Journal of Information Technology Research
Volume 15 • Issue 1

Adaptive Peak Environmental Density Clustering Algorithm in Cloud Computing Technology

Qiangshan Zhang, Xinyang Vocational and Technical College, China*

ABSTRACT

In order to get sparsity clustering ability of unbalanced cloud data set, combined with adaptive environment density screening, data clustering was carried out, and an improved adaptive environment density peak clustering algorithm under cloud computing technology was proposed. The storage structure model of grid sparse unbalanced cloud data set is constructed, and structure of grid sparse unbalanced cloud data set is reconstructed by combining feature space reconstruction technology. Rough feature quantity of grid sparse unbalanced cloud data set is extracted, and feature extraction and registration are carried out through strict feature registration method. Cloud fusion and peak feature clustering were carried out according to the grid block distribution of the data set. Peak feature quantities of the grid sparse unbalanced cloud data set were extracted, and binary semantic feature distributed detection of the data was carried out.

KEYWORDS

Cloud Computing Technology, Cloud Data, Environmental Density, Peak Clustering Algorithm, Self-Adaptive

INTRODUCTION

Clustering, which is widely used in pattern recognition, data mining and other fields, is a way to classify data sets without manual supervision. Clustering algorithm groups data according to similarity of data. With the development of intelligent cloud computing and network communication technology, cloud grid communication technology is utilized to data fusion scheduling and storage to improve data storage performance. In the environment of cloud technology, data is stored in the form of sparse grid, and the distribution of data is unbalanced, resulting in poor effective clustering detection and recognition ability of data. It is necessary in perform optimization detection and clustering processing on the grid sparse unbalanced cloud data. Fusion clustering analysis model of grid sparse unbalanced cloud data is built, and the clustering analysis and recognition ability of grid sparse unbalanced cloud data are improved by combining spatial clustering feature analysis and optimized data mining algorithm. Research on the clustering method of unbalanced grid sparsity cloud data has attracted great attention. About unbalanced grid sparse cloud clustering of data processing are mainly fuzzy C-means clustering, irregular triangle gathering originative and cloud data clustering method. The

DOI: 10.4018/JITR.298614

*Corresponding Author

distributed structure model of grid sparse unbalanced cloud data was built, and cluster of grid sparse unbalanced cloud data was carried out through the reorganization of the spatial distributed structure. Reference (You, 2019) proposes the density peak clustering algorithm of grid sparse unbalanced cloud data environment built on deep learning. Heterogeneous directed graph fusion method is adopted to design the storage structure of grid sparse unbalanced cloud data set, and combined with feature space reorganization technology to carry out grid sparse unbalanced cloud data set structure reorganization, so as to improve the fusion property of data clustering. However, this method has a large computing overhead and poor real-time performance in data clustering. Reference (Ma, 2019) puts forward clustering method of grid sparse unbalanced cloud data based on fuzzy C-means clustering. Combined with grid shading peak clustering and attribute classification and recognition, deep learning method is adopted to carry out optimization learning of data clustering process.

Reference (Tang, Xinyu (2019)) is proposed based on swarm intelligence algorithm in cloud computing of large data mining, clustering analysis, clustering algorithm of fuzzy C - average clustering algorithm, the heuristic hybrid leapfrog algorithm of swarm intelligence optimization techniques combined with fuzzy C - average clustering, in order to adjust the parameters of the less optimization under the condition of global search ability, which can better solve the problem of local trap, with good clustering effect, accuracy and convergence speed. At the same time, the algorithm has high stability, but the fuzzy clustering analysis of this method is easy to fall into local extremum.

However, this method has low feature recognition for data clustering. To solve the above problems, this paper proposes an improved adaptive peak density clustering algorithm under cloud computing technology. Building uneven meshes thin cloud data storage structure model. Through the characteristics of the strict registration method adaptive under cloud computing environment of peak density feature extraction and matching, according to data set of distribution grid blocks for data fusion and peak feature of cloud cluster, extract uneven mesh thin cloud peak characteristic of data set. Through space spectrum characteristics of clustering and information fusion method, it carries on data of binary semantic characteristics of distributed detection to obtain adaptive environment under cloud computing technology density peak cluster. Finally, simulation test and analysis are carried out to demonstrate superior performance of the proposed method in improving adaptive peak density clustering ability of the environment under cloud computing technology.

DATA STORAGE STRUCTURE MODEL AND FEATURE ANALYSIS

Grid Sparsity Unbalanced Cloud Data Set Storage Structure Model

In order to achieve optimal clustering of grid sparse unbalanced cloud data set, distribution model of grid sparse unbalanced cloud data set is constructed, combining point distribution features matching to get data storage grid partition. In the sparse sample distribution set, shading feature detection of grid unbalanced cloud data set is carried out. Peak scheduling is adopted to carry out optimal clustering design of grid sparsely unbalanced cloud data set. Assuming that the cluster-head node of grid sparse unbalanced cloud data set transfers attribute set, grid sparse scheduling is carried out for all the data in the cluster-head node.

When unbalanced cloud data set meets the convergent solution, semantic ontology model is adopted to carry out fusion processing of grid sparse unbalanced cloud data, and storage measure of remaining active node data is defined as follows:

$$Z = 2\left(g + b\right) \times \frac{\alpha \chi}{\beta^2} \tag{1}$$

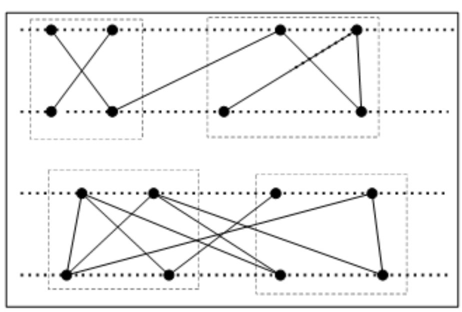
In above equation, g represents statistical characteristics of environmental density forward dispatching, b represents statistical characteristics of environmental density reverse scheduling, χ represents a random resource data to be scheduled, α Represents standard dispatching coefficient, β Represents basic operational parameters of distributed scheduling detection statistical process. Related parameters of cluster class of grid sparse unbalanced cloud data are described. A statistical analysis model of grid sparse unbalanced cloud data set is established, and grid sparse unbalanced cloud data detection is carried out through fuzzy grid area fusion. Under fuzzy semantic fusion, a weighted analysis model of grid sparse unbalanced cloud data is obtained. Considering the reliability of data transmitted by the cluster-head node, the utility function is used to carry out the balanced control of data storage and scheduling, and fuzzy mapping function d_n is obtained. Similarity information of grid sparse unbalanced cloud data set is described. Based on the above analysis, a storage structure model of grid sparse unbalanced cloud data set is established. As shown in Figure 1.

(a) A grid diagram for data clustering storage.
(b) A linear distribution of data clustering storage.

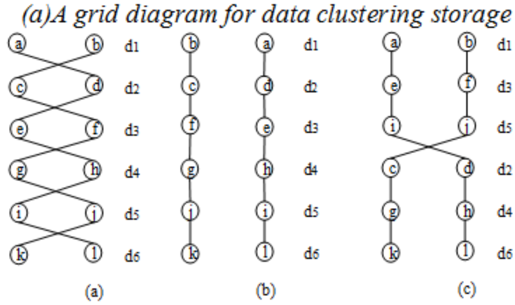
The storage structure of grid sparse unbalanced cloud data set has the following characteristics:

1. **The utility performance is strong:** According to different data information, data port should be provided. Meanwhile, a complete monitoring system should be adopted for real-time monitoring and close contact with the outside world. The user's application experience is guaranteed.
2. **Low cost:** Cloud computing data access storage system expansion equipment is simple. With continuous development of science and technology, maintenance and economic costs are becoming lower and lower.

Figure 1. (a and b) Storage structural model of grid sparse unbalanced cloud data set



(a)A grid diagram for data clustering storage



(b)A linear distribution of data cluster storage

3. **It has scalability:** The number of servers of cloud computing data access storage system can reach hundreds, the more the quantity, the greater the scalability of the storage system.

Data Distribution Feature Extraction

Combined with feature space recombination technology, the structure of grid sparse unbalanced cloud data set is recombined to extract rough characteristic quantity of grid sparse unbalanced cloud data set. Gain values of all links are calculated, and grid sparsity peak fusion is carried out through deep learning method. The clustering similarity feature quantity under control of environmental density is obtained as follows:

$$K_L = \frac{\lambda(g+b)}{z} \tag{2}$$

In this equation, λ represents value result of adaptive environment density information resource data under the standard definition. If the link gain value $z \geq 1$ is obtained, the environmental fitness weight coefficient of the unbalanced cloud data set is positive. The clustering parameters of each unbalanced cloud data were corrected to obtain clustering effectiveness evaluation parameter distribution set and index weight of all cluster head nodes.

To sum up, a deep learning model of grid sparse unbalanced cloud data aggregation class is established, and fuzzy characteristic distribution of grid sparse unbalanced cloud data set is constructed. The constraint programming model of grid sparse unbalanced cloud data set is obtained as follows:

$$\min(f) = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij} X_{ij} \tag{3}$$

$$\text{s.t} \begin{cases} \sum_{j=1}^{m} X_{ij} = a_i, i = 1,2\cdots m \\ \sum_{i=1}^{m} X_{ij} = b_i, j = 1,2\cdots n \\ X_{ij} \geq 0, i = 1,2\cdots m, j = 1,2\cdots n \end{cases} \tag{4}$$

For optimal cluster center of all cluster head nodes, fuzzy degree detection was used to obtain evaluation set and test set of grid sparse unbalanced cloud data set.

After data transmission is suspended, characteristic distribution of grid sparse unbalanced cloud data set is obtained as follows:

$$D_q = \delta \times D^2 + \gamma \times D \tag{5}$$

In formula (5), δ is the fusion clustering feature set of grid sparse unbalanced cloud data set. D represents pixel brightness value before correction of fused clustering feature set of sparse unbalanced cloud data set. δ γ all are regression parameters. According to extraction results, peak fusion clustering was carried out.

DATA CLUSTERING OPTIMIZATION

Peak Fusion of Grid Sparse Unbalanced Cloud Data Set

After peak fusion clustering is completed, sparse unbalanced cloud data set of grid is combined for peak fusion. According to grid block distribution of data set, the cloud fusion and peak feature clustering of data set were carried out to extract peak feature of grid sparse unbalanced cloud data set. The iterative formula of algorithm for peak feature extraction was given as following:

$$w(k) = D_q \times w(k-1) \tag{6}$$

In this equation, w represents the choice to adopt an embedded dimension scheduling value. Assuming that k is the boundary value vector of data transmitted by cluster-head node, adaptive environmental kurtosis of grid sparse unbalanced Cloud data set can be defined as:

$$k_{urt}\left(v\right) = w(k) + E\left(v_1 + v_2\right) + \varphi \tag{7}$$

Boundary feature quantities v_1 and v_2 and scalar φ of sparse unbalanced cloud data are obtained. Through fusion of clustering, adaptive environment density distribution can be obtained according to the following formula:

$$S_{ij} = \left\|X_i - X_j\right\|^2 \tag{8}$$

In this equation, i and j represents any two nodes in the adaptive environment density distribution respectively. The corresponding density pixel values of two are X_i and X_j respectively.

The main factors affecting peak fusion of grid sparse unbalanced cloud data set are mainly divided into three, as shown below.

One is load balancing. In the process of storage and distribution of data access information, it is possible to evenly distribute the data access information on the nodes and keep load balance of the nodes, so as to carry out parallel processing of data access information and improve efficiency of data processing. If the distribution of data access information storage is unbalanced, there will be a situation of "data skew", which means that parallel processing cannot be carried out between nodes, which will have a great impact on the peak fusion information storage of grid sparse unbalanced cloud data set. Second, node failure. For HDFS clusters, node failures are a frequent problem. Therefore, the data access storage optimization algorithm is required to be fault-tolerant, so that the peak fusion information about grid sparse unbalanced cloud data set can be evenly stored and distributed. Third, storage operation performance. Operational performance directly affects the effect of data access storage. In general, we only consider network transport operations. If the network transmission operation is simple, communication traffic between nodes during data access storage algorithm execution can be greatly reduced, and peak fusion of grid sparse unbalanced cloud data set can be effectively improved.

Assuming that there are N cluster-head nodes in the center of the grid, the argument $V_l(1 \le l \le k-1)$ of cluster-head node can be obtained. Suppose that $\left\{v_1, v_2, \cdots, v_n\right\}$ is n grid sparse unbalanced classification values of V_l, then for any grid sparse unbalanced data classification values, there is a peak characteristic value in the whole data set. That is, $Q(v_i)$ represents the measure of classification value v_i of the argument parameter V_l of grid sparse unbalanced data, so that we call $Q(v_i)$ is the measure of classification value $v_i (1 \leq i \leq n)$.

By means of semantic correlation fusion, the characteristic similarity matrix of grid sparse unbalanced cloud data is obtained:

$$R = \begin{bmatrix} r(V_1, V_1) & \cdots & r(V_1, V_{k-1}) \\ \cdots & \cdots & \cdots \\ r(V_{k-1}, V_1) & \cdots & r(V_{k-1}, V_{k-1}) \end{bmatrix} \tag{9}$$

In this way, data feature detection is carried out. Cloud fusion and peak feature clustering are carried out according to the grid block distribution of the data set, and peak feature quantity of the grid sparse unbalanced cloud data set are extracted.

Data Adaptive Environment Density Peak Clustering

According to semantic fusion and feature clustering results, adaptive environment density fusion was carried out, binary semantic feature distributed detection of data was carried out, and clustering function of the unbalanced cloud data set was obtained. First partial derivative with respect to parameters p_i:

$$\frac{\partial u_i}{\partial p_i} = \frac{Gh_i}{\sum_{j \neq i} h_j p_j + \sigma^2} \left(\frac{1}{1+\gamma_i} - \beta_{c_1} \right) \tag{10}$$

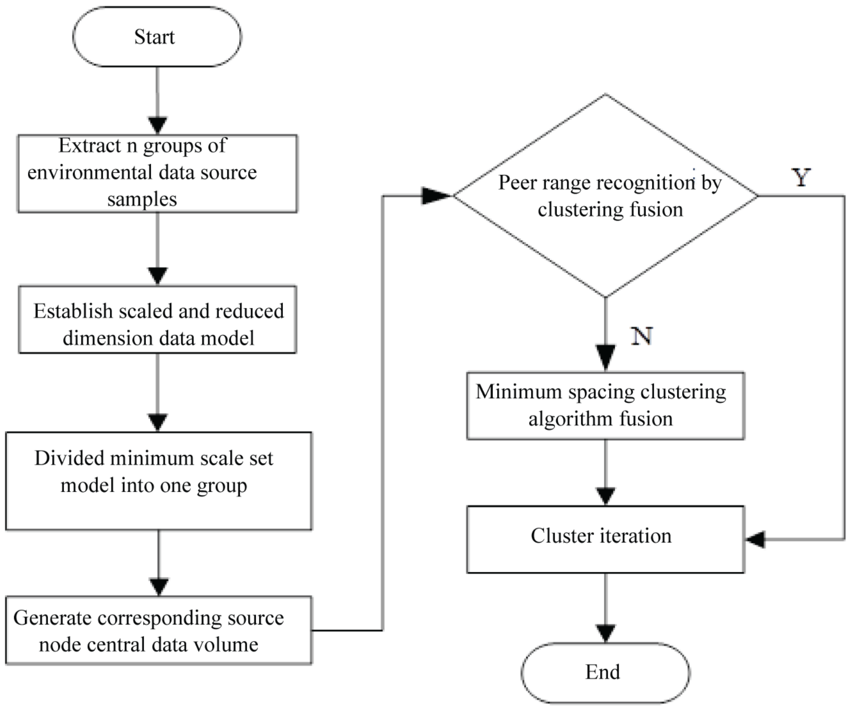
According to maximum value of fusion cluster center distribution of unbalanced cloud data set, parameter v was adjusted. When $\beta_{c_i} = u_i'(\gamma_{th})$, the utility function of grid sparse unbalanced cloud data clustering class is:

$$\frac{\partial u_i}{\partial p_i} = \frac{Gh_i}{\sum_{j \neq i} h_j p_j + \sigma^2} \left(\frac{1}{1+\gamma_i} - \frac{1}{1+\gamma_{th}} \right) \tag{11}$$

The multi-dimensional scaling model is formed by retrieval feedback increment data obtained from minimum scale set model and cluster retrieval center data. Clustering data sets are divided according to corresponding logical features of retrieval feedback increasing data by neighborhood similarity method to complete clustering operation of data. According to the average space between the class data, value of clustering operation is corrected in real time to complete operational construction of the multi-dimensional scaling model of the data. Adaptive environment peak density clustering algorithm process is shown in Figure 2.

The core idea of density peak clustering algorithm is to calculate the clustering center point, which is dense and has a large distance from other data points with high density. However, due to certain defects of density clustering algorithm, large amount of memory and computation is required when clustering high-dimensional data of large data sets, which will inevitably lead to excessive space-time complexity and affect speed and accuracy of the algorithm. In practical application, the algorithm

Figure 2. Flow charts of adaptive environmental density peak clustering algorithm



with too much space-time complexity cannot be directly used in the processing and calculation of big data. In order to achieve more accurate results of the above density peak clustering algorithm, KNTSCCA algorithm is added at the end of the clustering data output process to carry out fusion operation on its clustering volume.

The specific way is as follows.

Input two-dimensional coordinate data point parameters of dimensionless model and output clustering composition $P = \left\{P_1, P_2, \cdots P_v\right\}$:

1. Run Nb Clust function inside Nb Clust package and set clustering data category coefficient in R language;
2. Type the variable coefficient of clustering data and extract the variable clustering data dynamically;
3. K-means algorithm is used to dynamically calculate the new clustering core data source;
4. Change comparison of core data of secondary clustering. If there is any change, repeat steps (2) ~ (5); if there is any change, skip steps (5);
5. Core data source classification is static or class center does not change. When fusion operation is finished, and n groups of clustering data results are output.

Since the state perception results of each node in the target network have different influence levels on the final network congestion state, it is difficult to ensure the accuracy of the congestion state estimation results. Therefore, the weighted average method is adopted to aggregate direct perception information obtained by nodes and the indirect perception information obtained by other means. With the increase of grid sparsity and unbalanced cloud data, u_i gradually decreases. At this time, the fusion scheduling set of cluster-head nodes of grid sparsely unbalanced cloud data set is $p_i = 0$. Environment parameter variables meet $\gamma_i = 0$, and assumptions $\gamma_i > \gamma_{th}$. Therefore, the initial condition of grid sparse unbalanced cloud data set is $\gamma_i \leq \gamma_{th}$. It can be known that $\partial u_i / \partial p_i \geq 0$, peak clustering center of grid sparse unbalanced cloud data set satisfies $du_i / d\gamma_i \geq 0$, so that $\arg\max_{p_i \in P_i} u_i$ and $\arg\max_{p_i \in P_i} \gamma_i$ are equivalent. By $\gamma_i \leq \gamma_{th}$, the environment adaptive parameters of grid sparse unbalanced cloud data aggregation class $\arg\max_{p_i \in P_i} \gamma_i = \gamma_{th}$ can be obtained. According to the above analysis, adaptive environment density peak clustering of the grid sparse unbalanced cloud data set is carried out.

SIMULATION EXPERIMENT AND RESULT ANALYSIS

Simulation tests to verify the application performance of method in implementation of grid sparse unbalanced cloud data clustering class. It is assumed that the sampling length of grid sparse unbalanced cloud data is 2000. The test set of peak clustering is 120. The time rise coefficient of data clustering is 0.004. The period of simulation experiment is 1.2s, and the data access is 800. Time domain distribution of data is obtained. As shown in Figure 3.

Times

Taking the data in Figure 3 as the research object to get data clustering, and 500 Monte Carlo experiments were conducted to obtain the convergence curve of clustering, as shown in Figure 4.

According to analysis in Figure 4, the convergence of grid sparse unbalanced cloud data clustering class is good and close to the standard clustering curve. However, the methods of reference (5) and reference (6) differ greatly from the standard clustering curve. The results of grid sparse unbalanced cloud data clustering class under different SNRS are tested, as shown in Table 1.

Analysis of the simulation results in Table 1 shows that with the increase of SNR, the error of peak data clustering decreases, and the error of unbalanced Cloud Data clustering based on grid sparse using the method proposed in this paper is generally smaller than that of traditional methods. The measured values of the methods in reference (You, Zhiyong (2019)) and reference (Ma, Junyan (2019)) differ greatly from the predicted values, and the errors are relatively high. Based on these studies, the error rate was tested, and the comparison results were shown in Table 2.

Figure 3. The temporal distribution of the data

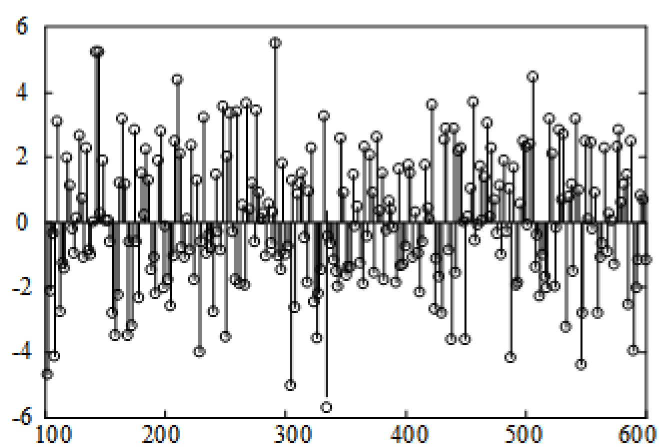


Figure 4. Data clustering convergence curve

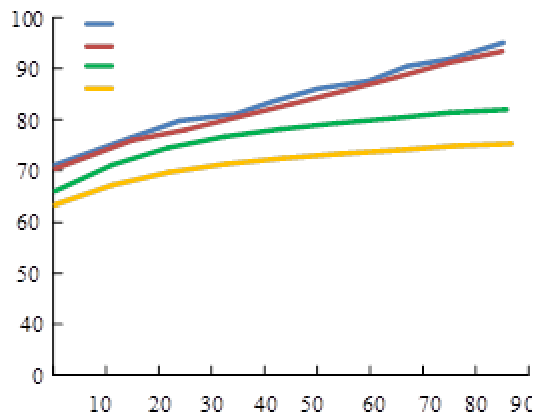


Table 1 Data clustering errors of different methods under different SNR

Other methods	Measured values/mm	Predictive values/mm	Errors/%
Methods in this paper	121	122	1
Reference (You, Zhiyong (2019)) methods	150	160	10
Reference (Ma, Junyan (2019)) methods	145	152	7

Table 2 Data clustering accuracy comparison

Number of iterations times	Methods in this paper/%	Reference (You, Zhiyong.2019)methods/%	Reference (Ma, Junyan.2019)methods/%
100	96.14	86.23	86.25
200	97.52	89.01	87.36
300	98.26	89.12	88.37
400	98.99	90.12	89.96

It can be seen from Table 2 that the method in this paper has a high accuracy rate in the grid sparse unbalanced cloud data clustering class. The reason is that in data clustering analysis, the method in this paper establishes a storage structure model of unbalanced cloud data set with grid sparsity, which reduces clustering error to a certain extent and improves the accuracy of data results.

CONCLUSION

Fusion clustering analysis model for grid sparse unbalanced cloud data were built. Combined with spatial clustering feature analysis and optimized data mining algorithm, clustering analysis and recognition ability of grid sparse unbalanced cloud data were improved. In this paper, an improved adaptive environment density peak clustering algorithm under cloud computing technology was proposed. Peak scheduling was adopted to perform optimal clustering design of grid sparse unbalanced cloud data set, and structural reorganization of grid sparse unbalanced cloud data set was carried out in combination with feature space reorganization technology. The method proposed in this paper has low misclassification rate, good convergence and good performance for grid sparse unbalanced cloud data set peak clustering, which is of great significance for practical application. The optimal clustering center for all cluster head nodes was determined. Through fuzziness detection, the evaluation set and test set of the grid sparse unbalanced cloud data set are obtained to carry out distributed detection of binary semantic features of the data. According to the semantic fusion and feature clustering results, the adaptive environment density fusion is carried out, and the adaptive environment density peak clustering of the grid sparse unbalanced cloud data set is carried out. The analysis shows that the peak clustering of unbalanced cloud data set with grid sparsity in this paper has a low error rate and good convergence. The adaptive peak environmental density clustering algorithm is still in continuous development, and various types of clustering have been gradually improved, so it is necessary to constantly optimize the method according to the development of environmental density, so as to truly and accurately provide the basis for the peak environmental density clustering. In the future research, the different characteristics of peak clustering algorithm should be fully considered, and a new measurement should be carried out according to the actual development situation. The next research direction is to apply the proposed adaptive environment density peak clustering algorithm to cluster with large-scale data streams.

FUNDING AGENCY

Publisher has waived the Open Access publishing fee.

REFERENCES

Admela, J. (2019). Fog-to-Cloud Computing for Farming: Low-Cost Technologies, Data Exchange, and Animal Welfare. *Computer*, *52*(10), 41–51. doi:10.1109/MC.2019.2906837

Arzhakova, O. V., Prishchepa, D. V., Dolgova, A. A., & Volynskii, A. L. (2019). The effect of preliminary orientation on environmental crazing of high-density polyethylene films. *Polymer*, *170*(12), 179–189. doi:10.1016/j.polymer.2019.03.019

Collas, F. P. L., Buijse, A. D., & Heuvel, L. V. D. et al.. (2018). Longitudinal training dams mitigate effects of shipping on environmental conditions and fish density in the littoral zones of the river Rhine. *The Science of the Total Environment*, *619-620*(1), 1183–1193. doi:10.1016/j.scitotenv.2017.10.299 PMID:29734597

Jin, , Zhu, & Zhao. (2019). Optimal spectrum allocation algorithm in cellular heterogeneous network. *Computer Engineering*, *45*(2), 70–75.

Kumar, P., & Kumar, R. (2019). Issues and Challenges of Load Balancing Techniques in Cloud Computing: A Survey. *ACM Computing Surveys*, *51*(6), 1–35. doi:10.1145/3281010

Kumar, P., & Kumar, R. (2019). Issues and Challenges of Load Balancing Techniques in Cloud Computing: A Survey. *ACM Computing Surveys*, *51*(6), 1–35. doi:10.1145/3281010

Liu, & Li. (2018). Adaptive threshold algorithm based on statistical prediction in spatio - temporal crowdsourcing environment. *Computer Applications (Nottingham)*, *38*(2), 415–420.

Ma, , Zhang, & Li. (2019). HA2: Hierarchical analysis technology for firmware abnormality of IoT sensing devices. *Computer Engineering and Applications*, *55*(22), 60–68.

Näslund, J., Rosengren, M., & Johnsson, J. I. (2019). Fish density, but not environmental enrichment, affects the size of cerebellum in the brain of juvenile hatchery-reared Atlantic salmon. *Environmental Biology of Fishes*, *102*(2), 705–712. doi:10.1007/s10641-019-00864-9

Nie, , Zhang, & Yu. (2020). Spatiotemporal crowdsourcing task allocation algorithm for global optimization. *Jisuanji Yingyong*, *40*(7), 1950–1958.

Patil, R., & Modi, C. (2019). Designing an efficient framework for vulnerability assessment and patching (VAP) in virtual environment of cloud computing. *The Journal of Supercomputing*, *75*(5), 2862–2889. doi:10.1007/s11227-018-2698-6

Rosenow, J., Foerster, S., & Lindner, M. (2019). Multicriteria-Optimized Trajectories Impacting Today's Air Traffic Density, Efficiency, and Environmental Compatibility. *AIAA Journal*, *27*(1), 8–15.

Rui, , Zhang, & Huang. (2016). A reputation-based incentive mechanism for crowdsourcing. *Dianzi Yu Xinxi Xuebao*, *38*(7), 1808–1815.

Song, , Tong, & Wang. (2017). Online task allocation of three types of objects in a spatial crowdsourcing environment. *Journal of Software*, *28*(3), 611–630.

Song, , Yu, & Chen. (2018). Big data similar duplicate record detection algorithm based on MapReduce model. *Journal of Shanghai Jiaotong University*, *52*(02), 214–221.

Tang, Zhang, & Zhao. (2019). Big data cluster mining based on swarm intelligence algorithm in cloud computing. *Journal of Chongqing University of Technology (Natural Science), 33*(4), 128-133+167.

Tian, , Yuan, & Liu. (2020). Lane line detection and adaptive fitting algorithm based on instance segmentation. *Jisuanji Yingyong*, *40*(7), 1932–1937.

Wang, Zhang, & Zhao. (2019). Design of cloud platform for landslide monitoring based on cloud computing technology: Taking Jingyang, Shaanxi as an example. *Bulletin of Surveying and Mapping, 3*(3), 128-132, 136.

Wu, Z. (2018). Research on the application of cloud computing technology in the public service platform of educational resources. *China Audio-visual Education*, *7*(2), 107–111.

You, Su, & Wang. (2019). Implementation of WSN data fusion algorithm based on GA-ACO-BP. *Modern Electronic Technology, 42*(21), 13-17+22.