ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Pre-Processing Approach for Discrimination Prevention in Data Mining

Mr. Pravin D. Kaware[*1], Prof. R. R. Shelke[2]
[*1,2] H.V.P.M's College of Engineering & Technology, Amravati, India
pravin.kaware@rediffmail.com

Abstract

Data mining is an important technology for extracting useful knowledge hidden in large collections of data. In data mining, 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. Especially when these type of attributes are used for decision making purpose such as giving them a job, loan. Insurance etc.. Discrimination can be either direct or indirect. 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. So we introduce an antidiscrimination techniques which including discrimination discovery and prevention. 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.

Keywords: Data mining, antidiscrimination, direct and indirect discrimination prevention, rule generalization, rule protection, privacy preservation.

Introduction

Data mining techniques are used in business and research are becoming more and more popular with time. There are, however, negative social perceptions about data mining, among which potential privacy invasion and potential discrimination. Discrimination can be viewed as the act of unfairly treating people on the basis of their belonging to a specific group. It involves denying to members of one group opportunities that are available to other groups.

There are several decision-making tasks which lend themselves to discrimination, For example, the European Union implements the principle of equal treatment between men and women in the access to and supply of goods and services in [1] or in matters of employment and occupation in [2]. Although there are some laws against discrimination, all of them are reactive, not proactive. The use of information systems based on data mining technology for decision making has attracted the attention of many researchers in the field of computer science. In consequence, automated data collection and a massive amount of data in data mining techniques such as association/classification rule mining have been designed and are currently widely used for making automated decisions. In [3], it is demonstrated that data mining can be both a source of discrimination and a means for discovering discrimination.

Discrimination can be either direct or indirect. Direct discriminatory rules indicate biased rules that are directly inferred from discriminatory items. Indirect discriminatory rules indicate biased rules that are indirectly inferred from non-discriminatory items because of their correlation with discriminatory ones.

Related Works

The existing literature on anti-discrimination in computer science mainly elaborates on data mining models and related techniques. Some proposals are oriented to the discovery and measure of discrimination. Others deal with the prevention of discrimination. The issue of antidiscrimination in data mining did not receive much attention until 2008 [8].Some proposals are oriented to the discovery and measure of discrimination. Others deal with the prevention of discrimination.

ISSN: 2277-9655
Impact Factor: 1.852

The discovery of discriminatory decisions was first proposed by Pedreschi et al. [3], [9]. Three approaches are conceivable:

• Preprocessing. Transform the source data 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 preprocessing approaches of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature [10], [11].

• In-processing. Change the data mining algorithms in such a way that the resulting models do not contain unfair decision rules. For example, an alternative approach to cleaning the discrimination from the original data set is proposed in [7].

• Post-processing**.** Modify the resulting data mining models, instead of cleaning the original data set or changing the data mining algorithms. For example, in [12].

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

Background

First, we recollect some basic definitions related to data mining [13]. After that, we concentrate on measuring and discovering discrimination.

A. Basic Definitions

•A data set is a collection of data objects (records) and their attributes. Let DB be the original data set.

• 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 →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, sup(X) is the fraction of records that contain the item set X. We say that a rule X → C is completely supported by a record if both X and C appears in the record.

•The confidence of a classification rule, conf(X → C), measures how often the class item C appears in records that contain X. Hence, if sup(X)> 0 then

$$\text{Conf}(X \rightarrow C) = \frac{supp\ (X,C,)}{\text{con f}(X)}$$

Support and confidence range over [0, 1].

•A frequent classification rule is a classification rule with support and confidence greater than respective specified lower bounds. Support is a measure of statistical significance, whereas confidence is a measure of the strength of the rule. Let FR be the database of frequent classification rules extracted from DB.

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 → 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

C. 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).

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

$$\text{Elift}\ (A, B \rightarrow C) = \frac{conf\,(A;B \rightarrow C)}{\text{con f}(B \rightarrow C)} \quad (2)$$

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). We call these rules direct α-discriminatory rules.

D. 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. To determine the redlining rules,

Pedreschi et al. in [3] stated the theorem.

ISSN: 2277-9655
Impact Factor: 1.852

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.

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 non redlining 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.

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. Definition5. 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:$conf\ r \geq p.(r')$
- Condition 2: $(r''{:}A,{\rightarrow}D){\geq}p$

C. Data Transformation for Indirect Discrimination

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 policy. To accomplish this, an appropriate data transformation with minimum information loss should be applied in such a way that redlining rules are converted to non-redlining rules. We call this procedure indirect rule protection (IRP).

D. Data Transformation for both Direct and Indirect Discrimination

We deal here with the key problem of transforming data with minimum information loss to prevent at the same time both direct and indirect discrimination. We will give a preprocessing solution to simultaneous direct and indirect discrimination prevention. First, we explain when direct and indirect discrimination could simultaneously occur. This depends on whether the original data set (DB) contains discriminatory item sets or not.

To provide both direct rule protection (DRP) and indirect rule protection (IRP) at the same time, an important point is the relation between the data transformation methods. Any data transformation to eliminate direct α-discriminatory rules should not produce new redlining rules or prevent the existing ones from being removed. Also any data transformation to eliminate redlining rules should not produce new direct α- discriminatory rules or prevent the existing ones from being removed. Indirect

discrimination also assumes that the background knowledge takes the form of classification rules connecting the item sets.

Utility Measures

The proposed solution should be evaluated based on two aspects:

• The success of the proposed solution in removing all evidence of discrimination from the original dataset (degree of discrimination prevention).

• The impact of the 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:

• Discrimination Prevention Degree (DPD).

This measure quantifies the percentage of α-discriminatory rules that are no longer α-discriminatory in the transformed dataset.

•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).

• 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).

• 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 in PPDM [14].

Conclusion

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.

References

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

[2] *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.*

[3] *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.*

[4] *S. Hajian, J. 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*

[5] *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.*

[6] *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*

[7] *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.*

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

[9] *S. Ruggieri, D. Pedreschi, and F. Turini, "Data Mining forDiscrimination*

ISSN: 2277-9655
Impact Factor: 1.852

Discovery," ACM Trans. Knowledge Discovery fromData, vol. 4, no. 2, article 9, 2010.

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

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

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

[13]P.N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining.Addison-Wesley, 2006.

[14]S. R. M. Oliveira and O. R. Zaiane. "A unified framework for protecting sensitive association rules in business collaboration". International Journal of Business Intelligence and Data Mining, 1(3):247287, 2006.