Research Article

Research on Intelligent Retrieval Method of Teaching Resources on Large-Scale Network Platform

Xiaofeng Su and Tianjing Zhang

1Xi’an Siyuan University, Xi’an, Shaanxi 330022, China
2Hebei Academy of Fine Arts, Shijiazhuang, Hebei 050000, China

Correspondence should be addressed to Tianjing Zhang; tianjing@hbafa.edu.cn

Received 12 April 2022; Revised 7 May 2022; Accepted 11 May 2022; Published 27 May 2022

Abstract

With the increase in information on various cloud computing platforms, there are more and more teaching documents and videos, which provide sufficient resources for people to learn. Facing the large-scale digital teaching resources, how to quickly and accurately retrieve the required content has become an important research direction in the information field. Especially in the face of heterogeneous, dynamic, and large-scale teaching resources stored in the cloud computing platform, the traditional cloud computing resource retrieval has poor performance and low work efficiency. To solve this problem, a cloud computing platform retrieval method based on genetic algorithm is proposed, which is suitable for intelligent retrieval of teaching resources. Firstly, the teaching resource storage system based on cloud computing platform is analyzed, and the overall architecture of the system and the network topology of cloud storage data are given. Then, a resource retrieval method suitable for cloud computing platform is designed by genetic algorithm, and the convergence performance of genetic algorithm is improved by ant colony algorithm. Finally, the selection algorithm in genetic algorithm is optimized by using random numbers and increasing the number of cycles. The experimental results show that the proposed intelligent retrieval method has greatly improved the Recall and Precision compared with the traditional retrieval methods.

1. Introduction

With the increase in information on various cloud computing platforms, there are more and more teaching documents and videos. Different from local storage, users using cloud data can greatly improve work efficiency and reduce hardware investment costs [1–4]. Many products and services based on cloud computing are constantly being introduced, and the scale and fields involved in the computer industry are constantly expanding. “Cloud Computing-Aided Instruction” (CCAI) has become a new means for colleges and universities to set up modern teaching [5–7], which has effectively improved the teaching quality technically. The features of CCAI mode are very beneficial to the information management of teaching, reduce the capital investment and maintenance costs, improve the network security, and help to build a personalized teaching environment.

However, with the continuous growth of digital teaching resources on the cloud computing platform, how to quickly and accurately retrieve the required content has become an important research direction in the information field [8, 9]. In most cases, these network resources are unorganized, or each has a different organizational structure, which brings a lot of pressure for users to inquire about resources. Although the emergence of search engines has eased the pressure of resource inquiry, most search engines are not satisfactory in recall and precision. Most of the time, users cannot find the resources they need from a large number of inquiry results. In this case, the user experience is poor, and the user still has not got rid of the trouble of too much information. Web information retrieval belongs to the category of information retrieval...
retrieval and is an important development stage in the field of information retrieval.

Genetic algorithm [10, 11] is a globally optimized intelligent probability search algorithm developed by referring to the natural selection and genetic evolution mechanism of organisms. The genetic algorithm is an effective method for finding the optimal solution in a large solution space. Searching for an optimal query in large-scale information retrieval system can also be regarded as a problem of searching for the optimal solution in a large solution space. Therefore, how to apply the retrieval method based on the genetic algorithm to the cloud computing teaching platform to improve the retrieval effect is the key research content of this paper. The results show that the genetic algorithm is effective in query optimization, and it can overcome the shortcomings of low Recall and Precision of the retrieval system, so that users can accurately and efficiently obtain the required network teaching resources.

2. Related Works

At present, the resource retrieval methods of cloud computing teaching platform are mostly based on manual classification or keyword matching technology [12, 13]. These two retrieval methods have not optimized the user’s query requirements, which lead to the unsatisfactory retrieval results of these teaching platforms.

There are many disadvantages in the retrieval method of manual classified catalog. The first is inefficient. Administrators of resource management systems need to upload resources based on manually categorized directories. However, once there is any objection to the manual classification catalog, the administrator shall be contacted to modify relevant catalog. The second is poor compatibility. Resources in one system are hard to reuse in another. To use these resources, the administrator needs to enter them one by one in another system. If we want to overcome these shortcomings, we need to provide a unified resource storage method, and the resource storage method based on cloud computing platform is a good solution.

The retrieval method based on keyword matching has great limitations in the semantic disclosure of information, and it is difficult to guarantee the accuracy and precision of information. The retrieval system simply matches the keywords entered by the user. Many resources that should be retrieved are not retrieved, while resources that should not be retrieved are retrieved. This requires query optimization. Global analysis is an early query optimization method with practical application value. Roui [14] proposed a global analysis method based on Latent Semantic Indexing, which realized effective semantic clustering and topic sorting of web documents. However, when the document set is very large, it is often infeasible in time and space to establish a global dictionary of word relations, and the update cost after the document set changes is huge.

At present, the popular local analysis methods mainly include Relevance Feedback and Pseudo Feedback. Pseudo Feedback is developed on the basis of Relevance Feedback. Relevance Feedback is a very important mechanism for query optimization in information retrieval. Because of the remarkable effect of relevant feedback, it has been widely applied and studied in information retrieval. Zhang et al. [15] proposed a method to improve the query effect by using relevant feedback. This method expands and shrinks the query at the same time, thus obtaining a high recall rate. Pseudo-relevance feedback does not need to interact with users. It directly regards the first N documents retrieved by the first query as relevant documents and optimizes the query based on this. Wang et al. [16] proposed a pseudo-relevance feedback framework for information retrieval, which combines relevance matching and semantic matching. However, the selection of keywords in Pseudo Feedback is more important. Generally speaking, keywords with higher weights are selected for query expansion. This selection method ensures the importance of keyword selection, but it does not guarantee that keywords are related to the topic.

Although information retrieval technology has made some progress, the performance of retrieval engines in large-scale network platforms still cannot meet users’ expectations. Because of the huge retrieval data set and the diversity and complexity of the factors that affect retrieval efficiency, the above optimization techniques are not ideal in practical application. The introduction of the genetic algorithm provides a new way to solve information retrieval problems. Therefore, a cloud computing platform retrieval method based on the genetic algorithm is proposed. The main innovations and contributions are as follows: (1) try to apply the genetic algorithm, which is suitable for finding the best solution in large space, to retrieval optimization, and design a resource retrieval method suitable for Spark platform, so as to overcome the low Recall and Precision of the retrieval system; (2) the ant colony algorithm is used to improve the convergence performance of the genetic algorithm, and the selection algorithm in the genetic algorithm is optimized by using random numbers and increasing the number of cycles.

3. Teaching Resource Storage System Based on Cloud Computing Platform


Cloud computing is a research hotspot in computer science and technology at present, which has attracted the attention of many enterprises and related Internet experts, and is an important trend of computer network technology development in the future. The concept of cloud computing was first put forward by Ehrlich Schmidt, CEO of Google Inc., at the Internet Conference in 2006. A typical cloud computing platform needs to have (1) a gridded data storage matrix network; (2) firewall equipment; and (3) computing resource equipment, allowing users to remotely use an expandable cloud storage space by leasing, to realize cloud application services [17], as shown in Figure 1.

A complete cloud computing architecture should include access layer, core layer, resource convergence layer,
API interface layer, and application layer [18], as shown in Figure 2.

3.2. Cloud Storage System Network Topology. The teaching resource system under CCAI mode needs to meet the requirements of all-weather, all-geographical, and all-connection. In this paper, C/S mode [19] is adopted to construct the service system architecture of network teaching resources, and all data are stored in the data server, as shown in Figure 3. In the teacher’s office, upload or access the online teaching server through the campus network. Students on campus can access learning resources through campus network in dormitory or library. On the other hand, off-campus personnel can also remotely access the training and learning resources through the Internet, thus realizing the efficient sharing of limited teaching resources, breaking the geographical space limitation, and reducing the input cost of manpower and material resources.

At present, there are many excellent learning resource banks, some of which are all open, and the construction of these network resource banks has laid the foundation for the improvement of network education. However, these learning resources have a disadvantage; that is, they are difficult to be compatible with each other, that is, different systems have different learning resources, and the construction standards of these resources are different, so they cannot share resources. If you want to use the resources of another system in one system, you need to rebuild the resources according to the resource construction scheme of this system. In this situation, the learning resource pool has not been shared in the real sense.

4. Intelligent Retrieval of Teaching Resources Based on Genetic Algorithm

4.1. Design of Resource Retrieval Method Based on Genetic Algorithm. As mentioned above, faced with the heterogeneous and large-scale teaching resources stored in the cloud computing platform, the traditional cloud computing resource retrieval has poor performance and low work efficiency. Therefore, this paper uses the genetic algorithm to realize the retrieval of cloud computing resources. First, suppose that there are $m$ hosts $H$ in the resource retrieval task, and $n$ virtual machines $V$ are installed on these hosts, and each genetic individual is coded $\{k_0, k_1, \ldots, k_{n-1}\}$ by coding mapping [20–22]. For example, as shown in Figure 4, in the mapping relationship between virtual machines and hosts, if the sequence length is 5, then the number of 0–5 of $V$ is $\{1, 0, 2, 0, 2\}$. The number in the sequence is the number of host $H$, and then the population is initialized.

Let the total number of constituent objects of a retrieval task be $N$ and the fitness of each of the $N$ constituent objects be $f_i$. Then, the probability of the $i$-th object being selected for evolution is as follows:

$$P = \frac{f_i}{\sum_{i=1}^{N} f_i}$$  \hspace{1cm} (1)

Let the position change of the retrieval task in a certain period of time be $\delta(H)$. While the probabilities of selection crossover and change of the genetic algorithm are $P_c$ and $P_m$, respectively, the expected value of the next generation belonging to the dynamic process of retrieval task is as follows:
Virtual machine 0  
Virtual machine 1  
Virtual machine 2  
Virtual machine 3  
Virtual machine 4

**Figure 4**: Schematic diagram of virtual machine and host mapping.

\[ E[m(H, t + 1)] \geq m(H, t) \cdot \frac{f(H, t)}{\tilde{f}(t)} \left( 1 - P_c \cdot \delta(H) \right) \left( 1 - O(H) \cdot P_m \right) \]

(2)

where \( O(H) \) is the dynamic order of task [23]. The longest distance of transmission is \( L \), and \( m(H, t) \) is the number of objects in the next generation that need to be transmitted by the retrieval task. \( f(H, t) \) and \( \tilde{f}(t) \) are the fitness and average fitness of the next generation of objects that need to be retrieved.

In the process of retrieval task, in order to ensure the integrity of the object and prevent the local data loss due to the change of retrieval task, the probability of selecting crossover operation must satisfy the following formula:

\[ P_s \geq 1 - P_c \cdot \delta(H) \frac{1}{L - 1} \]

(3)

Then, according to formulas (2) and (3), we can get the following:

\[ E[m(H, t + 1)] \geq m(H, t) \cdot \frac{f(H, t)}{\tilde{f}(t)} - P_c \cdot \delta(H) \frac{1}{L - 1} \cdot \left( 1 - O(H) \cdot P_m \right) \]

(4)

In formula (4), generally, the value of \( P_m \) is very small and then formula (4) can be further optimized to obtain as follows:

\[ (1 - P_m)^O(H) = 1 - O(H) \cdot P_m \]

\[ \left( 1 - P_c \cdot \delta(H) \frac{1}{L - 1} \right) \left( 1 - O(H) \cdot P_m \right) \geq 1 - P_c \cdot \delta(H) \frac{1}{L - 1} - O(H) \cdot P_m \]

(5)

If \( f(H, t)/\tilde{f}(t) > C \) (\( C \) is constant), it means that the operation has not reached the optimal solution calculated by the algorithm. Let \( K \) be given as follows:

\[ K = C \cdot \left( 1 - P_c \cdot \delta(H) \frac{1}{L - 1} - O(H) \cdot P_m \right) \]

(6)

If \( K > 1 \), there are:

\[ E[m(H, t + 1)] \geq m(H, t) \cdot K. \]

(7)

From this, we can recursively get the following:

\[ E[m(H, t + 1)] \geq m(H, 0) \cdot K. \]

(8)

After the object of the retrieval task is iteratively calculated by the genetic algorithm, the position change of the resource object required by the retrieval task in a certain period of time can be obtained. During the training of position change, the ant colony algorithm is used to improve the convergence performance of the genetic algorithm.

4.2. Genetic Algorithm after Ant Colony Optimization. Let the number of ants in the nest be \( R \) and the set of elements to be optimized be \( D_i \), where \( D_i \) represents the \( i \)-th element. In order to solve the initial population problem, the number of all parameters to be optimized in this paper is \( n \). Assuming that there are \( K \) possible values of these elements \( \phi \), then \( \zeta^j(D_\phi) \) is the pheromone of the \( j \)-th element under the initial condition.

According to formula (9), the \( t \)-th ant calculated its parameters to distinguish the probability of each possible value [24–26].

\[ k(\zeta^j(D_\phi)) = \frac{\zeta^j(D_\phi)}{\sum_{i=1}^{n} \zeta^i(D_\phi)} \]

(9)

Then, elements are selected from the set \( D_\phi \) with high probability and adjusted according to the following formula:

\[ \zeta_j(D_\phi)(t + \Delta t) = \zeta_j(D_\phi)(t) + \Delta \zeta_j(D_\phi) \]

(10)

where \( \Delta \zeta_j(D_\phi) \) is the information increment on element \( \phi \), representing the sum of pheromones left by all ants passing through this element. Its calculation method is as follows:

\[ \Delta \zeta_j(D_\phi) = \sum_k \Delta \zeta_k^j(D_\phi) \]

(11)

The above process was repeatedly performed until the maximum allowed number of iterations was reached, or all the ants could obtain the unique element, thus obtaining the optimized initial population-related parameters.

After the initial population is generated by the ant colony algorithm, it is necessary to continue the genetic operation. The main contents of genetic operation are selection operator, crossover operator, and mutation operator. The operation of the traditional genetic process will lead to
4.3. Design of Fitness Function. In order to achieve the performance balance (reduce the energy consumption) on the premise of improving the work efficiency, this paper combines the service quality constraint and the energy consumption constraint to construct the fitness function. Among them, the total QoS violation $Q_{\text{total}}$ of virtual machines is calculated as follows:

$$Q_{\text{total}} = 1 - \frac{\text{MIPS}_{\text{total,L}} - \text{MIPS}_{\text{total,M}}}{\text{MIPS}_{\text{total,L}}}$$

where MIPS$_{\text{total,L}}$ and MIPS$_{\text{total,M}}$ are all allocated millions of instructions per second and those that are not allocated on time, respectively.

Total system energy consumption $E$ is calculated as follows:

$$E = \sum_{i=0}^{m-1} \text{Host}_i,$$

where Host$_i$ is the energy consumption of the $i$-th host in cloud computing retrieval. A double index constraint composed of quality and cost is adopted as the fitness function. The fitness function is defined as follows:

$$\text{fitness} = 1 - a \times Q_{\text{total}} - b \times E,$$

where $a$ and $b$ are the weights corresponding to service quality violations and total energy consumption, respectively.

5. Experimental Results and Analysis

5.1. Experimental Setup. In order to test the performance of the proposed retrieval method based on the genetic algorithm, it is compared with Pseudo Feedback and extended retrieval method based on local context analysis (LCA). Experimental data were from the CISI test set. The CISI test set is a test set on information science, which consists of 1460 documents and 112 searches. The test set source url is http://www.dcs.gla.ac.uk/idom/irresources/test_collections/. The test set contains the full text of the document, the retrieved initial text, and a list of document relationships. In the list of retrieved and document relationships, each retrieved related document has been given.

Each document and the initial retrieval are preprocessed (stop words eliminated). A word stem extraction algorithm is adopted to extract the word stem and establish a keyword dictionary. Extract keywords from the dictionary and calculate their weights. At the same time, the retrieval of each document is vectorized. The cosine similarity calculation method is adopted to calculate the similarity between the initial retrieval and the documents, and the documents are sorted in descending order according to the size of the similarity. The more advanced the document is, the closer it is to retrieval. In genetic algorithms, generally speaking, selecting a larger initial population can handle more solutions at the same time, so it is easy to find the global optimal solution. The disadvantage is that it increases the time of each generation selection [27, 28], so the population size is generally 20–100. In the optimization process, the crossover probability always controls the crossover operator which plays a dominant role in genetic operations. The crossover probability controls how often crossover operations are used. The higher the frequency is, the greater the probability for each generation to produce new individuals is, and the better the diversity of the population is, and the faster it can converge to the optimal solution region. However, too high a frequency may also lead to premature convergence, generally taking the value of 0.4–0.9. When the maximum evolutionary algebra is used as the termination condition of the genetic algorithm, it is generally between 100 and 500 generations. In the experiment of this paper, the setting parameters of the genetic algorithm were as follows: initial population was 30, crossover probability was 0.4, mutation probability was 0.3, and maximum evolution algebra was 100.

5.2. Evaluation Indicators. Recall and Precision are widely used evaluation criteria of Web information retrieval effect [29]. Recall is the ratio of the number of relevant documents retrieved to all relevant documents in the document collection, and Precision is the ratio of the number of relevant
documents retrieved to the total number of documents retrieved. Recall and Precision are defined as follows:

\[
\text{Recall} = \frac{\text{number of related documents retrieved}}{\text{all related documents in the document set}} \times 100\%,
\]

\[
\text{Precision} = \frac{\text{number of related documents retrieved}}{\text{total number of documents retrieved}} \times 100\%.
\]  

(15)

In search engines, the first 10 or 20 documents usually reflect the results of the first page and the first two pages. Therefore, this paper uses the Recall and Precision of the first 10 or 20 retrieved documents as the evaluation indicators.

5.3. Result Analysis. Figures 5 and 6 show the Recall and Precision (the first 10 documents) of 10 different searches using three different algorithms, respectively. Note that, as mentioned in the previous section, only the first 10 documents retrieved are counted here.

Figures 7 and 8 show the Recall and Precision (the first 20 documents) of 10 different searches using three different algorithms, respectively. Table 1 shows the comparison results of retrieval performance of the three algorithms.

As can be seen from Table 1, compared with Pseudo Feedback and local context analysis-based retrieval extension method (LCA), the retrieval method based on the optimized genetic algorithm has higher Recall and Precision, that is, better retrieval performance. The number of false feedback detected is very large, and the proportion of the relevant literature is relatively low; that is, its Precision is relatively low. This retrieval mode has poor user experience, and users need to find the information they need by themselves from a large number of check-out results. However, the algorithm proposed in this paper does a good job in this respect, and most of the checked-out documents are related to the retrieval topic; that is, most of the checked-out results are the information that users need. It does not take users too much time to pick and choose the materials they need from the results.

6. Conclusions

Aiming at the poor performance of traditional cloud computing resource retrieval, this paper proposes a cloud computing platform retrieval method based on the genetic algorithm. The genetic algorithm is used to design a resource retrieval method suitable for cloud computing platform, so as to overcome the problems of low Recall and Precision. In
addition, the ant colony algorithm is used to improve the convergence performance of the genetic algorithm, and the selection algorithm in the genetic algorithm is optimized by using random numbers and increasing the number of cycles. The proposed method has achieved good retrieval performance in Recall and Precision, which verifies its feasibility. However, because there are not enough teaching resources in this system, the performance of the system has not improved much. If more types of teaching resources such as video resources and audio resources can be provided, the Recall and Precision of the proposed retrieval method will be obviously improved.

**Data Availability**

The experimental 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 conflicts of interest to report regarding the present study.

**References**


