IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. I (Jan. 2014), PP 60-65
www.iosrjournals.org

Protection of Direct and Indirect Discrimination using Prevention Methods

S. Rajeswari[1], R. Poonkodi[2], Dr.C. Kumar CharliePaul[3]

Abstract: *Along with privacy, discrimination is a very important issue when considering the legal and ethical aspects of data mining. It is more than observable that the majority people do not want to be discriminated because of their gender, nationality, religion, age and so on, particularly when those aspects are used for making decisions about them like giving them a occupation, loan, insurance, etc. determining such possible biases and eliminating them from the training data without harming their decision-making utility is therefore extremelypopular. For this reason, antidiscrimination methods containing discrimination detection and prevention have been introduced in data mining. Discrimination prevention consists of suggestmodels that do not lead to discriminatory decisions even if the original training datasets are essentially biased. In this section, by focusing on the discrimination prevention, we present taxonomy for classifying and examining discrimination prevention schemes. Then, we begin a group of pre-processing discrimination prevention schemes and indicate the special features of each approach and how these approaches deal with direct or indirect discrimination. A production of metrics used to estimate the performance of those approaches is also specified. In conclusion, we finish our learn by specifying interesting future directions in this research body.*

I. Introduction

In social sense, discrimination refers to an action basedon prejudice resulting in unfair treatment of people,where the distinction between people is operated onthe basis of their membership to a category or minority, without regard to individual merit or circumstances.Examples of social discrimination include racial/ethnic,religious, gender, nationality, disability, and age-related discrimination; a large body of international laws and regulations prohibit discrimination in socially-sensitive decision making tasks, including credit scoring/approval, house lending, and personnel selection. In order to prove (or disprove) a discrimination charge before a court, or to perform a social analysis of discrimination in a given context, it isclearly needed to rely on quantitative measures of thephenomenon under study: for this reason, discrimination has been the subject of a large body of research inlegal, economic and social sciences, as well as the subject of empirical analysis in a large number of juridicalcases.

For example, the EuropeanUnion implements the principle of equal treatment betweenmen and women in the access to and supply of goods andservices in [3] or in matters of employment and occupationin [4]. Although there are some laws against discrimination,each and every one are reactive, not positive. Technology can includeproactivity to legislation by contributing discriminationdiscovery and prevention techniques.

Services in the information society allow for automatic androutine collection of large amounts of data. Those data areoften used to train association/classification rules in view ofproduction automated decisions, similar to loan granting/denial,insurance premium computation, personnel selection, etc.At initialprospect, automating decisions may provide a sense offairness: classification rules do not guide themselves bypersonal preferences. Though, at anearlierseem, one realizesthat classification rules are actually learned by the system(e.g., loan granting) from the training data. If the training dataare essentially biased for or against a particular community(e.g., foreigners), the learned model may show a discriminatorybiasedactivities. In additionalterms, the scheme mayassume that just being foreign is a legitimate reason for loandenial. Discovering such potential biases and eliminatingthem from the training data without harming their decisionmakingutility is therefore highly desirable. One mustprevent data mining from becoming itself a source ofdiscrimination, owing to data mining responsibilities producing discriminatorymodels from biased data sets as part of theautomated decision making. In [12], it is demonstrated thatdata mining can be both a source of discrimination and ameans for discovering discrimination.

Discrimination can be either direct or indirect (also calledsystematic). Direct discrimination consists of rules orprocedures that explicitly mention minority or disadvantagedgroups based on sensitive discriminatory attributesrelated to group membership. Indirect discriminationconsists of rules or methods that, while not explicitlypoint out discriminatory aspects, purposely or unintentionallycould generate discriminatory decisions. Redliningby financial institutions (refusing to grant mortgagesor insurances in urban areas they consider as deteriorating)is an archetypal example of indirect discrimination,although certainly not the only one. With a slight abuse oflanguage for the sake of compactness, in this paper indirectdiscrimination will also be referred to as redlining and rulescausing indirect discrimination will be called redlining policies. Indirect discrimination

could occursince of theaccessibility of some background knowledge (rules), forexample, that a certain zip code corresponds to a deterioratingarea or an area with mostly black population. Thesurroundingsinformation might be available from publiclyavailable data (e.g., census data) or might be obtained fromthe original data set itself because of the existence ofnondiscriminatory attributes that are highly correlated withthe sensitive ones in the original data set.

II. Related Work

The discovery of discriminatory decisions was firstproposed by Pedreschi et al. [12], [15]. The approach isbased on mining classification rules (the inductive part) andreasoning on them (the deductive part) on the basis ofquantitative measures of discrimination that formalize legaldefinitions of discrimination. For instance, the US Equal PayAct [16] states that: a selection rate for any competition, gender, ornational group which is less than four-fifths of the rate for thegroup with the highest rate will generally be considered asevidence of difficultconflict. This evolution has beenabsolute to include statistical importance of the extractedpatterns of discrimination in [13] and to reason aboutaffirmative action and favoritism [14]. Moreover it has beenimplemented as an Oracle-based tool in. Currentdiscrimination discovery methods consider each rule individuallyfor measuring discrimination without consideringother rules or the relation between them. However, inthis paper we also take into account the relation betweenrules for discrimination discovery, based on the existence ornonexistence of discriminatory attributes.

Discrimination prevention, the other major antidiscriminationaim in data mining, consists of inducing patterns thatdo not lead to discriminatory decisions even if the originaltraining data sets are biased. Three approaches areconceivable:

Preprocessing:
Preprocessing approaches of data sanitization and hierarchy-basedgeneralization from the privacy-preserving literature. Along this line, adopts a controlled distortionof the training set.Transform the source data in such away that the discriminatory biases contained in theoriginal data are removed so that no unfair decisionrule can be mined from the transformed data andapply any of the standard data mining algorithms.The preprocessingapproach is useful for applications in which adata set should be published and/or in which datamining needs to be performed also by external parties(and not just by the data holder).

In-processing:
Change the data mining algorithms insuch a way that the resulting models do not containunfair decision rules. For example, an alternativeapproach to cleaning the discrimination from theoriginal data set is proposed in [2] whereby thenondiscriminatory constraint is embedded into adecision tree learner by changing its splittingcriterion and pruning strategy through a novel leafrelabeling approach. But, it is observable that in-processingdiscrimination prevention methods mustrely on new special-purpose data mining algorithms;standard data mining algorithms cannot be used.

Post-processing:
Modify the resulting data miningmodels, instead of cleaning the original data set orchanging the data mining algorithms. For example,in [13], a confidence-altering approach is proposedfor classification rules inferred by the CPAR algorithm.The post-processing approach does not allowthe data set to be published: only the modified datamining models can be published (knowledge publishing),hence data mining can be performed by thedata holder only.

Although some methods have already been proposedfor each of the above-mentioned approaches (preprocessing,in-processing, post-processing), discrimination preclusioncontinues a mostly unexplored research avenue. In thispaper, we concentrate on discrimination prevention basedon preprocessing, because the preprocessing approachseems the most flexible one: it does not require changingthe standard data mining algorithms, unlike the in-processingapproach, and it allows data publishing (ratherthan just knowledge was publishing), unlike the post-processingapproach.

III. Prevention Of Direct And Indirect Discrimination

We present our approach, including the datatransformation methods that can be used for direct and/orindirect discrimination prevention. For each method, itsalgorithm and its computational cost are specified.

Direct and indirect discriminationprevention can be described in terms of two phases:

Discrimination measurement

Direct and indirectdiscrimination discovery includes identifyingα-discriminatory rules and redlining rules. To thisend, first, based on predetermined discriminatoryitems in DB, frequent classification rules in FRare divided in two groups: PD and PND rules.

Second, direct discrimination is measured by identifyingα-discriminatory rules among the PDrules using a direct discrimination measure (elift)and a discriminatory threshold (α). Third, indirectdiscrimination is measured by identifying redliningrules among the PND rules combined with backgroundknowledge, using an indirect discriminatorymeasure (elb), and a discriminatory threshold(α). Let MR be the database of direct α-discriminatoryrules obtained with the above process. Inaddition, let RR be the database of redlining rulesand their respective indirect α-discriminatory rulesobtained with the above process.

Data transformation

Transform the original dataDB in such a way to remove direct and/or indirectdiscriminatory partialities, with lowestcollision on thedata and on genuine decision policy, so that nounfair decision rule can be mined from the transformeddata. In the following sections, we presentthe data transformation methods that can be used forthis purpose.

Direct Discrimination of Data Transformation

The proposed solution to prevent direct discrimination isbased on the fact that the data set of decision rules would befree of direct discrimination if it only contained PD rulesthat are α-protective or are instances of at least onenon-redlining PND rule. Therefore, a suitable data transformationwith minimum information loss should be appliedin such a way that each α-discriminatory rule eitherbecomes α-protective or an instance of a non-redliningPND rule. We call the first procedure direct rule protection(DRP) and the second one rule generalization.

Direct Rule Protection

There are two methods that could be appliedfor direct rule protection. One method (Method 1) changesthe discriminatory item set in some records (e.g., genderchanged from male to female in the records with grantedcredits) and the other method (Method 2) changes the classitem in some records (e.g., from grant credit to deny creditin the records with male gender). Similar data transformationmethods could be applied to obtain direct ruleprotection with respect to other measures (i.e., slift andolift).

Rule Generalization

Rule generalization is an additional data revolution methodfor direct discrimination prevention.

Case 1: Inthis case, r0 is a p-instance of r for p ≥0:8 and no transformation is required. **Case 2:** In this case, the PND rule rb in Dpn should be selectedwhich requires the minimum data transformation. A smaller differencebetween the values of the two sides for each r in Dpn indicates a smaller required datatransformation. In this case, the α-discriminatoryrule is transformed by rule generalization.

Case 3: No rule in Dpn satisfies. In this case, rule generalizationis not possible and direct rule protectionshould be performed.

For the α-discriminatory rules to which rule generalization can be concerned, it is feasible that rule protection canbe achieved with a smaller data transformation. For theserules the algorithm should select the approach with minimumtransformation.

Indirect Discrimination of Data Transformation:

The proposed solution to prevent indirect discrimination isbased on the fact that the data set of decision rules would befree of indirect discrimination if it contained no redliningpolicy. To accomplish this, a appropriate data transformation withminimum information loss should be applied in such a waythat redlining rules are converted to non-redlining rules. Wecall this procedure indirect rule protection (IRP).

Indirect Rule Protection

There are two methods that could be appliedfor indirect rule protection.One method (Method 1) changes the discriminatory item setin some records (e.g., from non-foreign worker to foreignworker in the records of hired people in NYC city with Zip6¼ 10451) and the other method (Method 2) changes the classitem in some records (e.g., from "Hire yes" to "Hire no" inthe records of non-foreign worker of people in NYC citywith Zip 6¼ 10451).

Both Direct and Indirect Discrimination of Data Transformation

We deal here with the key problem of transforming datawith minimum information loss to prevent at the same timeboth direct and indirect discrimination. We will give apreprocessing solution to simultaneous direct and indirectdiscrimination prevention. First, we explain when direct andindirect discrimination could simultaneously occur. Thisdepends on whether the original data set (DB) containsdiscriminatory item sets or not.

To provide both direct rule protection (DRP) andindirect rule protection (IRP) at the same time, an importantpoint is the relation between the data transformationmethods. Any data transformation to eliminate direct α-discriminatory rules should not produce new redliningrules or prevent the existing ones from being removed. Alsoany data transformation to eliminate redlining rules shouldnot produce new direct α-discriminatory rules or preventthe existing ones from being removed. Indirect discrimination also assumes that the background knowledge takes the form of classification rules connecting the itemsets.

Prevention Algorithms for Direct Discrimination

We start with direct rule protection. Algorithm 1 detailsMethod 1 for DRP. For each direct α-discriminatory rule r´ in MR, after finding the subset DBc, records in DBc should be changed until the direct ruleprotection requirement is met for each respectiverule.

Algorithm1. Direct Rule Protection and RuleGeneralization

1: Inputs: $\mathcal{DB}$, $\mathcal{FR}$, $\mathcal{TR}$, $p \geq 0.8$, α, DI_s
2: Output: $\mathcal{DB}'$ (transformed data set)
3: **for** each $r' : A, B \rightarrow C \in \mathcal{TR}$ **do**
4: $\quad \mathcal{FR} \leftarrow \mathcal{FR} - \{r'\}$
5: $\quad$ **if** $TR_{r'} = RG$ **then**
6: $\quad\quad$ // Rule Generalization
7: $\quad\quad \mathcal{DB}_c \leftarrow$ All records completely supporting $A, B, \neg D \rightarrow C$
8: $\quad\quad$ Steps 6-9 Algorithm 1
9: $\quad\quad$ **while** $conf(r') > \frac{conf(r_b : D, B \rightarrow C)}{p}$ **do**
10: $\quad\quad\quad$ Select first record in $\mathcal{DB}_c$
11: $\quad\quad\quad$ Modify class item of db_c from C to $\neg C$ in $\mathcal{DB}$
12: $\quad\quad\quad$ Recompute $conf(r')$
13: $\quad\quad$ **end while**
14: $\quad$ **end if**
15: $\quad$ **if** $TR_{r'} = DRP$ **then**
16: $\quad\quad$ // Direct Rule Protection
17: $\quad\quad$ Steps 5-14 Algorithm 1 or Steps 4-9 Algorithm 2
18: $\quad$ **end if**
19: **end for**
20: Output: $\mathcal{DB}' = \mathcal{DB}$

Algorithm 1 takes as input TR, it containing all r´Є MR andtheir respective TRr´ and rb. For each α-discriminatory ruler´ in TR, if TRr´ shows that rule generalization should be performed, after determining the records thatshould be changed for impact minimization, theserecords should be changed until the rule generalizationrequirement is met.

Prevention Algorithms for Indirect Discrimination

A detailed algorithm implementing Method 2 for IRP is provided, from which an algorithm implementingMethod 1 for IRP can be easily derived. For the sake ofbrevity and due to similarity with the previous algorithms.

Prevention Algorithms for Direct and Indirect Discrimination

Algorithm 2 details our proposed data transformationmethod for simultaneous direct and indirect discriminationprevention. The algorithm starts with redlining rules. Fromeach redlining rule (r : X C), more than one indirect α-discriminatory rule (r´ : A, BC) might be generatedbecause of two reasons: 1) existence of different ways togroup the items in X into a context item set B and anondiscriminatory item set D correlated to some discriminatoryitem set A; and 2) existence of more than one item inDIs. Hence, as shown in Algorithm, given aredlining rule r, proper data transformation should beconducted for all indirect α-discriminatory rules.

Algorithm2. Direct and Indirect Discrimination Prevention:

If some rules from database can be extracted as direct and indirect α –discriminatory rules, then it means that it has an overlap between the MR and RR, where the transformation of data is to be performed until both the direct rule protection and indirect rule protection necessities are satisfied.For each indirect α –discriminatory rule resulting from each redlining rule in *RR* which can be complete without any impact on the direct discrimination prevention. Given a redlining rule (r), the correct transformation of data should be carried out for all α –discriminatory rules resulting from r. For each direct α –discriminatory rule r´ Є MR /RR where the data transformation is carry out to satisfy the necessities direct discrimination prevention, which does not have any impact on indirect discrimination prevention. The algorithm achieves the rule protection and rule generalization for each rule in MR has no unfavorableoutcome on protection for other rules for the following reasons: transformation of data for each rule is the same and no α –discriminatory rules in MR.

1: Inputs: $\mathcal{DB}$, $\mathcal{FR}$, $\mathcal{RR}$, $\mathcal{MR}$, α, DI_s
2: Output: $\mathcal{DB}'$ (transformed data set)
3: **for** each $r : X \rightarrow C \in \mathcal{RR}$, where $D, B \subseteq X$ **do**
4: $\gamma = conf(r)$
5: **for** each $r' : (A \subseteq DI_s), (B \subseteq X) \rightarrow C \in \mathcal{RR}$ **do**
6: $\beta_2 = conf(r_{b2} : X \rightarrow A)$
7: $\Delta_1 = supp(r_{b2} : X \rightarrow A)$
8: $\delta = conf(B \rightarrow C)$
9: $\Delta_2 = supp(B \rightarrow A)$
10: $\beta_1 = \frac{\Delta_1}{\Delta_2}$ $//conf(r_{b1} : A, B \rightarrow D)$
11: Find $\mathcal{DB}_c$: all records in $\mathcal{DB}$ that completely support $\neg A, B, \neg D \rightarrow \neg C$
12: Steps 6-9 Algorithm 1
13: **if** $r' \in \mathcal{MR}$ **then**
14: **while** $(\delta \leq \frac{\beta_1(\beta_2+\gamma-1)}{\beta_2 \cdot \alpha})$ and $(\delta \leq \frac{conf(r')}{\alpha})$ **do**
15: Select first record db_c in $\mathcal{DB}_c$
16: Modify the class item of db_c from $\neg C$ to C in $\mathcal{DB}$
17: Recompute $\delta = conf(B \rightarrow C)$
18: **end while**
19: **else**
20: **while** $\delta \leq \frac{\beta_1(\beta_2+\gamma-1)}{\beta_2 \cdot \alpha}$ **do**
21: Steps 15-17 Algorithm 4
22: **end while**
23: **end if**
24: **end for**
25: **end for**
26: **for** each $r' : (A, B \rightarrow C) \in \mathcal{MR} \setminus \mathcal{RR}$ **do**
27: $\delta = conf(B \rightarrow C)$
28: Find $\mathcal{DB}_c$: all records in $\mathcal{DB}$ that completely support $\neg A, B \rightarrow \neg C$
29: Step 12
30: **while** $(\delta \leq \frac{conf(r')}{\alpha})$ **do**
31: Steps 15-17 Algorithm 4
32: **end while**
33: **end for**
34: Output: $\mathcal{DB}' = \mathcal{DB}$

IV. Conclusions

Along with privacy, discrimination is a very importantissue when considering the legal and ethical aspects of datamining. It is other than observable that the majority people do notdesire to be discriminated because of their gender, religious conviction,ethnic group, age, and so on, particularly when those aspectsare used for making decisions about them like giving thema job, loan, insurance, etc.The purpose of this paper was to develop a new preprocessingdiscrimination prevention methodology includingdifferent data transformation methods that can preventdirect discrimination, indirect discrimination or both ofthem at the same time. To attain this objective, the first step isto measure discrimination and identify categories andgroups of individuals that have been directly and/orindirectly discriminated in the decision-making processes;the second step is to transform data in the proper way toremove all those discriminatory biases. Finally, discrimination-free data models can be produced from the transformeddata set without seriously damaging data quality. Theexperimental results reported demonstrate that the proposedtechniques are quite successful in both goals ofremoving discrimination and preserving data quality.

In future we extend the work by combining the other attributes. For example consider load granting application, the manager can reject the application based on their sensitive or non-sensitive attributes. So we introduce a new classification rule by means other than sensitive and non-sensitive attributes, we will add insurance policy details also. The proposed method mainly prevents the indirect discrimination process.

References

[1] R. Agrawal and R. Srikant, "Fast Algorithms for MiningAssociation Rules in Large Databases," Proc. 20th Int'l Conf. VeryLarge Data Bases, pp. 487-499, 1994.

[2] T. Calders and S. Verwer, "Three Naive Bayes Approaches forDiscrimination-Free Classification," Data Mining and KnowledgeDiscovery, vol. 21, no. 2, pp. 277-292, 2010.

[3] European Commission, "EU Directive 2004/113/EC on Anti-Discrimination," http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?ui=OJ:L:2004:373:0037:0043:EN:PDF, 2004.

[4] 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.

[5] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "DiscriminationPrevention in Data Mining for Intrusion and CrimeDetection," Proc. IEEE Symp. Computational Intelligence in CyberSecurity (CICS '11), pp. 47-54, 2011.

[6] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "RuleProtection for Indirect Discrimination Prevention in DataMining," Proc. Eighth Int'l Conf. Modeling Decisions for ArtificialIntelligence (MDAI '11), pp. 211-222, 2011.

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

[8] F. Kamiran and T. Calders, "Classification with no Discriminationby Preferential Sampling," Proc. 19th Machine Learning Conf.Belgium and The Netherlands, 2010.

[9] F. Kamiran, T. Calders, and M. Pechenizkiy, "DiscriminationAware Decision Tree Learning," Proc. IEEE Int'l Conf. Data Mining(ICDM '10), pp. 869-874, 2010.

[10] R. Kohavi and B. Becker, "UCI Repository of Machine LearningDatabases," http://archive.ics.uci.edu/ml/datasets/Adult, 1996.

[11] D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, "UCIRepository of Machine Learning Databases," http://archive.ics.uci.edu/ml, 1998.

[12] D. Pedreschi, S. Ruggieri, and F. Turini, "Discrimination-AwareData Mining," Proc. 14th ACM Int'l Conf. Knowledge Discovery andData Mining (KDD '08), pp. 560-568, 2008.

[13] D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring Discriminationin Socially-Sensitive Decision Records," Proc. Ninth SIAMData Mining Conf. (SDM '09), pp. 581-592, 2009.

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

[15] S. Ruggieri, D. Pedreschi, and F. Turini, "Data Mining forDiscrimination Discovery," ACM Trans. Knowledge Discovery fromData, vol. 4, no. 2, article 9, 2010.

[16] United States Congress, US Equal Pay Act, http://archive. eeoc.gov/epa/anniversary/epa-40.html, 1963.