Review Paper on Prevention of Direct and Indirect Discrimination

Trupti N. Mahale
SPPU, PG Student,
Computer Engineering Department,
SITRC, Sandip Foundation Nashik.

Amol D. Potgantawar
SPPU, H.O.D,
Computer Engineering Department,
SITRC, Sandip Foundation Nashik.

ABSTRACT
Data mining is very important technology for extracting useful knowledge from large data. The discrimination is nothing but the unfair treatment given to an individual or group according to particular characteristics. For data mining classification rules are performing very important role but discrimination comes into picture because of biased classification rules. The training data sets are biased so we need to firstly discover discrimination and then need to prevent that discrimination to make it discrimination free. Discrimination can be of two types, direct and indirect. When decisions are made based on sensitive attributes, Direct Discrimination occurs. While decisions based on non-sensitive attributes, Indirect Discrimination occurs. The experimental evaluations demonstrate that the proposed techniques are effective at removing direct and/or indirect discrimination in the original data set while preserving data quality.

Keywords
Data mining, rule protection, rule generalization, antidiscrimination

1. INTRODUCTION
In general, discrimination is the unfair treatment of an individual based on their particular attribute. It means rejecting members of one particular group having that specific attribute (e.g. color, caste, nationality etc.). Different antidiscrimination acts are provided to avoid discrimination. For example, the European Union applies the principle of same treatment between men and women[3]. But almost all the laws provided are reactive. They should be proactive. Hence applying pro active laws to avoid discrimination. Classification rules are actually learned by the system (e.g., loan granting) from the training data. If the training data are biased then the learned model may show a discriminatory behavior. Hence to identify such biases and remove them to make data set discrimination free. Discrimination can be either direct or indirect (also called systematic). When decisions are made based on sensitive attributes, Direct Discrimination occurs. While decisions based on non-sensitive attributes, Indirect Discrimination occurs.

This work is explained in different chapters. Chapter 2 introduces Literature survey where previous work on Discrimination is given. Chapter 3 gives brief idea about Architecture. Chapter 4 summaries the concept under Analysis section. Finally chapter 5 Concludes the overall concept.

2. LITERATURE REVIEW
The two new algorithms were proposed to identify association rules to extract information from large database in 1994[2]. But data privacy was not supported by those algorithms. Hence new privacy preserving data mining algorithms were specified with the general survey of data mining models[3]. Eventually proposed a systematic framework for measuring discrimination to investigate whether evidence of discrimination can be found in given set of decisions[4]. The conceptual study regarding discrimination presence was done. Next to that to achieve classification with no discrimination is been achieved by introducing a sampling scheme for making data discrimination free instead of relabeling the dataset[5]. The discrimination-aware classification problem is illustrated and motivated. In this way discrimination free dataset concept came into picture. But simply removing the sensitive attribute from the training dataset does not solve the problem, due to the so called indirect discrimination rules[6]. Indirect discrimination rules are based on non-sensitive attributes. Though removal of sensitive attributes does not solve the problem of discrimination free dataset. A classification model based on direct as well as indirect discrimination is been introduced which is learnt on biased training data but works impartially for future data[7]. After study of discrimination introducing anti-discrimination in the context of cyber security; a new discrimination prevention method based on data transformation was introduced. That method considered several discriminatory attributes and their combinations to propose some measures for evaluating the proposed method in terms of its success in discrimination prevention and its impact on data quality[8]. To create discrimination free dataset its quality was key constraint to develop the method. Next to that introducing how to clean training datasets and outsourced datasets in such a way that legitimate classification rules can still be extracted but indirectly discriminating rules cannot[9]. In this way different approaches were designed for direct discrimination free model but was not for indirect discrimination free dataset. Biased decisions were introduced by Pedreschi et al.[15][17]. Considering all discriminatory rules, oracle based tool was implemented [18].Generally discrimination finding methods consider discriminatory rules individually. But this technique is considering all the relationships between discriminatory rules to make it effective.

To prevent discrimination, three approaches are there-

1) Preprocessing:- The dataset is been transformed to new dataset which is discrimination free i.e. as the name suggests, preprocessing is done on the dataset before applying any rule.[1].So that the results generated would be discrimination free.

2) In-processing:- the different algorithms which works on dataset needs to be changed so that the discriminatory dataset will give discrimination free results.[1]. Likewise , in-processing cannot use standard data mining algorithms as it is dependent on special featured algorithms to give discrimination free results.

3) Post-processing:- In spite of changing original data set or changing mining algorithms, modify the resulting mined data set [1].In post processing, modified mined data set can be shown publically not original resulting data set. Hence data holder can have rights to mine the data.

As per the general thinking, preprocessing approach looks suitable for removal of discriminatory attributes. It would solve direct discrimination problem but introduces large amount of data loss as well it can't solve indirect discrimination[15]. Hence, there are two main challenges in front of discrimination prevention: first one is to prevent direct as well as indirect discrimination prevention and second one is to maintain data quality with no data loss. Here concentrating on preprocessing approach discrimination prevention is achieved as it looks flexible one among all.

Discrimination prevention based on preprocessing having some limitations[5][7].

1) This approach can't guarantee that combination of transformed data sets are discrimination free because it only consider single rule to prevent discrimination instead of combinations of all rules.

2) It only consider direct discrimination not indirect one.

3) It doesn't give any measure to evaluate how much discrimination is been removed or how much data loss has been occurred.

As per this paper used preprocessing approach overcomes all the above mentioned limitations[1]. New data transformation methods are based on direct as well as indirect discrimination removal.

3. ARCHITECTURE ANALYSIS

New utility measure is defined to describe the different discrimination prevention methods in terms of maintaining data quality and removal of discrimination for both direct and indirect discrimination. To study this measure firstly need to study basic definitions regarding data mining concept:

1)Dataset is collection of records and their attributes.

2)An attribute with its value is called as an item, e.g. Nationality = Indian.

3)Collection of items with their values is considered as an item set i.e. A = {City = Mumbai, Gender = Male}.

4)A *support* of an item set is nothing but the fraction of records from item set A. It is given by the term *sup*(A).

5)*Cf*(A→B) is nothing but the confidence of classification rule, where A→B is nothing but the classification rule. It is also said that,

$$Cf(A \rightarrow B) = \frac{sup(A,B)}{sup(A)}$$

Depending on the given measures, showing experimental results for two well known data sets i.e. adult data set and German data set. Also comparing the different methods for direct and indirect discrimination prevention to find best method with respect to low information loss and high discrimination removal. The method is based on mining classification rules as per the basic definitions.

As shown in the Fig.1 the data transformation flow is mainly dependent on direct rule generalization and indirect rule generalization[3]. In rule generalization, it considers the relationship between classification rules in spite of different measures. Let's assume that a discrimination occurs if there are foreign workers applying for a job. Likewise it can be said that foreign workers are rejected because they are foreigner. But if instead of nationality, company can consider their experience. Experience is the valid reason or a good reason to give a job. Discrimination is of two types, direct and indirect[5]. When decisions are made based on sensitive attributes, Direct Discrimination occurs. While decisions based on non-sensitive attributes, Indirect Discrimination occurs.

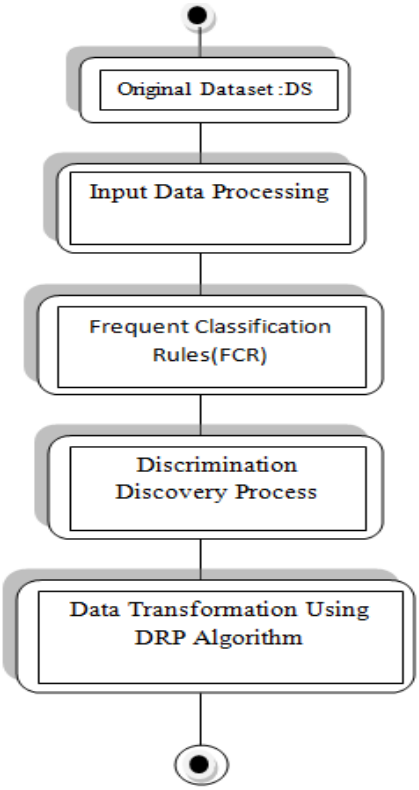

Fig.1: Process Flow of Discrimination Prevention

3.1 Algorithm and Process Flow

3.1.1 Input Data Preprocessing

As dataset may contains numerical values it needs to be processed before performing some functions on attributes.

3.1.2 Frequent Classification Rule extraction

Apriori algorithm generates frequent item sets. These generated item sets are needed to generate frequent classification rules[19].

3.1.3 Discrimination Discovery Process

The frequent classification rules are then categorized into Potentially Discriminatory and Potentially Nondiscriminatory groups in discrimination discovery[19].

3.1.4 Data Transformation using DRP Algorithm

As a result of data transformation, the transformed dataset is obtained as an output[16]. Data transformation is next step in discrimination prevention where the data is actually modified to make it biased free[18]. In this step modifications are done using the definition of elift i.e. equality constraint to satisfy the definition of corresponding discrimination prevention measure[19].

3.2 Modules Used

The given system is divided into following modules

3.2.1 Direct Discrimination Prevention

When decisions are based on sensitive attributes, Direct Discrimination occurs. Basically it combines rules and procedures that explain disadvantaged groups related to group membership[4]. Prevention of direct discrimination is based on the data set of classification rules that must be free of direct discrimination if it only contained PD rules that are protective or are instances of at least one non redlining PND rule. In this module, applying direct rule protection and direct rule generalization[14]. It is one of the basic module of given system.

3.2.2 Indirect Discrimination Prevention

When decisions are based on non sensitive attributes, indirect discrimination occurs i.e. biased sensitive attributes. Prevention of indirect discrimination is based on the data set of classification rules must be free of indirect discrimination i.e. contained no redlining rules[7]. To achieve this, transformation of suitable data with minimum information loss should be applied. So that redlining rules are converted to non redlining rules. To overcome this need to apply indirect rule protection and indirect rule generalization about data set and redundant data.

3.2.3 Rule Protection in Data Mining

The data transformation is based on the concept of direct rule protection as well as indirect rule protection. Classification rules doesn't support themselves by personal preferences[5]. However, one realizes that system actually learns classification rules (e.g., loan granting) from the training data. If the training data sets are biased against a particular community (e.g., foreigners), the learned model may show a discriminatory biased behavior.

3.2.4 Rule Generalization in Data Mining

The data transformation is mainly based on direct rule generalization and indirect rule generalization. In rule generalization, it considers the relationship between classification rules in spite of different measures[15]. Let's assume that a discrimination occurs if there are foreign workers applying for a job. Likewise it can be said that foreign workers are rejected because they are foreigner. But if instead of nationality, company can consider their experience. Experience is the valid reason or a good reason to give a job. 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[13]. Indirect discrimination consists of group of rules or procedures which are not explicitly mentioning discriminatory attributes, and could generate discriminatory decisions. Unfair treatment by financial institutions (refusing to grant insurances in urban areas they consider as not capable to refund) is an general example of indirect discrimination, although certainly not the only one. 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.

4. ANALYSIS

This section presents the analysis of the defined direct and/or indirect discrimination prevention approaches and algorithms. All algorithms and utility measures implemented here using the C programming language.

Table 1: Analysis of methods used in Discrimination Prevention

Method used	Advantages	Limitations
Frequent Classification Rule Extraction	Apriori algorithm is used to generate classification rule	Can use faster algorithm for classification
Discrimination Discovery Process	Rules are divided into PD and PND rules	Rules must be divided with preprocessing approach
Data Transformation	DRP algorithm is used for data transformation	Data Transformation might gives data loss.

5. CONCLUSION

It can be said that, discrimination is so important issue when considering the ethical points of data mining. The main aim was to develop a new pre processing discrimination prevention methods including modified data transformation methods which can prevent direct as well as indirect discrimination or both of them at the same time. To achieve this objective, the first step is to measure discrimination and define grouping of individuals that have been directly and/or indirectly discriminated in the decision-making processes; the second step is to transform the extracted 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.

Future work will include the exploration of relationship between prevention of discrimination and preservation of privacy in data mining.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge Computer Engineering department, SITRC and all the people who provided with the facilities being required and conductive conditions for completion of the review paper.

7. REFERENCES

[1] Sara Hajian and Josep Domingo-Ferrer "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining", Data Mining and Knowledge Discovery, vol. 25, no. 7, pp. 1445-1459, 2013

[2] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", Proc. 20th Intl Conf. Very Large Data Bases, pp. 487-499, 1994.

[3] V. 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, Springer, 2008.

[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] F. Kamiran and T. Calders, "Classification without Discrimination", Proc. IEEE Second Intl Conf. Computer, Control and Comm.(IC4 09), 2009.

[6] 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.

[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. Hajian, J. Domingo-Ferrer, and A. Martnez-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.

[9] S. Hajian, J. Domingo-Ferrer, and A. Martnez-Balleste, "Rule Protection for Indirect Discrimination Prevention in Data Mining", Proc. Eighth Intl Conf. Modeling Decisions for Artificial Intelligence (MDAI 11), pp. 211-222, 2011.

[10] 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.

[11] 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.

[12] F. Kamiran, T. Calders, and M. Pechenizkiy, "Discrimination Aware Decision Tree Learning," Proc. IEEE Int'l Conf. Data Mining (ICDM '10), pp. 869-874, 2010.

[13] R. Kohavi and B. Becker, "UCI Repository of Machine LearningDatabases,"http://archive.ics.uci.edu/ml/datasets/Adult, 1996.

[14] 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.

[15] 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.

[16] 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.

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

[18]United States Congress, US Equal Pay Act, http://archive.eeoc.gov/epa/anniversary/epa-40.html, 1963.

[19] D.P. Jagtap, "Classification with No Direct Discrimination", IJCAT, vol 3, no. 7, pp.464-467,2014