ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Technique for Avoiding Discrimination in Data Mining using Antidiscrimination Model

Satyendra B. Kothari, Mansi Bhonsle

PG Scholar, Department of Computer Engineering, G.H. Raisoni C.O.E.M., Wagholi, SPPU, Pune, India

Assistant Professor, Department of Computer Engineering, G.H. Raisoni C.O.E.M., Wagholi, SPPU, Pune, India

ABSTRACT: Insurance premium computation, loan granting/denial, etc are examples which use classification rule mining for taking certain decisions. Classification rule mining techniques for automated data collection have been used for taking automatic decisions. Some attributes are sensitive attributes which causes discriminations, like religion, gender or race. It may lead to discriminatory decisions if the training data set is unfair with respect to sensitive attributes. This paper introduces an antidiscrimination technique which deals with discrimination avoidance by means of discovery and prevention. There are two types of discrimination-indirect and direct. In case of direct discrimination technique decisions are biased based on some sensitive attributes. In indirect discrimination nonsensitive attributes which are related with sensitive attributes may cause decisions get biased. This paper measures and prevents discrimination from data mining. It is useful in preserving data quality.

KEYWORDS: Discrimination detection and avoidance, data mining, rule protection, rule generalization indirect and direct discrimination prevention, antidiscriminations.

I. INTRODUCTION

Data mining is generally used for extracting useful information from large amounts of data such as trends or patterns. To gain insight into methods of suspects or potential suspect's activities various governments are collecting large amounts of data [10]. This can be very useful, but usually at least part of the data set is confidential on which data mining is applied and privacy sensitive. Examples are race, religion, marital status, gender, disability, nationality, and age, etc. This raises the question how privacy [4]. Data mining techniques and automated data collection like classification rule mining used to make automated decisions. If the training data sets are biased, it results in discriminatory decisions. For this antidiscrimination technique is introduced. Automated unfair decisions may result due discriminatory rules extraction from data set used in antidiscrimination. On the other hand if discriminatory attributes list is *DB* can go through process of antidiscrimination so that the rules learned from classification are out of discrimination which results into unbiased automated decisions which get enabled [5]. Discrimination is of two types (indirect and direct). Direct discrimination means having classification rules or procedures that obviously describes minority groups having disadvantage based on discriminatory attributes (sensitive) related to particular member of a group. Indirect discrimination means procedures or rules that are not explicitly showing or includes discriminatory (sensitive) attributes but may generate discriminatory decisions intentionally or unintentionally [12]. Because of some availability of background knowledge indirect discrimination could occur. Publicly available data can be used for accessing or extracting the background knowledge

II. RELATED WORK

Work does not discuss widely and does not ensure quality measures of the different models which are related with antidiscrimination techniques from data mining. Some papers discuss discrimination measurement and discovery while some discusses prevention. In Pedreschi et al. [8], [9] discusses discrimination decision discovery. They formalize legal definition of discrimination. Approach is based on mining association rule. Rakesh Agrawal [1] prepared combinations of two proposed algorithms by combining their best features and proposed new one i.e. AprioriHybrid Quantities of the items considered in a transaction processing are not considered in this work, which are useful for some applications. They did not find such rules.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Toon Calders and Sicco Verwer [2] present three approaches which makes discriminated free Naive Bayes classifier. To handle the problem of classification without any discrimination they use Naive Bayes classifiers. Three approaches are generally used for discrimination prevention in data mining are discussed in literature. First, preprocessing [6] adapted from privacy preservation. Secondly In-Processing technique which cannot use standard data mining algorithms and rely on new special-purpose data mining algorithms. Without changing the algorithms used for data mining which are used as standards, this paper describes discrimination prevention which requires preprocessing, unlike the inprocessing method, and it allows data publishing. Whereas postprocessing model approach discusses only knowledge publishing. The preprocessing approach looks the most convenient as compared to two other techniques.

III. Proposed System

A. *System Architecture:*

In the system, the original dataset i.e. *DB* is first analyzed for detecting discriminatory decision rules (if any). And the detected discriminatory decision rules are stored in some other database. Then the list of discriminatory attributes and original dataset *DB* are given as input to antidiscrimination. With the help of measure discrimination and data transforming the transformed data set (*DB*') i.e. discrimination free data set get generated. Then again that transformed data set will be analyzed for ensuring that there are no more discriminatory decision rules present in it. An antidiscrimination model for avoiding discrimination which includes considerations for both indirect and direct discriminations is discussed in this section of paper. Two phases are used for this approach.

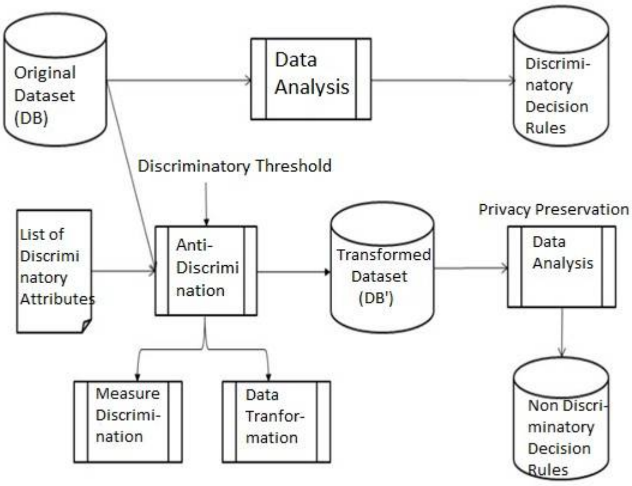

Fig. 1. System Architecture.

B. *Discrimination Measurement:*

To identify Direct and Indirect discrimination our system categorizes classifications into two groups, based on predetermined items in *DB*: PD and PND rules. By identifying vales of α and elift from the PD rules direct discrimination can be measured. By combining identified redlining rules with background knowledge from the PND rules, using elb, and α, indirect discrimination can be measured.

C. *Data Transformation*

Data transformation means process of converting a collection of data values from the data format of a original data in the case of a metadata and a data warehouse, into the destination system's data format. For removing indirect and/or direct discriminatory attributes, which are legitimate, transform the original data *DB*. This techniques which is used for this purpose is described in subsequent sections. For avoiding indirect and direct discrimination may or may not occur at the same time we consider real problem in transforming data which causes minimum information loss measures.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Preprocessing technique out of three processing solutions for indirect and/or direct discrimination avoidance is considered in this paper. Two different cases arise:

First, due to privacy constraints discriminatory item sets have been removed previously from it but, indirect discrimination still get possible because of data available publicly (e.g., census data). In such cases, only rules that are extracted from *DB* are PND so could result in only indirect discrimination. Second, if from the original database (*DB*) at least one sensitive item set cannot be removed, PD rules can be extracted from *DB* and may result in direct discrimination. But, indirect discrimination might occur along with direct discrimination because of knowledge extracted from background from *DB* itself. It happens because of sensitive (discriminatory) attributes which are related with non sensitive items. Thus, it causes indirect and direct discrimination. This is an important aspect for providing indirect rule protection and direct rule protection simultaneously. Moreover, it may not create new nondiscriminatory classification rules or avoid existing rules from removal while transforming data for eliminating direct discriminatory rules. Also while transforming data for eliminating indirect discriminatory rules or redlining rules, it should not produce new sensitive rules.

D. *Mathematical Model*

Let S be a database requested by user, antidiscrimination module which gives discrimination free database such that

$$S= \{I, F, O\}$$

Where, I represent the set of inputs;

$$I= \{I_1, I_2, I_3, I_4, I_5\}$$

I_1= Adult Database
I_2= FR be the database of frequent classification rule extracted from DB.
I_3= MR be the database of direct discriminatory rules.
I_4= α be a threshold value.
I_5= DIs is a collection of sensitive attribute.

And F is the set of functions:

$$F= \{F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9\}$$

F_1 = Login to the system.
F_2 = Request for database.
F_3 = Data extracted by using DPM by Admin.
F_4 = Calculate confidence and support.

$$\text{supp}(x) = \frac{\text{Occurance}(x)}{\text{Total}}$$

$$\text{conf}(x \rightarrow c) = \frac{\text{support}(x, c)}{\text{supp}(x)}$$

F_5 = Calculate elift.

$$\text{elift}(A, B \rightarrow C) = \frac{\text{conf}(A, B \rightarrow C)}{\text{conf}(B \rightarrow C)}$$

F_6 = Identify discriminatory rules.
F_7 =Transform database.
F_8 = Apply DDP algorithm.
F_9 = Get discrimination free database

And O is the set of outputs;

$$O = \{O_1\}$$

O_1=Discrimination free database without affecting original database.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

E. *Cost of computation*

Let,

- f is number of indirect α discriminatory rules in *RR*
- n is number of direct α discriminatory rules in *MR*
- m be the number of records in *DB*.
- k be the number of rules in *FR*
- h the number of records in subset DB_c.
- d is number of iterations.

Then,

- Cost [finding subset DB_c]= *O(m)*.
- Cost [computing impact (db_c) for all records]= *O(hk)*.
- Cost [sorting DB_c by ascending impact]= *O(h log h)*.
- Cost [impact minimization procedure in algorithms] = *O (hk +h log h)*.
- Total Computation time of algorithm is

$O((f + n) * \{m + hk + h \log h + dm\})$

IV. PSEUDO CODE

ALGORITHM: DIRECT AND INDIRECT DISCRIMINATION PREVENTION

Step 1: Accept Inputs: *DB, FR, RR, MR,* α, DI_s
Step 2**:** System Output: *DB'* (transformed data set)
Step3: for each $r : X \rightarrow C \in RR$, where $D,B \subseteq X$ do
Step 4: $\gamma = conf(r)$
Step5: for each r' : $(A \subseteq DI_S), (B \subseteq X) \rightarrow C \in RR$ do
Step 6: $\beta_2 = conf\,(rb_2 : X \rightarrow A)$
Step 7: $\Delta_1 = supp\,(rb_2 : X \rightarrow A)$
Step 8: $\delta = conf\,(B \rightarrow C)$
Step 9: $\Delta_2 = supp\,(B \rightarrow A)$
Step 10: $\beta_1 = \Delta_1 / \Delta_2$
Step 11: Find DB_c: all records in *DB* that completely support $\neg A, B, \neg D \rightarrow \neg C$
Step 12: for each $db_c \in DB_c$ do
{
Compute $impact(db_c) = |\{r_a \in FR | db_c \text{ supports the premise of } r_a\}|$
}
Sort DB_c by ascending impact
Step 13: if $r' \in MR$ then
Step 14: while $\delta \leq \frac{\beta 1\,(\beta 2+\gamma-1)}{\beta 2.\alpha}$ and $\delta \leq \frac{conf(r')}{\alpha}$ do
Step 15: Select first record db_c in DB_c
Step 16: Modify the class item of db_c from $\neg C$ to *C* in *DB*
Step 17: Recompute $\delta = conf(B \rightarrow C)$
Step 18: end while
Step 19: else
Step 20: while $\delta \leq \frac{\beta 1\,(\beta 2+\gamma-1)}{\beta 2.\alpha}$ do
Step 21: Steps 15 to 17
Step 22: end while

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Step 23: end if
Step 24: end for
Step 25: end for
Step 26: for each $r' : (A,B \rightarrow C) \in MR \setminus RR$ do
Step 27: $\delta = conf(B \rightarrow C)$
Step 28: Find DB_c: all records in DB that completely support $\neg A,B \rightarrow \neg C$
Step 29: Step 12
Step 30: while $\delta \leq \frac{conf(r')}{\alpha}$ do
Step 31: Steps 15 to 17
Step 32: end while
Step 33: end for
Step 34: Output: $DB' = DB$

V. EXPERIMENT RESULTS

In our system algorithms and utility measures are implemented using java as front end, NetBeans IDE and back end as MySQL. System implemented on Pentium Processor-IV, 1 GB RAM, 200 MB free HDD space. 80 GB Hard Disk. Adult data set (census income) is used for evaluation purpose. The data set contains 48,842 records which is split into test part and train part records. This dataset has 14 attributes. Prediction task involved is whether a person can make 50K dollar per year. In our proposed model main aspect used for system evaluations is to consider whether a proposed system is able to remove discriminations either indirect or direct or both from the given data set used for evaluation. Firstly data is accepted from adult data set and preprocessing is performed to remove any null value items from dataset. Then we calculate confidence of sensitive, non-sensitive and class attributes. Data is transformed and prevented from discrimination which is shown in *Fig.2*. For achieving results desired from an antidiscrimination model which is expected from running Algorithm 1 different utility measures are shown in *Fig.3*. The results are shown for different values of $\alpha \in [1, 1.9]$ by considering predetermined discriminatory items for data set. Utility Measures are for Confidence 10 Percent and Minimum Support considered 2 Percent for Direct and Indirect Rule Protection.

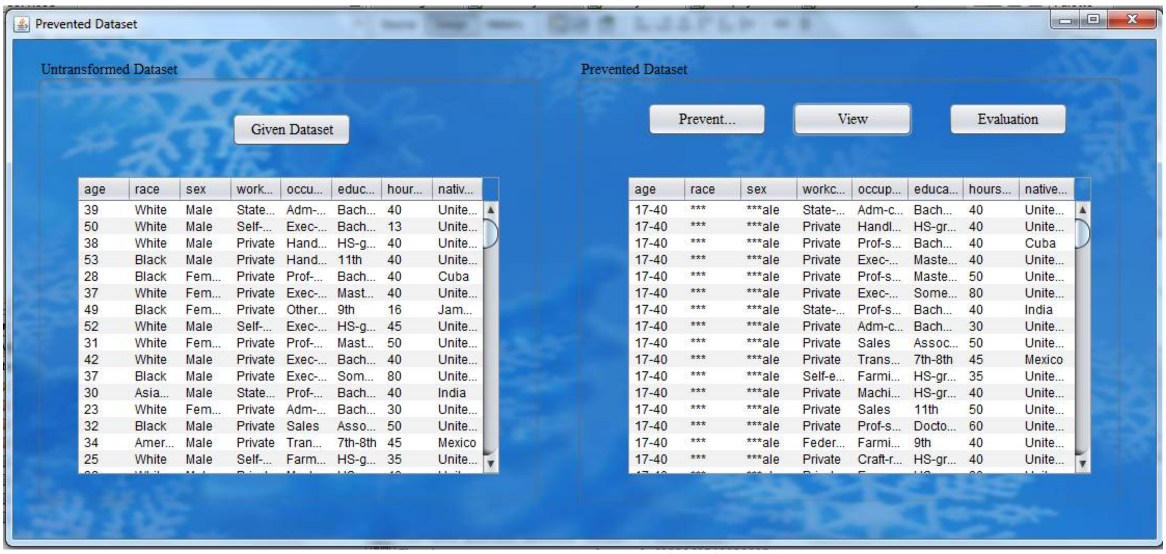

Fig.2. Anti-Discrimination Algorithm

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

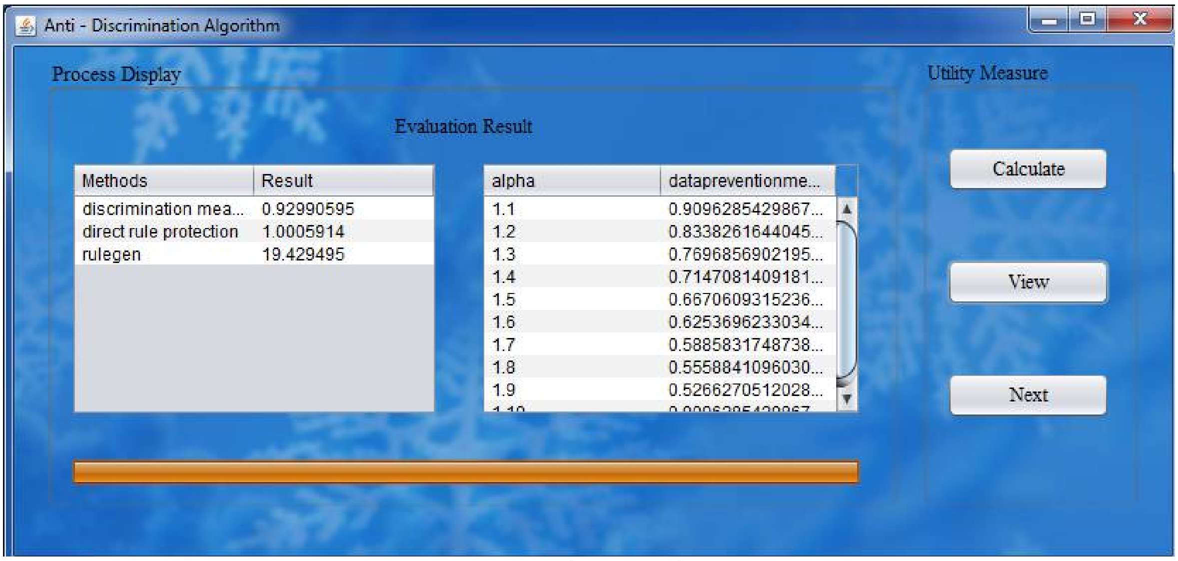

Fig.3. Evaluation Result

VI. CONCLUSION AND FUTURE WORK

While making certain decisions discriminations may occur because of sensitive attributes like sex, gender, religion and so on. We measure discrimination and identify categories and groups of individuals that have been directly or indirectly discriminated in the decision-making processes; secondly, for removing all those discriminatory biases we have to transform data in a proper way. Lastly, without seriously damaging data quality discrimination-free data system models can be devised from the data set after transformation.

By studying different definitions described in literature new data transformation techniques can be designed in future which will cover both legal and cultural conventions of that system.

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. 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.
3. 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.
4. 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.
5. F. Kamiran and T. Calders, “Classification without Discrimination,” Proc. IEEE Second Int’l Conf. Computer, Control and Comm. (IC4 ’09),2009.
6. F. Kamiran and T. Calders, “Classification with no Discrimination by Preferential Sampling,” Proc. 19th Machine Learning Conf. Belgium and The Netherlands,2010.
7. F. Kamiran, T. Calders, and M. Pechenizkiy, “Discrimination Aware Decision Tree Learning,”Proc. IEEE Int’l Conf. Data Mining (ICDM ’10),pp. 869-874, 2010.
8. 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.
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. European Commission, “EU Directive 2006/54/EC on AntiDiscrimination,”http://eurlex.europa.eu/LexUriServ/LexUriServ.do?Uri=OJ:L:2006:204:0023:0 036:en:PDF, 2006.
11. R. Kohavi and B. Becker, “UCI Repository of Machine LearningDatabases,”http://archive.ics.uci.edu/ml/datasets/Adult, 1996.
12. D. Pedreschi, S. Ruggieri, and F. Turini, “Measuring Discrimina-tion in Socially-Sensitive Decision Records,” Proc. Ninth SIAM Data Mining Conf. (SDM ’09),pp. 581-592, 2009.
13. Sara Hajian and Josep Domingo-Ferrer, "A methodology for direct and indirect discrimination prevention in data mining", Fellow, IEEE, 2013

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

14. D. Pedreschi, S. Ruggieri, and F. Turini, “Integrating Induction and Deduction for Finding Evidence of Discrimination,”Proc. 12th ACM Int’l Conf. Artificial Intelligence and Law (ICAIL ’09),pp. 157-166, 2009.
15. S. Ruggieri, D. Pedreschi, and F. Turini, “DCUBE: Discrimination Discovery in Databases,”Proc. ACM Int’l Conf. Management of Data (SIGMOD ’10), pp. 1127-1130, 2010..

BIOGRAPHY

Mr. Satyendra B. Kothari received B.E. in Computer Science and Engineering from Dr. Babasaheb Ambedkar Marathwada University in 2008 and working towards M.E. in Computer Networks from Savitribai Phule Pune University.

Mrs. Mansi Bhonsle persuing Ph. D. from K.L. University (A.P.). She received B.E. in CSE from DYPCSE-Kolhapur and M.Tech in CE from DBATU-Lonere. Her area of interest are graph theory, web mining and algorithm. She is LMISTE Life member of ISTE.