

An Approach for Discrimination Prevention in data Mining – A Review

Rupanjali Dive
Matoshri College of Engineering
& Research Center, Nashik
University of Pune
rupanjalidive@gmail.com

Anagha Khedkar
Matoshri College of Engineering
& Research Center, Nashik
University of Pune
anaghap2@yahoo.com

Abstract — 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. Therefore, in this paper we review the recent state of the art approaches for antidiscrimination techniques and also focuses on discrimination discovery and prevention in data mining. On the other hand, we also study a theoretical proposal for enhancing the results of the data quality.

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

1. 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.

In this paper, we review the issue of direct and indirect discrimination. The rest of the paper is organized as follows. The section 2 discussed the existing literature review of the various approaches. Section 3 discussed the analysis of the existing approaches. Section 4 presented a theoretical proposal of new approach. At the end, conclusion is presented in section 5.

2. 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. In this method, the author introduced a sampling scheme for making the data discrimination free instead of relabeling the dataset. The issues the author did not consider such as they do not proposing discrimination model which is used in many cases. Also, it is acceptable from an ethical and legal point of view to have some discrimination.

D. Pedreschi, S. Ruggieri, and F. Turini [4] presented the issue of discrimination in social sense viz. against minorities and disadvantaged groups. The author attempt to handle a dataset of decision records. In this approach, the author uses a classification rule for solving a problem. On the other hand, a measure of quantitative discrimination is also introduced.

D. Pedreschi, S. Ruggieri, and F. Turini[5] presented a method that is used find the evidence of discrimination in datasets of historical decision records in socially sensitive tasks viz. Access to credit, mortgage, insurance, and labor market. They also focus on the rule based framework process for direct and indirect discrimination. In this, they also focus on the quantitative measures.

S. Hajian, J. Domingo-Ferrer, and A. Martinez-Balleste[6] introduced an anti-discrimination in the context of cyber security. And proposed data transformation method for discrimination prevention and considered several

discriminatory attributes and their combinations. The issue of data quality is also addressed. But, the limitation of this method is that first, they does not run method on real datasets and also do not consider background knowledge (indirect discrimination).

Faisal Kamiran, Toon Calders and Mykola Pechenizkiy [7] presented a model for decision making in data mining. the author proposed a new techniques viz. *discrimination aware*. The main objective is to learn classification model by using potentially biased historical data. The care has been taken in such a way that it will generate accurate predictions for future decision making. However, the author introduced two techniques viz. Dependency-Aware Tree Construction and Leaf Relabeling for incorporating discrimination awareness into the decision tree construction process.

Faisal Kamiran, Toon Calders [8] introduced a classification model which works impartially for future data. The limitation of this approach is that they do not deem other classification models for discrimination-free classification. Also, do not incorporate numerical attributes and groups of attributes as sensitive attribute(s).

Sara Hajian and Josep Domingo-Ferrer, Fellow, IEEE[9] proposed preprocessing methods which overcome the above limitations and issues. The author introduced a new data transformation method which uses rule protection and rule generalization. This method handles both the issue such as direct and indirect discrimination and also can deal with several discriminatory items.

Thus, based on the issue and limitation investigated in the literature, new data transformation methods for discrimination prevention need to be designed.

3. Our 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.

Second, it cannot guarantee that the transformed data set is really discrimination free.

Third, the literature focuses on the direct discrimination.

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.

4. Direct and Indirect Discrimination :

The issues has been investigated in the recent literature and discussed in the section 3. Based on investigation, presented a new preprocessing discrimination prevention methodology. Thus, the central theme of our approach is to use data transformation methods that help to prevent direct discrimination, indirect discrimination or both of them at the same time.

To meet this objective the following steps need to be carried out.

- First step is to measure discrimination and identify categories. Based on the same theme, make groups of individuals that have been directly and/or indirectly discriminated in the decision-making processes.
- Second step is to transform data in the proper way to remove all those discriminatory biases.
- Third, discrimination-free data models can be generated by using the transformed data. However, the data transformation is been conducted in such a way that data quality should be hurtful.

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.

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] F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling," Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.
- [4] European Commission, "EU Directive 2006/54/EC on Anti-Discrimination," <http://eurex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:204:0023:0036:en:PDF>, 2006.
- [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. Hajian, J. Domingo-Ferrer, and A. Martínez-Ballester, "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.
- [7] 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.
- [8] 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.