Tabu Search based Association Rule Hiding

S.Vijayarani
Assistant Professor
School of Computer Science
and Engineering
Bharathiar University,
Coimbatore, India

Dr. A.Tamilarasi
Professor & Head
Department of MCA
Kongu Engg. College
Erode, India

R.SeethaLakshmi
M.Phil Research Scholar
School of Computer Science
and Engineering
Bharathiar University,
Coimbatore, India

ABSTRACT

Data mining algorithms are used for extracting the hidden knowledge from the large databases. Privacy preserving data mining is a new research area in the field of data mining which mainly deals with the side effects of the data mining techniques. The term privacy is denotes the individual's information should be protected. Nowadays, privacy protection has turn out to be an essential issue in data mining research. A primary constraint of privacy-preserving data mining is to prevent the sensitive knowledge extraction by protecting the input data, yet still allow the data miners to pull out the useful knowledge models. Hiding sensitive association rule is an important research problem in privacy preserving data mining. Sensitive association rules are protected by modifying the sensitive items in the original data set. In this research work, tabu search optimization technique is used for modifying the sensitive items for hiding the sensitive association rules.

General Terms

Data mining, Privacy, Security, Tabu search.

Keywords

Privacy, Association Rule, Sensitive item, Modification, Tabu search.

1. INTRODUCTION

Recently, sharing data becomes a frequent business practice. Data may be exchanged between organizations or it may be released publicly. Applying data mining techniques with these shared data, useful patterns, or knowledge can be discovered. With this knowledge, the business operations can be improved. However, data sharing can also causes the privacy issue. Data mining systems can be categorized according to the kinds of knowledge they mine, that is based on data mining functionalities such as characterization, discrimination, association and correlation analysis, classification, prediction, clustering, outlier analysis, and evolution analysis. A comprehensive data mining system usually provides multiple and/or integrated data mining functionalities. Data mining systems can be distinguished based on the granularity or levels of abstraction of the knowledge mined, including generalized knowledge at a high level of abstraction, primitive-level knowledge at a raw data level, or knowledge at multiple levels considering several levels of abstraction.

Privacy preserving data mining is the emerging field that protects sensitive knowledge discovery. It includes the collection of personal data's such as shopping habits, criminal records and credit cards. In conventional data analysis there is a limit to threat privacy. These techniques mainly present the results based on the mathematical characteristics associated with the data. By making use of these techniques does not reveal the interesting patterns which are hidden in data. The data mining techniques are used to explore the hidden patterns but the threat to privacy becomes real since data mining techniques derives highly sensitive from unclassified data which are not even known to database holders. On the other hand analyzing these data opens a new threat to privacy and autonomy of the individuals.

The problem of privacy-preserving data mining has turn into more significant in recent years because of the growing capability to accumulate private data about users, and the ever-increasing sophistication of data mining algorithms to influence this information. A number of techniques such as statistical disclosure control, distributed data privacy, randomization and k-anonymity, etc., have been recommended in recent years in order to execute data mining operations in a privacy preserving way. In addition, the problem has been discussed in database community, the statistical disclosure control community and the cryptography community.

In many cases, the results of data mining applications such as association rule or classification rule mining can compromise the privacy of the data. This has generated a field of privacy in which the results of data mining algorithms such as association rule mining are modified in order to safeguard the privacy of the data. A classic example of such techniques are association rule hiding methods, in which some of the association rules are suppressed in order to preserve privacy.

The rest of the paper is organized as follows. In Section 2 association rule hiding and the related works are discussed. Section 3 gives the general problem formulation and the basic definitions of association rule mining. In Section 4, the proposed tabu search optimization technique for sensitive item modification is given. The effectiveness of the algorithm is evaluated and the experimental results of the proposed technique are discussed in Section 5. Conclusions are given in Section 6.

2. RELATED WORK

The association rule hiding problem can be considered as a variation of the well known database inference control problem in statistical and multilevel databases. The main objective in database inference control problem is providing shield for accessing sensitive information that can be obtained through non sensitive data and inference rules. In association rule hiding, we consider that it is not the data itself but rather the sensitive association rules that create a breach to privacy. The set of association rules which can be extracted from a large data set and some of the association rules are considered to be sensitive. The main job of association rule hiding algorithms are correctly modifying the original data so that the association rule mining algorithms that may be applied to this modified data (i) will be incapable to discover the sensitive rules (ii) will be able to mine all the non sensitive rules that appeared in the original dataset and (iii) will be incapable to discover false rules.[1]

There are three types of association rule hiding algorithms namely heuristic approaches, border-based approaches and exact approaches. Heuristic algorithms are very efficient and fast algorithms that modify the selected transactions from the database for hiding the sensitive knowledge. In border based approach, the sensitive rule hiding can be done through the modification of the original borders in the lattice of the frequent and the infrequent patterns in the data set. Sensitive knowledge is hidden by implementing the revised borders in the modified database. Exact approaches are considered as non-heuristic algorithms which envisage the hiding process as a constraint satisfaction problem that may be solved using integer programming or linear programming [3]. Among these three approaches, the heuristic approaches have been the focus of interest for the majority of researchers in the knowledge hiding field because of their efficiency and scalability.

In this research work, we have used heuristic algorithms for hiding the sensitive rules. In order to develop the hiding algorithms, two forms of modification techniques are used namely distortion and blocking. Distortion is the process of replacing 1’s by 0’s and 0’s by 1’s, while blocking refers to replacing original values by question marks ‘?’.

In the paper “Disclosure limitation of sensitive rules” [6] has projected a technique for hiding the sensitive association rules through the reduction in the support of their generating item sets. The paper “Hiding association rules by using confidence and support” [4] has proposed the algorithm for hiding both the sensitive frequent item sets and sensitive rules. Three single rule heuristic hiding approaches are proposed that are based on the reduction of either the support or the confidence of the sensitive rules, but not both. In all three approaches, the goal is to hide the sensitive rules while minimally affecting the support of the non-sensitive item sets.

In the paper “Association Rule Hiding” [14] extended the work [4] by improving and evaluating the algorithms for their performance under different sizes of input datasets and different sets of sensitive rules. The paper “Privacy preserving frequent item set mining” [10] has introduced multiple rule hiding approaches. The proposed algorithms are efficient and require two scans of the database, regardless of the number of sensitive item sets to hide. “An experimental study of distortion-based techniques for association rule hiding“[5] paper has proposed two distortion-based heuristics to selectively hide the sensitive rules. On the positive side, the proposed schemes use effective data structures for the representation of the rules and effectively prioritize the selection of transactions for sanitization The first algorithm, called *Priority-based Distortion Algorithm* (PDA), reduces the confidence of a rule by reversing 1’s to 0’s in items belonging in its consequent. On the other hand, the second algorithm, called *Weight-based Sorting Distortion Algorithm* (WDA), concentrates on the optimization of the hiding process in an attempt to achieve the least side-effects and the minimum complexity.

The paper “A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms” [7], has proposed a new multi-objective method for hiding sensitive association rules based on the concept of genetic algorithms. The method developed in this paper uses binary transactional dataset as an input and modifies the original dataset for hiding sensitive association rules based on the concept of tabu search optimization algorithm in such a way that all the sensitive rules are to be hidden without any loss of data. The most possible technique for modifying the transaction in the form of distorting the original database. The performance of the algorithm is evaluated by verifying whether the modification process can affect the original set of rules, that can be mined from the original database, either by hiding rules which are not sensitive (*lost rules*), or by introducing rules in the mining of the modified database, which were not supported by the original database (*ghost rules*) and by hiding all the sensitive rules. [1]

3. PROBLEM FORMULATION

3.1 Formulation of Association Rule

Association rule hiding refers to the process of modifying the original database in such a way that certain sensitive association rules disappear without seriously affecting the data and the non-sensitive rules. Association rule mining is defined as: Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of *n* binary attributes called *items*. Let $D = \{t_1, t_2, \ldots, t_m\}$ be a set of transactions called the *database*. Each transaction in *D* has a unique transaction ID and contains a subset of the items in *I*. A *rule* is defined as an implication of the form X→Y where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The sets of items (for short *item sets*) *X* and *Y* are called *antecedent* (left-hand-side or LHS) and *consequent* (right-hand-side or RHS) of the rule respectively.

For example T = {T1, T2, T3, T4, T5}. I= {crème, sugar, coffee, beer, bread, chips, cheese, milk, oranges, apples, eggs}.

Support measure of *X* is denoted as *Support(X)*.

$$Support(X) = (Support\ count(X)/n) * 100$$

The *confidence* of a rule is defined

$$\mathrm{conf}(X \Rightarrow Y) = \mathrm{supp}(X \cup Y)/\mathrm{supp}(X)$$

Table 1. Transactional Database

T id	Items
1	{crème , sugar , coffee , beer}
2	{bread, chips , cheese , milk}
3	{oranges , sugar , crème , beer}
4	{apples , beer , crème , sugar}
5	{eggs , milk , coffee , sugar }

From the table 1 the item set {milk, sugar} has a support of 1 / 5 = 0.2 since it occurs in 60% of all transactions (3 out of 5 transactions).

The rule {milk , sugar}→ coffee has a confidence of 1 / 2 = 0.5 in the database, which means that for 50% of the transactions containing milk and sugar the rule is correct.

3.2 Apriori Algorithm

Apriori algorithm is the most popular algorithm to find all the frequent sets. It makes use of the downward closure property. Apriori algorithm is a bottom-up search, moving upward level-wise in the lattice. Before reading the database at every level it graciously prunes many of the sets which are unlikely to be frequent sets.

The apriori frequent item set discovery algorithm uses the two functions namely candidate generation and pruning at every iteration. It moves upward in the lattice starting from level I till level k, where no candidate set remains after pruning.[8] It has two processes such as Candidate Generation, Pruning.

Table 2. Apriori Algorithm

```
$L_1$: = {frequent 1-itemsets};
k:= 2; // k represents the pass number
While ($L_k$-1)
$C_k$ = New candidates of size k generated from $L_k$-1For all transactions t∈D Increment count of all candidates in $C_k$ that are contained in t
$L_k$ = All candidates in $C_k$ with minimum support
k = k+1
```

The first pass of the algorithm calculates single item frequencies to determine the frequent 1-itemsets. Each subsequent pass *k* discovers frequent itemsets of size *k*. To do this, the frequent itemsets L_{k-1} found in the previous iteration are joined to generate the candidate itemsets C_k. Next, the support for candidates in C_k is calculated through one sweep of the transaction list. From L_k-1, the set of all frequent (k-1) itemsets, the set of candidate k-itemsets is created.

Consider a given transactional database D, minimum support threshold value SUP_{min}, minimum confidence threshold value $CONF_{min}$, a set of association rules AR can be mined from D and a set of sensitive association rules AR_{sen} mined from D and a set of sensitive rules $AR_{sen} \subseteq AR$ to be hidden, generate a new database D', such that the rules in $AR_{non-sen}$=AR-AR_{sen} can mined from D' under the same SUP_{min} and $CONF_{min}$ C. No normal rules in $AR_{non-sen}$ are falsely hidden (lost rules) and no extra fake rules are (ghost rules) are mistakenly will mined after the rule hiding process.

4. PROPOSED SOLUTION

The following steps are required for the proposed solution.

Step 1: Consider a transactional database with set of items and transactions

Step 2: Apriori algorithm is used to find the frequent item sets based on the minimum support threshold.

Step 3: From the frequent item sets, the set of association rules can be generated based on the minimum support and confidence thresholds.

Step 4: Select the sensitive rules from the set of association rules.

Step 5: Tabu Search Optimization algorithm is used for modifying the items based on the cost function

Step 6: Repeat the steps 2 and 3 for the modified data set

Step 7: Verify (i) all the sensitive rules are hidden, (ii) no non-sensitive rules are hidden (iii) no false rules

Table 3. Proposed Solution

```
Input : Set of association rules, min_supp, min_conf.
Output : modified data set

Step 1: Finding the sensitive association rules from the set of
        association rules based on the threshold values
        minimum support and minimum confidence
Step 2: Tabu search optimization technique is used for
        modifying the sensitive items
Step 3: Apriori algorithm is applied to the modified data set
Step 4: Performance Analysis
           Verify (i) all the sensitive rules are hidden
                  (ii) no false rules generation
                  (iii) non-sensitive rule protection
Step 5: Stop.
```

4.1 Optimization Techniques

An optimization technique refers to choosing the best element from some set of available alternatives. Optimization techniques are used for solving problems in which one seeks to minimize or maximize a real function by systematically choosing the values of real or integer variables from within an allowed set. It is finding the "best available" values of some objective function given a defined domain, including a variety of different types of objective functions and different types of domains. Many types of optimization techniques and algorithms are used to get the optimal solutions. In this work, tabu search optimization algorithm is used for modifying the data items.

4.2 Tabu Search

Tabu search is a mathematical optimization method, belonging to the class of local search. Tabu search enhances the performance of a local search method by using memory structures: once a potential solution has been determined, it is marked as “taboo" ("taboo" being a different spelling of the same word) so that the algorithm does not visit that possibility repeatedly.

Tabu search is a metaheuristic algorithm that can be used for solving combinatorial optimization problems, such as the traveling salesman problem (TSP).[9] Tabu search uses a local or neighborhood search procedure to iteratively move from a solution x to a solution x' in the neighborhood of x, until some stopping criterion has been satisfied. To explore regions of the search space that would be left unexplored by the local search procedure, tabu search modifies the neighborhood structure of each solution as the search progresses. The solutions admitted to N * (x), the new neighborhood, are determined through the use of memory structures. The search then progresses by iteratively moving from a solution x to a solution x' in N * (x). In this research work, the general tabu search algorithm is used for modifying the sensitive data items.

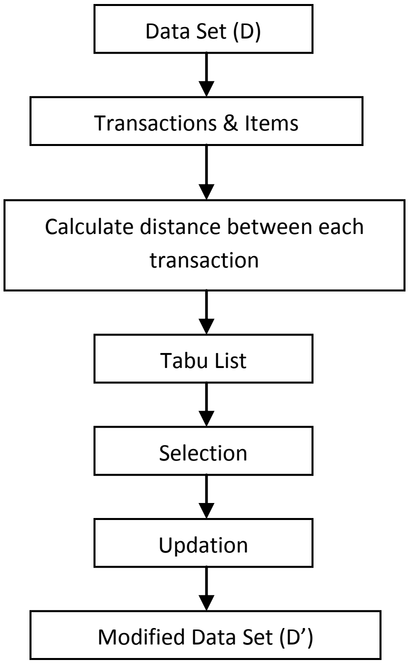


Fig 1: Tabu Search

In the first step, all the sensitive items, sensitive transactions and number of modifications required for the sensitive items are initialized. In the second step, the cost i.e. distance of each sensitive transaction is calculated. Based on this cost value the transactions are selected for sensitive item modification. If the particular sensitive item is found in the transaction then modify it as 1 to 0. These modified values are moved into tabu list. This process is repeated until the number of modifications becomes 0. In step 3 is a termination process, before terminating the algorithm ensures that all the sensitive items are modified and the number of modification is 0 then only the algorithm is terminated.

After performing the Tabu search optimization algorithm, the apriori algorithm is applied to the modified database for finding the frequent item sets and generates the sensitive rules from the database. In the modified data set, we ensure all the sensitive rules are hidden, no false rules are generated and non sensitive rules are not affected. The tabu search optimization algorithm for sensitive item modification is given in the table 4.

Table 4. Tabu Search Optimization for Sensitive Item Modification

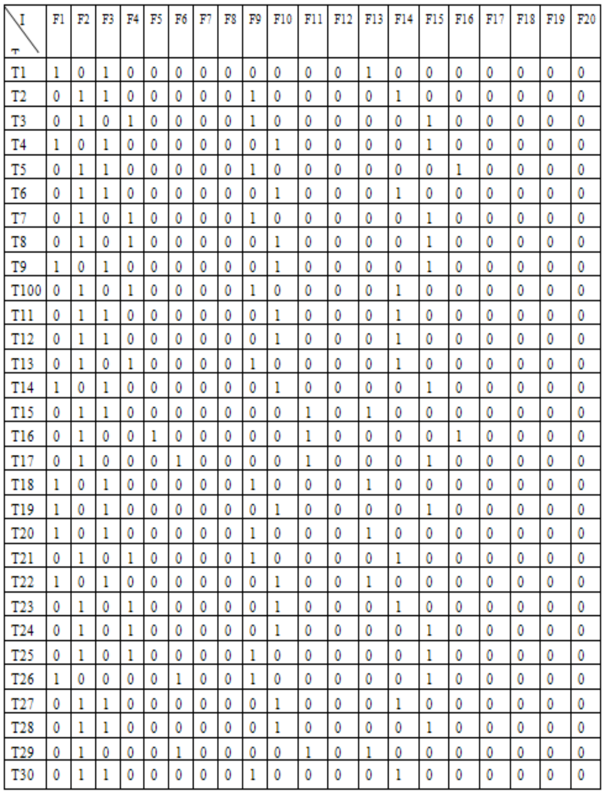
```
Procedure TS_MetaHeuristic

Notation
        S,  Current transaction
        S*  Total number of items
         f*,   Value of f(S*)
        N(P),  Number of items presented
        N(Q), Number of items not presented

1. Choose (construct) an initial solution Set S= $ST_{i,}$
   f*=f($ST_{i)\,,}$ S*= N(P) +N(Q), T=S;
    1.1 Initializing the sensitive items $S_i$, where $S_i \in I$,
          I={$i_1,i_2,\ldots i_n$})
   1.2 Initializing the sensitive transactions contains
            sensitive items $ST_i$ where $ST_i \in T$, T = {$t_1,t_2,\ldots t_m$})
2. {Search}
     While termination criterion not satisfied
      do
            $f(ST_{i)\,=}\,(\,N(P_i)\,_{-}\,N(Q_i)\,)^2$
            If  f ($ST_i$ ) < S*( if any ($S_i$) found in f($ST_i$) then
               modify $S_i$ as0.
            else
               Select next transaction
                 Record tabu for the modified current value
                 move in T.
               (delete oldest entry if necessary),
            Repeat the steps 1.2 until all the modification
            becomes 0
      End while
3. {Terminating}
      Ensure all the sensitive items are modified
      Number of modification becomes 0 then the process
      Completed

4. Exit
```

5. EXPERIMENTAL EVALUATION

5.1 Dataset

Dataset is collected from the website www.fimi.cs.helsinki.fi. Various types of datasets are available in this website such as Mushroom, Chess etc. In this work, mushroom dataset is used; it contains 119 items and 8124 transactions. From this mushroom dataset 50 items and 1500 transactions are chosen for this work. The table shows the sample dataset of 30 transactions and 20 items**.**

Table 5. Original Transactional Data Base



I T	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
T1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
T2	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T3	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T4	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T5	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
T6	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T7	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T8	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T9	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T100	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T11	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T12	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T13	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T14	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T15	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
T16	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
T17	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
T18	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
T19	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T20	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
T21	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T22	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
T23	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T24	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T25	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T26	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T27	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T28	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T29	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0
T30	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0

5.2 Phase I

Association rule algorithms show co-occurrence of variables. It occurs frequently together in a given dataset. It tends to generate many more rules than in decision trees. It gives the ability to find rare and less clear patterns in the data. It attempts to use the rules generated by association rule algorithms. Based on the support value 20% and confidence 60% the frequent items are generated and sensitive rules are retrieved based on the minimum support and minimum confidence. The table shows the sensitive rules and the items found in these rules are considered as sensitive items. In this example, the items F1, F3, F4, F10 and F13 are sensitive items.

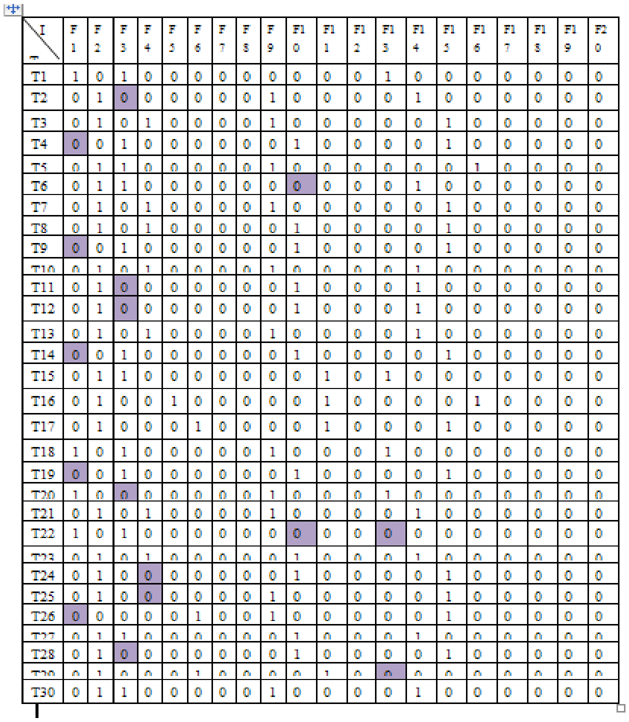
Table 6. Sensitive Rules

Sensitive Rules	Confidence
F1 → F3	88.89%
F4 → F13	100%
F1 → F13	72.73%
F13 → F3	90%
F3,F10 → F1	72%
F4 → F3	100%
F4 → F13	66.67%
F1,F10 → F13	63%
F10 → F13	75%
F4 → F10, F13	76%

5.3 Phase II

Tabu search excludes solutions in the tabu list from $N^*(x)$. A variation of a tabu list prohibits solutions that have certain attributes (e.g., solutions to the traveling salesman problem (TSP) which include undesirable arcs) or prevent certain moves (e.g. an arc that was added to a TSP [11] tour cannot be removed in the next n moves). Selected attributes in solutions recently visited are labeled "tabu-active." Solutions that contain tabu-active elements are "tabu". This type of short-term memory is also called "recency-based" memory. Table 6 shows the sensitive rules and the items found in these rules are considered as sensitive items. In this example, the items F1, F3, F4, F10 and F13 are sensitive items.

Table 7. Modified Data Base



I T	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
T1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
T2	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T3	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T4	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T5	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
T6	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
T7	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T8	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T9	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T10	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T11	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T12	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T13	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T14	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T15	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
T16	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
T17	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
T18	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
T19	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T20	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
T21	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
T22	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T23	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T24	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T25	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T26	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
T27	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
T28	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
T29	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
T30	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0

The above table represents the modified database based on the Tabu search optimization algorithm. In this algorithm transactions are selected based on the distance value. This distance values are based on number of items are presented and does not present in each transaction. Frequent items are taken from the association rule mining and applied the algorithm in the original database. This result is same as the genetic algorithm but modification of the items in a particular transaction is varied with the genetic algorithm. It makes faster during the modification process.

5.4 Analysis of Results

In this section, the result of tabu search optimization algorithm is analysed. The experimental results are analysed based on the following performance factors.

1. Sensitive Rule Protecction
2. Non-Sensitive Rule Protection
3. Fake Rule generation
4. Efficiency

5.4.1 Sensitive Rule Protection

We have to find out whether all the sensitive rules are protected or not. Three different sizes of data sets i.e 20 items 30 transactions, 30 items 500 transactions and 50 items 1500 transactions are used for this analysis. The accuracy of sensitive rule protection of the different data sets are given in the chart.

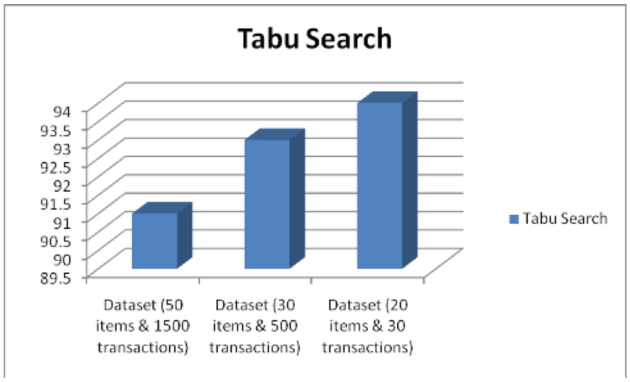


Fig 2: Sensitive Rule Protection

5.4.2 *Non Sensitive Rule Protection*

This performance factor is nothing but any side effect has occurred in the form of protecting non-sensitive rules during the hiding process. The accuracy of non-sensitive rule protection is depicted in the chart.

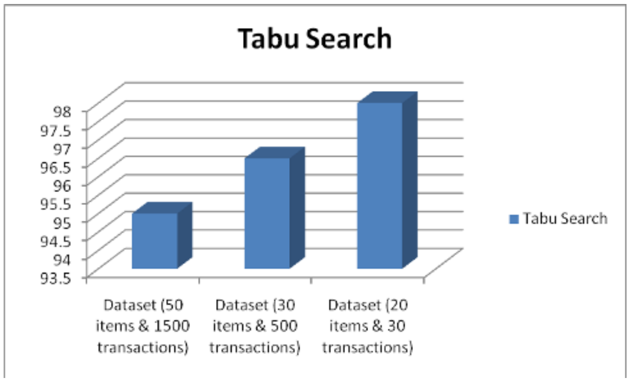


Fig 3: Non-Sensitive Rule Protection

5.4.3 *Fake Rules Generation*

During the hiding process we have to verify any false or fake rules can be generated. In this technique, no false rules are generated. The accuracy of fake rule generation is shown in the figure.

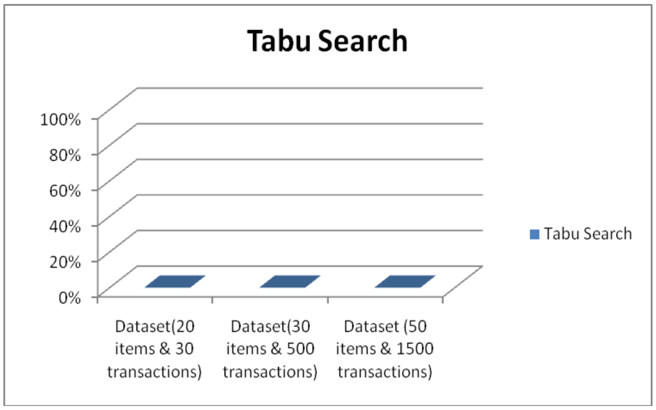


Fig 4: Fake Rules Generation

5.4.4 *Efficiency*

Efficiency of the tabu search algorithm measured in terms of number of iterations required during modication process. Different data sets and different thresholds values are applied to find out the number of iterations required.

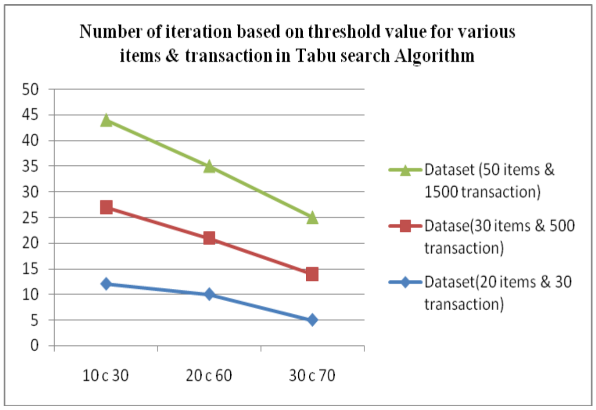


Fig 5: Efficiency of Tabu Search

6. CONCLUSIONS

Sensitive association rule hiding is a vital research problem in privacy preserving data mining. In this research work, we have used tabu search optimization technique for protecting sensitive association rules. Compared with the previous approaches this approach has modified the sensitive rules accurately without affecting the non-sensitive rules and no false rules are generated. In this work, it needs several iterations for selecting the optimal transaction for modification. In future, our goal is to develop new fitness functions and applying other optimization techniques to minimize the iterations**.**

7. REFERENCES

[1] Agarwal CC. and Yu PS., "Privacy-preserving data mining: Modeland Algorithms, (editors)CharuC.Aggarwal and Philip S. Yu, ISBN: 0-387-70991-8, 2008.

[2] Y. Saygin, V. S. Verykios, and A. K. Elmagarmid. "Privacy preserving association rule mining", In Proceedings of the 2002 International.

[3] Assaf Schuster, Ran Wolff, Bobi Gilburd " Privacy Preserving data mining on data Grids in the presence of Malicious Participants" IEEE International Symposium on High Performance Distributed Computing - HPDC 2004.

[4] Nan Zhang, Shengquan Wang, and Wei Zhao "A New Scheme on Privacy Preserving Association Rule Mining", Principles of Data Mining and Knowledge Discovery – PKDD, Volume 3202, Pg: 484-495, 2004.2004.

[5] Matthew Eric Otey, Chao Wang, Srinivasan Parthasarathy, Adriano Veloso, Wagner Meria , "Mining Frequent itemsets in Distributed and Dynamic Database"s, IEEE international conference on Data Mining,2003.

[6] Yucel Saygin, Vassilios S.Verkios, Ahmed K. Elmagarmid, "Privacy Preserving Association Rule Mining", Conference of Research Issues in Data Engineering - RIDE 2002.

[7] GUO Yu-hong, TONG Yun-Hai,TANG Shi-Wei, YANG Dong-Qing “Knowledge hiding in Database, Journal of Software, Vol.18, no 11,PP.2782-2799.Nov 2007.

[8] Oliveira S. R. M., Zaiane O., Saygin Y., “Secure Association-Rule Sharing.Advances in Knowledge Discovery and Data Mining,”Lecture Notes in Computer Science, Vol.3056, Pages.74-85, 2004.

[9] Glover, F. and M. Laguna (1993), “Tabu Search”, in C.R Reeves (ed.), in Modern Heuristic Techniques for Combinatorial Problems, C.R. Reeves (ed.), Blackwell, pp. 70-150.

[10] Glover, F., É. Taillard and D. de Werra (1993), “ A User's Guide to Tabu Search”, Annals of Operations Research **41**, 3-28.

[11] Glover, F., M. Laguna, É. Taillard and D. de Werra (eds.) (1993), “Tabu Search”, Annals of Operations Research **41**, J.C. Baltzer Science Publishers, Basel, Switzerland.