An Assembly of Discrimination Prevention Techniques in Data Mining

Anuja C. Tikole, Vikash V. Kamle, Shekhar J. Jadhav, Aditya S. More, Asmita Mali

Abstract— Data mining is the extraction of implicit, previously unknown, and potentially useful information from available data. The idea is to make computer programs that come through databases automatically, seeking regularities or patterns. In data mining, the data is stored electronically and search is automated by computer. Data mining is about solving problems by analyzing data already present in databases. There are, however, negative social perceptions about data mining, among which unjustifiable access and potential discrimination. Discrimination consists of unfairly treating people on the basis of their belonging to a particular group. Automated data collection and data mining techniques such as classification rule mining gives the way to making automated decisions, for e.g., loan granting/denial, insurance premium computation, etc. If the training data sets are biased in what regards discriminatory (sensitive) attributes as gender, race, religion, etc., discriminatory decisions may happen. Due to this, antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining .Discrimination is a presuppose privileges where provide to the each separate group for the safety of the data which is stored . Discrimination can be either direct or indirect. Direct discrimination finds 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. In this paper, proposed system covers discrimination prevention in data mining and propose new techniques applicable for direct and indirect discrimination prevention both at the same time.

Keywords—Data mining, Direct and Indirect Discrimination prevention, Antidiscrimination.

I. INTRODUCTION

Data mining is the process of discovering patterns in data. The process must be automatic or more easily semiautomatic. The available patterns must be meaningful in that they lead to some advantage, as many times an economic advantage. At most of the times data is present in of considerable importance quantities. Data mining refers to using a variety of techniques to identify information or decision-making knowledge in data, taken these in such a way that they can be put to use in the areas such as decision support, prediction, forecasting and estimation. The data is often detailed, but as it stands of low value as no direct use can be made of it; it is the hidden information in the data that is useful.

Manuscript Received on March 2015.

Anuja C. Tikole, Department of Information Technology, Pd. Dr. D. Y. Patil Institute of Engineering and Technology, Pimpri, Pune, India.

Vikash V. Kamle, Department of Information Technology, Pd. Dr. D. Y. Patil Institute of Engineering and Technology, Pimpri, Pune, India.

Shekhar J. Jadhav, Department of Information Technology, Pd. Dr. D. Y. Patil Institute of Engineering and Technology, Pimpri, Pune, India.

Aditya S. More, Department of Information Technology, Pd. Dr. D. Y. Patil Institute of Engineering and Technology, Pimpri, Pune, India.

Prof. Asmita Mali, Department of Information Technology, Pd. Dr. D. Y. Patil Institute of Engineering and Technology, Pimpri, Pune, India.

Data mining is not specific to one type of media or data. Data mining should be valid to any kind of information repository. However, algorithms and approaches may changes when applied to different types of data. But the challenges presented by different types of data changes significantly. Data mining is keep into use and studied for databases, having relational databases, data warehouses, object-relational databases and object oriented databases, unstructured and semi structured repositories such as the World Wide Web, transactional databases, advanced databases such as spatial databases, multimedia databases, time-series databases and textual databases, and even plane documents files.

The word *discrimination* originates from the Latin *discriminare*, which means to 'distinguish between'[4]. In generally, discrimination refers specially to an action based on partiality resulting in unfair treatment of people on the basis of their membership to a category, without considering individual merit.

Discrimination can be the act of unfairly treating people on the basis of their belonging to a specific group. For example, individuals may be discriminated because of their race, ideology, gender, etc. In economics and social sciences, discrimination studied for over half a century. There are several decision-making tasks which offer themselves to discrimination, *e.g.* loan granting and staff selection. In the last decades, anti-discrimination laws have been adopted by many democratic governments. In[2], Some examples are the US Equal Pay Act [1], the UK Sex Discrimination Act [2], the UK Race Relations Act [3] and the EU Directive 2000/43/EC on Anti discrimination [4].

Regarding the research side, the issue of discrimination in credit, mortgage, insurance, labor market, education and other human activities has attracted much interest of researchers in economics and human sciences since late '50s, when a theory on the economics of discrimination was proposed [3]. The literature has given evidence of unfair treatment in racial profiling and redlining , mortgage discrimination, personnel selection discrimination , and wages discrimination [3].

Discrimination can be either direct or indirect. Direct discrimination finds when decisions are taken based on sensitive attributes. Indirect discrimination finds when decisions are taken based on non-sensitive attributes which are strongly correlated with sensitive ones. So it is needed to give an antidiscrimination techniques including discrimination discovery and prevention. In the discrimination prevention method, there is 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. So there is requirement to checks how to clean training data sets and outsourced data sets in such a way that direct and/or

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

indirect discriminatory decision rules are converted to nondiscriminatory classification rules. Some metrics are used to evaluate the performance of those approaches is also given. In this paper with respect to these concepts, implementing system for banking application i.e. loan granting application. According to concept of discrimination firstly system will check for discriminatory attributes and then hide that attributes by encryption so that discrimination cannot occurs. For this purpose, set rule for applications which match with given set of input attributes. The set of rule which is discrimination free previously. Likewise the set of input is put forward the discrimination causing attributes will get encrypted. So that discrimination can be avoided. For indirect discrimination, the attributes which are directly dependent on sensitive attributes which may cause the discrimination are totally hidden and new set of rules is generalized so that necessary information is also taken and discrimination can be prevented. The aims of this paper are as follows: (1) Introducing anti-discrimination technique regarding banking database and its security; (2) proposing a new discrimination prevention method based on data transformation that can consider several discriminatory attributes and their combinations; (3) proposing some measures for evaluating the proposed method in terms of its success in discrimination prevention and its impact on original database. Here, Section II discusses related work; Section III introduces basic definitions required for applications based on data mining; Section IV reviews discrimination discovery; Section V presents the method for discrimination prevention and its evaluation; a discussion is given in Section VI; conclusions are drawn in Section VII.

II. RELATED WORK

The previous literature on discrimination prevention techniques in computer science mainly elaborates on data mining models and related techniques. Some are oriented to the discovery and measure of discrimination. Others deal with the prevention of discrimination.

• *Discrimination discovery* is based on formalizing legal definitions of discrimination1 and proposing quantitative measures for it. Data mining is a powerful aid for discrimination analysis, capable of discovering the patterns of discrimination that taken from the data.

• *Discrimination prevention* consists of making patterns that do not lead to discriminatory decisions even if trained from a dataset containing them. Three approaches are used: (i) adapting the preprocessing approaches of data transformation and hierarchy-based generalization from the privacy preservation literature (ii) changing the data mining algorithms (in-processing) by integrating discrimination measure evaluations within them and (iii) post-processing the data mining model to reduce the possibility of discriminatory decisions.

i. Pre-processing :-
In preprocessing data is transformed from original database in such a way that discriminatory attributes are removed and no unfair decision rule can be mined. It requires to apply standard data mining algorithms. This approach is useful when dataset are predefined or when data mining requires to be performed by third party also that not only by data holder.

ii. In-processing :-
In this approach algorithms are changed in such a way that new models do not contain discriminatory rules. Obviously in in-processing discrimination methods must depend on new special algorithms whereas standard algorithms are not used.

iii. Post-processing :-
In this approach instead of changing data mining algorithms or changing the original database we have to modify resulting data mining models. This approach does not allowed to dataset to be predefined.

Hence, one challenge regarding discrimination prevention is considering indirect discrimination other than direct discrimination and another challenge is to find an optimal trade-off between anti-discrimination and usefulness of the training data.

III. BACKGROUND

In this section we gather the basic knowledge required for this paper starting with basic concepts of data mining and then elaborate measures and discover the discrimination attributes.

- A dataset is a collection of records and their attributes. Let DB be the original dataset. An item is an attribute along with its value, e.g. Race=black. An itemset is a collection of one or more items. A classification rule is an expression X -> C, where X is an itemset, containing no class items, and C is a class item, e.g. Class=bad.
- The support of an itemset, supp(X), is the no. of records X contains in total record. So a rule X -> C completely supported by a record if both X and C appear in the record.
- The confidence of a classification rule, conf(X -> C), measures Show how many time C appear with X.
- A frequent classification rule is a classification rule with a support or confidence greater than a specified lower bound.

Let FR be the database of frequent classification rules extracted from DB. With the assumption that discriminatory items in DB are predetermined (e.g.Race=black), rules fall into one of the following two classes with respect to discriminatory and non-discriminatory items in DB:

(i) A classification rule is potentially discriminatory (PD) when X = A, B with A a non-empty discriminatory itemset and B a non-discriminatory itemset.

(ii) A classification rule is potentially non-discriminatory (PND) when X = D, B is a non-discriminatory itemset Let assume that the notation X(D,B) means X = D,B.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

(iii) Let PR a database of frequent classification rules with PD and PND classification rules.

The word "Potentially Discriminated" means that a PD rule could probably lead to discriminatory decisions, so some measures are needed to quantify the discrimination potential (direct discrimination). Also, a "Potentially Non Discriminated" that is PND rule could lead to discriminatory decisions if combined with some background knowledge (indirect discrimination);

e.g., if the premise of the PND rule contains the Zip=10451 itemset, rely on additional background knowledge one knows that zip 10451 is mostly inhabited by black people. System have already known a family of measures of the degree of discrimination of a PD rule.

One of these measures is extended lift measure (elift):

elift (A, B-> C) = conf (A, B-> C) /Conf (B->C)

Whether the rule is to be considered discriminatory can be assessed by using a threshold:

A. 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 → 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 → 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

B. Direct Discrimination Measure

Pedreschi et al. [3], [8] translated the qualitative statements in existing laws, regulations, 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. One of these measures is the extended lift (elift)[1].

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

Elift (A, B→C) = $co(A;B \rightarrow C)$con f$(B \rightarrow C)$ (2)

The idea here is to evaluate 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. Whether 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). It is called a direct α-discriminatory rules.

C. Indirect Discrimination Measure:

The purpose of indirect discrimination discovery is to identify redlining rules. In fact, redlining rules indicate biased rules that are indirectly inferred from nondiscriminatory items (e.g., Zip = 10451) because of their correlation with discriminatory ones. The proposed solution to prevent indirect discrimination is dependent on data set of decision rules would be free of indirect discrimination if it contained no redlining rule. To obtain this a suitable data transformation with minimum information loss should be applied as that redlining rules are converted into non redlining rules. This is called as indirect rule protection(IRP).

D. Data Transformation for Direct and Indirect Discrimination

It is important to deal with the problem of transforming data with minimum information loss to prevent at the same time both direct and indirect discrimination. With this regard it is needed to find out when direct and indirect discrimination occurs simultaneously. That depends on if original dataset contains discriminatory item sets or not.

1. If discriminatory data sets do not exist in the original database or that are already removed then only Indirect discrimination could occur.
2. At least one discriminatory item set is not removed from original database that is PD rules are extracted from database then direct discrimination could occur. And also indirect discrimination may occur due to existence of nondiscriminatory items that are closely related with sensitive attributes. Hence both direct and indirect discrimination can occur here in this case.

To get both direct rule protection (DRP) and indirect rule protection (IRP), simultaneously it is required the relation between data transformation methods. Any data transformation to eliminate the direct α-discriminatory rules should not give new redlining rules. Also any data transformation to eliminate redlining rules should not give new α-discriminatory rules.

IV. System Architecutre

By Using concepts of this paper, it is easy to make an system for most important banking transaction that is loan granting or denying. For this system there is login one for manager and one for end user. End user will apply for loan by filling the registration form and manager having original database as well as new database of recently added user/loan applicants and proposes 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, system 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. The

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

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. Following figure indicates exactly how proposed system is going to worked.

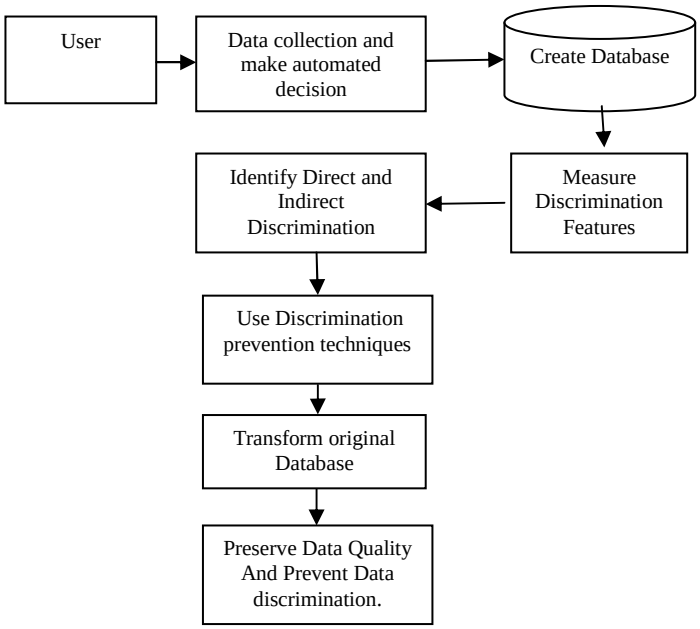

Fig1: Architecture of the Proposed System

V. DATA SETS

As previously work is done on this concept, regarding this two types of data sets are used.

1. Adult Data Set: It also called as census income, consisting of 48,842 records divided into 'train' part with 32,561 records and 'test' part with 16,281.the data set has 14 attributes. The main task is to determine income of particular person based on census and demographic information about people.
2. German Credit Data Set: It consists of 1000 records and 20 attributes of bank account holders. It is real-life data set , containing both numerical and categorical attributes.

VI. RESULTS AND UTILITY MEASURES

The solution should be evaluated based on two aspects:

1. Achieving the proposed solution in removing all attributes of discrimination from the original dataset (degree of discrimination prevention).
2. The impact of the achieving proposed solution on data quality (degree of information loss). A discrimination prevention method should provide a good trade-off between both aspects above. The following measures are proposed for evaluating our solution:
3. Discrimination Prevention Degree (DPD). This measure quantifies the percentage of α-discriminatory rules that are no longer α-discriminatory in the transformed dataset.
4. Discrimination Protection Preservation (DPP). This measure quantifies the percentage of the α-protective rules in the original dataset that remain α-protective rules in the transformed dataset (DPP may not be 100% as a side-effect of the transformation process).
5. Misses Cost (MC). This measure quantifies the percentage of rules among those extractable from the original dataset that cannot be extracted from the transformed dataset (side-effect of the transformation process).
6. Ghost Cost (GC). This measure quantifies the percentage of the rules among those extractable from the transformed dataset that could not be extracted from the original dataset (side-effect of the transformation process).

The DPD and DPP measures are used to evaluate the success of proposed method in discrimination prevention; ideally they should be 100%. The MC and GC measures are used for evaluating the degree of information loss (impact on data quality); ideally they should be 0%. MC and GC were previously proposed as information loss measures for knowledge hiding.

VII. CONCLUSIONS AND FUTURE WORK

As discrimination is a very important issue of data mining. The purpose of this paper was to develop a new preprocessing discrimination prevention including different data transformation methods that can prevent direct discrimination, indirect discrimination along with both at the same time. Along with privacy, discrimination is a very important issue when considering the legal and ethical aspects of data mining. 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 damaging data quality. The basic purpose of this paper was to develop a new banking application for loan granting purpose on the basis of preprocessing discrimination prevention methodology including different data transformation methods that can prevent direct and indirect discrimination at the same time without affecting original database. This will definitely helpful for maintaining privacy as well as security to the database by maintaining relationship between privacy preservation and discrimination prevention.

REFERENCES

[1] S. Hajian, J. Domingo- Ferrer, "A Methodology For Direct And Indirect Discrimination Prevention In Data Mining," Proc. IEEE transact. knowledge and data engineering, vol. 25, no. 7,pp.1041-4347, 2013.

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

[3] 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.

[4] 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.

[5] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "Rule Protection for Indirect Discrimination Prevention in Data Mining,"

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

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] S. Ruggeri, D. Pedreschi, and F. Turini, "Data Mining for Discrimination Discovery," ACM Trans. Knowledge Discovery from Data, vol. 4, no. 2, article 9, 2010.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.