Preventing Discrimination In Data Mining

Bhagyesh P. Asatkar[1], Asst. Prof. Pushpanjali M. Chouragade[2]

Government College of Engineering, Amravati, India

[1]bhagyeshasatkar@gmail.com, [2]pushpanjalic3@gmail.com

***Abstract*—For extorting the helpful comprehension concealed in the biggest compilation of a database the data mining technology is used. Data mining is an increasingly important technology for getting useful knowledge hidden in large collections of data. Rules extracted from datasets by data mining techniques, such as classification or association rules, when used for decision tasks such as benefit can be discriminatory in the above sense. discrimination refers to unfair or unequal treatment of people on the basis of their membership to a category or a minority, without regard to individual merit. Discrimination is a very important issue when we are considering the legal and ethical aspects of data mining. It is more than obvious that most people don't want to be discriminated because of their gender, religion, nationality, age and so on, especially when these kind of attributes are used for making decisions about them like giving them a job, loan, insurance, etc. There are some negative approaches occurred about the data mining technology, among which the potential privacy incursion and potential discrimination. The latter consists of irrationally considering individuals on the source of their fitting to an exact similar group. Data mining and automatic data collection methods like, the classification covered the way for making the automated judgment like granting or denying the loan on the basis of race, creed, etc. If the training datasets are unfair in what respects discriminatory attributes like masculine category, race, creed, etc., discriminator decisions may ensue. Because of this reason the data mining technology introduced anti-discrimination methods with including the discovery of discrimination and its avoidance. The discrimination can direct or indirect. When any decisions were made to the sensitive attributes at that time direct discrimination are occurring. While the indirect discrimination are occurring when the decision are made on the basis of non-sensitive attributes which are strongly associated with the sensitive.**

*Keywords—**Direct discrimination prevention, , Indirect discrimination prevention, CPAR, rule protection, rule generalization.***

I. INTRODUCTION

If considering in a social sense, discrimination refers to an action based on prejudice resulting in unfair treatment of people, where the distinction between people is made on the basis of their membership to a category or minority, without regard to individual merit or circumstances. For example, in a certain organization black people might systematically have been denied from offering a jobs. As such, the previous employment information of this company concerning job applications will be biased towards assign jobs to white people while denying jobs from black people. The European Union implements the law of equal treatment between men and women in the access to and providing of goods and services in [3] or in matters of employment and occupation in [4] Services in the information society allow for automatic and collection of large amount of data. Those data are often used to train association/classification rules in view of making automated decisions, like loan granting/denial, personnel selection, insurance loan etc. the European Commission promoted legal studies that explains the importance of data collection and data analysis for the fight against discrimination. The probability of accessing to historical data regarding decisions made in socially-sensitive tasks is the starting point for discovering discrimination. However, if available decision data/records accessible for inspection increase, the data available to decision makers for making their decisions increase at a much higher quantity, together with more intelligent decision support systems, having capacity of assisting the decision process, and sometimes to operate the entire process automatically. As a result, it is most difficult task to discover the discriminatory situations and practices, hidden in the decision records which are under analysis, may reveal. Discrimination can be either direct or indirect (also called systematic). Direct discrimination is nothing but the rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive attributes of discrimination related to group membership. Personal data in decision records are emphasized by many multiple valued variables: as a result, a huge number of possible contexts may, or may not, be the responsible for discrimination. To understand this point, consider the case of gender discrimination in credit approval: although an analyst may observe that, their is no discrimination occurs in general, that is, when analyzing the whole available decision records, it may turn out that it is most difficult for aged women to obtain car loans. Many small or large problems may exist that conceal discrimination, and therefore all possible specific situations should be considered as candidates, containing of all possible combinations of variables and variable values: personal data, demographics, social, economic and cultural indicators, etc.

Clearly, a huge range of possibilities are faced by the anti-discrimination analyst , which make him work hard: the task of checking some known suspicious situations can be conducted using available statistical methods, the task of discovering problems of discrimination in the data is not supported. Indirect discrimination contains the rules or procedures that are not explicitly mentioning discriminatory attributes, intentionally or unintentionally may generate discriminatory decisions. Discrimination by financial

institutions is a typical example of indirect discrimination, but not the only one. With a slight abuse of for example, the race or ethnicity, is not directly recorded in the data. Nevertheless, discrimination on the basis of race may be equally hidden in the data, for instance in the case where a redlining practice is adopted: frequently denial of credit for people living in a certain neighborhood, and by demographic data we can say that most of people living in such a neighborhood belong to the same ethnic minority. Once again, the anti-discrimination analyst is faced with a huge space of possibly discriminatory situations: How can they mention all interesting discriminatory situations that are emerge from the data, both directly and in combination with further background knowledge in their possession (e.g., Sensitive data)?

II. Related work

The existing work on anti-discrimination in computer science mainly explained on data mining models and related techniques. The measures and discovery of discrimination some proposals are defined. And some remaining deal with the prevention of discrimination. Discrimination discovery can be possible by formalizing legal definitions of discrimination and proposing quantitative measures for it. This approach has been extended to encompass statistical significance of the extracted patterns of discrimination, and it has been implemented as reported. Data mining is a powerful technique for discrimination analysis, capable of discovering the patterns of discrimination that emerge from the data.

Three approaches for discrimination prevention consists of inducing patterns that do not lead to discriminatory decisions even if trained from a dataset containing them.

- Using the preprocessing approaches of data transformation and hierarchy based generalization from the privacy preservation literature.
- Make a Change in the data mining algorithms (in-processing) by enhancing the discrimination measure evaluations within them.
- For reducing the possibility of discriminatory decisions use the Post processing of data mining.

III. Methods

Discrimination prevention, the other major anti-discrimination aim in data mining, consists of inducing patterns which will do not lead to discriminatory decisions even if the original training data sets are biased. Three approaches are introduced.

A. pre-processing

Data preprocessing is an often neglected but important step in the data mining technique. The phrase "Garbage In, Garbage Out" is particularly applicable to data mining and machine learning. Data collecting methods are often loosely controlled, resulting in out of range values(e.g., Income: -200), impossible data combinations such as Gender:-Male, Pregnant:- Yes, missing values, etc. Analyzing data that has not been carefully handled for such problems can create misleading results. Thus, the representation and quality of data is most important thing that has to be run before analysis. If there is much irrelevant and redundant information present or noisy and unreliable data, then during the training phase the knowledge discovery is more difficult. Data preparation and filtering steps can take large amount of processing time. Data pre-processing contains cleaning, normalization, transformation, feature like selection and extraction, etc. The product of data pre-processing is the final training set[3].It change the source data in such a way that the discriminatory records that are includes in the original data are removed so that no unfair decision rules can be mined from the transformed data and apply any of the standard data mining algorithms. The privacy preservation technique is use for the preprocessing approaches of data transformation and hierarchy-based generalization. Along this line, perform a controlled distortion of the training data from which a classifier is learned by making minimally intrusive modifications in data leading to an unbiased data set. The pre processing approach is useful for applications in which a data set should be published and in which data mining needs to be performed also by external parties (and not just by the data holder). Raw data is highly susceptible to noise, missing values, and inconsistency. The quality of data, affects the data mining results. In order to help improve the quality of the data and, consequently, of the mining results raw data is preprocessed so as to improve the efficiency and ease of the mining process.

Data preprocessing methods are divided into following categories[13]:

1. Data Cleaning
2. Data Integration
3. Data Transformation
4. Data Reduction

B. In-processing

The data mining algorithm is updated in such way that the unfair decision rule does not contained by the final resulting models. For example, for cleaning the discrimination from the original data set is proposed[9], though which the nondiscriminatory restriction is embedded into a decision tree learner by altering its gashing criterion and trimming strategies during a novel leaf relabeling approaches. However, it is obvious that in-processing discrimination prevention methods must depends on new special-purpose data mining algorithms; standard data mining algorithms cannot be used because they has to be adapted to satisfy the non-discrimination requirement[3].

C. Post-processing

The post-processing modifies the resulting data mining models it does not cleans the genuine data set or alters the data mining algorithm. For example, in, a confidence alternative approach is proposed for classification rules inferred by the CPAR algorithm. The authority to publishing

the data is to the modified data mining models only, not for the post-processing, hence data mining can be performed by the data holder only. One might think of a straightforward preprocessing approach consisting of just removing the discriminatory attributes from the data set. Although this would solve the direct discrimination problem, it would cause much information loss and in general it would not solve indirect discrimination. As stated in there may be other attributes (e.g., Zip) that are highly correlated with the sensitive ones (e.g., Race) and allow inferring discriminatory rules.

Hence, for discrimination prevention there are two important challenges:

- the first is to consider both direct and indirect discrimination instead of only direct discrimination.
- The second challenge is to find a good tradeoff between discrimination removal and the quality of the resulting training data sets and data mining models.

In spite of some methods have already been introduced for each of the above-mentioned approaches (pre-processing, in-processing, post processing), discrimination prevention stays a largely unexplored research avenue. In this paper, we are concentratinge on discrimination prevention based on preprocessing approach, because the preprocessing approach acts the most flexible one: it does not require updation in the standard data mining algorithms, unlike the in-processing approach, and it allows only data publishing (rather than just knowledge publishing), unlike the post processing approach[3].

IV. METHODOLOGYIES USED

Discrimination prevention can be done in three ways based on when and in which phase data or algorithm is to be changed. There are three ways for preventing discrimination: Pre-processing method, In-processing method and Post-processing method. Discrimination can be of 3 types: Direct, Indirect or combination of both, categorized by the presence of discriminatory attributes and other attributes that are strongly related with discriminatory one.

A. Discrimination Prevention by Pre-processing Method

The preprocessing method is use to remove direct and indirect discrimination from original dataset. It employees 'elift' as discrimination measure to prevent discrimination in crime and intrusion detection system, there is method based on “data massaging” proposed by Kamiran and Calder, where class label of some of the records in the dataset is changed but as this method is intrusive, concept of "Preferential sampling" was introduced. This method uses Ranking function and there is no need to change the class labels. This method first divides data into four groups that are DP, PP, DN and PN, where first letters D and P are for Deprived and Privileged class respectively and second letters P, N are for positive and negative class label. The ranker function then sorts data in ascending order with respect to non-negative(positive) class label. Later it changes sample size in respective group to make that data biased free. Preprocessing method is very useful in applications where data mining is to be performed by third party and data needs to publish for public usage.

There are four preprocessing techniques are being used for prevention of discrimination as given below:

1. *Suppression:* It finds those attributes which correlate most with the sensitive attribute S. Then remove sensitive attribute S and most correlated attribute, to minimize the discrimination between the class levels and attribute.
2. *Massaging the dataset:* By changing the labels of some objects in dataset we can remove discrimination from the dataset. The best candidates for relabeling can be select with the help of ranker.
3. *Reweighing:* Instead of change in some of the labels of some objects, allocating the weights in training data set’s tuples. For making dataset by carefully assigning the weights and the training data set can be made discrimination free without changing the labels in the dataset.
4. *Sampling:* This method used where weights cannot be used directly. Sample sizes for the four combinations of sensitive attribute S and values of class will make the dataset discrimination free. Applying samples for four groups will make two of the groups are under sampled and two will be over sampled. Then with help of two techniques, Uniform Sampling and Preferential sampling for selecting the objects to matching, and to remove. The above four methods are based on preprocessing the dataset after which any standard classification tools can be used.

B. Discrimination Prevention by In-processing Method

In-processing method based on decision tree where data mining algorithm is modified instead of modifying original dataset. Thos method is introduced by Faisal Kamiran, Toon Calders and Mykola Pechenizkiy. This method consists of two techniques for the decision tree construction process, first is Dependency-Aware Tree Construction and another is Leaf Relabeling. The Dependency-Aware Tree Construction technique focuses on splitting criterion for tree construction to construct a discrimination aware decision tree. For working properly, it first calculates the information gain with respect to class & sensitive attribute represented by IGC and IGS respectively.

For determining the best split that uses different mathematical operation, there are three alternative criteria: (i) IGC-IGS; (ii) IGC/IGS; (iii) IGC+IGS.

The IGC/IGS approach consists of processing of decision tree with discrimination-aware pruning and it re-label the tree leaves.

C. Discrimination Prevention by Post-processing Method

In post processing method resultant mining model is modified instead of modifying original data or mining algorithm but there are some disadvantages of this method and that are, this method do not allow for public usage original data is to be

published, also the task of data mining should be performed by data holder only. The naive bayes classifier is modified to perform classification that is independent with respect to a given perspective attribute this approach is proposed by Toon Calders and Sicco Verwer. There are three approaches(viewpoints) in order to make the naive bayes classifier discrimination free: (i) was modifying the probability of the decision being positive where the possibility of distribution of the sensitive attribute is modified. This method has disadvantage of either always increasing or always decreasing the number of positive labels assigned by the classifier, depending on how commonly the sensitive attribute is present in dataset, (ii) training one model for every sensitive attribute value and balancing them. This is done by partitioning the dataset into two separate sets and the model is learned generaly only the tuples from the dataset that have a favored sensitive value, (iii) adding a latent variable to the Bayesian model. This method, models the actual class labels using a latent variable. Sara Hajian, Anna Monreale, Dino Pedreschi ,Josep Domingo Ferrer proposed post processing method that derive frequent classification rule and modifies the final mining model using α-Protective k-Anonymous pattern sanitization to remove discrimination from Mine Model.

V. CONCLUSION

The perseverance of this topic was to develop unique post-processing discrimination prevention method. On the basis of a review of existing laws, we have studied and formalized a family of discrimination measures for classification rules. Discriminatory classification rules are the basic tool for uncovering direct and indirect discriminatory decisions in datasets of historical records, and in the output of classifiers. As a forthcoming work, discovering trials of discrimination dissimilar from the ones reflected in this topic laterally with privacy conservation in data mining.

.

REFERENCES

[1]"Data Preprocessing Techniques for Data Mining", online available:http://www.iasri.res.in/ebook/win_school_aa/notes/Data_Preprocessing.pdf

[2] E. C. 2006, "EU Directive 2006/54/EC on Anti-Discrimination,"2006.[Online].Available:http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:204:0023:0036:en:PDF

[3] D. Pedreschi, S. Ruggieri, and F. Turini, "Discrimination-aware data mining," in Proc. 14th ACM Intl Conf. Knowledge Discovery and Data Mining (KDD 08), 2008, pp. 560–568.

[4] S. Ruggieri, D. Pedreschi, and F. Turini, "Data mining for discrimination discovery," in ACM Trans. Knowledge Discovery from Data, vol. 4,2010.

[5]U.S.Congress,"USEqualPayAct,"1963.[Online].Available:http://archive.eeoc.gov/epa/anniversary/epa-40.html

[6] D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring discrimination in socially-sensitive decision records," in Proc. Ninth SIAM Data Mining Conf. (SDM 09), 2009, pp. 581–592.

[7] F. Turini, D. Pedreschi, and S. Ruggieri, "Integrating induction and deduction for finding evidence of discrimination," in Proc. 12th ACMIntl Conf. Artificial Intelligence and Law
(ICAIL 09), 2009, pp. 157–166.

[8] S. Ruggieri, D. Pedreschi, and F. Turini, "Dcube: Discrimination discovery in databases," in Proc. ACM Intl Conf. Management of Data(SIGMOD 10), 2010, pp. 1127–1130.

[9] F. Kamiran and T. Calders, "Classification without discrimination," inProc. IEEE Second Intl Conf. Computer, Control and Comm.(IC4 09),2009, p. 2009.

[10] T. Calders and F. Kamiran, "Classification with no discrimination by preferential
sampling," in Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

[11] T. Calders and S. Verwer, "Three naive bayes approaches for discrimination-free classification," in Data Mining and Knowledge Discovery, vol. 21, 2010, pp. 277–292.

[12] S. Hajian and J. Domingo-Ferrer, "A methodology for direct and indirect discrimination prevention in data mining," in Ieee Transactions On Knowledge And Data Engineering, vol. 25, 2013, pp. 11451157.

[13] E. C. 2004, "EU Directive 2004/113/EC on Anti-Discrimination," 2004. [Online]. Available:http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:373:0037:0043:
EN:PDF