Survey on Direct and Indirect Discrimination Prevention Attribute Method and Evaluation Parameter

Rajni Mishra, Asst. Prof. Sandeepkumar, Prof.Aishwarya Mishra

Abstract— **Data mining is a technology that extracts useful information, such as patterns and trends, from large amounts of data. The privacy sensitive input data and the output data that is often used for selecting deserve protection against abuse. The focus is on preventing that selection rules turn out to discriminate particular groups of people in unethical or illegal ways. This paper give information to protect such attribute which create discrimination directly and allow unfair information to propagate. Various method have been develop for hiding such attribute or perturb so no knowledge will be gather from these.**

Index Terms— ***Data mining, Data Perturbation, Multiparty Privacy Preserving.***

I. INTRODUCTION

The aim of data mining [1, 10, and 11] is to extract useful information, such as patterns and trends, from large amounts of data. In their fight against crime and terrorism, many governments are gathering large amounts of data to gain insight into methods and activities of suspects and potential suspects. This can be very useful, but usually at least part of the data on which data mining is applied is confidential and privacy sensitive. Examples are medical data, financial data, etc. This raises the question how privacy, particularly of those who are innocent, can be ensured when applying data mining. Furthermore, the results of data mining can lead to selection, stigmatisation and confrontation [7]. False positives and false negatives are often unavoidable, resulting in the fact that people are frequently being judged on the basis of characteristics that are correct for them as group members, but not as individuals as such [18]. In the context of public security, false positives may result in investigating innocent people and false negatives may imply criminals remain out of scope.

Main purpose of data mining is to retrieve important information from the dataset inform of patterns of items, it is like an trend that thing get repeat regularly in the dataset. In order to find the pattern which indicate the normal activity of the terrorist, crimals, customers, viruses, etc[7, 11].. This mining is very useful. This kind of information gathering from the raw data is harmul in many areas as well because it lead to the kind of separation from the uncommon part tend to generate the information which may give information in other sense as well. Suppose an intruder need to gather personal information from the dataset like medical, financial, social etc. This lead to new area of how to protect the personal information of the people from the data miners. So in order to release such kind of data which are fruitful for those people who want to get illegal information then it need to make some modification in the dataset. So in order to provide security for the public false negative may imply criminals out of scope.

A priori protection may be realised by protecting input data and access to input data. However, removing key attributes such as name, address and social security number of the data subject is insufficient to guarantee privacy; it is often still possible to uniquely identify particular persons or entities from the data, for instance by combining different attributes. Since the results of data mining are often used for selection, a posteriori protection is also desirable, in order to ensure that the output of data mining is only used within the imposed ethical and legal frameworks. This implies, for instance, that data mining results on terrorism, where data was collected within extensive jurisdiction of secret services, cannot be used just like that for shoplifting or car theft, where data was collected within limited jurisdiction of the police.

So to provide protection from the unfair activities some steps need to taken that modify the data in form of removing important coloum such as name, address, social connectivity, etc. But this might not sufficient because by means of some kind of the relation, pattern it might possible to gather unfair information that is harmful. So to what extent legal and ethical rules can be integrated in data mining algorithms need to find while that may not break any of the security. Now it is required to protect the legal, ethical rules and principles be translated in a format understandable for computers but at the same time it should retain anti-discrimination. For reducing the discrimination of the people on the basis of ethnic background or gender. So by mining these models may include discrimination of people on the basis of different category. It might required to hide such kind of information that can help the discriminator to identify it easily.

II. DISCRIMINATION AND OTHER RISKS

The search for patterns and relations in data by means of data mining may provide overviews of large amounts of data, facilitate the handling and retrieving of information, and help the search for immanent structure in nature. More closely

related to the goals of particular users, group profiles resulting from data mining may enhance efficacy (achieving more of the goal) and efficiency (achieving the goal more easily). Data mining may be useful in many different areas [15]. In the fight against crime and terrorism, analysing large amounts of data may help to gain insight into methods and activities of suspects and potential suspects. In turn, this may help police and justice departments to address these (potential) suspects more adequately for instance, by redirecting funding, people and attention to these particular groups.

As by data mining one can generate frequent pattern from it which generate different results or decision for fight against crime and terrorism, analysing large amounts of data may help to gain insight into methods and activities of suspects and potential suspects. It is very helpful for police and justice departments to identify these suspects more adequately for instance, by redirecting funding, people and attention to these particular activities. Because of lack of the dataset values one might not able to discriminate from the group, this may lead to one bigger problem as well which tends to generate false positives, i.e., people who are part of the group described in the risk profile but who do not share the properties of the group as an individual, and false negatives, people who are not part of the group described in the risk profile even though they constitute the risk the profile tries to describe.

This can be understand by an example where a general trend of the terrorist is a people have following gesture of a large black beard, wearing traditional Islamic clothing, and arriving from other country. When searching the terrorist on such kind of pattern is unethical as there are many religion in the world and may heart the feeling of the innocent one. This generate false positives: they are wrongly selected on the basis of their profile. Now it is not necessary that the terrorist have this pattern only they may be different one depend on the requirement of the activity , situation, place , etc. This may falls in false negatives category to the decision: they are wrongly not selected on the basis of their profile.

When selecting individuals or groups of people on particular characteristics, this may be unwanted or unjustified or both. Selecting for jobs on the basis of gender, ethnic background, etc., is considered unethical and, in many countries, forbidden by law. When risk profiles constructed by companies, governments or researchers become 'public knowledge', this may also lead to stigmatisation of particular groups [8].

III. BACKGROUND

A)Basic Notions

A dataset is a collection of data objects (records) and their attributes. Let DB be the original dataset.

An item is an attribute along with its value, e.g. {Race=black}.

An itemset, i.e. X, is a collection of one or more items, e.g. {Foreign work er=Yes, City=NYC}.

A classification rule is an expression X→ C, where C is a class item (a yes/nodecision), and X is an itemset containing no class item, e.g. {Foreign worker=Yes, City=NYC} --> {hire=no}. X is called the premise of the rule.

Support(s) of an association rule is defined as the percentage/fraction of records that contain X U Y to the total number of records in the database. The count for each item is increased by one every time the item is encountered in different transaction T in database D during the scanning process. It means the support count does not take the quantity of the item into account. For example in a transaction a customer buys three bottles of beers but we only increase the support count number of {beer} by one, in another word if a transaction contains a item then the support count of this item is increased by one. Support(s) is calculated by the following

Support (X➔Y) =(XUY) / D

Confidence: Confidence of an association rule is defined as the percentage/fraction of the number of transactions that contain X U Y to the total number of records that contain X, where if the percentage exceeds the threshold of confidence an interesting association rule X➔Y can be generated.

Confidence (X➔Y) = (XUY) / X

confidence is a measure of strength of the association rules, suppose the confidence of the association rule X➔Y is 80%, it means that 80% of the transactions that contain X also contain Y together, similarly to ensure the interestingness of the rules specified minimum confidence is also pre-defined by users.

These rules are classify into two group potential discriminate rules and non discriminate rules depend on the presence of the rule in the database if X→Y rule is present then it is consider as the potential discriminate rule while other is potential non discriminate rule. These rules are classify into two category depend on the support and confidence values such as those rules that cross the minimum value of support and confidence is consider as the frequent rules while other are consider as the non frequent rules.

There are some attribute in the database that can directly discriminate one from other such as : race, color, religion, nationality, sex, marital status, age and pregnancy which was also come in law of U. S. [1]. Base on these attribute one can evaluate different important informationso it is very necessary to remove, hide, modify these attributes. There are some more attributes that do not discriminate directly they are consider as the non discriminatory attributes so except discriminate attribute other are come into non discriminate ones.

IV. TAXONOMY OF DISCRIMINATION PREVENTION METHODS

As the dataset contain information that is fruitful for many people both for fair and unfair activities so care should be taken to release that dataset into the public place now, there is a taxonomy regarding that which is follow for protecting from such kind of activities below figure represent this that one has to identify the direct discrimination as well as indirect so the challenge increases if we want to prevent not only direct discrimination but also indirect discrimination or both at the same time.

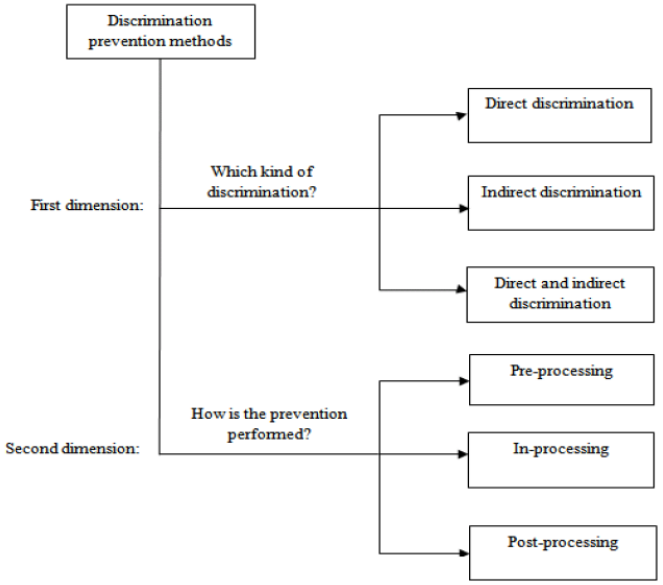

Fig. 1. Taxonomy of discrimination prevention

First Dimension: In this step one has to analyze different attributes into three category direct discriminate, indirect discriminate, and non discriminate attributes. Direct discriminate attributes has been already mention they can categorize the attributes directly and make decision quickly, while in case of indirect attributes which are generate by the association rule then the frequent one are consider as the indirect discriminate rules. Those rule and attribute which are left are consider as the non discriminate rule or attributes. There are many rule which use direct attributes while they are generate from the association these rules are consider as the direct indirect both.

Second Dimension : This consist of the three parts pre-processing, post-processing, in-processing. Consider each thing one by one. Although manyalgorithm have been already developedfor all the above -processing, post-processing, in-processing. Many researchers have given details, algorithms and experimental results on these methods are presented in [4,5]. The aim of all these methods is to transform the original data so that it will make minimum impact on the data and on legitimate decision rules, and the main purpose the work will be fulfill of making unfair decision rule that can be mined from the transformed data. The measure of the algot=rithm are depend on the, the metrics that specify which records should be changed, how many records should be changed and how those records should be changed during data transformation are developed so that it will make mininmumimapact on the original data. Few works are done base on the assumptions such as the class attribute in the original dataset is binary other is the database of discriminatory and redlining rules as output of a discrimination measurement phase based on measures proposed in [1,2].

In case of Pre-processing there are methods that can identify those rules or attributes in the database that is obtained from the source data then remove, modify those discriminatory rules or attributes biases contained in the original data so that no unfair decision rule can be mined from the transformed dataset by using any of the data mining algorithms. The pre-processing approaches of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature [5,11].

In case of the In-processing there are many approaches that change the data mining algorithms in such a way that the obtaining models is free from unfair decision rules [10]. For example, an alternative approach to cleaning the discrimination from the original dataset is proposed in [10] whereby the non-discriminatory constraint is embedded into a decision tree learner by changing its splitting criterion and pruning strategy through a novel leaf re-labeling approach. Although it is found that in-processing discrimination prevention algorithms are depends on the special purpose data mining approaches as standard data mining algorithms cannot be used because they ought to be adapted to satisfy the non-discrimination requirement.

V. EVALUATION PARAMETERS

There are two approaches to evaluate the discriminating algorithm developed which can specify the quality of the work first is Discrimination Removal while second is data quality after the implementation of the algorithm. Normally balancing both is quit difficult as if data quality need to maintain then some of the rules will be unaffected and over all purpose will be not be solve while in case of maintaining discriminating rule less data [11], dataset the quality will definite degrade as it need to either change or remove from the dataset.

i)Direct Discrimination Prevention Degree (DDPD). This measure quantifies the percentage of discriminatory rules that are no longer discriminatory in the transformed dataset.

ii)Direct Discrimination Protection Preservation (DDPP). This measure quantifies the percentage of the protective rules in the original dataset that remain protective in the transformed dataset.

*iii) Indirect Discrimination Prevention Degree (IDPD).*This measure quantifies the percentage of redlining rules that are no longer redlining in the transformed dataset.

*iv) Indirect Discrimination Protection Preservation (IDPP).*This measure quantifies the percentage of non-redlining rules in the original dataset that remain non-redlining in the transformed dataset.

Since the above measures are used to evaluate the success of the proposed methods in direct and indirect discrimination prevention, ideally their value should be 100%.

A) Measuring Data Quality

The second aspect to evaluate discrimination prevention methods is how much information loss (i.e. data quality loss) they cause. To measure data quality, two metrics are proposed in Verykios and Gkoulalas-Divanis (2008):

i) Misses Cost (MC). This measure quantifies the percentage of rules among those extractable from the original dataset that cannot be extracted from the transformed dataset (side-effect of the transformation process).

ii) Ghost Cost (GC). This measure quantifies the percentage of the rules among those extractable from the transformed dataset that were not extractable from the original dataset (side-effect of the transformation process). MC and GC should ideally be 0%. However, MC and GC may not be 0% as a side-effect of the transformation process.

CONCLUSION

Due to the right to privacy in the information ear, privacy-preserving data mining (PPDM) has become one of the newest trends in privacy and security and data mining research. Inthis paper, this work introduced the related concepts of privacy-preserving data.

REFERENCES

[1] Pedreschi, D., Ruggieri, S. &Turini, F. (2008). Discrimination-aware data mining. Proc. of the14th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2008),pp. 560-568. ACM.

[2] Pedreschi, D., Ruggieri, S. &Turini, F. (2009a). Measuring discrimination in socially-sensitivedecision records. Proc. of the 9th SIAM Data Mining Conference (SDM 2009), pp. 581-592.SIAM

[3] Ruggieri, S., Pedreschi, D. &Turini, F. (2010). Data mining for discrimination discovery. ACMTransactions on Knowledge Discovery from Data, 4(2) Article 9.

[4] Hajian, S., Domingo-Ferrer, J. & Martinez-Ballesté, A. (2011a). Discrimination prevention indata mining for intrusion and crime detection. Proc. of the IEEE Symposium on ComputationalIntelligence in Cyber Security (CICS 2011), pp. 47-54. IEEE.

[5] Hajian, S. & Domingo-Ferrer, J. (2012). A methodology for direct and indirect discriminationprevention in data mining.Manuscript.

[6] Yin, X. & Han, J. (2003). CPAR: Classification based on Predictive Association Rules. In Proc.of SIAM ICDM 2003. SIAM.

[7] Verykios, V. &Gkoulalas-Divanis, A. (2008). A survey of association rule hiding methods forprivacy. In C. C. Aggarwal and P. S. Yu (Eds.), Privacy-Preserving Data Mining: Models andAlgorithms. Springer.

[8] Meij, J. (2002) *Dealing with the data flood; mining data, text and multimed*ia, The Hague: STT Netherlands Study Centre for Technology Trends.

[9] Pedreschi, D., Ruggieri, R., and Turini, F. (2008) *Discrimination-aware Data Mining*. In Proceedings ofthe 14th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

[10] Calders, T., &Verwer, S. (2010). Three naive Bayesapproaches for discrimination-free classification. Data Mining and Knowledge Discovery, 21(2):277-292.

[11] Sara Hajian and Josep 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