Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 4, April 2015, pg.823 – 828

RESEARCH ARTICLE

Direct and Indirect Discrimination Prevention in Data Mining

Nickesh Rochlani[1], Prof. A.D.Chokhat[2]

[1]Department of CSE, P.R.Patil COET, Amravati, SGBAU, Amravati, Maharashtra, India

[2]Department of CSE, P.R.Patil COET, Amravati, SGBAU, Amravati, Maharashtra, India

[1] nicknrochlani@live.com; [2] amoldchokhat@gmail.com

***Abstract** Data mining is an increasingly important technology for extracting useful knowledge hidden in large collections of data. There are, however, negative social perceptions about data mining, among which potential privacy invasion and potential discrimination. The latter consists of unfairly treating people on the basis of their belonging to a specific group. Automated data collection and data mining techniques such as classification rule mining have paved the way to making automated decisions, like loan granting /denial, insurance premium computation, etc. If the training data sets are biased in what regards discriminatory (sensitive) attributes like gender, race, religion, etc., discriminatory decisions may result. For this reason, antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. Discrimination is a presuppose privileges where provide to the each separate group for the safety of the data which is stored. Discrimination can be either direct or indirect. 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. The propose work is to clean training data sets and outsourced data sets in such a way that direct and/or indirect discriminatory decision rules are converted to legitimate (non-discriminatory) classification rules.*

***Keywords:** Data mining, antidiscrimination, direct and indirect discrimination prevention.*

I. INTRODUCTION

Discrimination can be viewed as the act of unfairly treating people on the basis of their belonging to a specific group. 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:-

[1]Age
[2]Disability
[3]Race
[4]Religion
[5]Sex
[6]Sexual Orientation

Age discrimination is what occurs when an employer uses age as a determining factor in a job related decision. For example, age discrimination take place when an employer denies an applicant a job based on age. Religious discrimination involves the persecution or harassment of a person because of his or her religious beliefs or practices. Sexual Orientation refers to the treatment of individuals based on their sexual orientation and public or private institutions. Sex discrimination refers to differential treatment based on sex.

In economics and social sciences, discrimination has been studied for over half a century. There are several decision-making tasks which lend themselves to discrimination, 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 pro-activity to legislation by contributing discrimination discovery and prevention techniques.

Surprisingly, discrimination discovery in information processing did not receive much attention until 2008, even if the use of information systems in decision making is widely deployed. Indeed, decision models are created from real data (training data) in order to facilitate decisions in a variety of environments, such as medicine, banking or network security. 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 behaviour. 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. One must prevent data mining from becoming itself a source of discrimination, due to data mining tasks generating discriminatory models from biased data sets as part of the automated decision making. In [5], it is demonstrated that data mining can be both a source of discrimination and a means for discovering discrimination.

Discrimination can be either direct or indirect (also called systematic). Direct discrimination consists of rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. 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. With a slight abuse of language for the sake of compactness, in this propose work indirect discrimination will also be referred to as redlining and rules causing indirect discrimination will be called redlining rules[5]. Indirect discrimination could happen because of the availability of some background knowledge (rules), for example, that a certain zip code corresponds to a deteriorating area or an area with mostly black population. The background knowledge might be accessible from publicly available data (e.g., census data) or might be obtained from the original data set itself because of the existence of non-discriminatory attributes that are highly correlated with the sensitive ones in the original data set. Information technologies could play an important role in discrimination discovery and prevention. In this respect, several data mining techniques have been adapted with the purpose of detecting discriminatory decisions.

II. LITERATURE REVIEW

The discovery of discriminatory decisions was first proposed by Pedreschi et al. [5], [8]. 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. For instance, the US Equal Pay Act states that: “a selection rate for any race, sex, or ethnic group which is less than four-fifths of the rate for the group with the highest rate will generally be regarded as evidence of adverse impact.” This approach has been extended to encompass statistical significance of the extracted patterns of discrimination in and to reason about affirmative action and favouritism. Moreover it has been implemented as an Oracle-based tool in.

The existing literature on anti-discrimination in computer science mainly elaborates on data mining models and related techniques. Some proposals are oriented to the discovery and measure of discrimination. Others deal with the prevention of discrimination. The issue of antidiscrimination in data mining did not receive much attention until 2008 [8].Some proposals are oriented to the discovery and measure of discrimination. Others deal with the prevention of discrimination.

Three approaches are conceivable:

• Pre-processing: Transform the source data 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 [6], [7].

• In-processing: Change the data mining algorithms in such a way that the resulting models do not contain unfair decision rules. For example, an alternative approach to cleaning the discrimination from the original data set is proposed in [3].

• Post-processing**:** Modify the resulting data mining models, instead of cleaning the original data set or changing the data mining algorithms. For example, in [9], a confidence-altering approach is proposed for classification rules inferred by the CPAR algorithm. Removing sensitive attributes from data solve the direct discrimination problem but fail to solve indirect discrimination. As stated in [7] there may be other attributes that are highly correlated with the other sensitive one. In this paper, we concentrate on discrimination prevention based on pre-processing, because the pre-processing approach seems the most flexible one.

III. PROPOSED WORK

The propose work concentrate on discrimination prevention based on pre-processing, because the pre-processing approach seems the most flexible one: it does not require changing the standard data mining algorithms, unlike the in processing approach, and it allows data publishing (rather than just knowledge publishing), unlike the post processing approach. The propose work overcome the limitation based on pre-processing publish so far. In the propose work new data transformation methods are based on measures for both direct and indirect discrimination and can deal with several discriminatory items. This propose approach guarantee that the transformed data set is really discrimination free. It includes measure to evaluate how much discrimination has been removed and how much information loss has been incurred. Hence, the propose work approach to discrimination prevention is broader than in previous work. Propose work present a unified approach to direct and indirect discrimination prevention, with finalized algorithms and all possible data transformation methods that could be applied for direct or indirect discrimination prevention also specify the different features of each method. The propose methods developed new 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, 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, present extensive experimental results 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 proposed system includes use of NLP (Natural Language Processing). NLP provides means of analyzing text and its goal is to make computers analyze and understand the languages that humans use naturally. It can be seen as an Interaction between computers and humans. The method takes input as an original data set which contains discriminated attributes and by using NLP as shown in architecture the output of this processing will be discrimination free data set.

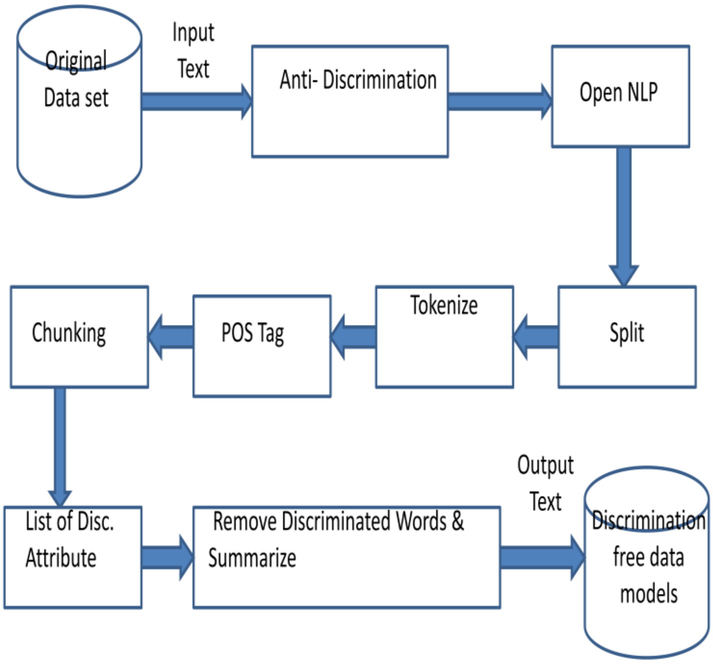

Fig.1: System Architecture

Our Proposed data transformation methods rule protection and rule generalization are based on measures for both direct and indirect discrimination and can deal with several discriminatory items. We demonstrate an integrated approaching to address and indirect discrimination prevention, on finalized algorithmic rule and all potential information shift ways confirmed rule protection and or convention generalization that could indirect discrimination prevention. We suggest fresh utility amounts to evaluate the different aimed favouritism prevention processes in terms by information quality and discrimination removal as some direct and indirect discrimination. Direct and indirect discrimination discovery includes identifying discriminatory rules and redlining rules.

Using the above transformation methods effectively to identify the categories and remove direct and indirect discrimination method. Finally, discrimination free data models can be produced from the transformed data set without seriously damaging data quality. Discrimination prevention techniques in conditions by information character and discrimination removal because some direct and indirect discrimination. The proposed techniques are quite successful in both goals of removing discrimination and preserving data quality.

IV. CONCLUSION AND FUTURE SCOPE

Along with privacy, discrimination is a very important issue when considering the legal and ethical aspects of data mining. The purpose of this paper was to develop a new pre-processing 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.

We plan to extend this in future as: Right now our definition of discrimination is quite brute force. No discrimination at all is allowed. We want to extend the notion of discrimination to that of conditional discrimination; e.g. instead of requiring that there is no discrimination at all, we could weaken this condition to no discrimination unless it can be explained by other attributes. Another extension we plan to consider are numerical attributes (e.g., income) as sensitive attribute. .We believe it might be the case that the maximum likelihood assignment does not correspond to a zero discrimination assignment. Investigating this behavior is left as future work.

REFERENCES

[1] S. Hajian, J. Domingo-Ferrer , "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining ," IEEE transactions on knowledge and data engineering, vol. 25, no. 7, july 2013

[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]A.Romei and S.Ruggieri. "A multidisciplinary survey on discrimination analysis". *The Knowledge Engineering Review*, Vol. 00:0, pp. 1–54, 2013.

[4]T. Kamishima, S. Akaho, H. Asoh, J. Sakuma, " Fairness-aware classifier with prejudice remover regularizer" *In ECML/PKDD, LNCS 7524*, pp. 35-50. Springer, 2012.

[5]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.

[6]F. Kamiran and T. Calders, "Classification without Discrimination," Proc. IEEE Second Int'l Conf. Computer, Control and Comm. (IC4 '09), 2009.

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

[8]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.

[9]. D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring Discriminationin Socially-Sensitive Decision Records," Proc. Ninth SIAMData Mining Conf. (SDM '09), pp. 581-592, 2009.

[10] P.N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Addison-Wesley, 2006.