ISSN: 2321-7782 (Online)
Volume 4, Issue 4, April 2016
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Various Methodologies in the discrimination of Data Mining

Ch.Vijaya kanaka Durga[1]
M.Tech Scholar
St. Mary's Group Of Institutions Guntur
Chebrole(V&M),Guntur(Dist), Andhra Pradesh – India

M. Ambarisha[2]
Assistant Professor
St. Mary's Group Of Institutions Guntur
Chebrole(V&M),Guntur(Dist), Andhra Pradesh – India

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. In the age of Database technologies a large amount of data is collected and analyzed by using data mining techniques. However, the main issue in data mining is potential privacy invasion and potential discrimination. One of the techniques used in data mining for making decision is classification. On the other hand, if the dataset is biased then the discriminatory decision may occur. If the training data sets are biased in what regards discriminatory (sensitive) attributes like gender, race, religion, etc., discriminatory decisions may ensue. For this reason, antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. Discrimination can be either direct or indirect[1]. 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 .In this paper 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.

Keywords: Antidiscrimination, data mining, direct and indirect discrimination prevention, rule protection, rule generalization, privacy policies.

I. INTRODUCTION

In data mining, discrimination is one of the issues discussed in the recent literature. Discrimination denies the members of one group with others. A law is designed to prevent discrimination in data mining. Discrimination can be done on attributes viz. religion, nationality, marital status and age. A large amount of data is collected by credit card companies, bank and insurance agencies. Thus, these collected data are auxiliary utilized by companies for decision making purpose in data mining techniques. The association and or classification rules can be used in making the decision for loan granting and insurance computation. Discrimination can be direct and indirect. 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[2]. 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.

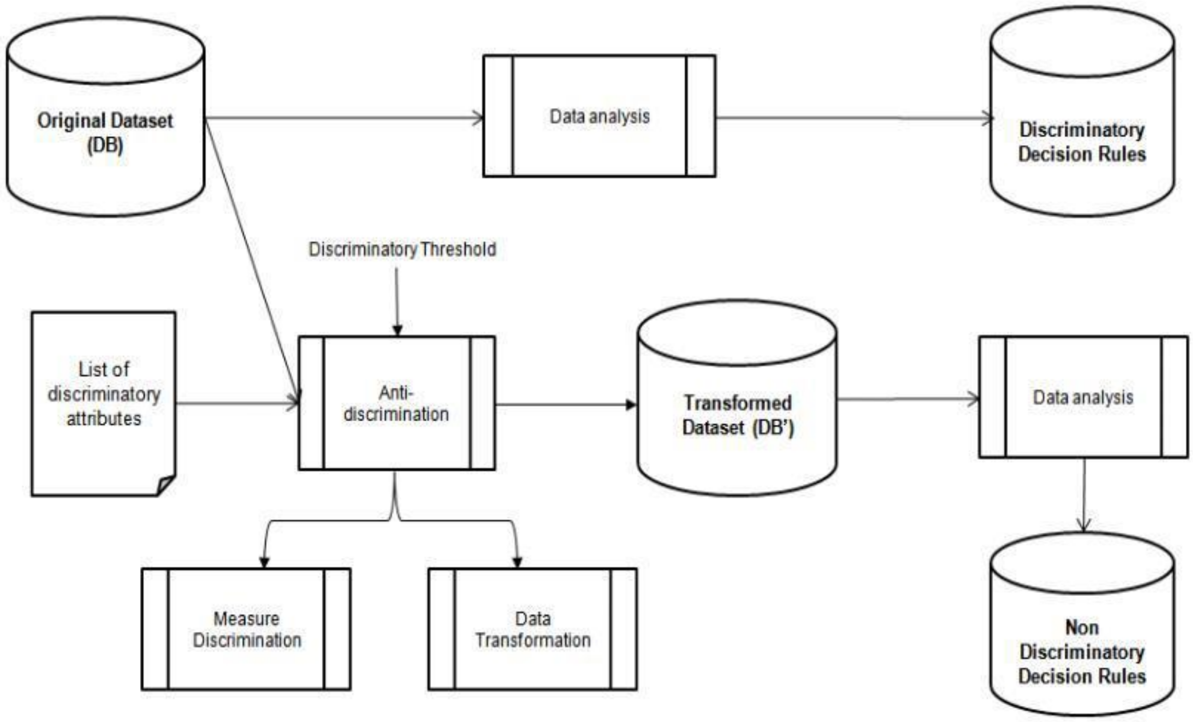

Fig: Architecture Diagram

Discrimination can be viewed as the act of unfairly treating people on the basis of their belonging to a specific group. For instance, individuals may be discriminated because of their race, ideology, gender, etc. 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, *e.g.* loan granting and staff selection. In the last decades, anti-discrimination laws have been adopted by many democratic governments. Some examples are the US Equal Pay Act [1], the UK Sex Discrimination Act [2], the UK Race Relations Act [3] and the EU Directive 2000/43/EC on Anti-discrimination [4]. Surprisingly, discrimination discovery in information processing did not receive much attention until 2008 [5], 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. Anti-discrimination also plays an important role in cyber security where computational intelligence technologies such as data mining may be used for different decision making scenarios. To the best of our knowledge, this is the first work that considers anti-discrimination for cyber security. Clearly, here the challenge is to avoid discrimination while maintaining data usefulness for cyber security applications relying on data mining, *e.g.* intrusion detection systems (IDS) or crime predictors.

II. Related Work

In this section, we discussed the state of the art approaches dealing with the antidiscrimination in data mining. However, we observe in recent literature, the issue of antidiscrimination is not attended by the several authors. R.Agrawal and R.Srikant [1] discussed the association rule method for the large database. Also they presented two algorithms that discover association between items in a large database of transactions. However, the performance gap is increases with the problem size. On the other side, they did not consider the quantities of the items bought in a transaction. T.Calders and S.Verwer [2] presented a modified Naive Bayes classification approach. In this, the author performs classification of the data in such a way that focuses on independent sensitive attribute. Such independency restrictions occur naturally when the decision process leading to the labels in the data-set was biased; e.g., due to gender or racial discrimination. This setting is motivated by many cases in which there exist laws that disallow a decision that is partly based on discrimination. This approach does not consider numerical attributes viz. Income as a sensitive attribute. F.Kamiran and T.Calders [3] proposed an approach which focuses on the concept of classification without discrimination. In this, the author introduced the idea of Classification with No Discrimination (CND). Thus, the author proposed a solution based on "massaging" the data to remove the discrimination from it with the least possible changes. On the other hand, the author also proposed a new solution to the CND problem.

III. ANALYSIS

During the investigation in the recent state-of-the art literature, we identified some of the issues.

First, the literature focus on the attempt to detect discrimination in the original data only for one discriminatory item and also based on a single measure.. Direct discrimination occurs when decisions are made based on sensitive attributes. It consists of rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. To prevent direct discrimination is based on the fact that the data set of decision rules would be free of direct discrimination if it only contained PD rules that are protective or are instances of at least one non redlining CND rule. In this we apply direct rule protection and direct rule generalization.

Second, it cannot guarantee that the transformed data set is really discrimination free. Indirect discrimination occurs when decisions are made based on no sensitive attributes which are strongly correlated with biased sensitive ones. It consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, 20 Intentionally or unintentionally could generate discriminatory decisions. To prevent indirect discrimination 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 achieve this, a suitable data transformation with minimum information loss should be applied in such a way that redlining rules are converted to no redlining rules. To overcome this we apply indirect rule protection and indirect rule generalization.

Third, the literature focuses on the direct discrimination. The data transformation is based on direct rule protection and indirect rule protection. 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.

Fourth, the state of the art approaches do not shows any measure to evaluate how much discrimination has been removed. Thus, the approaches did not concentrate on the amount of information loss generated. The data transformation is based on direct rule generalization and indirect rule generalization. In rule generalization, we consider the relation between rules instead of discrimination measures. Assume that a complainant claims discrimination against foreign workers among applicants for a job position. In other words, foreign workers are rejected because of their low experience, not just because they are foreign.

IV. CONTRIBUTION AND PLAN OF THIS PAPER

Discrimination prevention methods based on preprocessing published so far [7], [8] present some limitations, which we next highlight: They attempt to detect discrimination in the original data only for one discriminatory item and based on a single measure. 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. In this paper, we propose preprocessing methods which overcome the above limitations. Our new data transformation methods (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. Also, we provide utility measures. Hence, our approach to discrimination prevention is broader than in previous work. In our earlier work [5], we introduced the initial idea of using rule protection and rule generalization for direct discrimination prevention, but we gave no experimental results. In [6], 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 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. We specify the different features of each method. Since methods in our earlier papers [5], [6] could only deal with either direct or indirect discrimination, the methods described in this paper are new. As part of this effort, we 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, 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 rest of this paper is organized as follows. Section 2 introduces some basic definitions and concepts that are used throughout the paper. Section 3 describes our proposal for direct and indirect discrimination prevention. Section 4 shows the tests we have 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.

V. DISCUSSION

Although there are some works about antidiscrimination in the literature, in this paper we introduced anti-discrimination for cyber security applications based on data mining. Pedreschi *et al.* in [5], [7], [8], [9], [12] concentrated on discrimination discovery, by considering each rule individually for measuring discrimination without considering other rules or the relation between them. However in this work, we also take into account the PND rules and their relation with α-discriminatory rules in discrimination discovery. Then we propose a new preprocessing discrimination prevention method. Kamiran *et al.* in [11], [10] also proposed a preprocessing discrimination prevention method. However, their works try to detect discrimination in the original data for only one discriminatory item based on a simple measure and then they transform data to remove discrimination. Their approach cannot guarantee that the transformed dataset is really discrimination-free, because it is known that discriminatory behaviors can often be hidden behind several items, and even behind combinations of them. Our discrimination prevention method takes into account several items and their combinations; moreover, we propose some measures to evaluate the transformed data in degree of discrimination and information loss.

VI. CONCLUSION

In this paper, we discussed the issues and limitation of the recent state of the approaches. Based on the same issues, we study an approach that uses transformation method. This approach helps to prevent direct discrimination and indirect discrimination. However, the care has been taken for maintaining the data quality and privacy during the transformation. Thus, our future work is to implement a transformation method such that the data quality will not be disturbed. We also propose new metrics to evaluate the utility of the proposed approaches and we compare these approaches. The experimental evaluations demonstrate that the proposed techniques are effective at removing direct and/or indirect discrimination biases in the original data set while preserving data quality. Our contribution concentrates on producing training data which are free or nearly free from discrimination while preserving their usefulness to detect real intrusion or crime. In order to control discrimination in a dataset, a first step consists in discovering whether there exists discrimination. If any discrimination is found, the dataset will be modified until discrimination is brought below a certain threshold or is entirely eliminated. In the future, we want to run our method on real datasets, improve our methods and also consider background knowledge (indirect discrimination). Last but not least, we want to explore the relationship between discrimination prevention and privacy preservation in data mining. It would be extremely interesting to find synergies between rule hiding for privacy-preserving data mining and rule hiding for discrimination removal. Just as we were able to show that indirect discrimination removal can help direct discrimination removal, it remains to be seen whether privacy protection can help antidiscrimination or vice versa. The connection with current privacy models, like differential privacy, is also an intriguing research avenue.

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://eur-lex.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. D. Pedreschi, S. Ruggieri and F. Turini, "Discrimination-aware data mining". Proc. of the 14th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2008), pp. 560-568. ACM, 2008.
6. F. Kamiran and T. Calders, "Classification without discrimination".Proc. of the 2nd IEEE International Conference on Computer, Control and Communication (IC4 2009). IEEE, 2009.
7. S. Ruggieri, D. Pedreschi and F. Turini, "Data mining for discrimination discovery". ACM Transactions on Knowledge Discovery from Data, 4(2) Articles 9, ACM, 2010.
8. D. Pedreschi, S. Ruggieri and F. Turini, "Measuring discrimination in socially-sensitive decision records". Proc. of the 9th SIAM Data Mining Conference (SDM 2009), pp. 581-592. SIAM, 2009.
9. S. Ruggieri, D. Pedreschi and F. Turini, "DCUBE: Discrimination Discovery in Databases". Proc. of the ACM International Conference on Management of Data (SIGMOD 2010), pp. 1127- 1130. ACM, 2010.
10. R. Kohavi and B. Becker, "UCI Repository of Machine Learning Databases," http://archive.ics.uci.edu/ml/datasets/Adult, 1996.
11. D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, "UCI Repository of Machine Learning Databases," http://archive. ics.uci.edu/ml, 1998.
12. 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.