Prevention Methods for Discrimination in Data Mining

K. Madhavi[1], Krishna Naik Mudavath[2]

[1]Assistant Professor, Department of IT, V R Siddhartha Engineering College

[2]Student, M. Tech CST, V R Siddartha Engineering College

Abstract: *Discrimination is a very important issue when considering as law and professional aspects of data mining. Data mining is an increasing crucial technology for extracting useful information hidden in large collections of data and negative social perceptions in data mining. This is providing along with privacy and security of the information and potential discrimination. The latter consists of unequal treating people depends on their belonging to a specific group of membership. Data mining methods used for data collection automating and data mining techniques that is classification rule mining have to cover the way to providing decisions automatically like loan granting or reject, insurance premium computation, etc. If the training data sets are cover in what regards discriminatory (sensitive) attributes like gender, race, religion, nationality etc. Discriminatory decisions may ensue because antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. Discrimination can be providing either direct or indirect. Direct discrimination takes as when decisions are made depends on sensitive attributes. Indirect discrimination takes sa when decisions are made depends on nonsensitive attributes which are strongly relation with biased sensitive ones. In this paper handled by discrimination prevention in data mining and propose new techniques applicable for direct or indirect discrimination prevention individually or both at the same time and how to clean training data sets are handle and outsourced data sets in used to cover way to that direct and/or indirect discriminatory decision rules are converted to perfect (nondiscriminatory) classification rules and providing loan is process is speed .the important advantages for manager how because he is easily check details and grating loan. The practical result demonstrate that the proposed methods are effective at removing direct and/or indirect discrimination biases in the original data set with preserving data quality.*

Keywords: Antidiscrimination, Rule protection, rule generalization,, Privacy direct and indirect discrimination prevention.

1. Introduction

Discrimination is the prejudicial treatment of an individual based on their membership in a certain group or category In sociology. It involves hiding to members of one group opportunities that are provides to other groups. In this approach list of antidiscrimination acts are used, which are rules designed to prevent discrimination on the basis of a number of attributes example like race, religion, gender, nationality, disability, marital status, and age in various attributes (e.g., employment and training, access to public services, credit and insurance, etc.). For example, the European Union implements the laws of equal treatment between men and women in the access to and supply of goods and services in [3] or in matters of employment and occupation in [4]. Although there are some rules against discrimination, all of them are reactive, not proactive. Technology can add proactivity to perfection by contributing discrimination discovery and prevention techniques.

The information society allow for automatic and routine collection of large amounts of data and services. Those data are used to train association/classification rules in represent of making automated decisions, like loan denial/granting, insurance premium computation, personnel selection, etc. The automating decisions may give a sense of fairness: classification rules do not provide to them by personal preferences. It is closer look, one classify that classification rules are actually provided by the system (e.g. loan granting) from the training data. If the training data are inherently cover for or against a particular community (e.g., foreigners). The learned model may show a discriminatory unequal behavior. In define other hand, the system may infer that just being foreign is a perfect reason for loan reject.

Finding such potential biases and eliminating them from the training data without change their decision- making use is highly desirable. One must prevent data mining from becoming itself a source of discrimination, because for data mining tasks getting discriminatory models from biased data sets as part of the automated decision making. In [12]. that data mining can be both a source of discrimination and a means for discovering discrimination can be introduce.

Discrimination can be classified into direct and indirect (also called systematic). Direct discrimination consists of rules or procedures that provide internally mention minority or disadvantaged groups based on sensitive discriminatory data sets related to group membership. Indirect discrimination existing of rules or procedures that provide while not internally view the discriminatory datasets. Which is intentionally or un-intentionally could deriving discriminatory decisions? Financial institutions are redlining (refusing to grant mortgages or insurances in urban areas they consider as deteriorating) is an example of indirect discrimination, although certainly not the only one. the sake of compactness for a slight abuse of language. In this project indirect discrimination will also be referred to as sequence and rules due to indirect discrimination will be known as redlining rules [12]. Indirect discrimination could happen for the availability of some background information (rules), for example for that that a certain zip code according to a deteriorating area or an area with mostly black population. The background knowledge might be taken from publicly available data (e.g., census data) or might be get from the original data set itself because of the consisting of anti discriminatory attributes that are highly related with the sensitive ones in the real data set.

Paper ID: OCT141091
918

Discrimination is a very important issue when considering the legal and ethical aspects of data mining Along with privacy. It is more than obvious that most people do not want to be discriminated because of their gender, religion, nationality, age and so on, separately when those attributes are used for making decisions about them like giving them a insurance, job, loan, job, 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, At present a taxonomy for classifying and examining discrimination prevention approaches. Then data mining introduce a group of pre-processing discrimination prevention methods and view the different features of each method and how these approaches deal with direct or indirect discrimination. A presentation of laws used to solve the performance of those approaches is also given. Finally, conclude our study by revolting interesting future directions in this research about project.

2. Literature Survey

Discrimination prevention has been recognized as some problem in a tutorial by (Clifton, 2003) [1] where the danger of construct classifiers capable of racial discrimination in home loans has been put provide. Data mining and machine learning models extracted from large data may find traditional unequal treatment of people. for example, mortgage redlining can be easily find as a common pattern in loan data but so solution was provided in this tutorial. Data mining techniques used for automating decision taken. The data mining techniques such as classification and association rules when used for decision making tasks such as benefit or credit approval, find that results are in discriminatory in work of behavior. This inadequacy of classification and association rules poses ethical and general or law issues, as well as impediment to hypothesis application. D. Pedreschi, S. Ruggieri, and F. Turini [2] have viewed the first kind of project, which address the discrimination issue in data mining models in 2008. They have investigated how discrimination may be hidden in data mining models and also measured the discrimination through a generalization and protection of lift. They have introduced α protection as a measure of the discrimination power of a classification rule containing one or more discriminatory items.

The data mining techniques such as classification and association rules used to taking decision tasks such as benefit or credit provide, find that results are in discriminatory in nature. This deficiency of classification and association rules poses ethical and legal issues, as well as obstacles to practical application. D. Pedreschi, S. Ruggieri, and F. Turini [2] have represented the first kind of papers, which details the discrimination issues in data mining models in 2008. They have underseek how expert's hypothesis. They have also implemented LP2DD [6] approach by integrating generation and deduction for finding

evidence of discrimination of the overall reference model. They had discussed integrating generation, through data mining classification rule extraction, and derivation, through a calculation logic implementation of the analytical tools in 2009.

3. Project Plan

Discrimination prevention methods dependent on preprocessing published so far [7], [8] present some limitations of prevention, which we next highlight to major limitations

They attempt to detect discrimination in the original data only for one discriminatory item and based on a single measure and modify. This approach cannot guarantee that the transformed data set is really discrimination free, because it is known that discriminatory behaviors can often be hidden behind several discriminatory items, and even behind combinations of them. They only consider direct discrimination. They do not include any measure to evaluate how much discrimination has been removed and how much information loss has been incurred.

In this project propose to be preprocessing methods which overcome the above limitations. The new data mining used for data transformation methods for protections (i.e., rule protection and rule generalization (RG)) are based on measures for both direct and indirect discrimination and can deal with several discriminatory items and in this project to provide utility measures. After, approach to discrimination prevention is broader than in previous work of the project.

In this is paper work [5], to be introduced the initial idea of using rule protection and rule generalization for direct discrimination prevention. but we gave no experimental results and no practical. In [6], they are introduced the use of rule protection in a different way for indirect discrimination prevention and in the project gave to some preliminary experimental results. In this paper, present a unified approach to direct and indirect discrimination prevention, 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. The special specify the different features of each technique. Since methods to be in this project [5], [6] could only deal with either direct or indirect discrimination; the methods described in this paper are new and provide successful result gave in experiment.

As part of this effort, gave to have developed metrics that specify which records should be changed, how many records should be changed, and how those records should be changed during data transformation. In addition, have to 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, in this present extensive experimental results for two well known data sets and compare the different possible methods and techniques for direct or indirect discrimination prevention to seek out which methods could be more successful in terms or data sets of low information loss and

high discrimination removal and low time taken for loan take out. In this project introduced to the successful result.

The rest of this paper is organized as follows. Section 2 introduces some basic definitions and concepts that are used throughout the project. Section 3 describes our proposal for direct and indirect discrimination prevention. Section 4 shows the tests to the performed to assess the validity and quality of our proposal and compare different methods. Finally, Section 5 summarizes conclusions and identifies future research topics in the field of discrimination prevention.

4. Prevention Methods

4.1 Methods For Direct And Indirect Discrimination Prevention

Unequal and Unfairly treating people on the basis of their belonging to a specific category, namely race, nationality ideology, gender, etc., is known as discrimination. In law of economics and social sciences. Discrimination has been studied cross the last decades and antidiscrimination rules have been adopted by many democratic governments. Some examples are the US Employment Non-Discrimination Act (United States Congress 1994), the UK Sex Discrimination Act (Parliament of the United Kingdom 1975) and the UK Race Relations Act (Parliament of the United Kingdom 1976).There are different type of decision-making tasks which lend themselves to discrimination, example loan providing, education, health insurances and staff selection. In many cases, decision-making tasks are supported by information systems. Given a set of information items on a perfect customer and it is very useful like an automated system decides whether the customer is to be recommended for a credit or a certain type of life insurance. Automating decisions reduces the workload of the staff of banks and insurance companies' staff and various organizations.

The use of information systems based on data mining approaches for decision making has attracted the attention of several researchers in the field of computer science. In this reason, automated data collection and a number of data mining techniques such as classification /association rule mining have been designed and are currently widely used for making automated decisions. The automating decisions may give a sense of fairness: classification rules (decision rules) do not rule themselves by personal preferences. However, at present a related view, one realizes that classification rules are generally learned by the system depends on training data set. If the training data are design biased for or against a particular community (for example, foreigners) ,the learned model may view a discriminatory unequal behavior. For example, in a certain loan providing companies, non residence people might systematically have been reject access to loans throughout the years. If this trained dataset is used as training data to learn classification rules for an automated loan providing system, the learned rules will also show cover behavior toward foreign people. In other hand the system may involve that just being foreign is a perfect reason for loan reject.

4.2 Classification of Discrimination Prevention Methods

The discrimination discovery, preventing information-based decision support systems from making discriminatory decisions (discrimination prevention) and it is a more challenging issue. In this project have to prevent not only direct discrimination but also indirect discrimination or both at the same time to the challenge increases. In this paper present taxonomy of discrimination prevention methods after having viewed a collection of independent works in the place. Figure 13.2 views these classifications.

In order to be able to classify the various approaches, these approaches consider two orthogonal dimensions based on which represent the existing approaches. As a 1 Note that α is a fixed threshold stating an acceptable level of discrimination according to laws and regulations.
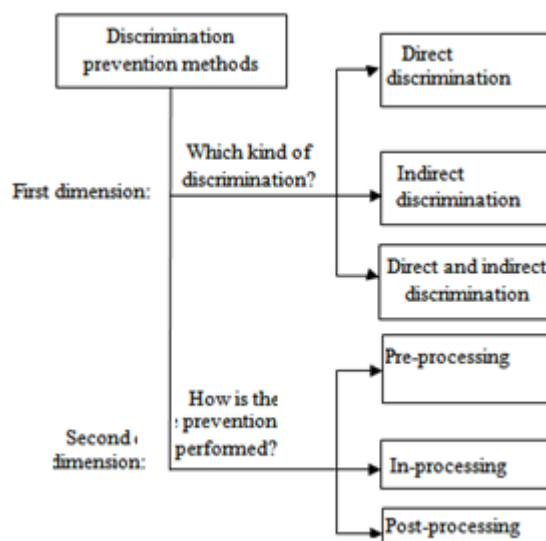

Figure 13.2: The taxonomy of discrimination prevention methods

First dimension, they have consider whether the approach handle with direct discrimination, and indirect discrimination, or both at the same time and several the discrimination prevention approaches into three clusters: *direct discrimination prevention approach* , *indirect discrimination prevention approach*, and *direct and indirect discrimination prevention approaches are used*. The second dimension in the classification relates to the phase of the data mining process in which discrimination prevention is done. Based on this second rule is discrimination prevention methods fall into three groups (Ruggeri et al. 2010).

5. Problem Definition

Number of Modules careful analysis the system has been identified to have the following modules:

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

5.1 Direct Discrimination Prevention Approach

Direct discrimination considers when decisions are made based on precise datasets. It should to be rules or procedures that internally display minority or disadvantaged groups depend on personal discriminatory attributes related to set off group collection. To keep direct discrimination is based on the exact values that the data set of decision laws would be free of direct discrimination if it has contained PD rules that are protective or are instances of at least one nonredlining PND rule. In this project apply direct rule protection and direct rule generalization.

5.2 Indirect Discrimination Prevention approach

Indirect discrimination internally when decisions are made based on nonsensitive attributes which are strongly related with cover sensitive ones. It consists of rules or process that, while not internally view discriminatory attributes, intentionally or unintentionally could evaluate discriminatory decisions for grating to loan. To keep or stopped indirect discrimination is based on the original that the data set of decision rules would be free of indirect discrimination. It contained no redlining rules. To reach this, a suitable data transformation with minimum information loss should be applied in such a way that redlining rules are converted to nonredlining rules. To overcome this approach applies indirect rule protection and indirect rule generalization. Rule Protection in Data Mining approaches.

5.3 Rule Protection in Data Mining approach

The data transformation is depends on direct rule protection and indirect rule protection. Classification rules do not handle themselves by sensitive activities. 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 cover for or against a particular community (e.g., foreigners), the learned model may show a discriminatory unequal behavior. In other hand, the system may infer that just being foreign is a legitimate reason for loan denial.

5.3 Rule Generalization In Data Mining approach

The data transformation is based on direct rule generalization and indirect rule generalization. In rule generating, the relation consider between rules instead of discrimination metric or perfect datasets. if it as complainant claims discrimination against foreign workers among customers for a job position. In other hand, foreign workers are rejected because of their have low experience and they are foreign so it will be rejected. The general rule rejecting low-experienced applicants are a perfect one, because they are foreign legitimate has experience. Here experience can be considered a genuine/legitimate requirement for some jobs.

6. Proposed System

The propose system new utility measures to solve the several proposed discrimination prevention methods in terms of data quality and discrimination removal for both direct and indirect discrimination. Based on the proposed measures, In this project represent introduce extensive practical results for two well known data sets and compare the several 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 methods are depends on mining classification rules and causes on them on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination.

7. Discussion and Future Work

Since most of effective decision models of data mining are constructed on the basis of historical decision records e.g., in credit scoring procedures and in credit card fraud detection systems, there is no guarantee that the extracted knowledge does not incur discrimination. This may be because the data from which the knowledge is extracted contain patterns with discriminatory bias. Hence, data mining from historical data may lead to the discovery of traditional prejudices. Thus prevention of discrimination knowledge based decision support systems; discovery is a more challenging issue. Some of the proposed techniques have been revived and based on the above review this is common that to some extent accuracy must be traded-off for lowering the discrimination. This trade-confirmed theoretically. This trade-off was studied and some of the future works in the area of discrimination prevention in data mining are to extend the discrimination prevention techniques to:

1) A multiple class problem by simply assuming one class as the desired class value and the rest of the class values as the not-desired category and vice versa.
2) To measure and evaluate how much discrimination has been removed by the above techniques

8. Conclusion

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, Those attributes are used for making decisions about them like giving them a job, loan, insurance, etc.

In this paper, we are displayed the existing work on discrimination discovery and prevention techniques in data mining and found that discrimination prevention in data mining is extremely difficult and challenging. Since most of us do not want to be discriminated based on our gender, religion, nationality, age and so on. Discrimination of any form must be detected and removed to get the unbiased results.

References

[1] 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.

[2] 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.

[3] European Commission, "EU Directive 2004/113/EC on Anti-Discrimination," http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:373:0037:0043:EN:PDF, 2004.

[4] European Commission, "EU Directive 2006/54/EC on Anti-Discrimination," http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:204:0023:0036:en:PDF, 2006.

[5] C. Clifton. Privacy preserving data mining: How do we mine data when we aren't allowed to see it? In Proc. of the ACMSIGKDD Int. Conf.on Knowledge Discovery and Data Mining (KDD 2003), Tutorial, Washington, DC (USA), 2003.

[6] D. Pedreschi, S. Ruggieri, and F. Turini,Discrimination-aware data mining. In Y. Li,B. Liu, and S. Sarawagi, editors, Proc. of the ACM SIGKDD Int. Conf. on Knowledge

[7] S. Ruggieri, D. Pedreschi, and F. Turini, Ifinntedginragt ienvgid einncdeu octfi odnis crainmdi nadtieodnu,c Atirotnif icfioarl Intelligence and Law, 18(1):1 − 43, 2010.

[8] B Luong, S Ruggieri, F Turini, K-nn as an implementation of situation testing for discrimination discovery and prevention, Technical Report TR-11-04: (2011), Dipartimento di Informatica, Universita di Pisa.

[9] F. Kamiran and T. Calders, Data preprocessing techniques for classification (w2i0th1o2u) t3 3:d1i −sc 3r3im. ination. Knowl Inf Syst

[10] S. Hajian and J. Domingo-Ferrer, A methodology for direct and indirect discrimination prevention in data mining. IEEE Transactions on Knowledge and Data Engineering, page to appear, 2012.

[11] A. Romei, S. Ruggieri, A multidisciplinary survey on discrimination analysis, The Knowledge Engineering Review, 2013.