A METHODOLOGY FOR DIRECT AND INDIRECT DISCRIMINATION PREVENTION IN DATA MINING

Mrs. K. Mohana, *1, Mrs. K. Sumathi, *2, C. Senthilraja *2

*1 M.Sc., M.Ed., Research Scholar, Dept. of CS, Siri Psg Arts & Science College for Women, Salem, Tamilnadu, India
*2 M.Sc., M. Phil., Asst.Professor/Computer Science, Siri Psg Arts & Science College for Women, Salem, Tamilnadu, India
*3 M.C.A., M.Phil., HOD/Computer Science, Siri Psg Arts & Science College for Women, Salem, Tamilnadu, India

Abstract—Security and privacy methods are used to protect the data values. Private data values are secured with confidentiality and integrity methods. Privacy model hides the individual identity over the public data values. Sensitive attributes are protected using anonymity methods. Discrimination is the prejudicial treatment of an individual based on their membership in a certain group or category. Antidiscrimination acts are designed to prevent discrimination on the basis of a number of attributes in various settings. Public data collections are used to train association/classification rules in view of making automated decisions. Data mining can be both a source of discrimination and a means for discovering discrimination. Automated data collection and data mining techniques such as classification rule mining are used to make automated decisions. Discriminations are divided into two types such as direct and indirect discriminations. Direct discrimination occurs when decisions are made based on sensitive attributes. Indirect discrimination occurs when decisions are made based on non sensitive attributes which are strongly correlated with biased sensitive ones. Discrimination discovery and prevention are used for anti-discrimination requirements.

Key terms: Index Terms—Antidiscrimination, data mining, direct and indirect discrimination prevention

I. INTRODUCTION

In sociology, discrimination is the prejudicial 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 and training, access to public services, credit and insurance, etc.). For example, the European Union implements the principle of equal treatment between men and women in the access to and supply of goods and services in or in matters of employment and occupation. Although there are some laws against discrimination, all of them are reactive, not proactive. Technology can add proactively to legislation by contributing discrimination discovery and prevention techniques.

Basic definition

Some basic definitions related to data mining. After that, we elaborate on measuring and discovering discrimination.

- A data set is a collection of data (records) and their attributes. Let DB be the original data set.
- An item is an attribute along with its value, e.g., Race = black.

II. RELATED WORK

Despite the wide deployment of information systems based on data mining technology in decision making, the issue of antidiscrimination in data mining did not receive much attention until 2008. Some proposals are oriented to the discovery and measure of discrimination. Others deal with the prevention of discrimination. The discovery of discriminatory decisions was first proposed by Pedreschi et al.,

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.

Statistical significance of the extracted patterns of discrimination in and to reason about affirmative action and favoritism. Moreover it has been implemented as an Oracle-based tool in. Current discrimination discovery methods consider each rule individually for measuring discrimination without considering other rules or the relation between them.
Three approaches are conceivable: preprocessing, in processing and post-processing approaches.

III. EXISTING 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. Some proposals are used to the discovery and measure of discrimination. But others deal with the prevention of discrimination. The discrimination discovery decision was first proposed by Pedreschi et al. This 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. In the US Equal Pay Act states that: “the selection rate for any ethnic group, race, or sex which is less than four-fifths of the rate for the group with the highest rate will generally be regarded as evidence of opposed effect.” It has been implemented as an Oracle-based tool.

In the current discrimination discovery methods consider each rule individually for measuring discrimination without considering other rules or the relation between them. In this paper we also take into account the relation between rules for discrimination discovery, based 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 original training data sets are biased. Three approaches are used:

Pre Processing

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 pre processing approaches of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature.

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 approach to cleaning the discrimination from the original data set is proposed in whereby the nondiscriminatory constraint is embedded into a decision tree learner by changing its splitting criterion and pruning strategy through a novel leaf relabeling approach.

1. Direct Discrimination Measure

Pedreschi et al. translated the qualitative statements in existing laws, regulations, and legal cases into quantitative formal counterparts over classification rules and they introduced a family of measures of the degree of discrimination of a PD rule. One of these measures is the extended lift (e lift).

Direct discrimination consists of rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. Discriminatory (sensitive) attributes like gender, race, religion, etc.,

2. Indirect Discrimination Measure

The purpose of indirect discrimination discovery is to identify redlining rules. In fact, redlining rules indicate Biased rules that are indirectly inferred from nondiscriminatory items (e.g., Zip = 10451) because of their correlation with discriminatory ones.

\[
\begin{align*}
\text{Pre Processing} & \\
\text{In Processing} & \\
\text{Indirect discrimination measure} & \\
\end{align*}
\]

Indirect discrimination consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, intentionally or unintentionally could generate discriminatory decisions. Redlining by financial institutions (refusing to grant mortgages or insurances in urban areas they consider as deteriorating) is
an archetypal example of indirect discrimination, although certainly not the only one.

\[ DDPD = \frac{|MR| - |MR'|}{|MR|} \]

Where \( MR \) is the database of \( \alpha \)-discriminatory rules from \( DB \) and \( MR' \) is the database of \( \alpha \)-discriminatory rules extracted from the transformed data set \( DB' \).

- Direct discrimination protection preservation (DDPP). This measure quantifies the percentage of the \( \alpha \)-protective rules in the original data set that remain \( \alpha \)-protective in the transformed data set. It is defined as

\[ DDPP = \frac{|PR| \cap |PR'|}{|PR|} \]

Where \( PR \) is the database of \( \alpha \)-protective rules extracted from the original data set \( DB \) and \( PR' \) is the database of \( \alpha \)-protective rules extracted from the transformed data set \( DB' \).

- Indirect discrimination prevention degree (IDPD).

- This measure quantifies the percentage of redlining rules that are no longer redlining in the transformed data set. Indirect discrimination protection preservation (IDPP).

- This measure quantifies the percentage of non-redlining rules in the original data set that remain non-redlining in the transformed data set.
IV. CONCLUSIONS AND FUTURE WORK

Along with 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.

The purpose of this paper was to develop a new preprocessing discrimination prevention methodology including different data transformation methods that can prevent direct 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.

REFERENCES


