The International Journal Of Engineering And Science (IJES)
|| Volume || 3 || Issue || 6 || Pages || 12-16 || 2014 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Preprocessing Technique for Discrimination Prevention in Data Mining

[1,]Jagriti Singh , [2,]Prof. Dr. S. S. Sane

[1,]*Department of Computer Engineering, K.K.Wagh Institute of Engineering Education and Research, Nashik, University of Pune, Maharashtra – 422003, India*

[2,]*Head of Department, Department of Computer Engineering, K.K.Wagh Institute of Engineering Education and Research, Nashik, University of Pune, Maharashtra – 422003, India*

----------ABSTRACT----------

Data mining is an important technology for extracting useful knowledge in large collections of data. To extract knowledge without violation such as privacy and non-discrimination is most difficult and challenging. Unjustified distinction of individuals based on their membership in a certain group or category are referred as discrimination. In this paper, we will learn a classifier that optimizes accuracy of its predictions on test data, but does not have discrimination. Data preprocessing techniques such as massaging the dataset by changing class labels, and reweighing or re-sampling the data, help in removing the discriminations or biased data.

KEYWORDS : *Anti Discrimination, Data Mining, Discrimination Prevention Technique, Preprocessing Technique, Massaging*

Date of Submission: 11 June 2014 Date of Publication: 20 June 2014

I. INTRODUCTION

Data mining is one of most important and useful technology in today world for extracting useful knowledge in large collections of dataset. Most of the organizations are having a large number of dataset but to extract useful and important knowledge is very difficult and extracting knowledge without violation such as privacy and non-discrimination is most difficult and challenging. Privacy refers to the individual right while discrimination refers to unfair or unequal treatment of people. From a legal perspective, discrimination arises only on application of different rules or of the same rule or practice to different situations or practices to comparable situations. There are two kind of discrimination. When any rules or practices explicitly favor one person than another, is known as direct discrimination and also called systematic discrimination. An apparently neutral provision, practice or criterion which results in an unfair treatment of a protected group called as indirect discrimination and also called as disparate impact. Government plays a vital role in the prevention and reduction of discriminations, by enforcing different type of anti-discrimination laws.

II. LITERATURE SURVEY

Discrimination discovery is the discovery of discriminatory situations and practices hidden in a large amount of historical decision records. The aim of discrimination discovery is to unveil contexts of possible discrimination. Discrimination prevention has been recognized as an issue in a tutorial by (Clifton, 2003) [1] where the danger of building classifiers capable of racial discrimination in home loans has been put forward. Data mining and machine learning models extracted from historical data may discover traditional prejudices for example, mortgage redlining can be easily recognized as a common pattern in loan data but so solution was provided in this tutorial. D. Pedreschi, S. Ruggieri and F. Turin [2] are the first to address the discrimination problem in data mining models. The problem of discrimination in data mining in a rule-based setting was tackled, by introducing the notion of discriminatory classification rules, as a criterion to identify the potential risks of discrimination. D. Pedreschi, S. Ruggieri and F. Turini [3] also stored out of a socially-sensitive decision task on a systematic framework for measuring discrimination, based on the analysis of the historical decision records, e.g. credit approval. They had worked on discovery of discrimination in a given set of decisions, by measuring the degree of discrimination of a rule that formalizes an expert's hypothesis. They had also implemented LP2DD [3] approach by integrating induction and deduction for finding evidence of discrimination of the overall reference model in 2009.

The problem of impartial classification was tackled by introducing a new classification scheme for learning unbiased models on biased training data in 2009 by F Kamiran, T Calders [5]. This method is based on massaging the dataset by making the least intrusive modifications which lead to an unbiased dataset. In this method the numerical attributes and group of attributes are not considered as sensitive attribute. How to modify the Naive Bayes classifier in order to perform classification that is restricted to be independent with respect to a given sensitive attribute has investigated by T Calders, S Verwer [6]. If independency restrictions occur naturally, the decision process leading to the labels in the data-set was biased; e.g., due to gender or racial discrimination. A conceptual foundation of data mining with incomplete data through classification was presented in 2010by H Wang, S Wang [7] which was relevant to a specific decision making problem. The proposed technique was to detect the interesting patterns of data missing behavior that are relevant to a specific decision making, instead of estimation of individual missing value. This focuses on to get high accuracy scores but do not take the discrimination into account.The discrimination discovery and prevention problems by a variant of k-NN classification was modeled by B Luong, S Ruggieri, F Turini [9] in 2011that implements the legal methodology of situation testing. Major advancements over existing models: a stronger legal ground, a global description of who is discriminated and who is not; overcoming the weaknesses of aggregate measures over undifferentiated groups; a discrimination prevention method.

III. DISCRIMINATION PREVENTION TECHNIQUES IN DATA MINING

The discovery of discriminatory situations and practices, hidden in a dataset of historical decision records, is an extremely difficult task due to typically highly dimensional data and indirect discrimination dataset. As most of effective decision models of data mining are constructed on the basis of historical decision records and automated, this extracted knowledge incur with discrimination. This may be because the data from which the knowledge is extracted contain patterns with discriminatory bias. Hence, data mining from historical data may lead to the discovery of traditional prejudices. Thus prevention of discrimination knowledge based decision support systems; discovery is a more challenging issue.

There are three different approaches for discrimination prevention in data mining:

Preprocessing: Removing of discrimination from original source data in such a way that no unbiased rule can be mined from the transformed data and applying any standard algorithm. This preprocessing approach is useful in such cases where data set should be published and performed by external parties.

In-processing: Change of knowledge discovery algorithm in such a way that resulting model do not contain biased decision rules. In-processing discrimination prevention depends on new special purpose algorithm. In this standard data mining algorithm cannot be used.

Postprocessing: Instead of removing biases from original data set or modify the standard data mining algorithm, resulting data mining models are modified. This approach does not allow the data set to be published, only modified mining models can be published. So this can be performed only by data holder.

A straight forward approach to avoid that the classifier's prediction be based on the discriminatory attribute would be to remove that attribute from the training dataset. This approach, however, does not work. The reason is that there may be other attributes that are highly correlated with the discriminatory one. In such a situation the classifier will use these correlated attributes to indirectly discriminate. The direct implementation of preprocessing techniques also will remove the discriminatory attribute from the original data set and solve the problem of discrimination but not solve the indirect discrimination. This would also cause much of information loss from the original data set. In this paper, we are focusing on the existing preprocessing techniques for discrimination prevention and the modified technique for discrimination prevention.

IV. PREPROCESSING TECHNIQUES FOR DISCRIMINATION PREVENTION

As stated above a straight forward approach to remove the discriminatory attribute does not work since other attributes are highly correlated with the discriminatory one. There are four preprocessing techniques are being used for prevention of discrimination as given below:

Suppression: Finding the attribute which correlate most with the sensitive attribute S. Remove S and most correlated attribute, to reduce the discrimination between the class levels and attribute.

Massaging the dataset: Discrimination can be removed from the dataset by changing the labels of some objects in dataset. The best candidates for relabeling can be select with help of ranker.

Reweighing: Instead of change in some of the labels of some objects, assigning the weights in training data set's tuples. By carefully assigning the weights, the training data set can be made discrimination free without changing the labels in the dataset.

Sampling: This method can be used where weights cannot be used directly. Sample sizes for the 4 combinations of sensitive attribute S- and Class-values will make the dataset discrimination free. Applying stratified sampling on the four groups will make two of the groups as under sampled and two will be over sampled. Then with help of two techniques, Uniform Sampling and Preferential Sampling for selecting the objects to duplicate, and to remove.The above four methods are based on preprocessing the dataset after which any standard classification tools can be used.

V. PROPOSED WORK

The proposed solution is to learn a non-discriminating classifier which use the sensitive attribute S only learning time and not at prediction time. The solution is for removing discrimination from training dataset. Logic for this approach is this, the classifier is learned on discrimination-free data, it is likely that its prediction will be more discrimination-free as well.

Assume a set of attributes

- $A = \{A_1, A_2, \ldots, A_n\}$ is a set of attributes with domains $dom(A_i)$, $i = 1, \ldots, n$.
- A tuple X is an element of $dom(A_1) \times \ldots \times dom(A_n)$ over the schema$(A_1, \ldots, A_n)$.The value of X for attribute A_i is denoted by $X(A_i)$.
- A dataset D is a finite set of tuples over scheme $(A_1, \ldots, A_n)$.
- A label dataset over schema (A1, . . ., An, class) is a finite set of tuples.
- Assume class has binary domain dom(class) = { - , +} where "+" is a desirable class.
- A special attribute $S \in A$, is sensitive attribute with multiple values.
- P is set of favored community values and Q is set of deprived community values.
- Domain of S is dom(S) = {P,Q}.
-

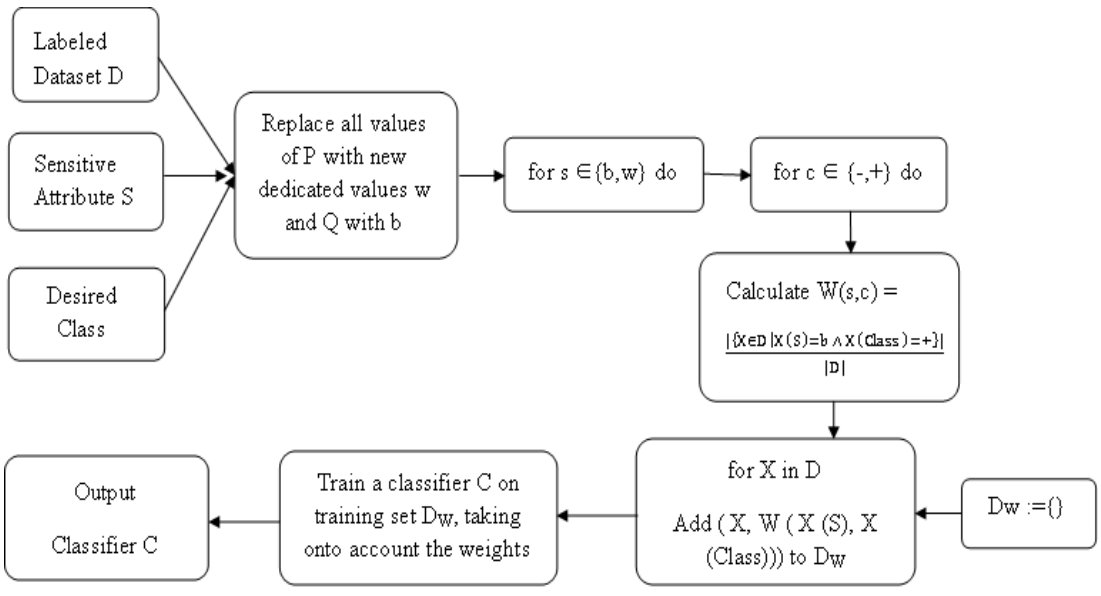

Fig.1 Block Diagram of Proposed Preprocessing Technique

In original dataset D, sensitive attribute S is categorical and its domain is non-binary. In dataset Class has binary domain dom(Class) ={-,+} where "+" is desirable class. The values of sensitive attribute for favored community is replaced with new dedicated value w and the values for deprived community with new dedicated value b.

In this approach different weight will be attracted to each objects in dataset. For example, the objects with X(S) = b and X(class) = - will get lower weight than object with X (S) = b and X (Class) = + and objects with X (S) = w and X (Class) = − will get higher weight than X (S) = w and X (Class) = +.

Weight calculation of objects can be defined as follows:

- If the dataset D is unbiased, i.e., S and Class are statistically independent, the expected probability Pexp(S = b ∧ Class = +) would be:

$$\text{Pexp}(S = b \wedge \text{Class} = +) \coloneqq \frac{|\{X \in D \mid X(S) = b\}|}{|D|} \times \frac{|\{X \in D \mid X(\text{Class}) = +\}|}{|D|}$$

But actually, the observed probability in D might be different to expected probability as below,

$$\text{Pobs}(S = b \wedge \text{Class} = +) \coloneqq \frac{|\{X \in D \mid X(S) = b \wedge X(\text{Class}) = +\}|}{|D|}$$

If the observed probability is lower than expected probability value, it mean the bias toward class – for those objects X with X(S) = b.

To compensate for the bias, assign lower weights to objects that have been favored or deprived. Every object X will be assigned weight:

$$W(X) \coloneqq \frac{\text{Pexp}(S = X(S) \wedge \text{Class} = X(\text{Class}))}{\text{Pobs}(S = X(S) \wedge \text{Class} = X(\text{Class}))}$$

VI. RESULTS AND DISCUSSION

6.1 Dataset

In experiment use the German credit dataset, available in the UCI ML repository. The German credit dataset has 1000 instances which classify the account holders into credit class. Each data object will be described by 21 attributes which include 14 categorical and 7 numerical attributes.

6.2 Performance Analysis

It may be suggested that removing sensitive attributes from databases is the best solution to prevent unethical or illegal discriminating data mining results. If sensitive attributes such as gender, ethnic background, religious background, sexual preferences, criminal and medical records are deleted from databases, the resulting patterns and relations cannot be discriminating anymore, it is sometimes argued. However, recent research has shown that this assumption is not correct.

VII. CONCLUSION

This paper has presented the Reweighing method of pre-processing technique for removing the discrimination and subsequently a classifier is learned on this unbiased data. This method will be extended for more sensitive attributes, deal with numerical sensitive attributes. With the proposed method, discrimination must be detected and removed to get the unbiased results.

REFERENCES

[1]C. Clifton. Privacy preserving data mining: How do we mine data when we aren't allowed to see it? *In Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2003)*, Tutorial, Washington, DC (USA), 2003.
[2]D. Pedreschi, S. Ruggieri and F. Turini, "Discrimination-aware Data Mining," *Proc. 14th Conf. KDD* 2008, pp. 560-568. *ACM*, 2008.
[3]D. Pedreschi, S. Ruggieri and F. Turini, "Measuring discrimination in socially-sensitive decision records," *SDM* 2009, pp. 581-592. *SIAM*, 2009.
[4]D. Pedreschi, S. Ruggieri and F. Turini, "Integrating induction and deduction for finding evidence of discrimination," *ICAIL* 2009, pp. 157-166. *ACM*, 2009.
[5]F Kamiran, T Calders, "Classifying without discriminating," *Proceedings of IEEE IC4 International conference on Computer, Control and Communication*. (2009a) IEEE Press.
[6]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.
[7]H Wang, S Wang, "Mining incomplete survey data through classification," Knowl Inf Syst, 1-13, 2010.
[8]F. Kamiran, T. Calders and M. Pechenizkiy," Discrimination aware decision tree learning," *ICDM* 2010, pp. 869-874. *IEEE*, 2010
[9]B Luong, S Ruggieri, F Turini, "K-nn as an implementation of situation testing for discrimination discovery and prevention," *Technical Report TR*-11-04: (2011), Dipartimento di Informatica, Universita di Pisa.
[10] F. Kamiran and T. Calders, "Data preprocessing techniques for classification without discrimination," *Knowledge and Information Systems*, 33:1–33, *Springer*, 2012.

[11] T. Calders and I. I. Zliobaite, "Why unbiased computational processes can lead to discriminative decision procedures," Discrimination and Privacy in the Information Society (eds. B. H. M. Custers, T. Calders, B. W. Schermer, and T. Z. Zarsky), volume 3 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*, pp. 4357. *Springer*,2013.
[12] B. Custers, T. Calders, B. Schermer and T. Z. Zarsky (eds.), "Discrimination and Privacy in the Information Society - Data Mining and Profiling in Large Databases," *Studies in Applied Philosophy, Epistemology and Rational Ethics* 3, *Springer*, 2013.
[13] A. Romei, S. Ruggieri, "A multidisciplinary survey on discrimination analysis," *The Knowledge Engineering Review*, 2013.

Jagriti Singh is Post Graduate Student of Department of Computer Engineering, K.K.Wagh Institute of Engineering Education and Research, Nashik, University of Pune, Maharashtra, India. She received her B.Tech. Degree in Information Technology from Uttar Pradesh Technical University, Uttar Pradesh.
(email id: pritijagriti@gmail.com)

Prof. Dr. S. S. Sane is the Head of Computer Engineering Department, K.K.Wagh Institute of Engineering Education and Research, Nashik, University of Pune, Maharashtra, India.