to further improve the user management capability for smart grid, this paper proposes a GMM (Gaussian Mixture Model) clustering method with joint KG (Knowledge Graph) and Expectation Maximization (EM), referred to as the KGEG method. The method mainly consists of 2 steps: firstly, using KG, the complex text is converted into the input required by the GMM method, and after that, EM method is used to perform GMM clustering on the data obtained in the previous step to realize the cluster analysis of the hidden correlations of the data. Finally, the proposed method is applied to the intelligent clustering of user behavior in electric power companies to obtain the user clusters of electric power companies, and the method is compared with other methods through evaluation indexes to verify the feasibility and effectiveness of the proposed method.

III. A. Knowledge mapping

III. A. 1) Knowledge map construction methods

This paper firstly introduces the elements and interconnections in the knowledge graph, then gives the representation of the knowledge graph, and finally proposes an ontology-based knowledge graph construction model. When constructing knowledge graph, it is necessary to establish corresponding models according to different objects. The relationship between entities can be expressed in the form of “entity-relationship-entity”, while entity-attribute-attribute value is used to express an attribute value of an entity. A complete knowledge structure system is formed by organizing these graph elements, and a knowledge mapping model is constructed based on it. As shown in Figure 4, these two levels together constitute the infrastructure of the knowledge graph. Knowledge mapping can organize all the knowledge into an organic whole, so as to achieve the role of knowledge sharing and reuse, improve work efficiency and quality, and reduce duplication of work. As a kind of database storing real data, the database structure of Knowledge Graph presents a pictorial form instead of the traditional phenotypic structure. Therefore, the knowledge graph can be represented as a network containing a large number of relationships, i.e., a tree-like network topology composed of many points. In order to achieve efficient organization and management of these types of data, it is necessary to utilize knowledge graph technology, which is schematically illustrated in Figure 4. Under this architecture, data mining algorithms are utilized to quickly and efficiently mine and discover potential knowledge that can have a significant impact on intelligent clustering of smart grid user behavior. On this basis, this knowledge is organized according to certain rules and the corresponding knowledge graph is drawn using machine learning algorithms and visualization techniques.

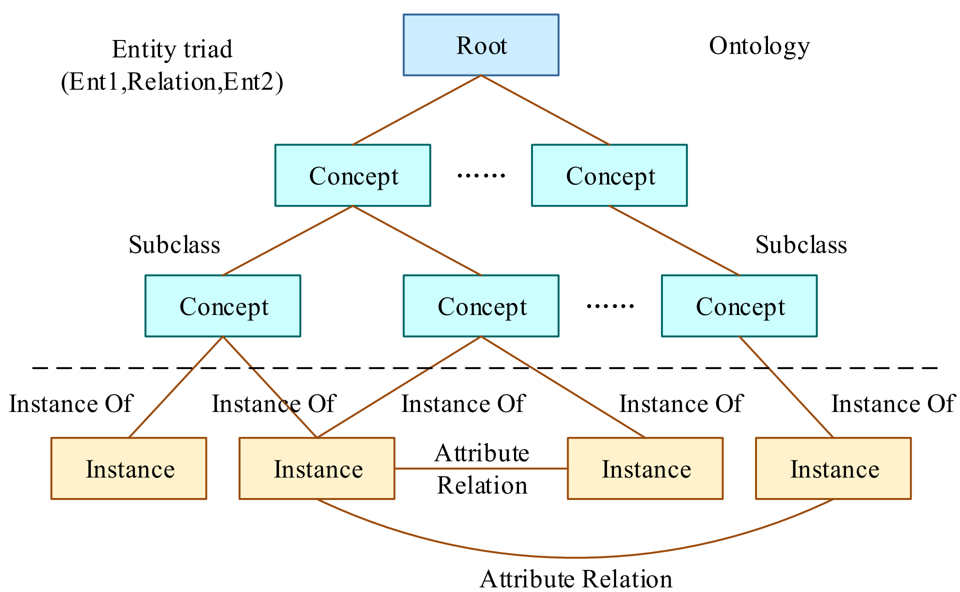


Figure 4: Logical architecture of the knowledge graph

III. A. 2) Knowledge modeling

Knowledge graphs contain rules for interconnecting concepts and a variety of information related to them, including the experience of experts in the field, researcher collaborations, and so on. Patterns are constructed by means of nodes and edges to further elucidate the limitations of the data. From both theoretical and practical perspectives, ontologies are an important information resource that can help people understand the laws and principles embedded in complex systems more effectively and gain useful knowledge from them. Modeling of ontology libraries is essentially modeling of knowledge. An ontology library is a formal representation based on rules and semantic similarity computation, which can effectively express the reasoning mechanism in human thinking. It mainly consists of ontology library construction methods based on a combination of top-down and bottom-up. These two types of ontology library construction methods are compared and analyzed a hybrid design method based on the combination of top-down and bottom-up to realize the mapping and sharing between various types of data in the ontology library. The bottom-up design approach is adopted, starting with entities, which are filtered and organized to form concepts of a certain size, and then abstracted. By dividing hierarchically, the utilization of data dimensions is improved, and the consistency between data on different levels is also ensured. Machine learning and statistical techniques are used to pre-analyze standard data sources from the perspective of the underlying entities to assist in the optimization of their organization.

III. A. 3) Knowledge extraction and knowledge representation

Knowledge graph is a complex system composed of different levels and domains, which covers important knowledge points in many disciplines from ontology to cognition and sociology, and also contains a large number of relevant application examples. Attribute names are meticulously analyzed and identified to obtain relevant information in an efficient way. Relational extraction, on the other hand, refers to obtaining information or results by calculating the similarity that exists between objects and other objects in the database without considering the semantic correlation between them.

In order to realize the rule- and template-based approach, domain experts are required to formulate relational rules and templates, and although these rules and templates are highly accurate, they require in-depth research and exploration due to their huge workload and difficulty in extending them to other domains. Aiming at this problem, a relationship extraction algorithm based on knowledge graph is proposed. With the continuous progress of machine learning technology, a supervised and weakly supervised relation extraction method is introduced to improve its efficiency and accuracy. In this paper, we propose a new feature-based relationship labeling algorithm. Similar to entity recognition, the feature-based relation extraction method first extracts features such as discriminative words, entity attributes, dependent syntax and syntactic trees, and then trains and constructs classification models for these features, and finally uses these trained models to perform relation extraction and dependent syntactic analysis of the text to obtain the final ternary relations.

The representation of knowledge is a computer symbolic simulation of a reasoning process that applies to both machines and humans to simulate knowledge in the human brain. The graph model is an efficient model that not only expresses relations but is also understood by computers and can be used to design models that simulate the reasoning process of knowledge in the human brain. The model uses a symbolic and vector-based representation that can be used for knowledge modeling.

III. A. 4) Knowledge integration and storage

Knowledge fusion needs to be supported by a complete and effective framework system, which involves key technologies in three aspects: knowledge fusion model, knowledge representation and processing mechanism, and knowledge base system. Between the data model layer and the data layer of the knowledge graph, the process of knowledge fusion presents a dual nature. Knowledge fusion is realized by transforming data into the data layer of the knowledge graph using ontologies. Knowledge mining, on the other hand, mines potential knowledge from this data. Knowledge is then a way to enhance the data itself. Thus knowledge storage is closely related to data storage techniques, that is, the interconnection of information through different forms. The efficiency of knowledge query is directly affected by knowledge storage, so when designing the storage requirements, it is necessary to consider the characteristics of the knowledge mapping graph model and combine it with the actual application scenarios, including the use of graph databases for efficient storage of knowledge. With the development of computer technology and the progress of Internet technology, graph database has gradually become a very popular research field.

III. B. GMM and EM algorithms

III. B. 1) GMM

The basic idea of GMM is to use a set of weighted sums of Gaussian-distributed PDFs to approximate the description of an arbitrary PDF, i. e:

$$f(x) \approx \sum_{k=1}^{K} w_k N\left(x; \mu_k, R_k\right) \tag{18}$$

where: x is the observation vector; K is the total number of Gaussian components, w_k is the weight coefficient of the k th Gaussian component; $N\left(x; \mu_k, R_k\right)$ is the PDF of the k th Gaussian component, and:

$$N\left(x; \mu_k, R_k\right) = \frac{1}{\sqrt{\det\left(2\pi R_k\right)}} \exp\left(-\frac{1}{2}\left(x-\mu_k\right)^T R_k^{-1}\left(x-\mu_k\right)\right) \tag{19}$$

μ_k and R_k are the mean vector and covariance matrix of the Gaussian components, respectively. It has been shown theoretically that for any $f(x)$, as the total number of Gaussian components K increases, $\sum_{k=1}^{K} w_k N\left(x;\mu_k,R_k\right)$ can be infinitely approximated by $f(x)$.

III. B. 2) EM algorithm

The EM algorithm is widely used in the multi-parameter solution problem of ML estimation.ML estimation, as one of the important methods of classical estimation theory, has been widely used in many engineering problems due to its simple and asymptotically effective solution idea.There are many typical methods to solve the parameters of ML estimation, and the most direct method is to solve the derivatives of the likelihood function with respect to each parameter to be solved, which can in turn be used to solve the differential equations or systems of equations. However, this method is often very related to the functional form of the likelihood function, and it is often difficult to find an analytical solution because of the complexity of the functional form. The grid search method has a simple solution logic, but the computational complexity increases exponentially with the number of parameters, and the Newton-Raphson iterative method may encounter convergence problems. Of course, modern optimization algorithms such as genetic algorithm, ant colony algorithm and simulated annealing algorithm can also be used to solve the ML parameter estimation problem, but since the EM algorithm is fast and guaranteed to converge when solving the GMM parameters, this paper adopts the EM algorithm to compute the ML parameters of GMM.

Assuming that the observations are $\{x_1,x_2,\cdots,x_M\}$, that the data are independent of each other, and that the parameter to be estimated is the vector θ, the The likelihood function can be written in the form of equation (20):

$$f(x)=\prod_{m=1}^{M} f(x_m;\theta) \tag{20}$$

Since the logarithmic function does not change the monotonicity of the function, the ML function can be maximized by maximizing the log-likelihood function. To wit:

$$\lg(f(x))=\sum_{m=1}^{M}\lg\left(f(x_m;\theta)\right) \tag{21}$$

To find, the key to the EM algorithm is to introduce the concept of hidden variables to decompose the estimation problem, the introduction of hidden variables should make the estimation problem simple and easy to solve, let the hidden variables be $z^{(m)}$, then equation (21) can be written as:

$$\begin{aligned}\lg\left(f(x)\right)&=\sum_{m=1}^{M}\lg\left(\sum_{z^{(m)}} f(x_m,z^{(m)};\theta)\right)\\&=\sum_{m=1}^{M}\lg\left(\sum_{z^{(m)}} Q_m(z^{(m)})\frac{f(x_m,z^{(m)};\theta)}{Q_m(z^{(m)})}\right)\\&\geq\sum_{m=1}^{M}\sum_{z^{(m)}} Q_m(z^{(m)})\lg\left(\frac{f(x_m,z^{(m)};\theta)}{Q_m(z^{(m)})}\right)\end{aligned} \tag{22}$$

where $Q_m(z^{(m)})$ is the PDF of the implied variable. The derivation of equation (22) uses Jensen's inequality. By the nature of Jensen's inequality, Eq. (22) holds when $f\left(x_m,z^{(m)};\theta\right)/Q_m\left(z^{(m)}\right)$ is a stationary constant, regardless of the arbitrary value of m. Also due to the fact that:

$$\sum_{z^{(m)}} Q_m(z^{(m)})=1 \tag{23}$$

Thus, it can be obtained:

$$Q_m(z^{(m)})=\frac{f(x_m,z^{(m)};\theta)}{\sum_{z^{(m)}} f(x_m,z^{(m)};\theta)}=\frac{f(x_m,z^{(m)};\theta)}{f(x_m;\theta)}=f\left(z^{(m)}\mid x_m;\theta\right) \tag{24}$$

Eq. (23) is the E step of the EM algorithm, while the M step of the EM algorithm is based on the decomposition of the estimation problem based on the $Q_m(z^{(m)})$ obtained from Eq. (24), which is then computed using the general ML solution method:

$$\theta_{\max} = \arg\max_{\theta}\left[\sum_{m=1}^{M}\sum_{z^{(m)}}\left(Q_m(z^{(m)})\lg\left(\frac{f(x_m, z^{(m)};\theta)}{Q_m(z^{(m)})}\right)\right)\right] \tag{25}$$

From equation (24), it can be found that the calculation of $Q_m(z^{(m)})$ in the E step requires prior knowledge of θ, while the calculation of $\theta_{\max}$ in the M step requires prior knowledge of $Q_m(z^{(m)})$. Therefore, the EM algorithm often adopts an iterative computation method in concrete operation, i.e., the initial value of the parameter θ_0 is given beforehand, and then substituting it into the equation (23) to get $Q_m(z^{(m)})$, and then substituting it into the equation (24) to get $\theta_{\max}$, so that Iterate using equations (23) and (24) alternately until the iteration is terminated when the $\theta_{\max}$ obtained from two adjacent iterations is essentially unchanged, and use the iteration result as the final θ estimate.

III. C. Design of intelligent clustering scheme

III. C. 1) KG pre-processing

KG adopts ternary to represent knowledge, and the knowledge element can be expressed as $ke_j = \{c_j, p_j, r_j, a_j\}$, where c_j, p_j, r_j, a_j stand for concepts, entities, relations, and attributes, respectively, and the knowledge domain, which consists of g knowledge elements, is denoted as $ku_d = \{ke_1, ke_2, \cdots, ke_g\}$, the relationship of knowledge domain, knowledge element is shown in Figure 5.

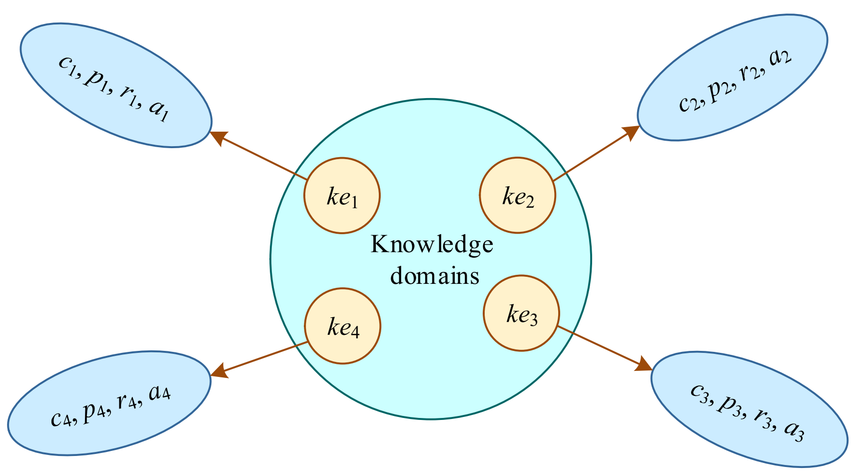
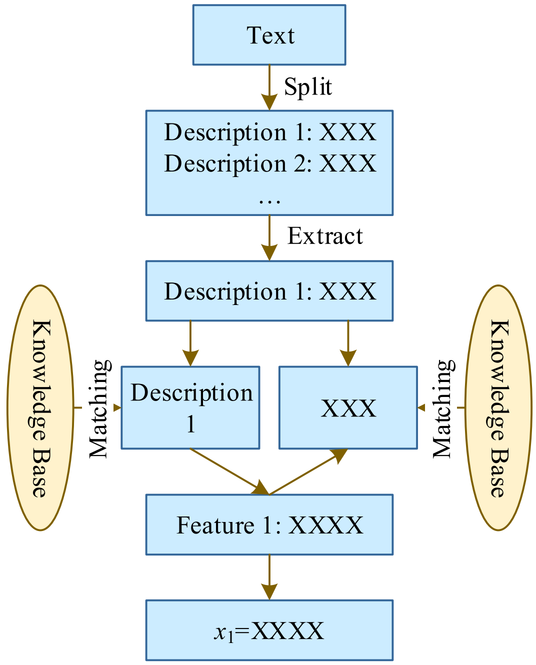


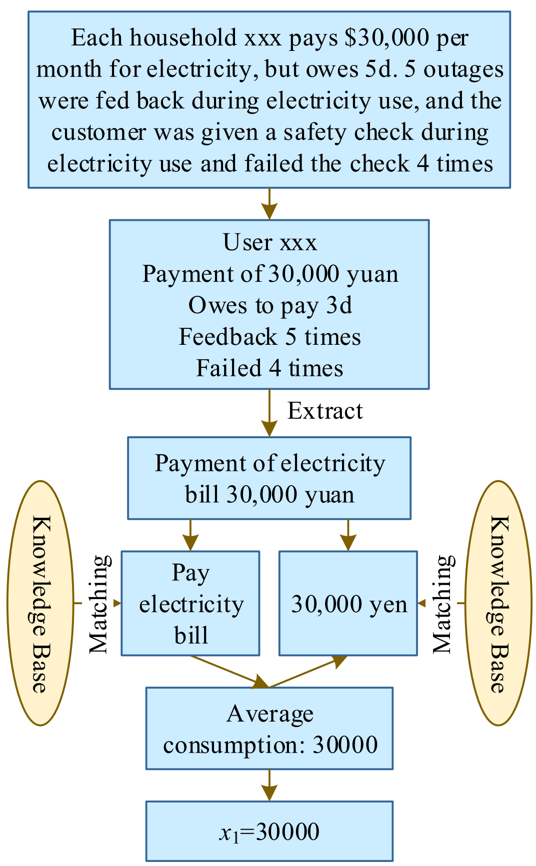
Figure 5: Knowledge domain and knowledge element diagram

For power user behavior, KG can effectively analyze the core concepts as well as the key contents of power user behavior, and for the clustering method, it needs canonical input data $x = [x_1, x_2, \cdots, x_n]^T$, and the knowledge meta-node of the power user behavior is obtained through KG before clustering, to avoid manually dealing with the data required for clustering , the principle of which is shown in Fig. 6.

Fig. 6(a) shows the abstract description of KG preprocessing process, and Fig. 6(b) shows the concrete process of KG preprocessing. Remember that the power user behavior data input is s, and the clustering method input $x = f_{KG}(s)$ is obtained after preprocessing the power user behavior data by KG, and the next subsection describes the GMM-based clustering method.



(a)Abstract description of KG pretreatment process



(b)The specific process of KG pretreatment

Figure 6: KG pretreatment process block diagram

III. C. 2) Gaussian hybrid clustering

For the set of all input samples $D=\{x_1,x_2,\cdots,x_m\}$, assume that these data are divided into k classes, and for one of the inputs x, it obeys a Gaussian mixture distribution, and hence the probability density function is:

$$p(x)=\sum_{i=1}^{k}\alpha_i\cdot p(x\,|\,v_i,C_i) \tag{26}$$

where $p(x\,|\,v_i,C_i)$ denotes one of the components obeying a Gaussian distribution with mean v_i and variance C_i:

$$p(x\,|\,v_i,C_i)=\frac{1}{(2\pi)^{\frac{n}{2}}\left|C_i\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-v_i)^T C_i^{-1}(x-v_i)\right] \tag{27}$$

Where: v_i and C_i are the mean and covariance, respectively, and $(\cdot)^T$ is the transpose operation. For Eq. (27), the parameters are v_i and C_i. Remember the hidden variable $p(y_j=i)=\gamma_{ji}$ as the probability that the sample x_i is from the j th Gaussian distribution component, and by using the following equation, it is possible to determine the i th Gaussian component of x_j:

$$\max\gamma_{ji},\quad j=1,2,\cdots,k \tag{28}$$

To solve for the above parameters, the following log-likelihood function is constructed:

$$LL(D)=\sum_{j=1}^{m}\ln\left[\sum_{i=1}^{k}\alpha_i\cdot p(x_j\,|\,v_i,C_i)\right] \tag{29}$$

For the mean value v_i, the partial derivatives are:

$$\frac{\partial LL(D)}{\partial v_i}=\sum_{j=1}^{m}\frac{\frac{\partial}{\partial v_i}[\alpha_i\cdot p(x_j\,|\,v_i,C_i)]}{\sum_{i=1}^{k}\alpha_i\cdot p(x_j\,|\,v_i,C_i)} \tag{30}$$

Substituting $\frac{\partial}{\partial v_i}\left[\alpha_i\cdot p(x_j\,|\,v_i,C_i)\right]=\alpha_i\cdot p(x_j\,|\,v_i,C_i)\cdot(x_j-v_i)$ into Eq. yields:

$$\frac{\partial LL(D)}{\partial v_i}=\sum_{j=1}^{m}\frac{\alpha_i\cdot p(x_j\,|\,v_i,C_i)\cdot(x_j-v_i)}{\sum_{i=1}^{k}\alpha_i\cdot p(x_j\,|\,v_i,C_i)} \tag{31}$$

Let $\frac{\partial LL(D)}{\partial v_i}=0$ and the update equation for the mean is:

$$v_i=\frac{\sum_{j=1}^{m}x_j\cdot\alpha_i\cdot p(x_j\,|\,v_i,C_i)}{\sum_{i=1}^{k}\alpha_i\cdot p(x_j\,|\,v_i,C_i)} \tag{32}$$

For the covariance C_i, the same can be obtained by solving for $\frac{\partial LL(D)}{\partial C_i}=0$:

$$C_i = \frac{\sum_{j=1}^{m} \alpha_i \cdot p(x_j \mid v_i, C_i)(x_j - v_i)(x_j - v_i)^T}{\sum_{i=1}^{k} \alpha_i \cdot p(x_j \mid v_i, C_i)} \tag{33}$$

Finally, for the parameter α_i, note that there is a constraint on δ_i that introduces the Lagrange multiplier method:

$$L(D,\lambda) = LL(D) + \lambda\left(\sum_{i=1}^{k} \alpha_i - 1\right) \tag{34}$$

Then $\frac{\partial L(D,\lambda)}{\partial \lambda}$ is:

$$\frac{\partial L(D,\lambda)}{\partial \lambda} = \frac{p(x_j \mid v_i, C_i)}{\sum_{i=1}^{k} \alpha_i \cdot p(x_j \mid v_i, C_i)} + \lambda \tag{35}$$

Let $\frac{\partial L(D,\lambda)}{\partial \lambda} = 0$, α_i be: (The document does not give the result here, and it is kept as it is). i.e:

$$\alpha_i = \frac{1}{m} \sum_{j=1}^{m} \alpha_i \cdot p\left(x_j \mid v_i, C_i\right) \tag{36}$$

All three parameters, v_i, C_i, and δ_i, need to be sub-modeled $p(x_j \mid v_i, C_i)$ probabilities, and according to Eq. (37), combining with the Bayesian formula, we can obtain $p(y_j = i \mid x_j)$ as:

$$p(y_j = i \mid x_j) = \frac{p(y_j = i) \cdot p(x_j \mid y_j = i)}{\sum_{i=1}^{k} \alpha_i \cdot p(x \mid v_i, C_i)} \tag{37}$$

Since the i th sub-model needs parameters v_i and C_i, Eq. (37) needs parameters v_i, C_i, and α_i again, and the parameters and the model are iterated alternately to achieve maximum likelihood estimation, which is the principle of the EM method. After initializing the model parameters (α_i, v_i, C_i), the three parameters, mean, covariance, and sub-model probability, are repeatedly updated until these parameters converge at the next iteration.

IV. Analysis of Power User Behavior and Intelligent Segmentation Effects

IV. A. Electricity User Behavior Analysis

IV. A. 1) Similarity analysis

The daily load characteristic curves of different categories provide a discriminatory basis for the initial screening of power users. The period to be tested is one month, and the daily load curve data of the users to be tested are extracted and processed as a weighted average to form a typical daily load curve of the period to be tested. Calculate and analyze the similarity between the characteristic curve and the load curve of the user to be tested. The similarity threshold should be set in combination with the comparison between the sample of power theft users and the characteristic curve in the experiment and the general situation, and the similarity threshold will be set to 0.8 through comparison and analysis, and the part of users with abnormal power consumption will be preliminarily screened out, i.e., the users whose similarity with the characteristic curve is less than 0.8. The abnormal power users need to be further identified and judged.

User A to be tested and the characteristic curve calculation results are shown in Table 1. Through the calculation results of user A to be tested, it can be concluded that the similarity between user A and load characteristic curve 1 is 0.8061112 when using the original DTW, and 0.8261376 when using the optimized DTW, which indicates that the optimized DTW is more suitable for similarity analysis, and it can better reduce the impact of the displacement changes between the power load curves on the similarity calculation. The similarity with other load characteristic curves is lower than 0.8, which reflects that the behavior of user A is in line with the first category and excludes abnormal electricity consumption.

Table 1: User A to be tested and the calculation result of the characteristic curve

User A and load to be measured Characteristic curve comparison	DTW		Optimized DTW	
	Distance	Similarity	Distance	Similarity
Load characteristic curve 1	0.2402704	0.8061112	0.2101452	0.8261376
Load characteristic curve 2	0.7153846	0.5824184	0.6242319	0.6152938
Load characteristic curve 3	0.4094443	0.7091379	0.3803748	0.7241221
Load characteristic curve 4	6.1210527	0.1402044	4.2090697	0.1914851
Load characteristic curve 5	3.8071693	0.2073514	3.5632313	0.2190654
Load characteristic curve 6	1.0473995	0.4881646	0.9764071	0.5054322
Load characteristic curve 7	1.5293665	0.3951466	1.3651995	0.4223808
Load characteristic curve 8	0.3294831	0.7514407	0.2964287	0.7710423

The results of the calculation of the user to be tested B and the characteristic curve are shown in Table 2. Through the calculation results of user B to be tested, it can be concluded that the similarity between user B and load characteristic curve 1 is 0.7792406 when using the original DTW, and the similarity is 0.8030638 when using the optimized DTW. Using the optimized DTW for the calculation and analysis of the similarity between load curves further reduces the influence of the displacement changes between electric load curves on the calculation of the similarity. The optimized DTW is used to calculate and analyze the similarity between load curves, which further reduces the influence of displacement variations between power load curves on the similarity calculation, and more closely reflects the morphological similarity between curves. The similarity calculated by user B using the optimized DTW is 0.8030319, which is greater than the set similarity threshold of 0.8, ruling out abnormalities in electricity consumption and reflecting that user B's electricity consumption behavior is in line with the first category.

Table 2: User B to be tested and the calculation result of the characteristic curve

User B and load to be measured Characteristic curve comparison	DTW		Optimized DTW	
	Distance	Similarity	Distance	Similarity
Load characteristic curve 1	0.2824523	0.7792406	0.2451654	0.8030319
Load characteristic curve 2	0.7362999	0.5754189	0.7062641	0.5854725
Load characteristic curve 3	0.5382612	0.6500103	0.4700925	0.6800932
Load characteristic curve 4	5.9632719	0.1433025	4.0190876	0.1991179
Load characteristic curve 5	4.0101788	0.1992931	3.3092436	0.2320231
Load characteristic curve 6	0.9821471	0.5042329	0.8291374	0.5463323
Load characteristic curve 7	1.5120985	0.3980239	1.2973922	0.4351009
Load characteristic curve 8	0.3820049	0.7232919	0.3563202	0.7370624

The calculation results of user to be tested C and the characteristic curve are shown in Table 3, and the calculation results of user to be tested D and the characteristic curve are shown in Table 4. Through the calculation results in Table 3 and Table 4, it can be concluded that the similarity between user to be tested C, user to be tested D and various types of load characteristic curves is less than the set similarity threshold of 0.8, and the power consumption abnormality exists, and the user to be tested C and user to be tested D are initially screened as users with abnormal power consumption, and further identification judgment is required for them. The process of calculating and analyzing the similarity of the users to be tested is the same as above, so only part of the experimental results are listed.

Table 3: User C to be tested and the calculation result of the characteristic curve

User C and load to be measured Characteristic curve comparison	DTW		Optimized DTW	
	Distance	Similarity	Distance	Similarity
Load characteristic curve 1	1.1821834	0.4581641	1.0541353	0.4863951
Load characteristic curve 2	2.0890828	0.3233554	1.1684284	0.4610311
Load characteristic curve 3	0.7243344	0.5794107	0.6681544	0.5992046
Load characteristic curve 4	4.4861334	0.1821366	3.9513940	0.2014736
Load characteristic curve 5	2.4882681	0.2863265	2.3282274	0.3002198
Load characteristic curve 6	0.8270601	0.5471546	0.7704214	0.5643514
Load characteristic curve 7	0.6471226	0.6070375	0.5940517	0.6271736
Load characteristic curve 8	1.5581919	0.3904353	1.3932206	0.4174042

Table 4: User D to be tested and the calculation result of the characteristic curve

User D and load to be measured Characteristic curve comparison	DTW		Optimized DTW	
	Distance	Similarity	Distance	Similarity
Load characteristic curve 1	0.8244284	0.5474941	0.7614035	0.5672995
Load characteristic curve 2	1.7394809	0.3644842	1.2723804	0.4394917
Load characteristic curve 3	0.3720996	0.7283785	0.3294464	0.7514701
Load characteristic curve 4	5.0943499	0.1640384	3.4091871	0.2263947
Load characteristic curve 5	2.9174007	0.2551222	2.8454619	0.2600127
Load characteristic curve 6	0.5493351	0.6451495	0.5012355	0.6660117
Load characteristic curve 7	0.6903463	0.5912367	0.6420404	0.6084916
Load characteristic curve 8	1.1572842	0.4632423	1.0914215	0.4780236

IV. A. 2) Results of segmentation of users' electricity behavior

A certain period of time or a certain period of time of the power load changes can be reflected through the customer power load curve. Through the user power load curve can intuitively characterize the power user's behavioral habits of electricity, but also the power user load by the passage of time to produce changes in the trend, the size of the depiction. The data samples of user electricity load are extracted from the power metering automation system, and the extracted data samples need to be pre-processed at multiple levels, and then the results are obtained through the clustering algorithm, and the trend of changes under different clustering numbers is shown in Fig. 7, and the characteristic curve of the user's daily load is shown in Fig. 8, in which (a) ~ (h) are respectively 1 to 8 categories of users. From Fig. 7, it can be seen that the optimal number of clusters is when the number of clusters is 8. When the number of clusters is selected as 8, the clustering center curve can be derived after the iterative operation of the clustering algorithm. The differentiated clustering center curves, i.e., the daily load characteristic curves representing different user types, are used to classify the electricity consumption behavior into different categories. For example, in Fig. 8(a), this type of user has higher loads during the daytime and slowly decreases from one peak to another, and may be an ordinary user. The remaining Fig. 8 (b) ~ Fig. 8 (h), with the data performance in the figure as the standard. The use of clustering means of analysis and research, the power consumption behavior law there are differences in the user into the corresponding multiple categories, so that the classification management efficiency has been improved, and the subsequent power Yonghua intelligent subgroups work to reduce the difficulty and complexity of the work can play a more effective role.

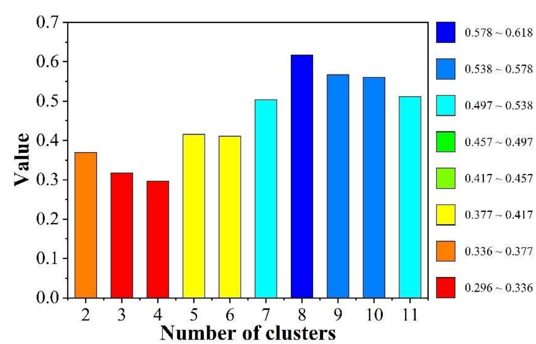
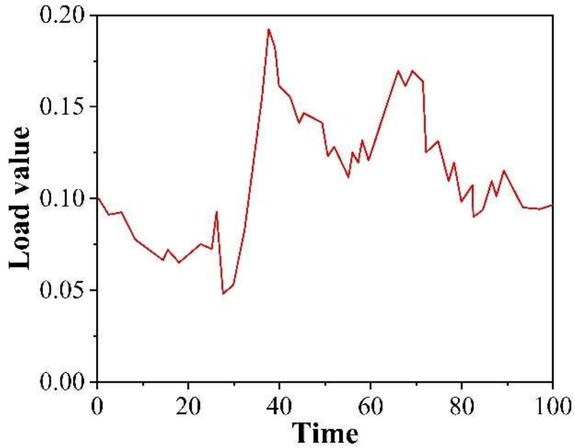
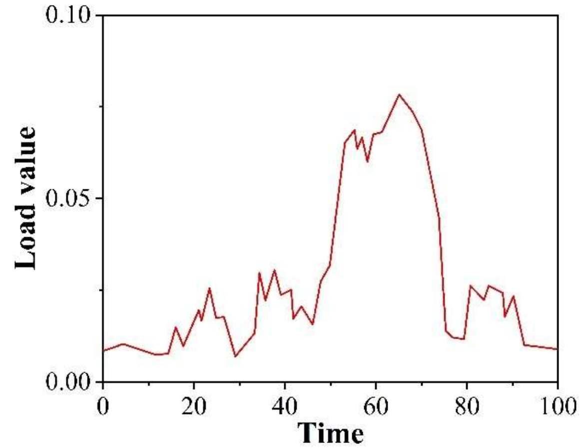


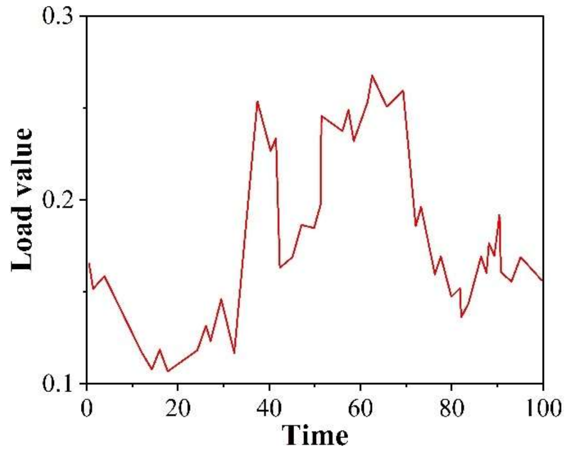
Figure 7: The changing trend under different cluster numbers



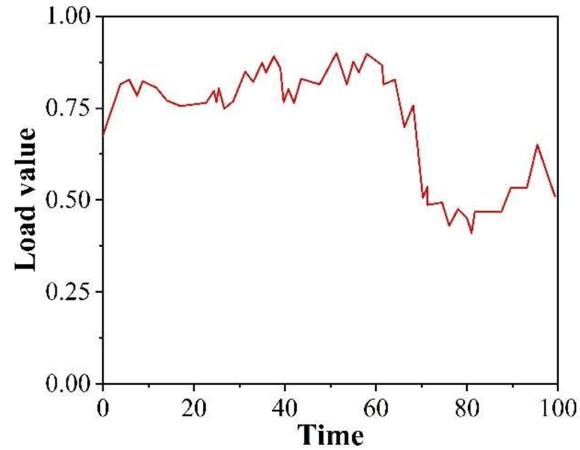
(a)First class user



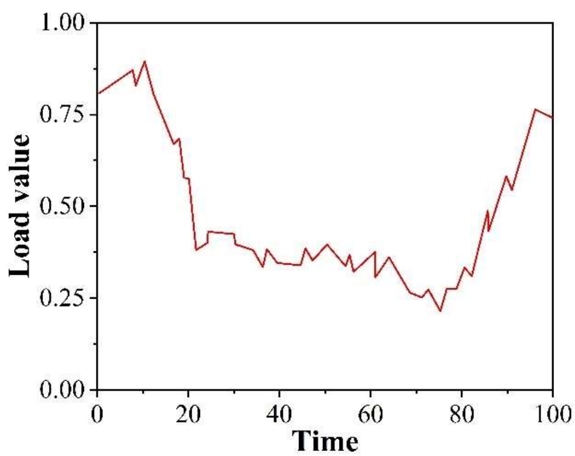
(b)Second class user



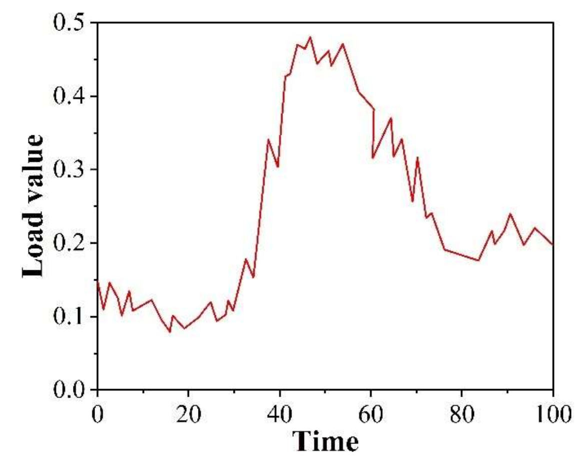
(c)Third user



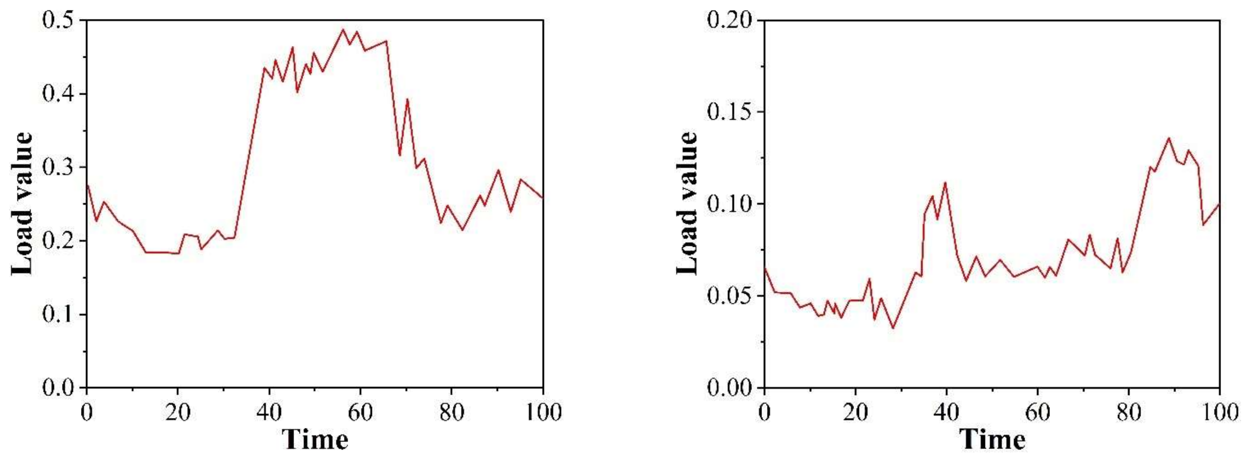
(d)The fourth type of users



(e)Fifth user



(f)Sixth user



(g)Seventh user (h)Eighth user

Figure 8: The results of user's electricity consumption behavior division

IV. B. Analysis of the effect of intelligent clustering of power users

IV. B. 1) Experimental environment setup

TensorFlow is a next-generation artificial intelligence learning system introduced by Google at the 2015 Google Research Blog, and the official definition given for it is that TensorFlow is an open-source software library for numerical computation based on data flow graphs, and its flexible framework is designed to allow users to deploy computation in a single-computer or distributed fashion on desktops, servers and even mobile devices such as cell phones. A data flow graph consists of nodes and lines, with each node representing a mathematical operation and each line representing an input or output relationship between nodes. Data is circulated between nodes, and the data structure is represented by a tensor and transmitted by a line.This experimental environment is built based on the system framework of TensorFlow.

IV. B. 2) Purpose and content of the experiment

For the smart grid user behavior dataset of large data size, high dimensionality, and the use of traditional clustering technology is difficult to achieve the real problem of effective intelligent clustering of smart grid users, this paper proposes a user behavior clustering scheme based on KGEG, through the way of comparative analysis to confirm the effectiveness of this paper's intelligent clustering scheme. The experimental contents are mainly:

(1) The traditional standard k-means clustering algorithm and GMM algorithm are used to cluster the user behavior dataset respectively and run 10 times, and the average of these 10 clustering results is taken as the final standard k-means clustering algorithm and GMM clustering results.

(2) The DAE-k clustering algorithm is used to cluster the user behavior dataset separately, set the hyperparameters and use linear search to find the best value, set the learning rate, and set the maximum number of iterations to 200.

(3) The KGEG algorithm proposed in this paper is used to cluster the user behavior dataset separately, set the hyperparameters, and use linear search to find the best value, set the learning rate, and set the maximum number of iterations to 200 times.

IV. B. 3) Evaluation indicators

In order to avoid the limitation of using a single evaluation method and to make the experimental results more convincing, this paper adopts three evaluation methods to evaluate the final clustering effect of the target clustering algorithm, which are Purity, ACC, and NMI, respectively.The evaluation formula for Purity is:

$$Purity\left(\hat{C}, C\right) = \frac{1}{N}\sum_{k}\max_{j}\left|\hat{c}_k \cap c_j\right| \tag{38}$$

where $\hat{C}(\hat{c}_1,\hat{c}_2,\cdots,\hat{c}_k)$ denotes the kth cluster after the algorithmic clustering, $\hat{c}_k$ denotes the kth cluster set, $C=\{c_1,c_2,\cdots,c_k\}$ denotes the set of sample points after clustering, c_j denotes the jth sample point, and N is the total number of sample points.ACC's evaluation formula is:

$$ACC=\frac{1}{N}\sum_{i=1}^{N}\delta\left(map\left(r_i\right),l_i\right) \tag{39}$$

where $map\left(r_i\right)$ is the label of the clustering result and l_i is the real label of the original data.

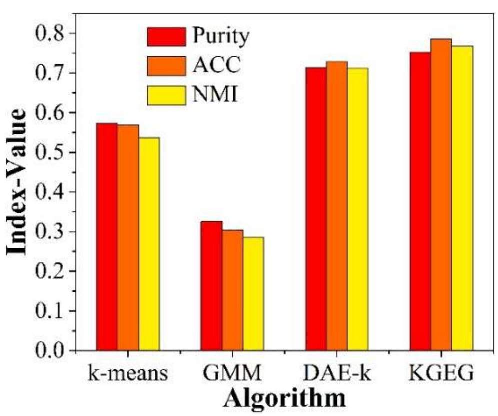
The evaluation formula of NMI is:

$$NMI\left(r,\hat{c}\right)=\frac{I\left(r,\hat{c}\right)}{\frac{1}{2}\left(H\left(r\right)+H\left(\hat{c}\right)\right)} \tag{40}$$

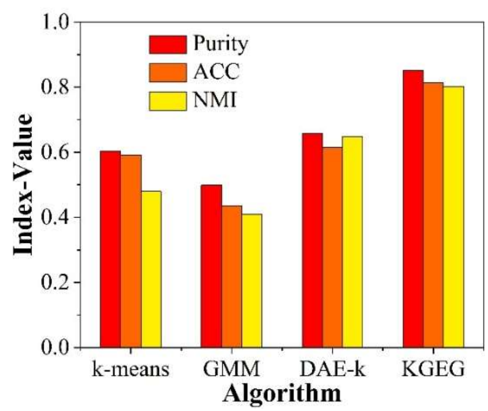
where r is the set of confirmed labels, $\hat{c}$ is the set of clustered combined labels, $I\left(r,\hat{c}\right)$ is the mutual information entropy of both, and $H\left(r\right)$ and $H\left(\hat{c}\right)$ are the information entropies, respectively.

IV. B. 4) Analysis of experimental results

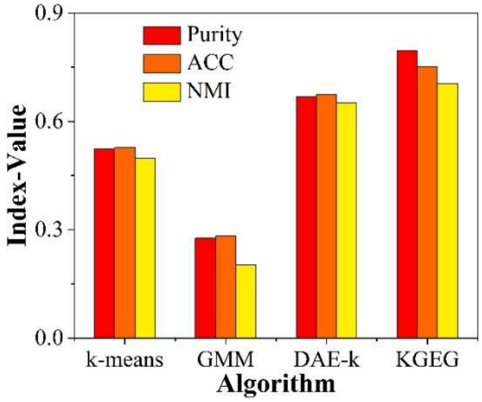
In this paper, the traditional standard k-means, GMM clustering algorithms as well as the DAE-k and KGEG algorithms based on deep learning techniques proposed in this paper are used to cluster each segmented power user behavior dataset, and the clustering results of the four algorithms on the dataset are shown in Fig. 9, in which (a) ~ (j) are datasets A ~ J, respectively. From the data in the above table, it can be seen that except for the GMM clustering algorithm, the k-means, DAE-k, and KGEG algorithms end up with almost all accuracies above 0.5. In each dataset in the table, this paper puts four algorithms, k-means, GMM, DAE-k, and KGEG, under the three evaluation indexes of Purity, ACC, and NMI, to carry out the intelligent clustering of electric power user behavior, and it can be seen that most of the data with the best clustering effect are concentrated on the KGEG algorithm. Among them, the highest clustering accuracy is on dataset J. The KGEG algorithm reaches 0.8926 on ACC evaluation index, which verifies that the algorithm of this paper has excellent application effect on top of smart grid user behavior clustering, and provides auxiliary decision-making for smart grid user management.



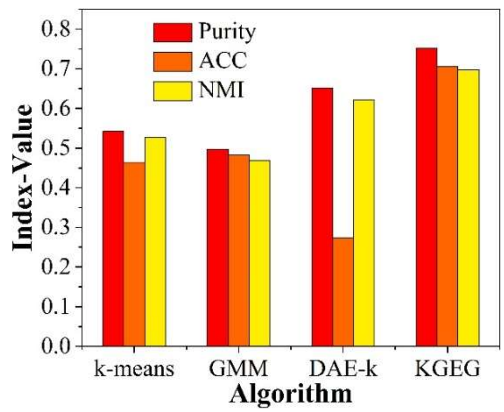
(a)Data set A



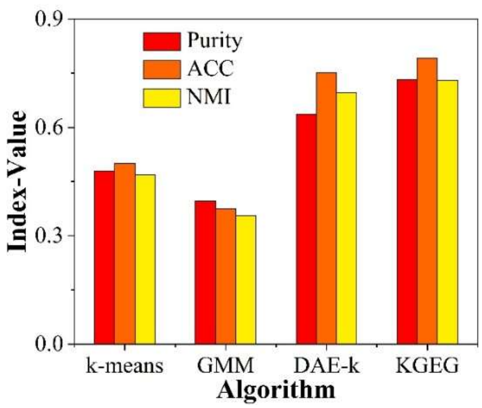
(b)Data set B



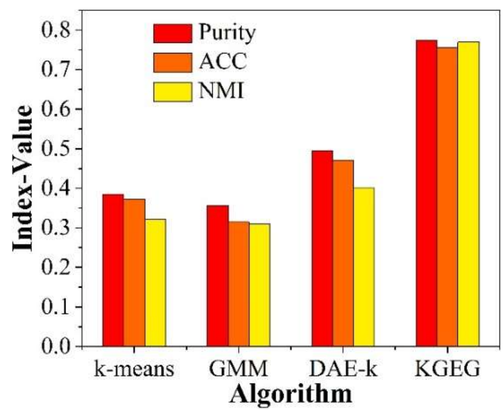
(c)Data set C



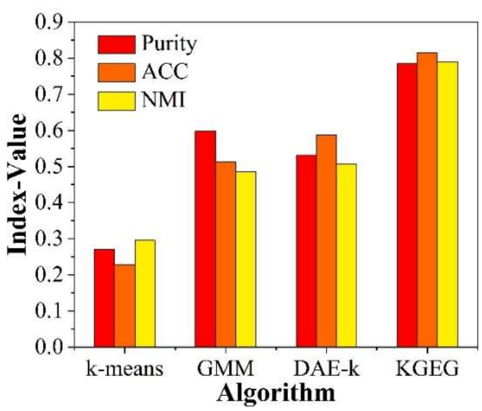
(d)Data set D



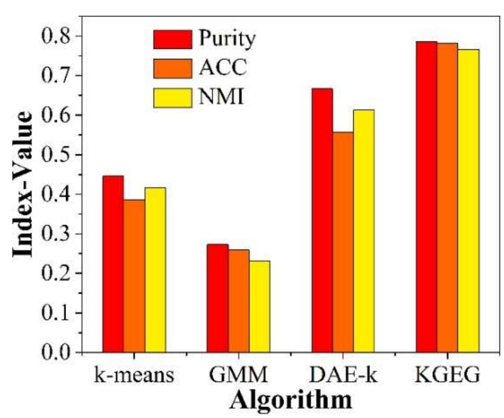
(e)Data set E



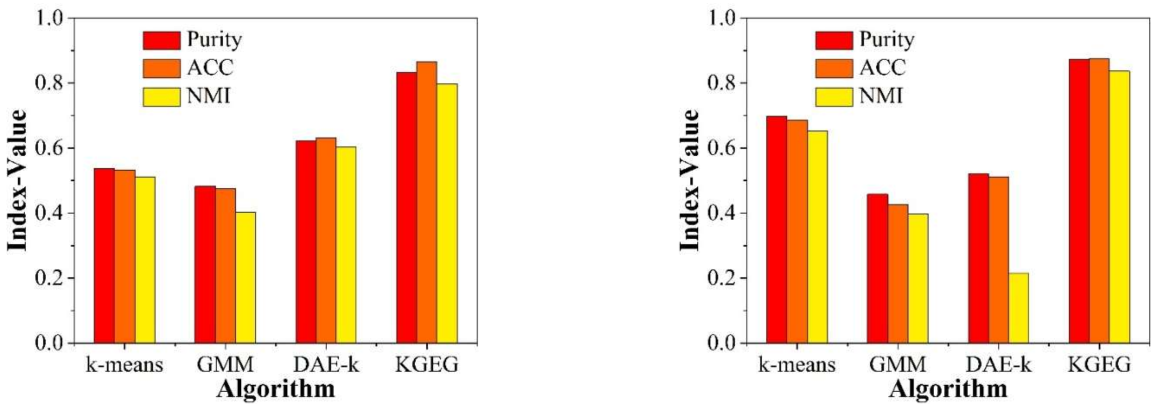
(f)Data set F



(g)Data set G



(h)Data set H



(i)Data set I (j)Data set J

Figure 9: the clustering results of four algorithms on the data set

V. Conclusion

In the context of smart grid research, the similarity of power user behavior is first calculated using the improved DTW algorithm to divide the realistic power user behavior, which ensures the rigor of the research work. On the basis of power user behavior division, the intelligent clustering dataset is prepared by constructing a knowledge graph, and an intelligent clustering research scheme based on the KGEG algorithm is purposely designed to achieve the goal of intelligent clustering of power user behavior. Finally, the comprehensive dataset and evaluation indexes are integrated to explore and analyze the power user behavior and intelligent clustering. On the three evaluation indexes of Purity, ACC and NMI, compared with k-means, GMM and DAE-k, the intelligent clustering effect of this paper's algorithm is the most excellent, and the values of the indexes are distributed in the range of 0.6~0.9, which proves that this paper's algorithm has a very outstanding application value in intelligent clustering of electric power users' behaviors, and it can meet the demand of smart grid well and promote the Smart grid high-quality development, so that it can better serve the users.

References

[1] Dileep, G. J. R. E. (2020). A survey on smart grid technologies and applications. Renewable energy, 146, 2589-2625.

[2] Hossain, M. S., Madlool, N. A., Rahim, N. A., Selvaraj, J., Pandey, A. K., & Khan, A. F. (2016). Role of smart grid in renewable energy: An overview. Renewable and Sustainable Energy Reviews, 60, 1168-1184.

[3] Kim, S. K., & Huh, J. H. (2018). A study on the improvement of smart grid security performance and blockchain smart grid perspective. Energies, 11(8), 1973.

[4] Colak, I., Bayindir, R., & Sagiroglu, S. (2020, June). The effects of the smart grid system on the national grids. In 2020 8th International Conference on Smart Grid (icSmartGrid) (pp. 122-126). IEEE.

[5] Faheem, M., Shah, S. B. H., Butt, R. A., Raza, B., Anwar, M., Ashraf, M. W., ... & Gungor, V. C. (2018). Smart grid communication and information technologies in the perspective of Industry 4.0: Opportunities and challenges. Computer Science Review, 30, 1-30.

[6] Kapse, M. M., Patel, N. R., Narayankar, S. K., Malvekar, S. A., & Liyakat, K. K. S. (2022). Smart grid technology. International Journal of Information Technology and Computer Engineering, 2(6), 10-17.

[7] Wang, Y., Ma, J., Gao, N., Wen, Q., Sun, L., & Guo, H. (2023). Federated fuzzy k-means for privacy-preserving behavior analysis in smart grids. Applied Energy, 331, 120396.

[8] Sun, Y., Jia, M., Lu, J., Zhang, B., & Yang, W. (2016). Research on smart grid users' power consumption behavior classification based on improved k-means algorithm. International Journal of Electrical Energy, 4(1), 6-10.

[9] Lv, Y. C., Xu, X., Xu, R. L., & Qiu, X. (2019, January). Quantitative Analysis Model of User's Electrical Behavior for Smart Grid. In 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 281-284). IEEE.

[10] Guo, Z., Zhou, K., Zhang, C., Lu, X., Chen, W., & Yang, S. (2018). Residential electricity consumption behavior: Influencing factors, related theories and intervention strategies. Renewable and Sustainable Energy Reviews, 81, 399-412.

[11] Wang, Y., Chen, Q., Kang, C., & Xia, Q. (2016). Clustering of electricity consumption behavior dynamics toward big data applications. IEEE transactions on smart grid, 7(5), 2437-2447.

[12] Kaur, R., & Gabrijelčič, D. (2022). Behavior segmentation of electricity consumption patterns: A cluster analytical approach. Knowledge-based systems, 251, 109236.

[13] Yang, W., Shi, J., Li, S., Song, Z., Zhang, Z., & Chen, Z. (2022). A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior. Applied Energy, 307, 118197.

[14] Li, H., Hu, B., Liu, Y., Yang, B., Liu, X., Li, G., ... & Zhou, B. (2021). Classification of electricity consumption behavior based on improved K-means and LSTM. Applied Sciences, 11(16), 7625.

[15] Motlagh, O., Paevere, P., Hong, T. S., & Grozev, G. (2015). Analysis of household electricity consumption behaviours: Impact of domestic electricity generation. Applied Mathematics and Computation, 270, 165-178.

[16] Zhang, W., Dong, X., Li, H., Xu, J., & Wang, D. (2020). Unsupervised detection of abnormal electricity consumption behavior based on feature engineering. Ieee Access, 8, 55483-55500.

[17] Siebert, L. C., Sbicca, A., Aoki, A. R., & Lambert-Torres, G. (2017). A behavioral economics approach to residential electricity consumption. Energies, 10(6), 768.

[18] Zhao, Q., Li, H., Wang, X., Pu, T., & Wang, J. (2019). Analysis of users' electricity consumption behavior based on ensemble clustering. Global Energy Interconnection, 2(6), 479-488.

[19] Zhu, L., Liu, J., Hu, C., Zhi, Y., & Liu, Y. (2024). Analysis of Electricity Consumption Pattern Clustering and Electricity Consumption Behavior. Energy Engineering, 121(9).

[20] Bedir, M., & Kara, E. C. (2017). Behavioral patterns and profiles of electricity consumption in dutch dwellings. Energy and Buildings, 150, 339-352.

[21] Acuner, E., & Kayalica, M. O. (2020). Behavior analysis of refugees' electricity consumption in developing countries: Case of Turkey. IEEE Transactions on Engineering Management, 69(4), 1206-1215.

[22] Wang, F., Liu, L., Yu, Y., Li, G., Li, J., Shafie-Khah, M., & Catalao, J. P. (2018). Impact analysis of customized feedback interventions on residential electricity load consumption behavior for demand response. Energies, 11(4), 770.

[23] Li, Q., Wang, G., Zhang, Y., & Yang, Q. (2023). Analysis of user electricity consumption behavior based on density peak clustering with shared neighbors and attractiveness. Concurrency and computation: practice and experience, 35(3), e7518.

[24] Yang, T., Ren, M., & Zhou, K. (2018). Identifying household electricity consumption patterns: A case study of Kunshan, China. Renewable and Sustainable Energy Reviews, 91, 861-868.

[25] Soares, A., Gomes, Á., & Antunes, C. H. (2014). Categorization of residential electricity consumption as a basis for the assessment of the impacts of demand response actions. Renewable and Sustainable Energy Reviews, 30, 490-503.

[26] Changjiang Niu. (2024). The application of improved DTW algorithm in sports posture recognition. Systems and Soft Computing,6,200163-200163.

[27] Junhong Chen & Kaihui Peng. (2024). Research on Text Similarity Algorithms Based on Interactive Attention. Academic Journal of Computing & Information Science,7(10),