A Survey on Discrimination Avoidance in Data Mining

Malpani Radhika S, Dr.Sulochana Sonkamble

PG Scholar, Department of Computer Engineering, JSPM NTC, Savitribai Phule Pune University, Pune, India

Head & Professor, Department of Computer Engineering, JSPM NTC, Savitribai Phule Pune University, Pune, India

ABSTRACT: For extracting useful knowledge which is hidden in large set of data, Data mining is a very important technology. There are some negative perceptions about data mining. This perception may contain unfairly treating people who belongs to some specific group. Classification rule mining technique has covered the way for making automatic decisions like loan granting/denial and insurance premium computation etc. These are automated data collection and data mining techniques. According to discrimination attributes if training data sets are biases then discriminatory decisions may ensue. Thus in data mining antidiscrimination techniques with discrimination discovery and prevention are included. It can be direct or indirect. When decisions are made based on sensitive attributes that time the discrimination is indirect. When decisions are made based on nonsensitive attributes which are strongly correlated with biased sensitive ones that time the discrimination is indirect. The proposed system tries to tackle discrimination prevention in data mining. It proposes new improved techniques applicable for direct or indirect discrimination prevention individually or both at the same time. Discussions about how 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 classification rules are done. New metrics to evaluate the utility of the proposed approaches are proposes and comparison of these approaches is also done.

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

I. INTRODUCTION

In sociology, discrimination is considered as the harmful treatment of an individual based on their membership in a certain group or category [1]. It includes refusing to the members of one group opportunities which are available to other groups. There is a list of an antidiscrimination acts. These laws are designed to restrict discrimination on the basis of a number of attributes in different settings. The basic attributes can be race, religion, gender, nationality, disability, marital status, and age. Different settings can be employment and training, access to public services, credit and insurance. Even if there are some laws against discrimination, all are reactive, not proactive. Technology can add proactively to legislation by contributing discrimination discovery and prevention techniques.

Discrimination is nothing but counterproductive treatment of an individual. It is based on their membership in the certain group of their membership in a certain group or categories. It absorbs contradicting to members of one group opportunities. These opportunities are those which are available to another group. Some laws are designed to prevent the discrimination.

For large amount of data automatic and routine collection is allowed by the information society. In view of making automated decisions, this data is used to train association/classification rules. Automated decisions can be like loan granting/denial, insurance premium computation, personnel selection, etc. A sense of fairness is given by automating decisions at first sight: classification rules do not guide themselves by personal preferences. But in closer look, it is realized that the classification rules are in reality learned by the system from the training data. The learned model can show a discriminatory prejudiced behavior if the training data are biased for or against a particular community. That is nothing but system may conclude that reason for loan denial is just a being foreign. It is highly desirable that the discovering potential biases and eliminating them from the training data without harming their decision making utility. As the data mining tasks generating discriminatory models from biased data sets as part of the automated decision

making, we must prevent data mining from becoming itself a source of discrimination. It is experimentally proved in [2] that the data mining can be both a source of discrimination and a means for discovering discrimination.

II. RELATED WORK

Pedreschi et al. [2], [3] proposed discovery of discriminatory decisions at very first. This was based on mining classification rules (the inductive part) and reasoning on them (the deductive part). It uses quantitative measures of discrimination that formalize legal definitions of discrimination. Consider example of US Equal Pay Act. It says 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 is advanced to encompass statistical significance of the extracted patterns of discrimination in [4] and to reason about affirmative action and favoritism [5]. It is implemented as an Oracle-based tool in [6]. Presently available discrimination discovery methods consider each rule individually. Each rule is used for measuring discrimination without considering other rules or the relation between them. Discrimination prevention is the other big antidiscrimination goal in data mining. It consists of inducing patterns which do not lead to discriminatory decisions even if the original training data sets are biased. There are three approaches.

1. Preprocessing: 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. The preprocessing approaches of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature. Along this line, [7], [8] perform a controlled distortion of the training data from which a classifier is learned by making minimally intrusive modifications leading to an unbiased data set. The preprocessing approach is useful for applications in which a data set should be published and/or in which data mining needs to be performed also by external parties (and not just by the data holder).

2. In-processing: The data mining algorithm is updated in such way that the unfair decision rule does not contained by the resulting models. For example, for cleaning the discrimination from the original data set is proposed in [9], though which the nondiscriminatory restriction is embedded into a decision tree learner by altering its gashing criterion and pruning strategies during a novel leaf relabeling approaches. Though, it is accessible that in-processing discrimination avoidance techniques must wait on new particular purpose data mining algorithms.

3. Post-processing: Rather than cleaning the genuine data set or altering the data mining algorithm process of post-processing modified the resulting data mining models. For example in [10], a confidence altering approached is proposed for the classification rule inferred by the CPAR algorithm. The authority to publishing the data is not given by the post-processing. They published only the modified data mining models. Since the process of data mining can be performed by data holder only.

In Classification with no Discrimination by Preferential Sampling (F. Kamiran and T. Calders, 2010) [8] classification with No Discrimination by Preferential Sampling is an expected solution for the discrimination issue. It gives guaranteeing results with both steady and shaky classifiers. It decreases the security level by keeping up a high precision level. It gives comparable execution to "massaging" however without changing the dataset and dependably beats the "reweighing" plan.

In Integrating Induction and Deduction for Finding Evidence of Discrimination by Pedreschi et al. (2009) [5] presented a reference model for the examination and revelation of discrimination in socially-sensitive choices taken by DSS. The methodology comprises first of extracting frequent classification rules, and afterward of examining them on the premise of quantitative measures of discrimination and their measurable significance. The key legitimate ideas of protected-by-law groups, direct discrimination, indirect discrimination, honest to goodness occupational prerequisite, affirmative activities and partiality are formalized as explanations over the set of concentrated runs and, perhaps, extra foundation information.

In Data Mining for Discrimination Discovery by Ruggieri et al. (2010) [3] present the issue of finding discrimination through data mining in a dataset of recorded choice records, taken by people or via programmed frameworks. They formalize the techniques of direct and indirect discrimination revelation by displaying protected-by-

law groups and connections where discrimination happens in a classification based extraction. Essentially, classification rules extracted from the dataset permit for divulging connections of unlawful discrimination, where the level of load over protected-by-law groups is formalized by an augmentation of the lift measure of a classification rules.

In DCUBE: Discrimination Discovery in Databases by Turini et al. (2010) [6] says that DCUBE is an analytical tool supporting the interactive and iterative procedure of discrimination detection. The future users of DCUBE include: antidiscrimination establishment, proprietors of socially sensitive decision databases, and auditors, researchers in social sciences, economics and law.

A Survey of Association Rule Hiding Methods for Privacy (V. Verykios and A. Gkoulalas-Divanis, 2008) [11] presents classification and a survey of recent approaches that have been practical to the association rule hiding difficulty. Association rule hiding refers to the procedure of adapting the original database in such a way that convinced sensitive association rules vanish without gravely affecting the data and the non-sensitive rules.

In Rule Protection for Indirect Discrimination Prevention in Data Mining by Hajian et al. (2011) [10] displayed the first technique for avoiding indirect discrimination in data mining because of biased training datasets. There contribution in this paper focuses on delivering training data which are free or about free from indirect discrimination while safeguarding their helpfulness to data mining algorithms. So as to avoid indirect discrimination in a dataset, a first step comprises in finding whether there exists indirect discrimination. If any discrimination is found, the dataset is altered until discrimination is brought underneath a certain threshold or completely removed.

Discrimination Prevention in Data Mining for Intrusion and Crime Detection by Domingo-Ferrer et al. (2011) [12] analyzed how discrimination could affect on cyber security applications, particularly IDSs. IDSs use computational knowledge advances, for example, data mining. It is evident that the training data of these frameworks could be discriminatory, which would bring about them to settle on discriminatory decision when foreseeing interruption or, all the more for the most part, wrongdoing.

In Three Naive Bayes Approaches for Discrimination-Free Classification (T. Calders and S. Verwer, 2010) [13] studied three Bayesian methods for discrimination-aware classification. They are altering the experiential probabilities in a Naïve Bayes model in such a way that its forecasts become discrimination-free. Next method concerned learning two dissimilar models; they are S_0 and S_1, and complementary these models afterwards. The last and most concerned method they initiated a latent variable L reflecting the latent "true" class of an object without discrimination.

In Fast Algorithms for Mining Association Rules in Large Databases (R. Agrawal and R. Srikant, 1994) [1] exhibited two new algorithms, Apriori and Aprioritid, for finding all huge association rules between items in a vast database of transactions. We contrasted these algorithms with the awhile ago known algorithms, the AIS [4] and SETM [13] algorithms.

As concluded in EU Directive 2004/113/EC on Anti-Discrimination (2004) [14] there might be another attributes for example zip code which are highly associated with the sensitive ones and permit inferring discriminatory rules. Since the two most significant challenges concerning the discrimination prevention is; first challenge is instead of considering only direct discrimination we want to consider both direct and indirect discrimination. Second challenge is to discovering the good arrangement among the discrimination removal and the quality of the resulting training data sets and the model of data mining.

In Classification without Discrimination (F. Kamiran and T. Calders, 2009) [7] suggest the thought of discrimination is non minor and poses ethical furthermore legitimate issues and also hindrances in commonsense applications. CND furnishes us with a basic yet influential beginning stage for the arrangement of the discrimination issue. CND classifies the future information (both discriminatory and non discriminatory) with least discrimination and high exactness. It likewise addresses the issue of redlining.

Since some methods are already proposed for each approach we mention above, the concept of prevention of discrimination is the topic of research for the researchers.

III. COMPARATIVE ANALYSIS

Here comparative study to understand how different methods affects the discrimination in mining. New different researches go under for finding reliable result with no discrimination present in databases without data loss.

Table 1: Different Mining Techniques for Discrimination with Advantages and Disadvantages

Paper References	Methods /Techniques used	Advantages	Disadvantages
"Fast Algorithms for Mining Association Rules in Large Databases", in 1994 [1]	- Apriori, AprioriTid are designed - comparison between AIS and SETM algorithms with Apriori, AprioriTid	- Association rule method for the large database. - get large item sets to work.	- Performance gap is increases with the problemsize increasing.
"Classification with no Discrimination by Preferential Sampling", 2010 [8]	- it's for less intrusive technique to get unbiased data without changing class labels as efficient as result of massaging and reweighing by sampling algorithm.	- introduced the idea of Classification with No Discrimination (CND).	- did not consider such as do not proposing discrimination model which is used in many cases.
"Rule Protection for Indirect Discrimination Prevention in Data Mining", in 2011 [10]	- study to find combinations of attributes to get indirect discrimination with preprocessing with algorithm- 1,2 [10] - Prevention degree and protection preservation (DPD and DPP) used for results	- antidiscrimination in the context of cyber security. - proposed data transformation method for indiscrimination preventionand considered several discriminatory attributes and their combinations.	- does not run method on real dataset. - don't consider background knowledge.
"Three Naive Bayes Approaches for Discrimination-Free Classification", in 2010 [13]	- Naive Bayes approach with 3 methods (1)change probability like no discrimination in predictions (2) balance models with S = 0/ 1 (3) true class with no discrimination by Latent variable.	- presented a modified Naive Bayes classification approach for taking decisions easily.	- doesn't consider numerical attributes viz. Income as a sensitive attribute.
" EU Directive 2006/54/EC on Anti-Discrimination" in 2006[14]	- Social information case discrimination and for that antidiscrimination study.	- Presented the issue of discrimination in social sense viz. against minorities and disadvantaged groups.	- Not focused on the rule based framework process for direct and indirect discrimination.

IV. CONCLUSION AND FUTURE WORK

As discrimination is a very important issue of data mining. The purpose of this paper was to develop new preprocessing discrimination prevention including different data transformation methods that can prevent direct discrimination, indirect discrimination along with both at the same time. Also it consists of supporting discrimination discovery, namely the unveiling of discriminatory decisions hidden, either directly or indirectly, in a dataset of historical decision records, possibly built as the result of applying a classifier. As future work we are exploring measures of discrimination different from the ones considered in this paper along with privacy preservation in data mining. Further, implementation of discrimination prevention in Post-processing will be done. Proposed Algorithm achieves high accuracy and efficiency.

REFERENCES

1. Sara Hajian and Josep 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. D. Pedreschi, S. Ruggieri, and F. Turini, "Discrimination-AwareData Mining,"Proc. 14th ACM Int'l Conf. Knowledge Discovery and Data Mining (KDD '08),pp. 560-568, 2008.
3. 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.
4. D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring Discrimination in Socially-Sensitive Decision Records," Proc. Ninth SIAM Data Mining Conf. (SDM '09),pp. 581-592, 2009.
5. D. Pedreschi, S. Ruggieri, and F. Turini, "Integrating Induction and Deduction for Finding Evidence of Discrimination,"Proc. 12th ACM Int'l Conf. Artificial Intelligence and Law (ICAIL '09),pp. 157-166, 2009.
6. S. Ruggieri, D. Pedreschi, and F. Turini, "DCUBE: Discrimination Discovery in Databases,"Proc. ACM Int'l Conf. Management of Data (SIGMOD '10), pp. 1127-1130, 2010.
7. F. Kamiran and T. Calders, "Classification without Discrimination," Proc. IEEE Second Int'l Conf. Computer, Control and Comm. (IC4 '09), 2009.
8. F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling," Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.
9. 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.
10. S. Hajian, J. Domingo-Ferrer, and A. Martı́nez-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.
11. VERYKIOS AND A. GKOULALAS-DIVANIS, "A SURVEY OF ASSOCIATION RULE HIDING METHODS FOR PRIVACY," PRIVACY-PRESERVING DATA MINING: MODELS AND ALGORITHMS, C.C. AGGARWAL AND P.S. YU, EDS., SPRINGER, 2008.
12. S. Hajian, J. Domingo-Ferrer, and A. Martı́nez-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.
13. 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.
14. 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.

BIOGRAPHY

Malpani Radhika S. is a PG Student in the Computer Engineering Department, JSPM NTC RSSOER, Savitribai Phule Pune University. She received Bachelor of Information Technology (B.E) degree in 2012 from Savitribai Phule Pune University, Pune, MS, India. Her research interests are Data Mining

Dr. Sulochana Sonkamble is a Head & Professor in the Computer Engineering Department, JSPM NTC RSSOER, Savitribai Phule Pune University.