

An Approach for Discrimination Prevention in Data Mining

Rupanjali T.Dive¹, Anagha P. Khedkar²

¹Computer Engineering, MCOER, Nashik, Pune University (MS), India First

²Computer Engineering, MCOER, Nashik, Pune University (MS), India

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. Discrimination can be either direct or indirect. Direct discrimination consist of rules procedures that explicitly mention minority group based on sensitive discriminatory attribute 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. 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. We discuss how to clean training data sets and outsourced data sets in such a way that direct discrimination decision rules are converted to legitimate classification rules.

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 system architecture of new approach. Section 5 presented mathematical model. Section 6 presented a algorithm for discrimination prevention. At the end, results and conclusion is presented in section 6 and 7.

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.

Sara Hajian, Josep Domingo-Ferrer and Antoni Martnez-Balleste introducing anti-discrimination in the context of cyber security [10].Proposing a new discrimination prevention method based on data transformation that can consider several discriminatory attributes and their combinations; proposing some measures for evaluating the proposed method in terms of its success in discrimination prevention and its impact on data quality.

Goce Ristanoski,Wei Liu, James Bailey consider imbalanced datasets as a special case of discrimination aware classification, and identify potential challenges other methods may encounter[11].Also analyze the nature of discrimination and propose a strategy of direct minimization of the model aided discrimination that can be applied to imbalanced datasets. And consider the case of the discriminated group being more present in the dataset and the effects this might have on the learning process.

Salvatore Ruggieri Dino Pedreschi Franco Turini[12] describe the architecture of DCUBE, and a demonstration which introduces the audience to the issue of discrimination discovery, by making them aware of the legal issues (their own) data can hide, and to an approach for discrimination analysis also guides the audience through the processes for discovering direct discrimination, affirmative actions, indirect discrimination, favoritism and respondent argumentation and allows the participants to directly interact by posing specific queries, through standard SQL, over the DCUBE database.

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

3. Summary of Literature Survey:

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..

So proposed work in data mining propose preprocessing methods which overcome the above limitations. And introduces 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.

4. IMPLEMENTATION DETAILS

A. System Architecture

The main purpose of this project is to get the discrimination free database for the purpose of equality in all application areas. Data goes from different modules and finally get the discrimination free database.

Consider user request to administrator to database. Administrator extracts the database from large collection of database using data mining techniques. Database passes to discrimination prevention module to get discrimination free database. In this module database passes to many subphases like calculation of support and confidence, calculation of elift ,data transformation and all. After passing database from this module administrator get discrimination free database and passes to it.

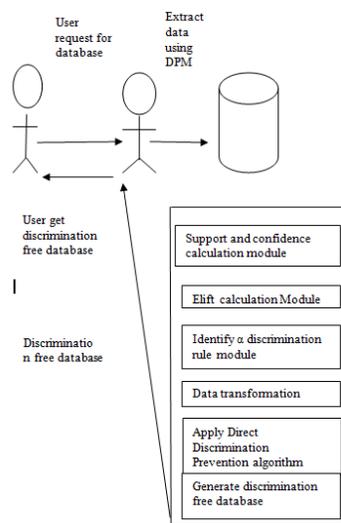


Figure: System Architecture of discrimination free database module.

B. 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.

C. Direct Discrimination Prevention Algorithm:

Step 1: Input: DB,FR, MR, alpha, DIs

Step 2: Output: DBdash

Step3: For each $r': A, B \rightarrow C$ in MR do

Step4: $FR \rightarrow FR - \{r'\}$

Step 5: $DBc =$ All records completely supporting

Step 6: A negated

Step 7: $B = C$ negated

Step 8: For each dbc in DBc do

Step 9: Compute $impact(dbc) = |\{ra \text{ in } FR | dbc$

Step 10: support the premise of $ra\}|$

Step 11: end for

Step 12: Sort DBc By ascending impact

Step13: while $conf(r') \geq \alpha * conf(B \rightarrow C)$ do

Step 14: Select first record in DBc

Step 15: Modify discriminatory item set of dbc from

Step 16: A negated to A in DB

Step 17: Recompute conf(r')
Step 18: Modify class item of dbc from C negated to
Step 19: C in DB // Contribution part
Step20: Recompute conf(B →C)
Step 21: end while
Step 22: end for

5. MATHEMATICAL MODEL

Let S be a database requested by user,
discrimination prevention module which gives discrimination free database.

Such that
S= {I, F, O} Where,

I represent the set of inputs;
I= {I1, I2,I3,I4,I5}

I1= Adult Database
I2= FR be the database of frequent classification
rule extracted from DB.
I3= MR be the database of direct discriminatory
rules.
I4= Alpha be a threshold value.
I4= DIs be a collection of sensitive attribute.

And F is the set of functions:

F= {F1, F2, F3, F4, F5, F6, F7, F8, F9}

F1 = Login to the system.
F2 = Request for database.
F3 = Admin extract database using DPM.
F4 = Calculate confidence and support.
F5 = Calculate ELIFT.
F6 = Identify discriminatory rules.
F7 =Transform database.
F8 = Apply DDP algorithm.
F9 = Get discrimination free database
And O is the set of outputs;

O = {O1}

O1=Discrimination free database without affecting original database.

Functions

F1: Login to the system

 If x = login
 F(x) = login successful
 If U [A-Z a-z] and P [A-Z a-z 0-9] and
 Length (P) >= 6
 Else
 Login fails
 Where,
 P: PasswordU: Username

F2 = Request for database.

 X: Given database.
 F(x) = Generate database with attribute.

F3 =Extract database by administrator using DPM.

 X : Database.
 F(X) = Take database and pass to DPM.

F4 = Calculate confidence and support.

X : Frequent classification rule.

F(X) = Calculate confidence and support of selected classification rule.

F5 =Elift calculation .

X : Confidence and support.

F(X) = Calculate Elift.

F6 = Identify -discriminatory rules.

X : -discriminatory rules.

F(X) =Extract -discriminatory rules.

F7 =Transform database.

X: Database.

F(X): Transformation of database.

F8 = Apply DDP Algorithm.

X: Discrimination free database.

F(X) = Generate database.

F9 = Display final output.

X: Database

F(X) = Display discrimination free database.

Schema of Database

Field Name	Data Type
workclass	Number
education	Text
education-num	Text
marital-status	Number
occupation	Text
relationship	Text
race	Text
sex	Text
capital-gain	Number
capital-loss	Number
hours-per-week	Number
native-country	Text

6. RESULTS AND DISCUSSION

In this discrimination prevention method we need to measure the impact of the method in terms of information loss i.e. data quality loss.

In the first module of discrimination prevention we are until calculating support and confidence and keep threshold value as 1.

We set class item and item set as some attributes form datasets.and calculate confidence and support of that rules having values greeated than threshold value i.e. 1.

In second module we check where the rules are affected by discrimination or not that gives the list of discriminative rules and protective rules according to confidence,support and elift calculation.

Date Set : Adult Dataset

Table 1: Adult Data Set Utility measures

	Alpha	No of direct discrimination rules
Removing Discrimination attribute	N.A.	N.A
Direct Rule Protection Method	1.2	274 (Total Records 5908)
Proposed Method	1	217 (Total Records 3289)

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://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, "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. Martí'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.
- [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.
- [9] Sara Hajian, Josep Domingo-Ferrer and Antoni Martnez-Balleste Universitat Rovira i Virgili, Discrimination Prevention in Data Mining for Intrusion and Crime Detection 2011 IEEE
- [10] Goce Ristanoski, Wei Liu, James Bailey, Discrimination Aware Classification for Imbalanced Datasets, CIKM13, Oct. 27Nov. 1, 2013, San Francisco, ACM 978-1-4503-2263-8/13/10 .
- [11] S. Ruggieri, D. Pedreschi, and F. Turini, DCUBE: Discrimination Discovery in Databases, Proc. ACM Intl Conf. Management of Data (SIGMOD 10), pp. 1127-1130, 2010.