ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)

International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)

Cochin College of Engineering & Technology

Vol. 4, Special Issue 1, June 2015

Integrating Discrimination Prevention and Privacy Preservation into Data Mining

Sujitha S.[1], Sreejith S.[2]

PG Scholar, Dept. of CSE, LBSITW, Thiruvananthapuram, India [1]

Assistant Professor, Dept. of CSE, LBSITW, Thiruvananthapuram, India [2]

Abstract: In information society, massive and automated data collection is required for different purposes in our daily life. There are mainly two threats for individuals whose information is published: privacy and discrimination. In data mining, decision models are mainly derived on the basis of records stored by means of various data mining methods. But there may be a risk that the extracted knowledge imposes discrimination. Many organizations collect a lot of data for decision making. The sensitive information of the individual whom the published data relate to, may be revealed, if the data owner publishes the data directly. Discrimination prevention and privacy preservation need to be ensured simultaneously in decision making process. In this paper, Discrimination Prevention Data Mining (DPDM) and Privacy Preservation Data Mining (PPDM) have been studied and their relationships have been explored. Different privacy models and its impact on the data have also been analysed.

Keywords: Anti-discrimination, discriminatory attribute, classification rule, rule generalization, rule protection, k-anonymity.

I. INTRODUCTION

Data mining is the process of discovering useful knowledge or patterns from large datasets. Data mining, while extracting hidden information, may impose the risk of violation of non-discrimination and privacy in the dataset. Privacy refers to the individual right to choose freely what to do with one's own personal information whereas discrimination refers to unfair or unequal treatment of people based on membership to a category, group or minority.

This paper is organized as follows. Section 2 overviews the background information related to discrimination prevention in data mining. Section 3 discuss some basic definitions and models related to privacy preservation in data mining. Section 4 describes the proposal for obtaining discrimination free privacy protective dataset. Finally, Section 5 summarizes conclusions of discrimination prevention and privacy protection.

II. BACKGROUND ON DISCRIMINATION PREVENTION IN DATA MINING

Sociologically, discrimination is the prejudicial treatment of an individual based on his/her membership in a certain group or a community. It denies opportunities, for members of one group that are available to other groups. In data mining, if the training data itself are biased for or against a particular community, then the data model may show discriminatory prejudiced behavior. Therefore, to discover and eliminate such biases from the data, without harming their decision making utility, is very important and crucial.

Discrimination in the dataset are of two types: direct discrimination and indirect discrimination. Direct discrimination are the rules or procedures that explicitly mention minority groups based on the discriminatory attributes whereas indirect discrimination are the rules or procedures that do not explicitly mention the discriminatory attributes but unintentionally generates discriminatory attributes.

A. Related Work

In this section, the existing work dealing with antidiscrimination in data mining is discussed.

D. Pedreschi, S. Ruggieri and F. Turini (2008) presented the first paper which addresses the discrimination problem in data mining models [2]. They investigated how discrimination is hidden in data mining models and measured discrimination through a generalization of lift. They also introduced α protection as a measure of the discrimination power and proposed the extraction of classification rules.

F. Kamiran and T. Calders (2009) tackled the problem of classification scheme for learning unbiased models on biased training data [6]. The method is based on massaging the dataset by making the least modifications that leads to an unbiased dataset. But the main drawback was that numerical attributes and group of attributes were not considered as sensitive attribute.

D. Pedreschi, S. Ruggieri and F. Turini (2009) presented a systematic framework for measuring discrimination, based on the analysis of decision records [3]. They investigated whether direct and indirect discrimination can be found in a given set of records. They discussed integrating induction, through classification rule extraction, and deduction through a computational logic implementation of the analytical tools. In 2010, they also presented the discrimination discovery in databases in which unfair practices are hidden in a dataset of historical decisions [4].

S. Hajian, J. D.Ferrer and A. Martinez-Balleste (2011) introduced an anti-discrimination in the context of cyber security [5]. They proposed data transformation method for discrimination prevention and considered several discriminatory attributes and their combinations. The issue of data quality was also addressed in the paper.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)

International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)

Cochin College of Engineering & Technology

Vol. 4, Special Issue 1, June 2015

B. Basic Definitions

Some of the basic definitions related to discrimination prevention data mining [1] are discussed below:

- A *data set* is a collection of data objects (records) and their attributes.
- An *item* is an attribute along with its value, e.g., Race = black.
- An *item set*, i.e., X, is a collection of one or more items, e.g., {Foreign worker = Yes; City = NYC}.
- A *classification rule* is an expression X → C, where C is a class item (a yes/no decision), and X is an item set containing no class item, e.g., {Foreign worker = Yes; City = NYC} → Hire = no.
- The *support* of an item set, supp(X), is the fraction of records that contain the item set X. A rule X → C is completely supported by a record if both X and C appear in the record.
- The *confidence* of a classification rule, conf(X → C), measures how often the class item C appears in records that contain X.
- The *negated item set*, i.e., ¬X is an item set with the same attributes as X, but the attributes in ¬ X take any value except those taken by attributes in X.

C. Potentially Discriminatory and Nondiscriminatory Classification Rules

Let DIs be the set of predetermined discriminatory items in DB (e.g., DIs = {Foreign worker = Yes; Race = Black; Gender = Female}).

1. A classification rule X → C is potentially discriminatory (PD) when X = A, B with A ⊆ DIs a nonempty discriminatory item set and B a non-discriminatory item set. For e.g.: {Foreign worker = Yes; City = NYC} → Hire = No.
2. A classification rule X → C is potentially non-discriminatory (PND) when X = D, B is a non-discriminatory item set. For e.g.: {Zip = 10451; City = NYC} → Hire = No, or {Experience = Low; City = NYC} → Hire = No.

Pedreschi et al. [2] translated the qualitative statements into quantitative formal counterparts over classification rules and they introduced a family of measures of the degree of discrimination of a PD rule.

Definition 1*: Let A, B → C be a classification rule such that conf (B → C) > 0. The extended lift of the rule is*

$$elift(A, B \rightarrow C) = conf(A,B \rightarrow C)/conf(B \rightarrow C) \quad (1)$$

Definition 2: *Let α ∈ R be a fixed threshold and let A be a discriminatory item set. A PD classification rule c = A, B → C is α-protective w.r.t. elift if elift(c) < α. Otherwise, c is α -discriminatory.*

D. Discrimination Measurement

Direct and indirect discrimination discovery is the process of identifying α-discriminatory rules and redlining rules. First, based on the predetermined discriminatory items in the dataset, frequent classification rules are divided into two groups: PD and PND rules. Direct discrimination can be found by identifying α-discriminatory rules among the PD rules using a direct discrimination measure and a discriminatory threshold. Indirect discrimination is measured by identifying redlining rules among the PND rules combined with background knowledge. Next, the original data is transformed for each respective α-discriminatory rule, without affecting the data or other rules and is evaluated to check whether they are free of discrimination.

E. Data Transformation

The data transformation methods are based on the fact that the data set of decision rules would be free of direct discrimination if it only contained PD rules that are α-protective or are instances of at least one nonredlining PND rule. Similarly, the data set of decision rules would be free of indirect discrimination if it contained no redlining rules.

1. Rule Protection

In order to convert each α-discriminatory rule into an α-protective rule, based on the direct discriminatory measure (Definition 2), there are two methods that could be applied. One method (Method 1) changes the discriminatory item set in some records and the other method (Method 2) changes the class item in some records.

2. Rule Generalization

Rule generalization is the second data transformation method for discrimination prevention. It is based on the fact that if each α-discriminatory rule r': A, B → C in the database of decision rules was an instance of at least one nonredlining PND rule in the form of r: D, B → C, it means that the data set would be free of direct discrimination.

III. Background on Privacy Preservation in Data Mining

Privacy is defined as the rights of individuals to determine for themselves when, how, and what information about them is used for different purposes. Privacy Preservation in data mining is important nowadays because they allow publishing and sharing sensitive data for analysis.

A. Basic Preliminaries

In this section, the background knowledge required for reviewing data privacy technologies is discussed.

Given the data table D (A_1… An), a set of attributes $A = \{A_1 \ldots A_n\}$, and a record/tuple t ∈ D.

- T $[A_i \ldots A_j]$: sequence of the values $A_i \ldots A_j$ in t where $\{A_i \ldots A_j\} \subseteq \{A_1 \ldots A_n\}$.
- D $[A_i \ldots A_j]$: the projection maintaining duplicate records of attributes $A_i \ldots A_j$ in D.
- |D|: the cardinality of D.
- Identifiers are attributes that uniquely identify individuals in the database, like Passport number.
- A quasi-identifier (QI) is a set of attributes that, in combination, can be linked to external identified information for re-identifying an individual, for example: Zip code, Birthdate and Gender.
- Sensitive attributes (S) are those that contain sensitive information, such as Disease or Salary. Let S be a set of sensitive attributes in D.

B. Privacy Models

Definition 3: *k*- anonymity[9]

Let $D(A_1, \ldots, A_n)$, be a data table and $QI = \{Q_1, \ldots, Q_m\} \subseteq (A_1, \ldots, A_n)$, be a quasi-identifier. D is said to satisfy k-anonymity w.r.t. QI if each combination of values of attributes in QI is shared by at least k tuples (records) in D.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)

International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)

Cochin College of Engineering & Technology

Vol. 4, Special Issue 1, June 2015

The probability of identifying an individual is reduced to 1/*k*. A larger *k* can bring a lower probability of a linkage attack. *k*-anonymity can be achieved by QI generalization or QI suppression. A generalization replaces QI attribute values with a generalized version of them using the generalization taxonomy tree of QI attributes. A suppression consists in suppressing some values of the QI attributes for some (or all) records.

The main attacks identified in k- anonymity model are: homogeneity attack and background knowledge attack.

Definition 4: *l*-diversity[8]

A q-block is l-diverse if it contains at least well-represented values for the sensitive attribute S.*

l- diversity principle can sometimes lead to similarity attack.

Definition 5: *t*- closeness [7]

An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t.

IV. Discrimination and Privacy Aware Pattern Discovery

We first describe an algorithm to obtain a *k*-anonymous version of the original dataset and then present our data transformation methods to obtain an α- protective version of the dataset. Since using anti-discrimination techniques cannot make the dataset k-anonymous, it is better to apply anti-discrimination techniques to a *k*-anonymous dataset to obtain an α-protective k-anonymous dataset. The algorithm used for achieving anonymity is Datafly [10] which is discussed below.

Datafly algorithm is an algorithm used to provide anonymity and it is achieved by automatically generalizing, substituting, inserting, and removing information as appropriate with minimum data loss. The Datafly algorithm continuously generalizes quasi-identifiers via a Domain Generalization Hierarchy on any particular attribute, noted DGH (Attribute). Pseudo code for the algorithm is as follows:

FREQ <-- list of quasi-id value frequencies from table
QUASI <-- the set of quasi-identifiers in FREQ
with count < k
While QUASI accounts for > k records:
    Choose attribute A with greatest number of distinct values
    Generalize attribute A according to the DGH (A).
    Re-calculate QUASI and FREQ
    Remove records with quasi-id set Q, where Q refers to < k records
Return resulting table

V. CONCLUSION

Data mining is an important technology for extracting useful knowledge hidden in large collections of data. Privacy preserving and anti-discrimination techniques have been introduced in data mining to protect the sensitive data. In this paper, the relationship between privacy preserving data mining and discrimination prevention in data mining have been explored to address both threats simultaneously during the knowledge discovery process.

ACKNOWLEDGMENT

In this paper I would like to thank the Dept. of CSE, LBSITW, for giving me an opportunity to write this paper and also to all the authors of the papers which I refer, to get enough information to write this paper.

REFERENCES

[1] Sara Hajian and Josep Domingo- Ferrer, "A methodology for Direct and Indirect Discrimination Prevention in Data Mining," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 7, pp. 1445-1459, July 2013.

[2] D. Pedreschi, S. Ruggieri, and F. Turini, "Discrimination-Aware Data Mining," Proc. 14th ACM Int'l Conf. Knowledge Discovery and Data Mining (KDD '08), pp. 560-568, 2008.

[3] D. Pedreschi, S. Ruggieri, and F. Turini, "Integrating Induction and Deduction for Finding Evidence of Discrimination," Proc. 12th ACM Int'l Conf. Artificial Intelligence and Law (ICAIL '09), pp. 157- 166, 2009.

[4] S. Ruggieri, D. Pedreschi, and F. Turini, "DCUBE: Discrimination Discovery in Databases," Proc. ACM Int'l Conf. Management of Data (SIGMOD '10), pp. 1127-1130, 2010.

[5] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "Rule Protection for Indirect Discrimination Prevention in Data Mining," Proc. Eighth Int'l Conf. Modeling Decisions for Artificial Intelligence (MDAI '11), pp. 211-222, 2011.

[6] F. Kamiran and T. Calders, "Classification without Discrimination," Proc. IEEE Second Int'l Conf. Computer, Control and Comm.(IC 4 '09), 2009.

[7] N. Li, T. Li and S. Venkatasubramanian." t-Closeness: privacy beyond k-anonymity and l-diversity", In IEEE ICDE 2007, pp. 106-115. IEEE, 2007.

[8] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam," l-Diversity: privacy beyond k-anonymity," ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), Article 3, 2007.

[9] L. Sweeney, "k-Anonymity: A model for protecting privacy", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557-570, 2002.

[10] L. Sweeney, "Datafly: a system for providing anonymity in medical data" ,Database Security, XI: Status and Prospects, T. Lin and S. Qian (eds), Elsevier Science, Amsterdam, 1998.

BIOGRAPHIES

Sujitha S., received B.Tech degree in Computer Science and Engineering from University of Kerala in 2011. At present she is a PG Scholar at LBS Institute of Technology for Women, Thiruvananthapuram, India.

Sreejith S., received B.Tech degree in Information Technology from M. S. University and M.E in Computer and Communication Engineering from Anna University. He is currently working as an Assistant Professor in Computer Science and Engineering Dept. at LBSITW, Thiruvananthapuram, India.