ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

Data preprocessing approach for removal of direct and indirect discrimination

Prakash Kumar[1], Kapil Patil[2], Aditya Patnurkar[3], Prof. Alka Londhe[4]

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India[1,2,3,4]

Abstract: Data mining is an important technology for extract the useful knowledge hidden in large collection of data. Automated data collection and data mining techniques such as classification rule mining have proved the way to making automated decisions, like loan granting, insurance premium computation, etc. If the training data set is biased in what regards sensitive attributes like gender, race, religion, colour, etc., . For this reason, anti-discrimination techniques including discrimination discovery and prevention have been introduced in data mining. Discrimination can be direct or indirect. Direct discrimination occurs only when decisions are made based on their sensitive attributes where indirect discrimination occurs when decisions are made based on their no sensitive attributes which are strongly related with biased sensitive ones. In this paper, we focus discrimination prevention in data mining and propose a new techniques applicable for direct or indirect discrimination prevention individually or both at the same time. Also we discuss how to clean training data set and outsourced data sets in such a way that direct and indirect discriminatory decision rules are converted to legitimate (non-discriminatory) classification rules. We also propose new metrics to evaluated the utility of the proposed approach and we compare these approach. The experimental evaluations implemented that the proposed techniques are effective removed by direct and/or indirect discrimination biases in the original data set while preserving data quality.

I. INTRODUCTION

In sociology, discrimination is the prejudicial treatment of an individual based on their different membership in a certain group or category. It involves denying to members of one group opportunities that are available to other group. There is a list of anti-discrimination act, which are laws

Designed for prevention of discrimination on the basis of a Number of attributes (e.g. race, religion, gender, nationality, disability, and age) in various settings (e.g. employment and training, access to public services, credit and insurance, etc.). For e.g. the European Union implement the principles of equal treatment between women and men in the access to and supply of goods and service in [3] or in matters of employment and occupation in [4]. Although there are some law against discrimination all of them are reactive, not proactive. The technology can add activity to legislation by contributing discrimination discovery and prevention techniques. Services in the information society that allows for automatic and routine collections of huge data. Those data are used to train classification rules in view of making automated decision, like loan granting. Personnel selection, etc. Personal preferences. After all at a closer look one realize that rules are actually learned by the system (e.g., loan denial) from the training data. If the training data are inherently discriminated for or against for a particular community (e.g., foreigner), the learned model show a discriminatory prejudiced behavior. In different words, the system may assume that just being foreign is a legitimate reason for loan granting. The discovering such potential eliminating and biases them from training data without affecting their decision-making utility is highly desirable. One most prevent data mining from becoming itself a source of discrimination due to data mining task creating discriminatory model from biased data set as part of the automated decision making. In the [12], it is implemented that data mining can be both a source of discrimination and mean for discovering discrimination. The discrimination can be either indirect or direct. Direct discrimination consist of rules or procedure that explicitly mention minority groups based on sensitive discriminatory attribute related to group membership. Indirect discrimination includes the rules or procedures that, while not explicitly mentioning discriminatory attribute, intentionally or un-intentionally could generate discriminatory decision. Red-lining by financial institutions is an archetypal example of indirect discrimination, although they certainly not the only one.

With a slight offence of language for the sake of tightness. In this paper indirect discrimination will also be referred as redlining and rules causing indirect discrimination will be called redlining rule [12]. Indirect discrimination can also happen because of the availability of some background knowledge for e.g. that a certain zip code corresponds to a deteriorating area. The background knowledge might be available from publicly accessible data or can also be obtained from the original data set itself because of the existence of non-discriminatory attribute that are highly related with the sensitive once in the original data set.

Related work: Although the wide deployment of information system based on data mining technologies in decision making the issue of anti-discrimination in data mining did not receive much focus until 2008 [12]. Some scenarios are oriented to the discovery and measure of

DOI 10.17148/IJARCCE.2015.4439

discrimination. Others are related with the prevention of discrimination The discovery of discriminatory decision was first proposed by Pedreschi et al. [12], [15]. The approach is based on mining classification rules and the deductive part on the basis of quantitative measure of discrimination that describe legal definitions of discrimination. For e.g., the US Equal Pay Act [18] states that: "a selection rate for any race, sex, color or ethnic group which is less than four-fifths of the rate for the group with the highest rate will generally be considered as evidence of adverse impact." This approach has been extended to enclose statistical importance of the extracted pattern of discrimination in [13] and the reason about approving action and favoritism [14]. it has been implemented as an Oracle-based tool in [16]. The current discrimination discovery method consider as each rule individually for measuring discrimination without consider the other rules or the relation between them. In this paper we also taken into consider the relation between rules for discrimination discovery, depending on the existence or nonexistence of discriminatory attributes. Discrimination prevention, the other major anti-discrimination goal in data mining, includes of inducing patterns that do not leads to discriminatory decision even if the original training data sets are biased. Three approaches are conceivable:

Preprocessing: Transform the source data in such way that the discriminated contained in the original data are removed so that no unequal decision rule can be mined from the transformed data and apply any of the standards data mining algorithm. The preprocessing approach of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature.

Along this line, [7], [8] perform a controlled distortion of the training data from which a classifiers are learned by making minimally intrusive updations leading to an unbiased data set. The preprocessing approach is useful for application in which a data set should be published or in which data mining needs to be performed also by external parties

In-processing: Change the data mining algorithm in such a way that the resulting models do not contains biased decision rule. For e.g., an another approach to removing the discrimination from the original data set is given in [2] whereby the non-discriminatory rules is embedded into a decision tree learner by changing its dividing criterion and eliminates strategy through a novel leaf relabeling approach. After all, it is obvious that in- processing discrimination prevention methods must depend on new special-purpose data mining algorithm; standard data mining algorithms cannot be used.

Post processing: Change the resulting data mining models, in place of cleaning the original data set or changing the data mining algorithms. For e.g., in [13], a confidence-altering approach is given for classification rules completed by the CPAR algorithm. The post processing technique does not allow the data set to be published: only the updated data mining models can be published and hence data mining can be performed by the data holder only. One might think of a direct pre-processing approach consisting of just eliminating the discriminatory attribute from the data set. Despite this would solve the direct discriminations problem, it would cause much information loss and in general it would not solve indirect discrimination. As listed in [12] there may be other attributes (e.g., Zip code) that are interrelated with the sensitive ones (e.g., Race) and allow inferring discriminatory rules. Hence, there are two important challenges regarding discrimination prevention: one challenge is to consider both direct and indirect discrimination instead of only direct discrimination; the other test is to find a good arrangement between discrimination removal and the qualities of the resulting training data sets and data mining models. Despite some methods have already been proposed for each of the above mentioned approach (preprocessing, in-processing, post-processing), discrimination prevention stays a huge unexplored research approach. In this paper, we concentrate on discrimination prevention based on preprocessing, because the preprocessing approach seems the most extensible one: it does not require changing the standard data mining algorithm, unlike the in-processing approach, and it allow data publishing (rather than just knowledge publishing), unlike the postprocessing approach.

Contribution and Plan of This Paper: Discrimination prevention methods depends on pre- processing published so far [7], [8] present some drawbacks, which we next highlight. They try to detect discrimination in the real data only for one discriminatory item and depend on a single measure. This approach cannot guarantee that the transposed data set is really discrimination free, because it is known that discriminatory behaviors can often be kept private behind many discriminatory item, and even behind collection of them. They do not include any measure to identify how much discrimination has been eliminated and how much information loss has been occurred. In this paper, we assert pre-processing techniques which affected the above limitation. Our new data transformations techniques are based on measures of both indirect and direct discrimination and can dealed with many discriminatory items. Moreover, we provide utility measures. Therefore, our intension to discrimination prevention is wider than in previous work. In our previous work [5], we proposed the initial idea of using rule protection and rule generalization for the direct discrimination prevention, but we achieved no experimental results. In [6], we introduced the use of rule protection in a another way for indirect discrimination prevention and we proposed some preliminary experimental results. In this paper, we present a unified technique to indirect and direct discrimination prevention, with final algorithms and all possible data transformation techniques that depend on rule protection and/or rule generalization that can applied for indirect or direct

DOI 10.17148/IJARCCE.2015.4439

discrimination prevention. We specify different features and use of each method. Since, techniques in our previous papers [5], [6] could only work with either indirect or direct discriminations; the techniques described in this paper are new and different. As part of this effort, we have developed metrics that shows which records should be changed, how many records should be changed, and how those records should be modified during data transformation. However, we gave new utility measures to find the different proposed discrimination prevention techniques in terms of data quality and discrimination removal for both indirect and direct discrimination. Depending on the given measures, we present extensive experimental result for two well defined data sets and compare the different possible techniques for indirect or direct discrimination prevention to identify which technique can be more successful in terms of low information loss and high discrimination elimination. The rest of this paper is arranged as follow. Section 2 lead into some basic definition and concept that are used in the paper. The Section 3 characterize our layout for direct and indirect discrimination prevention.

II. BACKGROUND

In this section, we briefly review of the background knowledge required of this paper. First, we revise some basic definitions of related data mining [17]. And than, we explain on measuring and discovering discrimination.

2.1 Basic Definitions

The data set is collection of data objects and their different attributes. Let DB is the original data set. An item is a attribute with its value, e.g., Race = black. An item set, i.e., X, is a collection of one or more items, e.g., {Foreign workers=Yes; City =DUBAI}. A classification rule is an expression X ->C, where C is class item and X is an item Set containing no class item, e.g., {Foreign worker = Yes; City = DUBAI -> Hire = no. X is called the premise of the rule.

- Apriori: Apriori algorithm is a well known algorithm in data mining world. Apriori algorithm requires an input which is called minimum support value. It generate frequent item sets which are having minimum support greater than user provided minimum support. It also calculates the minimum support for each frequent item sets(FIS).
- PD and PND Rules: The FIS generated by Apriori are divided into PD and PND rules. The PD rule contain those rules on which direct discrimination prevention is possible. And, PND rule contain those rules on which indirect discrimination prevention is possible.
- Alpha discriminated Rules: The PD rules are then categorized in alpha discriminated and alpha protected rules. Elift is calculated for each PD rules and each Elift value is compared with user defined threshold value called alpha. If elift > alpha, then PD rule is categorized into alpha discriminated rule else it is categorized into alpha protected rule.
- Non-Redlining Rules: PND rules are categorized in redlining and non red-lining rules.Elb is calculated for each PND rules and each Elb value is compared with user defined threshold value called alpha. If Elb < alpha, then PND rule is categorized into non redlining rule else it is categorized into redlining rule.
- Direct Rule Protection: For each alpha discriminated rules, direct rule protection is applied which transform the dataset which user had provided.
- Rule Generalization: For each alpha discriminated rules, direct rule generalization is applied which transform the dataset which user had provided.

Algorithms:-

1. Direct Discrimination prevention Algorithm

We start with direct rule protection. Algorithm 1 details Method 1 for DRP. For each direct discriminatory rule r0 in MR (Step 3), after finding the subset DBc (Step 5), records in DBc should be changed until the direct rule protection requirement (Step 10) is met for each respective rule (Steps 10-14).

1. Direct Rule Protection(Method 1)

```
1)  Input -DB, FR, MR, a, DIs
2)  Output -DB'(transformed data set)
3)  for each r : A,B C MR do.
4)  FR FR fr.
5)  DBc All records completely supporting A, B C .
6)  for each dbc DBc do.
7)  Compute impact (dbc) = — ra FR— dbc supports the
    premise of ra —.
8)  end for.
9)  Sort DBc by ascending impact.
10) while con f(r)¿= a.con (B C) do.
11) Select first record in DBc.
12) Modify discriminatory item set of dbc from A to An
    in DB.
13) Recomputed con f(r).
14) end while.
15) end for.
16) Output: DB= DB.
```

Among the records of DBc, one should change those with lowest impact on the other(protective or nonredlining) rules. Hence, for each record dbc DBc, the number of rules whose premise is supported by dbc is taken as the impact of dbc (Step 7), that is impact (dbc); the rationale is that changing dbc impacts on the confidence of those rules.

Then, the records dbc with minimum impact(dbc) are selected for change (Step 9), with the aim of scoring well in terms of the utility measures proposed in the next section. We call this procedure (Steps 6-9) impact minimization and we reuse it in the pseudo codes of the rest of algorithms specified in this paper.

Algorithm 2 details Method 2 for DRP. The parts of Algorithm 2 to find subset DBc and perform impact minimization (Step 4) are the same as in Algorithm 1. However, the transformation requirement that should be met for each discriminatory rule in MR (Step 5) and the kind of data transformation are different (Steps 5-9).

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

2. Direct Ruel Protection(Method 2)

```
1)  Input -DB, FR, MR, DIs
2)  Output -DB (transformed data set)
3)  for each r : A,B C MR do
4)  Steps 4-9 Algorithm 1
5)  while con f(B C)(conf(r'))/a do
6)  Select first record in DBc
7)  Modify the class item of dbc from C to C in DB
8)  end while
9)  end for
10) Output: DB= DB
```

As mentioned rule generalization cannot be applied alone for solving direct discrimination prevention, but it can be used in combination with Method 1 or Method 2 for DRP. In this case, after specifying the discrimination prevention method (i.e., direct rule protection or rule generalization) to be applied for each a-discriminatory rule based on the algorithm in Algorithm 3 should be run to combine rule generalization and one of the two direct rule protection methods.

3. Direct Rule Protection and Rule Generalization

Algorithm 3 takes as input T R, which is the output of the algorithm in, containing all r0 2MR and their respective TRr and rb. For each discriminatory rule r in T R, if TRr shows that rule generalization should be performed (Step 5), after determining the records that should be changed for impact minimization (Steps 7-8), these records should be changed until the rule generalization requirement is met (Steps 9-13).

Also, if TRr0 shows that direct rule protection should be performed (Step 15), based on either Method 1 or Method 2, the relevant sections of Algorithms 1 or 2 are called, respectively (Step 17)

Algo 3. Direct Rule Protection and Rule Generalization

```
1)  Input -DB, FR, RR, MR, DIs.
2)  Output-DB0 (transformed data set)
3)  for each r : X C RR where D,B-¿x
4)  for each r : A DIs;BXCRRdo:2 = conf(r)b2 : (XA):
5)  1 = support(rb2 : X A).
6)  = conf(B C).
7)  2= supp(B A)
8)  1 = 1/2 //conf(rb1 : A;B D)
9)  Find DBc: all records in DB that completely
10) Support: A; B; D C
11) Steps 6-9 Algorithm 1
12) if r MR then
13) while (1(2+-1))/(2-)
14) Select first record dbc in DBc
15) Modify the class item of dbc from: C to C in DB
16) Recomputed = conf(B C)
17) end while
18) else
19) while 121do
20) Steps 15-17 Algorithm 4
21) end while
22) end if
23) end for
24) end for
25) for each r : (A;B C) MRn RR do
26) = conf(B C)
27) Find DBc: all records in DB that completely support
    A;BC
28) Step 12
29) while( (conf(r'))/)
30) Steps 15-17 Algorithm 4
31) end while
32) end for
33) Output:DB'=DB
```

If some rules can be extracted from DB as both direct and indirect a-discriminatory rules, it means that there is overlap between MR and RR; in such case, data transformation is performed until both the direct and the indirect rule protection requirements are satisfied (Steps 13-18).

This is possible because, as we showed in Section 3.4, the same data transformation method(Method 2 consisting of changing the class item) can provide both DRP and IRP. However, if there is no overlap between MR and RR, the data transformation is performed according to Method 2 for IRP, until the indirect discrimination prevention requirement is satisfied (Steps19-23) for each indirect a-discriminatory rule ensuing from each redlining rule in RR; this can be done without any negative impact on direct discrimination prevention, as justified in. Then, for each direct a-discriminatory rule r0 2MRnRR (that is only directly extracted from DB), data transformation for satisfying the direct discrimination prevention requirement is performed(Steps 26-33), based on Method 2 for DRP; this can be done without any negative impact on indirect discrimination prevention, as justified in. Performing rule protection or generalization for each rule in MR by each of Algorithms 1-4 has no adverse effect on protection for other rules (i.e., rule protection at Step i x to make r0 protective cannot turn into discriminatory a rule r made protective at Step i) because of the two following reasons: the kind of data transformation for each rule is the same (change the discriminatory item set or the class item of records) and there are no two a-discriminatory rules r and r0 in MR such that r r0.

Appriori algorithm:

$$conf(X \rightarrow C) = \frac{supp(X, C)}{supp(X)}.$$

$$\frac{conf(r' : A, B \rightarrow C)}{conf(B \rightarrow C)} < \alpha.$$

These are used whenever required

$$conf(r' : A, B \rightarrow C) = \frac{supp(A, B, C)}{supp(A, B)},$$

$$conf(B \rightarrow C) > \frac{conf(r' : A, B \rightarrow C)}{\alpha}.$$

DOI 10.17148/IJARCCE.2015.4439

$$conf(B \rightarrow C) = \frac{supp(B,C)}{supp(B)},$$

Computational cost

$$d > \left(\frac{N_{ABC}}{DRP_{req1}} - N_{BC} \right).$$

$$O\left(m* \left\lceil \frac{N_{ABC}}{DRP_{req1}} - N_{BC} \right\rceil \right).$$

$$O((f+n) * \{m + hk + h \log h + dm\}),$$

where $d = \lceil (N_B * max_{req}) - N_{BC} \rceil$ and

$$max_{req} = \max\left(\frac{\beta_1(\beta_2 + \gamma - 1)}{\beta_2 \cdot \alpha}, \frac{conf(r')}{\alpha} \right).$$

REFERENCES

[1]. S.Hajian and J. Domingo Federer ”Methodologies For Direct And Indirect Discrimination Prevention In Data Mining”, IEEE Transactions On Knowledge And Mining, vol.25, No.07, Jul 2013

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

[3]. T. Calders and S. Verwer, Naive Bayes Approaches for Discrimination- Free Classification,”Data Mining and Knowledge Discovery, pp. 277-292, 2010.

[4]. S.Hajian and J. Domingo Federer, Naive Bayes Approaches for Discrimination- Free Classification,” Discrimination Prevention In Data Mining For Intrusion And Crime Detection, Proc. IEEE Symp.Computational Intelligence in Cyber Security (CICS ’11),pp. 47-54, 2011.

[5]. F. Kamiran, T. Calders, and M. Pechenizkiy , Aware Decision Tree Learning,” Proc. IEEE Int’l Conf. Data Mining (ICDM ’10), pp. 869-874,22

[6]. S. Ruggieri, D. Pedreschi, and F. Turini, “DataMining for Discrimination Discovery,” ACM Trans. Knowledge Discovery from Data, vol. 4, no. 2, article 9, 2010.

DOI 10.17148/IJARCCE.2015.4439