[Mangayarkkarasi* 4(3): March, 2017] ISSN 2349-4506
Impact Factor: 2.785

Global Journal of Engineering Science and Research Management

RANK-BASED WEIGHTED ASSOCIATION RULE MINING TECHNIQUE FOR SECURE CLOUD COMPUTING ENVIRONMENT

K. Mangayarkkarasi*, M. Chidambaram
* Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, Tamilnadu, India
Assistant Professor, Rajah Serfoji College, Bharathidasan University, Thanjavur, Tamilnadu, India

DOI: 10.5281/zenodo.376580

KEYWORDS: Apriori, Association Rule Mining, Cloud Storage, Data Mining, Security.

ABSTRACT
Currently, the online services are used for all business activities. This produces enormous amount of data every day and all these data is to be stored in different cloud data storages. To identify the best and useful information out of these data and to take a wise decision from the available information is considered as one of the challenging issues. To address this major issue, few researchers adopted data mining techniques such as clustering, classification and association models for mining the useful information from the cloud environment. Although Cloud Computing is a powerful means of achieving high storage and computing services at a low cost, it is revealed that due to challenging security issues, many cloud users are not showing interest to use cloud based services. This research work proposes an efficient Secure Cloud Data Mining Model through Apriori-HUDS Rule Miner to address these issues. This secure cloud data mining model is designed in such a way that it facilitates the cloud users to identify the best frequently used data sets from the cloud data centers for their queries and also provides high security to the cloud servers. Thus, the service providers can protect the sensitive data of clients. The proposed model is implemented in the cloud environment and thoroughly studied. From our experimental results, it is noticed that the proposed model performs well in terms of predicting frequently accessed data sets by cloud users, ranking data sets with highest support and achieving high memory utilization as compared with existing cloud data mining model.

INTRODUCTION
With the growth in the Internet applications, there is an increase in the business activities conducted over the web. Many business organizations prefer to build their own IT infrastructure to host databases or software or have a third party cloud service provider to host them on the large servers. The cloud computing environment acts as a platform to convert the network resources into usable services for the end users. The integration of data mining approaches to the cloud environment enables significant discovery of the useful knowledge. The application of data mining in the cloud computing environment allows the business organizations to centralize the data storage management, while guaranteeing the reliable and secure services for their users. The implementation of the data mining techniques in cloud computing enables the user to obtain significant information from a virtually integrated data warehouse at a minimum infrastructure and storage cost. The data mining uses statistical and machine learning techniques to discover the knowledge and present the knowledge in an easily understandable way to the humans. In large databases, data mining solves the problem to discover the hidden and useful knowledge from the data. This helps the government and business enterprises to make decisions.

Cloud Computing Challenges
From the available literature survey, it is noticed that the Cloud Computing Environments to enable the Future Internet of Services faces the following prime challenges. They are

- High Availability and Energy Efficiency
  - Lack of Resource Allocations for Users Demands and Needs
  - Lack of Schemes to identify the best On-Demand Service Provides
  - Sharing and Adopting Resources
  - Lack of Resource Usage and Utilization
- Lack of Power Efficient Resource Allocation Schemes
- Load Balancing

- Data Transfer Bottlenecks
- Infrastructure Security Threats
    - Authentication and Authorization
    - Data Confidentiality and Integrity
    - Distributed Denial of Service
    - Botnet

As listed above, to provide High Availability to the best resources, we need to maintain resource usage and utilization database to predict the most frequently accessing resources. To enable security to those best resources is another challenge.

This research work is planned to develop an efficient cloud data mining based model to address the above mentioned challenges, particularly to identify the best resources with respect to user demands and to provide security with less traffic congestion and data transfer bottleneck. This will facilitate and enable cloud users to comfortably access future cloud computing environment. The cloud data mining method developed for this purpose is discussed in the following section.

Security issues in cloud computing
The security issues are classified as

- Data Issues
- Privacy issues
- Infected Application
- Security issues

Data issues: The sensitive data arise as a major security issue with the cloud computing system. Firstly, when a data is stored on a cloud, the data can be accessed from the cloud anywhere and anytime. Hence, there is a need for a data integrity approach. Data theft is one of the serious issues in a cloud computing environment. If the service provider shut down the cloud services due to some financial or legal problem, there will be data loss for the user. Moreover, data can be lost or damage or corrupted due to natural disaster and fire outbreak.

Privacy issues: The cloud service provider should guarantee that the personal information of the user is highly secured from the third party agents and other users.

Infected Application: The cloud service provider should have the complete rights for the purpose of monitoring and maintenance of the server. This prevents uploading of any infected application onto the cloud. Hence, the user and cloud computing service are prevented from being affected by the infected application.

Security issues: The service provider should secure the server from the external threats. Though the service provider ensured best security, the user should ensure that there should not be data loss, theft or tampering.

A classification protocol is proposed for the privacy-preserving classification of two-class and multi-class problems using support vector machine [1]. Dai and Ji [2] applied a C4.5 decision tree algorithm using MapReduce programming model. Ikram et al. [3] presented a novel neural network framework for cloud computing. Lin et al. [4] proposed a novel approach for linear regression prediction for dynamic currency exchange rates in cloud computing. A neuroevolutionary algorithm is presented to choose an optimum Artificial Neural Network (ANN) architecture for the analysis of data in medical information systems [5]. A dynamic Virtual Machine (VM) consolidation algorithm is developed to reduce the number of physical servers on a data center [6]. The energy cost is reduced. A framework for selecting the optimal composition of QoS-aware cloud service is proposed [7]. The association rule mining algorithm is proposed for data perturbation before uploading to the server [8]. A survey on Association Rule Mining (ARM) algorithms in the cloud computing environment is presented [9]. The distributed association rule algorithm based on MapReduce programming model is proposed [10]. The problem of outsourcing the ARM task within a privacy-preserving framework is analyzed and an attack model is introduced based on the background knowledge [11]. An algorithm for scheduling workflow task is presented on the basis of

fuzzy clustering of resources in the cloud computing environment [12]. A novel cloud-based service is introduced for ARM on a distributed computing model [13]. A task scheduling optimization strategy and resource hybrid clustering are proposed based on the fuzzy clustering [14]. Zhang and Chen [15] proposed a weighted kernel Possibilistic C-Means (PCM) algorithm based on MapReduce for clustering the data objects in appropriate groups. A new resource allocation strategy is proposed based on the cost model of the parallel task performance, using the improved Fuzzy Clustering Algorithm [16]. A hierarchical agglomerative clustering algorithm is implemented for the parallel execution of task in distributed cloud computing environment [17]. A secure k-means approach is proposed to preserve the privacy of data to be distributed among various hosts [18]. An efficient cloud workload management framework is presented for the analysis and clustering of workloads through K-means based on the weights assigned and Quality of Service (QoS) requirements [19]. A fuzzy clustering method is applied for the effective preprocessing of the cloud resources [20].

Motivation, Objective and Problem Definition

As discussed in the previous sections, it is noticed that the resource owners and data owners configure and upload their data to the cloud environment. This causes a rise in the concerns for high availability with data confidentiality, privacy, authentication and access control. The use of cloud computing is continuously rising by cloud users. It facilitates few features such as mobility, huge availability and low cost. It is noticed that providing high availability resources' details to cloud users and providing security for those popular resources are considered as the major issue and cloud environment need an efficient model to address these demands also. To address the above mentioned issues, this research work proposed an efficient Secured Cloud Data Mining Model which is providing rank of the best and frequently used data sets and resources and providing security to the cloud environment as well.

The remaining sections of the manuscript are systematized as follows: Section II presents an overview of the existing data mining techniques. Section III explains the proposed cloud data mining framework and Section IV illustrates the performance analysis result of the proposed framework including Apriori-HUDS Algorithm. Section V states the conclusion of the cloud data mining framework.

MATERIALS AND METHODS

The data mining techniques are very much required for the next generation cloud computing environment as it required entrepreneurs to predict the future demands. This research work proposes association rule miner based cloud architecture called as secure cloud data mining. The association rule mining technique called as Apriori-High Utility Discovery Data Set (Apriori-HUDS) integrates with cloud environment, which is shown in the Fig. 1. The proposed secured cloud data mining model in a cloud computing environment allows cloud users or organizations to centralize the management of servers, data storage, hardware and software. This proposed model reduces to processing time to identify the best recourse for cloud users demands/queries also ensure that it provides secure services for its users and organizations as well. This proposed model will allow cloud users to retrieve meaningful and required information from the cloud data stores. The proposed Cloud Data Mining offers tremendous information about customers, their habits, interests, their frequently accessing resources and security as well.

Rank-based Weighted Association Rule Miner: Apriori-HUDS

The Weighted Condensed Support Confidence (wcs) is defined as

$$Support\ wcs(Z) = \begin{cases} \frac{\sum_{k=1}^{m} W_k(Z)}{m'(Z)}, if\ |Z| > 1 \\ \frac{\sum_{k=1}^{m} W_k(Z)}{m}, if\ |Z| = 1 \end{cases} \quad (1)$$

Where $m'(Z)$ is described as follows

$$m'(Z) = \max_{(\forall g_i \notin Z, Q=|Z|)} \left\{ \sum_{k=1}^{m} BIT_{k1}, \sum_{k=1}^{m} BIT_{k2}, \dots\dots, \sum_{k=1}^{m} BIT_{Q} \right\} \quad (2)$$

Where Q denotes the total number of datasets in data storage $S|Z| > 1$. The Weighted Condensed Confidence (wcc) is defined as

$$Confidence\ wcc(A \rightarrow C) = \frac{wcs(A \cup C)}{wcs(A)} = \frac{wcs(Z)}{wcs(A)} \quad (3)$$

Apriori-HUDS Algorithm

Input: Data Matrix $\boldsymbol{D(rows = Datasets,\ columns = samples)}$
Output: Set of rules ***Rules,*** support ***RuleSupp***, Confidence ***RuleConf***
Step 1: RankApriori-HUDS()
Step 2: {
Step 3: Normalize the Datamatrix D
Step 4: Calculate rank of Datasets
Step 5: Assign Weights wt(:) to all datasets to their ranks rank(:)
Step 6: Select initial seed value using k-means clustering
Step 7: Initialize k=1
Step 8: Find Frequent datasets
Step 9: $FDS_k = \{i | i \in A1\text{^}wcs(i) \geq \min\_wsupp\}$
Step 10: Repeat
Step 11: k=k+1
Step 12: Generate Candidate datasets, CDS_k *from* FDS_{k-1}
Step 13: For each Candidate dataset, $c \in CDS_k$
Step 14: {
Step 15: Calculate wcs(:) for each Candidate dataset c
Step 16: if $(wcs(:) \geq \min\_wsupp)$ *then*
Step 17: {
Step 18: $FDS_k \leftarrow [\, FDS_k : c]$
Step 19: Generate rules, rule(:) from the frequent datasets c
Step 20: Determine wcc(:) for every rule(:)
Step 21: For every rule, $r \in rule(:)$
Step 22: {
Step 23: if $(wcc(r) \geq \min\_wconf)$ *then*
Step 24: {
Step 25: Store the value of r in resultant rule-list Rules with its wcs and wcc
Step 26: $Rules \leftarrow r, RuleSupp \leftarrow wcs(r) and\ RuleConf \leftarrow wcc(r)$
Step 27: }
Step 28: }
Step 29: }
Step 30: }
Step 31: until $KDS_k = \phi$
Step 32: }

Cloud Data Mining Architecture-Framework Design

The cloud data mining model as shown in the Fig.1, the cloud computing is combined with data mining association rule miner Apriori-HUDS. This proposed model can provide powerful capacities of management in terms of i. Ranking Data Sets and ii. Providing security. As this proposed model has the features to identify the best frequently used data sets, the cloud users will get the best data sets for their queries and the system will achieve secure communication between cloud users and data sets.

ISSN 2349-4506
Impact Factor: 2.785

Global Journal of Engineering Science and Research Management

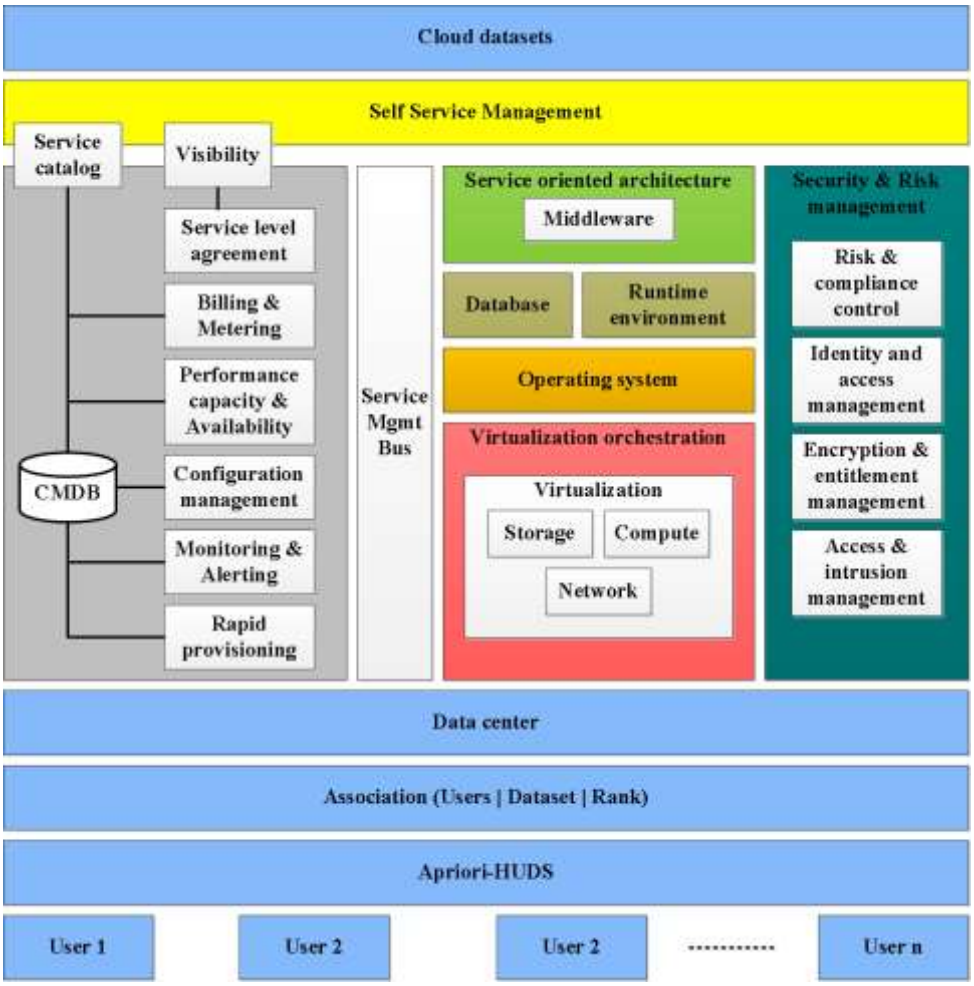

Figure 1 Cloud Data Mining Architecture integrated with Apriori-HUDS

RESULTS AND DISCUSSION

The experiments used around 375 Data Sets in around 25 distributed servers. The model was developed similar to GrenchMark. The workloads are generated as synthetic workloads to the cloud based cluster servers and distributed environments. The proposed model is implemented and cloud-specific measures such as resource acquisition time, experiment cost, support, execution times with respect to threshold, memory usage and memory utilization are studied. The threshold is optimized in such a way to obtain useful frequently accessed datasets.

From Fig.2, it is clearly noticed that the Apriori-HUDS Rule Miner based proposed cloud data mining model ranked various data sets based on users' queries and demands and found that selected data sets are frequently used by more of the cloud users. The Support value of the proposed model ensures that the predicted rank for each and every dataset is fair. Fig.3 depicts the minimum support threshold analysis. Fig.4 illustrates the variation in the minimum support threshold to choose datasets versus the memory usage of the proposed cloud data mining architecture. Fig.5 shows the graph illustrating the relationship between ranked cloud data sets utilization and cloud users.

ISSN 2349-4506
Impact Factor: 2.785

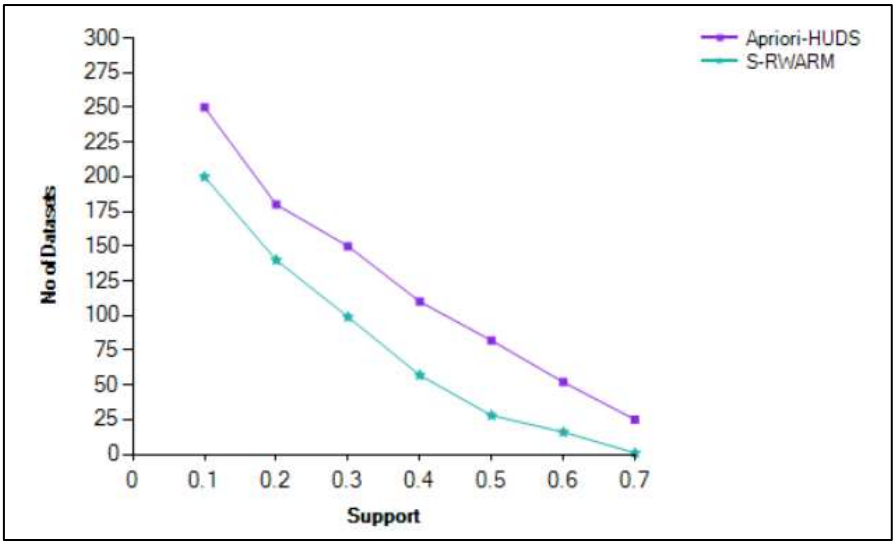

Fig.2 Efficiency of Dataset Selection based on Cloud Users' Queries

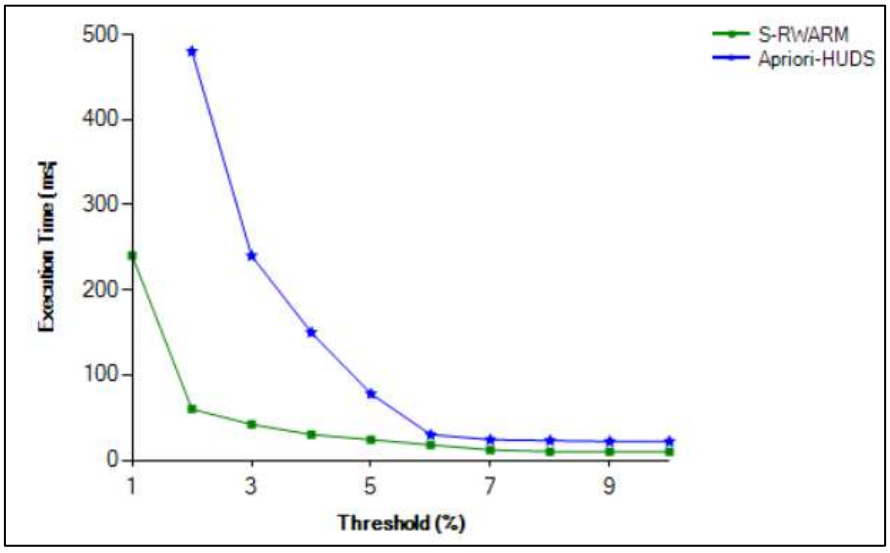

Fig.3 Minimum Support Threshold to Choose Datasets vs Execution Time for Cloud Users' Queries

ISSN 2349-4506
Impact Factor: 2.785

Global Journal of Engineering Science and Research Management

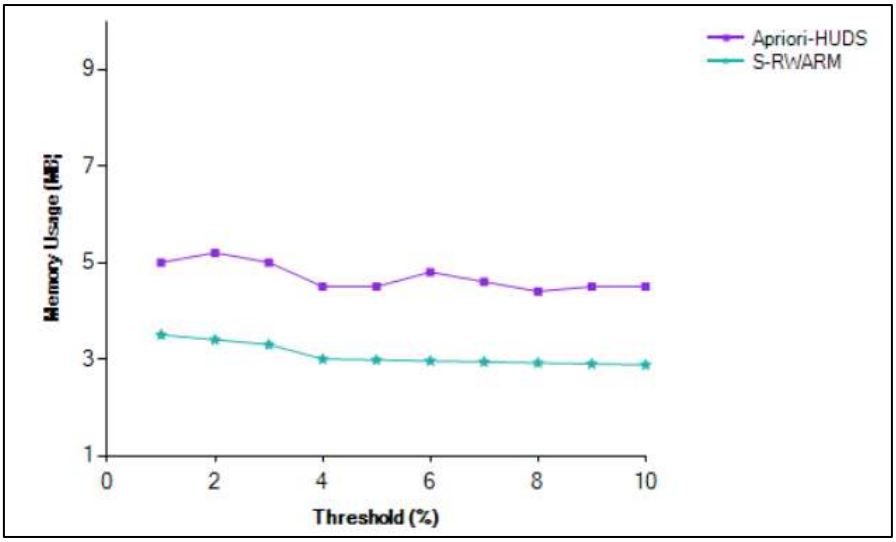

Fig.4 Minimum Support Threshold to Choose Datasets vs Memory Usage by Proposed Cloud Data Mining Architecture

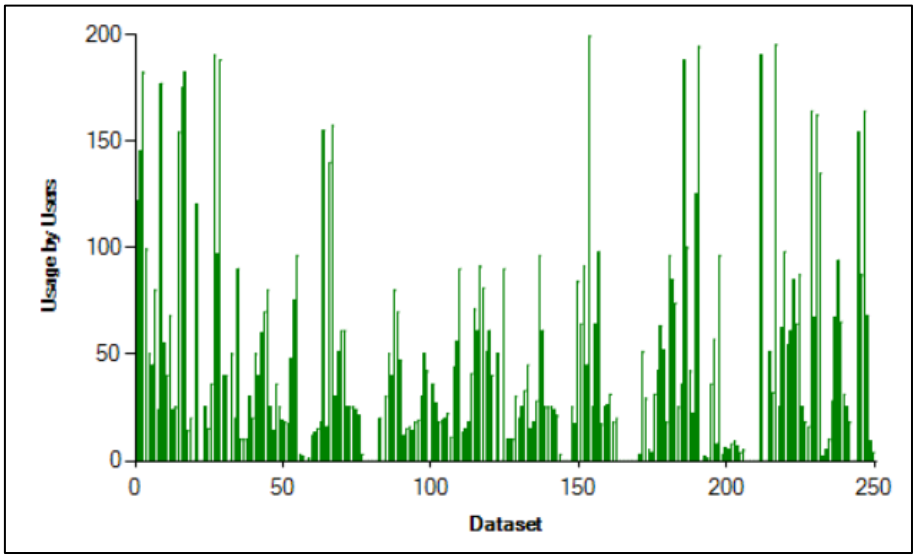

Fig.5 Ranked Cloud Data Sets Utilization vs Cloud Users

CONCLUSION

This research work has proposed an efficient secured cloud data mining model through Apriori-HUDS Rule Miner. This secured cloud data mining model is designed to facilitate cloud users to identify the best frequently used data sets from cloud data centers for their queries and security to the cloud environment. The proposed model is implemented in Cloud Environment and thoroughly studied. From our experimental results, it is noticed that the proposed model is performing well in terms of i. Predicting frequently accessed datasets by Users, ii. Ranking Datasets with highest Support and iii. Memory Utilization as compared with existing cloud data mining model.

REFERENCES

1. Y. Rahulamathavan, R. C.-W. Phan, S. Veluru, K. Cumanan, and M. Rajarajan, "Privacy-preserving multi-class support vector machine for outsourcing the data classification in cloud," IEEE Transactions on Dependable and Secure Computing, vol. 11, pp. 467-479, 2014.
2. W. Dai and W. Ji, "A mapreduce implementation of C4. 5 decision tree algorithm," International Journal of Database Theory and Application, vol. 7, pp. 49-60, 2014.

3. A. A. Ikram, S. Ibrahim, M. Sardaraz, M. Tahir, H. Bajwa, and C. Bach, "Neural network based cloud computing platform for bioinformatics," in IEEE Long Island Systems, Applications and Technology Conference (LISAT), 2013, pp. 1-6.
4. S.-Y. Lin, C.-H. Chen, and C.-C. Lo, "Currency exchange rates prediction based on linear regression analysis using cloud computing," system, vol. 6, 2013.
5. A. A. Zakharov, E. A. Olennikov, T. I. Payusova, and D. S. Silnov, "Cloud service for data analysis in medical information systems using artificial neural networks," International Journal of Applied Engineering Research, vol. 11, pp. 2917-2920, 2016.
6. F. Farahnakian, T. Pahikkala, P. Liljeberg, and J. Plosila, "Energy aware consolidation algorithm based on k-nearest neighbor regression for cloud data centers," in IEEE/ACM 6th International Conference on Utility and Cloud Computing (UCC), 2013, pp. 256-259.
7. Z. Ye, S. Mistry, A. Bouguettaya, and H. Dong, "Long-term QoS-aware cloud service composition using multivariate time series analysis," IEEE Transactions on Services Computing, vol. 9, pp. 382-393, 2016.
8. X. Yi, F.-Y. Rao, E. Bertino, and A. Bouguettaya, "Privacy-preserving association rule mining in cloud computing," in Proceedings of the 10th ACM symposium on information, computer and communications security, 2015, pp. 439-450.
9. Z. Qureshi, J. Bansal, and S. Bansal, "A survey on association rule mining in cloud computing," International Journal of Emerging Technology and Advanced Engineering, vol. 3, pp. 318-321, 2013.
10. X. Lin, "Mr-apriori: Association rules algorithm based on mapreduce," in 5th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2014, pp. 141-144.
11. F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, "Privacy-preserving mining of association rules from outsourced transaction databases," IEEE Systems Journal, vol. 7, pp. 385-395, 2013.
12. F. Guo, L. Yu, S. Tian, and J. Yu, "A workflow task scheduling algorithm based on the resources' fuzzy clustering in cloud computing environment," International Journal of Communication Systems, vol. 28, pp. 1053-1067, 2015.
13. D. Apiletti, E. Baralis, T. Cerquitelli, S. Chiusano, and L. Grimaudo, "Searum: A cloud-based service for association rule mining," in 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2013, pp. 1283-1290.
14. Q. Zhang, H. Liang, and Y. Xing, "A parallel task scheduling algorithm based on fuzzy clustering in cloud computing environment," International Journal of Machine Learning and Computing, vol. 4, p. 437, 2014.
15. Q. Zhang and Z. Chen, "A weighted kernel possibilistic c-means algorithm based on cloud computing for clustering big data," International Journal of Communication Systems, vol. 27, pp. 1378-1391, 2014.
16. W. Xiaojun, W. Yun, H. Zhe, and D. Juan, "The Research on Resource Scheduling Based on Fuzzy Clustering in Cloud Computing," in 8th International Conference on Intelligent Computation Technology and Automation (ICICTA),, 2015, pp. 1025-1028.
17. K. Srivastava, R. Shah, D. Valia, and H. Swaminarayan, "Data mining using hierarchical agglomerative clustering algorithm in distributed cloud computing environment," International Journal of Computer Theory and Engineering, vol. 5, p. 520, 2013.
18. D. Mittal, D. Kaur, and A. Aggarwal, "Secure Data Mining in Cloud Using Homomorphic Encryption," in IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2014, pp. 1-7.
19. S. Singh and I. Chana, "QRSF: QoS-aware resource scheduling framework in cloud computing," The Journal of Supercomputing, vol. 71, pp. 241-292, 2015.
20. Z. Liu, W. Qu, W. Liu, Z. Li, and Y. Xu, "Resource preprocessing and optimal task scheduling in cloud computing environments," Concurrency and Computation: Practice and Experience, vol. 27, pp. 3461-3482, 2015.