International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 10, October- 2014

A Probabilistic Approach to Study Antidiscrimination in Data Mining

Mrunal Pendharkar
Department of Information Technology,
RMD Sinhgad school of Engineering
Pune, India

Prof. Sweta Kale
Department of Information Technology,
RMD Sinhgad school of Engineering
Pune, India

***Abstract*— Data Mining is a technology of extracting useful information which is hidden in huge databases. As data is in huge amount, discrimination in data based on caste, sex, was observed to distinguish data. Hence, antidiscrimination techniques including discrimination discovery and prevention were introduced in data mining. The purpose of the paper is to tackle the problems of discrimination and security and to propose new techniques in antidiscrimination to neglect the two kinds: direct and indirect discrimination in data mining. Also to clean training data sets from direct and indirect discrimination.**

***Keywords*— Data Mining, Discrimination, Antidiscrimination**

I.INTRODUCTION

In Sociology, discrimination is an action that denies social participation or human rights to categories of people based on prejudice. It involves discrimination on the basis of sex, caste, religion etc. Discrimination is categorised as either direct or indirect (also called systematic). Direct discrimination consists of procedures that directly mention minority or distinguished groups considering some discriminatory attributes. For example, opening in company only for female candidates criteria generate discrimination in sex. Indirect discrimination consists of procedures which do not directly mention minority, but intentionally or unintentionally generate discrimination decisions. For example, using background knowledge of zip codes and discriminating black people just to cut them off without letting people know publicly is indirect discrimination. To overcome discrimination in data mining, three approaches for discrimination prevention are considerable:

- Pre-processing: Transform the source data in such a way that discrimination attributes are removed and no unfair decisions can be taken, using standard data mining algorithms. The approaches are discussed in detail in [2] [3].
- In-Processing: Changing the data mining algorithm in such a way that the results have unbiased decisions. The approach is discussed in [4] where the non-discriminatory constraint is embedded in decision tree by pruning strategies and splitting criterion using a novel leaf relabeling approach.
- Post processing: Changing the data mining models results, instead of cleaning the original data sets or data mining algorithms [5].

The Data Mining process depicted using Figure 1 shows the steps involved during the process of mining [6].

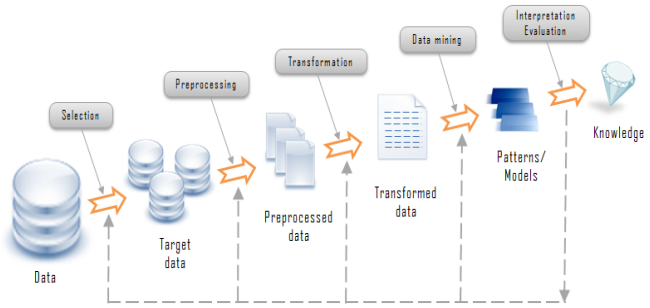

Fig.1 Steps involved in Data Mining process

The discrimination prevention process often starts with the process of collecting large amount of data. Then the collected data is analyzed and some discriminatory decision rules are extracted using standard data mining techniques. Using these discriminatory rules and already existing list of discriminatory attributes, a discriminatory threshold is decided. The process of anti-discrimination includes measurement of the amount of discrimination present and transformation of source data by pruning discrimination biases using data mining algorithms. Using these transformation methods, a transformed data set is obtained which is non-discriminatory.

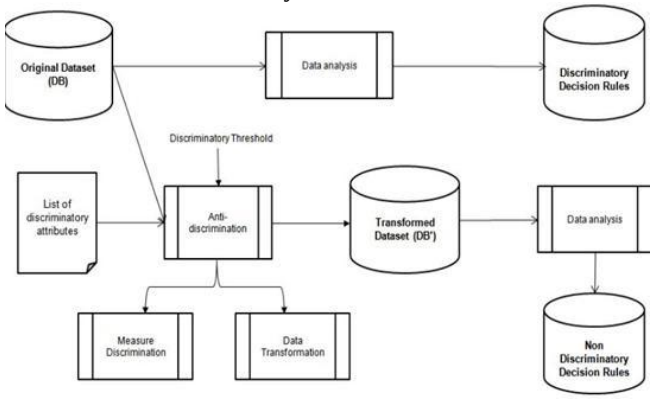

Figure 2. The process of anti-discrimination

The above Figure 2 shows the process of antidiscrimination [7]. The transformed data set may have some attributes which can be further used as non-discriminatory decision rules. Hence, they are extracted for research work. Efficiency and security of data are still an

issue in anti-discrimination approach. Some more approaches are proposed which can clean data sets as well as maintain the quality of data in data sets.

II. METHODS OF DISCRIMINATION PREVENTION

Discrimination can be either direct or indirect. Direct discrimination happens when decisions are made based on sensitive attributes. Indirect discrimination happens when decisions are made based on non-sensitive attributes which are strongly correlated with biased sensitive ones. The proposed system has discrimination prevention methods in terms of data quality and discrimination removal for both direct and indirect discrimination.

- Direct Discrimination Prevention Module
- Indirect Discrimination Prevention Module
- Rule Protection in Data Mining Module
- Rule Generalization in Data Mining Module

A. Direct Discrimination Prevention Module

The general idea of Discrimination measurement is identifying discriminatory and non-discriminatory rules which are useful for antidiscrimination process. Let DB be the original database which contains all non-discriminated data that is source data. Based on existing discriminatory sets in DB, frequent classification rules are categorised into two groups as Potentially Discriminatory and Potentially Non-discriminatory rules. Direct discrimination is calculated using Potentially Discriminatory rules and discriminatory threshold. Indirect discrimination is calculated using Potentially Non-Discriminatory rules and discriminatory threshold. There will be a database which will be collecting all direct discriminatory rules and one for indirect discriminatory rules.

Direct discrimination happens when decisions are made based on perceptive attributes. It consists of rules or measures that directly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group bias. To avoid direct discrimination is based on the fact that the data set of decision rules would be free of direct discrimination if it only contained Potentially Discriminatory rules that are protective or are instances of at least one non-redlining Potentially Non-Discriminatory rule. Here, direct rule protection and direct rule generalization are applied [1].

B. Indirect Discrimination Prevention Module

The process includes transforming of original data present in DB so that all direct and indirect discriminatory biases are removed, with less impact on the data and on decision rules to avoid unfair decision rules to be mined in transformed data set. In the following sections, data transformation for direct discrimination and data transformation of indirect discrimination is discussed.

Indirect discrimination occurs when decisions are made based on non-sensitive attributes which are strongly correlated with biased perceptive ones. It consists of rules or measures that, while not directly mentioning discriminatory attributes, intentionally or unintentionally could produce discriminatory decisions. To avoid indirect discrimination, it is based on the fact that the data set of decision rules would be free of indirect discrimination if it contained no redlining rules. To accomplish this, an appropriate data transformation with minimum information loss should be applied in such a way that redlining rules are converted into non-redlining rules. To overcome this, indirect rule protection and indirect rule generalization are applied [1].

C. Rule Protection in Data Mining Module

To convert discriminatory rule to protective rule, two methods are applied. One method changes discriminatory item set in some records like gender is changed from male to female with granted credits. Another method changes the class item in some records like changing grant credit to deny credit for male gender. The data transformation is based on direct rule protection and indirect rule protection. The classification rules do not guide themselves by personal decisions. However, one may realize that classification rules are actually learned by the system from the training data. If the training data sets are naturally biased for or against a particular membership, the learned model may show a discriminatory biased nature. In other words, the system may gather that just being foreign is a legitimate reason for loan denial.

Like in direct rule protection, there are two methods in indirect rule protection. One method changes discriminatory item set in some records like non-foreign worker to foreign worker. Another method changes the class item in some records like Hire which says no will change to hire which says yes.

D. Rule Generalization in Data Mining Module

Rule generalization is data transformation method in discrimination prevention for both direct and indirect discrimination. It is based on the fact that if each discriminatory rule has at least one non-redlining Potentially Non-discriminatory rule, the data is free from direct discrimination. In rule generalization, instead of discrimination measures, the relation between rules is considered.

The data transformation is based on direct rule generalization and indirect rule generalization. Assume that a complainant claims discrimination against foreign workers among applicants for a job position. In other words, foreign candidates are rejected because of their low experience in the field, not just because of their community. The general rule rejecting low-experienced candidates is a lawful one, because experience can be considered a lawful requirement for some jobs. As rule generalization might not work for all discriminatory rules, it is combined with rule protection for accurate results. There are various algorithms proposed for discrimination prevention which are applicable for various categories in direct and indirect discrimination.

III. CONCLUSION

Data mining concept is extracting useful knowledge from huge data. As data collected is huge, it is very difficult to identify which data is correct and which is useful. There are some attributes which distinguish data in different categories which can be intentional or unintentional but known to users.

This process is called Discrimination. To avoid discrimination in data mining, there are various approaches which are proposed for discrimination prevention. The above methods discussed for anti-discrimination helps to clean data sets as well maintain the quality of data. Privacy is also an important content in data mining. The purpose of paper was to create antidiscrimination methods and analyze them which include data transformation methods that can avoid direct, indirect discrimination or both. To gain this objective, it is necessary to measure the extent of discrimination present in the database and calculate a threshold so that it can be used to prevent discrimination. Last but not the least, some methods are proposed to find relationship between discrimination prevention and privacy preservation in data mining.

REFERENCES

[1] Sara Hajian and Josep Domingo-Ferrer, Fellow, IEEE, "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining." IEEE Transactions on Knowledge and Engineering, Vol. 25, No. 7, July 2013

[2] F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling," Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010

[3] F. Kamiran and T. Calders, "Classification with no Discrimination," Proc. IEEE Second Int'l Conf. Computer, Control and Comm. (IC4 '09), 2009

[4] T. Calders and S. Verwer, "Three Naïve Bayes Approaches for Discrimination-Free Classification," Data Mining and Knowledge Discovery, vol. 21, no. 2, pp. 277-292, 2010

[5] D. Prdreschi,S. Ruggieri, and F. Turini, "Discrimination-Aware Data Mining," Proc. 14th ACM Int'l Conf. Knowledge Discovery and Data Mining (KDD '08), pp. 581-592, 2009.

[6] Jiawei Han, Micheline Kamber, Data Mining: Concept and Techniques, Second Edition, University of Illinois at Urbana-Champaign, Morgan Kaufmann Publishers, 2006.

[7] Jiawei Han, Micheline Kamber, Data Mining: Concept and Techniques, Third Edition, University of Illinois at Urbana-Champaign, Morgan Kaufmann Publishers, 2006.