International Journal of Research in Advent Technology, Vol.4, No.4, April 2016
E-ISSN: 2321-9637
Available online at www.ijrat.org

Prevention of Direct and Indirect Discrimination using Rule Protection and Rule Generalization

Shruti A. Vagare 1, Prof. P.D. Thakare 2
1PG Student, Department of CSE, JCOET Yavatmal, India, vagareshruti@gmail.com,
2Professor, Department of CE, JCOET Yavatmal, India, paragthakare2003@yahoo.co.in

Abstract- In database large amount of data is collected then that data is analyze using various data mining techniques. In data mining issues are there like potential discrimination and privacy invasion. Discrimination can be done on attributes such as religion, nationality, marital status and age. In data mining classification technique is used for decision making. There are two types of discrimination as direct and indirect. Decision is based on sensitive attribute known as direct discrimination where as decision which is based on non sensitive attribute but highly correlated with sensitive ones. Anti-discrimination methods are adopted to eliminate the discrimination in data mining.

Index Terms : Data mining, discrimination discovery, discrimination prevention, rule protection and rule generalization**.**

1. INTRODUCTION

Data mining technology which is used for retrieving knowledge which is hidden in large collection of data. Data mining involved the extraction of implicit previous unknown and also potentially data which is a useful knowledge from large databases. The process of performing data analysis could reveal important data patterns,that leads to adapt business strategies, knowledge bases, and scientific research.

Discrimination becomes an important issue when assumping the legal and ethical aspects of data mining. Discrimination can be viewed as the act of illegally treating people on the basis of their belonging to a specific group. Discrimination can be carried out on the basis of sensitive attributes are male or female, nationality, religion, age, marital status etc. Discrimination does not grant opportunities which are available to the member of one group to member of another group. Discrimination is nothing but counter productive treatment of an individual. There are some systems like granting loan and premium computation where large amount of data is collected and analysed for decision making process in data mining technology. Here system like classification or association rules are used for purpose of decision making. [1]

So now coming the problem of discrimination, different types of new techniques were proposed. Discrimination carry two categories one is direct and second is indirect. Decisions based on sensitive attributes are termed as direct discrimination and the decisions those based on non-sensitive attributes where termed as indirect discrimination which is strongly correlated with biased sensitive once.

2. RELATED WORK

This section discusses the various techniques related to data-mining to our work to determine the associations between them.

D.Predeshi, S. Ruggieri, F. Turini where the first researchers that to address the problem of discrimination in data mining models. This approach is based on identifying the discriminatory rules that are present in the dataset, and the specific set of that data where they exist, other than on learned like a classifier with independency constraints for future predictions.

In Data Mining for Discrimination Discovery by Ruggieri presents the issue of finding discrimination through the data mining in the dataset of recorded choice of records, carried by people or programmed frameworks. They too formalulated the methods of direct and indirect discrimination by displaying the protected by law of groups and connections where the discrimination happens in a classification based extraction. Essentially, classification rules extracted from the dataset permit for divulging connections of unlawful discrimination, where level of the load over the protected by law of groups are formalized by the augmentation of lift measure of a classification rules.

S. Hajian and J. Domingo-Ferrer have addressed the problem of direct and indirect discrimination prevention in data mining. They had developed the rule protection and also generalization methods which are applicable for direct or indirect discrimination prevention individually or

simultaneously. This method uses various methods to clean training data sets and outsourced data sets in a way that direct or indirect discriminatory decision rules are properly converted to the legal classification rules. [4]

In Rule Protection for not direct Discrimination avoidance in information Mining where displayed the first technique for eliminating not the direct discrimination in information mining because of the biased training datasets. The involvement in paper centered on delivering training data which are to about free from oblique discrimination while safeguarding their helpfulness to data mining algorithms. So as to eliminate indirect discrimination in the dataset, here the first step comprises in the discovering data whether they exists indirect discrimination. If any of the discrimination is original, then the dataset is changed until the discrimination is brought beneath to a certain threshold or must be completely removed.

2.1 EXISTING SYSTEM

As the existing system get effected when removing the direct and indirect discrimination in original dataset and also preserves the data quality.Here in existing system data mining algorithm is not required. It is generally based on the classification rules. Discrimination prevention techniques are used in terms of data quality.Drawbacks of the existing system are shown below; it takes more time to handle the decision tree. It cannot handle more data and predict attributes. Mining data's are not trustworthy and system cannot handle more amounts of data.[3]

3. PROPOSED SYSTEM

Sara Hajian and Josep Ferror proposed pre-processing approach for discrimination prevention. It can handle 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. In Proposed system more data can be discriminated. Discrimination is main thing of this process by the way of this process more people can serve without any partiality. Nondiscriminatory constraint is embedded into a decision tree learner by changing its splitting criterion.

4. ALGORITHM USED

Rule protection and rule generalization are used to generate the discriminatory values. The algorithms where produces the support and confidence values for the discrimination measures. The output of the algorithm is the transformed dataset which is discrimination prevented data set. Rule protection algorithm always produces support and the confidence values. Rule generalization is used to find the relationship between the rules instead of the discrimination measures. Both rule protection and rule generalization are effective at removing the discrimination in the original dataset.

5. IMPLEMENTATION DETAILS

The main objective of the proposed system is to improve the efficiency of discrimination prevention while maintaining original data with minimum information loss.

1. Data Analysis:

Data analysis is used for data from the external disk. Dataset contains the real life and synthetic datasets. First we have to check if all the attributes are placed in a correct manner if any nill values are present then that dataset attributes will not

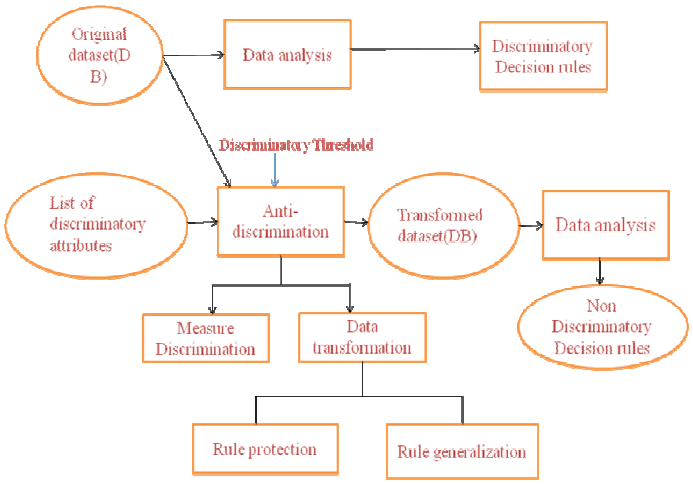

Fig. 1. Flow of Proposed system

be processed by the metric and other computation process. Data analysis is generally termed as the process of gathering and analysis of dataset individually in a given two dataset.

2. Utility Measuring:

Utility measures where taken to erase the discrimination on the given dataset. Dataset are analyzed with certain measures to remove the discrimination from the specified data's. Indirect discrimination removed and measuring of data quality where processed are computed by mathematical functionality like metric and rule protection and rule generalization. With the use of these techniques and algorithm records are filter in short time.

3. Transformation of Source Data:

The main purpose of transforming the original data to clean direct or indirect discrimination from the given data set. To attain this process algorithm should be developed to specify whichrecords should be changed during the transmission process. Discrimination data's are specified and changed during the data transmission.

4. Modified Discriminatory Methods:

In modified discriminatory data's where converted to anonymized data. Anonymized data's does not behave like sensitive attributes but this data's can be processed.This modified data can be done on the sensitive attributes like gender, race, religion, sex, marital status and so on to anonymized the data's. In this administrator can make the data to free from the sensitive attributes.

5. Decision Making:

Decisions can be depend on the attributes like gender, race and religion and so on. Each user gets score for their personal attributes in direct discrimination. Indirect discrimination can be done with the help of anonymized data's. Decisions should be done with the help of algorithm. The resulting data's are free from the discrimination removal and the data quality is maintained.

5. CONCLUSION

In this paper, we have carried out the different approaches for discrimination prevention, and analyses. Different transformations are used for the discovery of discrimination. The process measures the discrimination and identifies the categories by decision-making processes. Discrimination with 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] 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.

[2] Anju sundaresan, lakshmi s, "survey on discrimination prevention Techniques in data mining", ijarse, vol. No.4, special issue (01), march 2015.

[3] Naveena M.S, Robert S , "A Survey on Anti-Discrimination Techniques", IJEIT Volume 4, Issue 3, September 2014.

[4] R.kayalvizhi, s.sarika ," A novel approach to prevent the discrimination in data mining ", ijeat issn: 2249 – 8958, volume-3, issue-3, february 2014.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] Priyanka C. Ghegade, Vinod S. Wadne, "A Survey on Facilitating Document Annotation Techniques", ISSN (Online): 2319-7064, IJSR, Volume 4 Issue 4, April 2015.
[14] 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.
[15] conf.Belgium and The Netherlands,2010.
[16] 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.
[17] Priyanka C. Ghegade, Vinod S. Wadne, "A Survey on Facilitating Document Annotation Techniques", ISSN (Online): 2319-7064, IJSR, Volume 4 Issue 4, April 2015.