Computing and Informatics, Vol. 42, 2023, 923–942, doi: 10.31577/cai_2023_4_923

RETRIEVAL TECHNOLOGY OF ENTERPRISE DATA CENTER RESOURCES BASED ON DENSITY PEAK CLUSTERING ALGORITHM

Jiaming JIANG*, Guoheng RUAN, Zhenggan DAI

Qingyuan Power Supply Bureau of Guangdong Power Grid Co., Ltd.
Qingyuan, 511500, China
e-mail: {jiangjiaming0620, ruanguoheng, daizenggan2022}@163.com

Abstract. In order to effectively ensure the retrieval effect of enterprise data center resources, improve the retrieval accuracy of enterprise data center resources, and shorten the retrieval time of enterprise data center resources, a retrieval technology of enterprise data center resources based on density peak clustering algorithm is proposed. Analytical clustering algorithms, density clustering algorithms, and density peak clustering algorithms are all types of clustering algorithms. To reduce the dimensionality of enterprise data center resources, the kernel principal component analysis method is used. The structure of the enterprise data center resource set is reorganized and the feature quantity of the enterprise data center resource distribution is extracted using feature space reorganization technology. On this basis, the density peak clustering is carried out on the data center resource set of enterprise, and the semantic association distribution model of data center resource retrieval in enterprise is constructed. Through the semantic registration and weighted vector combination control method, the retrieval of enterprise data center resources is realized. The experimental results show that the proposed algorithm has a good effect on the retrieval of enterprise data center resources, which can effectively improve the resource retrieval accuracy and shorten the resource retrieval time.

Keywords: Density peak clustering algorithm, enterprise data center, kernel principal component analysis method, resource retrieval, semantic correlation distribution

* Corresponding author

1 INTRODUCTION

The data center is the precipitation of the existing/new information system business and data, and is the middle and supporting platform for realizing data empowerment of new business and new applications [1, 2, 3]. Enterprises build a data center, and through the aggregation and reuse of business and data, guide the company's power grid, industry, finance and international sector resources, systems and data integration. Effectively improve the rapid response and flexible adjustment capabilities of the information system, effectively empower front-end business applications and enhance innovation capabilities. As an important platform for data sharing and analysis within the enterprise, the enterprise data center carries the massive data resources of the enterprise, and at the same time realizes rapid response support for various data application needs. With the advancement of the construction of the enterprise data center, the demand for front-end business applications will also increase rapidly. Faced with more and more data resources and data services, how to quickly support business applications and help users find, see, and understand data in an extremely fast way is particularly important. Therefore, it is necessary to apply data resource retrieval technology to quickly and accurately find the desired content from massive information [4, 5, 6]. The pros and cons of the data resource retrieval technology largely determines the utilization rate of the data resources in the data center, which makes the data resource retrieval technology more important.

At present, scholars in related fields have carried out research on data resource retrieval, and have made great progress. Reference [6] proposes a retrieval algorithm for online tourism resources based on Page Rank search and ranking algorithm. A topic collection algorithm is constructed, and a starting point, topic keywords, and prediction mechanism are established. The algorithm consists of three stages: the first climbing stage, the learning stage and the continuous climbing stage. Open directory search was used for similarity judgment and result evaluation. Wood extraction algorithm is based on network tourism resource density. The algorithm calculates the proportion of Internet tourism resource labels by row, and uses a threshold extraction algorithm to distinguish regions from private non-Internet tourism resource regions. This paper takes tourism network resource monitoring as the research object, and establishes a tourism network resource monitoring system, which can provide users with customizable, all-round, real-time tourism network resource collection, extraction and retrieval services, so as to monitor tourism resources. This method can successfully extract the main content of articles from various web pages. The research results of this paper can promote the construction of tourism informatization, help users master the latest tourism information, and bring great convenience to the tourism industry. The system only downloads tourism-related information through theme collection technology, reducing the interference of irrelevant redundant web pages. Reference [7] proposed an image network teaching resource retrieval algorithm based on deep hashing algorithm. A pixel big data detection model of multi-view attribute coding image network teaching resources is constructed, and the pixel information collected by multi-view attribute coding image network teaching resources is reconstructed. The fuzzy information feature components of the multi-view attribute-encoded images are extracted, and the edge contour distribution images are combined. Distributed fusion network teaching resource view image edge contour, realizes the construction of view feature parameter set. The gray invariant moment feature analysis method is used to complete the information encoding, and the deep hash algorithm is used to realize the retrieval of multi-view attribute-encoded image network teaching resources. The algorithm has a high level of resource fusion for multi-view coded image network teaching resource retrieval. However, the above methods still have the problems of poor resource retrieval effect, low precision and long time.

Aiming at the above problems, this paper proposes a data center resource retrieval technology for enterprise based on density peak clustering algorithm. The kernel principal component analysis method is used to reduce the dimensionality and process of enterprise data center resources. Combined with feature space reorganization technology, the feature quantity of resource distribution in enterprise data center is extracted. On this basis, the density peak clustering is carried out on the enterprise data center resource set, and the enterprise data center resource retrieval is realized through semantic registration and weighted vector combination control method. The resource retrieval effect of the algorithm is good, which can effectively improve the resource retrieval accuracy and shorten the resource retrieval time.

2 RELEVANT BASIC THEORIES

2.1 Clustering Algorithm

Clustering is an important data mining technique. Clustering can find hidden patterns and trends in data without any supervised information such as data labels [8]. Graph analytics is considered as most influential tool able to guide to unwrap the hidden patterns and relationships in the data. Therefore, from a machine learning point of view, clustering is a form of unsupervised learning, where the cluster classes correspond to latent structures. Here, a simple definition can be presented: given a set of data points, they are divided into multiple clusters, where similar data points are in the same cluster, and dissimilar data points are in different clusters.

A clustering algorithm usually involves four steps: data representation, modeling, optimization, and validation. The data representation predetermines which data type is used to analyze the data. Data representation defines the form in which data is stored, processed and transmitted. On the basis of data representation, the modeling phase defines the concept of cluster class, that is, the data objects are divided. Typically, the quality of this partition is assessed by an approximate metric. For clustering analysis, the Gaussian mixture of models is considered as the most prominent model which says that the dataset is generally modelled with fixed number of Gaussian supplies. The optimization stage is to optimize or approximate the above-mentioned quality evaluation criteria when finding these hidden cluster structures according to the clustering goal. This mass is optimized or approximately optimized to produce a clustering solution. The Elbow method is the common method for defining the optimal number of clusters. The verification stage is to evaluate the clustering results obtained through the above scheme.

Let a dataset contain Q data objects, which can also be called data records or data points, expressed as:

$$W = \{w_1, w_2, \ldots, w_Q\}. \tag{1}$$

Each data object can be represented by an E-dimensional vector as:

$$w_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,E})^T. \tag{2}$$

In Equation (2), $w_{i,E}$ is the E attribute of w_i, which can also be called a feature or dimension. The number of attributes E is also known as the dimension of the dataset. The dataset can be divided into clusters and expressed as:

$$R = \{R_1, R_2, \ldots, R_T\}. \tag{3}$$

In Equation (3), Y is the number of clusters, and the division satisfies the following conditions:

$$W = R_1 \cup \cdots \cup R_Y \cup R_O. \tag{4}$$

And for all $i, E = 1, 2, \ldots, Y, i \neq E$, hard clustering is expressed

$$R_i \cap R_E = \emptyset. \tag{5}$$

Hard clustering is a process of grouping the data such that an item can exist in various clusters. This method is also known as non-fuzzy clustering. The data points usually belong to the various clusters in hard clustering.

Since clustering is a fairly common problem, clustering algorithms can be divided into several categories based on clustering techniques: Model-based clustering algorithms, distance-based clustering algorithms, density and grid-based clustering algorithms, graph-based clustering, subspace clustering, etc. The classification of clustering algorithms is shown in Figure 1.

1. Model-based clustering is based on the assumption: the data is generated by a mixture of multiple probability distributions. Probability distributions are utilized to explain the real-life variables populations and it is also used in hypothesis testing for determining values of p. It is commonly considered as the statistical function which defines all the conceivable values. Therefore, this type of clustering method estimates the parameters of each distribution in order to perfectly fit the observed data. The whole clustering process is to first assume a specific generative model, and then use the expectation maximization algorithm (EM, Expectation Maximization) to estimate the parameters of

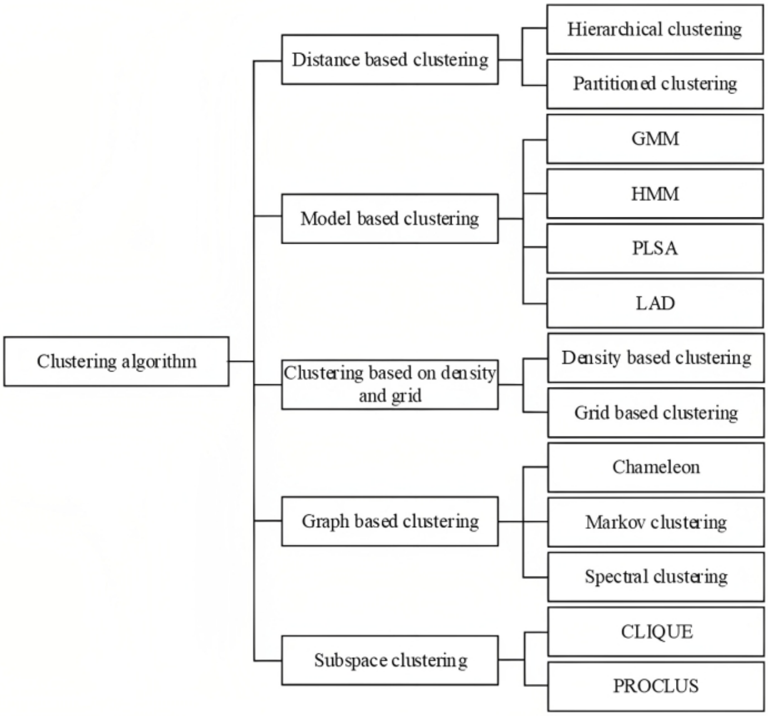


Figure 1. Classification of clustering algorithms

the model. Expectation Maximization is an estimation of mixture model which is the technique of probability density used in several applications. For implementing the distribution-based algorithm, the Data mining used the expectation maximization.

2. Distance-based clustering method can be regarded as a special form based on model. For example, the k-Means algorithm is closely related to the Gaussian distribution. Because, distance-based clustering methods are easy to implement and very simple in various scenarios such as consumer datasets, statistical profiles (age, gender, income, etc.), purchase history and web browsing activities. In order to understand a broad group of customers' behavior and preferences, such data are utilized to identify them together with their attributes. Therefore, such clustering algorithms are widely used. Distance-based clustering algorithms can generally be divided into two categories: hierarchical clustering algorithms and partition clustering algorithms.

   Hierarchical clustering algorithms: An algorithm is the analysis of hierarchical cluster which combines the comparable object into groupings called as clusters. A series of partitions of data objects has been developed by this method.

Partition clustering algorithm: This kind of algorithm is generally converting a distribution of set of data objects into form of clusters and so that every data object is in the form of one subset. A partitional clustering can be able to decay the set of data into set of disjoint clusters.

3. Density- and Grid-based clustering are two closely related clustering methods. Both types of approaches attempt to explore the data space at a high granularity level.

   Density-based clustering: Density-based clustering methods are built on a fundamental assumption: clusters are defined as dense regions separated by low-density regions. In k-means algorithm, the distance between the points has been determined by using distance calculation technique. And the Euclidean distance method is considered as the most frequently used method. By applying such phases, the density-based algorithm can be able to discover high density regions and disperse them from low density regions. It partitions the data in the data space at a higher granularity and "merges" dense regions together. At a high level of granularity, the solutions of density and grid-based clustering are the major two closely correlated categories which attempt two categories attempt to discover the data space. The density and grid-based cluster allows the arbitrary shape cluster and categorize the outliers in the data.

   Grid-based clustering: Grid-based clustering methods are a special type of density-based clustering methods in which the explored data space is partitioned into a grid-like structure.

4. Graph-based clustering is a class of clustering algorithms based on graph theory. The main idea of this type of clustering is to regard data objects as complete graphs or nodes in connected graphs and define the weight between nodes as the distance between data objects. In the simplest way to describe, all nodes form a complete graph connected to each other, using the similarity between data objects to represent the edges of the graph, then the data can be represented in the form of a weighted complete graph. When similar data samples are gathered into only one cluster, the data objects are frequently used in clustering method. Other data samples are gathered into various ones. After the graph is created, some of the edges are removed by some rules. Based on the weights of the cluster all the edges of the graphs are categorized in the form of descending order. And so, for every edge, starting from the maximum weighted edge, all the edge has been removed when the weight of the current edge is much greater than the number or the sum of adjacent edges. Further, remove each edge that has a weight greater than the average adjacent edge weight. After the above process, a disconnected graph is obtained, in which each subgraph represents a cluster class.

5. Subspace clustering methods can be considered as a form of local feature selection or local dimensionality reduction, which performs feature selection or

transformation for different subsets of the data. The techniques of dimensionality reduction might be well-defined as it was path of changing the higher dimensions dataset into lower dimension dataset which also ensured that it delivers alike information's. when resolving the classification, these methods are broadly utilized in machine learning for gaining good fit predictive model. Subspace clustering can be seen as an extension of traditional clustering. The subspace clustering algorithm is a kind of algorithm which is able to locate clusters within various subspaces. Generally speaking, subspace clustering can be divided into two categories: Bottom-up algorithms and Top-down algorithms.

2.2 Density Clustering Algorithm

Density-based clustering is a class of nonparametric methods that treat regions of high density as clusters [9]. Density-based clustering methods are based on the assumption of density-based clusters. A density-based cluster is a contiguous set of high-density regions separated by low-density regions. The density-based clustering method is one of the most famous unsupervised learning methods which is used in the algorithm of machine learning and model building. By the two low point density clusters, the data points have been separated and such clusters are regarded as noise.

This class of methods relies on two important parameters: the distance threshold and the density threshold. The distance threshold and the density threshold parameters play vital roles in the density-based clustering algorithm. Further, the distance threshold defines that the separation between two major consecutive elements in the cluster is curved to the following decimal point whereas the density threshold has been well-defined by two parameters such as the neighborhood radius, i.e. eps, and data points numbers, i.e. minPts. These two can be comprehended if prospecting the two concepts called density reachability and connectivity. One of the most important concepts in density-based methods is the nearest neighbor of a data point, which is defined as follows: the nearest neighbor is considered the most essential in cluster algorithms. Also, the analysis of the nearest neighbor is expressive statistics which displays locating feature patterns by contrasting the clearly observed nearest neighbor distance. It is also considered one of the simplest procedures which can be able to classify the rule of the nearest neighbor.

The α nearest neighbor of a data point $u \in U$ is:

$$Q_\alpha(u) = \{o \in U : \mathrm{dist}(u, o) \leq \alpha\}. \tag{6}$$

In Equation (6), $dist(,)$ is a specific distance function.

1. Directly density-reachable: If a data point $u \in U$ satisfies $u \in Q_\alpha(o)$ and $|Q_\alpha(o)| \geq p$, then the data point u is directly density-reachable from the data point o.
2. Density-reachable: If there is a sequence of data points denoted by $\{u_{a1}, u_{a2}, \dots, u_{an}\}$, $u_{a1} = u$, $u_{an} = 0$, such that for $i = 1, 2, \dots, n-1$, there is a direct density

reachable from data point u_{an+1} to u_{an}, and data point u from data point o is density reachable.

3. Density-connected: If there is a data point with the given parameters α and p and data points are density reachable from the data point, the two data points are then density-connected using these parameters.

A cluster class can be defined as the largest set of densely connected points. Mathematically, it can be defined as follows: Cluster: Given parameters α and p, a cluster needs to satisfy the following two properties:

1. Maximality: For $\forall u, o$, if $u \in C$ is density-reachable from u to o, then $o \in C$.
2. Connectivity: For $\forall u, o \in C, u$ and o are densely connected.

Based on the two parameters α and p and the above definitions, density-based clustering can identify three different types of data points:

1. Core point: A core point is a data point with high-density neighbors, that is, the number of α nearest neighbors is greater than or equal to the density threshold p.
2. Border point: The border point also belongs to a certain cluster, but its neighbors are not dense. According to the border point of the clusters, the latent cross cluster is utilized for removing the edges of the cross-cluster.
3. Noise point: The noise point does not belong to any cluster whereas the noise point has been considered as neither a core point nor a border point. Also, it classifies arbitrary size clusters in the database along with outliers. A noise point is not assigned for two points such as core and border points.

For data points that are both core points, the direct density reachability is a symmetric relationship, but if a core point and a boundary point are involved, the direct density reachability is usually not symmetrical. Density accessibility, as an extension of direct density accessibility, is also asymmetric.

2.3 Density Peak Clustering Algorithm

The density peak clustering algorithm is a new clustering algorithm that can find non-convex clusters. Density peak cluster classify the centers of cluster at ease without any preceding data. The non-convex clusters are the clusters which are usually incapable to recognize the clusters along with the shapes of non-convex specially the manifold ones. The core idea of the algorithm is based on two important assumptions about the cluster center point or the density peak point [10, 11].

Assumption 1. The local density of the cluster center point is greater than the local density of its surrounding neighbors.

Assumption 2. There is a relatively large distance between the cluster center point and other center points.

The above two assumptions not only describe the cluster center point but also give a criterion for detecting the center point. To quantify the above assumptions, two important values are introduced for each data point u_i: the distance δ_i of the data point whose local density χ_i is greater than it and is closest to it.

The local density χ_i is defined as:

$$\chi_i = \sum_{u_j \in U} \varepsilon\left(\text{dist}\left(u_i, u_j\right) - d_c\right),$$
$$\varepsilon(u) = \begin{cases} 1, & u < 0, \\ 0, & u \geq 0. \end{cases} \tag{7}$$

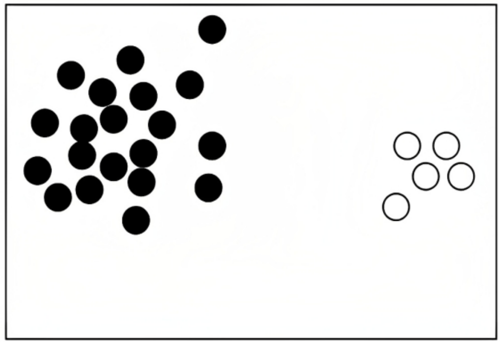
In Equation (7), d_c is called the cutoff distance, which is the only input parameter of the density peak clustering algorithm and actually plays the role of the distance threshold.

The data point distance δ_i is defined as:

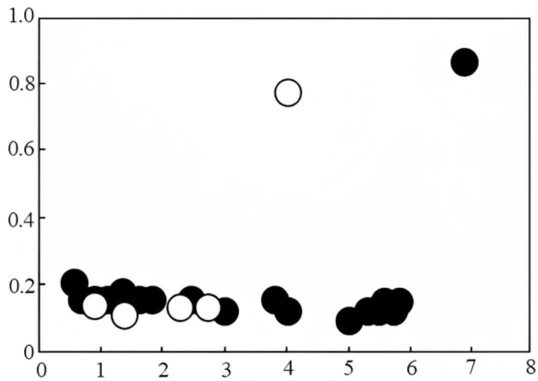
$$\delta_i = \begin{cases} \min_{j:\chi_i < \chi_j}\left(\text{dist}\left(u_i, u_j\right)\right), & \text{if Jjs.t.}\chi_i < \chi_j, \\ \min_j\left(\text{dist}\left(u_i, u_j\right)\right), & \text{otherwise.} \end{cases} \tag{8}$$

Similar to the clustering algorithm based on the center point, the density peak clustering algorithm also needs to find the center point (density peak) of the cluster. In order to obtain the cluster center point, the density peak clustering algorithm constructs the decision graph of χ and δ, and selects the data point with larger χ and δ as the cluster center. Simply put, the data point at the top right of the decision diagram is manually selected as the cluster center. The 2D medium density peak clustering algorithm is shown in Figure 2.

In Figure 2, 25 data points are shown, and the decision graph for this data contains χ and δ for each point. Figure 2 b) is above the density peak points, these two points have high δ and relatively high χ. However, the distribution of the remaining points is different from the clustering algorithm based on the center points. In terms of allocation strategy, this algorithm is similar to density-based clustering algorithm or agglomerative hierarchical clustering algorithm, and the attribution of data points depends on the attribution of surrounding points. The agglomerative clustering is defined as most common type of hierarchical clustering which is used for collection of objects in cluster based. It is also known as AGNES which defines Agglomerative Nesting. By treating every object as a single cluster, this algorithm has been initiated. For the remaining data points u_i, the density peak clustering algorithm classifies them into the cluster class of the data points whose density is greater than u_i and is the closest to u_i, and only needs to complete the assignment of the remaining data points in one step.



a) Data point distribution



b) Decision start

Figure 2. 2-dimensional medium density peak clustering algorithm

3 RETRIEVAL ALGORITHM OF ENTERPRISE DATA CENTER RESOURCE

3.1 Dimensionality Reduction Processing Enterprise Data Center Resources

Because there are many kinds of characteristic data in the enterprise data center resources, the distribution of data points in space is relatively sparse, and the similarity measurement method in low-dimensional space cannot effectively process high-dimensional data. The similarity measurement method measures data mining building blocks and utilized in recommendation engines, methods of clustering and detecting anomalies. This method is a distance along with dimensions to express the object features. The similarity between two data objects is determined by using the measurements method. To this end, the kernel principal component analysis method [12, 13] is used to reduce the dimensionality of enterprise data center resources. In kernel principal component analysis, for the reduction of dimensionality, the linear, polynomial, sigmoid kernel methods are used. As the reduction of dimensionality k-means clustering is put in to the document's reduction. Assuming that the resource set in the original enterprise data center is $D = \{d_1, d_2, \ldots, d_n\} \in R^m$, the nonlinear mapping is:

$$\begin{aligned} &\phi : R^m \to F, \\ &d \mapsto g = \phi(d), \end{aligned} \tag{9}$$

Centralize the enterprise data center resources mapped in the high-dimensional space, and then solve the covariance matrix of the data center resources of the high-dimensional enterprise:

$$H_J = \frac{1}{n-1} \sum_{i=1}^{n} \phi\left(d_i\right) \phi\left(d_i\right)^T = \frac{1}{n-1} \sum_{i=1}^{n} g_i g_i^T. \tag{10}$$

Eigendecompose H_J to solve the eigenvector matrix:

$$H_J h_r = \varphi_r h_r. \tag{11}$$

But:

$$g_i^T H_J h_r = \varphi_r g_i^T h_r. \tag{12}$$

Substitute Equation (10) and Equation (11) into Equation (12) to get:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_k^r \sum_{k=1}^{n} \left(g_i^T g_j\right) \left(g_i^T g_j\right) = \sum_{i=1}^{n} (n-1) \varphi_r \sum_{k=1}^{n} \gamma_k^r \left(g_i^T g_k\right). \tag{13}$$

Introduce the kernel function:

$$K_{jj} = {g_i}^T g_j. \tag{14}$$

Substitute Equation (14) into Equation (13), and simplify to get:

$$K \gamma^r = \varphi_r \gamma^r. \tag{15}$$

Using the Equation (15) to eigendecompose the kernel function matrix K, γ^r can be obtained, and further eigenvectors can be obtained. Use the obtained feature vector to perform feature transformation on the high-dimensional enterprise data center resource set:

$$g = \eta^T g_i = \eta^T \phi\left(d_i\right) = \sum_{k=1}^{n} \gamma_k K\left(d_k, d_i\right). \tag{16}$$

When non-linear mapping is performed on the resource set in the original enterprise data center, the specific form of the non-linear mapping is not clear. Therefore, by selecting an appropriate kernel function K, performing eigendecomposition, and obtaining the eigenvector γ^r, further, the resource g_i of the enterprise data center after dimensionality reduction can be obtained.

3.2 Extracting the Distribution Characteristics of Enterprise Data Center Resources

After dimensionality reduction and processing of enterprise data center resources, combined with feature space reorganization technology, the enterprise data center resource set structure reorganization is carried out, and the feature quantity of enterprise data center resource distribution is extracted. Calculate all link gain values and use the deep learning method [14] to obtain the clustering similarity feature of enterprise data center resources as:

$$K_L = \frac{\lambda\left(g_i + b\right)}{z}. \tag{17}$$

In Equation (17), λ represents the value result of enterprise data center resources defined by the standard. If the link gain value is $z \geq 1$, the fitness weight coefficient of enterprise data center resources is obtained as a positive number. Modify the clustering parameters of the resource set in each enterprise data center, and obtain the clustering effectiveness evaluation parameter distribution set and index weight of all cluster head nodes. A cluster head is considered as node which collects information's from the cluster sensors and sending information's to the base stations. The cluster heads are the nodes that pretend as a head of cluster. Based on the above analysis, a deep learning model for clustering enterprise data center resource sets is established, and the fuzzy feature distribution of enterprise data center resource sets is obtained, and the constraint programming model of enterprise data center resource sets is obtained as follows:

$$\min(F) = \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij}K_{ij}. \tag{18}$$

For the best cluster centers of all cluster head nodes, through ambiguity detection, the evaluation set and test set of enterprise data center resource set are obtained. After suspending the transmission of enterprise data center resources, the characteristic distribution of the enterprise data center resource set is obtained as follows:

$$D_q = \kappa \times \mu^2 + \upsilon \times \mu. \tag{19}$$

In Equation (19), κ represents the fused clustering feature set of enterprise data center resources, μ represents the pixel brightness value before correction of the fused clustering feature set of the resource set in the enterprise data center, and υ is the regression parameter. According to the extraction results of the distribution characteristics of the resources in the above enterprise data center, the density peak fusion clustering is performed.

3.3 Peak Clustering of Set Density in Enterprise Data Center Resources

On the basis of extracting the distribution characteristics of enterprise data center resources, the density peak clustering is performed on the resource sets of the enterprise data center. According to the grid block distribution of the resource set in the enterprise data center, the density peak feature quantity of the resource set in the enterprise data center is extracted, and the iterative formula of the algorithm for extracting the density peak feature is given as follows:

$$\varpi(\theta) = D_q \times \varpi(\theta - 1). \tag{20}$$

In Equation (20), ϖ represents the choice to adopt the embedded dimension scheduling value. Assuming that θ is the boundary value vector of the resource set in the enterprise data center transmitted by the cluster head node, the kurtosis of the resource set in the enterprise data center is defined as:

$$\theta_{kurt}(v) = \varpi(\theta) + E\left(v_1 + v_2\right) + \vartheta. \tag{21}$$

In Equation (21), v_1 and v_2 represent the boundary feature quantities of the resources in the enterprise data center, respectively, and ϑ is represented as a scalar. Through density peak clustering, the density distribution of resources in enterprise data center can be obtained to satisfy the following formula:

$$S_{xy} = \left\| B_x - B_y \right\|^2. \tag{22}$$

In Equation (22), x and y respectively represent any two nodes in the resource density distribution of enterprise data center resources, and B_x and B_y respectively represent the corresponding density pixel values of the two.

3.4 Realization of Resource Retrieval in Enterprise Data Center

Based on the clustering results of the peak density of resource sets in enterprise data center, a semantic correlation distribution model for resource retrieval in enterprise data center is constructed. Through the semantic registration and weighted vector combination control method [15], the retrieval of enterprise data center resources is realized. Using the joint quantitative feature analysis method, the least squares fitting function for the retrieval of enterprise data center resources is obtained [16]. The least square method finds a regression line or best-fitted line for any data set which is labelled by an equation. The core functions of the least square method reduce the sum of the squared errors. The description is as follows:

$$\mid \rho(x) = \varpi\left(S_{xy}\right) \sigma_{xy}. \tag{23}$$

In Equation (23), $\sigma_{xy} = 1$ indicates that the output of resource retrieval in enterprise data center satisfies convergence, and $\sigma_{xy} = 0$ indicates that the output of resource retrieval in enterprise data center diverges. Therefore, the retrieval of enterprise data center resources is constructed, and the entropy function of the information distribution of enterprise data center resources is as follows:

$$\zeta(x)=\frac{\rho(x)-\vartheta}{k+1}+\left(1-\theta_{\text{tot}}\,(v)\right)\tau(k). \tag{24}$$

In Equation (24), $\tau(k)$ is the characteristic function. According to the entropy distribution, combined with the mean function detection method, the retrieval of resources in the enterprise data center is carried out, and the following results are obtained:

$$v(x)=g(x)(k+1)+e(t)+\xi\zeta. \tag{25}$$

In Equation (25), $e(t)$ is the set of attribute names, and ξ and ζ are the closeness functions of the distribution of enterprise data center resources. According to the above analysis, the retrieval of enterprise data center resources is realized. The algorithm implementation process is shown in Figure 3.

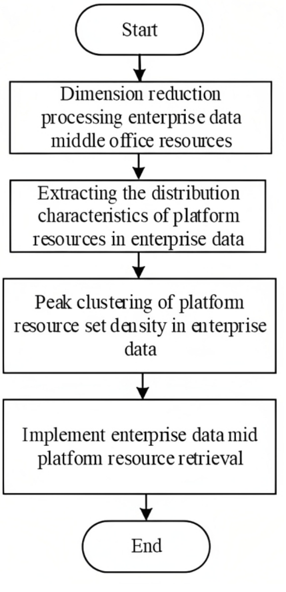


Figure 3. Algorithm implementation flow chart

4 EXPERIMENTAL SIMULATION AND ANALYSIS

4.1 Setting Up the Experimental Environment

In order to verify the effectiveness of the resource retrieval technology in enterprise data center based on the density peak clustering algorithm, the operating system used in the experiment is Windows 10, and the integrated development environment is MATLAB 2014a. The hardware conditions are: CPU Intel (R) Core (TM) i7-7700, main frequency 3.6 GHz, memory 8 GB. Selecting 5 000 MB enterprise data center resources as the experimental data, the proposed algorithm, the algorithm of reference [6] and the algorithm of reference [7] are compared to verify the effectiveness of the proposed algorithm.

4.2 Comparative Analysis of Retrieval Effect of Enterprise Data Center Resources

In order to verify the retrieval effect of the proposed algorithm in enterprise data center resources, the retrieval coverage is taken as the evaluation index. The higher the retrieval coverage, the better the retrieval effect of the algorithm's enterprise data center resources. The algorithm of reference [6], the algorithm of reference [7] and the proposed algorithm are used to compare, and the comparison results of the retrieval coverage of enterprise data center in different algorithms are shown in Figure 4.

Analysis of Figure 4 shows that when the amount of enterprise data center resources is 5 000 MB, the average retrieval coverage of enterprise data center resources of the algorithm of reference [6] is 84.6 %. The average enterprise data center resources retrieval coverage rate of the algorithm of the reference [7] is 79.7 %. The average enterprise data center resources retrieval coverage rate of the proposed algorithm is as high as 98.2 %. From this, it can be seen that the retrieval coverage of enterprise data center resources in the proposed algorithm is relatively high, indicating that the proposed algorithm has a better retrieval effect on the retrieval of enterprise data center resources.

4.3 Comparative Analysis of Retrieval Accuracy of Enterprise Data Center Resources

In order to further verify the retrieval accuracy of the proposed algorithm in enterprise data center resources, the retrieval accuracy rate is used as the evaluation index. The higher the retrieval accuracy is, the higher the retrieval accuracy of the algorithm's enterprise data center resources is. The algorithm of reference [6], the algorithm of reference [7] and the proposed algorithm are used to compare, and the comparison results of the retrieval accuracy of enterprise data center resources of different algorithms are shown in Figure 5.

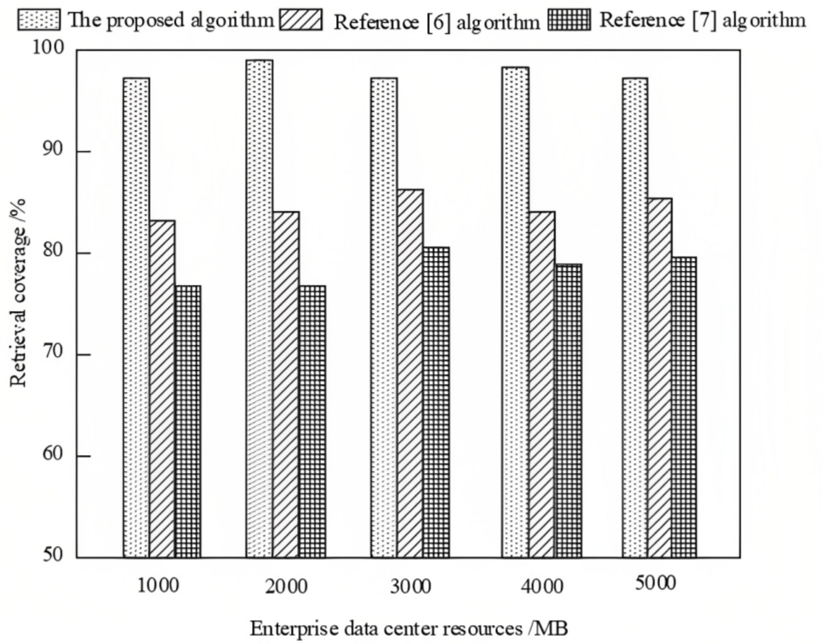


Figure 4. Comparison results of the retrieval coverage ratio of enterprise data center resources of different algorithms

Analysis of Figure 5 shows that when the amount of enterprise data center resources is 5 000 MB, the average retrieval accuracy rate of enterprise data center resources of the algorithm of reference [6] is 89.1 %. The average retrieval accuracy rate of data center resources in the algorithm of reference [7] is 86.4 %. And the average enterprise data center resources retrieval accuracy rate of the proposed algorithm is as high as 98.9 %. It can be seen that the proposed algorithm has a higher accuracy rate of retrieval of enterprise data center resources, indicating that the retrieval accuracy of the proposed algorithm in enterprise data center resources is higher.

4.4 Comparative Analysis of Retrieval Time of Enterprise Data Center Resources

On this basis, verify the retrieval time of the proposed algorithm in enterprise data center resources. The proposed algorithm, the algorithm of reference [6] and the algorithm of reference [7] are used to compare, and the comparison results of the retrieval time of enterprise data center resources of different algorithms are shown in Table 1.

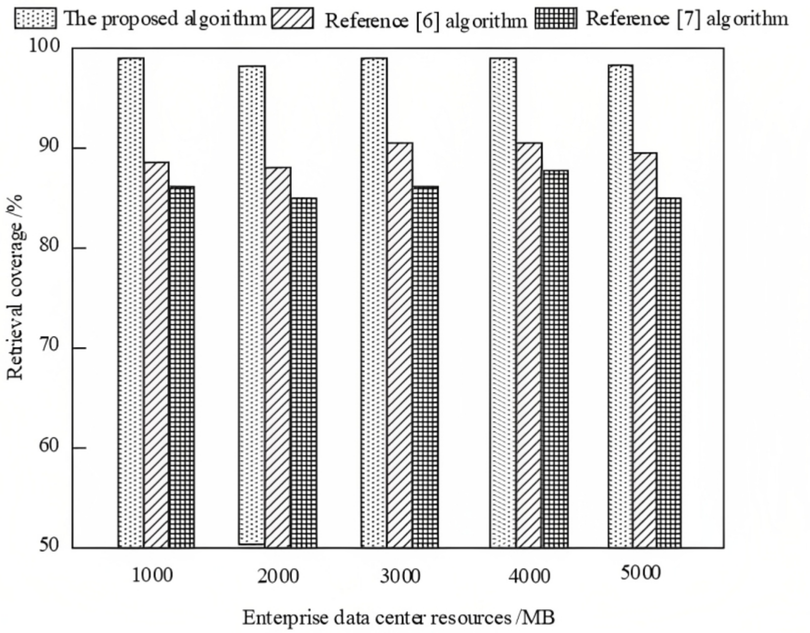


Figure 5. Corresponding author Comparison results of the accuracy rate of resource retrieval in enterprise data center in different algorithms

From the analysis of Table 1, it can be seen that with the increase of the amount of resources in the enterprise data center, the retrieval time of the enterprise data center resources of different algorithms increases accordingly. When the amount of enterprise data center resources is 5 000 MB, the retrieval time of enterprise data center resources in the algorithm of reference [6] is 12.3 s. The retrieval time of enterprise data center resources of the algorithm of reference [7] is 15.8 s. However, the retrieval time of the proposed algorithm in enterprise data center resources is

Enterprise Data Center Resources [MB]	The Proposed Algorithm [s]	The Algorithm of Reference [6] [s]	The Algorithm of Reference [7] [s]
1 000	1.2	4.5	7.2
2 000	2.6	6.1	9.8
3 000	4.1	8.8	11.6
4 000	5.2	10.6	13.2
5 000	6.7	12.3	15.8

Table 1. Comparison results of retrieval time of enterprise data center resources of different algorithms

only 6.7 s. It can be seen that the retrieval time of enterprise data center resources of the proposed algorithm is shorter.

5 CONCLUSION

In this paper, the retrieval technology of enterprise data center resources based on the density peak clustering algorithm is proposed, and the density peak clustering algorithm is used to realize enterprise data center resource retrieval. The algorithm has better retrieval effect of enterprise data center resources, which can effectively improve the retrieval accuracy of enterprise data center resources and shorten the retrieval time of enterprise data center resources. But the time complexity of this algorithm is high. Therefore, in the following research, how to reduce the time complexity of this process is the focus of the research.

Acknowledgments

Science and technology project of the China Southern Power Grid Corporation (Grant No. 031800kk52200001 (gdkjxm20200339)).

REFERENCES

[1] Zhou, A.—Wang, S.—Hsu, C. H.—Kim, M. H.—Wong, K. S.: Virtual Machine Placement with (m, N)-Fault Tolerance in Cloud Data Center. Cluster Computing, Vol. 22, 2019, No. 5, pp. 11619–11631, doi: 10.1007/s10586-017-1426-y.

[2] Mirza, J.—Raza, A.—Atieh, A.—Iqbal, S.—Ghafoor, S.: Self Restorable Intra Data Center Interconnect Based on Multimode Fiber and Free-Space Optics. Optical Engineering, Vol. 60, 2021, No. 3, Art. No. 036113, doi: 10.1117/1.OE.60.3.036113.

[3] Iwagami, M.—Kumazawa, R.—Miyamoto, Y.—Ito, Y.—Ishimaru, M.—Morita, K.—Hamada, S.—Tamiya, N.—Yasunaga, H.: Risk of Cancer in Association with Ranitidine and Nizatidine vs Other H2 Blockers: Analysis of the Japan Medical Data Center Claims Database 2005–2018. Drug Safety, Vol. 44, 2021, No. 3, pp. 361–371, doi: 10.1007/s40264-020-01024-0.

[4] Liu, X.—Zhu, F.—Fu, Y.—Liu, Q.: A Resource Retrieval Method of Multimedia Recommendation System Based on Deep Learning. International Journal of Autonomous and Adaptive Communications Systems, Vol. 13, 2021, No. 4, pp. 400–418, doi: 10.1504/IJAACS.2020.112606.

[5] HajiAminShirazi, S.—Momtazi, S.: Cross-Lingual Embedding for Cross-Lingual Question Retrieval in Low-Resource Community Question Answering. Machine Translation, Vol. 34, 2020, No. 4, pp. 287–303, doi: 10.1007/s10590-020-09257-7.

[6] Li, H.: Internet Tourism Resource Retrieval Using PageRank Search Ranking Algorithm. Complexity, Vol. 2021, 2021, Art. No. 5114802, doi: 10.1155/2021/5114802.

[7] Zhao, G.—Ding, J.: Image Network Teaching Resource Retrieval Algorithm Based on Deep Hash Algorithm. Scientific Programming, Vol. 2021, 2021, Art. No. 9683908, doi: 10.1155/2021/9683908.

[8] Agersted, M. D.—Khodabandeloo, B.—Liu, Y.—Melle, W.—Klevjer, T. A.: Application of an Unsupervised Clustering Algorithm on in Situ Broadband Acoustic Data to Identify Different Mesopelagic Target Types. ICES Journal of Marine Science, Vol. 78, 2021, No. 8, pp. 2907–2921, doi: 10.1093/icesjms/fsab167.

[9] Louhichi, S.—Gzara, M.—Ben-Abdallah, H.: MDCUT2: A Multi-Density Clustering Algorithm with Automatic Detection of Density Variation in Data with Noise. Distributed and Parallel Databases, Vol. 37, 2019, No. 1, pp. 73–99, doi: 10.1007/s10619-018-7253-1.

[10] Gu, Z.—Li, P.—Lang, X.—Yu, Y.—Shen, X.—Cao, M.: A Multi-Granularity Density Peak Clustering Algorithm Based on Variational Mode Decomposition. Chinese Journal of Electronics, Vol. 30, 2021, No. 4, pp. 658–668, doi: 10.1049/cje.2021.03.001.

[11] Parmar, M. D.—Pang, W.—Hao, D.—Jiang, J.—Liupu, W.—Wang, L.—Zhou, Y.: FREDPC: A Feasible Residual Error-Based Density Peak Clustering Algorithm with the Fragment Merging Strategy. IEEE Access, Vol. 7, 2019, pp. 89789–89804, doi: 10.1109/ACCESS.2019.2926579.

[12] Siddique, M.—Samantara, T.—Mishra, S. P.: A Hybrid Prediction Model of Kernel Principal Component Analysis, Support Vector Regression and Teaching Learning Based Optimization Techniques. Current Journal of Applied Science and Technology, Vol. 40, 2021, No. 20, pp. 17–25, doi: 10.9734/cjast/2021/v40i2031460.

[13] Li, X.—Wu, H.—Yang, X.—Xue, P.—Tan, S.: Multiview Machine Vision Research of Fruits Boxes Handling Robot Based on the Improved 2D Kernel Principal Component Analysis Network. Journal of Robotics, Vol. 2021, 2021, Art. No. 3584422, doi: 10.1155/2021/3584422.

[14] Pan, C. S.—Mao, J. L.—Yang, Y.: Active Sonar Target-Echo Recognition Research Based on Deep Learning. Computer Simulation, Vol. 37, 2020, No. 11, pp. 179–183 (in Chinese).

[15] Xue, F.—Lu, W.—Chen, K.—Webster, C. J.: BIM Reconstruction from 3D Point Clouds: A Semantic Registration Approach Based on Multimodal Optimization and Architectural Design Knowledge. Advanced Engineering Informatics, Vol. 42, 2019, Art. No. 100965, doi: 10.1016/j.aei.2019.100965.

[16] Ahmed, N. S.—Ahmed, S. H.: Enhancement RC4 Algorithm Based on Logistic Maps with Multi-Parameters. Journal of Al-Qadisiyah for Computer Science and Mathematics, Vol. 11, 2019, No. 4, pp. 58–67.

Jiaming Jiang is a Senior Engineer. She received her Bachelor of economics from the Guangdong University of Foreign Studies in the Department of E-Commerce in 2010. Currently, she is working in the Guangdong Qingyuan Power Supply Bureau of the China Southern Power Grid. Her research interests include enterprise digital transformation and digital innovation application. She has published around 13 academic articles. Meanwhile, she has participated in 8 scientific research projects.

Guoheng Ruan is a Senior Engineer. He received his Bachelor degree in 2004 from the Sun Yat-Sen University in electronic information science and technology. He received his M.Sc. degree in 2009 from the South China University of Technology in control engineering. Now he is working in the Guangdong Qingyuan Power Supply Bureau of the China Southern Power Grid and his research areas include enterprise digital construction. He has published five academic articles. Meanwhile, he has presided over three scientific research projects.

Zhenggan Dai is working as a Senior Engineer. He received his Bachelor degree from the Hunan Institute of Science and Technology in the Department of Computer Science and Technology, 2008. Currently, he is working in the Guangdong Qingyuan Power Supply Bureau of the China Southern Power Grid. His research areas include enterprise digital transformation. He has published around three academic articles. Meanwhile, he has participated in four scientific research projects.