IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2017

Hybrid Model for Privacy Preservation Using Clustering and Classification Technique

Priyank Pathak[1], Rajat Paliwal[2], Chetan Grawal[3]

M.Tech Scholar, Department of Computer Science and Engineering, RITS, Bhopal,India[1]

Supervisor, Department of Computer Science and Engineering RITS, Bhopal, India[2]

HOD, Department of Computer Science and Engineering RITS, Bhopal, India[3]

Abstract: The security of data over the internet is major issue in current scenario. For the security of data used various technique such as cryptography, steganography and watermarking. Privacy preservation is new generation of data security. For the privacy preservation used various methods and algorithm. In this paper proposed the data mining based privacy preservation algorithm. The proposed algorithm is combination of clustering and classification. The process of clustering used the vector decomposition technique for the portioning of data. The partition and distributed data transfer the classifier for the validation and publication of data. The proposed algorithm implemented in MATLAB software and used reputed dataset form UCI machine learning. Our experimental result shows the better value instead of pervious methods.

Keywords: Privacy Preservation, Data Mining, Clustering, Classification, Vector

I. INTRODUCTION

The privacy of data over the internet is big challenge in current scenario. For the privacy of data used security constrains based function and algorithm. For the preservation of privacy used some conventional technique such as cryptography, data potation, noise addition and many more algorithm. The applicability and diversity of data mining increase in different filed of data science [1,2]. In concern of that data mining technique is used for the process of privacy and preservation. The data mining gives two most promising algorithm clustering and classification. The process of clustering and classification imposed the data transformation and validation of data in terms of publication. In scenario of privacy preservation various process are used such as PPDM [3,4,5]. The PPDM frameworks is to keep up the information respectability of the information distributed and to accomplish effective information mining comes about .there is the requirement for an exceptionally secured framework for averting psychological warfare, the information accumulation and investigation of gathered information from the migration office is required to be done from each taking an interest country for breaking down the groups and track renegades or fear based oppressors to counter any dread exercises. The proposed algorithm utilized KNN base classifier and thickness based bunching strategy. The KNN classifier characterized information on the premise of closest neighbor. The thickness based bunching system assemble the information as indicated by their epic esteem. The epic esteem chooses the scope of information as per their predefined procedure. [6,7] For the disintegration of information and change of information utilized single esteem vector deterioration system. The single esteem deterioration strategy is procedure of information disintegration without loss of information. The information change procedure is a versatile strategy of information randomization for the blending of information as indicated by the way toward grouping and characterization. In worry of information mining application as security saving different methods are utilized, for example, affiliation control mining, grouping strategy and order system. And furthermore utilized a few information blended method for versatile clamor information in unique information [8,9].

Lattice decay is huge part in protection saving in information mining order. The sorts of grid decay are even vertical and slanting of file information of protection. In information mining application, the utility of outsider has been evacuated. During the time, spent network disintegration particular and different qualities are included. The particular esteem decay keeps the loss of blended information and removed information in deterioration of framework. Test choice keep up proportion of information between blended information and unique information amid handling of grouping. The proportion of test determination is 1:3 as a matter of course procedure of test choice. In this exposition, we proposed a half breed grouping method for protection saving procedure for information arrangement [10, 11]. In half and half arrangement is blend of bunching and grouping strategy such techniques are called troupe classifier this paper is divided into five sections. Section-1. Gives the introduction of privacy preservation and data mining. Section-2. Gives the information about SVD. Proposed method in section-3... In section-4. Discuss experimental work and finally discuss conclusion and future work in section 5.

DOI 10.17148/IJIREEICE.2017.51016

II. SINGLE VALUE DECOMPOSITION (SVD)

the singular value decomposition (svd) is a matrix of data factorization for the value of data transformation. the single value decomposition reduces the size and dimensions of data. the svd methods used the partition of data table for the processing of clustering algorithm. the utilization of svd strategies in information annoyance for protection saving information mining is proposed in [12,17]. the svd of the first $n * m$ information framework an is composed as here u is a $n * n$ ortho-typical network, s = diag[σ1, ..., σs], where s = min(n, m), without the loss of sweeping statement, and nonnegative corner to corner sections σis are in a non-expanding request. the corner to corner sections σ1, ..., σs are known as the solitary qualities. what's more, v t is likewise an ortho-typical network with measurement $m * m$. the quantity of nonzero corner to corner sections of s is equivalent to the rank of the network a. characterize
ak = ukskv t
k ; for a positive number k min(n;m); where uk just contains the main k segments of u, sk contains the principal k nonzero particular estimations of s, and v t k contains the primary k lines of v t. clearly, the rank of the grid ak is k, and ak is frequently called the rank-k truncated svd. ak has an outstanding property that it is the best k-dimensional (rank-k) guess of an as far as the frobenius standard. in data recovery, ek = a - ak can be considered as the commotion of the first information lattice. in security protecting information mining, ak can be utilized as a bothered adaptation of a [15]. thus, ak speaks to a decent estimation which keeps comparative examples of a, while it gives security to information protection.

III. PROPOSED ALGORITHM

The hybrid model is combination of clustering and classification technique. The clustering technique used the SVD data. The SVD function reduces the dimensions of data matrix and process of attribute grouping in single point. The single point transform the data in cluster for the privacy. The clustered data select and mapped according to their classes for the publication of data.

```
Input: data matrix for the process of preservation
Output : a mixed transform table data
class: E={},the set of the equivalence classes QIC={},set of equivalence classes with similar QI sets
CIP{}, set of attributes with similar class
DIP=number of different class values in the remaining dataset
Begin
While CIP >= attribute
Cluster T to m tables according QI For i=1 to m
Bucketize attributes according SA values
While |DIPi|>=ℓ
Create_equivalence_classes ()
E=E U Create_equivalence_classes()
return E
Incorporate the remaining attributes to E End
Generate equivalence class with prototype is
Input: CIP Output :E Begin
Randomly selection of a attributes tm from the smallest group
E={tm}
For p=1 until attribute-1
Select a attributes tp that minimizes the gcp
E=E U tp
Remove tp from T Remove tm from T Return E
End
Process of cluster generation in prototype classification
Input: data set used defined
Output: QIC={}, set of tables with attributes with similar QI sets
Begin
Insert T to the decision tree classification
QIC={ QIC1, QIC2,… QICm }
return QIC End.
```

IV. EXPERIMENTAL RESULT

For the evaluation of performance of proposed method used MATLAB software and two well know data set are used one is breast cancer dataset and another dataset is glass dataset. These datasets obtained from UCI machine learning repository. For the validation of result used three parameter such as utility measure, CP and accuracy. This parameter shows that effectiveness of proposed method [15].

For the performance evaluation procedure Data, has been taken from the University of California (UCI). To perform experiment work two datasets has been taken is Breast Cancer dataset and Glass Dataset. We used some parameters to measures the performance of data mining techniques with parameters for the privacy preservation classification, the description of used some parameters are given below.

UM (Utility Measures): The data utility measures assess whether a dataset keep the performance of data mining technique after the data distortion.

CP: To define change at the rank of the arrange value of the attribute.
$i=1$

TABLE I COMPARATIVE PERFORMANCE EVALUATION

Type of method	Cp	Accuracy	Elapsed time
Cc	74.45	89.93	5.31
Proposed method	67.45	87.38	5.46
Cc	65.93	88.64	3.80
Proposed method	68.24	90.64	5.38
Cc	67.56	89.00	5.55
Proposed method	69.45	91.00	5.52
Cc	78.50	92.52	5.42
Proposed method	81.24	94.52	5.35
Cc	84.44	94.66	5.43
Proposed method	87.45	96.66	5.49

Table 1: Comparative performance evaluation for the performance parameter using CC and Proposed method.

Comparative result graph for CC and Proposed method using all input value

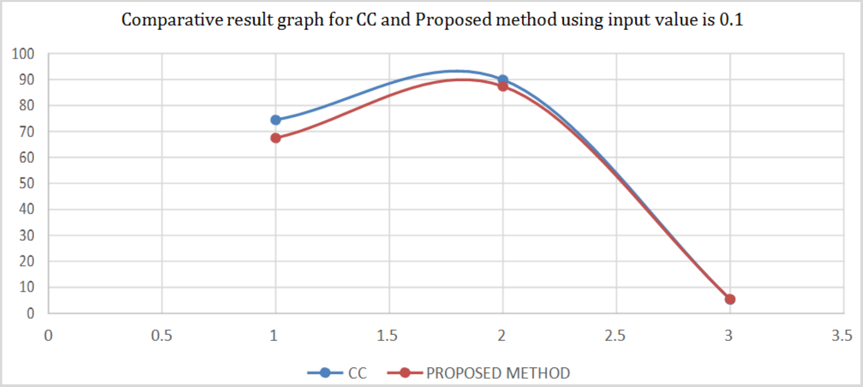

Figure 1: Shows that the comparative performance graph for CC and Proposed Method with the input value is 0.1, here we find t he value of utility CP, Accuracy and Elapsed time and our result in the terms of proposed method is always better than the existing method.

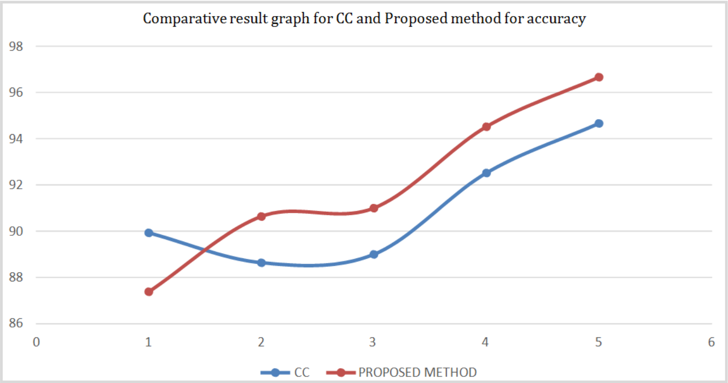

Figure 2: Shows that the comparative performance graph between CC and Proposed Method with the input value is 0.1, 0.2, 0.3, 0.5, and 0.8 for Accuracy and our result in the terms of proposed method is always better than the existing method.

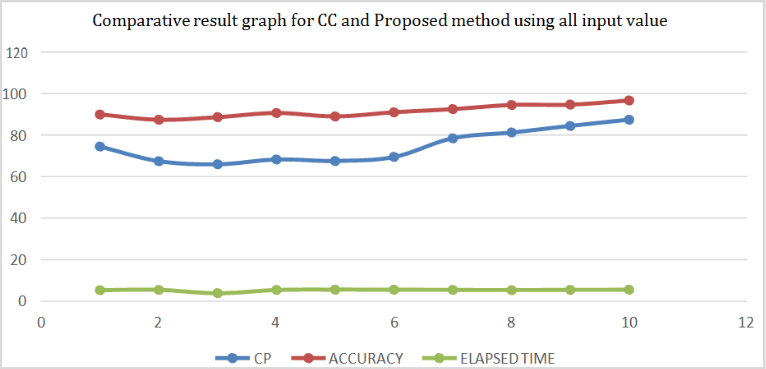

Figure 3: Shows that the comparative performance graph between CC and Proposed Method with the input value is 0.1, 0.2, 0.3, 0.5, and 0.8 for CP, Accuracy and Elapsed Time.

V. CONCLUSION AND FUTURE WORK

In this paper proposed a privacy preservation method using clustering and classification technique. The proposed model is called hybrid model. The hybrid model used two algorithms one is KNN classifier and other is K-MEANS algorithm. The K-MEANS algorithm process for data transformation and classifier are used for the grouping of data. Proposed a vector decomposition based method for privacy preservation. The proposed methods used hybrid clustering and classification process for data transformation and data publishing. The process of vector decomposition used selects the different attribute of given dataset for the process of transformation. The transform data cerates the bucket of index attributes and creates all sub id for the next similar data. After the creation of index build the class for the process of classification. The process of classification used nearest neighbor classifier (KNN). The classifier creates the number of published attribute for the recovery process. The proposed algorithm gives better result in compression of KPPDM technique. The proposed algorithm gives average classification ratio is 98%. Privacy and accuracy is a pair of contradiction; improving one usually incurs a cost in the other. How to apply various optimizations to achieve a trade-off should be deeply researched.

REFERENCES

[1] S Kumara Swamy, Manjula S H, K R Venu gopal, Iyengar S S , L M Patnaik “Association Rule Sharing Model for Privacy Preservation andCollaborative Data Mining Efficiency” IEEE 2014.

[2] Murat Kantarcioglu , Wei Jiang “Incentive Compatible Privacy-Preserving Data Analysis” IEEE 2013 PP 1323-1335. [3] Tamir Tassa “Secure Mining of Association Rules in Horizontally Distributed Databases” 2011 PP 1-18.

[4] lIS. Sasikala, IIS. Nathira Banu “Privacy Preserving Data Mining Using Piecewise Vector Quantization (PVQ)” International Journal of Advanced Research in Computer Science & Technology IJARCST 2014 PP 302-306.

[5] Shipra Agrawal, Jayant R. Haritsa, B. Aditya Prakash “FRAPP: a framework for high-accuracy privacy-preserving mining” Springer 2008.

[6] Julien Freudiger, Shantanu Rane, Alejandro E. Brito ,Ersin Uzun “Privacy Preserving Data Quality Assessment for High-Fidelity Data Sharing”.

[7] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, Wei Ding “Data Mining with Big Data” Department of Computer Science, University of Vermont, USA

[8] Nirali R. Nanavati, Devesh C. Jinwala “Privacy Preserving Approaches for Global Cycle Detections for Cyclic Association Rules in Distributed Databases” SECRYPT 2012 PP 368-371.

[9] Ms.R.Kavitha, Prof.D.Vanathi “A Study Of Privacy Preserving Data Mining Techniques” International Journal of Science and Applied Information Technology 2014 PP 71-77.

[10] Kumaraswamy S Ȧ, Manjula S HȦ, K R Venugopal Ȧ , L M Patnaik “A Data Mining Perspective in Privacy Preserving Data Mining Systems” 2014 PP 704-717.

[11] Somayyeh Seifi Moradi, Mohammad Reza Keyvanpour “Classification and evaluation The Privacy Preserving Distributed Data Mining techniques” Journal of Theoretical and Applied Information Technology 2005 PP 204-211.

[12] Kenampreet Kaur, Meenakshi Bansal “A Review on various techniques of hiding Association rules in Privacy Preservation Data Mining” International Journal Of Engineering And Computer Science 2015, PP. 12947-12951.

[13] S.D. Gordon, J. Katz, “Rational Secret Sharing, Revisited,” Int’l Conf. Security and Cryptography for Networks, 2006 PP 229

[14] X. Lin, C. Clifton, M. Zhu, “Privacy Preserving Clustering with Distributed EM Mixture Modeling,” Knowledge and Information Systems, 2005 PP. 68-81.

[15] M.J.Freedom, K.Nissim, B.Pnkas, “Efficient private matching and set intersection”, Advances in Cryptography: Eurocrypt, 2004, PP 1-19.

DOI 10.17148/IJIREEICE.2017.51016