A Literature Survey on Discrimination Prevention in Data Mining

Shubhangi Gaikwad[1], Ranjeet Parihar[2], Vinayak Pottigar[3]
[1,2,3]*Department of Computer Science & Engg, Solapur University, Solapur*
SKN Sinhgad College of Engineering, Korti, Solapur, MS, India
[1]gaikwad.shubhangi28@gmail.com

Abstract - Data mining is most significant technology for retrieving convenient data from large amount of data. Privacy irruption and discrimination becomes major issues in data mining. Discrimination comes into picture when people are given unjust or prejudicial treatment based on their certain characteristics. There are two main types of discrimination which are direct discrimination and indirect discrimination. Direct discrimination is depend upon sensitive attribute like race, gender, religion, age sex, etc. Indirect discrimination depends upon non-sensitive attributes which are highly correlated to sensitive attributes. This paper reviews different papers which gives the different methods that are used for discrimination prevention.

Keywords - Data Mining, direct and indirect discrimination, discrimination prevention, Discrimination measure

I. INTRODUCTION

Data mining is the process of gathering, searching through, and analyzing a large collection of data in the database, as to find out pattern or relationship. With the time data mining becoming more and more popular technique. Privacy invasion and potential discrimination becomes two major concerns in data mining. While considering legal and ethical aspects of data mining, discrimination becomes main concern. Discrimination is nothing but prejudicial treatment given to person or group on the basis of their belonging. People does not get some opportunity because of their race, religion, gender, etc. and those attributes are used in some decision making system like granting the loan to them, giving job, etc. Classification rule has very important role but discrimination happens due to biased classification rules. If the training data set is biased then first task is to find out discrimination and then apply the methods to prevent that discrimination.

There are mainly two types of discrimination which are direct discrimination and indirect discrimination. Direct discrimination happen when decision making is based on the basis of sensitive attributes like age, sex, race, religion, etc. Indirect discrimination happens due to non-sensitive attributes which are closely related to sensitive attributes like date of birth. Date of birth is may not be sensitive attribute but by using that attribute we can calculate age of that person and if age is the sensitive attribute for that decision making system, then indirect discrimination will happen. Apart from discovery of discrimination, making knowledge-based decision support system free from making discriminatory decisions is become more challenging issue. The challenge is increases when there is need to prevent both direct as well as indirect discrimination. Every discrimination prevention approach falls into one of the three approaches which are given as below:

- Preprocessing: It transform the data in such manner so that discriminatory biases from the original data are removed and we will get transformed data set and from which no unfair decision rule can be mined.
- In-processing: In this approach it changes data mining algorithm in such a way that resulting models does not contain any unfair decision rule.
- Postprocessing: In this approach it modify the resulting data mining model rather than cleaning original dataset or changing the data mining algorithm

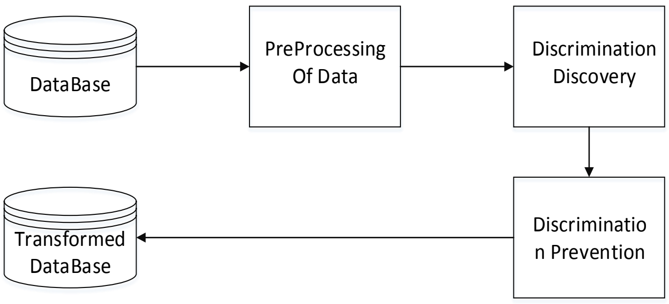

Fig 1: System Architecture

Many discrimination prevention methods were published so far [4],[9] which are based on preprocessing, but those are having some limitations, which are:

1. They does not considered combination of all the rules rather they considered single rule to prevent the discrimination, due to this reason it does not give guarantee that transformed dataset is discrimination free.

2. They only considered the direct discrimination.
3. They have not introduced any measure to calculate how much information loss has been incurred.

In earlier work[5], they introduced initial idea of some algorithm to prevent the direct discrimination, but they are not able to give some experimental results. In[2] they proposes use of rule protection method in some different manner to avoid the indirect discrimination and gives some preliminary experimental results, but it does not consider about direct discrimination prevention.

The proposed system is based on preprocessing approach which overcomes all the above limitations. This proposed approach will consider all the combination of rules. It will also consider the both types of discriminations that is direct as well as indirect discrimination. This proposed system removes both type of discrimination with minimum data loss. They also provide utility measures to calculate the accuracy of results.

II. LITERATURE SURVEY

Although the wide deployment of information systems based on data mining technology in decision making, the issue of discrimination did not receive much attention until 2008. Apart from discrimination discovery, the most challenging issue is to avoid knowledge-based decision support system from making discriminatory decisions. Some of these approaches are deal with discrimination discovery and measure of discrimination. Others deal with discrimination prevention.

It proposes modified Naive Bayes classification approach. Proposed approach based on post processing method of discrimination prevention. Independent sensitive attributes are mainly considered while doing classification. Such type of behavior occurs, when the decision process leading to the label in the dataset was biased on the basis of sensitive attributes. This technique is motivated by many case studies of decision making, in which there exists laws that does not allow a decision that is partly based on discrimination. There are three methods which are based on Bayesian classifier which are used for discrimination-aware classification.

The first approach proposes modifying probability in a Naive Bayes models in such a way that decision being positive. The second approach proposes learning of two different model. The third approach proposes adding of latent variable reflecting the latent “true” of an object discrimination free. The limitation of this approach is it does not support to the indirect discrimination and also it does not deal with numeric data[3].

Anti-discrimination having important part in cyber security in which computational intelligence technologies like data mining may be used for various decision making scenarios. It proposes antidiscrimination technique for the cyber security. It also invented new discrimination prevention approach which is based on the data transformation that can consider certain discriminatory attributes and their combinations. This approach concentrate on obtaining training data which are completely or nearly discrimination free. In order to overcome discrimination in dataset, the first step is discovery of discrimination. If any discrimination is observed then original training dataset is modified until discrimination is brought below a certain discriminatory threshold or is entirely erased. The main disadvantage of this approach is that it does not deal with indirect discrimination [5].

It gives new solution to the Classification with No Discrimination(CND) by proposing sampling scheme to get the discrimination free data rather than relabeling dataset. The proposed approach uses pre-processing method of discrimination prevention. This sampling scheme is used to get unbiased dataset.

For given dataset we can use this approach and change the distribution of different data objects so that we will get unbiased dataset. It considered that the data objects which are close to the decision making boundaries those are more responsible for discrimination. To making our dataset discrimination free we have to change the distribution of those borderline data objects. This approach gives the favorable results with the both stable as well as unstable classifiers. The main disadvantage of this approach is that is consider only those data object which are close to the borderline [4].

It proposes new pre-processing approach to deal with the indirect discrimination prevention. This is the first approach which deals with the indirect discrimination and which is based on the data transformation. Indirect discrimination comes into picture due to non-sensitive attributes which are highly correlated to the sensitive one. To deal with the indirect discrimination by using this approach firstly we have to find out whether given data set is having discrimination?

If we discover the discrimination then we have to apply the rule protection method so that discrimination is below the some threshold or it will be entirely removed. We have to apply this method again and again until our data set become discrimination free. This approach considers only indirect discrimination and it does not consider direct discrimination [2].

III. COMPARATIVE ANALYSIS

This section gives the comparative analysis of different approaches which are used to prevent discrimination .It also gives innovation as well as limitations.

TABLE
COMPARATIVE ANALYSIS

Paper	Innovation	Limitations
Rule protection for indirect discrimination.[2]	This is first approach which proposes discrimination prevention method for indirect discrimination .	It does not consider direct discrimination.
Three Naive Bayes approaches for discrimination Free Classification [3]	This approach cope up with direct discrimination by using modified Naive Bayes Classifier.	It does not consider indirect discrimination. It does not deal with numeric data.
Classification with no Discrimination by Preferential Sampling[4]	This approach gives the favorable results for stable and unstable classifier. This approach is highly accurate.	This approach consider borderline data only.
Discrimination prevention for intrusion and crime detection[5]	Data quality is retained.	This approach does not deal with indirect discrimination.

IV. CONCLUSION

In this paper, we have discrimination prevention approaches. After a review of existing techniques related to discrimination prevention, we point out that these methods are not powerful enough to avoid the both direct and indirect discrimination at a time. The proposed work presents an approach to remove both the direct and indirect discrimination at a time without much affecting on the data quality.

REFERENCES

[1] S. Hajian and J.Domingo-ferrer, “A Methodology for Direct and Indirect Discrimination Prevention in Data Mining” IEEE Transaction on Knowledge and Data Engineering, Vol. 25, No. 7, July 2013.

[2] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-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.

[3] 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.

[4] F. Kamiran and T. Calders, “Classification with no Discrimination by preferential Sampling,” Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

[5] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-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.

[6] 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.

[7] 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.

[8] 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, eds., Springer, 2008.

[9] F. Kamiran and T. Calders, “Classification without Discrimination,” Proc. IEEE Second Int’l Conf. Computer, Control and Comm.(IC4 ’09), 2009.

[10] 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.