International Journal of Computer Applications Technology and Research
Volume 3– Issue 7, 464 - 467, 2014, ISSN: 2319–8656

Classification with No Direct Discrimination

Deepali P. Jagtap
Department of Computer
Engineering,
KKWIEER, Nashik,
University of Pune, India

Abstract: In many automated applications, large amount of data is collected every day and it is used to learn classifier as well as to make automated decisions. If that training data is biased towards or against certain entity, race, nationality, gender then mining model may leads to discrimination. This paper elaborate direct discrimination prevention method. The DRP algorithm modifies the original data set to prevent direct discrimination. Direct discrimination takes place when decisions are made based on discriminatory attributes those are specified by the user. The performance of this system is evaluated using measures MC, GC, DDPP, DPDM etc. Different discrimination measures can be used to discover the discrimination.

Keywords: Classifier, Discrimination, Discrimination measures, DRP algorithm, Mining model.

1. INTRODUCTION

The Latin word 'Discriminare' is origin of the word Discrimination, its meaning is 'Distinguish between'. Discrimination is treating people unfairly based on their belonging to particular group. It restrict members of certain group from opporunities that are available to others [1]. Discrimination can also be observed in data mining. In data mining, large amount of data is collected and is used for training classifier. Classification model act as support to decision making process and the basis of scoring system. This makes business decision maker's work more easier [1,11]. If that historical data itself is biased towards or against certain entity such as gender, nationality, race, age group etc. Then resulting mining model may show discrimination. The use of automated decision making system gives sense of fair decision as it does not follow any personal preference but in actual results may be discriminatory [9]. Publishing such data leads to discriminatory mining results. The simple solution for discrimination prevention would consist neglecting or removing discriminating attributes, but even after removing those attribute, the discrimination may persist as many other nondiscriminatory attributes may strongly co-related with discriminatory attributes. The publicly available data may reveal co relation between them. Also removal of sensitive attribute results into loss of quality of original data [11].

Direct discrimination is observed when decisions are made based on the input data containing protected groups, whereas Indirect discrimination occurs when decisions are made based on nondiscriminatory input data but it is strongly or indirectly co-related with discriminatory one. For example, If discrimination occurs against foreign worker and even after removing that attribute from data set, one cannot guarantee that discrimination has been prevented completely as publicly available data may reveal nationality of that individual, hence shows indirect discrimination.

There are various laws against discrimination, but those are reactive not proactive. The use of technology and new mining algorithm helps to make them proactive. Along with mining algorithm, some algorithm and methods from privacy preservation such as data sanitization helps to prevent discrimination where original data is modified or support and confidence of certain attributes is changed to make them discrimination free. This system is useful in various applications such as credit/insurance scoring, lending, personnel selection and wage discrimination, job hiring, crime detection, activities concerning public accommodation, education, health care and many more. Benefit and services must be made available to everyone in a nondiscriminatory manner.

Rest of the paper is organized as follows. Section one provides introduction, survey of literature along with pros and cons of some of the existing methods are discussed in section two. Section three highlights basic terminology associated with this topic and section four describes algorithm as well as block diagram for discrimination prevention. Section five contains results and discussion about data set and finally last section presents conclusion along with the future scope of system.

2. LITERATURE SURVEY

Various studies have been reported the discrimination prevention in the field of data mining. Pedreschi noticed the discriminatory decisions in data mining based on classification rule and discriminatory measure. The work in this area can be traced back from year 2008. S. Ruggieri, Pedreschi and F. Turini [14] have implemented the DCUBE. It is oracle based tool to explore discrimination hidden in data. Discrimination prevention can be done in three ways based on when and in which phase data or algorithm is to be changed. Three ways for Discrimination prevention are: Preprocessing method, Inprocessing method and Postprocessing method. Discrimination can be of 3 types: Direct, Indirect or combination of both, based on presence of discriminatory attributes and other attributes that are strongly related with discriminatory one. Dino Pedreschi, Salvatore Ruggieri, Franco Turini[3] has introduced a model used in Decision Support System for the analysis and reasoning of discrimination that helps DSS owners and control authorities in the process of discrimination analysis [3].

Discrimination Prevention by Preprocessing Method

In preprocessing method, the original data set is modified in such a manner that it will not result in discriminatory classification rule. In this method any data mining algorithm can be applied to get mining model. Kamiran and Calder[4] proposed a method based on "data massaging" where class label of some of the records in the dataset is changed but as this method is intrusive, concept of "Preferential sampling" was introduced where distribution of objects in a given dataset is changed to make it non-discriminatory[4]. It is based on the idea that, "Data objects that are close to the decision boundary are more vulnerable to be victim of discrimination." This method uses Ranking function and there is no need to change the class labels. This method first divides data into 4 groups that are DP, DN, PP, PN, where first letters D and P indicate Deprived and Privileged class respectively and second letters P, N indicates positive and negative class label. The ranker function then sorts data in ascending order with respect to positive class label. Later it changes sample size in respective group to make that data biased free. Sara Hajian and Josep Domingo-Ferrer[9] proposed another preprocessing method to remove direct and indirect discrimination from original dataset. It employees 'elift' as discrimination measure to prevent discrimination in crime and intrusion detection system[10].

Preprocessing method is useful in applications where data mining is to be performed by third party and data needs to publish for public usage [9].

Discrimination Prevention by Inprocessing Method

Faisal Kamiran, Toon Calders and Mykola Pechenizkiy [5] introduced algorithm based on inprocessing method using decision Tree where instead of modifying original dataset data mining algorithm is modified. This approach consists of two techniques for the decision tree construction process, first is Dependency-Aware Tree Construction and another is Leaf Relabeling. The first technique focuses on splitting criterion for tree construction to build a discrimination-aware decision tree. In order to do so, it first calculates the information gain with respect to class & sensitive attribute represented by IGC and IGS respectively. There are three alternative criteria for determining the best split that uses different mathematical operation: (i) IGC-IGS (ii) IGC/IGS (iii) IGC+IGS. The second approach consists of processing of decision tree with discrimination-aware pruning and it relabel the tree leaves [5]. This methods requires special purpose data mining algorithms.

Discrimination Prevention by Postprocessing Method

Sara Hajian, Anna Monreale, Dino Pedreschi ,Josep Domingo Ferrer[12] proposed algorithm based on postprocessing method that derive frequent classification rule and modifies the final mining model using α-Protective k-Anonymous pattern sanitization to remove discrimination from Mined Model. Thus in postprocessing method, resultant mining model is modified instead of modifying original data or mining algorithm. The disadvantage of this method is, it doesn't allow original data to be published for public usage, and also the task of data mining should be performed by data holder only. Toon Calders and Sicco Verwer[15] proposed approach where the Naive Bayes classifier is modified to perform classification that is independent with respect to a given sensitive attribute. There are three approaches in order to make the Naive Bayes classifier discrimination-free: (i) modifying the probability of the decision being positive where the probability distribution of the sensitive attribute is modified. This method has disadvantage of either always increasing or always decreasing the number of positive labels assigned by the classifier, depending on how frequently the sensitive attribute is present in dataset, (ii) training one model for every sensitive attribute value and balancing them. This is done by splitting the dataset into two separate sets and the model is learned using only the tuples from the dataset that have a favored sensitive value, (iii) adding a latent variable to the Bayesian model. This method models the actual class labels using a latent variable[15].

3. THEORETICAL FOUNDATION

The basic terms in data mining are described in short as below:

3.1 Discrimination Measures

a. elift

Pedreschi [2] introduced '*elift*' called extended lift as one of the discrimination measure. For a given classification rule, Extended lift can be calculated as below. Elift provides gain in confidence due to presence of discriminatory item [1].

$$elift\ (A, B \rightarrow C) = \frac{\text{Confidence}\ (A, B \rightarrow C)}{\text{Confidence}\ (B \rightarrow C)}$$

b. slift

The selection lift i.e. '*slift*' for a classification rule of the form (A, B → C) is given as,

$$slift\ (A, B \rightarrow C) = \frac{\text{Confidence}\ (A, B \rightarrow C)}{\text{Confidence}\ (\neg A, B \rightarrow C)}$$

c. glift

Pedreschi [2] introduced '*glift*' to strengthen the notion of α-protection. For a given classification rule glift is computed as,

$$glift(\beta, \gamma) = \begin{cases} \beta/\gamma & \text{if } \beta \geq \gamma \\ (1-\beta)/(1-\gamma) & \text{otherwise} \end{cases}$$

3.2 Direct discrimination

Direct discrimination consists of rules or procedures that explicitly mention disadvantaged or minority groups based on sensitive discriminatory attributes [9]. For example, the rule r: (Foreign_worker = Yes, City = Nasik → Hire = No) shows direct discrimination as it contains discriminatory attribute Foreign_worker = yes.

3.3 Indirect discrimination

Indirect discrimination consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, intentionally or un-intentionally could generate discriminatory decisions [9], for example, the rule r: Pin_code = 422006, City = Nashik → Hire = No shows indirect discrimination, as attribute Pin_code corresponds to area with mostly people belonging to particular religion.

3.4 PD Rule

A classification rule is said to be Potentially Discriminatory rule if it contains discriminatory item in premise of a given rule.

3.5 PND Rule

A classification rule is said to be Potentially Non-discriminatory rule if it doesn't contains any discriminatory item in premise of a given rule [1, 9].

4. DISSERTATION WORK

4.1 Algorithm and Process flow

The diagram showing overall process flow of discrimination prevention is shown in fig 1. The system takes original dataset containing discriminatory items as an input.

4.1.1 Input Data Preprocessing

The original dataset contains numerical values so it should be preprocessed i.e. discretization is performed on some attributes.

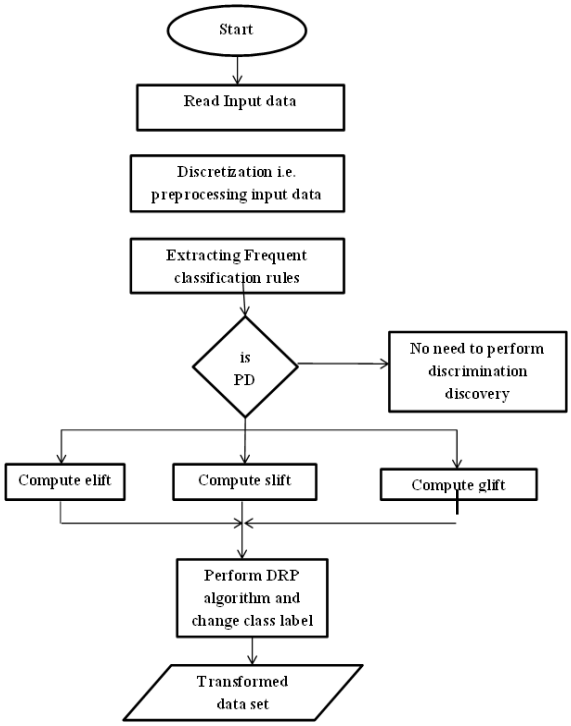

Fig. 1. Flow chart for Discrimination Prevention

4.1.2 Frequent Classification Rule extraction

Later on using Apriori algorithm frequent item sets are generated. In Apriori algorithm candidate set generation and pruning steps are performed. The resultant frequent item sets are used to generate frequent classification rules.

4.1.3 Discrimination Discovery Process

The frequent classification rules are then categorized into Potentially Discriminatory and Potentially Nondiscriminatory groups in discrimination discovery. For discrimination discovery each classification rule is examined and is placed into either PD or PND group based on presence of discriminatory items in premise of the rule. If the rule found to be PD then for every PD rule elift, glift and slift is calculated. If that calculated value is greater than threshold value (α) then that rule is considered as α-discriminatory.

4.1.4 Data Transformation using DRP Algorithm

Data transformation is carried out in the next step as α-discriminatory rules need to be treated further to remove discrimination where class label of some of the records is perturbed to prevent discrimination. As a result of above process finally the transformed dataset is obtained as an output.

Data transformation is second step in discrimination prevention where the data is actually modified to make it biased free. In this step modifications are done using the definition of elift/glift/slift i.e. equality constraint on rule are enforced to satisfy the definition of corresponding discrimination prevention measure.

Direct Rule Protection (DRP) algorithm is used here that converts α-discriminatory rules into α-protective rule using the definition of elift. It can be done in following way:

let r': α-discriminatory rule, condition enforced on r' is:

= elift(r') < α

$$= \frac{\text{Confidence}(A,B \rightarrow C)}{\text{Confidence}(B \rightarrow C)}$$

= confidence(r': A, B→C)/confidence(B→C) <α

= confidence(r': A, B→C/α <confidence (B→ C)

Here one needs to increase confidence (B→C), so change the class item from ¬C to C for all records in original DB that supports the rule of the form(¬A,B → ¬C). In this way this method changes the class label of class item in some records[9]. Similar method for slift as well as glift can be carried out.

4.2 Performance measures

To measure the success of the method in removing all evidence of Direct Discrimination and to measure quality of the modified data, following measures are used:

4.2.1 Direct discrimination prevention degree (DDPD)

The DDPD counts the percentage of α-discriminatory rules that are no longer α-discriminatory in the transformed data set.

4.2.2 Direct discrimination protection preservation (DDPP)

This measure counts the percentage of the α-protective rules in the original data set that remain α-protective in the transformed data set.

4.2.3 Misses cost (MC)

This measure helps to find the percentage of rules that are extractable from the original data set but cannot be extracted from the transformed data set. This is considered as side effect of the transformation process.

4.2.4 Ghost cost (GC)

This ghost cost quantifies the percentage of the rules that are extractable from the transformed data set but were not extractable from the original data set.

This MC and GC are the measures that are used in the context of privacy preservation. As similar approach of data sanitization is used in some methods for discrimination prevention, the same measures that are MC and GC can be applied to find out the information loss [16].

5. RESULTS AND DISCUSSION

German Credit Data set

This data set consists of 1000 records as well as 20 attributes. Out of those 20 attributes 7 are numerical and remaining 13 are categorical attributes. The class attributes indicates good or bad class for given bank account holder. Here the attribute foreign worker = Yes, Personal status = Female but not single and age = old are considered as discriminatory items where age > 50 is considered as old.

The Table I show the partial results computed on German credit dataset containing total number of classification rules generated and number of α-Discriminatory rules and the number of lines modified in original data set.

Table 1. German Credit dataset: Columns show the partial results for No. of α-Discriminatory rules, No. of lines modified.

Total No. of Classification rules	No. of α-Discriminatory rules	No. of Lines modified
9067	45	49

6. CONCLUSION

It is very important to remove discrimination, which can be observed in data mining, from original data. The removal of discriminatory attributes does not solve the problem. In order to prevent such discrimination, Discrimination Prevention by preprocessing technique is advantageous over the other two methods. The approach mentioned in this paper works in two steps: first is the discrimination discovery where α-Discriminatory rules are extracted and then in second step data transformation is performed in which the original data is transformed to prevent direct discrimination. This second step follows similar approach of Data Sanitization that is used in privacy preservation context. Many such algorithms uses 'elift' as a measure of discrimination, but instead of that one may use slift, glift as a measure of discrimination. The performance measure metrics i.e. DDPD, DDPP, MC, GC analyses data to check quality of transformed data as well as presence of direct discrimination. The less number of classification rules will be extracted from transformed data set as compare to original data set. The use of different discrimination measures such as slift, glift results into varying number of discriminatory rules and it have varying impact on original data.

In the future, one may explore how rule hiding in privacy preservation or other privacy preserving algorithms helps to prevent discrimination.

7. ACKNOWLEDGMENTS

With deep sense of gratitude I thank to my guide **Prof. Dr. S. S. Sane**, Head of Department of Computer Engineering, (KKWIEER, Nashik, University of Pune, India) for guiding me and his constant support and valuable suggestions that have helped me in the successful completion of this paper.

8. REFERENCES

[1] 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. (Cited by 56)

[2] D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring Discrimination in Socially-Sensitive Decision Records," Proc. Ninth SIAM Data Mining Conf. (SDM '09),pp. 581-592, 2009.

[3] D. Pedreschi, S. Ruggieri, and F. Turini, "Integrating Induction and Deduction for Finding Evidence of Discrimination,"Proc. 12thACM Int'l Conf. Artificial Intelligence and Law (ICAIL '09), pp. 157-166, 2009.

[4] F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling,"Proc. 19th Machine Learning Conf. Belgium and TheNetherlands, 2010.

[5] 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.

[6] M.Kantarcioglu, J. Jin and C. Clifton. When do data mining results violate privacy? In KDD 2004, pp. 599-604. ACM, 2004.

[7] P. N. Tan, M. Steinbach and V. Kumar, “Introduction to Data Mining”.Addison-Wesley, 2006.

[8] 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.

[9] S. Hajian and J. Domingo, "A Methodology for Direct and Indirect Discrimination prevention in data mining." IEEE transaction on knowledge and data engineeing, VOL. 25, NO. 7, pp. 1445-1459, JULY 2013. (Cited by 12)

[10] S. Hajian, J. Domingo-Ferrer, and A. Martnez Balleste, “Discrimination Prevention in Data Mining for Intrusion and Crime Detection,"Proc. IEEE Symp. Computational Intelligence in Cyber Security (CICS '11), pp. 47-54, 2011.

[11] S. Hajian, J. Domingo-Ferrer, and A. Martı´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,

[12] S. Hajian, A. Monreale, D. Pedreschi, J. Domingo-Ferrer and F. Giannotti. “Injecting discrimination and privacy awareness into pattern discovery,” In 2012 IEEE 12th International Conference on Data Mining Workshops, pp. 360-369. IEEE Computer Society, 2012.

[13] S. Ruggieri, D. Pedreschi and F. Turini. “Data mining for discrimination discovery,” ACM Transactions on Knowledge Discovery from Data (TKDD), 4(2), Article 9, 2010.

[14] S. Ruggieri, D. Pedreschi, and F. Turini, "DCUBE: Discrimination Discovery in Databases,"Proc. ACM Int’l Conf. Management of Data (SIGMOD’10), pp. 1127-1130, 2010.

[15] T. Calders and S. Verwer. “Three naive Bayes approaches for discrimination-free classification,” Data Mining and Knowledge Discovery, 21(2):277-292, 2010. (Cited by 44)

[16] 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.