

Retraction

Retracted: Construction of an Intelligent Analysis Model for Website Information Based on Big Data and Cloud Computing Technology

Discrete Dynamics in Nature and Society

Received 22 August 2023; Accepted 22 August 2023; Published 23 August 2023

Copyright © 2023 Discrete Dynamics in Nature and Society. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] R. Chen and B. Yang, "Construction of an Intelligent Analysis Model for Website Information Based on Big Data and Cloud Computing Technology," *Discrete Dynamics in Nature and Society*, vol. 2022, Article ID 7876119, 10 pages, 2022.

Research Article

Construction of an Intelligent Analysis Model for Website Information Based on Big Data and Cloud Computing Technology

Ronghua Chen ¹ and Bei Yang²

¹School of Information Engineering, Jiangxi Vocational College of Finance & Economics, Jiujiang, Jiangxi 332000, China

²School of Foreign Languages, Jiangxi Vocational College of Finance & Economics, Jiujiang, Jiangxi 332000, China

Correspondence should be addressed to Ronghua Chen; 53712126@qq.com

Received 11 November 2021; Accepted 7 December 2021; Published 7 January 2022

Academic Editor: Gengxin Sun

Copyright © 2022 Ronghua Chen and Bei Yang. 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.

Based on big data and cloud computing technology, the development process of this system includes hardware cluster deployment of the delivery system, optimization of website delivery strategy, and development of the delivery management background system. The main functions in the following aspects are realized: first, we built the website back-end delivery subsystem and data collection and analysis subsystem to realize the control of website delivery and data collection; second, we designed and developed a process management subsystem for booking, management, and delivery of website resources and developed the contract management and order management subsystems to realize the accurate placement of user portraits on the website; then, the placement data monitoring and effect feedback subsystems, and the data inventory subsystem of the website system were designed and developed. Finally, based on the research of Android and based on the Eclipse platform, this article has completed the construction of the Android 4.0.3 version environment, successfully used the Java development language to develop a website information intelligent analysis and navigation system, and analyzed the various functional modules of the entire system. Experimental results show that the system not only realizes ordinary route and site query functions but also combines map API, integrates big data and cloud computing technology, and realizes congestion avoidance query, time optimal query, population heat map, and real-time viewing. The nearby use of this software's personnel density distribution and other functions provides great convenience for personal travel, which can facilitate the real-time planning of travel plans, and has great practical and practical significance. Through the different levels of testing of various subsystems, the website delivery system meets the functional and nonfunctional requirements proposed by the network and on this basis realizes the use of group wisdom based on Pearson correlation coefficient, Cosine similarity, and Tanimoto coefficient for collaborative filtering website recommendation algorithm.

1. Introduction

In recent years, cloud computing technology has developed rapidly, and virtualization technology, as the key technology of cloud computing, has always been one of the hot research issues [1]. Virtualization technology includes multiple aspects, such as network virtualization, virtual machine placement, and virtual machine migration. This paper mainly focuses on the related issues of virtual machine placement in data centers. The online website delivery system is developed with the rapid development and gradual improvement of the online website market and big data

technology. It aims to help website vendors and online video websites achieve professional and accurate placement and page-based website targeting. System software is for tasks such as delivery process management and website operation data statistics [2–5]. As a leading video sharing website, online websites have played more than tens of millions of times a day. The company built an online website delivery management system with the purpose of using the cloud computing platform to strictly control the website delivery process to ensure accurate delivery of the website. Under the premise, at the same time, it can integrate the company's existing sales, product, and technical teams to form a

reliable, diverse, and customizable website sales program [6–8]. The website system can effectively optimize the website delivery plan and automatically adjust the delivery parameters during the delivery process, ensuring the most effective use of the website resources from the technical level.

In a cloud data center, in order to improve resource utilization, a scheme of placing multiple virtual machines on the same physical server is usually adopted, and the placement of virtual machines will consider many factors, including reliability, energy consumption, and network resources. Consumption and a reasonable and efficient virtual machine placement algorithm can improve the operation efficiency of the data center and save operation and maintenance costs. In the existing research, most of the virtual machine placement algorithms use traditional algorithms or heuristic algorithms [9–11]. Some researchers apply machine learning technology to the virtual machine placement problem, but most of them are just as auxiliary strategies, such as using machine learning technology for demand forecasting and so on. There are few algorithms that directly apply machine learning technology to decision-making. Yao et al. [12] proposed an energy-aware virtual machine placement algorithm based on the hybrid genetic algorithm, which comprehensively considers the energy consumption of physical servers in the data center and the resource consumption of network communication. Based on tabu search, Singh et al. [13] proposed a network-aware virtual machine placement algorithm and a heuristic virtual machine placement algorithm based on convex optimization. The goal is to optimize energy consumption and network bandwidth consumption and try to reduce the network through graph segmentation flow. Based on the improved particle swarm algorithm, Jebble et al. [14] proposed an energy-aware virtual machine placement algorithm. In addition to the optimization goals of energy consumption and network consumption, studies have also considered factors such as the reliability, quality of service, and security of virtual machine placement. Wu et al. [15] proposed a reliability-aware virtual machine placement algorithm. The algorithm is based on network topology, selects a series of available servers from candidate server clusters, then selects an optimization strategy based on K-Fault-Tolerance Guarantee, places the master and slave virtual machines on the selected available servers, and finally uses a heuristic strategy to solve the problem of remapping and optimization of tasks to virtual machines. Rehman et al. [16] added the consideration of load balancing and proposed a virtual machine placement algorithm based on energy consumption and service quality perception. Scholars proposed a virtual machine placement algorithm based on the particle swarm optimization algorithm. Unlike other studies, this paper considers the heterogeneity of physical servers and puts forward a corresponding virtual machine placement strategy while taking into account energy consumption and ensuring global QoS [17–19]. Focusing on the security issues in virtual machine placement, in response to attacks on virtual machines that have occurred in recent years, researchers have proposed a security-aware multiobjective optimization virtual machine placement algorithm to reduce

the security risk of data centers in the cloud environment. At the same time, the utilization efficiency of CPU, memory, disk storage resources, and network bandwidth is considered. Due to the rapid development of 5G technology in recent years, some scholars have considered the issue of virtual machine placement in the mobile edge network environment. During the operation of the virtual machine, resource requirements are constantly changing, so some researchers have proposed a placement algorithm based on the demand forecast model for the dynamic changes of demand [20–22].

This system is distributed on a large number of distributed computers instead of local computers or remote servers. At the same time, the mechanism has greatly reduced the management cost of the website system data center, but the improvement of the system resource utilization rate has greatly guaranteed the overall availability of the website system, and the reliability of the system has also been greatly improved. On the other hand, cloud storage can ensure that the core data of the system are stored at multiple points, and the cold and hot backups are automatically switched, which reduces the probability of affecting the entire system due to a single-node failure. This paper also presents an algorithm improvement plan based on the multimodel decision-making. Decisions are made by training multiple models at the same time, and the optimal strategy is selected from all decisions as the final decision. This paper selects the classic FFD algorithm and the adopted algorithm as the comparison algorithm and evaluates the performance of the proposed algorithm model. The experimental data show that the performance of the algorithm can meet expectations, and the improved algorithm based on the multimodel decision mechanism has obtained better performance.

2. Construction of an Intelligent Analysis Model for Website Information

2.1. Hierarchical Architecture of Big Data Space. In the big data space level, the user-defined map function accepts input key-value pairs. After completing the necessary data calculations, a set of intermediate key-value pairs will be generated. The framework class library will all have the intermediate value data with the same key name. It aggregated together and then passed to the reduction function through an iterator. The user-defined reduction function accepts the intermediate key name and the value set associated with it, and appropriately simplifies or merges these values to form a smaller value set [23]. Figure 1 shows the hierarchical architecture of the big data space.

Virtual Machine (VM) is a core concept in data center resource management. A virtual machine is a logical container that puts a part of the resources on the physical server together as an independent operating unit, allowing users to perform their own computing tasks on the virtual machine just like using a normal computer. Through the use of virtualization technology, each physical server can run multiple virtual machines at the same time. These virtual machines share the resources of the physical server and run

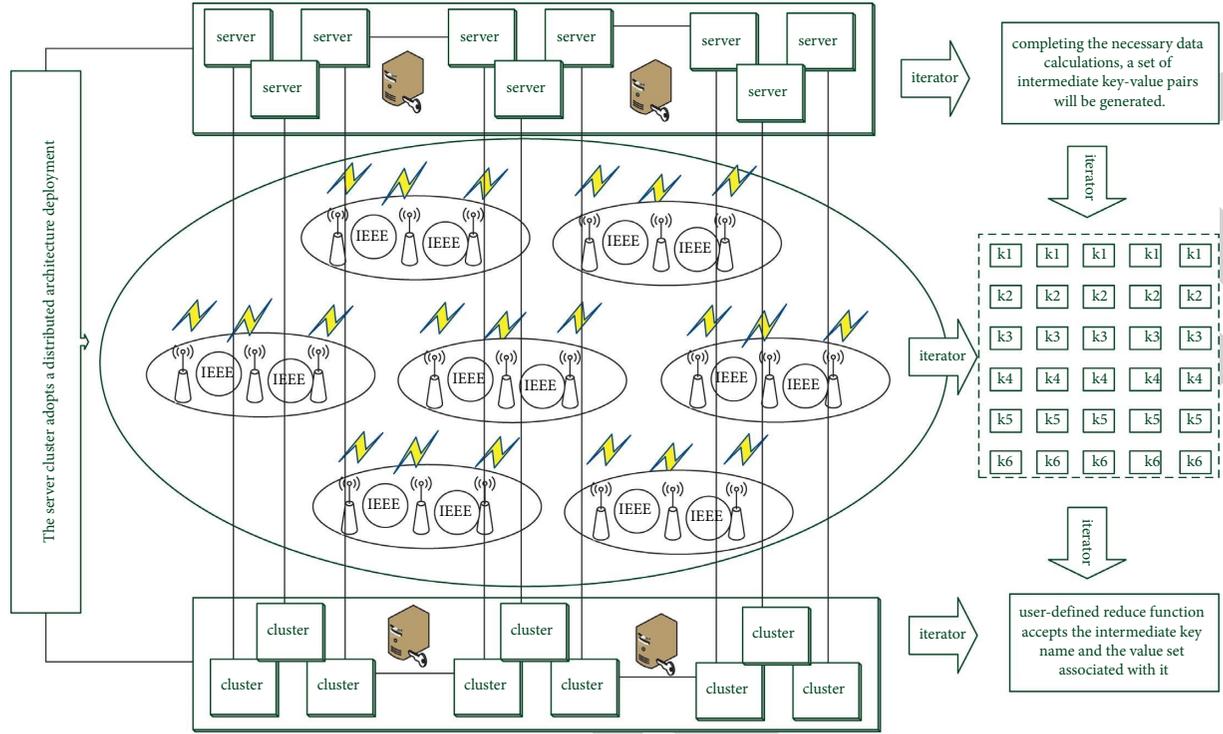


FIGURE 1: Hierarchical structure of big data space.

independently and safely, which improves the resource utilization of the physical server:

$$\sum t f(w_i, D_i) - N_{w_i, D_j} \times \sum_{n=1}^K N_{w_n, D_j} N = 0. \quad (1)$$

Cosine similarity uses the cosine value of the angle between two vectors in the vector space to judge the difference between different objects. Compared with the distance measure, Cosine similarity pays more attention to the difference of two vectors in direction, rather than distance or length. Therefore, Cosine similarity is not sensitive to absolute numerical information. The Tanimoto coefficient is also called the Jaccard similarity coefficient, which is used to compare the probability of similarity and dispersion of a concentrated sample:

$$\begin{cases} U''(\xi, t) - \Theta'(\xi) \leq 0, \\ U_{FR}''(\xi, t) - \Theta_{F1}'(\xi) \leq 0, \\ U''(\xi, t) - S'(\xi) > 0, \\ U_{FR}''(\xi, t) - S_{F1}'(\xi) > 0. \end{cases} \quad (2)$$

Pearson correlation is also called product-difference correlation (PPMCC or PCCs). This algorithm is used to measure the correlation (linear correlation) between two variables X and Y . It is a method of calculating linear correlation, and its value is between -1 and 1 . The Pearson coefficient ranges from 0.8 to 1.0 for very strong correlation, 0.6 – 0.8 for strong correlation, 0.4 – 0.6 for moderate correlation, 0.2 – 0.4 for weak correlation, 0.0 – 0.2 for very weak or no correlation, and 0 for irrelevant, less than 0 means

negative correlation. The relevant and necessary explanations are given in

$$F''(\xi, S) + F'(\xi, S) \times \Theta(\xi) - w^2(a - b\varphi a) = 0. \quad (3)$$

After obtaining the similarity of users, users can be grouped by threshold-based neighborhoods and K -neighborhoods algorithms to find users with similar interests among a large number of users and make them neighbors, and then combine them into one according to the websites they are interested in. The sorted directory is recommended to these users first.

2.2. Comprehensive Output of Cloud Computing Technology. Cloud computing provides a flexible mechanism that unifies and abstracts basic resources such as bandwidth, storage, computing power, and software packages into a computer resource sharing pool, and only needs to pay the polar management and configuration costs to quickly call these resources.

In the era of big data, massive amounts of data are the objects processed by the system, and the cloud computing platform is the tool for processing big data. The MADS system server cluster plan adopts the Feitian's large-scale distributed computing system independently researched and developed by cloud. The system integrates IT resources through virtualization technology and has functions such as elastic computing, distributed storage, automatic failure recovery, and network attack prevention, which can simplify develop and deploy the process, reduce operation and maintenance costs, and adopt cloud service architecture

systems such as ECS cloud server, CDN content distribution network, SLB load balancing, RDS database, and OCS open caching to build an on-demand network architecture, which can quickly respond to all kinds of complex website delivery requirements. Figure 2 shows the comprehensive output process of cloud computing technology.

This system realizes the extraction of web page information. At the same time, combined with the formal description model of web page information proposed in this article, it realizes the organization of data after web page extraction. The system is mainly divided into the following functional modules: information extraction module, web preprocessing module, DOM tree structure processing module, visual information processing module, page information block processing module, and hyperlink processing module. The placement algorithm is to provide a solution that attempts to divide a given sequence of virtual machines into many subsets, and the virtual machines in each subset are placed on the same physical server.

This research considers the problem of virtual machine placement under homogeneous servers. It is assumed that the resource configuration of any physical server in the data center DC is the same, and the bandwidth between any two physical servers is the same. Therefore, the placement of virtual machines can be reduced to the division of virtual machine sets. The relevant and necessary explanations are given in

$$\oint R * \frac{R_{FR} + R_{F1}}{R_{FR}R_{F1}} dx dy - \int_0^{\infty} \left(R_{FR} + R_{F1} - \frac{R_{FR} + R_{F1}}{R_{FR}R_{F1}} \right) dt = 0. \quad (4)$$

After fully investigating more than ten functions such as power function, logarithmic function, and logistic function, it is found that through mathematical modeling of each function, the system uses an elliptic function most in line with the cashback operation direction of “watching video to generate value.” Incentivizing users to watch more videos and increase the number of exposures of patch websites, users can get a lot of money after watching a small number of websites, but control it within a reasonable range, as the number of videos watched increases. The amount of cash back gradually decreases, but the general trend of cash back remains unchanged.

2.3. Cluster Analysis of Website Information. Among the current mainstream website information clustering technologies and algorithms, the recommendation method based on Collaborative Filtering (CF) is the most widely recognized and adopted. The collaborative filtering algorithm finds users who have similar interest orientations to the current user in the entire user group to form a user set. By analyzing the user’s interest, the system considers the evaluation of similar users for a specific information and the system’s preference for the information. Judgment can be made by evaluation, which is different from the traditional way of directly advancing based on content analysis. There are many ways to identify users: IP address identification, cookie identification, embedded SessionID, physical address

identification, and use of embedded proxy, where V is the set of points, E is the set of edges, and then the vertex set V is divided into two sets of points, namely the point set S and the point set U . S is the point set for which the shortest path has been found, and U is the point removed. The points other than the points in the set S are in the set. At the beginning, initialize a point to S . At this time, there is a source point in S . After that, only a shortest path is required to add the corresponding point, that is, the point in the point set U to the point set S , repeat this step, and follow the shortest path.

In the process of adding points in point set U to point set S , note that the size of the shortest path from the source point V to any point in the point set S is the same as the size of the shortest path from the source point V to any point in the point set U . The comparison is less than or equal to the relationship. In addition, all points should have a distance value. The distance value corresponding to each node of S should be the shortest path distance value from V to this node. The shortest path distance value of this node and the intermediate nodes in the process only include the nodes in S . Figure 3 shows a cluster analysis of website information based on big data.

The application calls the `bindService()` method to bind to the service, and the service is in the bound state at this time. The bound service will only run when there are other components and the service that realize the binding operation. A service can be bound by multiple components at the same time. When all these components and services are unbound, the service will be terminated and then destroyed. The bound service can provide a client-server interface so that only interactions between components and services can be realized, and information transmission can be operated. The relevant and necessary explanations are given in

$$\sum_{D=1} G(m, g(x, t)) \times \sum_{i=1}^n \frac{h(i, t)}{k(i, t)} = n \times \left(\frac{h(m, t)}{k(i, t)} \right). \quad (5)$$

On this basis, cross-process operations such as IPC (interprocess communication) can also be realized. The bound service belongs to the service class implementation class, which allows other applications to bind and interact with it. Developers must implement the `onBind()` callback method to provide the binding function. The `onBind()` method returns an `IBinder` object, which defines the interface of the program that the client interacts with the service. When the client binds to the service through `bindService()`, the client must provide the implementation class of the `ServiceConnection` interface, which can be used to monitor the connection between the client and the service. When the Android system creates a connection between the client and the server, it calls the `onServiceConnection()` method of the `ServiceConnection` interface to send the `IBinder` object, thereby realizing the communication between the client and the service.

2.4. Weighting of Intelligent Analysis Model. Collaborative filtering is a typical method of using collective wisdom to make intelligent analysis model predictions. The core idea of

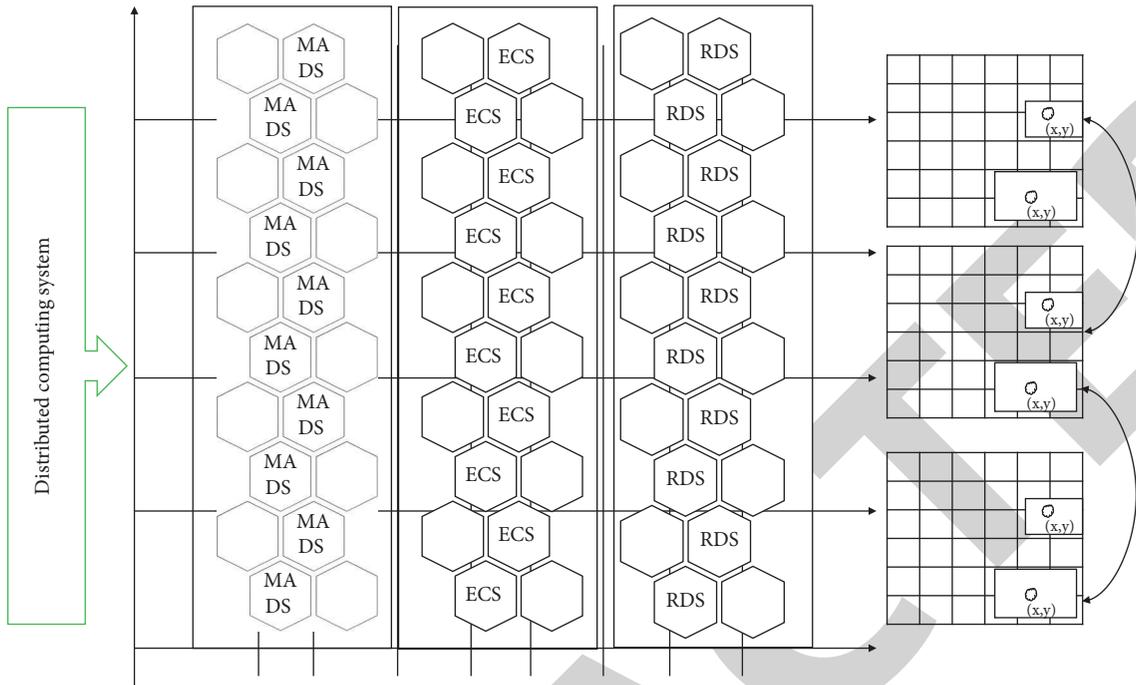


FIGURE 2: Comprehensive output process of cloud computing technology.

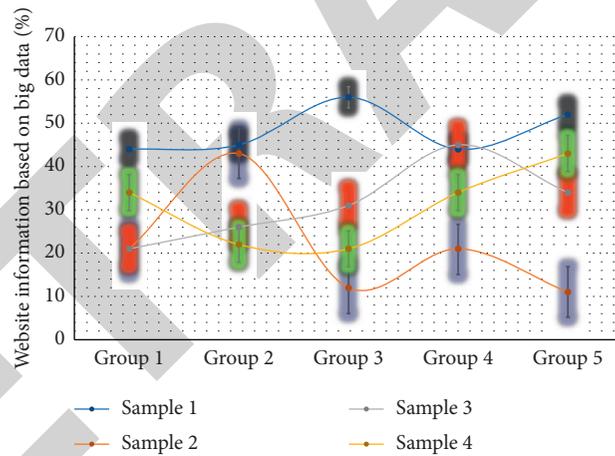


FIGURE 3: Cluster analysis of website information based on big data.

the algorithm is to make recommendations based on the interests of users and draw on the views of their related people to implement the collaborative filtering recommendation algorithm.

Collected data refer to the user’s historical behavior data. For example, the user’s viewing history, favorite behavior, like behavior, comment, etc. can be used as data for recommendation algorithms. These data include preferences that can be quantified in integers, such as website viewing. The duration may be $[0, n]$, and Boolean quantified preferences, such as whether there is website click behavior, the value may be 0 or 1. Different data have different accuracy and granularity, so the impact of noise needs to be considered. After preprocessing these data such as weighting, noise reduction, and normalization, a two-dimensional

matrix of user preferences is obtained. One dimension is the user list and the other is the website list. The value is the user’s preference for the website, generally floating point value between $[0, 1]$ or $[-1, 1]$. Finding similar users and websites is to calculate the similarity between users and websites based on the vector, that is, to calculate the distance between two vectors. Figure 4 shows the weight distribution of the intelligent analysis model.

The system information extraction module mainly extracts the HTML source code of a web page from the Internet. The source code mainly contains the HTML code part and the web page data information part, which is used as the data stream for subsequent processing. In this module, it also includes the related functions included in the information extraction. Any data of interest contained in the Content

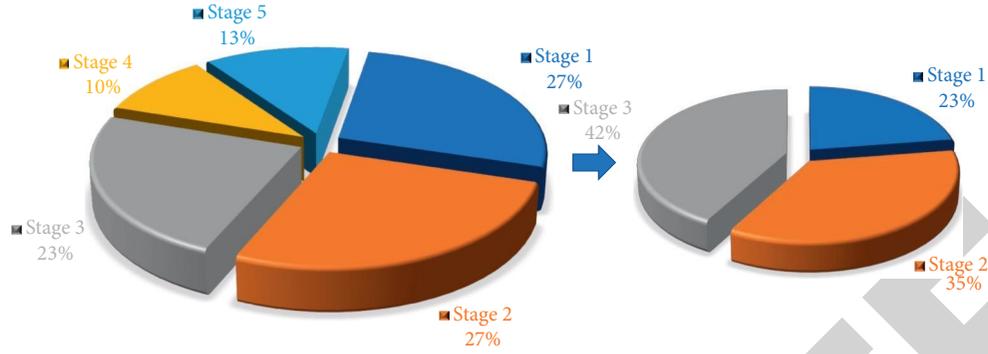


FIGURE 4: Intelligent analysis model weight distribution.

Provider can be obtained by certain methods, that is, by using the methods provided by the ContentResolver. When the query starts, the Android system not only needs to confirm the query target Content Provider and ensure that it is running, but also initialize all Content Provider objects. Under normal circumstances, each Content Provider corresponds to a separate instance, which can communicate with one or more ContentResolver class objects of different applications and processes. The communication between processes is handled by the Content Provider class and the ContentResolver class. Each Content Provider uniquely identifies its data set by using a URI, where the URI is packaged by the class. A Content Provider can also manage multiple data sets at the same time and provide a separate URI for each data set. All these URIs are prefixed with content, which is used to indicate that the data are managed by the Content Provider.

3. Results and Analysis

3.1. Cloud Computing Platform Data Rectification. The experiment in this paper is mainly divided into two parts: the algorithm experiment of web page information extraction and the verification of the formal description model of extraction result information. Among them, the web page information extraction algorithm is improved for the proposed formal description model of web page information, and the web page information extraction evaluation criteria proposed above are used to determine the accuracy of information extraction. The cloud server platform virtualizes more than 1,000 server clusters into multiple performance-configurable virtual machines (KVM). The relevant and necessary explanations are given in

$$\begin{cases} u(s, t) = \int (x(s, t) + y(s, t) + z(s, t)) ds dt, \\ v(s, t) = \int (a(s, t) + b(s, t) + c(s, t)) ds dt. \end{cases} \quad (6)$$

The performance of these virtual machines completely exceeds the physical limit of a single machine. The system is responsible for monitoring all computers in the cluster and scheduling, flexible configuration, and dynamic adjustment according to the actual use of resources, when a certain computing node needs to be maintained, it can achieve

nonstop service and nonstop, and all virtual machines on the node can be hot migrated to the cluster. At the same time, the cloud server overcomes the impact on the cluster caused by the single point of failure of the storage device and realizes the stability and high availability of the system at the storage level. The data cleaning machine is responsible for regularizing and cleaning the website access logs in the data acquisition machine according to a certain format, and the cleaned data will be stored in the cloud database. The cloud database is a relational database service that is built on SSD solid-state printing disks and is fully compatible with MySQL, SQLServer, and PostgreSQL protocols. It should adopt a master-slave dual-system hot backup architecture and have professional database backup and recovery solutions. Figure 5 shows the trend of cloud computing platform data regulation.

From the overall architecture diagram of the data analysis module, it can be seen that Redis is the link for the data exchange of the entire module and the center for data calculation. Redis's publish/subscribe mechanism makes the data results calculated by the computing nodes of each function independent and independent of each other, which makes the display of the final data results clear and clear at a glance.

When the HTTP message sending rate is lower than 20,000 per second, the memory usage of the Redis database is about 40%, while the memory usage of Node-red is only about 10%. When the arrival rate of HTTP packets reached 30,000 pieces/second, the memory usage of the Redis database also increased sharply, while the memory usage of Node-red hardly changed. After the message sending rate exceeded 35,000 pieces per second, the memory usage of Redis was as high as 89%, and the memory usage of Node-red was only 14%. The reason for this result is that almost all data processing work is concentrated in the Redis database, while Node-red is just data analysis, encapsulation, and data flow management. For this reason, it is necessary to regularly clean up the intermediate result set on the Redis server to reduce its calculation and storage pressure. The flow in the architecture diagram is specifically used to clean the intermediate result set on the Redis server.

3.2. Simulation of Website Information Intelligent Analysis Model. The experimental samples in this article are mainly for web pages of news topics on the Internet. We selected 5

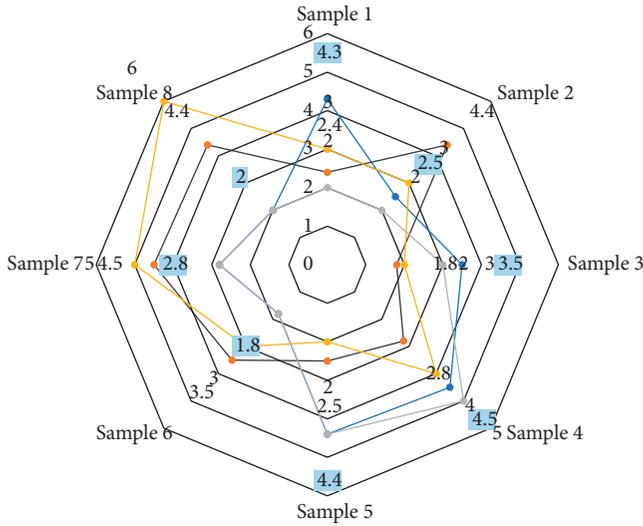


FIGURE 5: Data regulation trends of cloud computing platforms.

websites, each of which selected web pages belonging to 6 different categories, with 100 web pages in each category, for a total of 3000 web pages. We used the algorithm proposed in this paper to extract data, compared the extraction results with other algorithms, and compared the efficiency of the web page extraction algorithm. In this paper, the Python language is used to implement the above algorithm and simulation test environment. For the simulation test of the algorithm, a random generation method is used to generate the virtual machine sequence and the related resources of the data center. The server resources in the simulation are standardized.

Among them, the historical data are divided into three time attribute data to describe the time dependence. The CPU resource R_{cpu} and RAM resource R_{ram} are set to 100. The bandwidth resources provided between the servers are controlled variables that change with the experimental environment. The resource requirements of virtual machines, including CPU resource demand and RAM resource demand, as well as the communication traffic demand between virtual machines, are randomly generated. In the proposed ant colony algorithm, the loop in the algorithm jumps out after finding the optimal solution. In the simulation process, assuming that a better solution is not obtained after 100 iterations, it can jump out. The simulation test program runs on a server with Intel E5 processor and 64 G memory. Table 1 shows the description of intelligent analysis of website information.

In this experiment, the number of virtual machines is fixed to 100. In order to ensure that all virtual machines can be successfully placed, the number of physical servers is also set to 100. This article has carried out many tests, and recorded and compared the test results. The article shows the energy consumption comparison of the algorithms under different bandwidth conditions. The data show that the performance of the clustering algorithm is not as good as the ant colony algorithm and FFD when the bandwidth constraint is relatively strong, that is, when the bandwidth

resources are less. When it is less, the performance of the algorithm has no advantage compared with other algorithms, and with the increase of bandwidth resources, the performance of the algorithm in terms of energy consumption gradually manifests:

$$\begin{cases} D = \{D_1, D_2, D_3, \dots, D_k\}, \\ E = \{E_1, E_2, E_3, \dots, E_k\}, \\ F = \{F_1, F_2, F_3, \dots, F_k\}. \end{cases} \quad (7)$$

From the network consumption under different bandwidth constraints, it can be seen that the placement strategy given by the clustering algorithm generates less network traffic between the servers than the ant colony algorithm and the FFD algorithm, and with the increase of bandwidth resources, the communication in the network showed a clear downward trend. Figure 6 shows the distribution of information and communication traffic on the website.

In terms of the accuracy of information extraction, the accuracy of the algorithm in this paper is 2852 (95%), and the accuracy of the corresponding VIPS algorithm is 2791 (93%). Therefore, the algorithm proposed in this paper has a more obvious improvement than with the VIPS algorithm. It is effective to extract web page data by using the method of combining web page structure and visual features.

It is also feasible to clean and merge information blocks with text visual features as block features. Finally, based on the proposed web page information description model, we organized the results of the extracted data to make it orderly, and provided convenience for the reuse of data, such as data mining, big data analysis, and information rearrangement. In addition, the time consumed by the system experiment in this article is not only the information extraction time. However, the time consumption of the algorithm proposed in this paper is still within an acceptable range. At the same time, with the rapid development of cloud computing and distributed systems, the efficiency of the algorithm can be improved through distributed and parallel technologies.

3.3. Analysis of Experimental Results. The goal of the algorithm proposed in this paper is to place as many virtual machine tasks as possible on the limited physical server resources. Therefore, for a given virtual machine task sequence VMs, try to place the virtual machines in the physical server in sequence. A virtual machine is successfully placed, and a fixed reward value is given to the current action as the current reward.

The network input includes two parts: historical data used to model time attributes and external condition data (external) used to simulate external factors. Here, we use cProfile and RunSnakeRun tools to analyze the time performance of the program. cProfile is a time performance analysis tool that can analyze the running time of each function during program operation and generate analysis reports. RunSnakeRun is a visualization tool that can convert reports generated by cProfile into visualization charts for analysis. Taking into account the performance factors of the algorithm implementation, the test results here are for reference only. In this

TABLE 1: Description of intelligent analysis of website information.

| | Server test 1 | Server test 2 | Server test 3 | Server test 4 | Server test 5 |
|--------------|---------------|---------------|---------------|---------------|---------------|
| Weight set 1 | 1.0 | 0.2 | 0.2 | 0.5 | 0.6 |
| Weight set 2 | 0.2 | 1.0 | 0.4 | 0.1 | 0.5 |
| Weight set 3 | 0.2 | 0.3 | 1.0 | 0.7 | 0.4 |
| Weight set 4 | 0.5 | 0.4 | 0.4 | 1.0 | 0.3 |
| Weight set 5 | 0.6 | 0.5 | 0.3 | 0.7 | 1.0 |

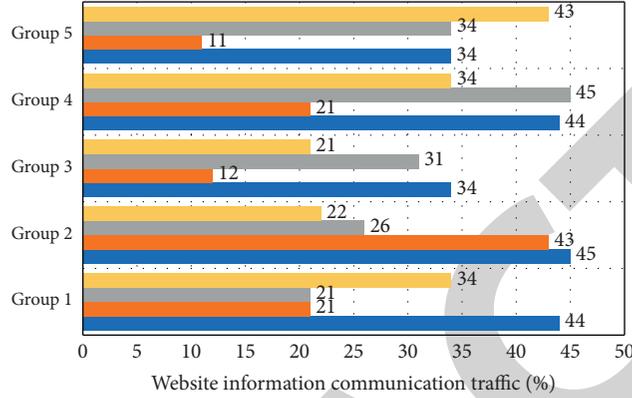


FIGURE 6: Distribution of information and communication traffic on the website.

paper, under the condition of 50 virtual machines, 50 servers, and unlimited bandwidth resources, 100 independent repeated experiments were performed, and time performance analysis was carried out with a time performance analysis tool. Figure 7 shows the information convergence based on cloud computing technology.

It can be seen that the curve convergence of multiagent is similar to that of single agent, but the result curve of multiagent is more stable, and it is significantly better than the decision result of single agent. It can be seen that in the case of a single-agent decision-making, the model can reach the same level as Fuzzy Logic after about 2000 iterations, but in the case of multiagent decision-making, the model has exceeded it after about 1,000 iterations. The relevant and necessary explanations are given in

$$E(p) = p(1 + w) * (1 - p) - p * w, \quad (8)$$

$$E(q) = q(1 - w) * (1 + q) - q * w. \quad (9)$$

The experimental test compares the effects of the multiagent model, FFD algorithm and Fuzzy Logic algorithm on CPU utilization, RAM utilization, and DISK utilization. As shown in the figure, it can be seen that the mul-DQN model is significantly ahead of the Fuzzy Logic algorithm in terms of algorithm performance, which is closer to the resource utilization of the FFD algorithm. Figure 8 shows the weight distribution of website information and communication elements.

It can be seen that according to the data elements contained in the web page information formal description model, the web page contains all data elements except the author and the comment. In the extraction result of the web page, the number of browsing times is not extracted, and the

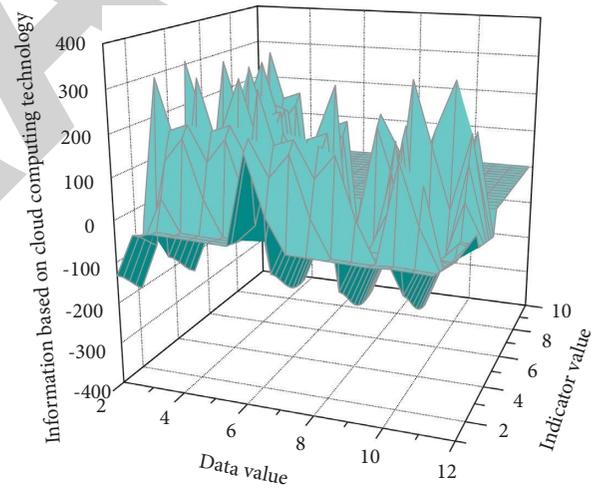


FIGURE 7: Information convergence based on cloud computing technology.

others are extracted. Therefore, according to the weight of the model elements, the accuracy of extracting information from a single page of this page is as good as the text. Considering that access traffic data are a kind of unstructured data, in order to collect effective information more accurately, it is necessary to preprocess the original data when collecting the data.

So, STResNet uses Convolutional Neural Network (CNN) as the main structure to capture spatial correlation. Because the original access traffic is the HTTP request and response messages, if these messages are only collected, they are all in the form of strings. The strings will cause problems whether in the data analysis process or in the final data visualization process. In order to facilitate analysis and

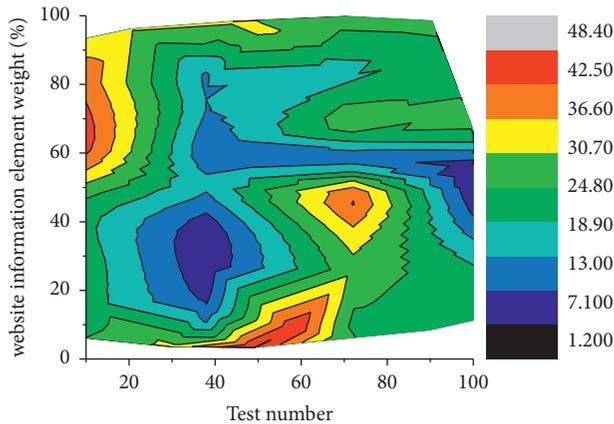


FIGURE 8: Weight distribution of website information and communication elements.

process these data better and more accurately, it is necessary to carry out preliminary structured processing. Because the data in JSON format can effectively reflect the characteristics of the data and can achieve lossless conversion with JavaScript objects, the JSON format is selected when formatting the data, and the JSON format is also selected when processing and storing the intermediate result set later.

4. Conclusion

This article briefly introduces the development background and trends of smart terminals, big data, and cloud computing in today's society. First, the principles of the Android platform are introduced and studied from two aspects. Service mainly performs background operation to provide corresponding services. Based on the comparative analysis of the Dijkstra algorithm, the Floyd algorithm, and the SPFA shortest path algorithm, this paper finally chooses the more efficient SPFA shortest path algorithm, which effectively reduces the waiting response time of users, and establishes an abstract model of the transportation network. Through a comparative analysis with the more popular AutoNavi navigation system on the market, the navigation algorithm of this system has reached the current level of navigation technology. Aiming at the lack of existing information extraction algorithms to support the formal organization of web information proposed in this paper and the shortcomings of existing extraction technologies, this paper proposes a web information extraction technology based on the VIPs algorithm for the formal organization of web information. This technology combines two aspects of DOM structure and visual features. It uses a top-down and reverse order to analyze the DOM structure. At the same time, it uses visual features and DOM structural features as the basis for information extraction. It combines label block and visual block with each other. The formal description structure of web page information classifies the blocks. For similar blocks that belong to the same formal description structure, the similar blocks are merged according to their tag path and other characteristics, and finally the important information extracted from the web page is divided into

different blocks. This technology combines the advantages of web page DOM structure and visual features and improves the accuracy of web page information extraction.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Acknowledgments

This work was supported by Science and Technology Support Project of Jiangxi Provincial Department of Education: Research on Web Data Mining Strategy Based on Cloud Computing (Subject no. GJJ204807).

References

- [1] D. Jiang, "The construction of smart city information system based on the Internet of Things and cloud computing," *Computer Communications*, vol. 150, pp. 158–166, 2020.
- [2] J. Wang, Y. Yang, and T. Wang, "Big data service architecture: a survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.
- [3] R. Massobrio, S. Nesmachnow, A. Tchernykh, A. Avetisyan, and G. Radchenko, "Towards a cloud computing paradigm for big data analysis in smart cities," *Programming and Computer Software*, vol. 44, no. 3, pp. 181–189, 2018.
- [4] J. H. Kim, "A review of cyber-physical system research relevant to the emerging IT trends: industry 4.0, IoT, big data, and cloud computing," *Journal of industrial integration and management*, vol. 2, no. 3, Article ID 1750011, 2017.
- [5] C. L. Stergiou, A. P. Plageras, K. E. Psannis, and B. B. Gupta, "Secure machine learning scenario from big data in cloud computing via internet of things network," *Computer Networks and Cyber Security*, pp. 525–554, Springer, Berlin, Germany, 2020.
- [6] P. Wazurkar, R. S. Bhadoria, and D. Bajpai, "Predictive analytics in data science for business intelligence solutions," in *Proceedings of the Communication Systems and Network Technologies*, pp. 367–370, Nagpur, India, November 2017.
- [7] C. Liu, Y. Feng, D. Lin, L. Wu, and M. Guo, "IoT based laundry services: an application of big data analytics, intelligent logistics management, and machine learning techniques," *International Journal of Production Research*, vol. 58, no. 17, pp. 5113–5131, 2020.
- [8] Y. Zhou, D. Zhang, and N. Xiong, "Post-cloud computing paradigms: a survey and comparison," *Tsinghua Science and Technology*, vol. 22, no. 6, pp. 714–732, 2017.
- [9] J. Chen, K. Li, H. Rong, K. Bilal, N. Yang, and K. Li, "A disease diagnosis and treatment recommendation system based on big data mining and cloud computing," *Information Sciences*, vol. 435, pp. 124–149, 2018.
- [10] N. Verma and J. Singh, "An intelligent approach to big data analytics for sustainable retail environment using Apriori-MapReduce framework," *Industrial Management & Data Systems*, vol. 12, 2017.

- [11] M. Elhoseny, A. Abdelaziz, A. S. Salama, A. M. Riad, K. Muhammad, and A. K. Sangaiah, "A hybrid model of internet of things and cloud computing to manage big data in health services applications," *Future Generation Computer Systems*, vol. 86, pp. 1383–1394, 2018.
- [12] X. Yao, G. Li, and J. Xia, "Enabling the big earth observation data via cloud computing and DGGs: opportunities and challenges," *Remote Sensing*, vol. 12, no. 1, Article ID 62, 2020.
- [13] A. Singh, S. Kumari, H. Malekpoor, and N. Mishra, "Big data cloud computing framework for low carbon supplier selection in the beef supply chain," *Journal of Cleaner Production*, vol. 202, pp. 139–149, 2018.
- [14] S. Jeble, S. Kumari, and Y. Patil, "Role of big data in decision making," *Operations and Supply Chain Management: International Journal*, vol. 11, no. 1, pp. 36–44, 2017.
- [15] S. M. Wu, T. Chen, and Y. J. Wu, "Smart cities in Taiwan: a perspective on big data applications," *Sustainability*, vol. 10, no. 1, Article ID 106, 2018.
- [16] M. H. U. Rehman, E. Ahmed, I. Yaqoob, I. A. T. Hashem, M. Imran, and S. Ahmad, "Big data analytics in industrial IoT using a concentric computing model," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 37–43, 2018.
- [17] K. Zhan, "Sports and health big data system based on 5G network and Internet of Things system," *Microprocessors and Microsystems*, vol. 80, Article ID 103363, 2021.
- [18] B. M. Balachandran and S. Prasad, "Challenges and benefits of deploying big data analytics in the cloud for business intelligence," *Procedia Computer Science*, vol. 112, pp. 1112–1122, 2017.
- [19] A. Mohamed, M. K. Najafabadi, Y. B. Wah, E. A. K. Zaman, and R. Maskat, "The state of the art and taxonomy of big data analytics: view from new big data framework," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 989–1037, 2020.
- [20] D. Blazquez and J. Domenech, "Big Data sources and methods for social and economic analyses," *Technological Forecasting and Social Change*, vol. 130, pp. 99–113, 2018.
- [21] R. H. Tsaih and C. C. Hsu, "Artificial intelligence in smart tourism: a conceptual framework," *Artificial Intelligence*, vol. 2, Article ID 23, 2018.
- [22] V. Chang, "A proposed social network analysis platform for big data analytics," *Technological Forecasting and Social Change*, vol. 130, pp. 57–68, 2018.
- [23] Q. Zhang, L. T. Yang, and Z. Chen, "PPHOPCM: privacy-preserving high-order possibilistic c -means algorithm for big data clustering with cloud computing," *IEEE Transactions on Big Data*, vol. 21, pp. 145–156, 2017.