ICEB 2004 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-5-2004

Bayesian Network Induction with Incomplete Private Data

Justin Zhan

LiWu Chang

Stan Matwin

Bayesian Network Induction with Incomplete Private Data

Justin Zhan[1], LiWu Chang[2] , Stan Matwin[3]

[1,3] School of Information Technology & Engineering, University of Ottawa, Canada
[2] Center For High Assurance Computer Systems, Naval Research Laboratory, USA
[13]{zhizhan, stan}@site.uottawa.ca, [2]lchang@itd.nrl.navy.mil

ABSTRACT

A Bayesian network is a graphical model for representing probabilistic relationships among a set of variables. It is an important model for business analysis. Bayesian network learning methods have been applied to business analysis where data privacy is not considered. However, how to learn a Bayesian network over private data presents a much greater challenge. In this paper, we develop an approach to tackle the problem of Bayesian network induction on private data which may contain missing values. The basic idea of our proposed approach is that we combine randomization technique with Expectation Maximization (EM) algorithm. The purpose of using randomization is to disguise the raw data. EM algorithm is applied for missing values in the private data set. We also present a method to conduct Bayesian network construction, which is one of data mining computations, from the disguised data.

Keywords: Bayesian network, randomization, EM algorithm, privacy.

1. INTRODUCTION

A Bayesian network is a high-level representation of a probability distribution over a set of variables that are used for constructing a model of the problem domain. It has been widely used in sales decision making, marketing systems, risk analysis, cost benefit factor inference in E-services, and other business applications. As Bill Gates mentioned, in Los Angeles Times on October 28, 1996, that Microsoft's competitive advantage was its expertise in Bayesian network. The Bayesian framework offers many advantages over alternative modeling approaches. A Bayesian network can be used to compute the predictive distribution on effects of possible actions since it is a model of the problem domain probability distribution. For instance, it can be used for the optimal procedure of making decision in risk analysis. Bayesian network model has been also found to be very robust to tolerate small alterations. Thus, it can be used for sales marketing systems since these systems need to be able to follow market changes rapidly without making much modification on the model. In addition, expert domain knowledge can be coded as prior distribution in Bayesian modeling.

Over the last decade, the Bayesian network has been widely utilized and various techniques have been developed to learn Bayesian networks from data. Although the techniques that have been developed are effective, new techniques dealing with Bayesian network induction over private data are required. In other words, we need methods to learn a Bayesian network with data privacy preserved. In particular, we try to solve the following type of problem: To improve their services, business agents want to obtain certain information about their customers. For instance, the information of interest can be the probability that a customer buys butter given that she bought bread. In order to collect the information from its customers, a business agent sends out a survey containing a set of questions. Customers are expected to answer those questions and send their answers back. However, because the survey contains questions regarding private information, not every user feels comfortable to disclose her answers to those questions. To protect customers' privacy, they need somehow mask their answers before sending them back to the agent. We observe that randomization can be used to hide customers' answers. This is because on one hand the raw data values of the original customer's information can be disguised via randomization, on the other hand, their probability distribution can be approximately estimated from the disguised data by agents for business analysis. In practice, even though randomization protection techniques are utilized, some customers still do not provide their data. Therefore, the collected data usually contain missing values. The challenge is how to conduct data mining from the disguised data with missing values?

To address this challenge, we propose to combine the proposed *Multi-Group Randomization* technique with the Expectation Maximization (EM) algorithm [2]. The basic idea of multi-group randomization technique is to partition attributes of a data set into several groups and randomize data in each group independently so that business agents can't tell with probabilities better than a pre-defended threshold whether the data from a customer contain truthful information or false information. Although information from each individual customer is scrambled, if the number of customers is significantly large, the aggregate information of these customers can be estimated with decent accuracy. Such property is useful for building a Bayesian network, since it is based on aggregate values of a data set, rather than individual data items. The EM algorithm is a general iterative algorithm for parameter estimation by maximum likelihood when some of the random variables involved are not completely observed. It formalizes an intuitive idea for obtaining parameter estimates when some of the data are missing. The term EM was introduced in [2] where proof of general results about the behavior of the algorithm was given as well as a large number of applications was introduced.

The rest of the paper is organized as the follows: We describe the background of Bayesian network in Section 2. In Section 3, we briefly describe how multi-group randomized response technique works. Then in section 4, we describe the EM algorithm. We give an algorithm to build a Bayesian network on disguised data with missing values in Section 5. In Section 6, we discuss related work. We give our conclusion in Section 7.

2. BACKGROUND OF BAYESIAN NETWORKS

A Bayesian network [10] is a representation of probabilistic conditional independence. The network model is a directed acyclic graph (DAC) in which a node represents a random variable and a directed link or an arrow denotes the conditional probability of a child node given its parent node with the arrow pointing from the parent to the child. Suppose a network BN_1 have n nodes. Associated with each node (e.g., X_i , $i \in [1,...,n]$) is a conditional probability table (CPT) that shows the probabilistic dependency relationships between X_i and its parent nodes, $Parent(X_i)$, i.e., $P(X_i = x_i \mid Parent(X_i) = pa_i, D, BN_1)$, where D is the dataset, BN_1 is the network model, and x_i and pa_i are values of X_i and $Parent(X_i)$, respectively. An entry of the CPT is a network parameter.

A Bayesian network may be inductively learnt from a data set [6]. As discussed in [6], inductive Bayesian learning involves the selection of network models and the estimation of parameters (e.g., the CPTs). Given a complete data set D, learning of the best Bayesian network model is carried out by computing the posterior probability $P(BN_1 \mid D)$ for different BN_1 , and selecting the one with the largest value. From the Bayes rule, $P(BN_1 \mid D)$ is determined by the likelihood $P(D \mid BN_1)$, the prior $P(BN_1)$, and the data probability D:

$$P(BN_1 \mid D) = \frac{P(D \mid BN_1) \cdot P(BN_1)}{P(D)}$$

Given the prior $P(BN_1)$, the model selection is determined by $P(D \mid BN_1)$. Assume the best model is denoted as BN. The network parameters are estimated with the data of D and the prior probability distribution of parameters. In this paper, we assume the Bayesian network model is known, and will mainly consider the induction of parameters or the conditional probability tables. Suppose the uniform prior is used. (See [6] for different prior probabilities.) The CPT entry associated with the values of $X_i = x_i$ and $Parent(X_i) = pa_i$ is as follows:

$$P(X_i = x_i \mid Parent(X_i) = pa_i, D, BN_1) = \frac{\#(X_i = x_i, Parent(X_i) = pa_i) + 1}{\#(Parent(X_i) = pa_i) + |X_i|} \quad \text{EQ.(1)}$$

for a complete data set D (i.e., D without missing value), where #(.) stands for the total number of the records that satisfies conditions in (.), and $|X_i|$ denotes the number of values that X_i has. Divide both the numerator and the denominator of EQ.(1) by the total number of data records (|D|) to obtain

$$\frac{\#(X_i = x_i, Parent(X_i) = pa_i)/|D| + 1/|D|}{\#(Parent(X_i) = pa_i)/|D| + |X_i|/|D|} = \frac{P(X_i = x_i, Parent(X_i) = pa_i) + 1/|D|}{P(Parent(X_i) = pa_i) + |X_i|/|D|}$$

In this paper, we will use $P(X_j = x_j)$ to denote the probability measure obtained from the frequency count $\#(X_j = x_j)/|D|$.

3. MULTI-GROUP RANDOMIZED RESPONSE TECHNIQUES

Randomized Response techniques were firstly introduced in [11] as a technique to solve the following survey problem: to estimate the percentage of people in a population that has attribute A, queries are sent to a group of people. Since the attribute A is related to some confidential aspects of human life, respondents may decide not to reply at all or to reply with incorrect answers. To enhance the level of cooperation, instead of asking each respondent whether she has attribute A, the interviewer asks each respondent two related questions, the answers to which are opposite to each other [11]. For example, the questions could be like the following. If the statement is correct, the respondent answers “yes”; otherwise, she answers ``no".

1. I have the sensitive attribute A.
2. I do not have the sensitive attribute A.

Respondents use a randomizing device to decide which question to answer, without letting the interviewer know which question is answered. The randomizing device is designed in such a way that the probability of choosing the first question is q , and the probability of choosing the second question is $1-q$. Although the interviewer learns the responses (e.g., ``yes" or ``no"), he does not know which question was answered by the respondents. Thus the respondents' privacy is preserved. Since the interviewer's interest is to get the answer to the first question, and the answer to the second question is exactly the opposite to the answer for the first one, if the respondent chooses to answer the first question, we say that she is telling the truth; if the respondent chooses to answer the second question, we say that she is telling a lie. To estimate the percentage of people who has the attribute A, we have

$$P^*(A=yes) = P(A=yes)\cdot q + P(A=no)\cdot(1-q)$$
$$P^*(A=no) = P(A=no)\cdot q + P(A=yes)\cdot(1-q)$$

where $P^*(A=yes)$ is the proportion of the ``yes'' responses obtained from the survey data. $P^*(A=no)$ is the proportion of the “no” responses obtained from the survey data. *P(A = yes)* is the estimated proportion of the “yes” responses to the sensitive questions. P(A=no) is the estimated proportion of the “no” responses to the sensitive questions. Getting *P(A = yes)* and *P(A = no)* is the goal of the survey. By solving the above equations, we can get *P(A = yes)* and *P(A = no)* if $q \neq 0.5$.

In this section, we develop a multi-group randomized response technique to deal with the problem of Bayesian network construction where multiple attributes are normally involved.

3.1 Notations

In this paper, we assume data are binary. Suppose there are N attributes ($A_1, \cdots, A_N$) in a data set. Let L represent any logical expression based on those attributes (e.g., $L=(A_1=1)\wedge(A_2=0)$; let $\bar{L}$ denote the logical expression that reverses the 1's in L to 0's and 0's to 1's; we call $\bar{L}$ the opposite of L in value assignments. For example, for the L in the previous example, $\bar{L} = (A_1 = 0 \wedge A_2 = 1)$. Let $P^*(L)$ be the proportion of the records in the whole *disguised* data set that satisfy L = *true*. Let *P(L)* be the proportion of the records in the whole *undisguised* data set that satisfy L = *true* (the undisguised data set contains the true data, but it does not exist). $P^*(L)$ can be observed directly from the disguised data, but *P(L)*, the actual proportion that we are interested in, cannot be observed from the disguised data because the undisguised data set is not available to anybody; we have to estimate *P(L)*. Our goal is to find a way to estimate *P(L)* from $P^*(L)$. In our multi-group scheme, we also divide each expression L into multiple sub-expressions. For example, in a m-group scheme, we write $L = L_1 L_2 L_3 \cdots L_m$, where L_i contains only the attributes in the group i.

3.2 One-Group Scheme

In the one-group scheme, all the attributes are put in the same group, and all the attributes are either reversed together or keeping the same values. In other words, when sending the private data to the central database, users either tell the truth about all their answers to the sensitive questions or tell the lie about all their answers. The probability for the first event is q , and the probability for the second event is $1-q$. The general model for the one-group scheme is described in the following:

$$P^*(L) = P(L)\cdot q + P(\bar{L})\cdot(1-q)$$
$$P^*(\bar{L}) = P(\bar{L})\cdot q + P(L)\cdot(1-q)$$

L denotes any logic expression among the attributes, for instance, L could be $(A_1=0, A_2=1, A_3=0, A_4=0)$ and represents the case when attribute 1 is 0, attribute 2 is 1, attribute 3 is 0, and attribute 4 is 0. The above model can be simplified. Let M_1 denote the coefficients matrix of the above equations, and let $p = q$ and $q = (1-q)$, then

$$\begin{Bmatrix} P^*(0Bar) \\ P^*(1Bar) \end{Bmatrix} = M_1 \begin{Bmatrix} P(0Bar) \\ P(1Bar) \end{Bmatrix} \qquad \text{Eq. (3)}$$

Where 0Bar = L, 1Bar = $\bar{L}$, and

$$M_1 = \begin{bmatrix} p & q \\ q & p \end{bmatrix}$$

3.3 Two-Group Scheme

In the one-group scheme, if the interviewer somehow knows whether the respondents tell a truth or a lie for one attribute, she can immediately obtain all the true values for all other attributes of a respondent's response. To improve the privacy level, data providers divide all the attributes into two groups (All the data providers should group the attributes in the same way, e.g., if one user lets attribute A_1 and A_2 be in group 1, then other users also let attribute A_1 and A_2 be in group 1). They then apply the randomized response techniques for each group *independently*. For example, the users can tell the truth for one group while telling the lie for the other group. With this scheme, even if the interviewers know information about one group, they will not be able to derive the information for the other group because they are disguised independently.

To show how to estimate $P(L_1 L_2)$, we look at all the contributions to $P^*(L_1 L_2)$. There are four parts that contribute to $P^*(L_1 L_2)$:

1. $P(L_1 L_2)$: users tell the truth about all the answers for both groups; the probability for this event is q^2.

2. $P(L_1 \bar{L}_2)$: users tell the truth about all the answers for group 1 and tell the lie about all the answers for group 2; the probability for this event is $q \cdot (1-q)$.

3. $P(\bar{L}_1 L_2)$: users tell the lie about all the answers for group 1 and tell the truth about all the answers for group 2; the probability for this event is $q \cdot (1-q)$.

4. $P(\bar{L}_1 \bar{L}_2)$: users tell the lie about all the answers for both groups; the probability of this event is $(1-q)^2$.

We then have the following equation:

$$P^*(L_1L_2) = P(L_1L_2)\cdot q^2 + P(L_1\bar{L}_2)\cdot q(1-q) +$$
$$P(\bar{L}_1 L_2)\cdot q(1-q) + P(\bar{L}_1\bar{L}_2)\cdot(1-q)^2$$

There are four unknown variables in the above equation $(P(L_1L_2), P(L_1\bar{L}_2), P(\bar{L}_1L_2), P(\bar{L}_1\bar{L}_2))$. To solve the above equation, we need three more equations. We can derive them using the similar method. The final equations are described in the following:

$$\begin{Bmatrix} P^*(0Bar) \\ P^*(1Bar) \\ P^*(2Bar) \end{Bmatrix} = M_2 \cdot \begin{Bmatrix} P(0Bar) \\ P(1Bar) \\ P(2Bar) \end{Bmatrix}$$

where $0Bar = L_1L_2$, $1Bar = L_1\bar{L}_2 + \bar{L}_1L_2$, $2Bar = \bar{L}_1\bar{L}_2$, M_2 is the coefficients matrix, and let $p = q$ and $q = 1-q$, then,

$$M_2 = \begin{bmatrix} p^2 & pq & q^2 \\ 2pq & p^2+q^2 & 2pq \\ q^2 & pq & p^2 \end{bmatrix}$$

By solving the above equations, we can get $P(L_1L_2) = P(0Bar)$

3.5 Multi-Group Scheme

Similar techniques can be employed to extend the above schemes to three-group scheme, four-group scheme, and so on. In the following, we will give a general formula for the m-group scheme, where we apply randomized response techniques for each group *independently*.

$$\begin{Bmatrix} P^*(0Bar) \\ P^*(1Bar) \\ \cdots \\ P^*(mBar) \end{Bmatrix} = M_m \cdot \begin{Bmatrix} P(0Bar) \\ P(1Bar) \\ \cdots \\ P(mBar) \end{Bmatrix}$$

Where $0Bar = L_1L_2L_3\cdots L_m$, $1Bar = \bar{L}_1L_2\cdots L_m + \cdots + L_1L_2\cdots\bar{L}_m$,…, $mBar = \bar{L}_1\bar{L}_2\bar{L}_3\cdots\bar{L}_m$ and M_m is the coefficients matrix. Let $p = q$ and $q = 1-q$, then

$$M_m = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & & a_{1(n+1)} \\ a_{21} & a_{22} & a_{23} & \cdots & & a_{2(n+1)} \\ a_{31} & a_{32} & a_{33} & \cdots & & a_{3(n+1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{(n+1)1} & & a_{(n+1)2} & a_{(n+1)3} & & \cdots a_{(n+1)(n+1)} \end{bmatrix}$$

The coefficiency matrix is different for different group scheme. The values of a_{ij} can be derived as we did for two-group scheme. After we derive the coefficiency matrix, we can solve the above equation and obtain $P(E_1E_2\cdots E_n) = P(0Bar)$ since $P^*(0Bar)$, $P^*(1Bar)$, …, $P^*(nBar)$, can be obtained from the randomized data.

4. EM ALGORITHM

The Expectation Maximization (EM) algorithm [2, 8] is a general algorithm for maximum likelihood estimation where the data are incomplete or the likelihood function involves hidden variables. The EM algorithm starts with randomly assigned values to all the parameters to be estimated. It then iteratively alternates between two steps, called the expectation step (i.e., the E-step) and the maximization step (i.e., the M-step), respectively. In the E-step, it computes the expected likelihood for the complete data (Q-function) where the expectation is taken with regards to the computed conditional distribution of the hidden variables given the current settings of parameters and our observed data. In the M-step, it re-estimates all the parameters by maximizing the Q-function. Once we have a new generation of parameter values, we can repeat the E-step and M-step. The above process continues until the likelihood converges.

Assume that the joint probability for hidden data Y and observed data Z is parameterized using m, as $P(y,z \mid m)$. The marginal probability for Z is then $P(z \mid m) = \sum_y P(y,z \mid m)$. Given observed data z, we want to find the value of m that maximizes the log likelihood, $L(m) = \log P(z \mid m)$. The procedures of the EM algorithm are as follows:

1. Initialize $m^{(0)}$ randomly or heuristically according to prior knowledge about what the optimal parameter value might be.

2. Iteratively improve the estimate of m by alternating between the following two steps:

(a) E-step: Compute a distribution $\tilde{P}^{(t)}$ over the range Y such that $\overset{\sim (t)}{P} = P(y \mid z, m^{t-1})$.

(b) M-step: Set $m^{(t)}$ to the m that maximizes $E_{\overset{\sim (t)}{p}}(\log P(y,z \mid m))$.

3. Stop when the algorithm converges to a local maximum.

5. CONSTRUCTION OF A BAYESIAN NETWROK

The problem of learning Bayesian networks comes in several varieties. The structure of the network can be either *known or unknown*, and the variables in the network can be either *fully observable or with missing values*. In this paper, we consider the case of *known structure* and *with missing values*. Since the structure is known, construction requires only the computation of the CPT for each node. We will describe how to compute the CPTs when data set is complete, and then present an algorithm to deal with Bayesian network construction when the data set D is disguised and incomplete. Let parent nodes of the ith node be $Y_1, \cdots, Y_p$. To obtain CPT entries of node i, we need to compute the two terms:
$P(X_i = x_i, Y_1 = y_1, \cdots; Y_p = y_p)$ and $P(Y_1 = y_1, \cdots, Y_p = y_p)$
Since

$$P(X_i = x_i \mid Parent(X_i) = pa_i, D, BN)$$
$$= \frac{P(X_i = x_i, Y_1 = y_1, \cdots, Y_p = y_p) + 1/\mid D\mid \cdot}{P(Y_1 = y_1, \cdots, Y_p = y_p) + \mid X_i \mid / \mid D \mid}$$

When data set D is undisguised and complete, these two terms can be easily computed. But, it is not the case if D is disguised and incomplete. We provide the following algorithm to deal with Bayesian network construction when the data set D is disguised and incomplete.

Step I: Estimate missing values
Use EM algorithm to estimate the missing values in the disguised data set and obtain a complete data set CD. Please note that the distribution of randomized data very likely differs from the data distribution before randomization. Therefore, we need test the data distribution after randomization and apply proper EM algorithm.

Step II: Compute CPTs
Apply the estimation model of the randomized technique described in Section 3 to estimate those terms needed for the CPT entries. Consider the Wet-grass example, where wet-grass (*W*) can be caused by either rain (*R*) or sprinkling (*S*). The particular CTP of grass being wet, given rain and sprinkling is as follows.

$$P(W = t \mid S = t, R = t, CD, BN)$$
$$= \frac{P(W = t, S = t, R = t) + 1/\left|CD\right|}{P(S = t, R = t) + \left|W\right| / \left|CD\right|}$$

(|CD| denotes the total number of records in the complete data set.) Assume the data set is disguised using one-group scheme. We compute $P^*(W = t, S = t, R = t)$ and $P^*(W = f, S = f, R = f)$ on the data set CD. We then apply one-group estimation model to compute *P(W=t, S=t, R=t).*

$$\begin{Bmatrix} P^*(W=t,S=t,R=t) \\ P^*(W=f,S=f,R=f) \end{Bmatrix} = M_1 \cdot \begin{Bmatrix} P(W=r,S=t,R=t) \\ P(W=f,S=f,R=f) \end{Bmatrix}$$

where $M_1 = \begin{bmatrix} q & 1-q \\ 1-q & q \end{bmatrix}$. Similarly, *P(S=t, R=t)* can be computed. Hence, we can calculate *P(W=t | S=t, R=t).* With a similar calculation, other CPTs can be obtained.

6. PREVIOUS WORK

There are two existing approaches to solve privacy preserving data mining problems. One is the perturbation approach, the other is the secure multi-party computation approach. Agrawal and Srikant proposed a scheme for privacy-preserving data mining using random perturbation [3]. In their scheme, a random number is added to the value of a sensitive attribute. For example, if x_i is the value of a sensitive attribute, $x_i + r$, rather than x_i, will appear in the database, where r is a random value drawn from some distribution. The paper shows that if the random number is generated with some known distribution (e.g., uniform or Gaussian distribution), it is possible to recover the distribution of the values of that sensitive attribute. With the independence assumption of the attributes, the paper then shows that a decision tree classifier can be built with the knowledge of distribution of each attribute. Du and Zhan [4] proposed to use randomized response techniques [11] to tackle privacy-preserving data mining problem. In their scheme, they do not assume the independence of the attributes. Rizvi and Haritsa presented a scheme called MASK to mine associations with secrecy constraints in [9], and Evfimievski et al. proposed an approach to conduct privacy preserving association rule mining based on randomization techniques [1]. Kargupta et al. further analyze the effectiveness of randomization approach in [5]. Several SMC-based privacy preserving data mining schemes have been proposed [7, 7, 13]. These studies mainly focused on two-party distributed computing, and each party usually contributes a set of records. In our proposed research, we focus on centralized computing, and each participant only has certain number of records to contribute. All records are combined together into a central database before the computation begins.

7. CONCLUSION AND FUTURE WORK

As mentioned, Bayesian network learning is proven to be valuable to the business world. In this paper, we have combined multi-group randomized response techniques with EM algorithm for constructing Bayesian networks over disguised data set where some data values are missing. In our future work, we will continue refining our technique to handle various risk-analyses in business. We will also extend our solution to data sets consisting of non-binary data types.

REFERENCES

[1] J. Gehrke, A. Evfimievshi, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining", In proceedings of the 22nd ACM SIGACT-SIGMOD -SIGART symposium on principles of database systems, San Diego, CA, June, 2003.

[2] N. Laird, A. Dempster, D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm, the royal statistical society, Series B, Methodological, 39:1-38, 1977.

[3] W. Barksdale, "New randomized response techniques for control of non-sampling errors in surveys", PhD thesis, University of North Carolina, Chapel Hill, 1971.

[4] W. Du, Z. Zhan, "Using randomized response techniques for privacy-preserving data mining", In proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining, Washington, DC, USA, August 24-27, 2003.

[5] Q. Wang, H. Kargupta, S. Datta, K. Sivakumar, "Random data perturbation techniques and privacy preserving data mining", In the IEEE international conference on data mining, Florida, USA, 2003.

[6] D. Heckerman, "Bayesian networks for knowledge discovery", In Advances in knowledge discovery and data mining, AAAI Press/MIT Press, 1996.

[7] Y. Lindell and B. Pinkas, "Privacy preserving data mining", Advances in cryptology-CRYPTO'00, 1880 of lecture notes in computer science, Spinger-Verlag: 36-54, 2000.

[8] R. Neal, G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants", Learning in graphical models, editor: M. Jordan, 1998.

[9] S. Rizvi, J. Haritsa, "Maintaining data privacy in association rule mining", In proceedings of the 28th VLDB conference, Hong Kong, China, 2002.

[10] S. Russell, P. Norvig, "Artificial Intelligence: A modern approach", Prentice Hall, Upper Saddle River, New Jersey 07458, 1995.

[11] S. Warner, "Randomized response: A survey technique for eliminating evasive answer bias", Journal of American Statistical Association, 60(309):63-69, March, 1965.

[12] A. Tamhane, "Randomized response techniques for multiple sensitive attributes", the American Statistical Association, 76(376):916-923, December, 1981.

[13] J. Vaidya, C. Clifton, "Privacy preserving association rule mining in vertically partitioned data", In proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, July 23-26, 2002.