A Survey on Discrimination Prevention Methods in Data Mining

Shruti A. Vagare 1, P.D. Thakare 2

1Department of Computer science & Engineering, JCOET Yavatmal, India, vagareshruti@gmail.com, 8856046076
2Professor, Department of Computer Engineering, JCOET Yavatmal, India, paragthakare2003@yahoo.co.in ,9764996757

Abstract- In the database large amount of data is collected and that data is analyzed using various data mining techniques. In data mining various issue are there like potential discrimination and privacy invasion. Discrimination is a partial treatment of individual or group based on their actual or perceived membership in certain group or category. Discrimination can be done on attributes like religion, nationality, marital status and age. In data mining classification technique is used for decision making. If training dataset are biased against particular community or nationality then it can be generate discriminatory decision. Two types of discrimination are direct and indirect. Decision based on sensitive attribute called as direct discrimination and decision which is based on non sensitive attribute but highly correlated with sensitive ones. To prevent such types of discrimination , discrimination measurement and different data transformation methods used.

Keywords: Antidiscrimination, data mining, direct and indirect discrimination prevention, rule protection, rule generalization.

Introduction

Data mining is technology which is used for retrieving knowledge which is hidden in large collection of data. There are negative perceptions about data mining. As discrimination is big issue in data mining. Discrimination does not grant opportunities which is available to the member of one group to member of another group .Discrimination can be done on sensitive attributes of individual like gender, nationality, religion, age, marital status etc. There are various systems like loan granting and premium computation in which large amount of data is collected and analysed for decision making process in data mining technology. In these system classification or association rule can be used for purpose of decision making. Accordingly, approaches for discrimination prevention fall into following three types

1) **Pre-processing**:- Approaches belonging to this group 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.

2) **In-processing**:- Approaches belonging to this group change the data mining algorithms in such a way that the

resulting models do not contain unfair decision rules.

3) **Postprocessing**:- Approaches belonging to this group modify the resulting data mining models, instead of cleaning the original dataset or changing the data mining algorithms.

To tackle the problem of discrimination, different types of new techniques were proposed. Discrimination comes under two categories one is direct and second is indirect. Decisions based on sensitive attributes are termed as direct discrimination and the decisions which are based on non-sensitive attributes are termed as indirect discrimination which is strongly correlated with biased sensitive once. In this paper we discussed various existing approaches for discrimination prevention and analysis of existing approaches of discrimination prevention.

Related Work

Here are the existing approaches for discrimination prevention in data mining technology：

1. Fast Algorithms for Mining Association Rules in Large Databases

These algorithms proposed by R. Agrawal and R. Srikant. Focuses in discovering association rules between items in large data base of transaction [1]. In this data mining is used to solve decision support problem which is occurred in retail organization. In retail organization bar code technology is used to collect and store large amount of sales data called as basket data. It contains the information about transaction date and items bought in transaction. This information is used by marketers to implement customize marketing strategies. The problem of mining association rule is introduced , cause these rules are important in cross marketing.

The Problem of finding out such association rules divided in to two phases in first phase find out all item sets that have transaction support greater than minimum support. Support is nothing but the number of transaction that contain the item sets, the item sets with minimum support are called as large item set and all other are small item sets. In second phase use large item sets to find out such rules.

Advantages of these algorithms are as it works on synthetic as well as real-life data, In this Execution time decreases a little as the number of item increases and disadvantages are As transaction size increases execution time increases gradually & did not consider the quantities of the items bought in a transaction.

2. Discrimination Aware data mining

This idea is proposed by Dino Pedreschi, Salvatore Ruggieri and Franco Turini. In this discrimination is nothing but unequal treatment of people based on their membership to a particular group or category. Rules extracted from databases by data mining techniques, such as classification or association rules, these rules are used for decision making tasks such as benefit or credit approval, can give discriminatory decision [2]. In this notion of discriminatory classification rules is introduced and studied. Discrimination can be direct or indirect discrimination. Direct discrimination consists of rules or procedures that explicitly mention disadvantaged groups. Indirect discrimination consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, intentionally or not impose the same disproportionate burdens. Direct discrimination is modelled through potentially discriminatory rules, which are classification rules A,B->C that contain potentially discriminatory item sets A in

their premises. Concerning indirect discrimination, we consider rules D, B-> C that are potentially non-discriminatory (PND), that do not contain potential discriminatory (PD) item sets. In this extended lift is introduced as measure of discrimination.

3. Three Naive Bayes Approaches for Discrimination-Free Classification

In this approach training data contains unwanted dependencies between the attributes. Given a labeled dataset and a sensitive attribute like ethnicity, the goal of this research is to learn a classifier for predicting the class label that does not discriminate with respect to the sensitive attribute[3]. This setting is motivated by many cases in which there exist laws that disallow a decision that is partly based on discrimination. It present three approaches for making the Naive Bayes classifier discrimination free: i) modifying the probability of the decision being positive, ii) training one model for every sensitive attribute value and balancing them and iii) adding a latent variable to the Bayesian model that represents the unbiased label and optimizing the model parameters for likelihood using expectation maximization. Main advantage is resulting decision is discrimination free but it does not consider numerical attributes as a sensitive attribute & does not identify indirect discrimination

4. Data mining for Discrimination discovery

In this paper discrimination refers to unfair treatment of people based on belonging to some disadvantaged group without looking their merits. Discrimination in credit, mortgage, insurance, labor market, and education has been investigated by researchers in economics and human sciences. This paper introduce the problem of discovering discrimination through data mining in a dataset of historical decision records[4], taken by humans or by automatic systems. It formalize the processes of direct and indirect discrimination discovery by modeling protected by law groups and contexts where discrimination occurs in a classification rule based syntax. The basic problem addressing can be as follows a dataset of historical decision records, |a set of potentially discriminated groups, |and a criterion of unlawful discrimination. This paper tackle the problem of discovering discrimination within a rule-based setting, by introducing the notion of discriminatory classification rules to find out the potential risk of discrimination. By mining all discriminatory classification rules from a dataset of historical decision records, practical method are used to find out direct and indirect discrimination hidden in the data, as well as a criterion to measure discrimination in any such contexts.

5. Rule protection for indirect discrimination prevention in data mining

In the information society different types of services are used to collect automatically large amounts of data. That data are used to train classification rules to make automated decisions, like loan granting or denial, insurance premium computation, etc[5]. If the training datasets are biased against sensitive attributes like gender, race, religion, etc. then discriminatory decisions may occur. Direct discrimination occurs when decisions are made based on biased sensitive attributes. Indirect discrimination occurs when decisions are made based on non-sensitive attributes which are strongly correlated with biased sensitive attributes. This paper discusses how to clean training datasets so that resulting datasets does not give indirect discriminating rules. Main contribution of this paper is proposed new pre-processing approach based method for indirect discrimination which is based on data

transformation. It also introduced some measure for finding out success of proposed method and impact on data quality.

6. Classification with no discrimination by prefential sampling

T.Calders and F.kamiran [6] present a new approach which focuses on the concept of classification without discrimination. They introduced an idea of classification with no discrimination. They give solution by messaging the data to remove the discrimination from it with minimum changes. An authors also present solution for CND problem by introducing sampling scheme for making the data discrimination free rather than relabeling the dataset.

7. A survey of Association rule hiding methods for privacy

This is survey of different approaches that are used for privacy preservation by hiding association rule given by Vassilios S. Verykios and Aris Gkoulalas-Divanis. In this Data and knowledge hiding are two research directions that investigate how the privacy of raw data, or information, can be maintained either before or after the course of mining the data. It presents survey of recent approaches that have been applied to the association rule hiding problem. Association rule hiding refers to the process of modifying the original database in such a way that certain sensitive association rules disappear without seriously affecting the data and the non-sensitive rules [7].
In this three approaches mention for association rule hiding. First one is Heuristic approach which is based on data distortion and data blocking method. In data distortion sensitive item is replaced by its negated item that is 1 replaced by0 and vice versa. In data blocking method sensitive item is replaced by question mark (?) to prevent its privacy. Second is Border based approach in that quality of the borders directly affects the quality of the sanitized database & in exact approach it is capable of providing superior solutions but at a high computational cost.

8. A Methodology for Direct and Indirect Discrimination Prevention in data mining

Sara Hajian and Josep Ferror [8] proposed pre-processing approach for discrimination prevention. They introduced new data transformation method like rule protection and rule generalization. This new data transformation handles both type of discrimination that is direct and indirect discrimination. It can handles several discriminatory item sets. Thus, based on various issues and limitations discussed in different existing approaches for discrimination prevention , new data transformation method need to be designed which prevent direct and indirect discrimination or both at same time as well as prevent discrimination present on right hand side of classification rule cause discriminatory item sets present on right hand side of classification rule whose confidence and support greater than minimum confidence and minimum support of classification rule. Therefore it can be give discriminatory decision. To meet this objective steps given below need to be carried out.

a. First measure discrimination of both types that is direct and indirect discrimination. Make groups of individuals that have been directly and indirectly discriminated in decision making process.

b. Transform data in proper way to remove discriminatory biases.

c. Data set which is discrimination free generated by using data transformation method without harming data quality.

9. Discrimination prevention in data mining for intrusion and crime detection

Antidiscrimination law refers to the law on the right of people to be treated equally. In the political participation people must be dealt with on equal basis in any case of sex, age, race, nationality. The approaches are used preprocessing, post processing. The preprocessing is data preprocessing is the important process in the data mining. In there is much irrelevant and redundant information present or noisy and unreliable data, and then knowledge discovery during the training phase is more difficult. The analyzing data that has not been carefully screened for such problems can produce misleading results. The post processing is data mining is the process of sorting through large amounts of data and picking our relevant information .Data mining in relation to enterprise resource planning is the statistical and logical analysis of large sets of transaction data.

Conclusion

In this paper, we have carried out a wide survey of the different approaches for discrimination prevention, and analyses. This paper presents a new pre-processing discrimination prevention method. Different transformations are used for the discovery of discrimination. The process measures the discrimination and identifies the categories by decision-making processes. Discrimination-free data models can be produced from the transformed data set without seriously damaging the data quality. More data's can be handled and the system result is trustworthy.

References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] F.Kamiran,T.Calders, *Classification with no discrimination by preferential sampling,*Proc.19th Machine learning conf.Belgium and The Netherlands,2010.

[7] V. Verykios and A. Gkoulalas-Divanis, *A Survey of Association Rule Hiding Methods for Privacy,*Privacy-Preserving Data Mining: Models and Algorithms, C.C. Aggarwal and P.S. Yu, eds. Springer, 2008.

[8] Sara Hajian and Joseph 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.