Discrimination Prevention Mechanism with Differential Privacy Scheme in Data Mining

S.Subbulakshmi
PG Scholar, Computer Science and Engineering
Vivekanandha College of Engineering for Women
Namakkal, India

B.Arulkumar
AP, Computer Science and Engineering
Vivekanandha College of Engineering for Women
Namakkal, India

Abstract--**Data mining is an important technology for extracting useful knowledge hidden in large collections of data. Privacy is a main issue in Data mining. The former is an unintentional or planned admission of a user profile or activity data as part of the output of a data mining algorithm or as a result of data sharing. For this reason, privacy preserving data mining has been introduced to trade-off the utility of the resulting data for protecting individual privacy. Along with privacy, discrimination is a very important problem when considering the legal and ethical aspects of data mining. Discriminations are divided into two types such as direct and indirect discriminations. In existing system discrimination discovery and prevention are used for anti-discrimination requirements. Direct and indirect discriminations prevention is applied on individually or both at the same time. The data values are cleaned to obtain direct and/or indirect discriminatory decision rules. But it contains the problem of low privacy assurance, because the data miner cannot worry about the privacy of data and also provide the measure of limited utility rate. The discrimination techniques cannot analyze the sensitive data from the transformed dataset. For getting a high privacy, a differential privacy scheme can be used in a discrimination prevention mechanism. The discrimination prevention model is integrated with the proposed method. Dynamic policy selection based discrimination prevention is adopted to generalize the systems for all regions. The data transformation technique is used to evaluate the utility rate, and also propose a rule hiding techniques for improving the removal process of discrimination. These proposed techniques are effective at shield sensitive data and removing direct and/or indirect discrimination biases in the original data set while preserving data quality.**

Index Terms-- **Data Mining, Discrimination, Privacy Preserving, Rules, Utility Rate, Dynamic Policy.**

I. INTRODUCTION

Civil right laws prohibit discrimination on the basis of race, color, religion, nationality, sex, marital status, age and pregnancy in a number of settings, including: credit and insurance; sale, rental, and financing of housing; personnel selection and wages; access to public accommodations, education, nursing homes, adoptions, and health care. For the key legal references, we refer the reader to the Australian Legislation; European Union Legislation; United Nations Legislation; U.K. Legislation; U.S. Federal Legislation. Several authorities monitor and report on discrimination compliances. For instance, the European Commission publishes an annual report on the progress in implementing the Equal Treatment Directives by the member states and in the U.S.A. the Attorney General reports to the Congress about the annual referrals to the Equal Credit Opportunity Act. Schiek et al. From the research side, the literature in economics and social sciences has given evidence of unfair treatment in racial profiling and redlining in Calem et al; Squires; mortgage lending in LaCour-Little; consumer market in Riach and Rich; Yinger; credit and housing in Dymski; personnel selection in Hunter and wages in Kuhn.

Given the current state of the art of decision support systems (DSS), socially sensitive decisions may be taken by automatic systems, e.g., for screening or ranking applicants to a job position, to a loan, to school admission and so on. For instance, data mining and machine learning classification models are constructed on the basis of historical data exactly with the purpose of learning the distinctive elements of different classes, such as good/bad debtor in credit/insurance scoring systems; Thomas or good/bad worker in personnel selection. When applied for automatic decision making, DSS can potentially guarantee more uniform decisions, but still they can be discriminating in the social, negative sense. Moreover, the decisions taken by those systems may be hard to be stated in intelligible terms, even if their internals are disclosed as in a case before a court.

A DSS is often the result of merging/weighting several hand-coded business rules and routinely built predictive models which are black-box software due to technical, legacy, or proprietary reasons. Currently, what the state of the art can offer is the verification of an hypothesis of possible discrimination by means of statistical analysis of past decision records. On the contrary, we aim at extracting contexts of possible discrimination supported by legally-grounded measures of the degree of discrimination suffered

ISSN: 2278 – 1323

by protected-by-law groups in such contexts. Reasoning on the extracted contexts can support all the actors in an argument about possible discriminatory behaviors. The DSS owner can use them both to prevent incurring in discriminatory decisions, and as a means to argument against allegations of discriminatory behavior. A complainant in a case can use them to find specific situations in which there is a prima facie evidence of discrimination against groups she belongs to. Finally, control authorities can base the fight against discrimination on a formalized process of intelligent data analysis.

The actual discovery of discriminatory situations and practices, hidden in the decision records under analysis, may reveal an extremely difficult task. The reason for this difficulty is twofold. On the one side, a huge number of possible contexts may, or may not, be the theatre for discrimination. To see this point, consider the case of gender discrimination in credit approval: although an analyst may observe that no discrimination occurs in general, i.e., when considering the whole available decision records, it may turn out that it is extremely difficult for aged women to obtain car loans. Many small or large niches may exist that conceal discrimination, and therefore all possible specific situations should be considered as candidates, consisting of all possible combinations of variables and variable values: personal data, demographics, social, economic and cultural indicators, etc. Clearly, the anti-discrimination analyst is faced with a huge range of possibilities, which make her work hard: even though the task of checking some known suspicious situations can be conducted using available statistical methods, the task of discovering niches of discrimination in the data is unsupported. We call this issue the inductive problem in discrimination discovery.

On the other side, discrimination is rarely defined in rigorous and universal terms. First, protected-by-law groups, such as minorities and poor people, are sometimes not fully known, leaving space for uncertain issues such as in the debate about multiple, intersectional and compound discrimination discussed in ENAR. Second, the interpretation of existing legislations leads to different quantitative measures of discrimination and, a fortiori, to dissimilar thresholds between what is legal and illegal. Third, discrimination can be hidden behind apparently neutral practices, known as indirect discrimination that must be unveiled by some deductive reasoning exploiting additional information, which we call background knowledge. Fourth, a few policies, known as positive actions, that favor minorities are allowed, encouraged or even imposed by laws. Finally, in case a prima-facie evidence of discrimination is found in the data, the anti-discrimination analyst has still to consider possible argumentations of the respondent, e.g., in rival a genuine occupational requirement justification. We call these issues the deductive problem in discrimination discovery.

II. RELATED WORK

Discrimination analysis from data should build over the large body of existing legal and economic studies [7], [8]. In this section, we review the under-representation principle that has inspired previous data mining proposals, and the situation testing methodology, which provides the legal grounds for the approach proposed in this paper.

Accordingly to laws, discrimination occurs when a group is treated \less favorably" [9], than others, or such that “a higher proportion of people not in the group is able to comply" [2] to a qualifying criterium. A general principle is then to consider group under-representation in obtaining a benefit [1] as a quantitative measure of discrimination against a protected-by-law group. Consider a dataset of historical decisions about granting or not a benefit. Let p_1 (resp., p_2) be the proportion of people in the protected group that were not granted the benefit, and let p be the proportion of all people that were not granted the benefit. Group under-representation can be measured as the difference $p_1 - p_2$, adopted in the U.K. legislation; or as the ratio p_1/p_2, called the selection lift and adopted in the U.S. legislation; or as one of the measures over a four-fold contingency table. Higher values of those measures denote higher under-representation of the protected group.

The under-representation principle has inspired the existing approaches for discrimination discovery and prevention. [10] Proposes to extract classification rules of the form $A,B \rightarrow C$ to unveil subsets B of the dataset where the protected group A suffered from under-representation with respect to the decision C. The approach is parametric to one of the measures and it is implemented on top of an Oracle database [3]. [5] Investigates three approaches for preventing discriminatory predictions in a Naive Bayes classifier. Discrimination is measured as the difference $p_1 - p_2$ calculated on the whole set of predictions over a test set.

On the technical side, the discrimination discovery approach [10] has two limitations. First, since it relies on frequent itemset mining, it deals with nominal attributes and nominal decisions only. Interval-scaled attributes and decisions must be discretized as a pre-processing step. Second, the result of the knowledge discovery process is a set of classification rules, which provide local niches of possible discrimination: a global description of who is discriminated and who is not is lacking. The discrimination prevention approach [5] shares the same limitation on nominal attributes. In addition,

ISSN: 2278 – 1323

it considers discrimination at top level, i.e., $p_1 - p_2$ is controlled only for the whole set of decisions. However, discrimination may still occur in some subset, e.g., as when discrimination of a bank branch manager against a minority group remains hidden in the large set of decisions of the whole bank.

In the legal field, situation testing is a systematic research procedure for creating controlled experiments analyzing decision maker's candid responses to applicant's personal characteristics. In situation testing, pairs of research assistants undergo the same kind of selection, for example they apply for the same job, and they present themselves at the same night club, and so on. Within each pair, applicant characteristics likely to be related to the situation are made equal by selecting, training, and credentialing testers to appear equally qualified for the activity. Simultaneously, membership to a protected group is experimentally manipulated by pairing testers who differ in membership - for example, a black and a white, a male and a female, and so on. Situation testing is being experimented worldwide as one of the tools that can assist victims to establish that discrimination may have occurred [4], [6].

III. EXISTING SYSTEM

A. Protecting Discrimination in Mining Process

In the literature study we have seen many methods for Discrimination prevention. Discrimination can be either direct or indirect. Direct discrimination consists of rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. Indirect discrimination consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, intentionally or unintentionally could generate discriminatory decisions. Redlining by financial institutions is an archetypal example of indirect discrimination, although certainly not the only one. With a slight abuse of language for the sake of compactness, in this paper indirect discrimination will also be referred to as redlining and rules causing indirect discrimination will be called redlining rules. Indirect discrimination could happen because of the availability of some background knowledge, for example, that a certain zip code corresponds to a deteriorating area or an area with mostly black population. The background knowledge might be accessible from publicly available data or might be obtained from the original data set itself because of the existence of non-discriminatory attributes that are highly correlated with the sensitive ones in the original data set. Discrimination prevention methods based on preprocessing published so far some limitations, which we next highlight:

They attempt to detect discrimination in the original data only for one discriminatory item and based on a single measure. This approach cannot guarantee that the transformed data set is really discrimination free, because it is known that discriminatory behaviors can often be hidden behind several discriminatory items, and even behind combinations of them. They do not include any measure to evaluate how much discrimination has been removed and how much information loss has been incurred.

In this system propose preprocessing methods which overcome the above limitations. Our new data transformation methods are based on measures for both direct and indirect discrimination and can deal with several discriminatory items. Also, we provide utility measures. Hence, our approach to discrimination prevention is broader than in previous work. As part of this effort, we have developed metrics that specify which records should be changed, how many records should be changed, and how those records should be changed during data transformation. In addition, we propose new utility measures to evaluate the different proposed discrimination prevention methods in terms of data quality and discrimination removal for both direct and indirect discrimination. Based on the proposed measures, we present extensive experimental results for two well known data sets and compare the different possible methods for direct or indirect discrimination prevention to find out which methods could be more successful in terms of low information loss and high discrimination removal.

Automated data collection and data mining techniques such as classification rule mining are used to making automated decisions. Discriminations are divided into two types such as direct and indirect discriminations. Direct discrimination occurs when decisions are made based on sensitive attributes. Indirect discrimination occurs when decisions are made based on non sensitive attributes which are strongly correlated with biased sensitive ones. Direct and indirect discriminations prevention is applied on individually or both at the same time. Data transformation techniques are applied to prepare the data values for the discrimination prevention. Rule protection and rule generalization algorithm and direct and indirect discrimination prevention algorithm are used to protect discriminations. The following drawbacks are identified in the existing system.

- Static discrimination policy based scheme
- Limited utility ratio
- Low privacy assurance
- Privacy association is not analyzed

ISSN: 2278 – 1323

IV. PROPOSED SYSTEM

A. Discrimination Prevention Mechanism with Differential Privacy

The discrimination prevention model is integrated with the differential privacy scheme to high privacy. Dynamic policy selection based discrimination prevention is adopted to generalize the systems for all regions. Data transformation technique is improved to increase the utility rate. Discrimination removal process is improved with rule hiding techniques. The discrimination prevention system is designed to protect the decisions that are derived from the rule mining process. The system is enhanced to improve the data utility rate and privacy preservation rate. Policy selection model is used to perform dynamic policy based discrimination prevention tasks. The system is divided into five major modules. They are data cleaning process, privacy preservation, rule mining, rule hiding and discrimination prevention. The data cleaning module is designed to prepare the data for mining process. Privacy preservation module is designed to protect sensitive attribute. Frequent pattern mining operations are performed under the rule mining module. Sensitive rules are protected under the rule hiding process. Discrimination prevention module is used to perform direct and indirect discrimination prevention process.

1) Data Cleaning Process

Data populate and missing value assignment operations are carried out in the data cleaning process. Textual data values are transferred into the Oracle database. Incomplete transactions are updated with alternate values. Aggregation based data substitution method is used for data assignment process. The proposed solution to prevent 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. We call the first procedure direct rule protection (DRP) and the second one rule generalization. The proposed solution to prevent indirect discrimination is based on the fact that the data set of decision rules would be free of indirect discrimination if it contained no redlining rules. To achieve this, a suitable data transformation with minimum information loss should be applied in such a way that redlining rules are converted to nonredlining rules. We call this procedure indirect rule protection (IRP).

2) Privacy Preservation

Privacy preservation is applied to protect sensitive attributes. Differential privacy technique is applied on sensitive attributes. Noise is added with the sensitive attributes. Data transformation process is applied to prepare the data for rule mining process. Differential privacy is a recent privacy definition that guarantees the outcome of a calculation to be insensitive to any particular record in the data set. These costs add up as more queries are executed, until they reach an allotted bound set by the data provider, at which point further access to the database will be blocked. The composition property also provides some protection from collusion: collusion between adversaries will not lead to a direct breach in privacy, but rather cause it to degrade gracefully as more adversaries collude, and the data provider can also bound the overall privacy budget. Typically, differential privacy is achieved by adding noise to the outcome of a query. One way to do so is by calibrating the magnitude of noise required to obtain ϵ-differential privacy according to the sensitivity of a function. The sensitivity of a real-valued function expresses the maximal possible change in its value due to the addition or removal of a single record.

3) Rule Mining

The rule mining process is performed to filter the frequent patterns. Candidate sets are prepared using attribute name and values. Support and confidence values are estimated using item sets. Frequent patterns are identified with minimum support and confidence values. A number of data mining algorithms have been recently developed that greatly facilitate the processing and interpreting of large stores of data. One example is the association rule-mining algorithm, which discovers correlations between items in transactional databases. Priori algorithm is an example of association rule mining algorithm. Using this algorithm, candidate patterns that receive sufficient support from the database are considered for transformation into a rule. This type of algorithm works well for complete data with discrete values. One limitation of many association rule-mining algorithms such as the Apriori algorithm is that only database entries, which exactly match the candidate patterns, may contribute to the support of the candidate pattern. This creates a problem for databases containing many small variations between similar patterns and for databases containing missing values. Given a set of transactions, each described by an unordered set of items, an association rule $X - Y$ may be discovered in the data, where X and Y are conjunctions of items. The intuitive meaning of such a rule is that transactions in the database,

ISSN: 2278 – 1323

which contain the items in X, tend to also contain the items in Y.

4) Rule Hiding

Rule hiding method is applied to protect the sensitive rules. Rules derived from sensitive attributes are not released directly. Rules are embedded with nearest rule intervals. Attribute ranges are adjusted with sensitive rules. The hiding strategies, that we propose, heavily depend on finding transactions that fully or partially support the generating itemset of a rule. The reason for this is that if we want to hide a rule, we need to change the support of some part of the rule. Another issue is that the changes in the database introduced by the hiding process should be limited, in such a way that the information loss incurred by the process is minimal. According to this, we try to apply minimal changes in the database at every step of the hiding algorithms that we propose. The decrease in the support of an itemset *S* can be done by selecting a transaction *t*, that supports *S* and by setting to 0 at least one of the non-zero values of *t.values of items* that represent items in *S*. The increase in the support of an itemset *S* can be accomplished by selecting a transaction *t* that partially supports it and setting to 1 the values of all the items of *S* in *t. values of items*. In order to be able to identify some viable ways for reducing either the support or the confidence of a rule, we need to analyze the formulas that we have already presented for the confidence and the support. Both the confidence and the support are expressed as ratios of supports of itemsets that support the two parts of a rule or its generating itemset.

5) Discrimination Prevention

Discrimination prevention process is designed to protect decisions. Rule generalization and rule prevention algorithms are enhanced for dynamic policy model. Direct and indirect discrimination prevention algorithm is also tuned for dynamic policy scheme. Discriminations are protected with reference to sensitive and non-sensitive attributes. One of these measures is the extended lift (elift). 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. We call these rules direct α-discriminatory rules. In addition to elift, two other measures slift and olift were proposed by Pedreschi et al. The reason is that different measures of discriminating power of the mined decision rules can be defined, according to the various antidiscrimination regulations in different countries. Yet the protection methods are similar no matter the measure adopted. To determine the redlining rules stated the theorem below which gives a lower bound for α-discrimination of PD classification rules, given information available in PND rules (γ,δ), and information available from background rules (β_1, β_2). They assume that background knowledge takes the form of classification rules relating a non discriminatory item set D to a discriminatory item set A within the context B.

V. CONCLUSION

Data mining techniques are applied to hidden knowledge from data bases. Discriminatory decisions are obtained and prevented with reference to the attributes. Direct and indirect discrimination prevention scheme is used to protect the decision rules and the discrimination prevention scheme is enhanced with dynamic policy selection model and differential privacy mechanisms. It mainly increases the data utility rate and preserving the privacy rate. Policy selection based discrimination prevention model can be applied for all regions. The further enhancement of the system is to protect discrimination on textual data values perform under the distributed database environment.

REFERENCES

[1]. Alberto Trombetta, Wei Jiang, Bossi, "Privacy-Preserving Updates to Anonymous and Confidential Databases", IEEE Transactions On Dependable And Secure Computing, Vol. 8, No. 4, July/August 2011.

[2]. Binh Thanh Luong, Salvatore Ruggieri and Franco Turini, "k-NN as an Implementation of Situation Testing for Discrimination Discovery and Prevention", February 20, 2011.

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

[4]. 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.

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

[6]. 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.

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

[8]. 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.

[9]. 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.

[10]. Sara Hajian and Josep Domingo-Ferrer, "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining", IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 7, July 2013.

ISSN: 2278 – 1323