RACISM Prevention in Data Mining

S. P. Santhoshkumar
Assistant Professor, Department of CSE,
SNS College of Technology, Coimbatore, India

Abstract-Data mining is a technique used to extract information from dataset and transforms it into understandable structure. In classification, discrimination is the major aspects in data mining. Discrimination is a type of treatment that includes denying the membership in one group opportunities that are available in another group. Discrimination based on age, religion, gender, caste, disability, employment, language, race and nationality. Automated collections of data are used to define classification rules by means of making automated decisions like loan granting, insurance premium and jobs. If the training data sets are biased against particular community, the transformed models also show discriminatory prejudiced behavior. Antidiscrimination techniques including discrimination discovery and prevention are used to prevent discrimination. Discrimination can be classified into direct discrimination and indirect discrimination. In direct discrimination, the extracted rules can be directly obtained by means of searching discriminatory contexts. Indirect discrimination consists of rules or procedures that not explicitly mentioning discriminatory attributes intentionally could generate discriminatory decisions. In proposed technique, direct and indirect discrimination is prevented using rule protection and rule generalization methods. In this process, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) is used to perform hierarchical clustering for huge data-sets. The proposed method has been implemented on the Adult dataset and shown promising results.

Keywords- Data mining, Classification, Direct discrimination, Indirect discrimination, rule protection, data transformation, birch algorithm, rule generalization.

I. INTRODUCTION

Discrimination involves the group's initial reaction that influencing the individual's actual behavior towards the group, restricting members of one group from privileges that are available to another group, leading to the rejection of the individual or entities based on logical decision making. Discrimination based on age, religion, gender, caste, disability, employment, language, race and nationality. There are several decision-making tasks which made them to discrimination, e.g. loan granting, education and health insurances. Given a set of information items on a customer, an automated system decides whether the customer is to be recommended for a credit or a certain type of life insurance.

Automating such decisions reduces the workload of the staff of banks and insurance companies, among other organizations. Age discrimination is discrimination that depends on beliefs and values which used to justify discrimination and subordination based on someone's age. Ageism defines that it directed towards old people, or adolescents and children. Disability discrimination is the process of individuals which treats as the standard of usual living that ends in public and private places, education, and social work that are built to survive best people, thereby rejecting those with various disadvantages. Denying someone job opportunities, or disallowing one from applying for particular jobs, is often considered as employment discrimination for such a rejection is not related to the requirements protected characteristics may include age, disability, ethnicity, weight, religion, gender, gender identity, height, nationality, gender orientation and skin color. In the beginning, automating decisions may give a sense of fairness but the decision rule does not learn itself by personal preferences. The classification rules are actually learned by the system model based on historical data. If the original data are inherently biased against a particular community, the learned model may also show the negative impact on it. For example, in a certain loan granting organization, foreign people are rejected for loan for the years. If this biased historical dataset is used as training data to learn classification rules for an automated loan granting system, the learned rules also suffered from biased behavior toward foreign people and also the system may consider that the foreign is a legitimate criterion for loan rejection.

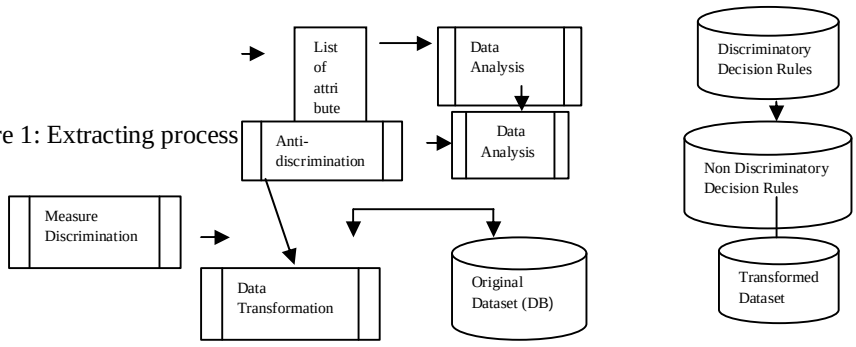
II. RELATED WORK

Numerous direct and indirect discrimination schemes have been proposed previously. Those schemes either eliminate direct or indirect discrimination. Fast algorithms for mining association rules that defines the issues of discovering association rules between items in a large database of sales transactions[1]. The proposed algorithms can be combined into a hybrid algorithm named as AprioriHybrid.

Figure 1 describes the process of extracting biased and unbiased decision rules. Data mining with discrimination sensitive and privacy sensitive attributes states that the privacy legitimate input data and the output data are used for selecting preserving protection against such discrimination. In this technique it describes the work in progress of research project based on legal and ethical rules can be combined in data mining algorithms to prevent such activities.

Toon Calders investigated that to modify the naive Bayes classifier in order to perform classification that is restricted to be independent with respect to a given sensitive attribute [2]. Such independency restrictions occur naturally when the decision process leading to the labels in the dataset was biased due to gender or racial discrimination. Preferential sampling introduced the idea of Classification with No Discrimination (CND*)* and proposed a solution based on "massaging" the data to remove the discrimination from it with the least possible changes [6]. Automatic Decision Support Systems (DSS) are used for cleaning purposes in socially sensitive tasks that includes access to credit, denial, insurance, and job opportunities. While less decisions can be approved, automatic DSS can still be legitimate in the socially negative sense of resulting in unequal treatment of people [11].It provides a guarantee of non-discrimination is treated to be an unusual task. A naıve approach, like taking away all discriminatory attributes, is shown to be not suitable when other background knowledge is available in databases [9].

Figure 1: Extracting process



Measuring discrimination can handle the problem that the historical training datasets of values of discrimination can be suffered by a given legitimate groups by means of decisions [10]. This problem is revised in a classification rule based process and quantitative measures of discrimination are introduced, on the basis of existing regulations. Calders states that in which the non-discriminatory constraint is taken deeply into a decision tree learner by changing its splitting process and pruning strategy by using a re-labeling approach [7]. Data mining detects the problem of discovering discrimination in a dataset of historical decision records by automatic systems and formulates the processes of direct and indirect discrimination discovery by modeling suitable groups and contexts that discrimination occurs in a classification rule based regulations [12].

Indirect discrimination occurs when automated decisions are taken, based on non-sensitive attributes that are correlated with biased sensitive attributes. In order to prevent indirect discrimination in a dataset, a first step consists in discovering whether database contains indirect discrimination. If any legitimate rules are found, the dataset is modified until discrimination is brought below a certain threshold or is entirely removed [4].

Sara Hajian investigated that how to clean training datasets and outsourced datasets in such a way that legitimate classification rules can still be extracted but discriminating rules based on sensitive attributes cannot[3]. To overcome this issue, anti-discrimination process is introduced. It includes discrimination discovery and prevention techniques. Discrimination discovery is based on mining classification rules and reasoning on them on the basis of quantitative and qualitative measures of discrimination [15]. This approach has been extended to encompass statistical significance of the extracted patterns of discrimination and reason about affirmative action. Discrimination prevention represents a set of patterns that do not allow for discriminatory decisions even if the historical data sets are discriminated. It involves three approaches:

1. *Pre-processing*
2. *In-processing*
3. *Post-processing*

In this paper, we concentrate on preprocessing techniques.

III. CLASSIFICATION BASED ON DISCRIMINATION PREVENTION USING DATA TRANSFORMATION TECHNIQUES

Classification is the task of generalizing known structures applies to new data. Classification is supervised learning. For example, classes are used to represent that a customer defaults on a loan decisions like 'Yes' or' No'. It is important that each record in the dataset used to represent the classifier already have a value for the attribute used to describe classes. Because each record has the attribute value used to define the classes. Classification is a machine learning technique used to predict group membership for data instances. It assigns items in a collection to target categories. The aim of classification is to accurately determine target class for each and every case in data. Direct discrimination restricts a particular community based on sensitive reasons. Indirect discrimination restricts certain number of peoples based on non sensitive one.

1. DISCRIMINATION MEASUREMENT

The purpose of Discrimination measurement is to identify discriminatory rules and redlining rules using Potentially Discriminatory (PD) and potentially non-discriminatory (PND) rules. Direct discrimination is measured by identifying α-discriminatory rules among the PD rules using a direct discrimination measure (elift) and threshold (α).

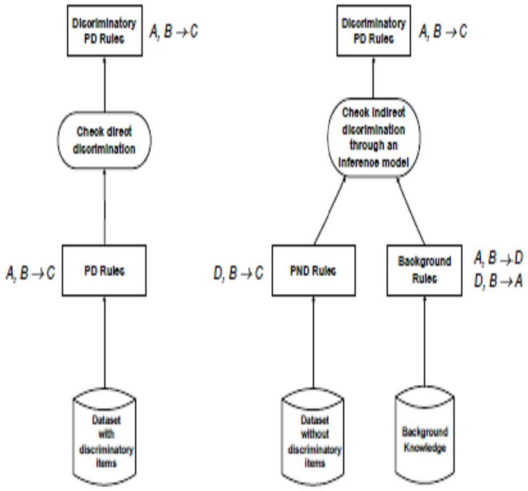


Figure 2: Process of direct discrimination and indirect discrimination

The extended lift of the rule can be calculated as

$$\mathbf{elift}(A, B \rightarrow C) = \frac{\mathbf{Conf}(A, B \rightarrow C)}{\mathbf{Conf}(B \rightarrow C)}$$

Figure 2 defines the process of direct discrimination and indirect discrimination that generate negative impacts.. The indirect discrimination is measured by identifying redlining rules among the PND rules that correlated with background knowledge based on an indirect discriminatory measure (elb), and a discriminatory threshold (α). In fact, redlining rules indicate biased rules that are indirectly inferred from nondiscriminatory items because of their correlation with discriminatory ones.

2. DATA TRANSFORMATION

Transform the original data DB in such a way to remove direct and indirect discriminatory biases, with minimum impact on the datasets. So there is no other negative impact can be discovered from transformed datasets[16].

The data transformation method should increase or decrease the confidence of the rules to the target values with minimum impact on data quality, maximize the disclosure prevention measures and minimize the information loss measures. Data transformation includes rule protection and rule generalization methods for both f-direct and indirect discrimination.

2.1 RULE PROTECTION ALGORITHM FOR DIRECT AND INDIRECT DISCRIMINATION

In direct discrimination, rule protection algorithm is used to convert each α-discriminatory rule into a α-protective rule based on the direct discriminatory measure. There are two methods that could be applied for direct rule protection.

1. Method 1 modifies the discriminatory item set in some records.

2. Method 2 modifies the class item in some records from grant credit to deny credit in the records with male gender.

Indirect rule protection is used to turn a redlining rule into a non-redlining rule, based on the indirect discriminatory measure. Rules that are associated with some background knowledge indirectly called as redlining rules.

RULE PROTECTION ALGORITHM

Input : Original data set DB
Output: Transformed data set DB'
1: For each r': A, B->C € MR do
// MR- database of direct discriminatory rules
2: FR<- FR-{r'}
// FR-database of indirect discriminatory rules
3: if $MR_{r'}$ = RG // then rule generalization
4: DB_c<- All records completely supporting ¬A
5: B->¬C
6: For each db_c€ DB_c do
7: Compute impact (db_c=|{r_a€ FR| db_c supports r_a}|
8: Sort DB_c by ascending impact
9: While conf(r') ≥ α.conf(B->C) do
10: Select first record in DB_c
11: Modify discriminatory item set of
db_c from ¬A to A in DB
12: Recompute conf(r')

For each direct α-discriminatory rule in MR, after finding the subset, records in DB_c should be changed until direct rule protection requirement met for each respective rule. For each record, the number of rules whose premise is supported is taken as the impact. If conf(r') ≥ α.conf (B->C), confidence of rule is greater than the discriminatory threshold can be considered as minimum impact. Then the records in db_c with minimum impact are selected for change.

2.2 RULE GENERALIZATION

Rule generalization is based on the fact that if each α-discriminatory rule r': A, B -> C in the database of decision rules was an instance of at least one non-redlining PND rule r: D, B -> C, the data set would be free of direct discrimination. In rule generalization, it considers the relation between rules instead of discrimination measures. A classification rule {Foreign worker = Yes, City = NYC} -> Hire = No with high elift supports the complainant's claim. The decision maker could argue that this rule is an instance of a general rule {Experience= Low, City=NYC} -> Hire =No. Usually foreign workers are rejected because of their low experience, not only they are foreign.

3. BIRCH ALGORITHM

BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) is an unsupervised data mining algorithm used to perform clustering in discrimination environment. It can be used in multi-dimensional datasets and it has minimized I/O cost than Apriori algorithm (1 or 2 scans). BIRCH has two concepts: CF and CF tree. CF (Clustering Feature) is represented as a triple CF = (N, LS, SS) N defines the number of points, LS represents linear sum of points and SS defines the square sum of points in cluster [14].

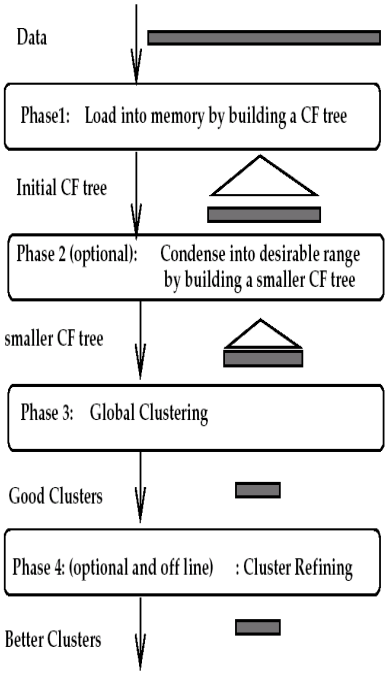


Figure 3: Birch Algorithm

First, it scans the data set and construct clustering feature tree in its memory as described in Figure 3. Then it condenses large clustering feature tree into smaller one and performs global clustering by using its centroid points. Finally it does cluster refining one more time for removing outliers. BIRCH algorithm can be divided into two phases: It scans the transformed data set in memory and generate model based on eligible and not eligible criteria as shown in Figure 5.

IV. RESULTS

Net Beans is an integrated development environment (IDE) for developing primarily with Java. It is also an application platform framework for Java desktop applications.

Net Beans IDE is an open-source integrated development environment. Net Beans IDE supports development of all Java application types (Java SE (including Java FX), Java ME, web and Enterprise Java Beans (EJB).

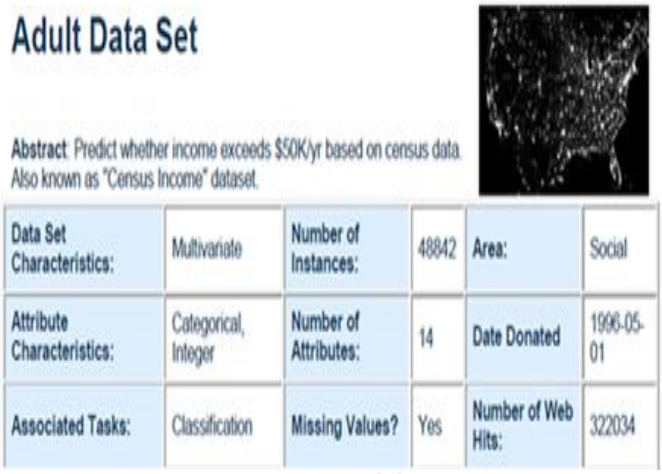


Figure 4: Adult Dataset

ADULT DATA SET

We used the Adult data set [8], also known as Census Income, in our experiments. Figure 3 describes adult data set consists of 48,842 records, split into a "train" part with 32,561 records and a "test" part with 16,281 records as shown in Figure 4.. The data set has 14 attributes. The prediction task associated with the Adult data set is to determine whether a person makes more than 50K$ a year based on census and demographic information about people.

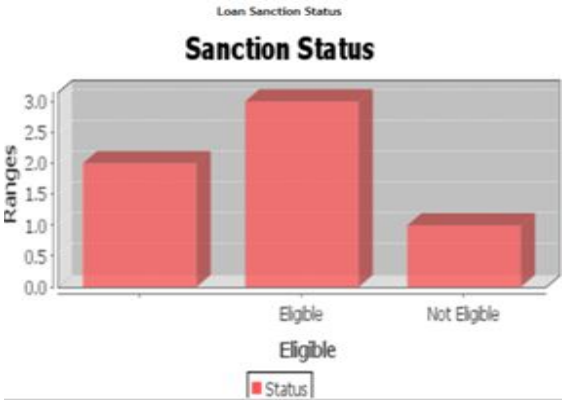


Figure 5: Loan sanction status based on eligible and not eligible criteria

V. CONCLUSION

The purpose of this paper was to develop a new preprocessing discrimination prevention technique that consists of various data transformation methods includes rule protection and rule generalization used to prevent direct discrimination and indirect discrimination To obtain this goal, the first step is to measure discrimination and identify the groups of individuals that have been directly and indirectly discriminated in the automated decisions processes; the second step is to transform data in the proper way to eliminate all those legitimate biases. Finally, discrimination- free data models can be produced from the transformed data set without damaging data quality. The experimental results reported demonstrate that the proposed techniques are efficient in both goals of removing discrimination and preserving data quality.

VI. REFERENCES

[1] Sara Hajian and Josep Domingo-Ferrer, "A Methodology for direct and indirect discrimination in data mining", Knowledge and Data Engineering, vol. 25, no. 7, pp. 1445-1459, 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] 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.

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

[5] 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.

[6] F. Kamiran and T. Calders, "Classification without Discrimination", Proc. IEEE Second Int'l Conf. Computer, Control and Comm.(IC4 '09), 2009.

[7] F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling", Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 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] R. Kohavi and B. Becker, "UCI Repository of Machine Learning Data bases", http://archive.ics.uci.edu/ml/datasets/Adult,1996.

[10] 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.

[11] 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.
[12] 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.
[13] 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.
[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] Rui Liu, Xiao-long Qian and Shu Mao,” Research on Anti-Money Laundering Based on Core Decision Tree Algorithm”, Control and Decision Conference (CCDC),pp .4322 – 4325, 2011.
[16] P.N.Tan, M. Steinbach and V. Kumar , “Introduction to Data Mining.”, Addison-Wesley, 2006.