

Discrimination prevention with improved data transformation method

Priti Gawade

Student, Department of Computer Engg.
BSIOTR, Savitribai Phule Pune University

Prof. Jayashree Chaudhari

Professor, Department of Computer engg.
BSIOTR, Savitribai Phule Pune University

Abstract— Discrimination is a very important issue in data mining. It is partial treatment of individual or group based on their actual or perceived membership in certain group or category. Discrimination can be done on attributes like religion, nationality, marital status and age. Most of the people do not want to discriminated based on their race, nationality, religion, age and so on. This problem mainly arises when these kind of attributes are used for decision making purpose such as giving them a job, loan, Insurance computation etc and give biased decision based on these attributes. For this reason finding out such attributes and removing them from the training data without affecting their decision-making utility is essential. So proposed system introduce an antidiscrimination techniques which including discrimination discovery and prevention. Discrimination prevention is used for bringing the patterns that do not lead to discriminatory decisions even if the original training datasets contain any of the discriminatory attributes. In the discrimination prevention, pre-processing discrimination prevention methods like discrimination measurement and data transformation are introduce and specify how these approaches deal with direct or indirect discrimination. Data transformation method like rule protection, rule generalization which prevent discrimination present on left hand side of classification rule and RHS discrimination prevention data transformation method for preventing discrimination present on right hand side of classification rule. Proposed technique are effective at finding out and removing direct and indirect discrimination prejudices from original data sets while preserving data quality.

Index Terms— Data mining, Discrimination, Discrimination prevention, Direct & Indirect discrimination, Discrimination measurement, Data transformation.

I. INTRODUCTION

Discrimination is nothing but partial treatment of people belonging to a specific group like Race, religion etc. It includes not granting permission to take opportunities to members of one group that are available to other group of people. Antidiscrimination laws and acts designed to prevent discrimination based on set of sensitive attributes like race, religion, gender, nationality and marital status in various settings for example credit and insurance, employment and training, access to public services etc. Different types of decision making tasks. Set of information is given about a customer, an automated system which decides whether to grant credit to customer or not. such type of automated decisions decreases the workload of the staff and insurance

companies. These information systems are used in decision making task. There are various data mining techniques such as classification or association rule mining and automated data collection have been designed and used for automated decisions making tasks but in this classification rule do not guide themselves. Classification rule are actually based on training data set learned by the system. If this training data set are biased against particular community (for example foreign workers) then there is chance of occurring discriminatory decision. There are two types of discrimination, Direct discrimination and Indirect discrimination.. Direct discrimination consists procedure or rule that explicitly mention minority or disadvantaged group membership. Indirect discrimination consists of procedure or rules that are not explicitly mentioning discriminatory attributes may or may not be generate discriminatory decision. Indirect discrimination referred to as redlining and rules causing indirect discrimination called as redlining rule. Indirect discrimination need some background knowledge.

II. LITARETURE SURVEY

In earlier work [2] focuses in discovering association rules between items in very large data base of transaction, focuses on problem of mining association rule is introduced because these rules are important in cross marketing. The Problem of finding out such association rules divided in to two phases in first phase find out all item sets that have transaction support greater than minimum support. In second phase use large item sets to find out such rules. In [4] this approach training data contains unwanted dependencies between the attributes. Given a labeled dataset and a sensitive attribute like ethnicity, the goal of this research is to learn a classifier for predicting the class label that does not discriminate with respect to the sensitive attribute it present three approaches for making the Naive Bayes classifier discrimination free: i) modifying the probability of the decision being positive, ii) training one model for every sensitive attribute value and balancing them and iii) adding a latent variable to the Bayesian model that represents the unbiased label and optimizing the model parameters for likelihood using expectation maximization.

In [3] this discrimination is nothing but unequal treatment of people based on their membership to a particular group or category. Rules extracted from databases by data mining techniques, such as classification or association rules, these

rules are used for decision making tasks such as benefit or credit approval, can give discriminatory decision introduces extended lift as measure of discrimination. In [5] discrimination refers to unfair treatment of people based on belonging to some disadvantaged group without looking their merits. Discrimination in credit, mortgage, insurance, labor market, and education has been investigated by researchers in economics and human sciences. This paper tackle the problem of discovering discrimination within a rule-based setting, by introducing the notion of discriminatory classification rules to find out the potential risk of discrimination This paper [6] discusses how to clean training datasets so that resulting datasets does not give indirect discriminating rules. Main contribution of paper is proposed new pre-processing approach based method for indirect discrimination which is based on data transformation. It also introduced some measure for finding out success of proposed method and impact on data quality. In [8] Data and knowledge hiding are two research directions that investigate how the privacy of raw data, or information, can be maintained either before or after the course of mining the data. It presents survey of recent approaches that have been applied to the association rule hiding problem.

III. RELATED WORK

Existing systems

Traditional discrimination discovery based on classification rule mining and reasoning on them based on quantitative measure of discrimination. Next it is extended to include numerical meaning of extracted patterns of discrimination. It has been developed as oracle based tool. Existing system only one rule for discrimination measure does not consider another rule or relationship between rules. If original data are partial then discriminatory decision may be generated. There are three approaches for discrimination prevention.

1. Preprocessing: In this approach input data is modified so that the discriminatory biases which contained in the input data are removed so that no unfair decision rule can be generated from the transformed data and after that apply any of the standard data mining algorithms [1].

2. In-Processing: In this approach data mining algorithm change so that generated model does not contain any unfair decision rules[1]. This discrimination prevention methods require new special-purpose data mining algorithms, cannot use standard data mining algorithms.

3. Post Processing: In this approach change data models of data mining rather than transforming input data and modifying data mining algorithm but cons is that only changed data mining models are published not data set.[1]

Proposed system

The Proposed system focuses on preprocessing discrimination prevention because the preprocessing approach is flexible one cause it does not require modifying the standard data mining algorithms and it allows data publishing rather than just knowledge publishing.

1. It will clean training data sets in such a way that direct and/or indirect discriminatory decision rules are identified

effectively and convert to legal (nondiscriminatory) classification rules with the minimum impact on data.

2. Propose preprocessing approach with improved data transformation method for preventing discrimination present on right hand side of classification rule and rule protection and rule generalization are based on measures for both direct and indirect discrimination and can deal with several discriminatory items. Also, provide utility measures.

Preprocessing direct and indirect discrimination prevention can be described in terms of two phases.

1) Discrimination Measurement

To discover direct and indirect discrimination first identify α discriminatory rule & redlining rule. To identify α discriminatory rule.& redlining rule, first there are some predetermine discriminatory items in database from that frequent classification rules extracted which is divided in to two groups Potential discriminatory (PD) and Potential nondiscriminatory(PND).From PD groups find out α discriminatory rule by using direct discrimination measure elift and discriminatory threshold α . For indirect discrimination identify redlining rules from PND groups by using indirect discrimination measure elb and discriminatory threshold α after finding out α discriminatory rules stored in MR database and redlining rule are stored in RR database.

2) Data Transformation

In the data transformation phase transform the original data DB in such a way to remove direct and indirect discriminatory biases, with less impact on the data and on legal decision rules, so that no unfair decision rule can be mined from the transformed data. In this different data transformation methods that can be used for this purpose like rule protection ,rule generalization and discrimination prevention on right hand side of classification rule.

System architecture

We consider preprocessing approach for discrimination prevention which has two phases first one is discrimination measurement and second one is data transformation. These two phases contains frequent classification rule extraction after that both type of discrimination discovery including direct discrimination measurement and indirect discrimination measurement and finally data transformation methods which prevent direct and indirect discrimination as well as discrimination present on right hand side of classification rule as shown in fig.1.

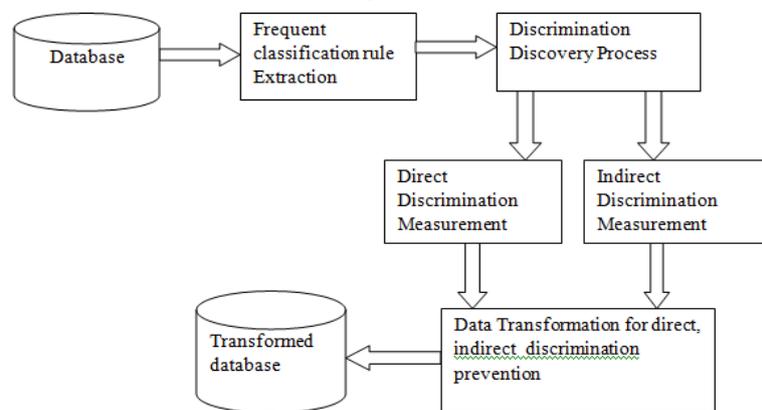


Fig 1 System Architecture

Algorithm

Algorithms structure for Direct and Indirect discrimination prevention.

Input :-DB = Original Data set
 FR = Frequent Item set
 RR = Redlining rules
 MR = Direct rule from DB
 Alpha(α) = Threshold
 DI_s = predetermined discriminatory items
 Output :-DB' = transformed data set

1: for each $r : X \rightarrow C \in RR$,
 where $D, B \subseteq X$
 do
 2: $\gamma = \text{conf}(r)$
 for each $r' (A \subseteq DI_s), (B \subseteq X) \rightarrow C \in RR$
 do
 3: $\beta = \text{conf}(r_{b2}: X \rightarrow A)$
 4: $\Delta_1 = \text{supp}(r_{b2}: X \rightarrow A)$
 5: $\delta = \text{conf}(B \rightarrow C)$
 6: $\Delta_2 = \text{supp}(B \rightarrow A)$
 7: $\beta_1 = \frac{\Delta_1}{\Delta_2} \quad // \text{conf}(r_{b1} : A, B \rightarrow D)$
 8: Find DBc: all records in DB that completely support $\neg A, B \rightarrow \neg C$
 9: for each db_c compute impact of DBc where $r' \in FR$
 , db_c supports premise of r'
 10: end for
 11: arrange impact in ascending order.
 12: If $r' \in MR$ then
 13: while $(\delta \leq \frac{\beta_1(\beta_2 + \gamma - 1)}{\beta_2 \cdot \alpha})$ and $(\delta \leq \frac{\text{conf}(r')}{\alpha})$
 14: Select first record db_c in DB_c
 15: Modify the class item of db_c from $\neg C$ to C in DB
 16: Recompute $\delta = \text{conf}(B \rightarrow C)$
 17: end while
 18: else
 19: while $(\delta \leq \frac{\beta_1(\beta_2 + \gamma - 1)}{\beta_2 \cdot \alpha})$ do
 20: Steps 13 - 15
 21: end While
 22: end if
 23: end for
 24: end for
 25: for each $(r' : A, B \rightarrow C) \in MR \setminus RR$, do
 26: $\delta = \text{conf}(B \rightarrow C)$

27: Find DBc: all records in DB that completely support

$\neg A, B \rightarrow \neg C$

28: Steps 8-11

29: while $(\delta \leq \frac{\text{conf}(r')}{\alpha})$ do

30: step 15 - 17

31: for each $(r' : B \rightarrow A) \in RR$ do

32: $\delta = \text{conf}(B \rightarrow A)$

33: Find DBc: all records in DB that completely support

$B \rightarrow \neg A$

34: Select first record db_c in DB_c

35: Modify A to $\neg A$

36: Recompute $\delta = \text{conf}(B \rightarrow A)$

37: end while

38: end for

39: Output $DB' = DB$

Modules

1) *Frequent classification rule extraction:*

Classification rule whose support and confidence of rule is greater than specified lower bounds called as frequent classification rule. Hence, if $\text{supp}(X) > 0$ then $\text{conf}(X \rightarrow C) = \frac{\text{supp}(X \rightarrow C)}{\text{supp}(X)}$. Frequent classification rule stored in FR database.

Direct discrimination measurement

In this module measure is used to determine degree of discrimination of potential discriminatory (PD) rule. In proposed system extend lift (elift) is used as measure of direct discrimination..

3) *Indirect discrimination measurement*

In this module measure is used to determine degree of discrimination of potential non discriminatory (PND) rule. In proposed system elb is used as measure of indirect discrimination.

4) *Direct discrimination prevention*

To prevent direct discrimination, decision rules would not contain direct discrimination if it only contain PD rules which are protective or are instances of at least one non redlining PND rule. In this apply direct rule protection and direct rule generalization.

5) *Indirect discrimination prevention*

To prevent indirect discrimination is based on the fact that the data set of decision rules would be free of indirect discrimination if it contained no redlining rules. To achieve this redlining rules are converted to non redlining rules for this purpose indirect rule protection is used.

6)RHS discrimination prevention.

In this module new data transformation method is used to prevent discrimination present on right hand side of classification rule. This is due to discriminatory item set is on right hand side and having confidence ,support value greater than specified lower bound so it can be give discriminatory decision.

Expected Result set

In this system we are taking German credit card data set as input and one more adult data set [9]. I am considering one data set as example in which 12 records are their which contains information about individual and 13 attribute as shown in below figure. Predetermined discriminatory item sets are given. Record number 7 has direct discrimination and 12 has indirect discrimination.

#	A	B	C	D	E	F	G	H	I	J	K
1	workclass	Age	education	marital-status	occupation	relationship	race	sex	hours-per-week	native-country	Label
2	State-gov	Old	Bachelors	Never-married	Adm-clerical	Not-in-family	White	Male	40	United-States	<=50K
3	Private	Young	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	13	United-States	<=50K
4	Private	Old	HS-grad	Divorced	Handlers-cleaners	Not-in-family	White	Male	40	United-States	<=50K
5	Private	Old	11th	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	40	United-States	>50K
6	Private	Old	Bachelors	Married-civ-spouse	Prof-specialty	Wife	Black	Female	40	Cuba	>50K
7	Private	Young	Masters	Married-civ-spouse	Exec-managerial	Wife	White	Female	40	United-States	<=50K
8	Private	Old	9th	Married-spouse-absent	Other-service	Not-in-family	Black	Female	16	Jamaica	<=50K
9	Self-emp-not-inc	Young	HS-grad	Married-civ-spouse	Exec-managerial	Husband	White	Male	45	United-States	>50K
10	Private	Old	Masters	Never-married	Prof-specialty	Not-in-family	White	Female	50	United-States	>50K
11	Private	Young	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	40	United-States	>50K
12	Local-gov	Young	Bachelors	Married-civ-spouse	Other-service	Husband	White	Female	56	Germany	>50K
13											
14											Method-2 Modify the class item negation of C to C in DB

#	A	B	C	D	E	F	G	H	I	J	K
1	workclass	Age	education	marital-status	occupation	relationship	race	sex	hours-per-week	native-country	Label
2	State-gov	Old	Bachelors	Never-married	Adm-clerical	Not-in-family	White	Male	40	United-States	<=50K
3	Private	Young	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	13	United-States	<=50K
4	Private	Old	HS-grad	Divorced	Handlers-cleaners	Not-in-family	White	Male	40	United-States	<=50K
5	Private	Old	11th	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	40	United-States	>50K
6	Private	Old	Bachelors	Married-civ-spouse	Prof-specialty	Wife	Black	Female	40	Cuba	>50K
7	Private	Young	Masters	Married-civ-spouse	Exec-managerial	Wife	White	Female	40	United-States	<=50K
8	Private	Old	9th	Married-spouse-absent	Other-service	Not-in-family	Black	Female	16	Jamaica	<=50K
9	Self-emp-not-inc	Young	HS-grad	Married-civ-spouse	Exec-managerial	Husband	White	Male	45	United-States	>50K
10	Private	Old	Masters	Never-married	Prof-specialty	Not-in-family	White	Female	50	United-States	>50K
11	Private	Young	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	40	United-States	>50K
12	Local-gov	Old	Bachelors	Married-civ-spouse	Other-service	Husband	White	Female	56	Germany	<=50K
13											
14				Dis-(sex-Female, Age-Young)							
15				No-7 Direct discrimination							
16				No-12 indirect discrimination							

This direct discrimination and indirect discrimination is prevented by changing discriminatory item set that is by Method 1 and by changing class item that is by Method 2as shown in below figures.

#	A	B	C	D	E	F	G	H	I	J	K
1	workclass	Age	education	marital-status	occupation	relationship	race	sex	hours-per-week	native-country	Label
2	State-gov	Old	Bachelors	Never-married	Adm-clerical	Not-in-family	White	Male	40	United-States	<=50K
3	Private	Young	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	13	United-States	<=50K
4	Private	Old	HS-grad	Divorced	Handlers-cleaners	Not-in-family	White	Male	40	United-States	<=50K
5	Private	Old	11th	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	40	United-States	>50K
6	Private	Old	Bachelors	Married-civ-spouse	Prof-specialty	Wife	Black	Female	40	Cuba	>50K
7	Private	Young	Masters	Married-civ-spouse	Exec-managerial	Wife	White	Female	40	United-States	<=50K
8	Private	Old	9th	Married-spouse-absent	Other-service	Not-in-family	Black	Female	16	Jamaica	<=50K
9	Self-emp-not-inc	Young	HS-grad	Married-civ-spouse	Exec-managerial	Husband	White	Male	45	United-States	>50K
10	Private	Old	Masters	Never-married	Prof-specialty	Not-in-family	White	Female	50	United-States	>50K
11	Private	Young	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	40	United-States	>50K
12	Local-gov	Young	Bachelors	Married-civ-spouse	Other-service	Husband	White	Female	56	Germany	<=50K
13											
14				using Method-1 negation of A to A in DB							

IV. CONCLUSION

The proposed preprocessing approach is to develop discrimination prevention methodology including different data transformation methods that can prevent direct discrimination, indirect discrimination or both of them at the same time and discrimination present in right hand side of (RHS) classification rule. For this , first step is to measure discrimination and identify categories and groups of individuals that have been directly and/or indirectly discriminated in the decision-making processes, the second step is to transform data in the proper way to remove all those discriminatory biases. Finally, discrimination- free data models can be produced from the transformed data set.

REFERENCES

- [1] Sara Hajian and Joseph 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] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] S. Hajian, J. Domingo-Ferrer, and A. Mart'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, Classification with no discrimination by preferential sampling, *Proc. 19th Machine learning conf. Belgium and The Netherlands*, 2010.
- [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] 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.