ISSN: 2348 9510

International Journal Of Core Engineering & Management (IJCEM)
Volume 2, Issue 9, December 2015

PRIVACY PRESERVING USING ASSOCIATION RULE MINING

Satish Choudhary
Institute of Engineering & Science IPS Academy, Indore (M.P) India
Satishchoudhary1989@gmail.com
9977725635

Arvind Upadhyay
Assistant Professor, Institute of Engineering & Science IPS Academy, Indore (M.P) India
Upadhyayarvind10@gmail.com
9827566736

ABSTRACT

The privacy preserving data mining issue is an important one made by wide available personal data. Some exist problems in new and rapidly emerging research field of privacy preserving data mining introduced in this thesis work. Some new work is analyzed and makes privacy reserved of data. In association rule mining and privacy protection data release, data distortion concept is important once were focused on discussion. Privacy preserving algorithm , which include algorithm performance, privacy protection degree, and data mining difficulty were illustrated . Finally, for further directions the development of privacy preserving data mining is prospected.

Nowadays, sometimes yet dangerous but useful technology is data mining through which sensitive information and the relationships among the items in a database are identified .for their progress the users and originations are need to share their data and they should manage the data for preserving the privacy of sensitive data. For managing information sharing, the concept of preserving the privacy of information in data mining was introduced. This paper presents a hybrid algorithm with distortion technique with both

support-based and confidence-based approaches for privacy preserving. The proposed algorithm tries to hide sensitive rules from the perspective of the database owner and maintain useful association rules.

Keywords: Classification, SLIQ, ID3, KNN, performance

I. INTRODUCTION

This thesis discusses privacy and security issues that are likely to affect data mining projects. It introduces solutions to problems, without violating privacy how to obtain data mining results, whereas a level of data access that violates privacy and security constraints standard data mining approaches would require. This thesis gives an overview of privacy preserving data mining technique and it also proposes a secure data publishing framework background about these goals is introduced in the coming chapters.

Computers is deluge of information have promised us a fountain of wisdom . This huge amount of data makes it crucial to develop tools to discover what is called interesting sensitive information. These tools are called data mining tools. So, data mining promises to discover what is hidden, but if this knowledge were exposed to the public or to adversaries because that the hidden knowledge is sensitive and owners would not be happy? This problem motivates research to develop algorithms, to assure data owners that privacy is protected by techniques and protocols while satisfying their need to share data for a common good.

Integration of several fields is represented by data mining, machine learning and database systems is included in it, data visualization, information theory data statistics. Data mining can be defined as a non-trivial process of identifying the valid and novel data item which is potentially useful and understandable ultimately .

The process which covers many interrelated steps in discovering the information is verycomplex. Key steps in the knowledge discovery process are:

1. Data cleaning: noise removing and elimination of inconsistent data during process.
2. Data integration: it is the process of combining the multiple data sources.
3. Data selection: the part of the data that are relevant for the problem is selected by data selection.
4. data transformation: transform the data into a suitable format.
5. Data mining: apply data mining algorithms and techniques.
6. Pattern evaluation: pattern that is found is fulfill the requirement is evaluated.
7. Knowledge presentation: mined knowledge is present to the user.

Data mining is the process of extracting hidden patterns from data. Amount of data doubling every three years as more data is gathered, an important tool to transform this data into knowledge is data mining. Wide range of applications used data mining for the marketing and fraud detection and the purpose of scientific discovery.

Data mining can be applied to data sets of any size, and to uncover hidden patterns it can be used, it cannot uncover patterns which are not already present in the data set.

For an effective analysis and decision means the useful knowledge is extracted by data mining.

An automated extraction of novel is knowledge discovery in databases (kdd) , in large databases understandable and potentially useful patterns implicitly is stored .

Data mining is an essential step in the process of knowledge discovery in databases, in which in order to extract patterns , intelligent methods are applied. Other steps in knowledge discovery process include pre-mining tasks such as data cleaning (inconsistency and noise removing of data) and data integration (to merge the data in single location from multiple source), as well as post mining tasks such as pattern evaluation (representing knowledge of interesting pattern that is evaluated) and knowledge presentation (the extracted rules are presented using visualization and knowledge representation techniques).

In data mining, the popular and well researched method is **association rule learning** for discovering interesting relations between variables in large databases. Analyzing and presenting strong rules discovered in databases is describe by piatetsky-shapiro using different measures of interestingness. Based on the concept of strong rules, agrawal et al introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (pos) systems in supermarkets.
Let us take an example , here the rule

$$\{onions, potatoes\} \Rightarrow \{beef\}$$

Of a supermarket data is found would indicate that if a customer buys onions and potatoes together, they may be also buy beef. Such information can be used as the basis for decisions about marketing activities such as, pricing of product or placement of product. In addition to the above example from market basket analysis many application areas in which association rules are employed today including web usage mining, intrusion detection and bioinformatics.

II. LITERATURE SURVEY

The amount of data kept in computer files is growing at a phenomenal rate. To discover unknown information the data mining field is used. Data mining is often defined as the process of discovering meaningful, and interesting patterns and trends through implicit extraction of non trivial set of data, and extraction of unknown information from repositories of large amount of data, using pattern recognition as well as statistical and mathematical techniques. An sql query is usually stated OR written to retrieve specific data, while data miner are not sure what they are exactly require.

Background

Association rule mining

Association rule mining finds interesting associations in large data base and/or correlation relationships among large sets of data items [1]. The data item that occure frequently together in a given dataset is shown by association rules . A typical and widely-used example of association rule mining is market basket analysis [2].

For example, data are collected using bar-code scanners in supermarkets. Large number of transaction records consists in such market basket data base. By a customer on a single purchase transaction , each record lists all items bought. The groups of items which are consistently purchased together managers would be interested to know if certain. They could use this data for adjusting store layouts (with respect to each other placing items optimally), and item cross-selling , for promotions of items, for catalog design and buying pattern is used to identify customer segments.

the information of this type in the form of "if-then" statements is provided by association rule. Unlike the if-then rules of logic, association rules are probabilistic in nature. These rules are computed from the data

In addition "if" part is called antecedent and the consequent is called the "then" part , and the degree of uncertainty about the rule is expressed by these two rules. During the rule analysis the sets of items (called item sets) are antecedent and consequent that are disjoint (do not have any items in common).

The first number is called the support for the rule. The percentage of the total number of records in the database is expressed as a support or it is the item in antecedent and consequent part of rult is support.

confidence of the rule is known by other number. Confidence is the ratio of the number of transactions that include all items in the consequent ("then" part) as well as the antecedent ("if" part or namely, the support) to the number of transactions that include all items in the antecedent.

A survey of privacy-preserving association rule mining

Ppdm can be attempted at three levels as shown in figure 2.4. The first level is raw data or databases where transactions reside. Techniques and algorithms that ensure privacy in data mining is at second level. The third level is the output of different data mining algorithms and techniques.

Level 1 **Raw data**
Level 2 **Data mining algorithms**
Level 3 **Rules**

Figure 2.4: privacy-preserving data mining attempted at three major levels.

At level 1, researchers have applied different techniques to raw data or databases for the sake of protecting the privacy of individuals (by preventing data miners from getting sensitive data or sensitive knowledge), or protecting privacy of two or more parties who want to perform some analysis on the combination of their data without disclosing their data to each other.

At level 2, privacy-preserving techniques are embedded in the data mining algorithms or techniques and may allow skillful users to enter specific constraints before or during the mining process.

Finally, at level 3, researches have applied different techniques to the output of data mining algorithms or techniques for the same purpose of level 1.

In the literature of privacy-preserving association rule-mining, researchers presented different privacy-preserving data mining problems based on the classifications of the authors. These classifications are good but we believe that from the point of view of the targeted people (individuals and parties who want to protect their sensitive data), it is difficult to understand. We believe these people are interested in the answer to the following questions: can this privacy-preserving algorithm or technique protect our sensitive data at level 1, level 2 or at level 3? The second important question is: which dimension does this algorithm or technique fall under? Is it the individuals or the ppdmsmc? In the following we review the work that has been done under each level.

level 1 (raw data or databases)

The individual's dimension

In 1996, clifton et al. [10] the privacy of individuals at level 1 is presented numbers of ideas. These include the following:

- Limiting access: the access to data can control so users can have access only to a sample of the data. We can lower the confidence of any mining that is attempted on the data. In other words, the control of access to data stops users from obtaining large chunks and varied samples of the database.
- Fuzz the data: altering the data by forcing, for example, aggregation to the daily transactions instead of individual transactions prevents useful mining and at the same time allows the desired use of data. This approach is used by the u.s. census bureau.
- Eliminate unnecessary data: unnecessary data that could lead to private information. For example, the first three digits in a social security number can indicate the issuing office, and therefore reveal the location of that number holder. Another example, if an organization assigns phone numbers to employees based on their location; one can mine the company phone book to find employees who for instance work on the same project. A solution to this problem is to give unique identifiers randomly to such records to avoid meaningful groupings based on these identifiers.

- Augment the data: which means adding values to the data without altering its usefulness. This added data is usually misleading and serves the purpose of securing the privacy of the owner of this data. For example, adding fictitious people to the phone book will not affect the retrieval of information about individuals but will affect queries that for instance try to identify all individuals who work in a certain company.
- Audit: to publish data mining results inside an organization rather than the world, we can use auditing to detect misuse so that administrative or criminal disciplinary action may be initiated.

Despite the potential success of the previous solutions to restrict access to or perturb data, the challenge is to block the inference channels. Data blocking has been used for association rule confusion. This approach of blocking data is implemented by replacing sensitive data with a question mark instead of replacing data with false incorrect values. This is usually desirable for medical applications. An approach that applies blocking to the association rule confusion has been presented in [11].

In 2001, saygin et al. [8] proposed an approach by replacing selected values or attributes with unknowns for hiding rules, instead of replacing them with false values. In the following we discuss the approach:

Using unknowns to prevent discovery of association rules

This technique depends on the assumption that in order to hide a rule a → b either the support of the item set asb should be decreased below the minimum support threshold (mst) or the confidence of the rule should be decreased below the minimum confidence threshold (mct). Based on the above, we might have the following cases for item set a which is contained in a sensitive association rule:

- A remains sensitive when minsup (a) $\geq$ mst.
- A is not sensitive when maxsup (a) $<$ mst.
- A is sensitive with a degree of uncertainty when minsup (a) $\leq$ mst $\leq$ maxsup (a).

According to [8] the only way to decrease the support of a rule a → b is to replace 1s by ?s for the items in asb in the database. In this process, the minimum support value will be changed while the maximum support value will stay the same. Also, the confidence of a rule a → b can be decreased by replacing both 1s and 0s by?s.

In the same year (2001), dasseni et al. [6] proposed another approach that is based on perturbing support and/or confidence to hide association rules.

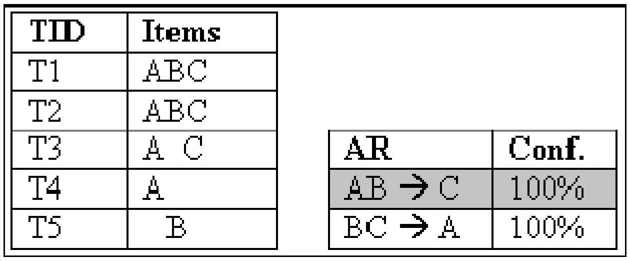
Hiding association rules by using confidence and support

The work in [6] proposed a method to hide a rule by decreasing either its support or its confidence. This is done by decreasing the support or the confidence one unit at a time by modifying the values of one transaction at a time. Since conf (a → b) = supp (ab) / supp (a), there are two strategies for decreasing the confidence of a rule:

- Increasing the support of a in transactions not supporting b.
- Decreasing the support of b in transactions supporting both a and b.

Also to decrease the support for a rule a → b, we can decrease the support for the item set (ab). An example mentioned by the people who proposed the method in [6] clarifies the method as follows; let us suppose that s = 20% and c = 80%.

Let us suppose that we have the database in the table shown in figure 2.5. With the values for s and c above, we can deduce that we have two rules ab → c and bc → a with 100% confidence.

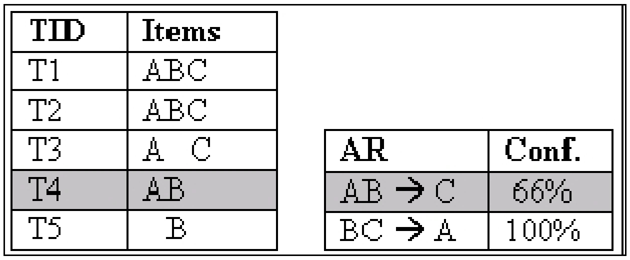


TID	Items
T1	ABC
T2	ABC
T3	A C
T4	A
T5	B

AR	Conf.
AB → C	100%
BC → A	100%

Figure 2.5: ab → c and bc → a with 100% confidence.

Let us suppose that we want to hide the rule ab → c by increasing the support of ab. Let us do that by turning to 1 the item b in transaction t4 so the database becomes as shown in figure 2.6.

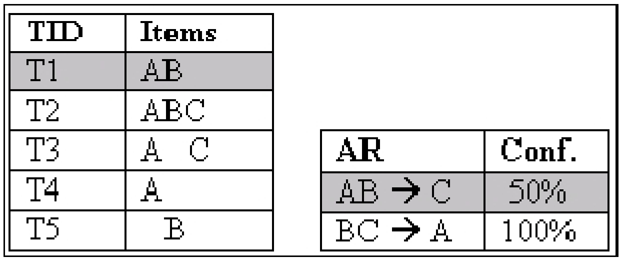


TID	Items
T1	ABC
T2	ABC
T3	A C
T4	AB
T5	B

AR	Conf.
AB → C	66%
BC → A	100%

Figure 2.6: hide the rule ab → c by increasing support of ab.

Notice that the confidence for the rule ab → c was decreased to 66%. Having in mind that c = 80%, we were successful in hiding the rule ab → c. We can



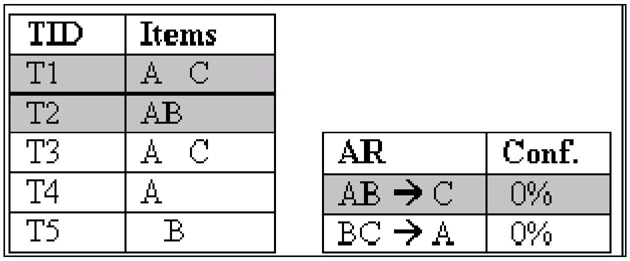
TID	Items
T1	AB
T2	ABC
T3	A C
T4	A
T5	B

AR	Conf.
AB → C	50%
BC → A	100%

Figure 2.7: hide the rule ab → c by decreasing the support of c.

Also hide the rule ab → c by decreasing the support of c by turning to 0 the item c in t1 as figure 2.7 shows.

Notice that the confidence for the rule was decreased to 50% which means that we were successful in hiding the rule ab → c. Finally we can also hide the rule ab → c by decreasing the support of abc by turning to 0 the item b in t1 and turning to 0 the item c in t2 as figure 2.8 shows.



TID	Items
T1	A C
T2	AB
T3	A C
T4	A
T5	B

AR	Conf.
AB → C	0%
BC → A	0%

Figure 2.8: hide the rule ab → c by decreasing the support of abc.

Notice that the confidence for the rule ab → c was decreased to 0%, so again this time in hiding the rule we were successful.

The work in [21] proposed a hybrid method to hide a rule by decreasing its support or decreasing its confidence. This method uses features of both isl & dsr algorithms. This is done by decreasing the support or the confidence n units at a time by modifying the values of transactions.

III. PROBELEM FORMULATION & METHODOLOGY

Formulation of problem sensitive rule hiding is described as follows:

To extract hidden and interesting rules or patterns from databases is the main objective of data mining. However, to hide certain sensitive information is main problem in data mining so that they cannot be discovered through data mining techniques.

In association rule mining , we assume that only sensitive itemset in transaction database are given and problem is to hide sensitive items so that it cannot be inferred through association rules mining algorithms. More specifically, given a transaction database , a minimum support, a minimum confidence and a set of items to be hidden, the objective is to modify the database such that no association rules containing on the right hand side or left hand side will be discovered.

The problem is to hide the sensitive and useful data from others by using the transaction database, min_sup_thr.(mst), min_conf_thr.(mct) , a set of strong rules(above the threshould value) is given with the set of sensitive items, how can we modify the database such that using the same mst and mct, in the modified database satisfies all the constraints with in the set of strong rules: 1) no sensitive rule, 2) no lost rule, and 3) no false rule?

Let transactions database is d and the set of items are as j = {j1, ..., jn}. One or more items in j is included in transaction t . An association rule has the form x → y , where x and y are non-empty sets of items (i.e. Subsets of j is x and y) such that they are disjoint set (there is no common item between them) x ∩ y = null. A set of items is called an itemset, while x is called the antecedent as we discussed in previous chapter.

The problem of mining association rule is to find associated rule that have greater support and greater confidence then user specified minimum support threshold (mst) and minimum confidence threshold (mct).

IV. PROPOSED ALGORITHM

The task of mining association rules over market basket data [1] is considered a core knowledge discovery activity. An useful mechanism is provide by association rule mining for item discovering correlations that belonging to customer transactions in a market basket database. Let d be the database of transactions 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 itemset, while x is called the antecedent. The support of an item (or itemset) x is the percentage of transactions from d in which that item or itemset 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.

This is knowledge discovery with balance the privacy is an effort . It seems that conflict with hiding sensitive data for discovery of itemsets. Sanitizing algorithms that work at level 1 take (as input) a database d and modify it to produce (as output) a database d′ where mining for sensitive itemsets are not shown by the rule . The alternative scenario at level 3 publish the rest after removing the sensitive item sets from the set of frequent itemsets. The database d does not need to be published is implies by this scenario. The problem (in both scenarios) is that sensitive knowledge can be inferred from non-sensitive knowledge through direct or indirect inference channels. This chapter focuses on the problem in the first scenario where a database d′ is to be published.

Proposed Algorithm

- **Step 1: transaction data base, rule data base, mct (minimum confidence threshold) are the inputs.**
- **Step 2: enter the sensitive element**
- **Step 3: find all those rules in the rule data base which contains sensitive element on the rhs & whose confidence is greater than the mct.**
- **Step 4: for each rule which contains a sensitive item on rhs repeat step 4**
- **Step 5: while the data set is not empty**
- **Find all those transactions where sensitive item = 1 and lhs = 1**
- **Then put sensitive item = 0 in all those transactions. In this way, the confidence will become less than the mct (minimum confidence threshold)**
- **Step 6: exit**

V. RESULT ANALYSIS

. A data set

Suppose there is a database of transactions as below:

Tid	items
t1	abd
t2	b
t3	acd
t4	ab
t5	abd

Fig 4.1: a data set

One has also given a mst of 60% and a mct of 70%. One can see four association rules can be found as below

A → b (60%, 75%)
B → a (60%, 75%)
A → d (60%, 75%)
D → a (60%, 100%)

Now there is a need to hide d and b.

Previous methods:

One can see that by simple **isl** algorithm if someone want to hide d and b, then the transaction t2 can be check after modification from b to bd (i.e. From 0100 to 0101).but still isl cannot hide the rule d → a. It can be see with following example

Tid	items	bit map
t1	abd	1101
t2	b	0100
t3	acd	1011
t4	ab	1100
t5	abd	1101

(rule d → a hide by isl approach)

T-id	items	bit -map
t1	abd	1101
t2	b	***0101***
t3	acd	1011
t4	ab	1100
t5	abd	1101

So by above explanation it is clear that rule d → a can not be hidden by isl approach because by modifying t2 from b to bd (i.e. From 0100 to 0101) rule d → a will have support and confidence 60% and 75% respectively.

By dsr approach:

T-id	items	bit -map
t1	abd	1101
t2	b	0100
t3	acd	1011
t4	ab	1100
t5	abd	1101

(rule d → a hide by dsl approach)

T-id	items	bit- map
t1	abd	***0101***
t2	b	0100
t3	acd	1011
t4	ab	1100
t5	abd	1101

Now the rule d → a is hidden by dsr technique as its support 40% its confidence is now 60% , but as a result the rule a → d is also hidden as a side effect.

Result analysis of proposed algorithm 2:

A data set

Suppose there is a database of transactions as below:

Table 1

Tid	items
t1	abc
t2	abc
t3	abc
t4	ab
t5	a
t6	ac

Fig 4.2: a data set

Suppose mct is 50%.

Table 2

Tid	abc
t1	111
t2	111
t3	111
t4	110
t5	100
t6	101

The all possible rules with confidences are:
a–>b(66.66%) ,
a–>c (66.66%),
b–>a(100%),
b–>c (75%),
c–>a(100%),
c–>b (75%).

By hybrid approach and proposed algorithm 2:

Suppose that the item a need to be hide , for this, first take rules in which a is in rhs. The rule are c–>a and b–>a , the confidence is greater in both the rule. Search in transaction data base by taking the rule b–>a first . And select the transaction which supports both b and a i.e., b = a = 1.the t1, t2, t3, t4 are four transactions with a = b = 1. Now on the place of item a put 0 in all the four transactions in transaction table. After modifying the transaction , here givin the table 3 as the resultant modified table.

Table 3

Tid	abc
t1	011
t2	011
t3	011
t4	010
t5	100
t6	101

Now 0% is the confidence of rule b–>a after calculation, it is which is less than minimum confidence so now this rule is hidden. Now search transactions with rule c–>a in which the value of a = c = 1, t6 only transaction which has the value a = c = 1, by putting 0 instead of 1 update transaction in place of
A. Now the confidence of rule c–>a is 0% after calculation, which is less than the minimum confidence so now this rule is hidden. Now those rules are taken in which a is in lhs.

Table 4

Tid	abc
t1	011
t2	011
t3	011
t4	010
t5	100
t6	001

The rules are a–>b and a–>c but confidence of both the rules is less than minimum confidence so these rules are week rule and its not required to hide. So after hiding item a table 4 shows the modified database . So the data base unnecessarily scans by hybrid algorithm . Because to find the same sensitive item a in lhs it scans the data base .and the item a is already hidden in the data base so that it doesn't make any difference . Proposed algorithm 2 removes this problem of hybrid algorithm.

VI. CONCLUSION & FUTURE WORK

In this thesis, a new algorithm was proposed to solve privacy- preserving data mining problems. The major contributions of this thesis work are summarized as follows: chapter II introduces a new categorization of ppdm techniques. In chapter 3, a new technique based on item-restriction that hides sensitive itemsets was proposed. Results were also displayed in the chapter 4.

Chapter 2 classified the existing sanitizing algorithms into three levels. The first level is either raw data or databases where transactions reside. Techniques and algorithm of mining data is done at level secound. The third level is at the output of different data mining algorithms and techniques. The focus in this thesis is on level 1 and level 3. Level 1 technique and algorithms take (as input) and database d and a database d′ produce as a output by modifying d. Where mining for rules will not show sensitive patterns. The alternative scenario at level 3 is to remove the sensitive patterns from the set of frequent patterns and publish the rest. And according to this scenario the publishing of database d does not need.

Privacy-preserving data mining can be applied in different domains. The focus in this thesis is on the association rule mining domain. The goal of association rule mining is to find (in databases) all patterns based on some hard thresholds, such as the min_ support and the min_confidence . Some patterns that are sensitive nature , might need to hide owners of the databases. The degree of sensitivity of item and sensitivity of item decided to help the data owner is decided by the data miner . Nowadays, determining the most effective way to protect sensitive patterns while not hiding non-sensitive ones as a side effect is a crucial research issue.

Chapter 3 introduced an effective privacy-preserving algorithm. It also studied the existing sanitizing algorithms and stated their drawbacks. We showed that previous methods remove more knowledge than necessary.

Future Work

Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs and increase sales of item and enhance research area. While data mining in general represents a significant advance in the type of analytical tools currently available, there are limitations to its capability. One limitation is that although data mining can help reveal patterns and relationships, it does not tell the user the value or significance of these patterns. It does not tell the users which patterns are sensitive and which are not. Also one limitation of proposed algorithm 2 is that it hides extra rules.

It can be said that software privacy failures can be direct result of one or more of the following points that are taken from risk management:

- Overestimation: to overemphasize data mining results which leads to false conclusions and incorrect decisions.
- Underestimation: failure to predict what adversaries could do with data mining results to penetrate privacy.
- Over-confidence: inaccurate assumptions based on software developer’s certainty on how they would handle the situation.
- Complacency: to feel quiet secure and be unaware of some potential danger.
- Ignorance: when there is a lack of knowledge and virtually no intelligence, we are at the mercy of events.

- Failure to join the dots: failure to assemble pieces of intelligence to make a coherent whole. After all, data mining tools output patterns but cannot interpret or analyze these patterns. The human intelligence is essential at this point.

Future research arising from the work presented in this thesis may focus on the following:

- Association rule mining is of relevance to e-commerce applications. In this thesis, the focus is on the accuracy of the data and blocking the inference channels. However, in practice, some commercial criteria may also be important and should be considered in the technique implementation.
- Chapter 4 contains a new technique that hides sensitive item. It is interesting to come up with more new techniques and compare them to this proposed technique.
- This thesis consistently preferred accurate data or patterns to be shared between parties or published to the public. However, there are cases where less accurate (distorted) data may be preferable. Whether and when to consider such less accurate data are questions that deserve further study.

Although privacy-preserving data mining has been studied for many years, and it will continue to be studied and will become the core of each data mining project. This is because security and privacy are necessary factors to convince data owners to share or publish their data for the common good.

REFERENCES

[01] r. Agrawal, t. Imielinski, and a. Swami. *Mining association rules between sets of items in large databases*. In proceedings of the acm sigmod conference on management of data, pages 207–216, new york, ny, usa, may 1993. Acm press.

[02] a. K. Pujari. *Data mining techniques* (book), 2001. University press (india) limited.

[03] r. Chen, k. Sivakumar, and h. Kargupta. *Distributed web mining using bayesian networks from multiple data streams*. In n. Cercone, t. Young lin, and x. Wu, editors, proceedings of the 2001 ieee international conference on data mining (icdm'01), pages 75–82, san jose, california, usa, november 2001. Ieee computer society.

[04] s. Goldwasser. *Multi-party computations: past and present*. In proceedings of the 16th annual acm symposium on the principles of distributed computing, pages 1–6, santa barbara, california, usa, 1997. Acm press.

[05] m. Atallah, e. Bertino, a. Elmagarmid, m. Ibrahim, and v. Verykios. *Disclosure limitation of sensitive rules*. In proceedings of 1999 ieee knowledge and data engineering exchange workshop (kdex'99 pages 45–52, chicago, illinois usa, november 1999. Ieee computer society.

[06] e. Dasseni, v. S. Verykios, a. K. Elmagarmid, and e. Bertino. *Hiding association rules by using confidence and support*. In i. S. Moskowitz, editor, proceedings of the 4th information hiding workshop, volume 2137, pages 369–383, pittsburg, pa, usa, april 2001. Springer veralg lecture notes in computer science.

[07] s. R. M. Oliveira and o. R. Zaiane. *Algorithms for balancing privacy and knowledge discovery in association rule mining*. In proceedings of the 7th international database engineering and applications symposium (ideas'03), pages 54–65, hong kong, china, july 2003. Ieee computer society.

[08] y. Saygin, v. S. Verykios, and c. Clifton. *Using unknowns to prevent discovery of association rules*. In acm sigmod record, volume 30(4), pages 45–54, new york, ny, usa, december 2001. Acm press.

.

[09] m. Kantarcioglu and c. Clifton. *Privacy-preserving distributed mining of association rules on horizontally partitioned data*. In ieee transactions on knowledge and data engineering journal, volume 16(9), pages 1026–1037, piscataway, nj, usa, september 2004. Ieee educational activities department.

[10] c. Clifton and d. Marks. *Security and privacy implications of data mining*. In workshop on data mining and knowledge discovery, pages 15–19, montreal, canada, february 1996. University of british columbia, department of computer science.

[11] y. Saygin, v. S. Verykios, and a. K. Elmagarmid. *Privacy preserving association rule mining*. In z. Yanchun, a. Umar, e. Lim, and m. Shan, editors, proceedings of the 12th international workshop on research issues in data engineering: engineering e-commerce/e- business systems (ride'02), pages 151–158, san jose, california, usa, february 2002. Ieee computer society.

[12] w. Du and m. J. Atallah. *Secure multi-party computation problems and their applications: a review and open problems*. In v. Raskin, s. J. Greenwald, b. Timmerman, and d. M. Kienzle, editors, proceedings of the new security paradigms workshop, pages 13–22, cloudcroft, new mexico, usa, september 2001. Acm press.

[13] j. Vaidya and c. Clifton. *Privacy preserving association rule mining in vertically partitioned data*. In proceedings the 8th acm sigkdd international conference on knowledge discovery and data mining, pages 639–644, edmonton, alberta, canada, july 2002. Acm press.

[14] y. Lindell and b. Pinkas. *Privacy preserving data mining*. In crypto-00, volume 1880, pages 36–54, santa barbara, california, usa, 2000. Springer verlag lecture notes in computer science.