Volume 3, Issue 12, December 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Privacy Preservation Using Discrimination Prevention Methods in Data Mining

Neenu Mary Kuruvila, V.Vennila
Ksr College Of Engineering
India

Abstract- In data mining, discrimination is a very important issue when considering the legal and ethical aspects of privacy preservation. It is more clear that most of the people do not have a wish to discriminated based on their race, nationality, religion, age and so on. This problem mainly arises when these kind of attributes are used for decision making purpose such as giving them a job, loan. Insurance etc.. For this reason discovering such attributes and eliminating them from the training data without affecting their decision-making utility is essential. So we introduce an antidiscrimination techniques which including discrimination discovery and prevention. Discrimination is two types. Direct and indirect. Direct discrimination is occurs when decision making is based upon some sensitive attributes. Indirect discrimination is occurs when decision making is based upon non sensitive attributes which are correlated with sensitive attributes. Discrimination prevention is mainly used for the purpose of inducing patterns that do not lead to discriminatory decisions even if the original training datasets contain any of the discriminatory attributes. In the discrimination prevention method, we introduce a group of pre-processing discrimination prevention methods and specify the different features of each approach and how these approaches deal with direct or indirect discrimination. We discuss how to clean training data sets and outsourced data sets in such a way that direct and/or indirect discriminatory decision rules are converted to nondiscriminatory classification rules. Some metrics are used to evaluate the performance of those approaches is also given.

IndexTerms- data mining, antidiscrimination, direct and indirect discrimination prevention, rule generalization, rule protection, privacy preservation.

I. INTRODUCTION

Discrimination is defined by the process of unfairly treating people on the basis of their belonging to a specific group, namely race, ideology etc. This involves denying opportunities to members of one group that are available to other group of people. Here some antidiscrimination acts are used , which are laws designed to prevent discrimination on the basis of a set of attributes (e.g., race, religion, gender, nationality, disability and marital status) in various settings (e.g., credit and insurance, employment and training, access to public services, etc.).Some examples are the US Employment Non-Discrimination Act (United States Congress 1994), the UK Sex Discrimination Act (Parliament of the United Kingdom 1975) and the UK Race Relations Act (Parliament of the United Kingdom 1976). Several decision-making tasks are there which lend themselves to discrimination, such as health insurances loan granting, education, and staff selection. In many applications, decision-making tasks are supported by information systems. Given a set of information items about a normal customer, an automated system decides whether the customer is to be recommended for a credit or a certain type of life insurance. This type of automated decisions reduces the workload of the staff of banks and insurance companies, among other organizations. The use of these information systems in data mining technology for decision making has attracted the attention of many persons in the field of computer applications. In consequence, automated data collection and data mining techniques such as association/classification rule mining have been designed and are currently widely used for making automated decisions. Automating decisions may give a sense of fairness: classification rules (decision rules) do not guide themselves by personal preferences. However in a closer look, one realizes that classification rules are actually learned by the system based on training data. If this training data are inherently biased for or against a particular community (for example, foreign workers),then ther is a chance of occurring discriminatory characteristics.

Discrimination is two types. Direct and indirect. Direct discrimination consists of procedures or rules that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. Indirect discrimination consists of procedures or rules that, not explicitly mentioning discriminatory attributes, directly or indirectly generate discriminatory decisions. An example of indirect discrimination is refusing to grant mortgages or insurances in urban areas they consider as deteriorating although certainly not the only one. In this paper indirect discrimination will also be referred to as redlining and rules causing indirect discrimination will be called redlining rules [1]. Indirect discrimination could need some background knowledge (rules), for example, that a certain zip code corresponds to a deteriorating area or an area with mostly female population. The background knowledge might be get

from publicly available data (e.g., census data) or might be obtained from the original data set itself because of the existence of nondiscriminatory attributes that are highly correlated with the sensitive ones in the original data set.

This paper is organized as follows. Section II introduces existing work. Section III discuss some basic definitions and concepts that are used throughout the paper. Section IV describes our proposal for direct and indirect discrimination prevention. Section V shows the tests we have performed to assess the validity and quality of our proposal and compare different methods. Finally, Section V1 summarizes conclusions of discrimination prevention.

II. EXISTING WORK

The wide deployment of information systems based on data mining technology in decision making, the important of antidiscrimination in data mining did not get much care until 2008 [1]. Some proposals are used to the discovery and measure of discrimination. But others deal with the prevention of discrimination.

The discrimination discovery decision was first proposed by Pedreschi et al. [1], [3]. This approach is based on mining classification rules (the inductive part) and reasoning on them (the deductive part) on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination. In the US Equal Pay Act [5] states that: "the selection rate for any ethnic group, race, or sex which is less than four-fifths of the rate for the group with the highest rate will generally be regarded as evidence of opposed effect." It has been implemented as an Oracle-based tool [4].In the current discrimination discovery methods consider each rule individually for measuring discrimination without considering other rules or the relation between them. In this paper we also take into account the relation between rules for discrimination discovery, based on the presence or absence of discriminatory attributes. In discrimination prevention, the other major antidiscrimination aim in data mining consists of introducing patterns that do not lead to discriminatory decisions even if the original training data sets are biased. Three approaches are used:

- Pre processing. The source data is transformed in such a way that the discriminatory biases contained in the original data are removed so that no unfair decision rule can be mined from the transformed data and apply any of the standard data mining algorithms. In this pre processing approaches of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature [6], [7].
- In processing. Change the data mining algorithms in such a way that the resulting models do not contain unfair decision rules. There is an alternative approach to cleaning the discrimination from the original data set is proposed in [2] whereby the nondiscriminatory constraint is embedded into a decision tree learner by changing its splitting criterion and pruning strategy through a novel leaf relabeling approach.
- Post processing. In this approach modify the resulting data mining models, instead of transform the original data set or changing the data mining algorithms. For instance, in [8], a confidence-altering approach is proposed for classification rules inferred by the CPAR algorithm.

In this paper, we focus on discrimination prevention based on preprocessing, because the preprocessing approach seems the most flexible one.

III. BACKGROUND

First, we recollect some basic definitions related to data mining [10]. After that, we concentrate on measuring and discovering discrimination.

A. Basic Definitions

- A data set is a group of data objects (records) and their attributes. Consider the original data set as DB.
- An item is an attribute along with its value, e.g: Race = black.
- An item set, i.e $.X$,is a collection of one or more items, e.g., {Foreign worker = Yes; City=NYC}.
- A classification rule is an expression, $X \rightarrow C$, where C is a class item (a yes/no decision), and $X$is an item set containing no class item, e.g., {Foreign worker =Yes; City = NYC} →Hire = no. X is called the premise of the rule.
- The support of an item set,$supp(X)$ is the fraction of records that contain the item set X. We say that a rule $X \rightarrow C$ is completely supported by a record if both X and C appear in the record.
- The confidence of a classification rule,$conf(X \rightarrow C)$, measures how often the class item Cappears in records that contain X. Hence, if supp(X) > 0 then

$$conf(X \rightarrow C) = \frac{supp\ (X,C)}{supp\ (X)} \qquad (1)$$

- A frequent classification rule is also a classification rule with support and confidence greater than respective specified lower bounds. Here support is a measure of statistical significance, and confidence is a measure of the strength of the rule. Let FR be the database of frequent classification rules extracted from DB.
- The negated item set, i.e., $\neg X$ is an item set with the same attributes in X, but the attributes in $\neg X$ take any value except those taken by attributes in X. In this paper, we use the ¬ notation for item sets with binary or non binary categorical attributes. For a binary attribute, e.g., {Foreign worker =Yes/No}, if X is {Foreign worker = Yes}, then $\neg X$ is {Foreign worker=No}. If X is binary, it can be converted to $\neg X$ and vice versa, that is, the negation works in both meaning. In the above example, we can select the records in DB such that the value of the Foreign worker attribute is "Yes" and change that attributes value to "No," and conversely.

B. Potentially Discriminatory And Nondiscriminatory Classification Rules

Let DIs be the set of predetermined discriminatory items in DB (e.g., DIs = {Foreign worker = Yes; Race = Black; Gender =Female}). Frequent classification rules in FR fall into one of the following two classes:

1. A classification rule $X \rightarrow C$ is potentially discriminatory (PD) when $X = A, B$ with A → DIs a nonempty discriminatory item set and B a nondiscriminatory item set. For example, {Foreign worker = Yes; City = NYC} → Hire = No.
2. A classification rule $X \rightarrow C$ is potentially nondiscriminatory (PND) when $X = D, B$ is a nondiscriminatory item set. For example, {Zip =10451; City = NYC} → Hire = No}, or {Experience = Low; City = NYC} → Hire =No.

C. Direct Discrimination Measure

Pedreschi et al. [8], [9] translated the qualitative statements in regulations, existing laws and legal cases into quantitative formal counterparts over classification rules and they introduced a family of measures of the degree of discrimination of a PD rule. From these measures one important one is the extended lift (elift).

Definition 1. Let $A, B \rightarrow C$ be a classification rule such that $conf(B \rightarrow C) > 0$. So the extended lift of the rule is

$$Elift(A, B \rightarrow C) = \frac{conf\ (A,B \rightarrow C)}{conf\ (B \rightarrow C)} \qquad (2)$$

The idea here is to measure the discrimination of a rule as the gain of confidence due to the presence of the discriminatory items (i.e., A) in the premise of the rule.If the rule is to be considered discriminatory can be assessed by thresholding elift as follows.

Definition 2. Let $\alpha \in R$ R be a fixed threshold and let A be a discriminatory item set. A PD classification rule $C = A, B \rightarrow C$ is α-protective w.r.t. elift if, elift(C) < α.Or C is α-discriminatory.

The purpose of direct discrimination discovery is to identify α-discriminatory rules. In fact, α-discriminatory rules indicate biased rules that are directly inferred from discriminatory items (e.g., Foreign worker = Yes). We call these rules direct α-discriminatory rules.

D. Indirect Discrimination Measure

The aim of indirect discrimination discovery is to recognize redlining rules. In fact, redlining rules point out biased rules that are indirectly inferred from nondiscriminatory items (e.g., Zip = 10451) because of their mutual relation with discriminatory ones. To find out the redlining rules, Pedreschi et al. in [9] stated the theorem below which gives a lower bound for α-discrimination of PD classification rules, based on information available in PND rules (γ,δ), and information available from background rules (β1, β2). They assume that background knowledge takes the within the context B.

Theorem1.Let $r: D, B \rightarrow C$ be a PND classification rule, and let γ = $conf(r: D, B \rightarrow C)$ δ =$conf(B \rightarrow C)$ >0

Let A be a discriminatory item set, and let β1, β2 such that

$conf(rb1: A, B \rightarrow D) \geq \beta 1$

$conf(rb2: D, B \rightarrow A) \geq \beta 2 > 0$

Call

$f(x) = \frac{\beta 1}{\beta 2}(\beta 2 + x - 1)$, then

$$elb(x, y) = f(x)\ if\ f(x) > 0 \qquad (3)$$

It holds that, for α ≥0, if elb(γ,β) ≥α the PD classification rule $r0: A, B \rightarrow C$ is α-discriminatory.

On the basis of the above theorem, the following subsequent interpretations of redlining and non redlining rules are presented:

Definition 3. A PND classification rule $r: D, B \rightarrow C$ is a redlining rule if it could yield an α-discriminatory rule $r0: A, B \rightarrow C$ in combination with currently available background knowledge rules of the form $rb1: A, B \rightarrow D$ and $rb2: D, B \rightarrow A$, where A represent a discriminatory item set. For example, {Zip = 10451; City = NYC} → Hire = No.

Definition 4. A PND classification rule $r: D, B \rightarrow C$ is a nonredlining or legitimate rule if it cannot yield any α-discriminatory rule $r0: A, B \rightarrow C$ in combination with currently available background knowledge rules of the form $rb1: A, B \rightarrow D$ and $rb2: D, B \rightarrow A$. Where A belongs to a discriminatory item set. For example, {Experience = Nil; City = NYC} → Hire = No. We call α-discriminatory rules that ensue from redlining rules indirect α-discriminatory rules.

IV. A PROPOSAL FOR DIRECT AND INDIRECT DISCRIMINATION PREVENTION

In this section, we introduce our approach, containing the data transformation methods that can be used for direct and/or indirect discrimination prevention. The algorithm and its computational cost for each method is indicated.

A. The Approach

The approach for direct and indirect discrimination prevention can be described in connection of two phases:

- Discrimination measurement: Direct and indirect discrimination discovery includes identifying α-discriminatory rules and redlining rules. For this, first, based on preordained discriminatory items in DB, frequent classification

rules in FR are divided in two groups: PD and PND rules. Second, direct discrimination is identified by identifying α-discriminatory rules among the PD rules using a direct discrimination measure (elift) and a discriminatory threshold (α). Third, indirect discrimination is measured by identifying redlining rules among the PND rules integrated with background knowledge, with the help of a indirect discriminatory measure (elb), and a discriminatory threshold (α). Let MR be the database of direct α-discriminatory rules obtained with the above process. In addition, let RR be the database of redlining rules and their respective indirect α-discriminatory rules obtained with the above process.

- Data transformation: Transform the original data DB by remove direct and/or indirect discriminatory biases, with minimum effect on the data and on legitimate decision rules, so that no uneven decision rule can be mined from the transformed data. In the coming sections, we introduce the data transformation methods that can be used for this purpose.

B.Data Transformation for Direct Discrimination

The recommended solution to avert direct discrimination is based on the fact that the data set of decision rules would be free of direct discrimination if it only contained PD rules that are α-protective or are instances of at least one nonredlining PND rule. For that, an appropriate data transformation with minimum information loss should be applied in such a way that each α-discriminatory rule either becomes α-protective or an instance of a nonredlining PND rule. The first procedure is direct rule protection (DRP) and the second one is rule generalization.

TABLE 1
Methods for Direct and Indirect Rule Protection

	Method1	Method2
Direct Rule Protection	$\neg A, B \rightarrow \neg C \equiv A, B \rightarrow \neg C$	$\neg A, B \rightarrow \neg C \equiv \neg A, B \rightarrow C$
Indirect Rule Protection	$\neg A, B, \neg D \rightarrow \neg C \equiv A, B, \neg D \rightarrow \neg C$	$\neg A, B, \neg D \rightarrow \neg C \equiv \neg A, B, \neg D \rightarrow C$

1) Direct Rule Protection

In order to convert each α-discriminatory rule into an α-protective rule, based on the direct discriminatory measure (i.e., Definition 2), we should enforce the following inequality for each -discriminatory rule $r0: A, B \rightarrow C$ in MR, where A is a discriminatory item set: there are two methods that could be applied for direct rule protection. One method (Method 1) changes the discriminatory item set in some records (e.g., gender changed from male to female in the records with granted credits) and the other method (Method 2) changes the class item in some records (e.g., from grant credit to deny credit in the records with male gender).

2) Rule Generalization

Rule generalization is secondary data transformation method for direct discrimination prevention. It is based on the fact that if each α-discriminatory rule $r0: A, B \rightarrow C$ in the database of decision rules was an instance of at least one nonredlining (legitimate) PND rule in the form of $r: D, B \rightarrow C$, it means that the data set would be free of direct discrimination. In rule generalization, we regard the relation between rules instead of discrimination measures.

Definition 5. Let $p \in [0,1]$. A classification rule $r': A, B \rightarrow C$ is a p-instance of $r: D, B \rightarrow C$ if both conditions below are true:

- Condition 1: $conf(r) \geq p.conf(r')$
- Condition 2: $conf(r'': A, B \rightarrow D) \geq p$

C. Data Transformation for Indirect Discrimination

An appropriate data transformation with minimum information loss should be applied in such a way that redlining rules are converted to nonredlining rules. We address this procedure indirect rule protection (IRP).

A) Indirect Rule Protection

In order to transform a redlining rule into an nonredlining rule, depends on the indirect discriminatory measure (i.e., elb in Theorem 1), we should compel the following inequality for each redlining rule in the form of $r: D, B \rightarrow C$ in RR:

$$elb(\gamma, \beta) < \alpha$$

Two methods are applied for indirect rule protection. One method (Method 1) will changes the discriminatory item set in some records and the other method (Method 2) changes the class item in some records.

D. The Algorithms

Algorithm 1. Direct Rule Protection (Method 1)
1: Inputs: DB, FR, MR, α, DIs
2: Output: DB' (transformed data set)
3: for each $': A, B \rightarrow C \epsilon MR$ do
4: $FR \leftarrow FR - \{r'\}$
5: $DBc \leftarrow$ All records completely supporting $\neg A, B \rightarrow C$
6: for each $dbc \in DBc$ do
7: Compute $impact(dbc) = |\{ra \in FR | dbc\ support\ the\ premise\ of\ ra\}|$
8: end for
9: Sort DBc by ascending impact

10: while $conf(r') \geq \alpha.conf(B \rightarrow C)$ do
11: Select first record in DBc
12: Modify discriminatory item set of dbc from $\neg A$ to $A\ in\ DB$
13: Recompute $conf(r')$
14: end while
15: end for
16: Output:$DB' = DB$

Algorithm 2. Direct Rule Protection (Method 2)
1: Inputs: DB, FR, MR, α, DIs
2: Output: DB' (transformed data set)
3: for each r' : $A, B \rightarrow C \epsilon MR$ do
4: Steps 4-9 Algorithm 1
5: while $conf(B \rightarrow C) \leq \frac{conf\ (r')}{\alpha}$do
6: Select first record in DBc
7: Modify the class item of dbc from $\neg C$ to C in DB
8: Recompute $conf(B \rightarrow C)$
9: end while
10: end for
11: Output: $DB' = DB$

V. DATA SETS

Adult data set: We used the Adult data set [11], also known as Census Income, in our experiments. This data set consists of 48,842 records, split into a "train" part consists of 32,561 records and a "test" part consists of 16,281 records. The data set has 14 attributes (without class attribute). We used the "train" part in our experiments. 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. Both categorical and numerical attributes are contained in the data set. For our experiments with the Adult data set, we set DIs = {Sex = Female, Age = Young}. Although the Age attribute in the Adult data set is numerical, we converted it to categorical by partitioning its domain into two fixed intervals: Age ≤30 was renamed as Young and Age > 30 was renamed as old.
German credit data set: we also used the German Credit data set [12]. This data set consists of 1,000 records and 20 attributes (without class attribute) of bank account holders. This is a well-known real-life data set, containing both numerical and categorical attributes. The class attribute in the German Credit data set takes values representing good or bad classification of the bank account holders. In our experiments with this data set, we set DIs = {Foreign worker = Yes, Personal Status = Female and not Single, Age = Old}; (cut-off for Age = Old: 50 years old).

A. Utility Measures

Discrimination removal is based on the following metrics:

- Direct discrimination prevention degree (DDPD).
  This measure quantifies the percentage of α-discriminatory rules that are no longer α-discriminatory in the transformed data set. DDPD can be defined as

$$DDPD = \frac{|MR| - |MR'|}{|MR|}$$

where MR is the database of α-discriminatory rules from DB and MR' is the database of α-discriminatory rules extracted from the transformed data set DB'.

- Direct discrimination protection preservation (DDPP).

This measure quantifies the percentage of the α-protective rules in the original data set that remain α-protective in the transformed data set. It is defined as

$$DDPP = \frac{|PR| \cap |PR'|}{|PR|}$$

where PR is the database of α-protective rules extracted from the original data set DB and PR' is the database of α-protective rules extracted from the transformed data set DB'

- Indirect discrimination prevention degree (IDPD).
  This measure quantifies the percentage of redlining rules that are no longer redlining in the transformed data set.
- Indirect discrimination protection preservation (IDPP).
  This measure quantifies the percentage of nonredlining rules in the original data set that remain nonredlining in the transformed data set.
- Misses cost (MC). This measure quantifies the percentage of rules among those extractable from the original data set that cannot be extracted from the transformed data set (side effect of the transformation process).
- Ghost cost (GC). This measure quantifies the percentage of the rules among those extractable from the transformed data set that were not extractable from the original data set (side effect of the transformation process).

B. Evaluation of the Methods

We implemented the algorithms for all proposed methods for direct and/or indirect discrimination prevention, and we evaluated them in terms of the proposed utility measures. We report the performance results in this section

TABLE 2
Adult Data Set: Utility Measures for Minimum Support 2 Percent and Confidence 10 Percent for All the Method

Methods	α	p	No. Red lining Rules	No. Indirect α-Disc. Rules	No. Direct α-Disc. Rules	Discrimination Removal				Data Quality	
						Direct		Indirect		MC	GC
						DDPD	DDPP	IDPD	IDPP		
Removing Disc. Attributes	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a	64.35	0
DRP(Method1)	1.2	n.a	n.a	n.a	991	100	100	n.a	n.a	15.44	13.52
DRP(Method 2)	1.2	n.a	n.a	n.a	991	100	100	n.a	n.a	0	4.06
DRP(Method1)+RG	1.2	0.9	n.a	n.a	991	100	100	n.a	n.a	13.34	12.01
DRP(Method2)+RG	1.2	0.9	n.a	n.a	991	100	100	n.a	n.a	0.01	4.06
IRP(Method1)	1.1	n.a	37	42	n.a	n.a	n.a	100	100	1.62	1.47
IRP(Method2)	1.1	n.a	37	42	n.a	n.a	n.a	100	100	0	0.96
DRP(Method2)+IRP(Method2)	1.1	n.a	37	42	499	99.97	100	100	100	0	2.07

Table 2 shows the results for minimum support 5 percent and minimum confidence 10 percent. The results of direct discrimination prevention methods are reported for discriminatory threshold p = 1.2 and, in the cases where direct rule protection is applied in combination with rule generalization, we used p = 0.9, and DIs ={Personal Status = Female and not Single, Age = Old} in the German Credit data set..

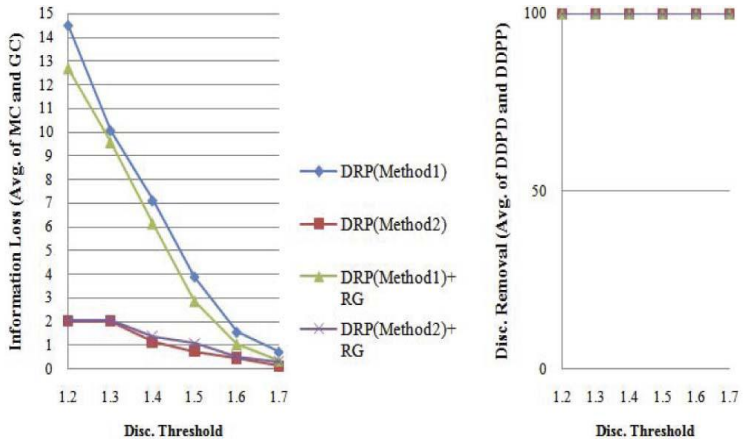

Fig. 1.
Information loss (left) and discrimination removal degree (right) for direct discrimination prevention methods for $\alpha \in [1.2,1.7]$.

After the above general results and comparison between methods, we now present more specific results on each method for different parameters α and p. Fig. 1 shows on the left the degree of information loss (as average of MC and GC) and on the right the degree of discrimination removal (as average of DDPD and DDPP) of direct discrimination prevention methods for the German Credit data set when the value of the discriminatory threshold α varies from 1.2 to 1.7, p is 0.9, the minimum support is 5 percent and the minimum confidence is 10 percent. The number of direct α-discriminatory rules extracted from the data set is 991 for α = 1.2, 415 for α=1.3, 207 for α = 1.4, 120 for α = 1.5, 63 for α = 1.6, and 30 for α = 1.7, respectively. As shown in Fig. 1, the degree of discrimination removal provided by all methods for different values of α is also 100 percent. However, the degree of information loss decreases substantially as

α increases; the reason is that, as α increases, the number of α -discriminatory rules to be dealt with decreases. In addition, as shown in Fig. 1, the lowest information loss for most values of α is obtained by Method 2 for DRP.

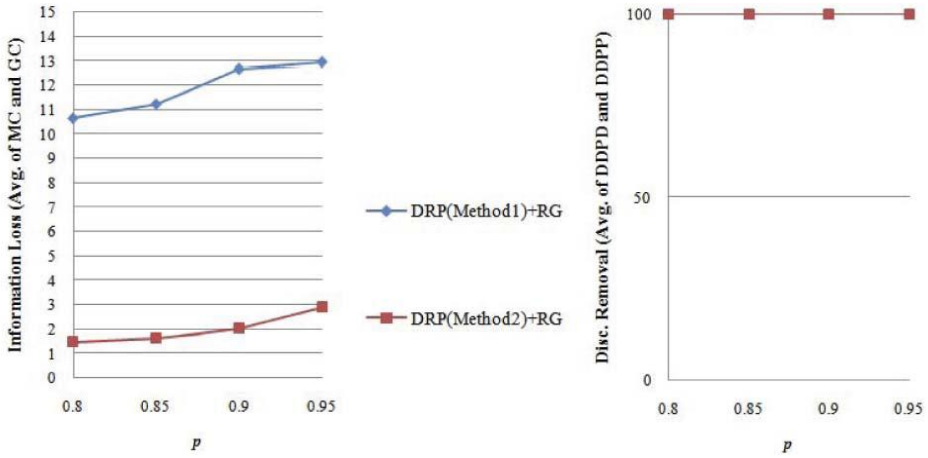

Fig. 2.
Information loss (left) and discrimination removal (right) degree for direct discrimination prevention methods for $p \in [0.8,0.95]$.

In addition, to demonstrate the impact of varying p on the utility measures in the methods using Rule Generalization, Fig. 2(left) shows the degree of information loss and Fig. 2(right) shows the degree of discrimination removal for different values of $p(0.8,0.85,0.9,0.95)$ and α = 1.2 for the German Credit data set. Although the values of DDPD and DDPP achieved for different values of p remain almost the same, increasing the value of p leads to an increase of MC and GC because, to cope with the rule generalization requirements, more data records must be changed.

VI. CONCLUSION

Along with privacy, discrimination is a very important issue when considering the legal and ethical aspects of data mining. The purpose of this paper was to develop a new preprocessing discrimination prevention methodology including different data transformation methods that can prevent direct discrimination, indirect discrimination or both of them at the same time. To attain this objective, the first step is to measure discrimination and identify categories and groups of individuals that have been directly and/or indirectly discriminated in the decision-making processes; the second step is to transform data in the proper way to remove all those discriminatory biases. Finally, discrimination- free data models can be produced from the transformed data set without seriously damaging data quality. The experimental results reported demonstrate that the proposed techniques are quite successful in both goals of removing discrimination and preserving data quality.

REFERENCE

[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.
[2] 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.
[3] 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.
[4] 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.
[5] United States Congress, US Equal Pay Act, eeoc.gov/epa/anniversary/epa-40.html, 1963.
[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] 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.
[9] 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.
[10] P.N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Addison-Wesley, 2006.
[11] R. Kohavi and B. Becker, "UCI Repository of Machine Learning Databases," http://archive.ics.uci.edu/ml/datasets/Adult, 1996.
[12] D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, "UCI Repository of Machine Learning Databases," ics.uci.edu/ ml, 1998.

