PSO-Enabled Privacy Preservation of Data Clustering

G. Asha Kiran¹*, Manimala Puri² and S. Srinivasa Suresh³

¹MBA Department, Rajarshi Shahu College of Engineering, Survey No.80, Pune-Mumbai Bypass Highway, Tathawade, Pune - 411033, Maharashtra, India; ashakiran45@rediffmail.com
²JSPM Group of Institutes, Survey No.80, Pune-Mumbai Bypass Highway, Tathawade, Pune - 411033, Maharashtra, India; manimalap@yahoo.com
³CSE Department, KMIT, 3-5-1026, Narayanguda, Hyderabad - 500029, Telangana, India; sssuresh74@gmail.com

Abstract

Background/Objective: Privacy is the vital issue when sharing of the data comes into picture. The demand and scope for privacy is increasing day-by-day as data storage techniques have emerged from standalone database to distributed database and then progressed to parallel databases. K-means and Fuzzy C-means (FCM) are the frequently used clustering algorithms for standalone database, distributed database and parallel databases. The current paper highlights Particle Swarm Optimization algorithm along with Fuzzy C-means clustering algorithm technique for preserving the privacy on distributed databases. Methods/Statistics Analysis: The experimentation is performed by means of the datasets accessible in the UCI machine-learning repository. The main benefit of the suggested technique is that, this technique will assess in terms of their privacy of cluster. Therefore, the technique plans to give improved visibility for the protected data. The technique is executed in the working platform of MATLAB and the effects will be examined to show the presentation of the suggested clustering technique. Findings: The performance of the proposed clustering technique based on privacy preserving is analyzed for accuracy and Database Different Ratio (DBDR) on six UCI medical related data sets namely Hugerian dataset, Cleveland data set, Reprocessed Hugerian data sets, Long Beach V.A data, BUPA and liver disorder data. Performance improvement observed in the range of 3%-6% on each of the six data sets compared to K-means algorithm. Application/Implementation: The main benefit of the suggested technique is that technique will have to assess in terms of their privacy of cluster. Therefore, the technique plans to give improved visibility for the protected data.

Keyword: Clustering, Distributed Data, K-means, PPSSI, PSO

1. Introduction

The amount of data storage and processing was raised by the technology development in storage and processing. Data mining has turn out to be an essential assignment at present in all fields to remove the unseen constructive information¹. The outcome of data mining (also called knowledge) is useful for improving decision making capacity, and in turns our quality of life. On the one hand, such data is a vital asset to enterprises and governments for decision-making processes and to offer social advantages, such as crime reduction, medical research, national security etc.². In addition, Privacy Preserving Data Mining (PPDM) is one of the significant areas of data mining that plans to offer security for secret information from unsolicited or unsanctioned revelation. Data mining techniques examines and forecasts constructive information. On the other side, examining such data becomes threat to privacy. The idea of privacy preserving data mining is mainly concerned with defending secret data against unsolicited access. It is vital issue nowadays. Data mining techniques are competent to forecast high sensitive information from large data⁴. However, new large data processing techniques are emerging at a high pace. In this regard, distributed and parallel processing paradigms are gaining the trend. The ability of applying data mining techniques on large volume of data stored on distributed storages give new dimensionality to researchers. To discover significant information, more number of data mining techniques have been developed, the manner of mining the data from the database is not concern about the privacy⁵. In order to avoid that, the PPDM is brought

*Author for correspondence
in, to sustain the privacy while mining the data. All over
the medical, hospitals, and clinics, enclose vast amount
of patient data and they require sustaining their informa-
tion. This data enclose complete medical data about each
of the patient.

For illustration, database contains characteristics like
patient id, disease id, time duration of the disease etc.
Currently numerous privacy preserving approaches are
obtainable in the data mining field such as randomiza-
tion method, k-anonymity model, l-diversity, sampling,
cell suppression, data swapping and perturbation. Not
many researchers have indicated the limitations of cryp-
tography in the area of privacy preserving data mining.
As a result, cryptography falls short of offering a complete
result in PPDM.

The k-anonymity model of privacy was regarded
intensively in the context of public data sharing. In data
sharing approach, commonly, databank owner hopes ano-
nymy to be maintained after data mining operation. As
a result, to decrease the risk of this kind of attack,
k-anonymity has been suggested. Hence, the most impor-
tant purpose of k-anonymization is to secure the solitude
of the individuals to whom the data relates. On the other
hand, it is as well necessary that the liberated data stay
as “useful” as feasible subjecting to this constraint. In
the literature, several recoding models have been sug-
gested for k-anonymization incessantly. On the other
hand, clustering is commonly used statistical technique
in business applications and many others. Clustering is
a separation of data into groups of related objects. Each
group, called cluster, contains objects that are alike among
themselves and different to objects of other groups. K-
Means, the most familiar and generally employed part-
titioning method, applied for the privacy preserving. By
employing fuzzy sets for privacy preserving, we can carry
out a slow evaluation of the data set specified to us and
this is prepared by employing a fuzzy membership func-
tion. Each lexical term can be symbolized as a fuzzy set
containing its own membership function. Fuzzy c-means
clustering is the one of the clustering approach frequently
adopted in the development of privacy preserving data
mining applications. However any element in the set may
have membership in more than one group.

The remaining of the paper is organized as follows:
the recent research works is determined in section 2;
proposed work described section 3; in the section 4, the
experimental output are described and the section 5 re-

2. Related Work

Different researchers have suggested many approaches
for privacy preserving in data mining. Among them a
handful of significant researches are presented in this
segment; Substitution cipher techniques has been sug-
gested by W.K.Wong et al. The author’s et al highlighted
encryption of transactional data for outsourcing asso-
ciation rules. After recognizing the non-trivial threats
to a straightforward one-to-one item mapping substitu-
tion cipher, they suggest a more secure encryption
scheme based on a one-to-n item mapping that converts
transactions non-deterministically, still promises right
decryption. They improved efficiency algorithm based
on this method. The algorithm carried out a single pass over
the database and appropriate for applications in which
data owners send streams of transactions to the service
provider. The effects demonstrated that the technique was
extremely secure with a low data transformation cost.

Moreover, an approach for preserving privacy in asso-
ciation rule mining has been suggested by Ling Qiu et
al. The most important plan was to employ keyed Bloom
filters to symbolize transactions and data items. The
approach can completely protect privacy while uphold-
ing the accuracy of mining results. The tradeoff between
mining precision and storage requirement was examined.
They as well suggested δ -folding technique to further
decrease the storage requirement without sacrificing
mining precision and running time.

Authors Jun Lin Lin and Yung Wei Cheng highlighted
the problem of privacy-preserving mining of frequent item
sets. They offered a process to defend the privacy of data
by attaching noisy items to each transaction. To renovate
frequent itemsets an algorithm is suggested from these
noise-added transactions. The experimental effects pointed
out that the method could accomplish a rather high level
of precision. For frequent itemset mining the method uses
presented algorithms, and thus takes full benefit of their
progress to mine frequent itemset competently.

Additionally, privacy-preserving distributed asso-
ciation rule mining protocol has been offered by Fun Yi
and Yunchun Zhang based on a new semi-trusted mixer
model. The protocol can guard the privacy of each distrib-
uted database against the coalition up to n-2 other data
sites or still the mixer if the mixer does not get together
with any data site. In addition, the protocol requires only
two communications among each data site and the mixer
in one round of data collection. Moreover, Chunhua Su
and Kouichi Sakurai\(^{23}\) have spotlighted on the privacy issue of the association rules mining and offered a secure frequent-pattern tree (FP-tree) based scheme to safeguard private information while doing the collaborative association rules mining. They demonstrated that their plan was protected and collusion-resistant for parties, which means that even if n-1 dishonest parties get together with a deceitful data miner in an effort to learn the association rules among honest respondents and their responses, they will be not capable to triumph.

In, Saeed Samet and Ali Miri\(^{23}\) have developed a model for preserving the horizontal and vertically portioned data using Neural Network techniques: Back Propagation and Extreme machine learning algorithm. Neural Network techniques emerging rapidly in areas where prediction is necessary. Here, the author’s primary objective is to perform prediction based on input data (online internet records) while preserving the privacy of incoming data and learning model. Also, the authors shared this work with all concerned, who can utilize it mutually to anticipate the comparing yield for their objective information.

In\(^{32}\) message passing through clusters method is proposed. It is a dynamic variant of AP clustering called Incremental Affinity Propagation with K-Medoid (IAPKM) is used along with the Fuzzy Density based clustering (DENCLUE) method. In Fuzzy DENCLUE, number of clusters is reduced and randomness in the form of noise is removed. The evaluation result defines that the effectiveness and efficiency of IAPKM and Fuzzy DENCLUE achieves comparable performance. The authors compared proposed method with IAPKM and Fuzzy DENCLUE. The proposed approach achieves 10% greater accuracy rate than the former approach.

In\(^{32}\) authors proposed nearest Neighbor Density based Clustering Approach on a Proclus Method to Cluster High Dimensional Data. Generally, clustering high data sets needs an efficient algorithm such as Proclus. The Proclus algorithm ignores cluster with small data points. The authors propose an ensemble of clustering that combines technique of two clustering algorithms to achieve a quality cluster of even small data points.

In\(^{32}\), the authors focuses on Clustering and Sequential Pattern Mining techniques. The objective is to mine small projected databases rejected by Frequent Pattern - Projected Sequential Pattern mining (Free Span) technique using a weighted distance metric clustering method; a process of finding the distance between the small data points and cluster it so that it cannot be rejected.

In\(^{32}\), the authors proposes a graph clustering method on a directed weighted graph network for detecting communities in social network, based on neighborhood nodes and the frequency of the path traversed. The algorithm is evaluated with small, medium and large sized data. The algorithm is experimented with the data and its performance is evaluated in terms of cluster quality. The results work represent a novel method for identifying quality clusters with high intra-similarity and low intersimilarity.

In\(^{32}\), the authors\(^{21}\) proposed a framework for improving the classification of customer reviews on products. It is a recommender system. Online data analysis and applying predictive techniques gaining demand day-by-day. Majorly, the work involves text processing, constraint based association rules ontology model and improved K-means algorithm. The major work done is consolidating reviews arrived from multiple sources including symbols.

### 3. Proposed Vertically Distributed Data Clustering using Fuzzy C-Means Algorithm

Clustering of allocated multiple database is a vigorous research. Formerly, the k-means algorithm is employed for clustering the distributed multiple database which has various disadvantages such as high computation cost and low clustering precision. Hence, in this document we suggest a distributed data clustering by means of FCM based on the particle Swarm Optimization Algorithm. In our suggested system we employ distributed medical data for clustering by means of KFCM algorithm. The medical data employed here is perpendicularly distributed data which must be safeguarded for privacy. Now, we employ a Particle Swarm Optimization algorithm to raise the level of privacy and to get improved result. Our suggested technique contains three imperative stage of processing: Input data, Dissimilarity matrix, and Encryption. The fundamental block architecture of the suggested privacy preserving distributed data clustering is exposed in the Figure 1 beneath.

#### 3.1 Input Data

Let us reflect on N medical data sites such as \(D_{S_1}, D_{S_2}, \ldots, D_{S_m}\) which are specified as input for clustering. In this each data sites \(D_{S_i}\) has \(n_i\) tuples. Each and every input medical data sites contain similar schema with \(m\)
numeric attributes. As all medical data sites contain the similar schema, union of the databases can be stated as $u_1, u_2, \ldots, u_p$ where $P = \sum_{i=1}^{N} n_i$ and where object $o_i$ has attribute values $a_1, a_2, \ldots, a_m$. The group database is vertically partitioned among $N$ medical data sites. The vertically partitioned database is database $D = \{a_{i1}, a_{i2}, \ldots, a_{im}\}_{1 \leq i \leq n, a_{11} \in R_1, \ldots, a_{m1} \in R_m}$, which is perpendicularly distributed among $m$ parties $P_{11}, P_{22}, \ldots, P_{mm}$ if $P_{ij}(1 \leq j \leq m)$ holds $D_{ij} = \{a_{ij}\}_{1 \leq i \leq n}$, where $R$ is the domain of $a_{ij}$ for $1 \leq i \leq n$. The subsequent stage of our suggested privacy preserving data mining is to compute the dissimilarity matrix of all objects in all databases.

3.2 Dissimilarity Matrix

A dissimilarity matrix accumulates a collection of proximities that are accessible for all pairs of objects. It is produced based on the perpendicularly portioned input database. This matrix is frequently symbolized by a $m \times m$ table.

In the above Figure 2., dissimilarity matrix, $x$ each element $d(i, j)$ symbolizes the difference among object $i$ and $j$. Now we employ Euclidean distance measure to compute the dissimilarity among object $i$ and $j$. If object $i = (x_{i1}, x_{i2}, \ldots, x_{im})$ and $j = (x_{j1}, x_{j2}, \ldots, x_{jm})$ are $n$ dimensional data objects, the Euclidean distance among objects $i$ and $j$ is computed based on the equation specified beneath

$$d(i, j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \quad (1)$$

Before clustering, one of the important thing in distributed data clustering is privacy preserving. For this now we encrypt the data based on the arbitrarily produced filter coefficient.

3.3 Encryption based on Filter Coefficient

In this, encryption is prepared based on the randomization process which avoids the user from learning sensitive data which can be simply executed as the filter coefficient generate from the initial solution of PSO is employed to the input medical data is independent from the other records. The size of the filter coefficient matrix is relying on the initial solution size which is big enough to conceal the original values. The randomization method is easy as compare to other methods as it does not need to knowledge of other records. Large randomization raises the ambiguity and user’s personal privacy. They declare that approaches may misplace data and not offer privacy by introducing arbitrary noise to the data by employing random matrix properties. It effectively divides the data from the filter coefficient and next releases the original data.

In our suggested system, vertically distributed medical data clustering filter coefficient is produced arbitrarily based on the input data for encryption purpose. Thus, the privacy of the input perpendicularly distributed medical data can be safeguarded.

A filter coefficient matrix is produced arbitrarily based on the initial solution of the PSO is exposed in the Figure 3 beneath.

For instance, we produce a $3 \times 3$ filter coefficient matrix arbitrarily to apply over the medical data for encryption purpose. Hereafter, the produced filter coefficient is employed over medical data and we obtain an encrypted medical data output which is specified as input for clustering. In the Figure 3, FC1, FC2, FC3, FC4, FC5, FC6, FC7, FC8 and FC9 symbolize the filter coefficient. At last we obtain an encrypted data for cluster. In our distributed data clustering system we employ a clustering method called fuzzy c-means clustering. We assume readers are aware of C-Means clustering algorithm.

3.4 Privacy Preserving Clustering using FCM based on PSO Algorithm

Step by step process of our suggested privacy preserving clustering algorithm is exposed beneath,

**Step 1**

The first stage of our suggested method is solution initialization. Now $N$ number of solution is initialized arbitrarily with a length $m$. further for each arbitrarily initialized solution we work out the fitness or clustering.
Step 2
For every particle, find out the consequent clustering center based on the value of first position. Next, the value of objective function can be calculated by the particle and its related clustering centers using the equation specified beneath,

\[ J_m(U, E) = \sum_{k=1}^{n} \left( \mu_k \right)^m \| x_k - e_i \|^2 \]  

(2)

In the above equation, \( m \) is constant and \( m > 1 \). Where \( e_i \) is cluster \( i = (1, 2, \ldots, c) \), \( k = (1, 2, \ldots, n) \).

For each particle \( i \),
- Store \( P_{best} [i] \) (best position) and current_particle_fitness [i],
- Select best _particle_fitness [i] and call it as \( G_{best} \).

Step 3
Find updated position and velocity based on the equation (3)

Step 4
Find the next clustering center based on each particle value. Subsequently, find the fitness and consequent clustering centers.

Step 5
This step finds the each particle individual best
- For each particle \( i \)
  - Compare (particle_fitness [i], earlier fitness \( P_{best} \)),
  - If particle_fitness [i] is better than \( P_{best} \) Then particle [i] = \( P_{best} \)

Step 6
This step finds the group best
- For each particle \( i \)
  - Compare (particle_fitness [i], group earlier fitness best fitness)
  - If particle_fitness [i] is better, particle_current_position [i] = \( G_{best} \).

Step 7
Search for preconditions. If preconditions are not satisfied then go to step3, else stop iterations and write the optimal output.

4. Result and Discussion

The proposed privacy preserving method is implemented on MATLAB tool. The current experiment uses six datasets for the processing privacy preserving, which are: Cleveland, Switzerland, Hungarian, Reprocessed Hungarian, Long Beach V.A data and BUPA liver datasets. The whole experiment is carried out on a PC with an Intel i5 processor having 4GB of main memory as this is commonly available configuration now a days.

4.1 Evaluation Metrics

The evaluation of clustering technique using privacy preserving data mining is carried out using the following metrics as suggested by below equations,

4.1.1 Clustering Accuracy

Clustering Accuracy is the measure of closeness of the cluster shaped as a consequence of the proposed calculation to the required value which implies how much exact the individuals from a cluster are. In our paper the clustering accuracy is computed using the following formula.

\[ CA = \frac{1}{N} \sum_{i=1}^{T} X_i \]  

(3)

Where, \( N \) is the number of data point and \( T \) is the number of classes.

4.1.2 Database Difference Ratio

Database difference ratio is defined as the dissimilarity between the original database \( D \) and the encrypted database \( D' \) is calculated as DBDR

\[ DBDR = \frac{|D - D'|}{|D|} \]  

(4)
Table 1. Description of Dataset

| Dataset                | Type of Disease | No. of Instances considered for analysis | No. of Features considered for analysis | Class Distribution |
|------------------------|-----------------|-----------------------------------------|----------------------------------------|-------------------|
|                        |                 |                                         |                                        |                   |
| Cleveland Data         | Heart Disease   | 244                                     | 14                                     | % of instances with disease absent (%) |
|                        |                 |                                         |                                        |                   |
| Switzerland Data       | Heart Disease   | 123                                     | 14                                     | % of instances with disease absent (%) |
|                        |                 |                                         |                                        |                   |
| Hungarian Data         | Heart Disease   | 257                                     | 14                                     | % of instances with disease absent (%) |
|                        |                 |                                         |                                        |                   |
| Re-Processed Hungarian Data | Heart Disease | 294                                     | 14                                     | % of instances with disease absent (%) |
|                        |                 |                                         |                                        |                   |
| Long Beach V.A data    | Heart Disease   | 200                                     | 14                                     | % of instances with disease absent (%) |
|                        |                 |                                         |                                        |                   |
| BUPA data              | Liver Disorder  | 345                                     | 7                                      | % of instances with disease absent (%) |

4.2 Dataset Description

The proposed system is experimented with the six types of dataset such as Cleveland, Switzerland, Hungarian, Reprocessed Hungarian, Long Beach V.A data and BUPA liver dataset. These are benchmark datasets and are taken from the UCI machine learning repository. The description of the dataset is given below:

4.3 Performance Evaluation

The basic idea of our research is to design and develop a technique for Clusters Technique Using Privacy Preserving Data Mining. Here, we utilize the Fuzzy C Means (FCM) algorithm for the clustering process and further utilize the PSO algorithm for improve the accuracy of the system. The privacy preservation is providing the proper balance between the privacy protection and knowledge discovery. Here, we prove our work efficiency we compare our work with k-means clustering algorithm. The following figures 4 to 15 shows the performance evaluation plots on the chosen data sets.

4.4 Findings

The performance of the proposed clustering technique based on privacy preserving is analyzed with the help of accuracy and Database Different Ratio (DBDR) and the same is shown in the Figures 4 to 15. The clustering accuracy mappings are plotted by varying the clustering size between 2 and 10.

Moreover, Figure 10 to 15 shows the performance of data base different ratio (DBDR) for different dataset. When analyzing the Figure 14, we obtain the maximum DBDR of 11% for using proposed approach FCM and
Figure 6. Performance Evaluation of accuracy Plot for Hungarian Dataset.

Figure 7. Performance Evaluation of Accuracy Plot for Reprocessed Hungarian Dataset.

Figure 8. Performance Evaluation of Accuracy Plot for Switzerland Dataset.

Figure 9. Performance Evaluation of Accuracy Plot for Long Beach V.A Dataset.

Figure 10. Performance Evaluation of DBDR for BUPA Liver Dataset.

Figure 11. Performance Evaluation of DBDR for Cleveland Dataset.
Figure 12. Performance Evaluation of DBDR for Hungarian Dataset.

Figure 13. Performance Evaluation of DBDR for reprocessed Hungarian Dataset.

Figure 14. Performance Evaluation of DBDR for Switzerland Dataset.

Figure 15. Performance Evaluation of DBDR for Long Beach V.A Dataset.

Table 2. Performance comparison: Proposed Technique Vs K-means

| Sno | Data set                   | Clusters | Accuracy | Comparison with K-means |
|-----|----------------------------|----------|----------|------------------------|
| 1   | BUPA liver dataset         | 10       | 44%      | 34%                    |
| 2   | Cleveland data             | 7        | 14%      | 8%                     |
| 3   | Hungarian dataset          | 9        | 15%      | 12%                    |
| 4   | Reprocessed Hungarian      | 9        | 15%      | 12%                    |
| 5   | Switzerland                | 10       | 16%      | 13%                    |
| 6   | Long Beach V.A data        | 9        | 13%      | 10%                    |

4% for using existing approach K-means clustering. In Figure 11, when using the cluster size is 7 we achieve the maximum DBDR is 8.5% for proposed approach using Cleveland dataset. In Figure 15, we use performance evaluation of DBDR for Long Beach V.A dataset. Here, we obtain the maximum DBDR of 13% which value is high compare to existing approach. Overall, the performance of the proposed technique is better compared to the k-means clustering using privacy preserving data mining using different dataset.

5. Conclusion

In this paper we have considered distributed medical data as input for clustering. Together with this a filtering
based encryption technique was used for the purpose of privacy preservation while clustering the input medical data. Also, here we have used an optimization algorithm called Particle Swarm Optimization algorithm. It optimizes the filter coefficient by ‘n’ number of iteration. This analysis has shown that our method protects privacy in every round of iteration of FCM clustering as long as the underlying cryptographic code is secure. The experimental result of our proposed method shows better result.

6. References

1. Diesburg SM, Wang A. A survey of confidential data storage and deletion methods. ACM Computing Surveys (CUSR). 2010; 43(1). Available from: Crossref
2. Aggarwal C, Yu P. Springer-Verlag: USA: Privacy-Preserving Data Mining Models and Algorithms. 2008. Available from: Crossref
3. Vaidya J, Clifton C. Privacy preserving k-means clustering over vertically partitioned data. The 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2003; p. 206-15. Crossref
4. Aruna Kumari D, Rajasekhar Rao K, Suman M. Chennai: International Conference on CNSA: Privacy preserving distributed data mining using Steganography. 2010; p. 263-69.
5. Anuradha T, Suman M, Aruna Kumari D. Data obscuration in privacy preserving data mining. ICWS’09: Proceedings International conference on web sciences. 2009.
6. Agrawal R, Srikant R. Privacy Preserving Data Mining. USA: Dallas, Texas: Proceedings of ACM SIGMOD Conference on Management of Data (SIGMOD’00). 2009; p. 439-50. PMid:19207352 PMCID:PMC2697821
7. Patel M, Richiari P, Shrivastava A. A review paper on Privacy Preserving Data Mining. Compusoft: International Journal of Advanced Computer Technology. 2013; 2(9):296-99.
8. Charu C. Aggarwal TJ. On k-Anonymity and the Curse of Dimensionality. Norway: Trondheim: Proceedings of the 31st VLDB Conference. 2005; p. 901-09.
9. Machanavajjhala A, Kifer D, Gehrke J, Venkita Subramanian M. ε-Diversity: Privacy Beyond K-Anonymity. Journal of ACM Transactions on Knowledge Discovery from Data (TKDD). 2007; 1(1).
10. Fung B, Wang K, Chen R, Yu P S. Privacy Preserving Data Publishing: A Survey of Recent Development. ACM Computing surveys. 2010; 42(4).
11. Bayardo RJ, Agrawal R. Data privacy through optimal k-anonymization. ICDE’05: 21st International Conference on Data Engineering. 2005; p. 217-28. Available from: Crossref
12. Bertino E, Ooi BC, Yang Y, Deng RH. Privacy and ownership preserving of outsourced medical data. ICDE’05: 21st International Conference on Data Engineering. 2005; p. 521-32. Available from: Crossref
13. Mukherjee S, Chen Z, Gangopadhyay A. A privacy-preserving technique for Euclidean distance-based mining algorithms using Fourier-related transforms. The VLDB Journal. 2006; 15(4):293-315. Available from: Crossref
14. Kantarcioğlu M, Clifton C. Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. IEEE Transactions on Knowledge and Data Engineering. 2004; 16(9):1026-37. Available from: Crossref
15. Wang J, Zhang J, Xu S, Zhong W. A Novel Data Distortion Approach via Selective SSVD for Privacy Protection. International Journal of Information and Computer Security. 2007; 2(1):48-70. Available from: Crossref
16. Kotsiatsis S, Kanellopoulos D. Association Rules Mining: A Recent Overview. GESTS International Transactions on Computer Science and Engineering. 2006; 32(1):71-82.
17. Kaoru S. SAS Institute: SAS Institute Best Practices Paper: Data Mining and the Case for Sampling. 1999; 18:361-80.
18. Kargupta H, Chan P. USA: MIT Press: Advances in distributed and parallel data mining. 2000. PMid:11017408
19. Zaki M, Ho C. (Eds.). Large-scale parallel data mining. Springer-Verlag: Berlin Heidelberg: Lecture Notes in Artificial Intelligence. 2000; 1759:1-8.
20. ChiaT, KannapanS. Strategically Mobile Agents. Germany: Berlin: Proceedings of the First International Workshop on Mobile Agent. 1997; p. 149-61. Available from: Crossref
21. Krishnaswamy S, Loke SW, Zaslavsky A. Cost Models for Distributed Data Mining. USA: Chicago: Proceedings of the 12th International Conference on Software Engineering & Knowledge Engineering, 2000; p. 1-8.
22. Han J, Kamber M. Data Mining: Concepts and Techniques. CA: San Francisco: Morgan Kaufmann Publishers. 2001.
23. Ross TJ. USA: McGraw Hill International Editions: Fuzzy Logic with Engineering Applications. 1997.
24. Wong WK, Cheung DW, Hung E, Kao B, Mamoulis. Security in outsourcing of association rule mining. Austria: Vienna: Proceedings of the 33rd International Conference on Very Large data bases. 2007; p. 111-22.
25. Qiu L, Li Y, Wu X. Preserving privacy in association rule mining with bloom filters. Journal on Intelligent Information Systems. 2007; 29(3):253-78. Available from: Crossref
26. Lin JL, Cheng Y-W. Privacy preserving item set mining through noisy items. International Journal on Expert Systems with Applications. 2009; 36(3):5711-17.
27. Yi X, Zhang Y. Privacy-preserving distributed association rule mining via semi-trusted mixer. Data and Knowledge Engineering. 2007; 63(2):550-67. Available from: Crossref
28. Su C, Sakurai K. A Distributed Privacy-Preserving Association Rules Mining Scheme Using Frequent-Pattern Tree. Berlin Heidelberg: Springer-Verlag: Advanced Data Mining and Applications. 2008; p. 170-81. Available from: Crossref

29. Samet S, Miri A. Privacy-preserving back-propagation and extreme learning machine algorithms. Data and Knowledge Engineering. 2012; 79-80:40-61. Available from: Crossref

30. Suganya M, Nagarajan S. Message Passing in Clusters using Fuzzy Density based Clustering. Indian Journal of Science and Technology. 2015 July; 8(16). Crossref

31. Gayathri S, Mary Metilda M, Sanjai Babu S. A Shared Nearest Neighbor Density based Clustering Approach on a Proclus Method to Cluster High Dimensional Data. Indian Journal of Science and Technology. 2015 Sep; 8(22). Crossref

32. Gayathri S, Mary Metilda M, Sanjai Babu S. A Weighted Distance Metric Clustering Method to Cluster Small Data Points from a Projected Database Generated from a Free span Algorithm. Indian Journal of Science and Technology. 2015 Sep; 8(22). Crossref

33. Parimala M, Daphne Lopez S. Kaspar. K-Neighbourhood Structural Similarity Approach for Spatial Clustering. Indian Journal of Science and Technology. 2015 Sep; 8(23). Crossref

34. Razia Sulthana A, Ramasamy Subburaj. An Improvised Ontology based K-Means Clustering Approach for Classification of Customer Reviews. Indian Journal of Science and Technology. 2016 Apr; 9(15). Crossref