Scholar Publishing Group
Distributed Processing System
https://doi.org/10.38007/DPS.2022.030307
ISSN 2790-0916 Vol. 3, Issue 3: 53-60

Resource Allocation of Distributed System Based on Hierarchical Clustering Algorithm

Amarine Charu*

American University of Afghanistan, Afghanistan

**corresponding author*

Keywords: Hierarchical Clustering, Clustering Algorithm, Distributed System, Resource Allocation

Abstract: The resource allocation of distributed systems has the properties of heterogeneous multi-source, autonomous peer-to-peer, and autonomous control, which is completely different from traditional computer resources. In order to use the resources of distributed systems better and more efficiently, it becomes crucial to design and implement a way to combine distributed systems and hierarchical clustering algorithms. In order to solve the problem of resource allocation in the existing distributed system, this paper summarizes the operation process of the allocation of data resources based on the centralized and distributed resource allocation strategy of system resources and the hierarchical clustering algorithm. Design and discuss the experimental environment and tools for resource allocation of distributed system based on hierarchical clustering algorithm, and design the computational flow chart of distributed font resource allocation by using hierarchical clustering algorithm. The processing time of resource allocation in the system is compared with other algorithms for experimental data analysis. The experimental data shows that the processing time of the resource allocation of the distributed system based on the hierarchical clustering algorithm is less than that of the other three algorithms, and the fastest processing time can reach 400 tasks. It takes 174.45 milliseconds, and the processing time gradually decreases with the increase of the task volume, so it verifies the superiority of the hierarchical clustering algorithm in the allocation of distributed system resources.

1. Introduction

The problem of resource allocation of the system has a very important position in both the traditional system theory and the current theory of centralized and distributed systems. Since the resource allocation problem of the system is essentially the process of establishing the relationship

between resources and users, on the basis of reasonably extracting data models, completely different system resource allocation problems can be represented in a similar abstract form.

Nowadays, more and more scholars conduct research on the practical application of hierarchical clustering algorithm and other algorithms in the resource allocation of distributed systems, and have achieved certain research results. By classifying a group of objects and making the shuffled objects more similar to other groups of objects, Botre M K can provide a search function for unclassified data, Botre M K analyzed in the study A variety of clustering methods are used, and the analyzed rice is classified, mainly including network clustering, hierarchical clustering, dense clustering, etc. Botre M K also adds the proposed clustering algorithm to the function. classification combination, thereby improving the performance of the algorithm [1]. Prasad A S proposes a multi-source distributed management system by classifying the sources of various energy management systems, and enables the proposed system to function in various power and circuit loads, and uses a clustering algorithm Calculations are performed to improve the support of experimental data for the proposed management system, and Prasad A S evaluates the performance in existing power and grids to varying degrees with the proposed distributed energy management system [2]. Hossen MB studied two important methods proposed in the experiment, and analyzed the two methods based on the robust queuing method. It mainly includes the random allocation of resources. On the basis of this idea, Hossen MB proposed another algorithm for finding specific cases of this method. Through real case data experiments, Hossen MB found that compared with the original algorithm, the algorithm proposed by Hossen MB can reduce the supply of resources by 30%, and can reduce resource waste by up to 60% in the allocation and utilization of resources. At the same time, Hossen MB calculates RConf based on the original greedy algorithm. Optimality is controlled to determine an approximation of the optimal solution for resource allocation [3]. Although the existing research on resource allocation of distributed systems is very rich, the research on resource allocation of distributed systems based on hierarchical clustering is still insufficient.

Therefore, in order to solve the existing problems in the resource allocation of distributed systems based on hierarchical clustering, this paper uses the hierarchical clustering algorithm for design and application. First, the steps of the data resource allocation function equation of the hierarchical clustering algorithm are introduced. And the two methods of resource allocation of the system, secondly, the experimental tools and parameter settings of the resource allocation of the distributed system based on the hierarchical clustering are discussed, and finally the process structure of the resource allocation of the distributed system based on the hierarchical clustering is designed, and By comparing the processing time of the distributed system based on hierarchical clustering with the other two algorithms through experimental data, the final experiment shows the effectiveness of the proposed multi-level clustering based distributed system in resource allocation.

2. Research on Resource Allocation of Distributed System Based on Hierarchical Clustering Algorithm

2.1. System Resource Configuration Method

(1) Centralized system resource configuration

In the centralized resource allocation mode, there are one or more central control layers, and the central control layer is responsible for collecting and saving the data resource information of other layers in the system [4]. The central control layer communicates with the ordinary layer, but the ordinary layers do not communicate. When the resource task configuration is to be performed, an instruction needs to be sent to the central control layer, and the central control layer makes a

decision and selects a suitable layer for the task. received [5].

(2) Distributed resource allocation strategy

The resource allocation strategy adopted in the resource allocation in this paper is the distributed resource allocation strategy. When you need to query the resource configuration of other layers, you can get it directly on the layer, unlike the centralized management method, you need to go to the central control node to adopt a distributed resource configuration strategy, which will not exist in the centralized strategy. The single point failure problem [6]. Although it will cause information redundancy, because this paper adopts a non-fixed period resource collection method, it can reduce the additional traffic generated by load information to a certain extent [7].

2.2. Hierarchical Clustering Algorithm

In hierarchical clustering analysis, because the initially acquired data may belong to different categories, it is impossible to directly calculate or compare the data [8]. Therefore, the initial data resources must be transferred to the same sub-category before the direct calculation or comparison of data can be gradually realized. The normalization of data resources is to compress or expand the original data resources according to the specified ratio, and finally standardize the original data within one to ten [9].The definition of the original resource for normalization of data resources is as follows: Suppose domain $G=\{u_1,u_2...u_k\}$ is all objects to be clustered, if each object u_x has a y characteristic, then the characteristic vector of object u_x is defined as follows:

$$u_x=(u_{x1},u_{x2},...u_{xy})(x=1,2,...k) \tag{1}$$

General data resource normalization methods include min-max normalization and Z-score normalization [10].

(1) min-max normalization

The min-max normalization method needs to select data $u_{\max}^n$ and data $u_{\min}^n$ in the n-th feature of the original data u_x [11]. The specific formula is as follows:

$$u_{xn}^*=\frac{u_{xn}-u_{\min}^n}{u_{\max}^n-u_{\min}} \tag{2}$$

(2) Z-score normalization

Z-score normalization Before the standard deviation normalization of all the original data u_{xn}, the mean ϑn and standard deviation u_{xn}^* of the original data of all the resources must be obtained [11]. The specific normalization formula is as follows:

$$u_{xn}^*=\frac{u_{xn}-\bar{u}n}{\vartheta n} \tag{3}$$

Where k is the clustered resource data.

$$\bar{u}_n=\frac{1}{k}\sum_{1}^{k}u_{xn} \tag{4}$$

3. Research on Resource Allocation of Distributed System Based on Hierarchical Clustering Algorithm

3.1. Resource Configuration Parameter Setting of Distributed System Based on Hierarchical Clustering Algorithm

In the experiment, the CloudSim library function is used to randomly generate a data set, and the number of tasks $k = \{100,200,300,400\}$ is set to take the average results of 200 task processing times for comparison [12]. In the experiment, the specific experimental configuration of the resource allocation of the distributed system using the hierarchical clustering algorithm is as follows:

(1) Task parameters: the task length is between $[500-4000]$ [13], the bandwidth range $[1000-2000]$, and the storage range is between $[512-2064]$ [14].

(2) Specific parameters of the VM are shown in Table 1:

Table 1. Vm parameters

Vm Configuration	Number of CPU	CPU speed range	Bandwidth range	Store the scope
Parameter configuration	$\{1,2,4\}$	$[500-1000]$	$[500-3000]$	$[512-4069]$

3.2. Experimental Tool for Resource Allocation of Distributed System Based on Hierarchical Clustering Algorithm

The cloud computing simulation software developed based on GridSim and SimJava can simulate the calculation of hierarchical clustering algorithm [15]. CloudSim can be divided into three layers from bottom to top: core simulation engine, simulation system and user code. The core simulation engine of CloudSim mainly includes GridSim and SimJava[16]. CloudSim takes full advantage of existing simulator resources to build its underlying layer. At the lowest level, SimJava is used to perform the core functions of the high-level simulation framework, such as query and processing of time, creation of system components (such as data centers, data center agents, virtual machines, user tasks, etc.), and communication between different components [17]. GridSim toolkit and its library functions are used to support high-level software, including network and grid components, etc. [18]. The specific structure is shown in Figure 1:

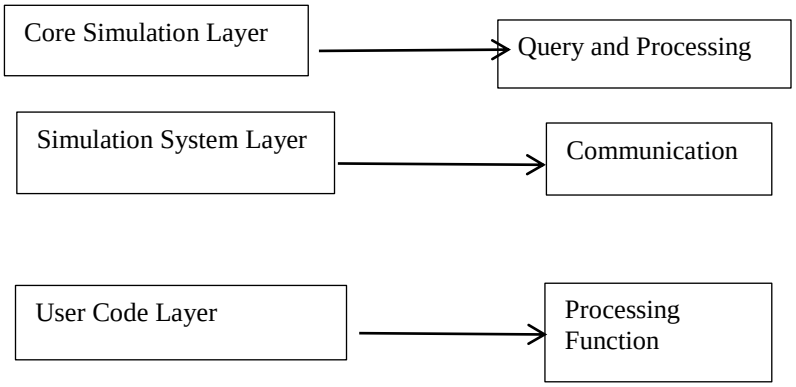

Figure 1. CloudSim hierarchy diagram

4. Research on Resource Allocation Design of Distributed System Based on Hierarchical Clustering Algorithm

4.1. Resource Allocation Design of Distributed System Based on Hierarchical Clustering Algorithm

Based on the improved resource allocation of distributed systems, this paper proposes a resource allocation scheme based on hierarchical clustering algorithm. Using hierarchical clustering algorithm, the resource blocks with the best quality are preferentially allocated to users. If the best resource block of the user is the same, the resource block is allocated to the user with the largest value, and other users who cannot obtain the best resource block are marked, and the resource block with the highest relative value among the remaining resource blocks is allocated to them. The specific algorithm flow As shown in the picture 2:

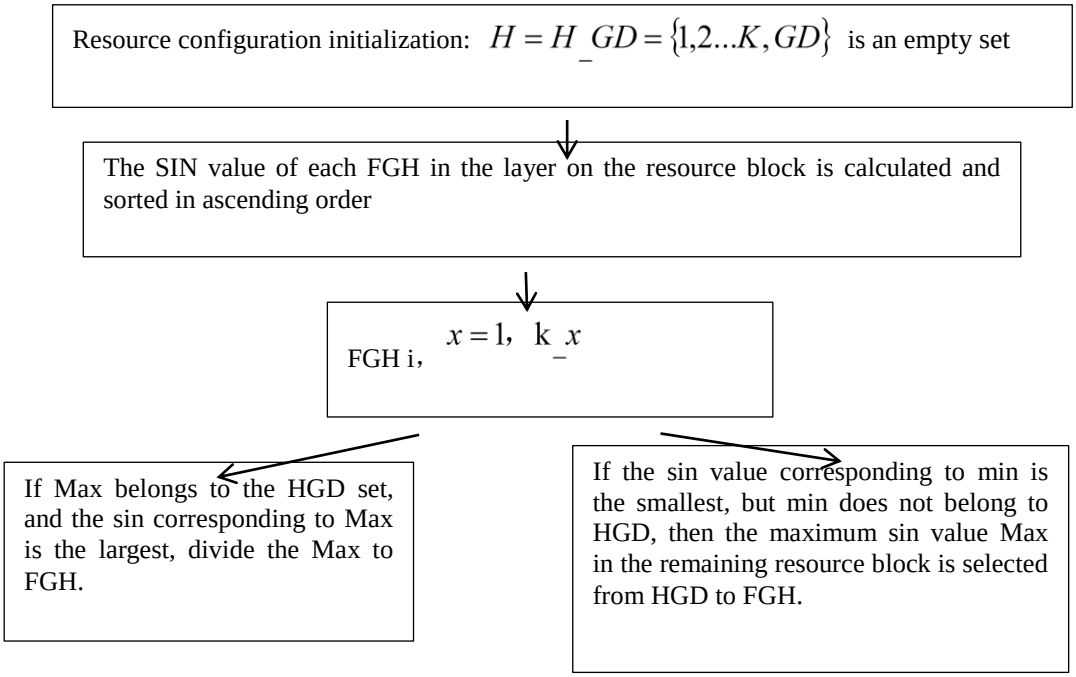

Figure 2. Resource configuration flowchart

It should be noted that in the resource allocation algorithm based on the hierarchical clustering distributed system, the worst interference situation is considered when calculating, that is, except for the FGH in the system level, the rest of $K - k_x$ are interference sources, and the result is obtained. The sin value will be smaller than the actual value. In addition, orthogonal resource allocation is implemented in the system level, and one resource block can only be allocated to one system level, so as to reduce interference. Since there are not many data resources at each level, the data in each level can be configured in the distributed system at the same time without being affected.

4.2 Application of Resource Allocation of Distributed System Based on Hierarchical Clustering Algorithm

In order to verify the feasibility of the hierarchical clustering algorithm in this paper in the

resource allocation of distributed systems, Table 2 and Figure 3 are used to show the change and comparison of the processing time in the resource allocation of the two algorithms.

Table 2. Processing time for resource configuration

Resource Configuration Overhead	CGBD	CDBB	Hierarchical Clustering	Homotopy Algorithm
100	204.67	194.89	204.67	204.67
200	204.67	194.89	184.56	184.56
300	204.67	194.89	177.05	184.56
400	204.67	194.89	174.45	183.67

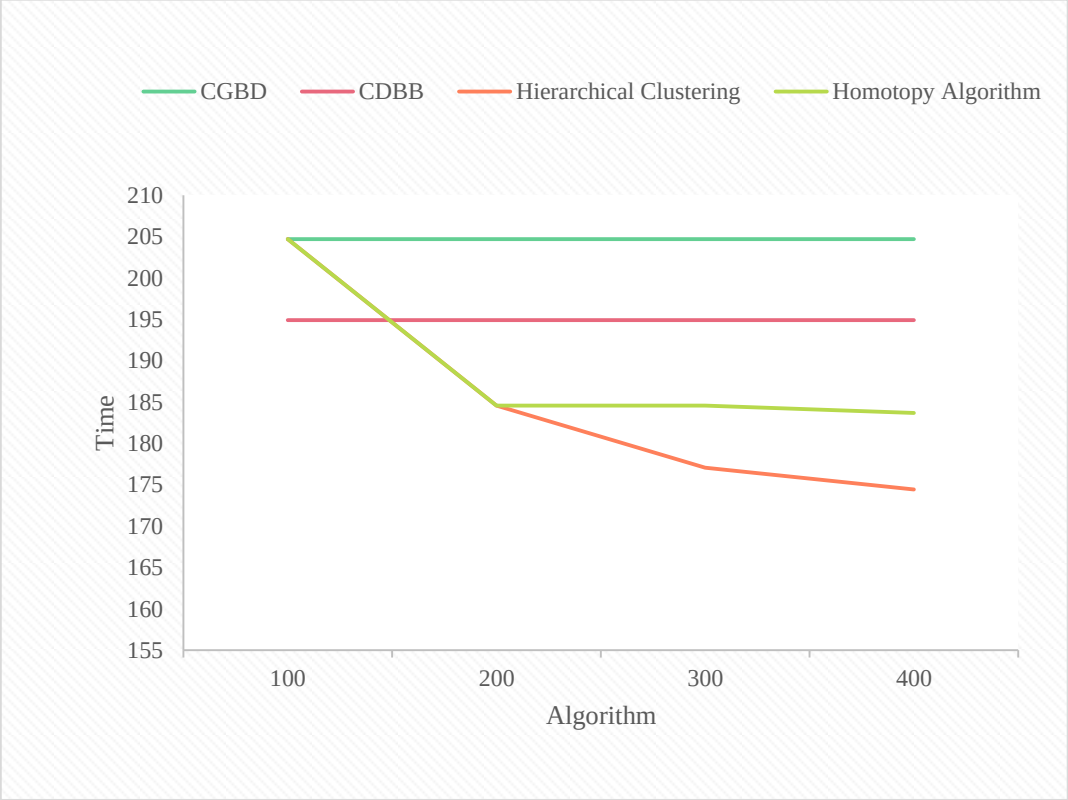

Figure 3. Processing time for resource configuration

It can be seen from Figure 3 and Table 2 that the processing time of the hierarchical clustering algorithm proposed in this paper tends to decrease in the resource allocation of distributed systems. When the number of tasks for resource allocation is 400, the processing time of the hierarchical clustering algorithm is about 174.45 milliseconds, while the processing time of the CGBD and CDBB algorithms is more than 204.67 milliseconds. The other homotopy algorithm also shows a downward trend, but its processing time is significantly slower than that of the hierarchical clustering algorithm. Therefore, the experimental data show the applicability of the hierarchical clustering algorithm in the allocation of distributed system resources.

5. Conclusion

Therefore, in order to enrich the research on the resource allocation of the distributed system of the hierarchical clustering algorithm, this paper first briefly introduces the functional equation of the multi-level clustering algorithm and the method of resource allocation of the distributed system, and then discusses the hierarchical clustering algorithm. Based on the analysis and discussion of the resource allocation technology of the distributed system, the experimental tools and parameters of the resource allocation of the distributed system based on the hierarchical clustering algorithm are investigated and designed. Secondly, the design and analysis of the process structure of the resource allocation of the distributed system of the hierarchical clustering algorithm is carried out. Finally, the experimental data is compared and analyzed for the processing time of the resource allocation of the distributed system of the hierarchical clustering algorithm designed in this paper, and the final experimental results are verified. In this paper, the applicability of integration techniques for distributed systems based on multi-objective homotopy algorithms.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Botre M K, In T. Modified clustering algorithms : Survey paper. International Journal of Development Research, 2018, 2(3):34-39.Botre

[2] Prasad A S, Koll D, Iglesias J O, et al. RConf(PD): Automated resource configuration of complex services in the cloud. Future Generation Computer Systems, 2018, 87(OCT.):639-650

[3] Hossen M B, Auwul M R. Comparative Study of K-Means, Partitioning Around Medoids, Agglomerative Hierarchical, and DIANA Clustering Algorithms by Using Cancer Datasets. Biomedical Statistics and Informatics, 2020, 5(1):20-25.

[4] Bhatnagar V, Majhi R, Jena P R. Comparative Performance Evaluation of Clustering Algorithms for Grouping Manufacturing Firms. Arabian Journal for Science & Engineering, 2018, 43(8):4071-4083.

[5] Botre M K, In T. Modified clustering algorithms : Survey paper. International Journal of Development Research, 2018, 2(3):34-39.Botre

[6] Igual J. Hierarchical Clustering of Materials With Defects Using Impact-Echo Testing. IEEE Transactions on Instrumentation and Measurement, 2020, PP(99):1-1.

[7] Safavizadeh, Arash, Yousefi, et al. Voltage Variation Mitigation Using Reactive Power Management of Distributed Energy Resources in a Smart Distribution System. IEEE Transactions on Smart Grid, 2019, 10(2):1907-1915.

[8] Carvallo J P, Taneja J, Callaway D, et al. Distributed Resources Shift Paradigms on Power System Design, Planning, and Operation: An Application of the GAP Model. Proceedings of the IEEE, 2019, PP(99):1-17.

[9] Seo B, Sim D, Lee T, et al. Efficient Time Synchronization Method With Adaptive Resource Configuration for FBMC Systems. IEEE Transactions on Communications, 2020, PP(99):1-1.

[10] Dalimunthe S, Hanafiah A. Implementation of Agglomerative Hierarchical Clustering Based on The Classification of Food Ingredients Content of Nutritional Substances. It Journal Research And Development, 2021, 6(1):60-69.

[11] Brunk E, Sahoo S, Zielinski D C, et al. Recon3D: A Resource Enabling A Three-Dimensional View of Gene Variation in Human Metabolism. Nature Biotechnology, 2018, 36(3):272-281.

[12] Lyons D J, Mack J, Ketsche P G, et al. the impact of inventory leanness and slack resources on supply chain resilience: an empirical study notice to borrowers the impact of inventory leanness and slack resources on supply chain resilience: an empirical study. Human Resource Management Review, 2019, 21(3):243-255.

[13] Challita U, Dong L, Saad W. Proactive Resource Management for LTE in Unlicensed Spectrum: A Deep Learning Perspective. IEEE Transactions on Wireless Communications, 2018, PP(99):1-1.

[14] Nicholas J, Boster A, Wu N, et al. Comparison of Disease-Modifying Therapies for the Management of Multiple Sclerosis: Analysis of Healthcare Resource Utilization and Relapse Rates from US Insurance Claims Data. PharmacoEconomics - Open, 2018, 2(1):31-41.

[15] Anand A, Veciana G D. Resource Allocation and HARQ Optimization for URLLC Traffic in 5G Wireless Networks. IEEE Journal on Selected Areas in Communications, 2018, 36(11):2411-2421.

[16] Atsawawaranunt K, Bru L C, Mozhdehi S A, et al. The SISAL database: A global resource to document oxygen and carbon isotope records from speleothems. Earth System Science Data, 2018, 10(3):1687–1713.

[17] Schild M, Weber V, Thai D, et al. Treatment Patterns and Healthcare Resource Utilization Among Patients with Atopic Dermatitis: A Retrospective Cohort Study Using German Health Claims Data. Dermatology and Therapy, 2021, 12(8):1925-1945.

[18] Fabris M, Michieletto G, Cenedese A. A Proximal Point Approach for Distributed System State Estimation. IFAC-PapersOnLine, 2020, 53(2):2702-2707.