

Research Article

An Intelligent Cloud Computing Data Processing System for College Innovation and Entrepreneurship Data Statistics

Hongjun Jia ¹ and Wenjing Chen²

¹College of Innovation and Entrepreneurship, Yellow River Conservancy Technical Institute, Kaifeng, 475004 Henan, China

²College of Information Engineering, Yellow River Conservancy Technical Institute, Kaifeng, 475004 Henan, China

Correspondence should be addressed to Hongjun Jia; 2007810480@yrcti.edu.cn

Received 12 March 2022; Revised 4 April 2022; Accepted 21 April 2022; Published 25 May 2022

Academic Editor: Liming Chen

Copyright © 2022 Hongjun Jia and Wenjing Chen. 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.

With the improvement of the enrollment rate in recent years, innovation and entrepreneurship in colleges and universities have become the focus of the country and people. However, the statistics of innovation and entrepreneurship in colleges and universities have become more and more complicated and difficult. The purpose of this paper is to study how to face the statistics of innovation and entrepreneurship in colleges and universities, then to study the intelligent cloud computing data processing system, and put forward the importance of innovation and entrepreneurship and cloud computing in colleges and universities. The experiment results show that it can be seen that the number of graduates has risen sharply from 2011 to 2020 and the percentage has risen from about 20% to about 90%, which has also led to a great increase in the difficulty of graduate entrepreneurial data management. The difficulty of graduate entrepreneurial data management has risen from around 18% at the beginning to around 87% in 2020. Therefore, it is necessary to study the data processing system to make data processing more efficient. The intelligent cloud computing data processing system significantly improves the quality and efficiency of entrepreneurial service management, and the data mining application in the system can also provide data support for the analysis and prediction of graduates' entrepreneurial situations. The intelligent cloud computing data processing system significantly improves the quality and efficiency of entrepreneurial service management, and the data mining application in the system can also provide data support for the analysis and prediction of graduates' entrepreneurial situations.

1. Introduction

With the rapid expansion of higher education in China in recent years, the number of students has also increased significantly, which has brought great pressure to graduates with entrepreneurial spirit, and also to the staff who manage their entrepreneurship. Faced with this pressure, more and more scholars hope to use the data processing system to improve the management of graduate information, so as to better provide guidance and help for graduates to start their own businesses. Cloud computing has 10 trillion operations per second. With such powerful computing power, it is possible to simulate nuclear explosions, predict climate change, and market trends. In the future, whether it is intelligent driving, emotional companion robots, and many other arti-

ificial intelligence products, it is inseparable from the support of cloud computing and even edge computing.

Entrepreneurship of college graduates has always been an important issue in the field of education. The state pays great attention and accordingly puts forward a series of policies for the graduate entrepreneurial service system. In order to respond to these policies in a timely manner, it is necessary to follow up the development of national informatization, and from the perspective of innovative work services, mainly by means of informatization services, to more efficiently and effectively promote the management of graduate entrepreneurship.

The innovations of this paper are as follows: (1) It introduces the theoretical knowledge of innovation and entrepreneurship and cloud computing in colleges and universities,

and uses data mining to analyze the importance of data mining in intelligent cloud computing data processing systems. (2) It expounds data mining and association rules. Through experiments, it is found that the intelligent cloud computing data processing system based on data mining algorithm can improve the work efficiency of data statistics.

2. Related Work

As the country attaches great importance to innovation and entrepreneurship, more and more people choose to start their own businesses after graduation. LIN found that in recent years, social entrepreneurship education has ushered in a period of vigorous development. However, in the field of entrepreneurship education for college students, social entrepreneurship education is not satisfactory. The scholar mentioned that innovation and entrepreneurship is the general trend and also found that entrepreneurship education is not very perfect, but he did not propose how to solve this problem [1]. Deng found that mobile users generally have high demand for localization and information services. However, retrieving data from remote locations is often inefficient, so there is edge computing, which is an extension of cloud computing. In this basic framework, it is very important to study the interaction and cooperation between edge computing and cloud technology. Although the scholar realized that it is important to study the relationship between edge computing and cloud technology, he did not mention what the relationship between the two is [2]. Wei found that the existing resource scheduling algorithms cannot meet the resource scheduling requirements required by cloud computing and the current cloud infrastructure solutions only provide operational support at the basic level. Considering the competitive nature in cloud computing, he proposed a cloud resource allocation model using the Hidden Markov Model in cloud computing environment. Although the scholar proposed a cloud resource allocation model, he did not mention the specific concept of this model, nor did experiments to prove its feasibility [3]. Jin finds that data sharing is an attractive service offered by cloud computing platforms because of its convenience and economy. As a potential technology to realize data sharing, attribute-based encryption has attracted a lot of attention. However, most of the existing encryption solutions have the disadvantages of high computational overhead and weak data security, which seriously hinders resource-constrained mobile device customization services. The scholar found that encryption is a promising technology and also found its shortcomings, but did not propose specific solutions for these shortcomings [4]. Hirai found that cloud computing provides a large-scale parallel distributed processing service, in which a huge task is split into multiple subtasks, and processing the subtasks takes a long time. An efficient way to mitigate this problem is to have another worker perform the same subtask, so he takes into account the efficiency of the backup task. However, the scholar has no specific experiments to prove whether the backup task can really effectively improve work efficiency [5]. Ghahramani found that the resources of cloud computing are provided by the user with minimal adminis-

trative effort; however, there are some obstacles and concerns in the use of the cloud. The scholar only mentioned some obstacles and concerns in the use of cloud computing, and did not give a clear introduction to these obstacles [6]. Barsoum found that more and more organizations are choosing to outsource data to remote cloud service providers (CSPs), where customers can pay to store large amounts of data. To improve scalability, availability, and durability, some customers may wish to replicate their data across multiple servers in multiple data centers. Therefore, the customer needs to have a strong guarantee, and he proposes a dynamic data ownership scheme. Although the scholar proposed corresponding solutions for the corresponding problems, there is no actual case to prove that this solution is really feasible [7]. Tsai found that in modern society, the number of mobile users increased dramatically, he proposed an efficient distributed mobile cloud computing service authentication scheme, and the proposed scheme provides security and convenience for mobile users. The security strength of the proposed scheme is based on the cryptographic system and dynamic random number generation. Although the scholar proposed a safe and reliable scheme, he did not enumerate the relevant data to support the authenticity of his proposed scheme [8].

3. Data Mining Algorithm Based on Big Data

3.1. The Concept of Cloud Computing and Innovation and Entrepreneurship. With the advancement of technology, the hardware composition of computers has undergone great development. The Internet is changing every day, and the amount of data on the network is increasing by leaps and bounds. No matter how many resources are added, the demand cannot be met. Due to the huge amount of data on the Internet, the computer is completely unable to perform the corresponding tasks [9]. How to integrate and optimize resources has become an important topic in the field of computer applications. Under the premise of such application, cloud computing is put on the agenda. The architecture of cloud computing is shown in Figure 1.

As shown in Figure 1, the large-scale data storage of cloud computing is mainly realized by the decentralized file system, and the design of the decentralized file system must meet the requirements of transparency, scalability, failure resistance, and security [10]. With the rapid development of China's economy, the number of university enrollments continues to increase, and many university graduates will face enormous pressure to start a business. The state's emphasis on innovation and entrepreneurship in universities is shown in Figure 2.

As shown in Figure 2, for China, how to deal with the times and develop innovative education is a major issue facing China's higher education. Now, China is actively adopting reforms to promote innovation and entrepreneurship. In order to promote entrepreneurship, the state has accelerated the construction of makerspaces and supported and encouraged enterprises, investment institutions, industry organizations, and other social forces to invest in construction in accordance with the principles of marketization. Therefore,

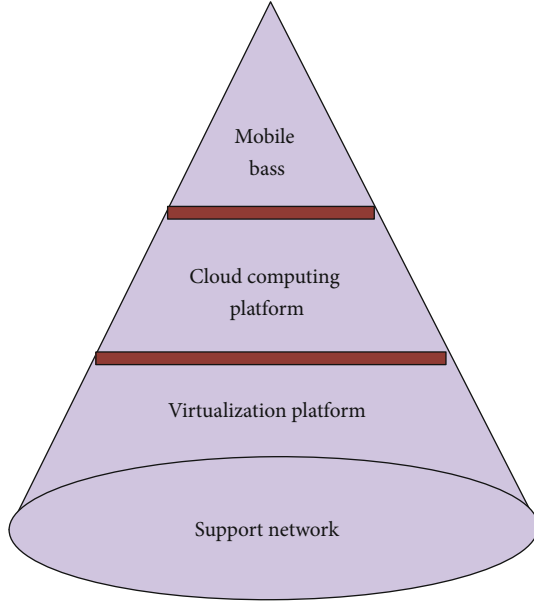


FIGURE 1: Architecture of cloud computing.

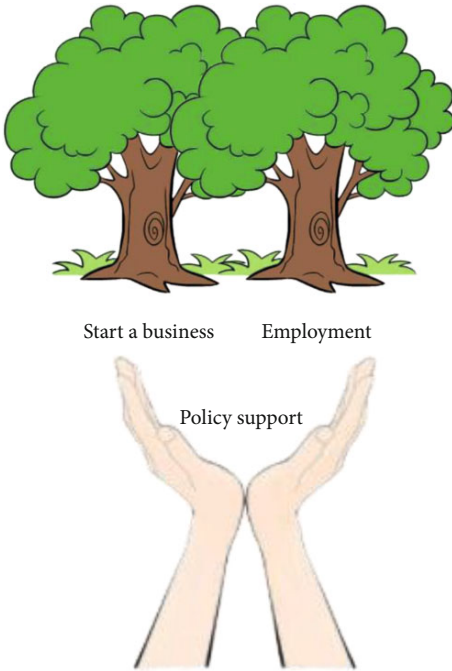


FIGURE 2: Innovation and entrepreneurship supported by the state.

in order to cultivate a talent team with innovative consciousness and entrepreneurial spirit and realize the sustainable development of China’s modern economy, the state has proposed a series of supporting policies [11].

The education and training of colleges and universities is of great significance to the cultivation of outstanding talents. Therefore, it is very critical and necessary to conduct data mining on the information of college students [12]. With the advancement of science and technology, the innovation and entrepreneurship data statistics system is also constantly

updated. Innovation and entrepreneurship in the information age are shown in Figure 3.

As shown in Figure 3, with the advent of the era of big data, especially the rapid development of big data-related technologies, universities are paying more and more attention to the analysis and mining of existing historical data, and innovation and entrepreneurship are carrying out more scientific educational activities [13].

There is a large amount of entrepreneurial data information stored in the student data management system, and the hidden relationship plays a crucial role in the reform of entrepreneurial methods and innovation systems [14]. The flowchart of the innovation and entrepreneurship data statistics system is shown in Figure 4.

As shown in Figure 4, this flowchart fully shows the process of data mining in the system. The system should have all kinds of information of students during school so that it can achieve data sharing and easy for users to analyze student information [15]. It provides efficient services for the staff engaged in education and at the same time realizes the systematization, standardization, and automation of student information management.

3.2. Parallel Apriori Algorithm Based on Weighted Itemsets.

The Apriori algorithm is a frequent itemset algorithm for mining association rules. Its core idea is to mine frequent itemsets through two stages: candidate set generation and plot downward closure detection. The parallel Apriori algorithm based on weighted itemsets implements the distributed Apriori algorithm. In order to achieve efficient parallel computing on the original database, preprocess the data, and filter out all infrequent itemsets, this paper transforms the database into weight matrix and Boolean matrix [16]. For any given transaction database D , let Equation (1) hold:

$$f : D \longrightarrow R. \tag{1}$$

Among them, $R = f(D) = (r_{ij})_{n \times m}$, n is the number of transactions, m is the number of items, and transaction set r_{ij} is shown below:

$$r_{ij} = \begin{cases} 1, & I_j \in T_i \\ 0, & I_j \notin T_i \end{cases}. \tag{2}$$

Apriori is a data mining algorithm that uses multiple loop iterations to find all frequent itemsets layer by layer, but it has two relatively more serious defects: multiple scans of the original data set and too large a candidate itemset per iteration [17]. In order to improve the main defects of Apriori algorithm, scholars proposed a distributed Apriori algorithm.

The Apriori algorithm under the distributed architecture is their key execution data parallelization strategy that is to cut and distribute the data to the corresponding processors, then perform processing operations on the local data, obtain the local processing results, then integrate the results on all processors, and finally get all the results [18]. This paper compares and analyzes the traditional Apriori algorithm

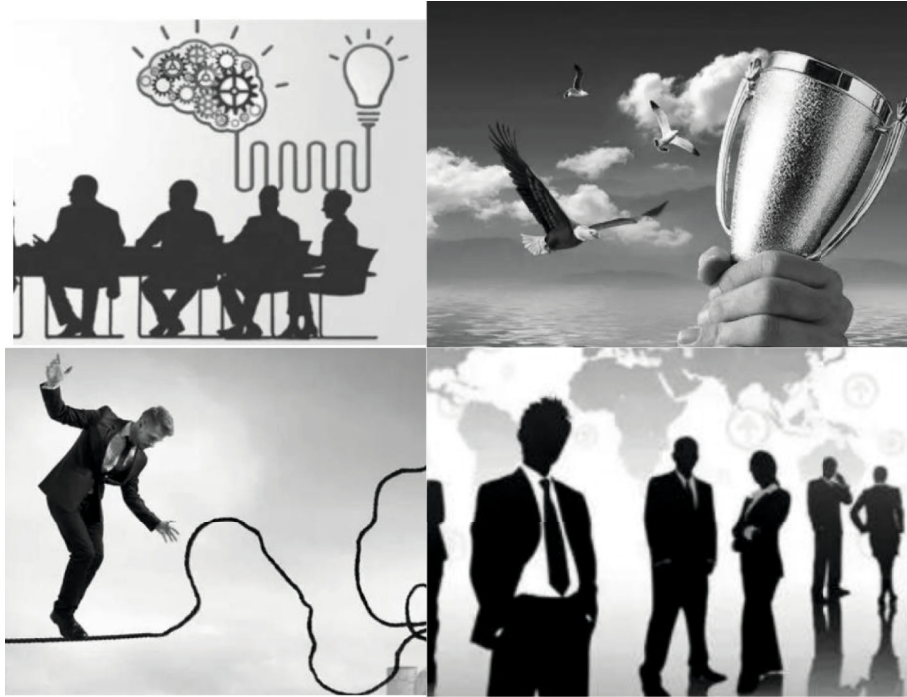


FIGURE 3: Innovation and entrepreneurship in the information age.

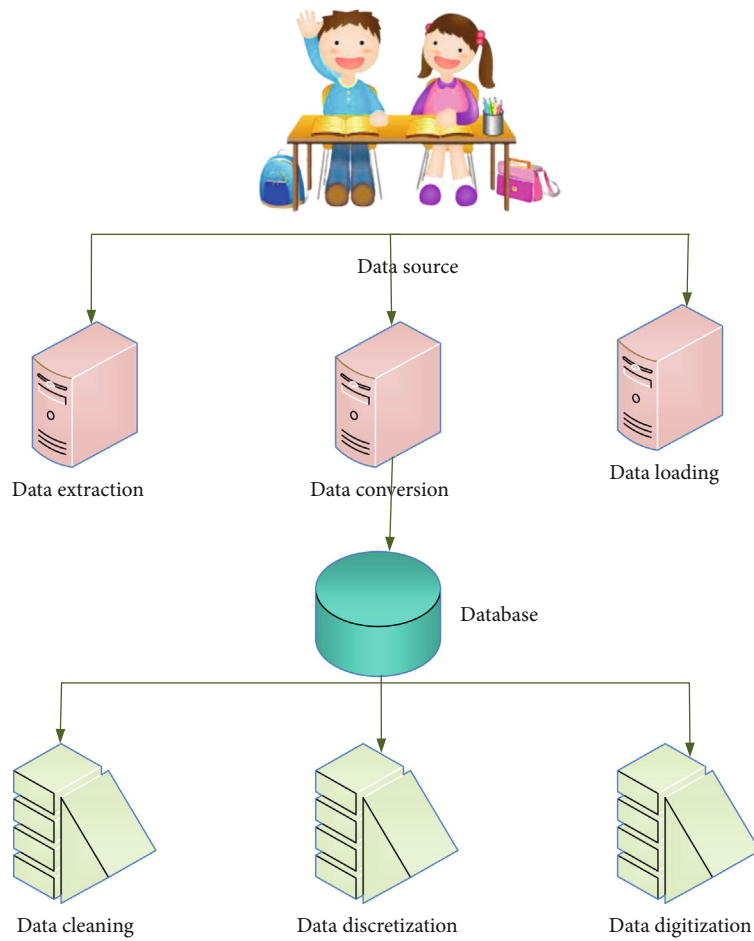


FIGURE 4: Flow chart of the innovation and entrepreneurship data statistics system.

and the distributed Apriori algorithm, as shown in Tables 1 and 2.

As shown in Tables 1 and 2, the running time of the traditional Apriori algorithm is between 28 min and 36 min, the computing power is between 39% and 47%, and the computing efficiency is between 28% and 32%. Compared with the traditional algorithm, the parallel Apriori algorithm shortens the running time, improves the computing power, and also improves the running efficiency of the algorithm. It can be seen that choosing the distributed Apriori algorithm is obviously more advantageous [19].

3.3. Data Preprocessing. Data preprocessing refers to some processing performed on the data before the main processing. For example, before converting or enhancing most of the data, firstly, the irregularly distributed measurement network is converted into a regular network through interpolation, so as to facilitate computer operations. If the data is directly calculated, the data is preprocessed first because the data is too large and the infrequent candidate itemsets are repeatedly calculated. After filtering out the non-frequent itemsets, the candidate itemsets and higher itemsets will be reduced by a large amount [20]. The efficiency of the algorithm is improved because the support calculation of unnecessary candidate itemsets is reduced and the running time of the algorithm is shortened.

First, the original database D is scanned once, and the itemsets in all transactions are traversed, and the support counts are superimposed on the occurrences. Record the support count of itemset A denoted as $A.\text{sup}$:

$$A.\text{sup} = \text{sum}A(T_i). \quad (3)$$

In the equation, $i = 1, 2, \dots, n$ and n are the number of transactions in the data, and $i = 1, 2, \dots, n$ is the sum of the occurrences of itemset A in all transactions.

The first stage of association rule mining must find out all high-frequency item groups from the original data set. High frequency means that the frequency of a certain item group must reach a certain level relative to all records. In the study of association rules, people often only care about the existence of itemset A , but ignore the quantity of A itself in a transaction. The degree of association between A and B should not only consider the number of times they appear in the transaction database at the same time, but also the number of A and B themselves [21].

In a certain transaction, the minimum value of the number of all items in the itemset A is the weight of A in the transaction, which is represented by $A.\text{weight}$:

$$A.\text{weight} = \min(A_i.\text{weight}), \quad (4)$$

where A_i represents any item in itemset A .

The average weight is the ratio of the cumulative sum of the weights of itemset A in all transactions to $A.\text{weight}$, represented by $A.\text{averageweight}$:

$$A.\text{averageweight} = \sum A.\text{weight}(A_i) / A.\text{sup}, \quad (5)$$

TABLE 1: Traditional Apriori algorithm.

Analysis object	Times 1	Times 2	Times 3
Operation hours	36 min	28 min	32 min
Calculate ability	39%	47%	45%
Operating efficiency	28%	32%	30%

TABLE 2: Distributed Apriori algorithm.

Analysis object	Times 1	Times 2	Times 3
Operation hours	10 min	8 min	9 min
Calculate ability	73%	79%	90%
Operating efficiency	68%	67%	70%

In Equation (5), $i = 1, 2, \dots, n$, and n is the number of transactions in the data, and $\text{sum}A.\text{weight}(A_i)$ is the cumulative sum of the weights of A in all transactions.

When it comes to parallel association rule mining algorithms, the amount of chunked data allocated to each host must be considered. If the data block is large, the computing pressure on the node will be large, and the data cutting is too small to lose the meaning of parallel computing.

When performing data mining on a given database, the process of association rule mining is generally divided into two processes, as shown in Figure 5.

As shown in Figure 5, allocating data of balanced size to nodes can make full use of the computing resources of the cluster and shorten the running time of the algorithm. The operation of Boolean matrix is the same as that of general matrix, for example, it can be connected and can be transposed. The weight matrix refers to the frequency of each number in the weighted average of complex or real sets arranged in a rectangular array. The Boolean matrix and the weight matrix are, respectively, cut into two small matrices and then transposed to the following:

$$BR_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, BR_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (6)$$

Dividing the data into n blocks and the sum of the local minimum support in n nodes is the global minimum support. Therefore, the size of the local minimum support is the product of the global minimum support and the number of transactions in the node data block, which is represented by $L \text{ min_sup}$:

$$L \text{ min_sup} = \text{min_sup} * |BR| / n. \quad (7)$$

Both the local average weight and the global average weight represent the correlation degree of a certain item combination within the itemset, which is represented by $A.\text{LaverageWeight}$:

$$A.\text{LaverageWeight} = \sum A_i.\text{weight} / A.L \text{ sup}, \quad (8)$$

where $A.L \text{ sup}$ is the local support count of A .

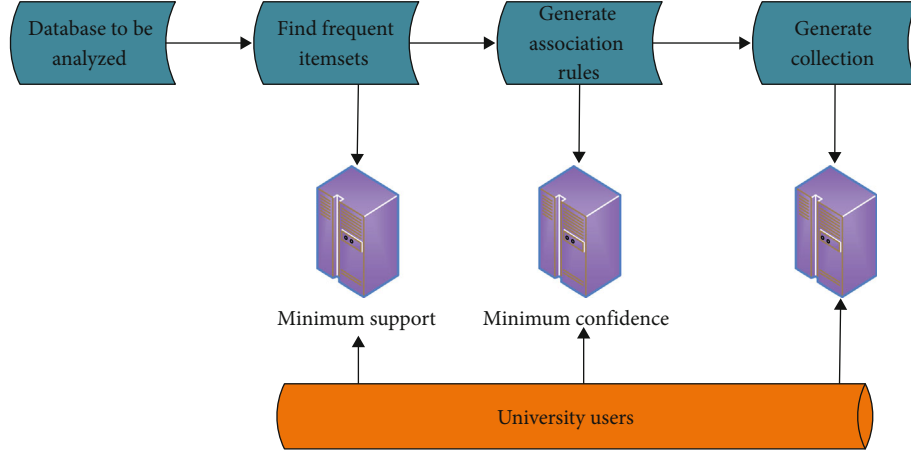


FIGURE 5: Association rule process.

The weight of an itemset is the number of times that each item contained in the itemset has its own occurrence in a transaction. Taking the minimum number of occurrences can represent the maximum number of common occurrences of all items in the set, and also the number of occurrences of the set in this transaction. The average weight is the average of the occurrences of an item in the minimum computing environment and represents the correlation within the set among all items.

Calculate the support count and average weight of each candidate itemset, which is represented by $A.GaverageWeight$:

$$A.GaverageWeight = \sum \min A_i \cdot (weight) / A.G \text{ sup}, \quad (9)$$

where $A.G \text{ sup}$ is the global support count of itemset A .

A frequent itemset is a set whose support degree is greater than or equal to the minimum support degree, where the support degree refers to the frequency of a certain set appearing in all transactions. All the itemsets that satisfy the minimum average weight and minimum support at the same time are another part of the global frequent itemsets, and these two frequent itemsets are all the global frequent itemsets.

3.4. K-Means Algorithm. K-means is one of the most commonly used clustering algorithms. The biggest feature of the algorithm is that it is simple, easy to understand, and fast in operation, but it can only be applied to continuous data. The k-means algorithm finally gathers the sample data into k clusters. Therefore, the algorithm first needs to select k initial cluster centers, then calculate the distances from the remaining data samples to the k cluster centers, respectively, and put the data sample into the category corresponding to the smallest cluster center in the distance from all k cluster centers.

As shown in Figure 6, although the k-means algorithm has many areas that need to be improved and optimized, when processing some basic large data sets, the obtained clustering effect and the efficiency of the algorithm processing data have been generally recognized.

When a specific function is used to evaluate the criterion of the pros and cons of the strategy adopted by the system, it is called a criterion function. The criterion function equation of the clustering algorithm is

$$E = \sum_{i=1}^k \sum_{a \in C_i} a = a_i^2. \quad (10)$$

Compared with other algorithms, the k-means algorithm is simpler in principle and operation and has high algorithm execution efficiency.

3.5. Fuzzy C-Means Clustering (FCM) Algorithm. The FCM algorithm is a partition-based clustering algorithm, and its idea is to maximize the similarity between objects that are divided into the same cluster, while the similarity between different clusters is the smallest. Scholars use fuzzy set theory to improve the k-means algorithm, finally replace the square of the distance between the sample and the cluster center in the criterion function of the k-means algorithm with the weight of the square of the membership degree, and obtain the objective function of the FCM algorithm.

The objective function is the target form that is pursued by the design variables, so the objective function is the function of the design variables, which is a scalar. Let n -dimensional sample data set be $X = \{x_1, x_2, \dots, x_n\}$, and each sample data has s -dimensional attributes. If the FCM algorithm is used to aggregate this sample data into class c , then the objective function corresponding to the algorithm is

$$J_m(X, Y) = \sum_{i=1}^c \sum_{j=1}^n x_{ij}^m D_{ij}^2, \quad (11)$$

where x_{ij}^m represents the degree of membership between a certain sample data j and the i th category and $0 \leq j \leq n$. In the k-means algorithm, the value of the membership degree x_{ij}^m of the sample object and the corresponding category is either 1 or 0. In the fuzzy C-means clustering FCM algorithm, the value of membership degree is a certain value in

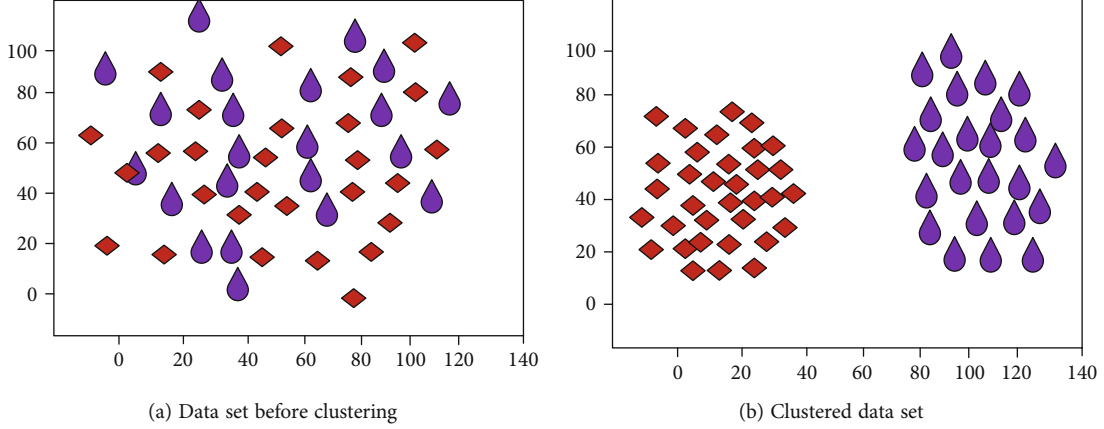


FIGURE 6: Data set performance before and after clustering.

the interval $[0..1]$, indicating the probability that the data object belongs to this class.

When the objective function satisfies the constraint condition $\sum_{j=1}^n x_{ij}^m = 1$ and the minimum value is obtained, the corresponding fuzzy matrix and cluster center are obtained when the optimal clustering effect is obtained. Use the Lagrange multipliers to establish the Lagrange equation:

$$J_m^*(X, Y) = \sum_{i=1}^c \sum_{j=1}^n x_{ij}^m D_{ij}^2 + \lambda \left(\sum_{j=1}^n x_{ij}^m - 1 \right). \quad (12)$$

Fuzzy matrix is used to represent the matrix of fuzzy relationship. If set X has m elements and set Y has n elements, the fuzzy relationship from set X to set Y can be represented by a matrix. The partial derivative of Equation (12) is calculated, and the fuzzy matrix and cluster center corresponding to the minimum value of the objective function are obtained through simplification, as shown below:

$$X_{ij} = \left[\sum_{k=1}^c \left[\begin{array}{c} D(a_j, v_i) \\ D(a_j, v_k) \end{array} \right]^{m-1} \right]^{-1}, \quad (13)$$

$$v_i = \sum_{j=1}^n x_{ij} a_j. \quad (14)$$

At this time, we need to pay attention to a special case: When the denominator under the fuzzy matrix in Equation (13) is zero, the distance $D(a_j, v_i)$ between the sample data and the cluster center at this time is 0. It is generally considered that the j th sample data at this time belongs to the i class. After this process is completed, the average value of the cluster is recalculated to obtain a new cluster center, and this process is repeated until the standard function converges. The distance calculation function is generally Euclidean distance, and the distance convergence criterion for data

objects is usually the square error criterion, which is

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2. \quad (15)$$

Among them, p represents any data object, and E represents the sum of squared errors of all data objects. m_i is the average value of the cluster C_i where the data p is located, that is, the sum of the squares of the distances between each data object and the center point of the cluster where it is located is finally obtained. The absolute deviation of the average can be calculated as

$$s_f = \frac{1}{n} (|a_{1f} - m_f| + |a_{2f} - m_f| + \dots + |a_{nf} - m_f|), \quad (16)$$

Here, $a_{nf} - m_f$ is the n measures of f , and m_f is the average value of f , which is

$$m_f = \frac{1}{n} (a_{1f} + a_{2f} + \dots + a_{nf}). \quad (17)$$

The normalized measure is calculated as

$$z_{if} = \frac{a_{if} - m_f}{s_f}. \quad (18)$$

Among them, the mean absolute deviation of a_{if} is more robust to abnormal data than the standard deviation. The most commonly used distance measure is the Euclidean distance. Euclidean metric, also known as Euclidean distance, is a commonly used definition of distance, which refers to the true distance between two points in m -dimensional space, or the natural length of a vector. Euclidean distance in 2D and 3D space is the actual distance between two points, which is

$$d(i, j) = \sqrt{|a_{i1} - a_{j1}|^2 + |a_{i2} - a_{j2}|^2 + \dots + |a_{ip} - a_{jp}|^2}. \quad (19)$$

Here, a_{ip} and a_{jp} are two P -dimensional data objects. Another commonly used distance calculation method is

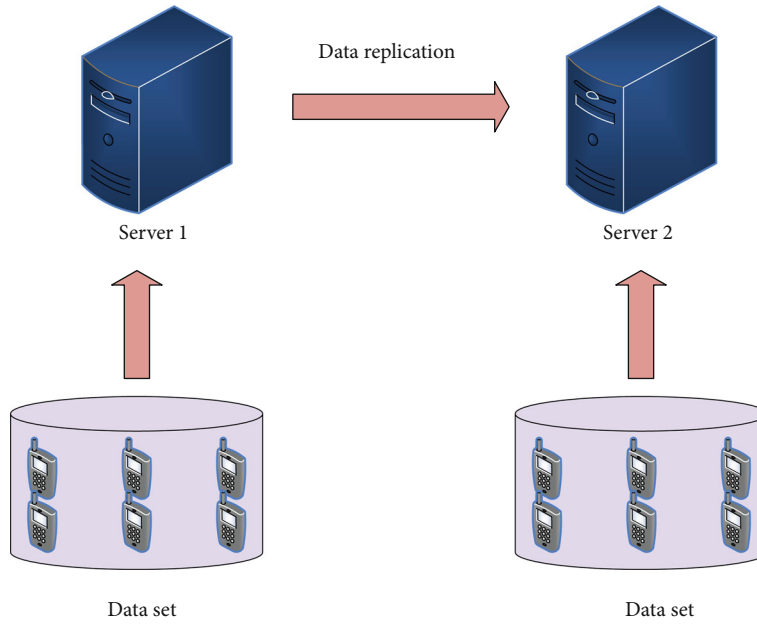


FIGURE 7: The general process of data replication technology.

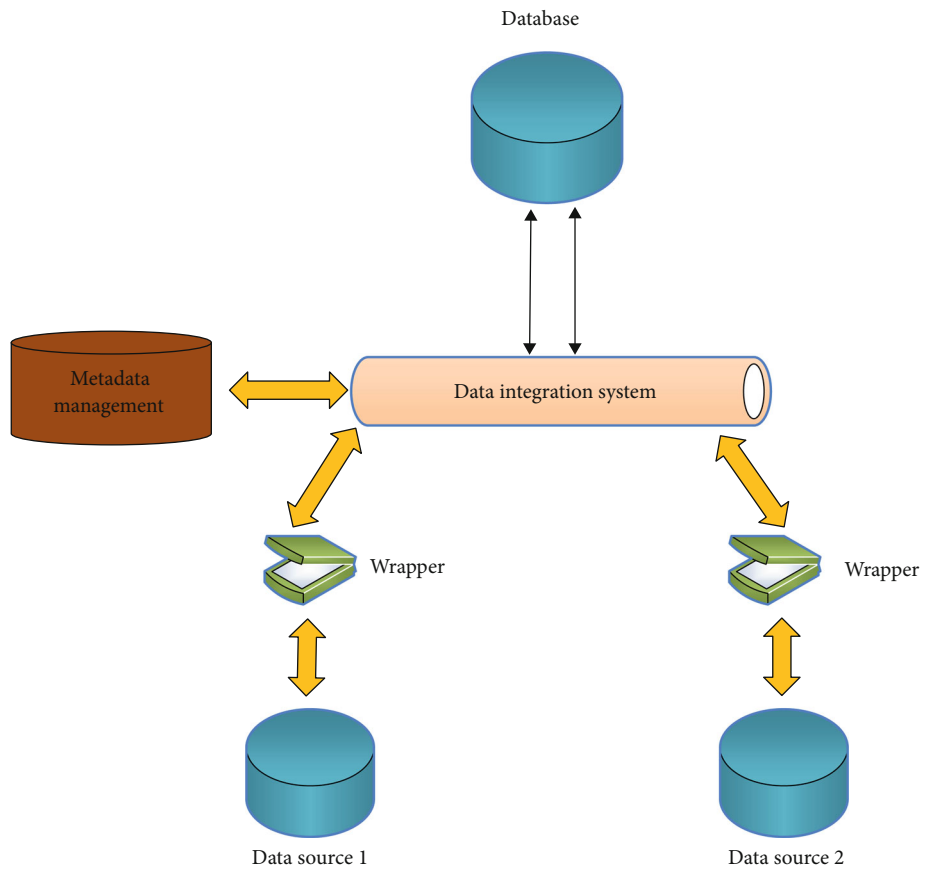


FIGURE 8: Data integration model based on data warehouse.

the Manhattan distance. The Manhattan distance, also known as Manhattan distance, is the distance between two points in the north-south direction plus the distance in the

east-west direction. When the coordinate axis changes, the distance between the points will be different, and its specific calculation equation is defined as

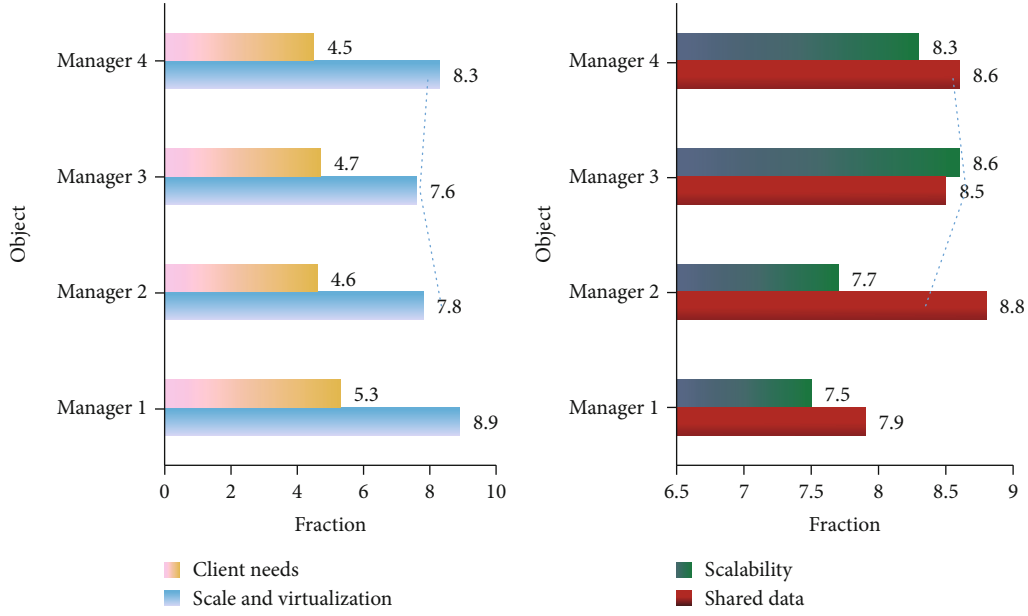


FIGURE 9: Characteristics of cloud computing.

TABLE 3: Problems existing in entrepreneurship data collection.

Years	Utilization level	Use awareness	How to use
2015	36%	52%	Single
2016	31%	49%	Single
2017	28%	38%	Single
2018	40%	42%	Single
2019	37%	45%	Single

$$d(i, j) = |a_{i1} - a_{j1}| + |a_{i2} - a_{j2}| + \dots + |a_{ip} - a_{jp}|. \quad (20)$$

Data replication technology solves the mismatch of different data to a certain extent by copying data from each data source to other related data sources, which greatly improves the efficiency of information sharing and use. The general process of data replication technology is shown in Figure 7.

As shown in Figure 7, data replication technology can ensure that the same data stored on different nodes is consistent. In this way, when a node fails, data can be obtained from other nodes that store the data to avoid data loss, thereby improving the reliability of the system. Data replication improves the performance of the data merging system by reducing the amount of different data sources and avoids frequent access to different data sources, which reduces query efficiency.

Because there are many different databases, the extraction of data information and the development of knowledge mining are limited to a certain extent. In order to lift this limitation, data warehouse technology came into being. The data integration model based on the data warehouse is shown in Figure 8.

As shown in Figure 8, the data warehouse itself is also a database, and it is very simple for users to access the data

warehouse. At the same time, the data warehouse stores all the data required for user decision support, mining and analysis, and provides users with powerful analysis and execution data support.

4. Data Processing System Experiment and Analysis of Intelligent Cloud Computing

4.1. *Characteristics of Cloud Computing.* This article interviewed 4 campus data system managers and analyzed their characteristics of cloud computing, as shown in Figure 9.

As shown in Figure 9, the characteristics of cloud computing are as follows:

- (1) **Hyperscale and virtualization:** Cloud computing provides the most reliable and secure data storage center. Users do not have to worry about data loss, virus intrusion, and other issues. At the other end of the “cloud” are the world’s most professional teams that support information management and the most advanced data centers that help keep data
- (2) **Client demand is low:** Cloud computing requires the least equipment on the client side and is the most convenient to use. For example, in order to prevent viruses from being introduced during downloading, antivirus and firewall software have to be installed repeatedly. With cloud technology, as long as you have a computer with Internet access and a favorite browser, you can enjoy the infinite fun that cloud computing brings to you
- (3) **Easily share data:** Cloud computing can simply realize the sharing of data and applications between different devices. In the network application model of cloud computing, all electronic devices can access

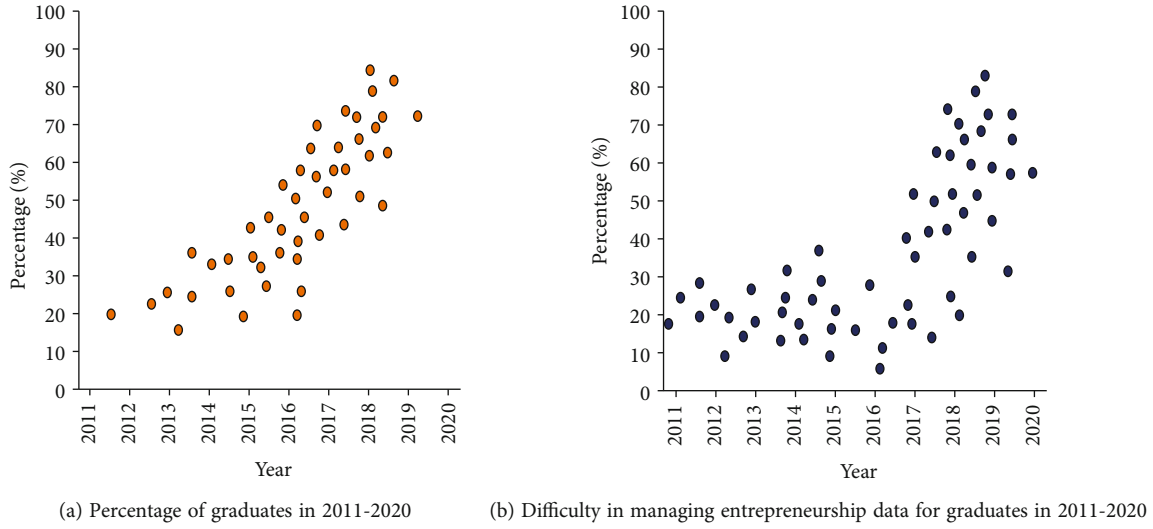


FIGURE 10: Percentage of graduates from 2011-2020 and difficulty of graduate entrepreneurial data management.

TABLE 4: Traditional data processing systems.

Number of experiments	Scientific	Practicality	Efficiency	Cost
5	47.8%	35.7%	27.9%	58.9%
10	43.6%	34.2%	28.5%	56.1%
15	46.7%	36.0%	26.6%	55.0%
20	42.9%	37.2%	29.0%	54.5%

the same data at the same time by simply connecting to the Internet

- (4) Strong scalability: Cloud computing is not application specific. By supporting “cloud,” various applications can be built according to the needs of users. The scale of the “cloud” can also be dynamically scaled to adapt to the needs of the application environment and the expansion of the user scale

4.2. *Traditional Data Processing System and Intelligent Cloud Computing Data Processing System.* At present, various colleges and universities generally have two problems: the lack of scientificity and accuracy in the collection of graduate entrepreneurial data and the low level of utilization of entrepreneurial data. For colleges and universities, it is difficult to obtain accurate entrepreneurial data and make effective use without a set of scientific entrepreneurial data collection processes and standards. The problems existing in the collection of entrepreneurial data are shown in Table 3.

As shown in Table 3, the utilization rate of entrepreneur data is low, the awareness of utilization is low, and the utilization method is simple. The lack of scientific and accurate data collection for entrepreneurs directly reduces the value and importance of data utilization.

How to make good use of the valuable resource of entrepreneurial data and provide a scientific basis for the transformation of schools and the reform of disciplines and majors is a problem that colleges and universities urgently

need to solve. The solution of these problems is of great significance to the country, universities, and graduate groups.

With the expansion of undergraduate enrollment in Chinese universities year by year, in an environment where the number of graduates increases significantly every year, the graduate department of the Talent Center not only needs to use the comprehensive management information system for entrepreneurial services to manage a large amount of graduate data, but also need to analyze these data. This paper analyzes the percentage of graduates from 2011 to 2020 and the difficulty of graduate entrepreneurial data management, as shown in Figure 10.

As shown in Figure 10, therefore, in order to intelligently and quickly obtain useful information and knowledge on entrepreneurial services such as entrepreneurial trainees from a large amount of data to help staff make overall planning for entrepreneurial services, data mining of historical data in the comprehensive management information system of entrepreneurial services, and analysis of the models and relationships in the mining results have become an indispensable link.

With the realization of global information resource sharing, the rapid growth of the amount of information and the improvement of the request for information. At present, it is difficult to meet the needs of practical applications based on traditional information extraction methods, so data mining technology is proposed. This paper compares the traditional data processing system and the intelligent cloud computing data processing system, as shown in Tables 4 and 5:

As shown in Tables 4 and 5, driven by today’s high-tech development, the application scope of data mining is getting wider and wider. Applying it to university management can not only promote the further reform, improvement and development of school management, but also provide a favorable basis for managers to make correct decisions. At the same time, it also improves the scientificity, practicability and efficiency of the school’s educational administration management methods. It uses data mining technology to

TABLE 5: Intelligent cloud computing data processing system.

Number of experiments	Scientific	Practicality	Efficiency	Cost
5	67.9%	89.4%	73.5%	20.1%
10	78.3%	76.7%	69.0%	15.7%
15	79.2%	81.5%	72.7%	12.9%
20	84.0%	77.3%	75.8%	13.8%

extract and analyze massive data, find hidden clues, and provide more valuable basis for further decision-making.

5. Discussion

This paper discusses how to study an intelligent cloud computing data processing system for college innovation and entrepreneurship data statistics, describes the theoretical knowledge related to intelligent cloud computing and college innovation and entrepreneurship, and focuses on the difficulty of data statistics. It explores a more scientific method of data statistics, discusses the effect of cloud computing on data processing system research through experimental analysis, and finally finds that cloud computing can make data processing systems more practical.

This paper also studies the cluster analysis and studies the cluster analysis before and after the improvement. It combines traditional clustering analysis with fuzzy theory and obtains an improved fuzzy clustering analysis algorithm. Cluster analysis plays an important role in data mining, which can make data classification more rapid.

It can be known from the experimental analysis in this paper that using the data mining method in the intelligent cloud computing data processing system can not only improve the practicability of the intelligent cloud computing data processing system, and the accuracy of data statistics has also been greatly improved.

6. Conclusions

With the increase of college graduates in recent years, the problem of entrepreneurship of college graduates has also become a national concern. The state not only encourages graduates to actively start their own businesses but also encourages them to actively innovate and start businesses and has given them a lot of support. However, with the increase of talents, it becomes more and more difficult to collect information and data for these graduates. Therefore, this paper proposes an intelligent cloud computing data processing system for college innovation and entrepreneurship data statistics. The system is based on the basis of data mining and is implemented in combination with cloud computing. However, because of the author's limited ability, this paper makes a general analysis and research on the framework of the data processing system. In the method part, it mainly expounds the association rule algorithm and cluster analysis method based on data mining. It applies two algorithms to the intelligent cloud computing data processing system, which can improve the efficiency of data processing

and classify data, thus making it easier for graduates to innovate and start businesses. In the experiment, this paper analyzes the advantages of cloud computing and finds that cloud computing not only has a large scale to process data, but also can share data, which makes the information collection of entrepreneurship more convenient. The experiment finally found that the data processing system based on cloud computing is more efficient than the traditional data processing system, and has a high degree of science.

Data Availability

No data was used to support this study.

Conflicts of Interest

The authors declare that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

Acknowledgments

This work were supported by the Key Scientific Research Projects of Colleges and Universities in Henan Province (20B880015) and Special research project on innovation and entrepreneurship teaching reform of Yellow River Conservancy Technical Institute in 2021(2021CXCYJG008).

References

- [1] A.-J. Lin and H. Tang, *Social Entrepreneurship Education: New Development in College Entrepreneurship Education*, no. 2017-3, pp. 125–125, 2021.
- [2] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "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.
- [3] 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.
- [4] L. Jin, Y. Zhang, X. Chen, and Y. Xiang, "Secure attribute-based data sharing for resource-limited users in cloud computing," *Computers & Security*, vol. 72, pp. 1–12, 2018.
- [5] T. Hirai, H. Masuyama, S. Kasahara, and Y. Takahashi, "Performance analysis of large-scale parallel-distributed processing with backup tasks for cloud computing," *Journal of Industrial & Management Optimization*, vol. 10, no. 1, pp. 113–129, 2014.
- [6] M. H. Ghahramani, M. C. Zhou, and T. H. Chi, "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.
- [7] A. Barsoum and M. A. Hasan, "Provable multicopy dynamic data possession in cloud computing systems," *IEEE Transactions on Information Forensics & Security*, vol. 10, no. 3, pp. 485–497, 2017.
- [8] 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, 2017.

- [9] L. Ping, J. Li, Z. Huang et al., "Multi-key privacy-preserving deep learning in cloud computing," *Future Generation Computer Systems*, vol. 74, pp. 76–85, 2017.
- [10] W. Tang and W. Feng, "Parallel map projection of vector-based big spatial data: coupling cloud computing with graphics processing units," *Computers, Environment and Urban Systems*, vol. 61, pp. 187–197, 2017.
- [11] M. Abdel-Basset, M. Mai, and V. Chang, "NMCDA: a framework for evaluating cloud computing services," *FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE*, vol. 86, pp. 12–29, 2018.
- [12] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "IoT-based big data storage systems in cloud computing: perspectives and challenges," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 75–87, 2017.
- [13] J. Du, L. Zhao, F. Jie, and X. Chu, "Computation offloading and resource allocation in mixed fog/cloud computing systems with min-max fairness guarantee," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1594–1608, 2018.
- [14] N. Kratzke and P. C. Quint, "Understanding cloud-native applications after 10 years of cloud computing - a systematic mapping study," *Journal of Systems & Software*, vol. 126, pp. 1–16, 2017.
- [15] Z. Cao, L. Jin, C. Wan, Y. Song, Y. Zhang, and X. Wang, "Optimal cloud computing resource allocation for demand side management," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1943–1955, 2017.
- [16] S. Wang, K. Liang, J. K. Liu, J. Chen, J. Yu, and W. Xie, "Attribute-based data sharing scheme revisited in cloud computing," *IEEE Transactions on Information Forensics & Security*, vol. 11, no. 8, pp. 1661–1673, 2016.
- [17] X. F. Liu, Z. H. Zhan, J. D. Deng, Y. Li, T. Gu, and J. Zhang, "An energy efficient ant colony system for virtual machine placement in cloud computing," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 1, pp. 113–128, 2018.
- [18] V. Sudarsan and N. Satyanarayana, "Secure and practical outsourcing of linear programming in cloud computing: a survey," *International Journal of Computer Applications*, vol. 159, no. 4, pp. 1–4, 2017.
- [19] A. Javed, H. Larijani, A. Ahmadinia, and D. Gibson, "Smart random neural network controller for HVAC using cloud computing technology," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 351–360, 2017.
- [20] M. S. Hasan, Y. Kouki, T. Ledoux, and J. L. Pazat, "Exploiting renewable sources: when green SLA becomes a possible reality in cloud computing," *IEEE Transactions on Cloud Computing*, vol. 5, no. 2, pp. 249–262, 2017.
- [21] F. Zafar, A. Khan, S. Malik et al., "A survey of cloud computing data integrity schemes: design challenges, taxonomy and future trends," *Computers & Security*, vol. 65, pp. 29–49, 2017.