Effect of Privacy Protection on Discrimination

Asmita Kashid
Post Graduate Student
Department of Computer Engineering, MIT
Pune, India

Vrushali Kulkarni
Head of Department
Department of Computer Engineering, MIT
Pune, India

Ruhi Patankar
Assistant Professor
Department of Computer Engineering, MIT
Pune, India

***Abstract*— Data is very important aspect in today's world. However mere data is not useful for a specific individual or organizations. Data mining is a technique which finds the knowledge hidden in a raw data. However there are two problematic aspects related to data mining: *potential privacy violation and potential discrimination.* According to civil and social rights law, discrimination means treating people unfairly or unequally only because they are the members of a particular category or a minority, without considering their individual merit. Data mining may lead to discriminatory decisions, if it uses a historical dataset which is biased towards a particular community, to extract classification or association rules. Discrimination Prevention Data Mining deals with discovering, preventing and measuring discrimination. Privacy means the right of a person to decide how to use her/his sensitive information. Privacy violation occurs if a person's sensitive information is displayed to an unauthorized entity as a result of data mining tasks. Privacy Preserving Data Miming provides methods and tools for publishing useful information while preserving data privacy. Recently, it is identified that these two fields are dependent on each other. It is important to bridge the gap between the individual researches in these two areas. In this paper, we are trying to identify the effect of privacy protection on discrimination. An architecture of proposed work is also specified. Also some future research ideas are specified.**

Keywords*— Discrimination discovery; discrimination prevention; data anonymization techniques; privacy preserving techniques

I. INTRODUCTION

Data is very important aspect in today's world. However mere data is not useful for a specific individual or organizations. It is necessary to find the knowledge hidden in a raw data. Data mining is a technique used to do this task. However there are two problematic aspects related to data mining: *potential privacy violation and potential discrimination.* These both can be thought of as side effects of the data mining tasks. Discrimination means treating people unequally just because they belong to a minority, without considering their individual merit. It is not a mere existence of statistical imbalance in the data, but a property of a decision that may lead to such an imbalance. Data mining may lead to discriminatory decisions, if it uses a historical dataset which is biased towards a particular community, to extract classification or association rules. Discrimination Prevention Data Mining (DPDM) deals with discovering, preventing and measuring discrimination. Privacy means a right of a person to decide how to use her/his sensitive information (e.g. salary). Privacy violation occurs when values of published sensitive attributes can be linked to specific individuals. It is an intentional or unintentional intrusion into personal data. Data mining faces the problem of privacy violation as a side effect of data mining tasks. Privacy Preserving Data Mining (PPDM) deals with developing techniques to modify the original data in some way, so that private data remain private even after data mining process. It protects identities of people under consideration. It deals with privacy attacks, privacy models and anonymization techniques.

Recently, it is identified that these two fields are dependent on each other. It is important to bridge the gap between the individual researches in these two areas. PPDM and DPDM can be combined for several reasons – 1) these two areas are dependent on each other. Hiding discriminatory attribute for privacy protection affects the discrimination caused. E.g. in case of employee hiring, it is interesting to investigate, what happens if the employer knows the race of job-seeking candidate and what happens if the race is unknown to the employer. 2) Both these areas have common challenges. E.g. trade-off occurs between achieved privacy and data utility loss. Trade-off also occurs between discrimination removal and data utility loss. 3) They have common methodological problems to be solved. E.g. privacy attacks may occur after releasing the data. Discrimination threats may occur after releasing the data. 4) As privacy preservation is a well explored area, many of the privacy preservation methods can be used for discrimination prevention.

Main aim of this paper is to analyze the effect of privacy protection on discrimination. The rest of the paper is organized as follows: section II shows the literature survey related to these two fields under the heading related work. Section III defines basic terminology in data mining, DPDM and PPDM. Section IV presents how privacy protection affects discrimination. Section V gives a brief overview of our proposed work. Section VI presents conclusions and future work.

II. RELATED WORK

The research of DPDM has been started in 2008[1]. Method for discrimination discovery is explained in [2]. There are three different approaches for discrimination prevention [3] – preprocessing, inprocessing and postprocessing. Preprocessing approach consists of creating methods to remove discrimination from the original dataset i.e. processing is done on the dataset and dataset is cleaned to remove discrimination. Inprocessing approach deals with modifying the standard data mining algorithms to remove discrimination. Here processing is not done on the datasets, however data mining algorithms are processed to incorporate discrimination removal. In postprocessing approach, neither dataset is changed nor are standard data mining algorithms changed. Changes are done on the final results of the data mining algorithms.

Research in DPDM deals with developing different discrimination prevention methods using any of the above three approaches. Methods for discrimination prevention using preprocessing approach are presented in [3] [4]. Discrimination prevention using decision tree techniques is shown in [5]. This uses both inprocessing and postprocessing approaches. Discrimination prevention using Naïve Bayes model is presented in [6]. Naïve Bayes model for discrimination prevention also uses both inprocessing and postprocessing approaches. Different metrics to measure amount of discrimination is given in [7].

Research of PPDM has been started long back since the year 2000 as shown in [8]. Different algorithms and techniques have been developed to preserve user's privacy. Generalization and suppression techniques are specified in [9]. The survey of different privacy preserving techniques is specified in [10]. The data anonymization technique called generalization replaces QI attribute values with a generalized version of them using the generalization taxonomy tree of QI attributes. Suppression consists in suppressing some values of the QI attributes for some (or all) records. Anatomy [11] does not modify the quasi-identifier or the sensitive attribute, but deassociates the relationship between the two. Precisely, the method releases the data on QID and the data on the sensitive attribute in two separate tables: a quasi-identifier table (QIT) contains the QID attributes, a sensitive table (ST) contains the sensitive attributes, and both QIT and ST have one common attribute, GroupID. Slicing [12] divides the data set vertically and horizontally and deassociates the relation between tuples in different columns.

Some research also exists to identify relation between PPDM and DPDM. The effect of data anonymization techniques (e.g. generalization and suppression) on anti-discrimination is given in [13]. The method to make data discrimination free using privacy preserving model (e.g. t-closeness) is depicted in [14]. The impact of knowledge publishing on anti-discrimination is shown in [15] [16]. Methods for discrimination discovery using privacy attack strategies are presented in [17].

III. BASIC TERMINOLOGIES

A. *Basic Definitions in Data Mining*

Some basic definitions [3] of data mining are mentioned below. There definitions are used as background knowledge for measuring and discovering discrimination:

- 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.
- The support of an item set, sup (X), is the fraction of records that contain the item set X. We say that a rule ⟶ X C is completely supported by a record if both X and C appear in the record.
- Confidence of a classification rule, conf (X ⟶ C), measures how often the class item C appears in records that contain X.
- Hence, if supp (X) > 0 then,

$$conf\ (X \to C) = \frac{supp\ (X, C)}{supp\ (X)} \qquad (1)$$

Support and confidence range over [0, 1].

- A frequent classification rule is a classification rule with support and confidence greater than respective specified lower bounds. Support is a measure of statistical significance, whereas confidence is a measure of the strength of the rule.

B. *Basic Definitions in DPDM*

Some definitions related to rule-based discovery and prevention of discrimination [2] [3] are mentioned below. They have significance throughout the discrimination discovery and prevention process.

- A data item is said to be potentially Discriminatory (PD) if it is decided as discriminatory according to laws and regulations.
- A classification rule X → C is potentially discriminatory (PD) when X = A, B with A ⊆ DIs, a nonempty discriminatory item set and B a nondiscriminatory item set.
- Let A, B → C be a PD classification rule extracted from DB with conf (B → C) > 0. The extended lift (elift) of the rule is,

$$elift(A, B \ \to C) = \frac{conf\ (A, B \ \to C)}{conf\ (B \to C)} \qquad (2)$$

- Let A, B → C be a PD classification rule extracted from DB with conf ($\tilde{A}$, B→ C) > 0. The selection lift (slift) of the rule is,

$$slift(A, B \to C) = \frac{conf(A, B \to C)}{conf(\tilde{A}, B \to C)} \quad (3)$$

  The *slift* is the ratio of the proportions of benefit denial, e.g. credit denial, between the protected and unprotected groups, e.g. women and men resp., in the given context, e.g. those who live in NYC.
- Let f be one of the measures i.e. elift or slift and α be a fixed threshold and let A be a PD itemset. A PD classification rule c = A, B → C is *α-protective* w.r.t. f if f (c) < α. Otherwise, c is *α-discriminatory*.
- Let DB (A_1, ...A_n) be a data table, DA a set of PD attributes associated with it, and f be one of the measures i.e. elift,, slift. DB is said to satisfy *α-protection* or to be *α-protective* w.r.t. DA and f if each PD frequent classification rule c = A, B → C extracted from DB is α-protective, where A is a PD itemset and B is a PND itemset.

C. *Basic Definitions in PPDM [10]*

- Explicit identifier is a set of attributes that explicitly/uniquely identifies record owners.
- Quasi_Identifier is a set of attributes that could potentially identifies record owners.
- Sensitive attributes contain sensitive person specific information such as disease, salary or disability status.
- Non-Sensitive attributes contain all the attributes which do not belong to other three categories.
- Data Anonymization is an approach of PPDP that hides the identity and/or sensitive data of record owners, assuming sensitive data must be retained for data analysis i.e. Hide sensitive data in such a way that they will be reverted back for analysis purpose.

TABLE 1. PRIVATE DATA TABLE WITH BIASED DECISION RECORDS

ID	Gender	Job	Age	Credit_approved
1	Male	Engineer	35	Yes
2	Male	Engineer	38	Yes
3	Male	Lawyer	38	No
4	Female	Writer	30	No
5	Male	Writer	30	Yes
6	Female	Dancer	31	No
7	Female	Dancer	32	Yes

IV. IMPACT ANALYSIS OF PRIVACY PROTECTION ON DISCRIMINATION

In this section, we will see how privacy protection can affect the discrimination caused by using an example. Consider TABLE 1, which represents raw customer credit data, where each record represents a customer's specific information [13].

Gender, Job, Age can be taken as Quasi_identifier attributes. Class attribute has two values Yes and No, to indicate whether a particular customer has received credit or not. Suppose Gender is taken as a discriminatory attribute.

Suppose α = 1.2 and slift is taken as a discriminatory measure. A frequent PD classification rule {Gender = Female} → Credit_approved = no is extracted from the table.

$$slift = \frac{2/3}{1/4} = 2.66 \quad (4)$$

The rule is α-discriminatory as slift > α.

If we apply data anonymization technique called slicing [12] on TABLE 1, then TABLE 1 will be transformed to TABLE 2 as below:

TABLE 2. TRANSFORMATION OF TABLE 1 AFTER APPLYING SLICING

ID	Gender	Job	Age	Credit_approved
1	Male	Engineer	35	Yes
2	Male	Engineer	38	Yes
3	Male	Lawyer	38	No
4	Male	Writer	30	No
5	Female	Writer	30	Yes
6	Female	Dancer	31	No
7	Female	Dancer	32	Yes

Slift of the same rule will become:

$$slift = \frac{1\,/\,3}{2\,/\,4} = 0.666 \quad (5)$$

The rule has become α-protective as slift < α, after applying slicing technique on the table. This proves that data anonymization methods can achieve α-protection.

V. PROPOSED WORK

In our proposed work, we are going to use the same concepts as discussed in section IV. The problem statement is, to analyze effect of different privacy preserving (data anonymization) techniques on discrimination prevention.

We are trying to compare the percentage of discrimination removal (α-protection achieved) by different data anonymization techniques. There are many data anonymization techniques used in the PPDP, such as generalization [9], suppression [9], permutation [10], slicing [12], bucketization [11] anatomy [11], etc. As we have seen in section IV, privacy protection and anti-discrimination are dependent on each other. So it is important to find impact of privacy protection on anti-discrimination. Although full-domain generalization technique is used to make data privacy protected and discrimination prevention, there is still a scope to test impact of other data anonymization techniques on discrimination. So we are planning to do comparative study of different data anonymization techniques. Scope of our proposed work is limited to direct discrimination and use of preprocessing approach. Proposed architecture is depicted in Fig 1.

Input to our proposed system will be discrimination threshold, discriminatory dataset, discriminatory attribute, sensitive attribute, quasi identifier attribute, and data anonymization method.

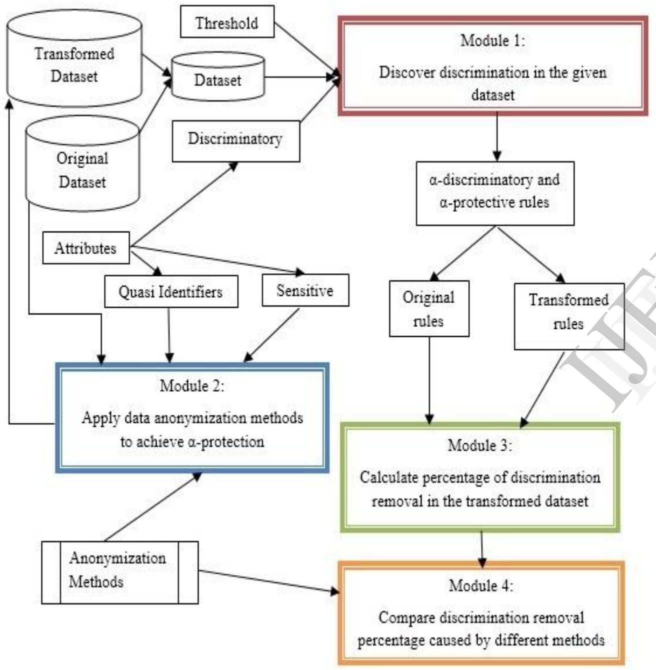
Transformed Dataset
Threshold
Dataset
Module 1:
Discover discrimination in the given dataset
Original Dataset
Discriminatory
α-discriminatory and α-protective rules
Attributes
Quasi Identifiers
Sensitive
Original rules
Transformed rules
Module 2:
Apply data anonymization methods to achieve α-protection
Module 3:
Calculate percentage of discrimination removal in the transformed dataset
Anonymization Methods
Module 4:
Compare discrimination removal percentage caused by different methods

Fig. 1. Architecture of the Proposed Work

After taking input dataset from the user, the first step is to discover discrimination in the dataset [2]. Then apply the inputted data anonymization method on the input dataset. The dataset will get transformed to a new dataset. Again discover discrimination from the transformed dataset. Finally calculate percentage of discrimination removal [3]. For each of the inputted data anonymization method, the same process will be repeated. We can compare percentage of discrimination caused by different methods. Main aim is the impact analysis, which can be done by using n number of methods depending upon time constraints. Though number of discriminatory datasets can be inputted to system, for testing purpose, we are going to use a single dataset. We are going to use two datasets for testing purpose: Adult dataset [18] and German Credit dataset [19].

VI. CONCLUSIONS AND FUTURE WORK

Privacy preserving and anti-discrimination are dependent on each other. Different data anonymization techniques can have different impact on discrimination. Some techniques may increase discrimination, some may decrease it or some may not have any effect on discrimination. Hence it might help if we find relation between them. The knowledge of this relationship, can help in making the original data protected against both privacy and discrimination risks. It is also observed that we cannot protect original data against privacy attacks without taking into account anti-discrimination requirement. Our proposed system can work as a tool for analyzing effect of privacy preserving techniques on discrimination. Our proposed system will provide a proper methodology to analyze effect of privacy preserving techniques on discrimination. The proposed tool can be extended to other data anonymization techniques in the privacy literature. Our system will also give an idea about which data anonymization techniques are best suitable for discrimination removal. This will be promising step towards making data both privacy protected and discrimination free. Our proposed system is scalable system, where a new research in privacy preserving area can be combined easily with discrimination research.

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] 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.

[3] S. Hajian & J. Domingo-Ferrer, “A Methodology for Direct and Indirect Discrimination prevention in data mining,” IEEE transaction on knowledge & data engg. pp. 1445-1459, July 2013.

[4] F. Kamiran and T. Calders, “Data preprocessing techniques for classification without discrimination,” Springer, 2011.

[5] F. Kamiran, T. Calders, and M. Pechenizkiy, “Discrimination Aware Decision Tree Learning,” Proc. IEEE Int’l Conf. Data Mining (ICDM ’10), pp. 869-874, 2010.

[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] D.Pedreschi, S.Ruggieri and F.Turini, “Measuring Discrimination in Socially-Sensitive Decision Records,” Proc. Ninth SIAM Data Mining Conf. (SDM ’09), pp. 581-592, 2009.

[8] R.Agrawal and R.Srikant, “Privacy-preserving Data Mining”, In Proc. of the ACM SIGMOD, pp. 439-450,2000.

[9] L.Sweeney, “Achieving k-anonymity privacy protection using generalization and suppression”, 2002.

[10] B.C.M Fung, K. Wang, R. Chen, P.S.Yu, “Privacy-preserving data publishing: A survey of recent developments,” *ACM Comput. Surv.* 42(4), Article 14, 2010.

[11] Xiao, X., Tao, Y, “Anatomy: Simple and effective privacy preservation”, In proc of VLDB, pp 139-150, 2006.

[12] Tiancheng Li, Ninghui Li, Jian Zhang, and Ian Molloy, “Slicing: A new approach for privacy preserving data publishing”, IEEE transactions on knowledge and data engineering, Vol. 24, no.3. 2012.

[13] S Hajian and J. Domingo-Ferrer, “A Study on the Impact of Data Anonymization on Anti-Discrimination,” Proc. I.EEE 12th International Conference on Data Mining Workshops, pp. 352-359, 2012.

[14] S.Ruggieri, “Data Anonymity Meets Non-Discrimination,” IEEE 13th International Conference on Data Mining Workshops (ICDMW), pp. 875-882, 2013.

[15] S. Hajian, A. Monreale, D. Pedreschi, J. Domingo-Ferrer and F. Ginnotti, “Injecting Discrimination and Privacy Awareness into Pattern Discovery,” Proc. IEEE 12th International Conference on Data Mining Workshops, pp. 360-369, 2012.

[16] S. Hajian, A. Monreale, D. Pedreschi, J. Domingo-Ferrer and F. Ginnotti, “Fair Pattern Discovery,” Proc. 29th Annual ACM Symposium on Applied Computing, pp. 113-120, 2014.

[17] S.Ruggieri, S. Hajian, F. Kamiran, and X. Zhang, “Anti-discrimination Analysis using Attack Strategies,” 25th European Conference on Machine Learning and 18th Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD-2014), to appear.

[18] R. Kohavi and B.Becker, “UCI Repository of Machine Learning Databases,” http://archive.ics.uci.edu/ml/datasets/Adult,1996.

[19] D.J. Newman, S.Hettich, C.L. Blake, and C.J. Merz, “UCI Repository of Machine Learning Databases,” http://archive.ics.uci.edu/ml, 1998.