International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 3 Issue 9 September 2016

# Survey Paper onan Updated & More Efficient Hybrid Algorithm for Privacy Preserving in Data Mining

Pooja Rathore[1], Santosh Varshney[2]
Department of Computer Science and Engineering, Lakshmi Narain College of Technology, Indore, Madhya Pradesh, India[1]
Department of Computer Science and Engineering, Lakshmi Narain College of Technology, Indore, Madhya Pradesh, India[2]
poorath8668@gmail.com[1], varshneysantosh.25@gmail.com[2]

***Abstract:*** *In this paper, we present an overview of modern privacy preserving in data mining algorithms. Privacy preserving in data mining is a heart favorite topic of research for many researchers over the years. Also privacy preserving in data mining takes a lot of data base scans. Therefore it is a computationally expensive task. So still there is a need to update and enhance the existing privacy preserving data mining techniques so that we can get the more efficient methods for the same task. In this paper, we have developed a method to hide sensitive items. This method hides all association rules which contain sensitive items. The proposed method takes less number of data base scans to hide sensitive items as compared to the existing hybrid algorithm for information hiding.*

## 1. INTRODUCTION

The use of data mining [1, 2] is placed in various decisions making task, using the analysis of the different properties and similarity in the different properties can help to make decisions for the different applications. Among them the prediction is one of the most essential applications of the data mining and machine learning. This work is dedicated to investigate about the decision making task using the data mining algorithms. Therefore an application of heart disease is reported for providing the fruitful results from the algorithms.

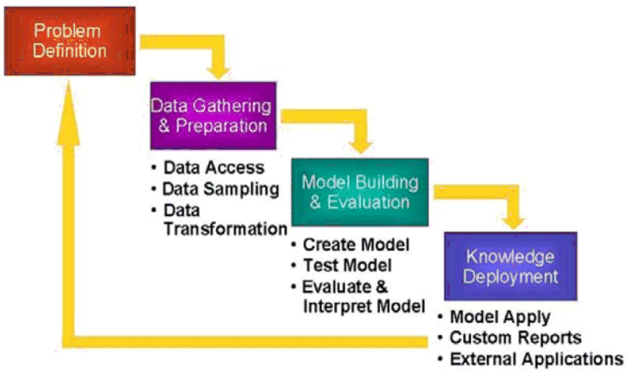


Figure 1: Data Mining

Finally, it enables them to "drill down" into summary information to view detail transactional data. With data mining, a retailer could use point-of-sale records of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop

products and promotions to appeal to specific customer segments.

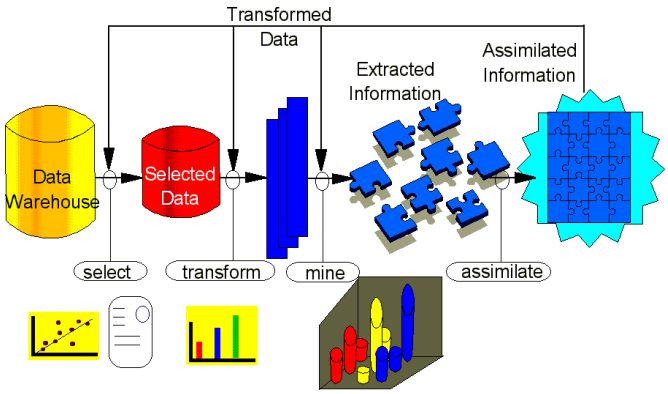


Figure 2: Key steps in data mining

The data mining is a process of analysis of the data and extraction of the essential patterns from the data. These patterns are used with the different applications for making decision making and prediction related task. The decision making and prediction is performed on the basis of the learning of algorithms. The data mining algorithms supports both kinds of learning supervised and unsupervised. In unsupervised learning only the data is used for performing the learning and in supervised technique the data and the class labels both are required to perform the accurate training. In supervised learning the accuracy [5, 6] is maintained by creating the feedbacks form the class labels and enhance the classification performance by reducing the error factors from the learning model.

Let D be the database of transactions [3, 4] and J = {J1,..., Jn} be the set of items. A transaction T includes one or more items in J. An association rule has the form X → Y, where X and Y are non-empty sets of items (i.e. X and Y are subsets of J) such that X ∩ Y = Null. A set of items is called an item set, while X is called the antecedent. The support of an item (or item set) x is the percentage of transactions from D in which that item or item set occurs in the database The confidence or strength c for an association rule X → Y is the ratio of the number of transactions that contain X or Y to the number of transactions that contain X.

The problem of mining association rule is to find all rules that have support and confidence greater then user specified minimum support threshold (MST) and minimum confidence threshold (MCT).

As an example, for a given database in following table, a minimum support of 33% and a minimum confidence of 70%, nine association rules can be found as follows: B=>A(66%, 100%), C=>A (66%, 100%), B=>C (50%, 75%),C=>B (50%, 75%), AB=>C

(50%, 75%), AC=>B (50%, 75%), BC=>A (50%, 100%), C=>AB (50%, 75%), B=>AC (50%, 75%).

| **TID** | **Items** |
|---|---|
| T1 | ABC |
| T2 | ABC |
| T3 | ABC |
| T4 | AB |
| T5 | A |
| T6 | AC |

The objective of privacy preserving data mining is to hide certain sensitive information so that sensitive information cannot be discovered through data mining techniques. Given a transaction database, a minimum support threshold and minimum confidence threshold and set of sensitive items X, the objective is to modify database in such a way that no predictive association rule containing X on the left hand side will be discovered. So if in above example element A is sensitive then rules AB=>C (50%, 75%), AC=>B (50%, 75%) should not be discovered by data mining algorithm.

## 2. PROPOSED METHODOLOGY

**Step 1: Transaction Data Base, Rule Data Base, MCT (Minimum Confidence Threshold) are the inputs.**

**Step 2: scan the rule data base and find the set of frequent items.**

**Step 3: Eliminate all the infrequent items from the Transaction Data Base.**

**Step 4: Enter the sensitive element.**

**Step 5: Find all those rules in the rule data base which contains sensitive element on the LHS & whose confidence is greater than the MCT.**

- **Sort rules in decreasing order of their Confidence and store in R.**

**Step 6: For each rule R from Repeat step 7**.

**Step 7: While the data set is not empty**

- **Find all those transactions where Sensitive item = 1 and RHS = 1**
- **Update the transaction data base: put sensitive item = 0 in all those transactions. In this way, the confidence will become less than the MCT (Minimum Confidence Threshold)**

**Step 8: Exit.**

## 3. CONCLUSION

In this paper, we surveyed the list of existing privacy preserving data mining techniques. We restricted ourselves to the classic privacy preserving data mining problem. It is the protection of all sensitive items from public disclosure with respect to minimal thresholds for support & confidence.

In this paper, we presented a novel algorithm for privacy preserving in data mining via association rule hiding. We have also evaluated the performance of our proposed algorithm. It is taking less number of data base scans to hide association rules containing sensitive items in comparison to previous algorithm.

## REFERENCES

[1] Ila Chandrakar, Manasa, Usha Rani and Renuka. Hybrid Algorithm for Association Rule mining. Journal of Computer Science 6(12), pages 1494-1498, 2010.
[2] Vi-Hung Wu, Chia-Ming Chiang and Arbee L.P. Chen, Senior Member, IEEE Computer Society Hiding Sensitive Association Rules with Limited Side Effects , VOL. 19, NO.1, JANUARY 2007. IEEE Transactions on Knowledge and Data Engineering.
[3] V. S. Verykios, A. K. Elmagarmid, B. Elisa, Y. Saygin and D. Elena. Association rule hiding. In IEEE Transactions on Knowledge and Data Engineering, volume 16(4), pages 434–447, Los Alamitos, CA, USA, April 2004. IEEE Computer Society.
[4] Arun K. Pujari, "Data Mining Techniques", 14th impression, 2008.
[5] Charu C. Aggarwal and Philip S. Yu,"Privacy-Preserving Data Mining: A Survey", IBM, T.J. Watson Research Center.
[6] http://www.thearling.com/text/dmwhite/dmwhite.htm