Data mining for Discrimination Prevention

Supriya M.Manglekar, V. K. Bhusari
Department of Computer Engineering, Pune University,
Pune, India

Abstract— For privacy, discrimination is a very important issue when considering the legal and ethical aspects of data mining. It is more than obvious that most people do not want to be discriminated because of their gender, religion, nationality, age and so on, especially when those attributes are used for making decisions about them like giving them a job, loan, insurance, etc. Discovering such potential biases and eliminating them from the training data without harming their decision-making utility is therefore highly desirable. For this reason, antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. Discrimination prevention consists of inducing patterns that do not lead to discriminatory decisions even if the original training datasets are inherently biased. In this chapter, by focusing on the discrimination prevention, we present a taxonomy for classifying and examining discrimination
prevention methods. Then, we introduce a group of pre-processing discrimination prevention methods and specify the different features of each approach and how these approaches deal with direct or indirect discrimination. A presentation of metrics used to evaluate the performance of those approaches is also given. Finally, we conclude our study by enumerating interesting future directions in this research body.

Keywords: discrimination, antidiscrimination

I. INTRODUCTION

In sociology, discrimination is the unfavourable treatment of an individual based on their membership in a certain group or category. It involves denying to members of one group opportunities that are available to other groups. There is a list of antidiscrimination acts, which are laws designed to prevent discrimination on the basis of a number of attributes (e.g., race, religion, gender, nationality, disability, marital status, and age) in various settings (e.g., employment, credit and insurance, etc.). Although there are some laws against discrimination, all of them are reactive, not proactive. Services in the information society allow for automatic and routine collection of large amounts of data. Those data are often used to train association/classification rules in view of making automated decisions, like loan granting/denial, insurance premium computation, personnel selection, etc. At first sight, automating decisions may give a sense of fairness: classification rules do not guide themselves by personal preferences. However, at a closer look, one realizes that classification rules are actually learned by the system (e.g., loan granting) from the training data. If the training data are inherently biased for or against a particular community (e.g., foreigners), the learned model may show a discriminatory prejudiced behavior. In other words, the system may infer that just being foreign is a legitimate reason for loan denial. Discovering such potential biases and eliminating them from the training data without harming their decision making utility is therefore highly desirable.

   Discrimination can be either direct or indirect. Direct discrimination. Indirect discrimination consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, intentionally or unintentionally could generate discriminatory decisions. Indirect discrimination could happen because of the availability of some background knowledge , for example, a certain zip code corresponds to a deteriorating area or an area with mostly black population.
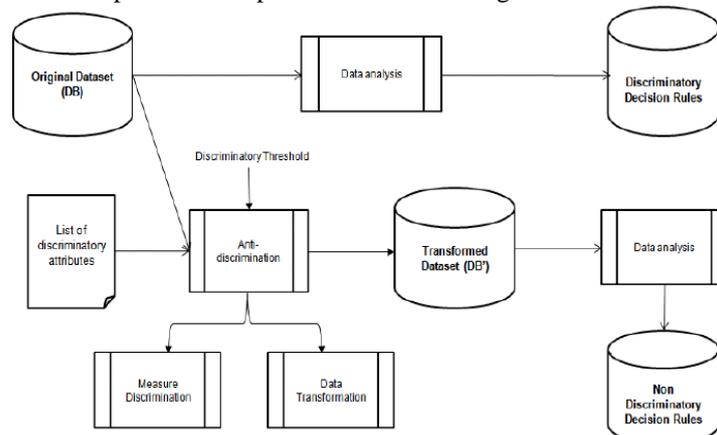

Fig. 13.1. The process of extracting biased and unbiased decision rules.

Figure 13.1 illustrates the process of discriminatory and non-discriminatory decision rule extraction. If the original biased dataset *DB* is used for data analysis without any anti-discrimination process (i.e. discrimination discovery and prevention),the discriminatory rules extracted could lead to automated unfair decisions. On the contrary, *DB* can go through an anti-discrimination process so that the learned rules are free of discrimination, given a list of discriminatory attributes (*e.g.* gender, race, age, etc.). As a result, fair and legitimate automated decisions are enabled

II. EXISTING SYSTEM

In Existing system the initial idea of using rule protection and rule generalization for direct discrimination prevention, but we gave no experimental results. We introduced the use of rule protection in a different way for indirect discrimination prevention and we gave some preliminary experimental results. In this paper, we present a unified approach to direct and indirect discriminationprevention, with finalized algorithms and all possible data transformation methods based on rule protection and/ or rule generalization that could be applied for direct or indirect discrimination prevention. We specify the different features of each method. Since methods in our earlier papers could only deal with either direct or indirect discrimination, the methods described in this paper are new.

DISADVANTAGES OF EXISTING SYSTEM:

- Profiling based on broad categories of race,ethnicity or religion may have several disadvantages.
- From the point of view in law enforcement greatest difficulty relates to the strain it can place on relations with minority communities.
- This, in turn, can undermine effective policing methods that rely on public cooperation, and it can also generate resentment among the communities concerned

III. PROPOSED SYSTEM

We propose new utility measures to evaluate the different proposed discrimination prevention methods in terms of data quality and discrimination removal for both direct and indirect discrimination. Based on the proposed measures, we present extensive experimental results for two well known data sets and compare the different possible methods for direct or indirect discrimination prevention to find out which methods could be more successful in terms of low information loss and high discrimination removal. The approach is based on mining classification rules (the inductive part) and reasoning on them (the deductive part) on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination.

ADVANTAGES OF PROPOSED SYSTEM:

- Privacy-invasion
- Potential-discrimination
- Discrimination-Measurement
- Data Transformation

IV. CONCLUSION

Compare the different possible methods for direct or indirect discrimination prevention to find out which methods could be more successful in terms of low information loss and high discrimination removal. The approach is based on mining classification rules (the inductive part) and reasoning on them (the deductive part) on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination. The purpose of this paper was to develop a new preprocessing discrimination prevention methodology including different data transformation methods that can preventdirect discrimination, indirect discrimination or both of them at the same time. To attain 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 biases. Finally, discrimination free data models can be produced from the transformed data set 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. We want to explore the relationship between discrimination prevention and privacy preservation in data mining.

ACKNOWLEDGEMENT

REFERENCES

[1]   S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "Discrimination Prevention in Data Mining for Intrusion and Crime Detection," Proc. IEEE Symp. Computational Intelligence in Cyber Security (CICS '11), pp. 47-54, 2011.

[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.

[4]   R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," Proc. 20th Int'l Conf. Very Large Data Bases, pp. 487-499, 1994

[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