INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Discrimination Prevention and Privacy Preservation in Data Mining

Kamal D. Kotapalle
Siddhant College of Engineering Sudumbare, Pune, Maharashtra India
Shyam Gupta
Assistant Professor, Siddhant College of Engineering Sudumbare, Pune, Maharashtra India

Abstract:
Data mining is an increasingly more vital technology for extracting useful information hidden in large collections of data. Along with privacy, discrimination is a very vital problem when taking into account the legal and ethical aspects of data mining. Most of the human being do not want to be discriminated just because of their, religion, gender, age, nationality and so on, when those attributes are used for building decisions about them especially like giving them a loan, job, insurance, etc. For this motivation, anti-discrimination methods including discrimination finding and avoidance have been introduced in data mining. It can be direct or indirect. When decisions are made depend on sensitive attributes then direct discrimination occurs. Indirect discrimination arises when decisions are based on non-sensitive attributes which are robustly correlated with biased sensitive ones. The proposed method gives samples of privacy preservation and potential discrimination in data mining, that privacy and discrimination risks should be handled collectively. We discover the association between privacy preserving data mining and discrimination avoidance in data mining to design methods capable of addressing both threats concurrently during the knowledge innovation process. This paper deals with the concepts for privacy preservation for direct and indirect discrimination avoidance in data mining.

Keywords*: Data mining, Antidiscrimination, direct and indirect discrimination prevention, rule protection, privacy preservation*

1. Introduction

Data mining is used to extract useful information, such as trends and patterns, from large amounts of data.Discrimination is defined by the process of unjustly treating public on the base of their belonging to a precise group, namely race, ideology etc. This involves denying opportunities to members of one group that are available to other group of people Discrimination is of two types. Direct discrimination and Indirect discrimination.

Direct discrimination consists of procedures or rules that unambiguously mention minority or disadvantaged groups based on sensitive discriminatory attributes interrelated to group membership. Indirect discrimination consists of procedures or rules that do not explicitly mentioning discriminatory attributes directly or indirectly generate discriminatory decisions. An instance of indirect discrimination is refusing to grant insurances or mortgages in urban areas they consider as deteriorating although certainly not the only one.

Indirect discrimination could need some background knowledge (rules), for example, that a certain zip code corresponds to a deteriorating area or an area with mostly female population. The background knowledge might be get from publicly available data (e.g., census data) or might be obtained from the original data set itself because of the existence of nondiscriminatory attributes that are highly associated with the sensitive ones in the original data set.

2. Related Work

The wide deployment of information systems based on data mining technology in decision making, the important of antidiscrimination in data mining did not get much care until 2008 [1]. Some techniques are used to the discovery and measure of discrimination. But some others deal with the prevention of discrimination. The discrimination discovery decision was first proposed by Pedreschi et al. [1], [3]. This technique is based on mining categorization rules (the inductive part) and analysis on them (the deductive part) on the root of quantitative actions of discrimination that formalize authorized definitions of discrimination. In the US Equal Pay Act [5] states that: “the selection rate for any ethnic race and sex which is less than four-fifths of the rate for the set with the highest rate will normally be regarded as evidence of conflicting effect”. In the existing discrimination discovery methods consider every rule

independently for measuring discrimination without allowing for other rules or the relation between them. In this paper we also consider relation between rules for discrimination discovery, depend on the presence or absence of discriminatory attributes. In discrimination prevention, the other major antidiscrimination aim in data mining consists of introducing patterns that do not lead to discriminatory decisions even if the unique training data sets are biased.
Three approaches are used:

2.1. Preprocessing
The source data is transformed in such a way that the discriminatory biases contained in the original data are removed so that no unfair decision rule can be mined from the transformed data and apply any of the standard data mining algorithms. In this preprocessing approach of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature [6], [7].

2.2. In processing
Change the data mining algorithms in such a way that the resulting models do not contain unfair decision rules. There is an alternative method to cleaning the discrimination from the original data set is proposed in [2] whereby the nondiscriminatory constraint is rooted into a decision tree learner by changing its splitting criterion and pruning policy through a novel leaf relabeling approach.

2.3. Post processing
In this approach modify the resulting data mining models, as an alternative of transform the original data set or changing the data mining algorithms.

3. Contributions
In more detail, our contributions are:

- We develop a new pre-processing discrimination prevention methodology including different data transformation methods that can prevent direct discrimination, indirect discrimination or together of them at the same time. To achieve this objective, the first step is to measure discrimination and identify categories and groups of individuals that have been directly and/or indirectly discriminated in the decision-making processes; the second step is to transform data in the proper way to remove all those discriminatory differences. Finally discrimination free data models can be produced from the transformed dataset without seriously damaging data quality. The experimental results reported demonstrate that the proposed techniques are quite successful in both goals of removing discrimination and preserving data quality.
- 2. We have investigated the problem of discrimination and confidentiality aware frequent sample innovation, i.e. the cleansing of the group of patterns mined from a transaction record in such a way that neither privacy-violating nor biased inferences can be inferred on the released patterns. We found that our discrimination preventing transformations do not interfere with a privacy preserving sanitization based on k-anonymity, thus accomplishing the task of combining the two and achieving a robust (and formal) notion of fairness in the resulting pattern collection. Further, we have presented extensive empirical results on the utility of the protected data. Specifically, we evaluate the distortion introduced by our methods and its effects on classification. It turns out that the utility loss caused by simultaneous anti-discrimination and privacy protection is only marginally higher than the loss caused by each of those protections separately. This result supports the practical deployment of our methods. Moreover, we have discussed the possibility of using our proposed framework while replacing k-anonymity with differential privacy.

4. Methods for Measuring Discrimination
Discrimination has been a central field for social scientific research for decades, resulting in a wide specter of methods and techniques used for measuring the phenomena. Which method is the most suitable depends on the question posed; each of the methods can provide important documentation of certain aspects of the dis
crimination complex, but used alone they all have obvious limitations if the purpose is to establish knowledge which may contribute to political action. In the next section, we will briefly discuss four different points of departure for discrimination as a examine field, using the applicability for policy intervention as a criteria for evaluation.
In this section, thus the approach, including the data transformation methods that can be used for direct and/or indirect discrimination prevention.

4.1. The Approach
Our approach for direct and indirect discrimination prevention can be described in terms of two phases:

4.1.1. Discrimination Measurement
Direct and indirect discrimination discovery includes identifying redlining rules and **α**-discriminatory rules. To this last part, primary, based on predetermined discriminatory items in DB, frequent classification rules in FR are divided in two groups: PD and PND rules.Secondt direct discrimination is calculated by identifying **α** -discriminatory rules among the PD rules using a direct discrimination measure (elift) and a discriminatory threshold (**α**). Third, indirect discrimination is measured by identifying redlining rules among the PND rules shared with background knowledge, using an indirect discriminatory measure (elb), and a discriminatory

threshold (**α**). Let MR be the database of direct **α** -discriminatory rules obtained with the above process. In addition RR be the record of red-lining rules and their respective indirect **α** -discriminatory rules obtained with the above process.

4.1.2. Data Transformation

Transform the original data DB in such a way to remove direct and/or indirect discriminatory bias, with minimum impact on the data and on legitimate decision rules, so that no unfair decision rule can be mined from the transformed data. In the following sections, present the data transformation methods that can be used for this purpose.

4.2. Data Transformation for Direct Discrimination

The proposed solution to prevent direct discrimination is 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 non-redlining PND rule. Therefore, a suitable data transformation with minimum information loss should be applied in such a way that each **α** -discriminatory rule either becomes **α** -protective or an instance of a no redlining PND rule. Call the first procedure direct rule protection (DRP) and the second one rule generalization.

4.2.1. Direct Rule Protection

In order to convert each **α** -discriminatory rule into an **α** -protective rule, base on the direct discriminatory measure (i.e., Definition 2), and should enforce the following inequality for each **α** -discriminatory rule r': A,B → C in MR, where A is a discriminatory item set:
Elift(r') < **α**
It is clear that Inequality can be satisfied by increasing the confidence of the base rule (B → C) of the **α** - discriminatory rule r' : A,B → C to a value higher than the right-hand side of Inequality, without affecting the value of conf(r' : A,B → C). A possible solution for increasing Expression

$$Conf(B \rightarrow C) = \frac{supp(B,C)}{Supp(B)} \quad \text{.....(1)}$$

4.2.2. Rule Generalization

Rule generalization is another data transformation method for direct 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 non redlining (legitimate) PND rule(r : D,B→ C), the data set would be free of direct discrimination.
In rule generalization, consider the relation between rules instead of discrimination actions. The following example illustrates this theory. Suppose a complainant claims discrimination against foreign workers among applicants for a job position. A classification rule {Foreign worker = Yes; City = NYC}→Hire = No with high elift supports the complainant's claim. However, the decision maker could disagree that this rule is an instance of a more general rule {Experience = Low; City = NYC}→ Hire= No. In other words, foreign workers are rejected because of their less experience, not just because they are foreign.
The two conditions can be comfortable in the following definition.

- **Definitions**. Let p ε[0,1]. A classification rule r' : A,B→ C is a p-instance of r : D,B→ C if both conditions below are true:
- **Condition 1**: conf(r) ≥ conf(r') .
- **Condition 2**: conf(r' : A,B → D) ≥P. Then, if r' is a p-instance of r (where p is 1 or a value near 1), r' is free of direct discrimination. Based on this concept, propose a data transformation method.

5. A Proposal for Privacy Aware Data Mining

Privacy is not just a goal or service like security, but it is the people's belief to reach a protected and controllable state, possibly without having to actively look for it by themselves. Therefore, privacy is defined as "the rights of individuals to determine for themselves what, how and when information about them is used for various goal". The protection of responsive data is a essential topic, which has involved many researchers in information technology. In information discovery, efforts at guaranteeing privacy when mining and sharing personal data have led to developing privacy preserving data mining (PPDM) techniques. PPDM have become more and more popular because they allow publishing and sharing sensitive data for secondary analysis. Various PPDM methods and models (measures) have been proposed to trade offs the service of the resulting data/models for
defending individual privacy against various kinds of privacy attacks.

5.1. Brief Review

The problem of protecting privacy within data mining has been extensively studied since the 1970s, when Dalenius was the first to formulate the statistical disclosure control problem. Research on data anonymization has carried on ever since in the official statistics community, and several computational procedures were proposed during the 1980s and 1990s, based on random noise addition, generalization, suppression, micro aggregation, bucketization, etc. .In that literature ,the approach was first to anonymize and then measure how much anonymity had been achieved, by either computing the probability of re-identification or performing record linkage experiments. In the late 1990s, researchers in the database community stated the k-anonymity model- a data set is k-anonymous if its records are indistinguishable by an intruder within groups of k. The novelty of this approach was that the anonymity

target was established ex ante and then computational procedures were used to reach that target. The computational procedures initially proposed for k-anonymity were generalization and suppression; micro aggregation was proposed later as a natural alternative. In 2000, the database community re-discovered anonymization via random noise addition, proposed in the statistical community as far back as 1986, and coined the new term privacy-preserving data mining (PPDM,[3]) is a more recent anonymity model that holds much promise: it seeks to render the influence of the presence/absence of any individual on the released outcome negligible. The computational approach initially proposed to achieve differential privacy was Laplace noise addition, although other approaches have recently been proposed. SDC and PPDM have become increasingly popular because they allow publishing and sharing sensitive data for secondary analysis.

5.2. Preliminaries

5.2.1. Basic Definitions

Given the data table D(A1,⋯ ,An), a set of attributes A = {A1,⋯ ,An}, and a record/tuple t ∈ D, t[Ai,⋯ ,Aj] denotes the sequence of the values of Ai,⋯ ,Aj in t, where {Ai,⋯ ,Aj} ⊆ {A1,⋯ ,An}. Let D[Ai,⋯ ,Aj] be the projection, maintaining duplicate records, of attributes Ai,⋯ ,Aj in D. Let |D| be the cardinality of D, that is, the number of records it contains. The attributes A in a database D can be classified into several categories. 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, Birthdates 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.

5.3. Models of Privacy

As mentioned in the beginning of this section, in the last fifteen years plenty of privacy models have been proposed to trade off the function of the resultant data/models for protecting individual privacy against different kinds of privacy attacks. Defining privacy is a difficult task. One of the key challenges is how to model the background knowledge of an adversary. Simply removing explicit identifiers (e.g., name, passport number) does not preserve privacy, given that the adversary has some background knowledge about the victim. It illustrates that 87% of the U.S. population can be uniquely identified based on 5-digit zip code, gender, and date of birth. These attributes are QI and the adversary may know these values from publicly available sources such as a voter list. An individual can be identified from published data by simply joining the QI attributes with an external data source (i.e., record linkage). 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, Zipcode, Birth

date and Gender. In order to prevent record linkage attacks between the released data and external identified data sources through quasi-identifiers, Samarati and Sweeney [79, 83]proposed the notion of k-anonymity.

Definition 1 (k-anonymity). Let D(A1,⋯ ,An) be a data table and QI = {Q1,⋯ ,Qm} ⊆ {A1,⋯ ,An} 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.

Consequently, the probability of linking a victim to a specific record through QI is at most 1/k. A data table satisfying this requirement is called k-anonymous. Other privacy measures to prevent record linkage include (X,Y)-anonymity [89] and multi-relational k-anonymity [66].k-Anonymity can protect the original data against record linkage attacks, but it cannot protect the data against attribute linkage (disclosure). In the attack of attribute linkage, the attacker may not precisely identify the record of the specific individual, but could infer his/her sensitive values (e.g., salary, disease) from the published data table D. Some models have been proposed to address this type of threat. The most popular ones are l-diversity [61] and t-closeness [56]. The general idea of these models is to diminish the correlation between QI and sensitive attributes. l-Diversity requires at least l distinct values for the sensitive attribute in each group of QI. Let q∗block be the set of records in D whose QI attribute values generalize to q.

Definition 2 (l-diversity). A q∗-block is l-diverse if it contains at least l well-represented values for the sensitive attribute S. A data table D is l-diverse if every q∗-block is l-diverse. t-Closeness requires the distribution of a sensitive attribute in any group on QI to be close to the distribution of the attribute in the overall table.

Definition 3 (t-closeness). A q∗-block is said to have t-closeness if the distance between the distribution of a sensitive attribute in this q∗-block and the distribution of the attribute in the whole table is no more than a threshold t. A data table D is said to have t-closeness if all q∗-blocks have t-closeness. Other privacy models for attribute disclosure protection include (α,k)-anonymity (k,e)-anonymity , (c,k)-safety, privacy skyline [13], m-confidentiality and (,m)- anonymity. Differential privacy is a privacy model that provides a worst-case privacy guarantee in the presence of arbitrary external information. It protects against any privacy breaches resulting from joining different databases. It guarantees that an adversary learns nothing about an individual, regardless of whether the individual's record is present or absent in the data.

5.3.1. Privacy-aware Frequent Pattern Discovery

In this section, we first describe the notion of k-anonymous frequent patterns and then we present a method to obtain a k-anonymous version of an original pattern set.

5.3.1.1. Anonymous Frequent Pattern Set

Given a support threshold σ, an itemset X is called σ-frequent in a database D if suppD(X) ≥σ. A σ-frequent itemset is also called σ-frequent pattern. The collection of all σ-frequent patterns in D is denoted by F(D,σ). The frequent pattern mining problem is formulated as follows: given a database D and a support threshold σ, find all σ-frequent patterns, i.e. the collection F(D,σ). Several algorithms have been proposed for finding F(D,σ). In this section we use the Apriori algorithm [2], which is a very common choice. In [5], the notion of k-anonymous patterns is defined as follows: a collection of patterns is k-anonymous if each pattern p in it is k-anonymous (i.e. supp(p) = 0 or supp(p) ≥ k) as well as any further pattern whose support can be inferred from the collection. The authors introduce a possible attack that exploits non k-anonymous patterns whose support can be inferred from the collection. Then they propose a framework for sanitizing patterns and block this kind of attacks.

Example 1. Consider again the motivating example and take
k=8.

The two patterns p1:{Job=veterinarian, Credit approved=yes} and p2: {Job=teacher 20000, Credit approved=yes} are 8-an, Salary > onymous because supp(p2) = 40 > 8 and supp(p1) = 41 > 8.However, an attacker can exploit a non-8-anonymous pattern {Job =teacher, ¬ (Salary > 20000), Credit approved=yes}, whose support he infers from supp(p1) − supp(p2) = 41 − 40 = 1. In order to check whether a collection of patterns is k-anonymous, in [5] the inference channel concept is introduced. Informally, an inference channel is any collection of patterns (with their respective supports) from which it is possible to infer non-k-anonymous patterns.

Definition 6. Given a database D and two patterns I and J, with I = {i1,...,im} and J = I ∪ {a1,...,an}, the set CJ I ={hX,suppD(X)i|I ⊆ X ⊆ J} constitutes an inference channel for the non k-anonymous pattern p = I ∪ {¬a1,...,¬an} if 0 < suppD(CJ I) < k where suppD(CJ I) = X I⊆X⊆J (−1)|X\I|suppD(X).

An example of inference channel is given by any pattern such as p : {b} which has a superset ps : {b,d,e} such that 0 < Cps p < k. In this case the pair hp,supp(p)i,hps,supp(ps)i constitutes an inference channel for the non-k-anonymous pattern {a,¬b,¬c}, whose support is given by supp(b)−supp(b,d)−supp(b,e)+supp(b,d,e).Then, we can formally define the collection of k-anonymous pattern set as follows.

Definition 4 (k-Anonymous pattern set). Given a collection of frequent patterns F(D,σ) and an anonymity threshold k, F(D,σ) is k-anonymous if (1) @p ∈ F(D,σ) s.t. 0 < supp(p) < k, and (2) @p1 and p2 ∈ F(D,σ) s.t. 0 < suppD(Cp2 p1) < k, where p1 ⊂ p2.

5.3.1.2 .Achieving an Anonymous Frequent Pattern Set

To generate a k-anonymous version of F(D,σ), Atzori et al. [5] proposed to first detect inference channels violating k-anonymity in F(D,σ) and then block them in a second step. The pattern sanitization method blocks an inference channel CJ I due to a pair of patterns I = {i1,...,im} and J = {i1,...,im, a1,...,an} in F(D,σ) by increasing the support of I by k to achieve supp(CI J) ≥ k. In addition, to avoid contradictions among the released patterns, the support of all subsets of I is also increased by k.

Example 2. Let us resume Example 1 and take k = 8. An inference channel due to patterns p1 and p2 can be blocked by increasing the support of pattern p1:{Job=veterinarian, Credit approved=yes} and all its subsets by 8. In this way, the non-8-anonymous pattern {Job=veterinarian, ¬(Salary > 15000), Credit approved=yes} is 8-anonymous.The privacy pattern sanitization method can avoid generating new inference channels as a result of its transformation. In this way, we can obtain a k-anonymous version of F(D,σ).

6. Conclusion

Along with privacy, discrimination is a very important issue when considering the legal and ethical aspects of data mining. Explore the relationship between discrimination prevention and privacy preservation in data mining considering alternative data anonymization techniques, other than those studied in this paper. In this paper, we have discussed different methods for discrimination, conclusions without sufficient knowledge of the institutional conditions for employments and thereby the danger of drawing inferences on the wrong basis. The purpose of this paper is to how we can preserve privacy for the different discrimination for different social aspects.

7. References

1. 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.
2. D. Pedreschi, S. Ruggieri, and F. Turini, “Measuring Discrimina- tion in Socially-Sensitive Decision Records,” Proc. Ninth SIAM Data Mining Conf. (SDM ’09), pp. 581-592, 2009.
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. 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.
5. S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, “Discri- mination Prevention in Data Mining for Intrusion and Crime Detection,” Proc. IEEE Symp. Computational Intelligence in Cyber Security (CICS ’11), pp. 47-54, 2011.
6. T. Calders and S. Verwer, “Three Naive Bayes Approaches for Discrimination-Free Classification,” Data Mining and Knowledge Discovery, vol. 21, no. 2, pp. 277-292, 2010.
7. S. Ruggieri, D. Pedreschi, and F. Turini, “Data Mining for Discrimination Discovery,” ACM Trans. Knowledge Discovery from Data, vol. 4, no. 2, article 9, 2010.

8. 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.
9. United States Congress, US Equal Pay Act, eeoc.gov/epa/anniversary/epa-40.html, 1963.
10. F. Kamiran and T. Calders, "Classification without Discrimination," Proc. IEEE Second Int'l Conf. Computer, Control and Comm. (IC4 '09), 2009