Pre-Post Processing of Discriminated Data in Data Mining

Vaibhav P Sonawane, P C Jaypal

Abstract— Data mining is increasingly important domain for retrieving knowledge from the large from the large databases. However potential privacy invasion and potential discrimination have made a negative social perception about data mining. The later consist of unjust or prejudicial treatment of different categories of people specifically on basis of their age, region or race. Biased decision may occur because of sensitive attributes which may infer in biased rules. Such discrimination is defined as direct discrimination. Some biased decision may occur because of some attributes are strongly connection with the sensitive attributes; such discrimination is called as indirect discrimination. In this paper, we try to summarize the techniques used to prevent this discrimination also some transformation are being focused. We also highlight the post-processing method to remove discrimination by using (CPAR) algorithm.

Index Terms— Preprocessing, Postprocessing, Association Rule Mining, Classification using predictive association rule, Direct Rule Protection, Discrimination.

1 Introduction

A prejudicial treatment to people on the basis of their belonging to an ethnic group, race, ideology, gender, etc., is known as discrimination. Discrimination has been studied on the various fields like social, political, employment and various other fields are also under investigation.

Loan granting, staff selection, education and life insurance are the decision making fields which tends to discrimination. Many decision making system are dependent on the information system (i.e. background Knowledge). By setting the sensitive attribute the customer can be judged that he can be credited the loan or the life insurance.

The workload of the staff will be greatly reduced in the banks, education system if the automated decision will be free from discrimination. In the automated decision making system classification/rule mining is being used at the large extent. At first we are wrong assumption that automated system are taking decision wisely without partializing, but it is not so in the practical. The automated decision making system makes unfair decision which may come into attack by the affected people. If the training data set are inherently biased for or against a particular community (for example, foreigners), the learned model may give bad result against that community which is not expected in the data mining. The learned rules will also show biased behavior toward foreign people, if biased historical dataset is used as training data to learn classification rules for an automated loan granting system. In other words, we can conclude that system may infer that just being foreign is a legitimate reason for loan denial. The discriminatory rules extracted could lead to automated unfair decisions, if the original biased dataset DB is used for data analysis without any anti-discrimination process (i.e. discrimination discovery and prevention). On the contrary, DB can go through an anti-discrimination process so that the learned rules are free of discrimination, given that a list of discriminatory attributes (e.g. gender, race, age) is present. As an output, fair and legitimate automated decisions are enabled.

There are two types of discrimination which are direct and indirect. Direct discriminatory rules shows that the biased rule are inferred from the sensitive attributes. Indirect discrimination rules are the rules which are inferred from the attributes which are strongly correlated with the sensitive attributes. Indirect discrimination could happen because of the availability of some background knowledge (rules).

2 Literature Survey

In previous studies about the discrimination prevention techniques only preprocessing and post processing techniques are addressed at different instance. The concept of discrimination was first highlighted by Pedreschi [3]. The method was related to mining classification rules (the inductive part) and reasoning (the deductive part) on the basis of quantitative measures of discrimination that create legal definitions of discrimination. The extracted patterns of discrimination in [4] and to reason about affirmative action and favoritism [3] approach has been extended to encompass statistical significance. There are three methods for preventing discrimination classified according to in which stages it is removed. These are as follows:

- Preprocessing: In this technique the discrimination is discovered and removed at pre-stage of the data mining. Here data is modified before the dataset is supplied to the mining algorithm.
- Inprocessing: In this process the data mining algorithm is changed so it will not result into discriminated data.

Postprocessing: In this process the data is modified at the end where the output of the mining algorithm is being modified.

3 BACKGRAOUND KNOWLEDGE

- Let *DIs* be the set of predetermined discriminatory items in *DB*.
- Frequent classification rules in *FR* fall into one of the following two classes:
    1. A Classification rule X → C is potentially discriminatory (PD) when X = A, B with A is a nonempty discriminatory item set and B a nondiscriminatory item set.
    2. A Classification rule X → C is potentially non- discriminatory (PND) when X = D, B is a non- discriminatory item set.
- Let *MR* be the database of the direct α discriminatory rules.
- Let *RR* be the database of redlining rules and their respective indirect α-discriminatory rules obtained.
- Let *BK* be a database of background rules that is defined as *BK* = {r: D, B → A | A is discriminatory item set and supp(D, B → A) >= ms}

4 PREPROCESSING TECHNIQUE

In this method there are two types of discrimination to deal with i.e. direct discrimination removal and indirect discrimination removal. The direct discrimination removal deal with the sensitive attributes in database and indirect discrimination removal refers to the attributes which are strongly correlated with the perceptive attributes. There are again two methods for removing direct data discrimination named as Direct Rule Protection and Rule Generalization. Again Direct Rule Protection is having two types of technique which are named as Direct Rule Protection method 1 and Direct Rule Method 2.

4.1 Direct Rule Protection

In [1], they have suggested a technique to remove direct discrimination rule by making direct discrimination rule as direct rule protected one. For this we assume that our classified rule is classified on the minimum support called α (alpha). With this we come to know that we are searching for the inferred rule with minimum support more than α. To make sure that our inference rule does not fall into α discriminated data we are converting α discriminated rule into α protective. This is done by using term called *elift* (Extended lift). The formula for extended lift is given by,

Definition1: [Elift] Let A, B → C be an association rule such that conf (B → C) > 0. We define the extended lift of the rule with respect to B as:

conf(A,B →C)/conf (B → C). (1)

We call B the context, and B → C the base-rule. By using this formula we are converting α discriminated rule into α protective by using following technique.

Definition 2: [α protection] c = A, B →C be a PD classification rule, where A is a PD and B is a PND itemset, and let: ɣ = conf(A,B →C) δ = conf(B → C) > 0.
For a given threshold $\alpha \geq 0$, we say that c is α-protective if elift(ɣ, δ) < α,
Where: elift(ɣ, δ) = ɣ/δ.

c is called α-discriminatory if elift(ɣ, δ) ≥ α. Thus we have to concentrate on the rules which are having elift greater than α.
In order to convert α-discriminatory rules into α-protective based on direct discriminatory measure (definition 2). We should enforce the following inequality for each α-discriminatory rule r0: A, B → C in MR, where A is a discriminatory item set:

elift (r0) <α

By using the statement of the elift definition, inequality above can be written as

conf (r0: A, B→C) / conf (B→C) < α. (2)

We can write inequality above as below:

conf (r0: A, B→C) < α. conf(B→C) (3)

It is clear that inequality (2) can be satisfied by decreasing the confidence of the α discriminatory rule r0 to a value less than the right hand side value of the inequality (3), without affecting the confidence of its base rule B→C. A possible solution to decrease conf (r0: A, B→C) is to modify the item set from ⌐A to A in the subset DB_c of the original data set which completely supports ⌐A,B→⌐C and will do minimum impact on the other rules [2].

There is another way to do so by increasing the confidence of the base rule. For this we have to make conversion of ⌐C to C which completely supports the rule ⌐A, B→⌐C. Thus we can remove the direct discrimination in our data mining.

4.2 Rule Generalization

In this method we are converting the original rule which is classified as direct discrimination rule into a rule which will come into indirect rule protection. It is based on the fact that if each α discriminatory rule r1: A, B→C in the database of the decision rule is the instance of at least non-redlining PND rule r: D, B→C, and then the original dataset would be set free from the direct discrimination. In order to formalize this dependency of the rules, Ruggieri et al [5] says that if PND rule r: D, B→C holds conf(r) ≥ conf (r1) where rule r1: A, B→C is the PD rule which will be instance of rule r where A is discriminatory item set and satisfies conf (r′: A, B →D) =1. The two conditions can be relaxed using definition given below.

Definition 3: Let p ϵ {0; 1}. A classification rule r0: A, B → C is
a p-instance of r : D, B → C if both conditions below are

true:

- Condition 1: conf(r)≥p. conf(r0)
- Condition 2: conf(r′′: A, B→C)≥p

If r0 is p instance of r and p is nearly or equal to 1 then r0 will be free from direct discrimination. Based on this rule [1] proposed some transformation in the r0 such that it will be free from the direct discrimination. For implementing rule generalization we have make sure that confidence of direct discriminatory rule r0: A, B→C should be less than rule r: D, B →C. To satisfy the condition ⌐C to C have to transformed which completely supports the rule A, B, ⌐D→C without having impact on the other rules.

4.3 Direct Rule Protection and Rule Generalization

When both direct rule protection and rule generalization are used then at that time *α* discriminated rule are classified into two categories:

- *α* discriminated rule r0: A, B →C for which their exist at least one PND rule so that condition 2 in definition 3 can be satisfied. For this we convert such rule into PND rules.
- *α* discriminated rule r0: A, B →C for which there is no PND rule then in that case we use direct rule protection.

Both these steps give us three more conditions two resolve direct discrimination which are related to definition 3.

- If there exist a discriminated rule r0 which is an instance of PND rule and p in definition is nearly equal to 1 then in that case there is no transformation in the rule.
- If there exist a discriminated rule r0 which is instance of PND and at least one rule is having its instance as r0 then we convert such rule by rule generalization.
- If there exist a discriminated rule r0 which is not having any instance of the r0 then such rule is converted by direct rule protection.

4.4 Indirect Rule Protection

Definition 4: [Redlining Rules] Redlining rules are those rules which are inferred from the background rules which contains background information about all generated rules which are having support than *α* (minimum support).

Until we have seen direct discrimination part which is removed by two techniques called direct rule protection and rule generalization. In this section discriminated rule in *MR* is related with non-sensitive attributes but still they are being part of discrimination because they are strongly correlated with the sensitive attributes.

Theorem 1: Let r: D, B → C be a PND classification rule, and let
ɣ=conf(r: D, B → C) δ=conf (B → C) > 0.
Let A be a discriminatory item set, and let _1, _2 such that
conf(r_{b1}: A, B → D)>β_1
conf(r_{b2}: D, B → A)≥β_2 > 0.
Call
f(x)=$\beta_1/\beta_2(\beta_2+x-1)$
elb(x, y) ={ f(x)/y if f(x)>0
0 otherwise
It holds that, for *α*≥ 0, if elb(ɣ, δ)≥*α*, the PD classification rule r0 : A, B→ C is *α*-discriminatory.

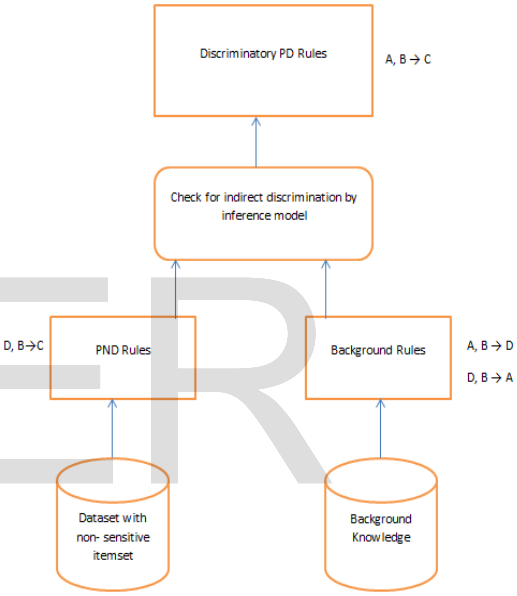

Figure 1 Indirect Discrimination Prevention

The figure 1 shows process of indirect discrimination prevention. In this process the nonredlining rules are removed by indirect rule protection (IRP). The IRP acts in such way that PND rules in our Discriminatory PD rule are converted into non redlining rule.

This can be done by decreasing the confidence of rule r_{b1}: A, B→D to values less than right hand side as in [1] without affecting both the confidence of the redlining rule or the base rule B→C and rule r_{b2}.

Above condition can be satisfied by transforming itemset ⌐A to A in the subset DB_c which completely supports the rule ⌐A, B, ⌐D→⌐C. Another way is to transform itemset ⌐C to C which completely supports the rule ⌐A, B, ⌐D→⌐C.

5 Post Processing Technique

This section reveals the discrimination prevention using post processing technique. This method actually works at the end of the mining of the database. The main idea of removing discrimination at post processing is by using Classification based on Predictive Analysis Rules (CPAR) algorithm instead of REPPER, FOIL, PRM and C4.5 since it is highly efficient than the rest [6].

CPAR algorithm is more efficient then FOIL (First Order Inductive Learner) and PRM (Predictive Rule Mining) algorithms, the basic difference in these strategies is in rule generation process. Foil generates rules which are not redundant but to achieve this, it loses some important rules. PRM extracts these rules but with cost of redundancy. Some rule may be extracted more than ones. CPAR also uses similar concept of PRM as to generate more rule with some redundant rules, but it can test more than one attribute at a time to judge whether this attribute can also give some useful rule or not. So more rules and less computation [4] is needed in CPAR for comparison to the PRM algorithm. To implement these algorithms, following three [5 steps are used;

1. Rule Generation.
2. Estimate Accuracy of rules.
3. Classification of rules
4. Result analyses.

The main difference between CPAR and PRM is that instead of choosing only one attribute to obtain best gain on each iteration(as in FOIL and PRM),CPAR choose a number of attributes if those attributes have similar best gain. This is done by applying GAIN_SIMILARITY_RATIO and by calculating the minimum gain.

CPAR takes input as (space separated) binary valued dataset R and produces a set of CARs. It also requires minimum gain constant which is user defined value, decay factor and TOTAL_WEIGHT_THRESHOLD. The resulting data is in the form of linked-list of rules ordered according to Laplace accuracy.

6 Conclusion

The motivation of this paper was to summarize the overall techniques for discrimination prevention at different stages. In this paper we have summarized preprocessing and postprocessing techniques for discrimination removal. In preprocessing technique direct rule protection, rule generalization and direct rule protection and rule generalization simultaneously are used for removing direct discrimination. For removal of indirect discrimination indirect discrimination protection technique which is based on background knowledge is used. Whereas in the post processing technique classification based on predictive association rule is used instead of other conventional algorithms.

Acknowledgment

We thank all who supported us during this survey. We specially thank Mrs Joshi for her valuable guidance and her invaluable suggestions in preparing this paper.

We also thank Mr M. Kshirsagar for his support during the survey of this paper. We also thank Mr A. K. Kureshi for providing such environment during this survey.

References

[1] A Methodology for Direct and Indirect Discrimination Prevention in Data Mining, Sara Hajian and Josep Domingo-Ferrer, IEEE transaction on knowledge and Data Engineering, Volume 25, No. 7, July 2013.

[2] Classifying without Discrimination, Faisal Kamiran and Toon Calders, IEEE xplore, 2009.

[3] Discrimination aware data mining, Dino Pedreschi and Salvatore Ruggieri and Franco Turini, ACM, 2008.

[4] Measuring Discrimination in Socially Sensitive Decision Records, Dino Pedreschi and Salvatore Ruggieri and Franco Turini, SIAM.

[5]Integrating Induction and Deduction for Finding Evidence of Discrimination, Dino Pedreschi and Salvatore Ruggieri and Franco Turini, ICAIL, 2009.

[6] CPAR: Classification based on predictive analysis Rules, Xiaoxin Yin and Jiawei Han.