

Research Article

Big Data Clustering Mining Based on Swarm Intelligence Algorithm in Cloud Environment

Yaning Yan 

School of Intelligent Science and Information Engineering, Xi'an Peihua University, Xi'an 710125, Shaanxi, China

Correspondence should be addressed to Yaning Yan; 150436@peihua.edu.cn

Received 28 March 2022; Revised 18 May 2022; Accepted 26 May 2022; Published 17 June 2022

Academic Editor: Gopal Chaudhary

Copyright © 2022 Yaning Yan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As science and technology progress and develop rapidly in this day and age, various industry applications have changed the data in a new way, and the explosive growth of data has made traditional data mining unable to perform the current data mining work. The aim of this paper is to study the aggregation digging of big data based around flocking automation algorithms under circumstances of cloud. There is a fuzzy-mean aggregation arrangement with mixed frog leaping in this paper. Through simulation analysis of different clustering algorithms, hybrid frog-hopping and as well as the merging arrangement introduced in this paper, it is concluded that the merging algorithm achieves an improvement of up to 90% accuracy on the iritual membrane dataset. It was demonstrated that the arrangement provides effective agglomerations. When the number of iterations is 500, the fitness value of the algorithm on Dataset 1 is 1.59×10^4 , and its convergence speed is faster than the other algorithms.

1. Introduction

Through the accelerated growth of modern science and engineering, the role played by information technology on the Internet in the life of human society has become increasingly significant. Internet technology has permeated all walks of life, facilitating the advent of the epoch of megadata. The era of big data has completed a series of technological changes in concepts, technologies, and applications centered on data. Such extensive and fundamental changes will inevitably lead to changes in human production and communication methods, as well as changes in social management methods and structures. With the continuous development and progress of society, services such as social networks and location-based services increasingly comprise an integral portion of the population's lifestyle. With the rapid development of science and technology, the data generated by these technologies and information services has become a problem that people have to consider. If the data cannot be effectively processed during a relatively brief interval, the accumulation of data will cause unimaginable disasters. As a result, there is an imperative in undertaking the study on big data clustering mining.

Following the accelerated expansion of the global economy and the rapid application of information technology, big data mining technology has emerged as the times require. It is a technology that can efficiently mine useful data from heterogeneous, large-scale, and rapidly arriving data sources. Therefore, in order to better adapt to the scale, persistence, and high speed of big data, it is essential to study crowd mapping of macrodata under a cloud-focused landscape supported by honeycomb intelligent routines. Cloud computing is attached to the network, and through distributed computing and storage, it can realize efficient computing of big data. Figure 1 shows the development process of cloud computing. As a result, it would be of significant real-world implications to explore big data clustering mining in combination with the cloud environment. The swarm intelligence optimization algorithm is a technology that has emerged in recent years. This technology is to imitate the daily behavior of animals to find the best solution for the global search ability. For example, the particle swarm optimization algorithm is one of them. According to the characteristics of swarm intelligence, the research content of this paper is novel, and the research results can be used in big data scenarios such as the Internet

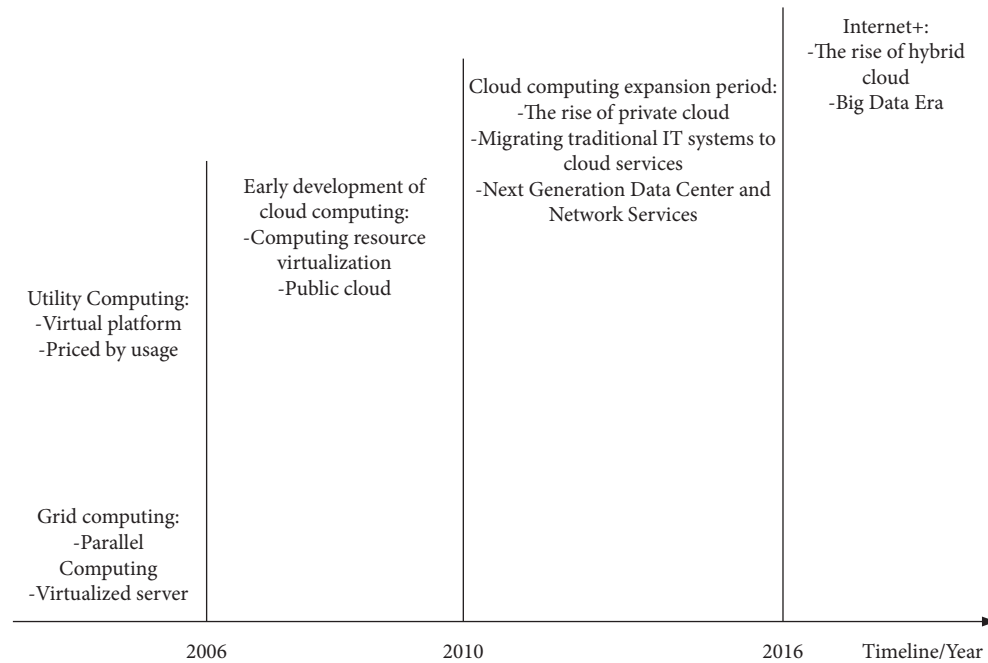


FIGURE 1: The development process of cloud computing.

of Things and e-commerce. It has certain reference value and excellent practicability.

In recent years, data mining technology has become the main demand of various industries such as artificial intelligence and online social networking. Under the social demand, data mining, as an emerging computer science technology, has developed rapidly. It is for this reason that the scholars have not paid much attention to the research of big data mining, but only talked about the advantages of swarm intelligence algorithm in data mining from the side, and lacked specific conceptual research. Therefore, this paper explores the research on big data clustering mining based on swarm intelligence algorithm in cloud environment, so as to further promote the development of swarm intelligence algorithm, and also provide reference for future theoretical research on big data mining.

2. Related Work

In latest years, it has been the case that clinical cloud consultancy study, in particular regarding information on statistical duplication technology and its applications, is constantly evolving. Maintaining data availability and stability of the application system will be expensive when the quantity of duplication is incremented and positioned in multiple locations. Awad et al. scholars proposed two biomimetic algorithms in order to provide a better representation of the collection and positioning of replica media reports from the domain of clouds. These two techniques are many goal granular system organisation and ant colonisation strategy. Simulated results demonstrate that there is excellent duplication of data provided by MOPSO than any available alternatives. In furtherance, MOACO realizes greater overhead costs, reduced density distribution, and

decreased media utilization over the previous versions of the strategy [1]. Fu et al. discussed the efficacy, toxicity, and technical characteristics of IMRT in the treatment of nasopharyngeal carcinoma. Sliding window dynamic CT image-guided IMRT technique was used to treat 31 cases of nasopharyngeal carcinoma with radical radiotherapy, and the irradiation time was 30–33 minutes. At a follow-up of 3 to 18 months, the localized patients had 1-year progression-free survival, long-term transplant-free storage, and overall survivorship values of 93.5%, 87.1%, and 93.5% respectively. The final results suggest that IMRT provides favorable levels of short-term benefits, noticeably minimises emergency radiological interactions, and offers an excellent compromise in cardiovascular outcomes [2], and also provides Parallel problem solving by the simulation of a bird's search for food. Component Particle Swarm Organisation (PSO) is a prototype flocking intelligent technique. Zheng et al. presented a creation computer system consisting of a Chinese traditional folk melody composition model supported by a particulate swarm programme. The empirical outcomes indicated that the module is practicable and useful in the creation procedure of Chinese traditional music [3]. With the passage of time and the evolution of knowledge in science and technology in the last years, the growth in the field of lesson picking founded on certain well-defined standards resulted in facilitating and more productive lesson picking mechanisms. Using an innovating viewpoint, flock intelligent technique has been applied to the intentional selection machinery for lesson picking. Taking English course selection as an example, Zhang et al. combined flock smart algorithms with comprehensive class picking and recommendation algorithms with English class picking intentions to explore the relevant decision-making mechanisms. The experimental results show there is high precision

in the PSO algorithm, which allows individual actions to perform better and helps to establish a class picking mechanism [4]. Hossain et al. examines cloud-based automated testing software capable of performing all aspects of grey-box printing, free-box printing, and integrated component and assembly printing. It also discusses some of the alternative approaches to automation that can be used for software testing in a climate of claw. Some of these methodologies are considered to be more productive and beneficial. One of these frameworks uses the Flock Smart Computing algorithm. Rather, there is a routing process that uses an antihill computing algorithm for total route coverage to reduce time and flocking and optimisation for backwards purposes to ensure backwards consistency [5]. In response to the challenges of early convergence, low recruitment performance and fragmentation of most particulate swarm inversion strategies with velocity terms, Li and Li propose a Stochastic Inertia Power and Performance Parameter Skewing Strategy. Finally, simulation experiments demonstrate the obvious improvement in search speed, confluence precision, and flexibility of the algorithm versus established retrofit solutions. As a result of the features of the IWPSO method, there is superior overall aggregation behavior of the BP neighbourhood managed with IWPSO, making it a highly productive part on the population utilization system [6]. To sum up, the development of swarm intelligence has been deeply explored by many researchers. However, the research on big data clustering mining based flocking smart operators for crowd mapping of aggregated datasets in clinical applications has not generally entered the field of view of scientific researchers. Therefore, more in-depth exploration is required.

3. Theories Related to Big Data Clustering Mining Based on Swarm Intelligence Algorithm in Cloud Environment

3.1. Cloud Computing. The computing power and storage of traditional computer technology are extremely limited. Traditional network storage platforms mainly use centralized storage servers for data storage. This mode makes the storage server a bottleneck of system performance and cannot meet the storage requirements of massive data. At this time, a new computational model was born: cloud computing [7]. Cloud computing needs the support of the network. By integrating storage and computing resources in different places, these resources form a resource pool, and computing tasks can be distributed on the resource pool. Cloud computing is the synthesis and expansion of parallel computing, grid computing, and distributed computing. Distributed computing and distributed storage are the two main cores of cloud computing. Distributed storage technology mainly uses location servers, namely, LBS, and a highly scalable system architecture to achieve load balancing of large-scale storage servers. This technology cannot improve system performance. Hardware resource virtualization technology is a key technology in cloud computing related technologies. Virtualization technology deploys virtual machines with different operating systems on the

same server to execute programs [8, 9]. Cloud computing data management technology can efficiently manage large datasets and how to find specific data in massive data. Bigtable is a classic cloud computing data management technology.

It is one of the most representative system platforms of cloud computing and a distributed system framework. It is made up of many different elements. HDFS (Hadoop Distributed File System) stores all files in Hadoop and is the bottom layer of Hadoop [10]. HDFS can be applied to low-cost equipment, and has the characteristics of strong fault tolerance and large data throughput. Because of this, HDFS has good feasibility and effectiveness in big data mining, as shown in Figure 2.

In the field of cloud computing, Map-Reduce is a well-known marketing strategy, which is a distributed programming model for massive data operations in huge concentrations [11, 12]. As shown in Figure 3, the programming model is relatively simple on the cloud platform, and can be directly displayed to the user, so that the user can prepare straightforward simultaneous routines employing the production code pattern to implement certain aspects of performance. Through the production code pattern, users can happily enjoy the services brought by the cloud [13]. Cloud computing can not only realize high-speed computing of big data but also quickly store big data. The storage and computing capabilities of distributed clusters enable comprehensive and efficient mining of big data. Cloud computing can fulfill requirements for two aspects of big data mining. First, as extensive computation is required for large database operations, it has high requirements for computing power [14].

Cloud can be divided into three categories according to the service mode: Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS). As shown in Figure 4, PaaS is a type of cloud service model that does not require users to download or install the operating system and related services through the Internet, but needs to manage the operating environment of the deployed applications and configure the application hosting environment, such as Google's App Engine and Microsoft's Web App Services. In an IaaS platform, users do not need to manage and control the underlying physical computing resources or other infrastructure, but can fully control the operating system, storage devices, and deployed applications. For example, Amazon's Elastic Cloud EC2 [15, 16]. SaaS is a service in which applications deployed by service providers on cloud platforms are directly provided to users. SaaS needs to be able to directly access and use the services provided by the software from the cloud platform through a client such as a browser, and users do not need to pay attention to the control and management of the software, including servers and storage. The professional SaaS product development and service team of Maker Artisan provides SaaS tool products including: education SAAS, live broadcast SAAS, corporate training SAAS, and content SAAS platform, providing comprehensive system services for enterprises [17].

Parallel computing refers to a computing mode in which computing tasks are distributed to several processors that

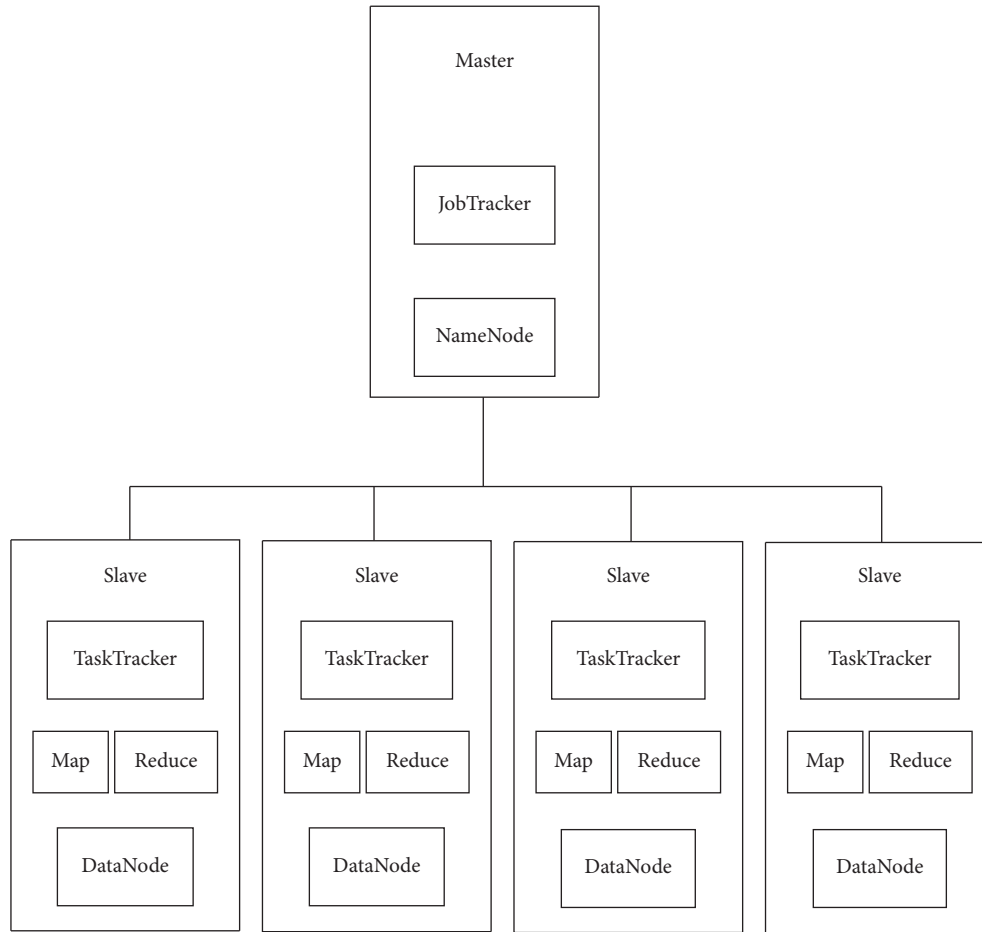


FIGURE 2: Hadoop system deployment framework.

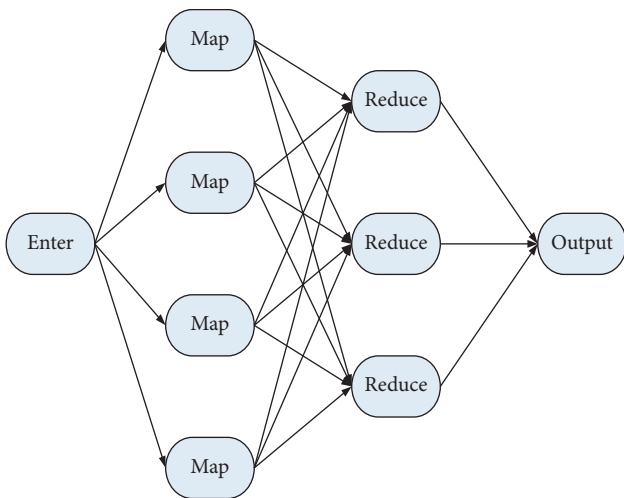


FIGURE 3: Map-reduce execution process diagram.

share the same memory. The architecture of a parallel computing system is usually characterized by the homogeneity of components; each processor is of the same type and has the same processing performance. Shared memory has a separate address space that is accessible by all processors. Parallel programs are divided into execution units

and assigned to different processors, which communicate with each other by means of shared memory. Initially, only architectures with multiple processors sharing the same physical memory could be called parallel systems. Over time, these constraints have been relaxed, and as long as an architecture is based on the concept of shared memory, whether a physical memory system, or a system consisting of libraries, specific hardware, and an efficient network infrastructure, it can be called a parallel system. For example, a cluster where nodes are connected by an unlimited bandwidth network and configured with a distributed shared memory system can be called a parallel system [18].

3.2. Data Mining. Data mining is the process of acquiring knowledge and information from incomplete, large-scale, random datasets. As shown in Figure 5, cluster analysis is one of the tasks of data mining [19]. Clustering analysis can construct a macroscopic data concept. Through data clustering, the relevant distribution state of the data can be directly given, and the association between data attributes can be found according to the data category [20].

Data mining is a technology that finds its laws from a large amount of data by analyzing each data. It mainly includes three steps: data preparation, law search, and law expression. Data preparation is to select the required data

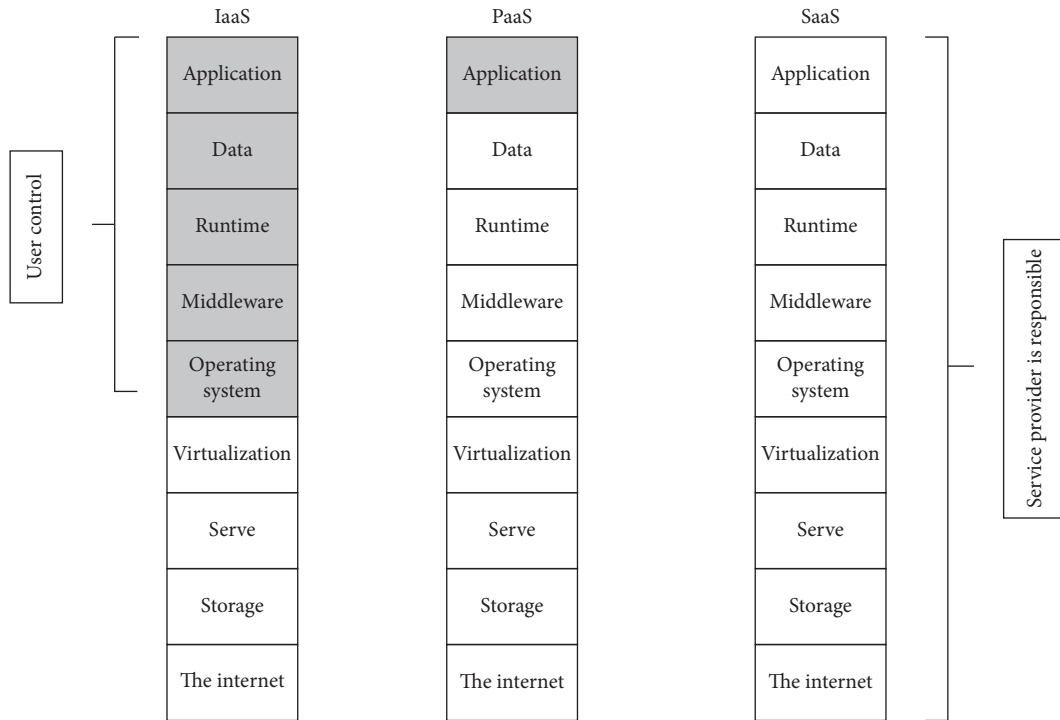


FIGURE 4: Comparative example of three cloud service models.

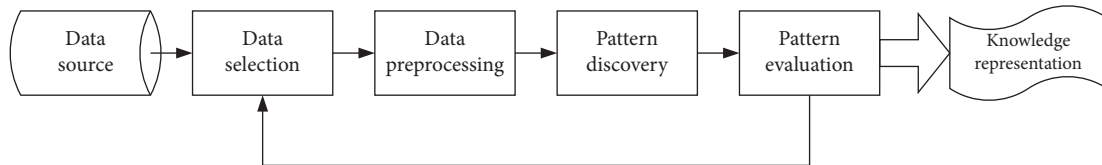


FIGURE 5: Data mining process.

from relevant data sources and integrate it into a dataset for data mining; regular search is to find out the regularity contained in the dataset by a certain method; regular expression is to be as user-readable as possible. The way of understanding (such as visualization) will represent the found patterns. The tasks of data mining include association analysis, cluster analysis, classification analysis, anomaly analysis, specific group analysis, and evolution analysis [21].

One of the most common uses of Hadoop is web search. While it is not the only software framework application, it stands out as a parallel data processing engine. One of the most interesting aspects of Hadoop is the Map and Reduce process, which was inspired by Google’s development. This process, called indexing, takes as input the textual web pages retrieved by the web crawler and reports the frequency of words on those pages as a result. This result can then be used throughout the web search process to identify content from the defined search parameters [22].

The steps of the data mining process model mainly include defining the problem, establishing the data mining library, analyzing the data, preparing the data, establishing the model, evaluating the model, and implementing.

There are many uncertain data in the dataset, which will interfere with the mining of the target data. Therefore, in the

process of data selection, all data related to the target must be obtained, which greatly increases the efficiency of data mining. The dataset has the characteristics of large scale, so the appropriate processing method should be selected during data preprocessing. Pattern evaluation can get the desired knowledge pattern. The final knowledge representation is to present the obtained knowledge pattern to the user in a clear way with the help of related technologies. Following the evolution of modern information technology, the emergence of giant statistics has been driven by the dramatic spread of applications. Giant statistics contains huge information, which can create huge value and play an important role in the development of society and enterprises. Similarly, due to the huge amount of data, big data needs to be processed at the appropriate time, otherwise new progress cannot be led [23]. The emergence of the era of giant statistics also brings huge challenges to data mining. For example, when processing voluminous datasets, the mining technology must complete the task efficiently under the condition of limited storage space, and when dealing with high-speed data flow, the mining technology must not analyze the data in a limited time. It is also necessary to keep the memory usage small. So as to maximize the benefits, there are also privacy issues that are also one of the

challenges faced by data mining, and there is no way to avoid this problem. Mining technology should accomplish its mission while ensuring data privacy and security.

Rough set method, also known as rough set theory, is a new mathematical tool for dealing with ambiguous, imprecise, and incomplete problems. The advantage is that the algorithm is simple, and no prior knowledge about the data is required in the processing process, and the inherent law of the problem can be automatically found; the disadvantage is that it is difficult to directly deal with the continuous attributes, and the attributes must be discretized first. Therefore, the discretization of continuous attributes is the difficulty that restricts the practical application of rough set theory. Rough set theory is mainly used in approximate reasoning, digital logic analysis and simplification, and establishment of predictive models.

3.3. Cluster Analysis. In the cloud environment, in order to divide the data well, cluster analysis can set up different groups, group similar data into one group, and group unrelated data or data with small correlation into one group [24]. Therefore, the introduction of the concept of "distance" can know the correlation between the data, that is, the similarity of the data, and at the same time, it can increase the efficiency of sample classification. The formula for calculating the absolute distance is shown in the formula.

$$d(X, Y) = \sum_{i=1}^p |x_i - y_i|. \quad (1)$$

Among them, let $X = \{x_1, x_2, \dots, x_p\}$ and $Y = \{y_1, y_2, \dots, y_p\}$ be the sample data of two dimensions.

$$q_{ij} = \frac{\sum_{k=1}^p x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^p x_{ik}^2 \sum_{k=1}^p x_{jk}^2}}, \quad (2)$$

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

where \bar{x}_i and \bar{x}_j are the mean values, and n represents the dimension of the sample space.

3.4. Fuzzy-Mean Clustering Algorithm. Calculate the similarity coefficient between samples or variables to establish fuzzy similarity matrix; use fuzzy operations to perform a series of synthetic transformations on the similarity matrix to generate a fuzzy equivalent matrix; finally, according to the different interception levels λ , the fuzzy equivalent matrix is intercepted and classified.

Clustering is to divide a dataset into different categories according to certain criteria (such as distance criteria), so that the data in the same category can be better correlated, and the data that are not in a category has a large difference. Fuzzy clustering methods are aimed at things that are ambiguous and inaccurate. The representation of the initial data matrix of this clustering method is shown in the calculation formula.

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}. \quad (3)$$

The data objects are

$$U = \{x_1, x_2, \dots, x_n\}, \quad (4)$$

where $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, $i = 1, 2, \dots, n$.

Fuzzy-mean clustering is a hard clustering algorithm, which deals with data mining problems by optimizing the objective function to the minimum value. The calculation formula of this algorithm is shown in the formula.

$$X_1 \cup X_2 \cup \cdots \cup X_C = X, \quad (5)$$

$$X_i \cup X_k = \emptyset, 1 \leq i \neq k \leq C,$$

$$X_i \neq \emptyset, X_i \neq X, 1 \leq i \leq C,$$

where $X = \{x_1, x_2, \dots, x_n\} \subset R^s$ is a specific dataset.

In the formula, u_{ij} means that x_j belongs to the membership degree of subset X_i , then the hard C grouping of dataset X is

$$M_{fc} = \left\{ U \in R^s \mid u_{ki} \in \{0, 1\}; \sum_{i=1}^C u_{ki} = 1, \forall k; 0 < \sum_{k=1}^n u_{ki} < n, \forall i; \right\}. \quad (6)$$

The constraints are

$$\sum_{i=1}^C u_{ij} = 1, 1 \leq j \leq n,$$

$$0 \leq u_{ij} \leq 1, 1 \leq i \leq c, 1 \leq j \leq n, \quad (7)$$

$$0 \leq \sum_{j=1}^n u_{ij} \leq n, 1 \leq i \leq c,$$

where: c : indicates the number of categories of clusters; s : indicates the number of attribute values in the sample.

The distance from x_j to v_i is

$$d_{ij} = \|x_i - y_j\|, \quad (8)$$

where: x_j : indicates the j sample data; v_i : indicates the i cluster center.

Then, v_i and membership u_{ij} are

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m \cdot x_j}{\sum_{j=1}^n u_{ij}^m}, j = 1, \dots, c, \quad (9)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij}/d_{ik})^{2/m-1}}, i = 1, \dots, n; j = 1, \dots, c.$$

3.5. Algorithm Fusion. Hybrid leapfrog algorithm is a combinatorial optimization by imitating the behavior of frogs catching food. As shown in Figure 6, this algorithm is a swarm

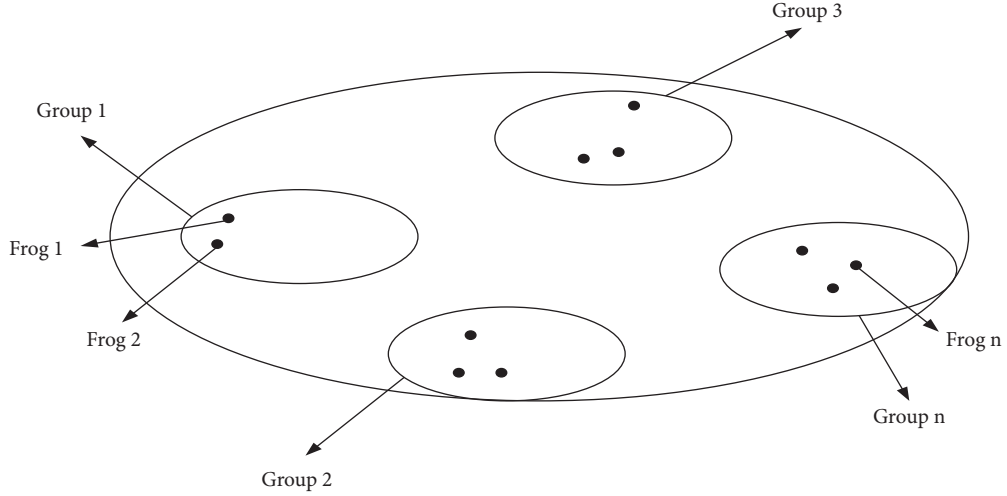


FIGURE 6: Schematic diagram of the hybrid frog leaping algorithm.

intelligence algorithm, which is usually used for local search and has a powerful global information search function. In this paper, the fuzzy-mean clustering algorithm is easy to fall into local minimum points, and the algorithm is combined with the hybrid frog leaping algorithm to achieve the enhancement of clustering effect and optimization accuracy. The calculation formula of the hybrid leapfrog algorithm is shown in the formula. where: x^b : indicates the best fitness of the subpopulation under each population; x^w : represents the worst fitness of the subpopulations under each population; D_{\max} : represents the farthest length of the frog jumping; D_i represents the length of each jump of the frog.

$$D_i = \text{rand}() \cdot (x^b - x^w), i = 1, 2, \dots, m; 0 < \text{rand}() < 1, \quad (10)$$

$$x^w = x^w + D_i, -D_{\max} \leq D_i \leq D_{\max}, i = 1, 2, \dots, m.$$

The flowchart of the fuzzy-means clustering algorithm based on hybrid frog leap is shown in Figure 7. Among them, in the behavior selection, the fitness value of all frogs needs to be obtained, and different combinations are divided and arranged in descending order. The calculation formula is shown in the formula, which is (11), that is, the objective function of fuzzy-mean clustering.

$$f = \frac{q}{J_m + 1}, 0 \leq q \leq 1. \quad (11)$$

3.6. *PSO – FCM Algorithm.* PSO – FCM algorithm is a clustering algorithm constructed according to the membership matrix between data.

$$\sum_{i=1}^p \mu_{ij} = 1, \mu_{ij} \in [0, 1],$$

$$\mu_{ij} = \begin{cases} \frac{1}{\sum_{k=1}^p (d_{ij}/d_{ik})^{2/m-1}}, & d_{kj} \neq 0, \\ 0, & d_{kj} = 0. \end{cases} \quad (12)$$

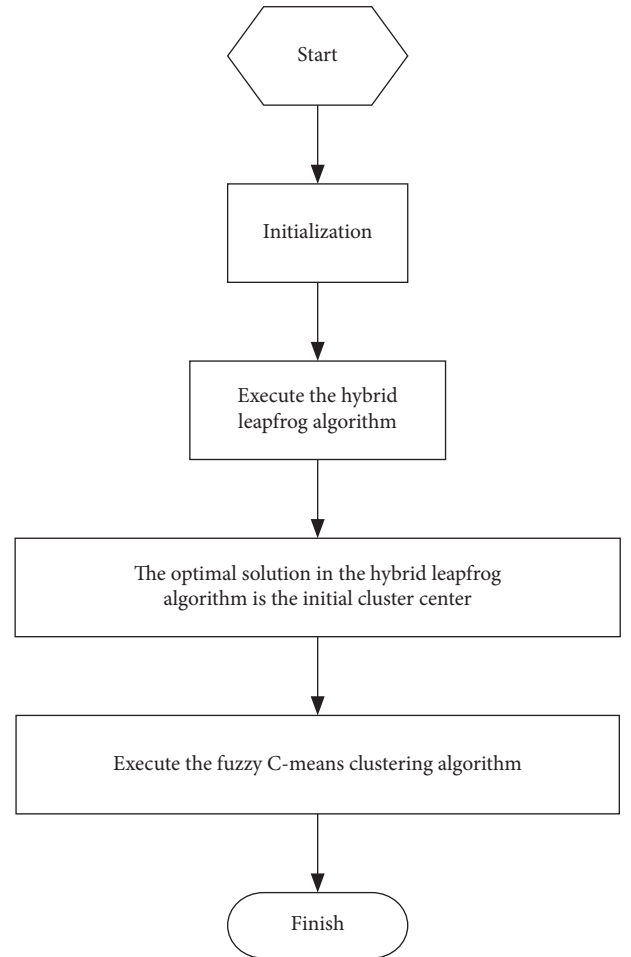


FIGURE 7: Flow of the fusion algorithm.

The fitness function value derived from all sample membership values:

$$J_m = \sum_{j=1}^n \sum_{i=1}^p (\mu_{ij})^m (x_j - c_i)^2. \quad (13)$$

where: m : indicates weighted index, $m > 1$; $x_j - v_i$: represents the clustering of sample x_j to i sample centers.

$$c_i = \frac{\sum_{j=1}^n (\mu_{ij})^m x_j}{\sum_{j=1}^n (\mu_{ij})^m},$$

$$v_i^{k+1} = \begin{cases} \omega v_i^k + c_1 r_1 (\rho_i - x_i^k) + c_2 r_2 (\sigma_i - x_i^k), & k > 1 \\ c_2 r_2 (\sigma_i - x_i^k), & k = 1 \end{cases},$$

$$v_i^{k+1} = x_i^k + v_i^{k+1}. \quad (14)$$

Among them, ω is the inertia weight, and σ_i is the best fitness function of the population.

4. Big Data Clustering Mining Simulation Experiment

In order to further verify the effect of the fusion algorithm proposed in this paper, this paper conducts simulation experiments on the four algorithms introduced in this paper. The experimental data comes from datasets and artificial datasets, as shown in Table 1. The first is the validity analysis of clustering, and the second is the comparison of the convergence rate.

4.1. Analysis of the Effectiveness of Clustering. The parameter settings in the experimental process of this paper are shown in Table 2. The article will use the correct rate to represent the clustering effect of the four algorithms. The calculation formula of the classification correct rate is shown in formula (15). All four different algorithms are repeated 50 times, and the average of all metrics is taken.

$$T = \frac{M}{N} \times 100\%, \quad (15)$$

where: M : represents the correct number of sample clusters; N : represents the total number of data objects in the dataset.

It can be found from Figure 8 that the correct number of sample clusters between PSO-FCM and the fusion algorithm proposed in this paper is not much different. It can be seen that the fusion algorithm proposed in this paper has better performance.

In the same dataset, it can be seen from Table 3 that compared with the other three algorithms, the clustering effect of hybrid frog leaping is obviously poor, and the correct rate of fuzzy C-means clustering has reached 73%, while the hybrid frog leaping algorithm has a significantly poorer clustering effect. The correct rate is only 68%.

It can be seen from Table 4 that on the Iris dataset, the fusion algorithm proposed in this paper has the best sample clustering correct number and correct rate, which are 125 and 90%, respectively. This shows that the algorithm has

TABLE 1: Parameters of the four experimental datasets.

Dataset	Number of categories	Number of samples	Number of attributes
Dataset 1	4	300	2
Dataset 2	7	400	5
Dataset 3	2	200	4
Iris	3	160	3

TABLE 2: Parameter settings of the experimental process.

Serial number	Description	Parameter
1	The maximum number of iterations	400
2	Population size	20
3	Number of subclusters	2
4	Number of subgroups	4
5	The number of iterations within the group	4
6	Fuzzy index	$m = 2$

high accuracy and its performance is better than the other three algorithms.

As shown in Table 5, the clustering results of each algorithm are compared.

4.2. Comparison of Convergence Speeds. In this paper, the convergence speed of the four algorithms is analyzed experimentally on the three artificial datasets. The four algorithms are: the first is the hybrid leapfrog algorithm, the second is the fuzzy C-means clustering, the third is the PSO-FCM algorithm, and the fourth is the fusion algorithm proposed in this paper. Figure 9

It can be seen from Figure 10 that on the Dataset 1, the fusion algorithm proposed in this paper has the fastest convergence speed. When the number of iterations is 500, the fitness value of the fusion algorithm is 1.59×104 .

Through the comparison of the convergence speed, it can be seen from Figure 11 that when the number of iterations is 0, the fitness values of PSO-FCM and the fusion algorithm proposed in this paper are 2.6×103 and 2.5×103 , respectively, and when the number of iterations is 400, the fitness values are 2.3×103 and 2.1×103 , respectively. This shows that the test results of the two algorithms are relatively similar. However, the convergence speed of the algorithm in this paper is still better than that of PSO-FCM.

It is not difficult to see from Figure 11 that the convergence speed of hybrid leapfrog and fuzzy C-means clustering is not fast. When the number of iterations is 100, the fitness values of the hybrid leapfrog and fuzzy C-means clustering algorithms are 3.9×103 and 4.7×103 , respectively.

5. Discussion

Through the simulation analysis of the hybrid frog leaping algorithm, fuzzy C-means clumping, PSO-FCM

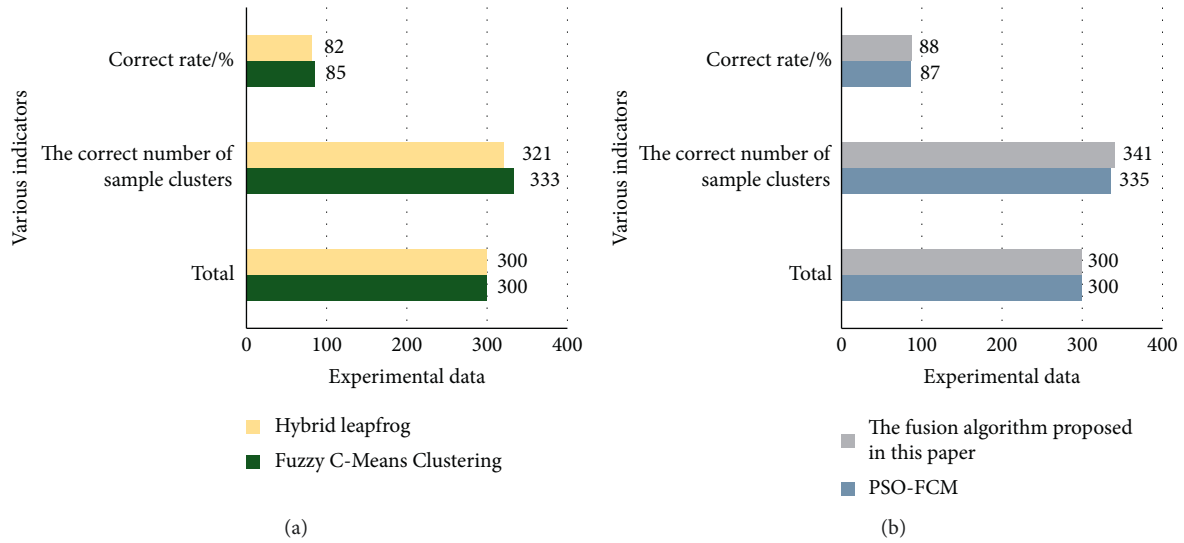


FIGURE 8: Experimental results on dataset 1. (a): The experimental effect of hybrid leapfrog and fuzzy C-means clustering on the Dataset 2; (b): Experimental results of PSO-FCM and the fusion algorithm proposed in this paper on the Dataset 2.

TABLE 3: Experimental data on dataset 3.

Algorithm	Total	The correct number of sample clusters	Correct rate (%)
Fuzzy C-means clustering	200	137	73
Hybrid leapfrog	200	127	68
PSO-FCM	200	145	77
The fusion algorithm proposed in this paper	200	150	78

TABLE 4: Experimental data on the iris dataset.

Algorithm	Total	The correct number of sample clusters	Correct rate (%)
Fuzzy C-means clustering	160	120	86
Hybrid leapfrog	160	117	84
PSO-FCM	160	122	88
The fusion algorithm proposed in this paper	160	125	90

TABLE 5: Experimental data on the iris dataset.

Dataset	Algorithm	Clustering accuracy	Number of iterations
Gla	Fuzzy C-means clustering	25.69	62
	Hybrid leapfrog	73.41	96
	PSO-FCM	89.01	30
	The fusion algorithm proposed in this paper	93.78	55

methodology, and merging mechanism proposed by the present study, it is possible to conclude as follows:

Through the analysis of the clustering effect of the four algorithms, it is verified that the clustering effectiveness of the merging method proposed in this paper is the best. The experimental data show that the clustering effect of the hybrid leapfrog and fuzzy C-means clumping method is not good on the same dataset. On the Dataset 2, the correct rates of hybrid leapfrog and fuzzy C-means clumping are 62% and 65%, respectively. It is also worse than PSO-FCM and the merging approach submitted in the current text on the Iris dataset. Compared with the correct rate of 90% of the

approach submitted in the present text, the correct rate of hybrid frog leap is only 84%.

Through the experimental analysis of the convergence speed of the four algorithms on the three artificial datasets, the data show that when the number of iterations is 300, the fitness value on the Dataset 3. The merging approach submitted in the present text is 2.5×10^3 , which is better than 3×10^3 of PSO-FCM. At the same time, the results show that the convergence speed of the fuzzy C-means clumping method is lower than that of the hybrid leapfrog. This proves that the method in this paper is better than the other three methods in terms of convergence speed.

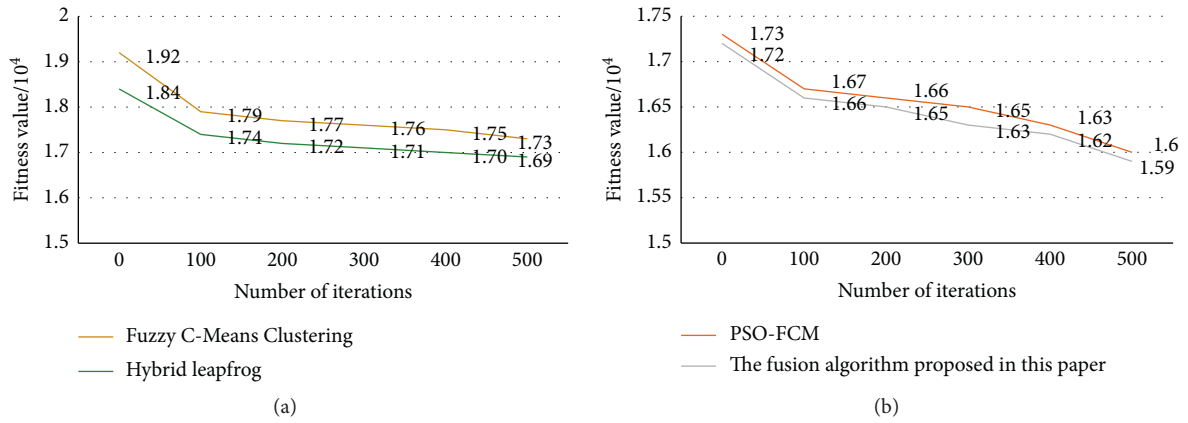


FIGURE 9: Test results on dataset 1. (a): Test results of hybrid leapfrog and fuzzy C-means clustering on Dataset 1; (b): Test results of PSO-FCM and the fusion algorithm proposed in this paper on Dataset 1.

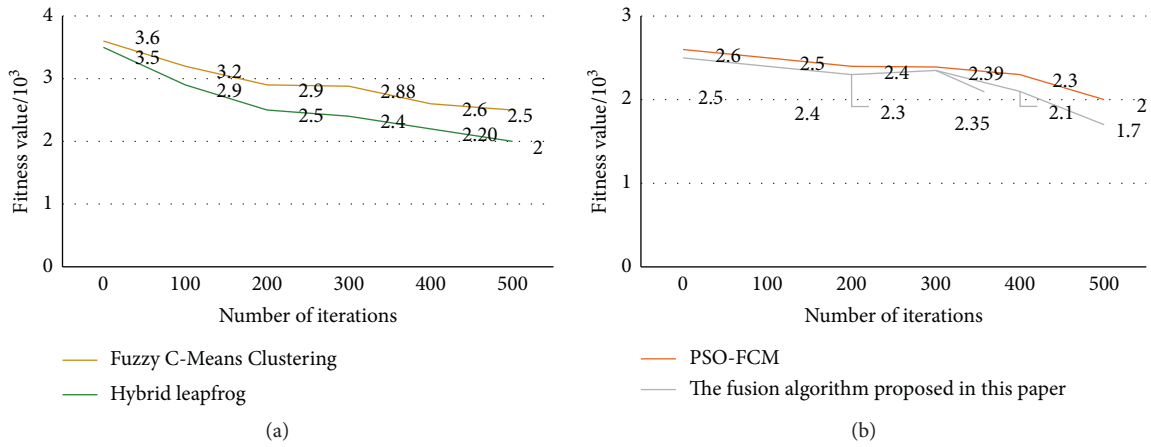


FIGURE 10: Test results on dataset 2. (a): Test results of hybrid leapfrog and fuzzy C-means clustering on Dataset 2; (b): Test results of PSO-FCM and the fusion algorithm proposed in this paper on Dataset 2.

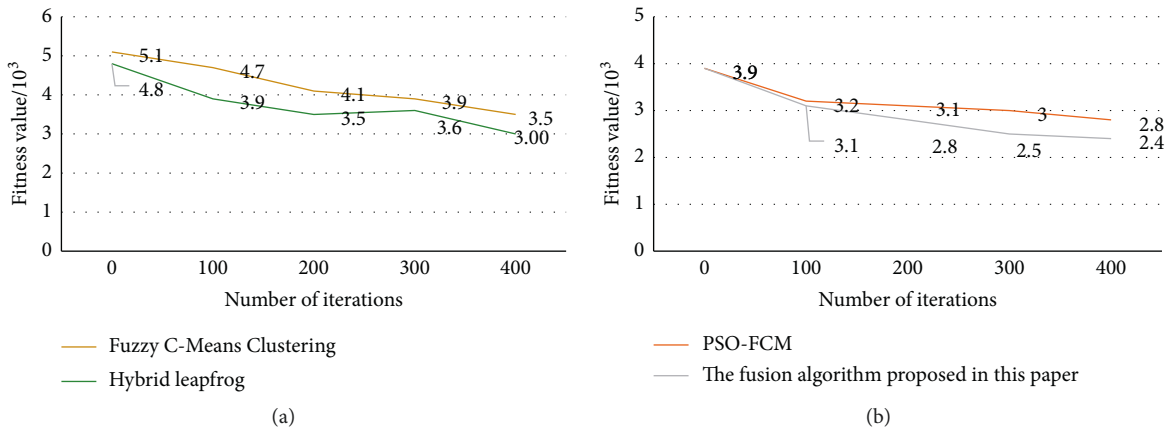


FIGURE 11: Test results on dataset 3. (a): Test results of hybrid leapfrog and fuzzy C-means clustering on Dataset 3; (b): Test results of PSO-FCM and the fusion algorithm proposed in this paper on the Dataset 3.

The whole comparative test data shows that the approach submitted in the present text not only has the characteristics of fast convergence speed and good clustering effectiveness but also outperforms other methods in accuracy. It can be seen that the swarm intelligence optimization method is excellent and effective.

6. Conclusion

This paper introduces different clustering methods and expounds clustering analysis, which is a tool with high usage rate of data mining in cloud computing. Therefore, this paper proposes a fusion of fuzzy-mean clustering method and swarm intelligence algorithm-hybrid leapfrog method, and completes the research on big data clustering mining based on swarm intelligence algorithm in cloud environment. Through the experimental analysis of the clustering effectiveness and convergence speed of the merging method proposed in this paper and the hybrid frog leaping algorithm, fuzzy C-means clustering, and PSO-FCM method, it is concluded that the merging algorithm proposed in this paper enhances the accuracy of the fuzzy C-means clustering method. The merging method can avoid the local obstacle problem of the fuzzy C-means clustering method, and has excellent clustering effectiveness and accuracy. The merging algorithm proposed in this paper can select the fitness function well, promote the improvement of the global search ability, and achieve good results in the clustering effect, which is similar to the PSO-FCM method. But the fusion method can get the cluster centers faster. Compared with the other algorithms, the merging method proposed in this paper is superior to the other three methods in terms of clustering effect, convergence speed, and accuracy. Big data clustering mining is very complex and involves a wide range of areas. Due to my limited time and energy, and the limitation of resources, there are some shortcomings in this paper, such as the refinement and expansion of the swarm intelligence method. From the perspective of some indicators, the clustering method incorporating swarm intelligence has more obvious advantages than the traditional method. However, some shortcomings are inevitable. Further theoretical analysis, application field expansion, model improvement, and other issues are future research focus and difficulty.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the special research project on the ideological and political teaching reform of Xi'an Peihua University in 2021. Project name: Reform and practice of ideological and political teaching based on "web front end development" course. Project No: PHKCSZ202130.

References

- [1] A. Awad, R. Salem, R. Salem, H. Abdelkader, and M. Salam, "A swarm intelligence-based approach for dynamic data replication in a cloud environment," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 2, pp. 271–284, 2021.
- [2] X. Fu, M. Li, M. Yin, and Q. Y. Li, "IMRT images based on swarm intelligence algorithm in the treatment of nasopharyngeal carcinoma and pain," *Scientific Programming*, vol. 2021, no. 39, pp. 1–7, Article ID 7007169, 2021.
- [3] X. Zheng, Y. Wang, L. Shen, W. Y. Guo, and D. L. Li, "Algorithm composition of Chinese folk music based on swarm intelligence," *International Journal of Computing Science and Mathematics*, vol. 8, no. 5, pp. 437–446, 2017.
- [4] D. Zhang and D. Wang, "Study on the intentional choice mechanism of course selection based on swarm intelligence algorithm," *Scientific Programming*, vol. 2021, no. 1, pp. 1–6, Article ID 5510039, 2021.
- [5] M. Hossain, S. Abufardeh, and S. Kumar, "Frameworks for performing on cloud automated software testing using swarm intelligence algorithm: brief survey," *Advances in Science, Technology and Engineering Systems Journal*, vol. 3, no. 2, pp. 252–256, 2018.
- [6] Q. Li and S. Li, "Optimization of artificial CNN based on swarm intelligence algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 4, pp. 6163–6173, 2021.
- [7] Z. Xia, X. Wang, L. Zhang, and Z. X. K. Qin, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594–2608, 2016.
- [8] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Lia, "Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1171–1181, 2017.
- [9] Z. Xu, "Computational intelligence based sustainable computing with classification model for big data visualization on map reduce environment," *Discover Internet of Things*, vol. 2, no. 1, p. 2, 2022.
- [10] T. Hirai, H. Masuyama, H. Masuyama, S. Kasahara, and Y. Takahashi, "Performance analysis of large-scale parallel-distributed processing with backup tasks for cloud computing," *Journal of Industrial and Management Optimization*, vol. 10, no. 1, pp. 113–129, 2014.
- [11] M. H. Ghahramani, M. Zhou, and C. T. Hon, "Toward cloud computing QoS architecture: analysis of cloud systems and cloud services," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 1, pp. 6–18, 2017.
- [12] W. Wei, X. Fan, H. Song, X. Fan, and J. Yang, "Imperfect information dynamic stackelberg game based resource allocation using hidden markov for cloud computing," *IEEE Transactions on Services Computing*, vol. 11, no. 1, pp. 78–89, 2018.
- [13] A. F. Barsoum and M. A. Hasan, "Provable multicopy dynamic data possession in cloud computing systems," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 485–497, 2015.
- [14] J.-L. Tsai and N.-W. Lo, "A privacy-aware authentication scheme for distributed mobile cloud computing services," *IEEE Systems Journal*, vol. 9, no. 3, pp. 805–815, 2015.
- [15] P. Li, J. Li, Z. Huang, and T. C. Li, "Multi-key privacy-preserving deep learning in cloud computing," *Future Generation Computer Systems*, vol. 74, no. SEP, pp. 76–85, 2017.

- [16] Z. Xu, G. Zhu, N. Metawa, and Q. Zhou, "Machine learning based customer meta-combination brand equity analysis for marketing behavior evaluation," *Information Processing & Management*, vol. 59, no. 1, Article ID 102800, 2022.
- [17] S. Rajendran, O. I. Khalaf, Y. Alotaibi, and S. Alghamdi, "MapReduce-based big data classification model using feature subset selection and hyperparameter tuned deep belief network," *Scientific Reports*, vol. 11, no. 1, Article ID 24138, 2021.
- [18] K. Sharma, A. Shankar, P. Singh, fnm au, A. Shankar, and P. Singh, "Information security assessment in big data environment using fuzzy logic," *Journal of Cybersecurity and Information Management*, vol. 5, no. 1, pp. 29–42, 2021.
- [19] H. L. Li, Y. Liang, and S. C. Wang, "Review on dynamic time warping in time series data mining," *Control and Decision*, vol. 33, no. 8, pp. 1345–1353, 2018.
- [20] X. Wang, T. Zhou, X. Wang, and Y. Fang, "Harshness-aware sentiment mining framework for product review," *Expert Systems with Applications*, vol. 187, Article ID 115887, 2022.
- [21] S. Zhang and C. Duan, "Clustering optimization algorithm for data mining based on artificial intelligence neural network," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–16, Article ID 1304951, 2022.
- [22] X. Yin, "Construction of student information management system based on data mining and clustering algorithm," *Complexity*, vol. 2021, pp. 1–11, Article ID 4447045, 2021.
- [23] H. Hong, P. Tsangaratos, I. Ilia, and J. A.-X. W. Liu, "Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China," *The Science of the Total Environment*, vol. 625, no. JUN.1, pp. 575–588, 2018.
- [24] A. Hassani, A. Iranmanesh, and N. Mansouri, "Text mining using nonnegative matrix factorization and latent semantic analysis," *Neural Computing & Applications*, vol. 33, no. 20, pp. 13745–13766, 2021.