ONTOLOGY BASED DATA MINING METHODOLOGY FOR DISCRIMINATION PREVENTION

Nandana Nagabhushana[1] and Dr.Natarajan S[2]

[1]M.Tech in Software Engineering,
Department of Information Science, PESIT, Bangalore, India
nandana.sgn@gmail.com
[2]Professor and Key Resource person,
Department of Information Science, PESIT, Bangalore, India
natarajan@pes.edu

ABSTRACT

Data Mining is being increasingly used in the field of automation of decision making processes, which involve extraction and discovery of information hidden in large volumes of collected data. Nonetheless, there are negative perceptions like privacy invasion and potential discrimination which contribute as hindrances to the use of data mining methodologies in software systems employing automated decision making. Loan granting, Employment, Insurance Premium calculation, Admissions in Educational Institutions etc., can make use of data mining to effectively prevent human biases pertaining to certain attributes like gender, nationality, race etc. in critical decision making. The proposed methodology prevents discriminatory rules ensuing due to the presence of certain information regarding sensitive discriminatory attributes in the data itself. Two aspects of novelty in the proposal are, first, the rule mining technique based on ontologies and the second, concerning generalization and transformation of the mined rules that are quantized as discriminatory, into non-discriminatory ones.

KEYWORDS

Ontology, Discrimination Prevention, Rule Protection, Rule Generalization, Postmining

1. INTRODUCTION

The unjust or prejudicial treatment of different categories of people, especially on the grounds of race, age, or gender is coined as Discrimination. It is the recognition and understanding of the difference between one quality and another, which might pave way for inequity and bigotry towards some particular classes of society in provision of certain services, which otherwise should be made obtainable to all the classes of the society. The Anti-Discrimination Acts proposed and institutionalized as a part of Law of the land by various nations, consist several clauses designed to prevent discrimination in numerous fronts like access to public services, loans, insurance, education, employment etc. based on attributes related to Gender, Nationality, Race, Religion, Marital Status, Physical Disability etc. Technology, particularly data mining can contribute to a fair extent in this arena, to discover and prevent discrimination by automating the routines used in many systems for decision making. Collections of data can be used to train association/classification rules to make decisions that are not influenced by the human decision maker who can be probably biased.

David C. Wyld et al. (Eds) : COSIT, DMIN, SIGL, CYBI, NMCT, AIAPP - 2014
pp. 35–48, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4904

Nevertheless, this is not sufficient to abrogate the plausibility of discrimination. A thoughtful contemplation about the data mining process reveals that rules indeed are mined and learnt from a training data-set, which can be inherently biased. This will lead to discovery of rules which are naturally prejudiced, and possibly discriminatory, thereupon necessitating the extermination of potential biases from the training dataset, thus preventing data mining itself being an agent for discrimination.

A novel solution to the problem has been suggested by Sara Hajian et al. [1] for all discovered types of discrimination. Discrimination can be classified as Direct and Indirect/Systematic [2], based on the nature of discrimination implication. Direct Discrimination can be defined as the process of differentiating based on evidently discriminatory attributes related to a disadvantaged group possessing sensitive discriminatory attributes. For example, denial of admission to an educational institution, based on the candidate's ethnicity, can be termed as Direct Discrimination. Indirect Discrimination is differentiation based on certain attributes of the individual that apparently are not discriminatory, but are highly correlated to discriminatory attributes. To exemplify, denial of admission to an educational institution based on the zip-code of the candidate due to the background knowledge that the dwelling of the candidate is mostly occupied by a particular ethnic group.

Drifting towards the technical aspects of the proposal, it is approving to mention that Association Rule Mining forms the backbone of Knowledge Discovery Process. But it is conspicuous that though rule mining aims at discovering implicative tendencies in the collected data, which can be valuable in decision making. It yields to rules whose usefulness is greatly influenced and limited due to their large numbers. Thus, an effort is required to be made to moderate the number of rules learnt from the training dataset.

Based on the literature in [3], a methodology is proposed to conceptualise the background knowledge possessed by the user, in the form of ontologies. Ontologies are constructed and used to formulate and mine rules into the rule schema, which are then subjected to certain transformations. As the last step, an attempt is made to quantize the discrimination present in the final set of rules, and these rules are validated against certain metrics. The rules which pass the threshold test are marked and allowed as non-discriminatory rules, which are collated as the final rule set.

1.1 Related Work

Data Mining has been extensively employed in numerous applications of various domains which inculcate decision making processes. T. Delenius [4] was the harbinger, who in 1970s, first studied and formulated the statistical disclosure control problem. Research has been carried on ever since then, and in 1990s k-anonymity model was proposed by P. Samarati and L. Sweeney [5]. In this approach, a data set is k-anonymous if its records are not distinguishable by an intruder within groups of k members. The novelty of this model was that the anonymity target was established ex-ante and then computational procedures were used to achieve that target.

Decision Models are mostly constructed by machine learning that happens on historical decision records, using data mining methods. Nevertheless, there is no recognizance that automation of decision making completely rules out the chances of production of discriminatory rules, because the extracted knowledge might contain implicit discriminatory bias. An upright approach to prevent this, is to avoid the classifier's prediction to be based on discriminatory attributes by removing them. But, research by F. Kamiran and T. Calders [6] has proved that this is not an effective and efficient method for discrimination prevention. The attributes which highly correlated to the discriminatory attribute can still exist, whose removal might cause information loss, leading to sub-optimal predictors as depicted in [6, 7].

In this accord, researchers have formalised many strategies among which three are popularised, and practised. The first approach is based on pre-processing, in which the data set is transformed so that the discriminatory rules do not ensue from mining. Kamiran and T. Calders[8, 9] have adopted hierarchy based generalization, to perform controlled distortion and learn the classifier by minimally intrusive modifications to the data sets. This results in an unbiased data set which can then be used to learn rules that are non-discriminatory. This approach proves to be useful in scenarios where the data sets should be published. The in-processing strategy states certain modifications on the data mining algorithms. A novel in-processing method proposed in [10], states that the non-discriminatory criteria are considered as the splitting criteria of a decision tree learner and relabeling is used for pruning. The third strategy, being the post-processing approach, proposes to modify the resulting data-mining model. That is, the rules that are mined as a result of learning the dataset are transformed to remove discrimination. D. Pedreschi, S. Ruggieri, and F. Turini [2, 11] propose a confidence altering approach on the CPAR algorithm. A more recent methodology by Sara Hajian et al. [1], proposes a unified approach to direct and indirect discrimination and also states utility measures to quantify the discrimination. Data transformation methods like rule-generalization and rule-protection are formulated.

In [12], the authors describe and adopt a discrimination discovery method, that not only addresses direct discriminatory attributes, but also those correlated indirect discriminatory attributes. The correlation information is implied as background knowledge, which takes the form of a set of association rules. The challenge of representing the user knowledge has been addressed in a novel way by Claudia Marinica and Fabrice Guillet [3]. It has been proposed that, ontologies can be formalized using specification languages, which can be understood by machines, and parsed in software programs. As a base to this proposal, Liu et al. [3] has proposed a specification language, which can be used to formalize ontologies. In [14], T.R. Gruber defines ontology as a formal, explicit specification of a shared conceptualization. It can be presumed that ontology describes an abstract model of some phenomenon by its important concepts. Also, the formal notion denotes that the formulation and representation of ontology is such that, it is machine interpretable. H. Nigro et al. [15] have classified ontologies into two qualitative categories - Domain and Background Knowledge Ontologies, and, Ontologies for Data Mining Process or Metadata Ontologies.

1.2 Contribution and Plan of this paper

Despite the fact that there have been many propositions of discrimination prevention methodologies, this avenue provides a greater scope for exploration. In this direction, this paper makes an effort to propose a data mining methodology for discrimination prevention using ontologies. This is believed to help in construction of background knowledge by design and offer native technological safeguard against discrimination. This is an attempt to go beyond discrimination discovery and prevention, and cope to the more challenging goal of preventing discrimination in the early stages of KDD process.

This paper is structured as the following: Section 2 introduces notations and definitions used throughout the paper. Section 3 presents the proposed framework and its elements. Section 4 is devoted to the results obtained during experimentation. Finally, Section 5 presents conclusions and shows directions for future research.

2. NOTATIONS AND DEFINITIONS

Let I = {i_1, . . . , i_n} be a set of items, where each item ij has the form attribute=value (e.g., Sex=female).

An item set X ∈ I is a collection of one or more items, e.g. {Sex=female, Credit history=not-taken}.

A ***database*** is a collection of data objects (records) and their attributes; more formally, a (transaction) database D = {r_1,...,r_m} is a set of data records or transactions where each r_i ⊂ I. Alternately, the database D can also be defined as a set of transactions D = {t_1,...,t_m}. Civil rights laws [6, 22] explicitly identify the groups to be protected against discrimination, such as minorities and disadvantaged people, e.g., women.

In the project context, these groups can be represented as items, e.g., Sex=female, which we call Potentially Discriminatory (PD) items; The discrimination is evident with respect to such attributes. A collection of PD items can be represented as an itemset, e.g., {Sex=female, Foreign worker=yes}, which we call PD itemset or protected by-law groups, denoted by DI_s.

An ***itemset*** X is Potentially Non-Discriminatory (PND) if X ∩ DI_s = Ø , e.g. {credit history=no-taken} is a PND itemset where DIs : {Sex=female, Race=black, Foreign worker=yes}.

A ***decision attribute*** is an attribute which takes as values “yes” or “no” to report the outcome of a decision made on an individual. An example for this type of attribute is “credit approved”, which can be yes or no. A class item is an item of class attribute, e.g., Credit approved=no.
The ***support*** of an itemset X in a database D is the number of records that contain X. That is, suppD(X) = | r_i ∈D|X ⊆r_i } |, where | . | is the cardinality operator.

An ***Association Rule*** is an implication X → Y, where X and Y are itemsets and X ∩ Y = Ø. The former is the antecedent and the latter is the consequent of the rule. X → Y is a classification rule if Y is a class item and X is an itemset containing no class item e.g. {Sex=female, Credit history=not-taken → Credit approved=no}. The itemset X is called the premise of the rule.

The rule X → Y is ***completely supported*** by a record if both X and Y appears in the record. Henceforth, due to generalization of the measures to the context of the considered database, this context suffix in discarded and generalized measures and rules are used.
The ***confidence*** of a classification rule, conf(X → Y), is the measure of frequency of the class item Y in records that contain X. Hence, if supp(X) > 0 then,

$$\boldsymbol{conf} = \frac{\boldsymbol{supp(X,Y)}}{\boldsymbol{supp(X)}} \text{...} (1)$$

The value of confidence ranges over [0, 1]

The ***lift*** of a classification rule $lift_D$(X → Y), is the measure of importance of the rule. The lift value of an association rule is the ratio of the confidence of the rule and the expected confidence of the rule.

$$\mathbf{lift_D} = \frac{\boldsymbol{conf(X,Y)}}{\boldsymbol{expected\_conf(X,Y)}} \text{...} (2)$$

The ***expected confidence*** of a rule is defined as the product of the support values of the rule antecedent and the rule consequent divided by the support of the rule antecedent.

$$\mathbf{expected\_conf_D(X,Y) = (supp_D(X) * supp_D(Y)) / supp_D(X)} \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

A ***frequent classification rule*** is a classification rule with support and confidence greater than respective specified lower bounds.

A ***negated itemset***, i.e. $\neg X$ is an itemset with the same attributes as X, but the attributes in $\neg X$ take any value except those taken by attributes in X. For a binary attribute, e.g. {Foreign worker=Yes/No}, if X is {Foreign worker=Yes}, then $\neg X$ is {Foreign worker=No}. For a non-binary categorical attribute, e.g. {Race=Black/White/Indian}, if X is {Race=Black}, then $\neg X$ is {Race=White} or {Race=Indian}. In the current context, only non-ambiguous negations are used.

A ***closed itemset*** [16] is defined as an itemset X which has the property of being the same as its closure, i.e., $X = c_{it}(X)$. The minimal closed itemset containing an itemset Y is obtained by applying the closure operator cit to Y. Let R_1 and R_2 be two association rules. We say that rule R_1 is more general [3] than rule R_2, denoted$R_1 \leq R_2$, if R_2 can be generated by adding additional items to either the antecedent or consequent of R_1. In this case, we say that a rule R_j is redundant [17] if there exists some rule R_i such that $R_i \leq R_j$.

Formally, an ***Ontology*** [18] is a quintuple O = {C, I, R, H, A}. C = {C1, C2,..., Cn} is a set of concepts and R = {R1, R2,…,Rm} is a set of relations defined over concepts. I is a set of instances of concepts and H is a Directed Acyclic Graph (DAG) defined by the subsumption relation (is-a relation, ≤) between concepts. We say that C2 is-a C1, C1 ≤ C2, if the concept C1 subsumes the concept C2. A is a set of axioms bringing additional constraints on the ontology.

3. DISCRIMINATION PREVENTION USING ONTOLOGIES : APPROACH

The dataset used in the case study is titled “Adult Data set” which was extracted by Barry Becker from the 1994 Census database. It comprises of 48842 instances of 14 attributes of type either categorical or integer. Some of the important attributes are age, education, race, sex, and native-country.

The proposed approach can be paraphrased in four phases namely –

1. Ontology construction and rule mining

2. Discrimination measurement

3. Data Transformation

The description of each of these phases follows in the sections 3.1 through 3.3 respectively.

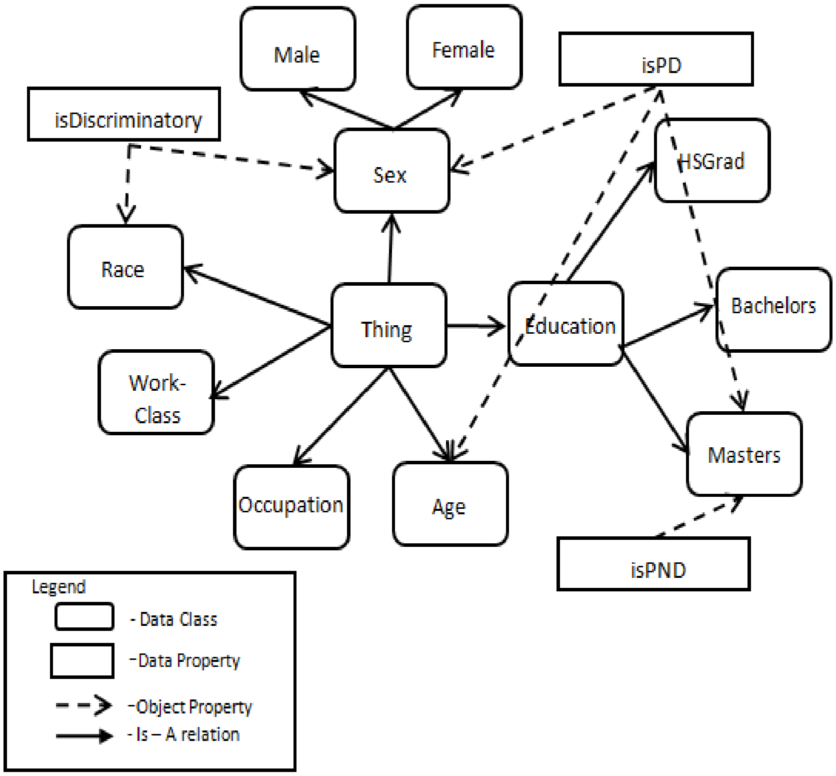

Figure 1: Graphical Representation of the Ontology for Adult Data Set Attributes

3.1 Construction of ontology based on the background knowledge and Association Rule Mining

Background knowledge represents the backbone of association/classification rule mining systems. It is proposed here that ontologies can contribute to a major extent in representing this knowledge. Generally ontologies represent subsumption relations (is-a). The proposal here is to represent background knowledge in the ontology in terms of relationship classes by defining data properties pertaining to discrimination prevention. That is, to represent the knowledge of PD, PND attributes and the subsumption attributes in the ontology. In this accord, four data properties –

- isDiscriminatory,
- isNonDiscriminatory
- isPotentiallyDiscriminatory
- isPotentiallyNonDiscriminatory

are defined in the ontology. The illustration is shown in the Figure 1, which is the representation of ontology construction for the attributes of Adult Data Set.

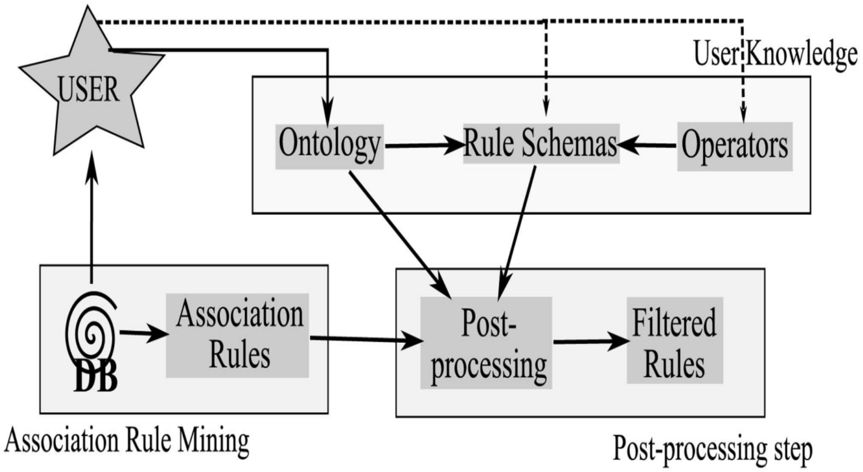

Figure 2: The ARIPSO framework [3]

The ARIPSO (Association Rule Interactive post-Processing using Schemas and Ontologies) framework [3] is used to learn the association rules. Shown in Figure 2 is the ARIPSO framework. One iteration of this process is used, against the suggestion of multiple iterations of user feedback. This is due to the assumption that most of the user knowledge is represented in the data properties of the ontology at one shot. Concepts like, interestingness measures, ontology based rule mining and filtering- results in comparatively less number of rules – rules that are interesting and relevant to the context. The ARIPSO framework chooses to employ FP-Growth algorithm to mine frequent itemsets and hence a set of association rules pertaining to the dataset.

3.2 Discrimination measurement

Although discrimination is discovered in terms of background knowledge during the rule learning phase, a reiteration of this activity is necessary to further classify and fine tune the discovered rules. The utility measures described by Pedreschi et al.[2, 20] over classification rules, for measuring the degree of discrimination of a PD rule (i.e. elift) for direct discriminatory discovery and a PND rule (i.e. elb) for indirect discrimination can be well utilized to quantize the amount of discrimination in each of the generated rule. Filtering of rules should be done based on the threshold values of these measures which further reduces the number of rules. A brief formal description of the terminology and the utility measures follows –

Let, DIs be the set of predetermined discriminatory items in DB (e.g. DI_s = {Foreign worker=Yes, Race=Black, Gender=Female}). Frequent classification rules fall into one of the following two classes:

1) A classification rule X → C is potentially discriminatory (PD) when X = A, B with A $\subset DI_{S,}$ a non-empty discriminatory itemset and B a non-discriminatory itemset. For example {Foreign worker=Yes; City=NYC} → Hire=No.
2) A classification rule X → C is potentially non-discriminatory (PND) when X = {D, B} is a non-discriminatory itemset. For example {Zip=10451, City=NYC} → Hire=No, or {Experience=Low; City=NYC} → Hire=No.

elift is a measure that can be used to assess whether the PD rule is potentially directly discriminatory. Based on a fixed threshold of this measure, a PD rule is judged to be either discriminatory or protective. Formal definition of elift is –

If A,B→ C is a classification rule such that conf(B→ C) > 0, extended lift of the rule is

$$elift(A,B \rightarrow C) = \frac{conf(A,B \rightarrow C)}{conf(B \rightarrow C)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

where, $A \subset DI_S$ and $B \cap DI_S = \emptyset$

Theoretically, elift is the evaluation of discrimination of a rule as a gain of confidence due to the presence of discriminatory items in the antecedent of the rule. If $\alpha \in R$ is a fixed threshold stating an acceptable level of discrimination, and if $A \subset DI_s$, and $B \cap DI_s = \emptyset$, then a PD classification rule R1: A, B→ C is α-protective w.r.t. elift if, elift(R1) < α, otherwise it is α-discriminatory.

The PND counterpart of elift is ***elb*** which is used to assess the quantization of discrimination in PND rules. Based on this measure, PND rules can be classified as either redlining or non-redlining (legitimate) rules. To determine the redlining rules, the value of elb is formally arrived at, by the following theorem which provides a lower bound for α-discrimination, using the information available in PND rules which are (γ, δ) and the information (β1 ,β2) available from background rules. The assumption is that the background knowledge takes the form of association rules relating a PND itemset D to a PD itemset A within the context B.

Let r: D, B→ C be a PND classification rule. Let γ = conf(r: D, B→ C) and δ = conf(r:B→ C) > 0 . Let A be a PD itemset, and let β1, β2 be such that,

$$conf\ (r_{b1} : A, B \rightarrow D) \geq \beta 1$$
$$conf\ (r_{b2} : D, B \rightarrow A) \geq \beta 2 > 0$$

Then,

$$f(x) = \frac{\beta 1}{\beta 2}(\beta 2 + x - 1)$$

$$\mathbf{elb(x,y)} = \{\frac{f(x)}{y}\ iff(x) > 0; 0\ \text{otherwise} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

It holds that, for α ≥ 0, if elb (γ, δ) ≥ α, the PD classification rule R1: A, B→C is α-discriminatory.

A PND classification rule r: D, B→ C is a redlining rule if it could yield an α - discriminatory rule r′: A, B→C in combination with currently available background knowledge rules of the form r_{b1}: A, B→D and r_{b2}: D, B→A, where A is a discriminatory item set. For example, {Zip = 10451; City = NYC} → Hire = No. Otherwise, it is a non-redlining or legitimate rule. For example, {Experience = Low; City = NYC} → Hire = No.

3.3 Data Transformation

Sara Hajian et al. [1] have proposed two data transformation methods – Rule Protection and Rule Generalization which when applied, transforms the data, with minimum information loss. The α – discriminatory rules are transformed to α – protective for direct discrimination, and to an instance of non-redlining PND rule in the case of indirect discrimination.

3.3.1 Rule protection

Rule protection for direct discrimination is termed as Direct Rule Protection (DRP) and is based on the direct discriminatory measure elift. This method simply states that the α – discriminatory rule after transformation, should exhibit an elift less than the value of α. That is if r′: A, B→ C is the transformed counterpart of the rule r, then

$$\mathbf{elift\ (r') < \alpha} \text{ ....}(6)$$

From equation 4, we can deduce that

$$\frac{conf(r':\mathrm{A,B} \rightarrow \mathrm{C})}{conf(\mathrm{B} \rightarrow \mathrm{C})} < \alpha \text{ ....}(7)$$

Thus, by inferring from equation 1, we can achieve the inequality by performing the transformation as stated in Table 1, for the measure elift. This method is a modified version of confidence altering approach stated in [11]. The DRP data transformation attempts to alter the confidence of the base rule B→ C.

On the same lines of DRP, but with the discriminatory measure elb, for indirect discrimination, the transformations are as stated for elb measure in Table 1. The inequality to be established for each redlining rule r: D, B→ C is

$$\boldsymbol{elb\ (\gamma, \delta) < \alpha} \text{..}(8)$$

The inference for this transformation is from equation 5. These transformations are elaborated and proved in [1].

Table 1: Rule Protection

Measure	Transformation	Condition
elift	⌐A, B→ ⌐C » A, B→ ⌐C	elift(A, B→ C)<α
	⌐A, B→ ⌐C » ⌐A, B→ ⌐C	
elb	⌐A, B, ⌐D → ⌐C » A, B, ⌐D → ⌐C	elb (γ, δ) < α
	⌐A, B, ⌐D → ⌐C» ⌐A, B, ⌐D →C	

3.3.2 Rule generalization

After performing rule protection, there might still exist some discriminatory content in the rule repository. Unlike the strategies suggested by Pedreschi et al.[12] and by Sara Hajian et al. [1], a simpler method of generalization which makes use of k-anonymity principle proposed and extended by P. Samarati and L. Sweeney[5, 20, 21] is employed. The recourse from the basic k-anonymity theory is, only those α – discriminatory rules that are not subjected to and remain after rule protection, are generalized by anonymization of the PD attribute to the level in the class hierarchy until the rule becomes α – protective or non-redlining. The graph denoted by Figure 3 is an example of the data classification hierarchy for an attribute "Race" in the Adult Data Set. Likewise, if any rule R1: {Race = Australian-White, Age = Young} is generalised to one level higher in the class hierarchy and measured for its discrimination,

Table 2: Adult Data Set Hierarchies

Attribute	No. of Distinct Values	Levels of Hierarchy
Education	16	5
Marital status	7	4
Native country	40	5
Occupation	14	3

Race	5	3
Relationship	6	3
Sex	2	2
Work-class	8	5

and "Hire = No" proves to be α – discriminatory, then it is transformed to R1′ : {Race = White, Age = Young} → Hire = No . If this rule exhibits high values of elift or elb, generalization is reiterated and the rule becomes R1″ : {Race = Any, Age = Young} → Hire = No. Since information loss is inherent with data transformations, the effect of data transformation on data quality should be quantified and measured. Two metrics have been proposed in the literature as information loss measures in the context of rule hiding for privacy-preserving data mining (PPDM) [19] namely Misses Cost and Ghost Cost can be effectively used for this purpose. ***Misses cost*** (MC) quantifies the percentage of rules among those extractable from the original data set that cannot be extracted from the transformed data set. ***Ghost cost*** (GC) is the measure that quantifies the percentage of the rules among those extractable from the transformed data set that were not extractable from the original data set. Generalizarion is performed on all the attributes listed in Table 2. The hierarchy for each of the attributes is obtained from [22]. Additionally, four utility measures [1] have been adopted to measure the discrimination removal. They are –

1. ***Direct Discrimination Prevention Degree*** (DDPD) – Quantifies the percentage of α – discriminatory rules that are transformed into α –protective rules, after the transformations

2. ***Direct Discrimination Protection Preservation*** (DDPP) – Quantifies the percentage of α–protective rules that remain α-protective, after the transformations

3. ***Indirect Discrimination Prevention Degree*** (IDPD) – Quantifies the percentage of redlining rules that transformed to non-redlining, after the transformations

4. ***Indirect Discrimination Protection Preservation*** (IDPP) – Quantifies the percentage of non-redlining rules that remain non-redlining, after the transformations

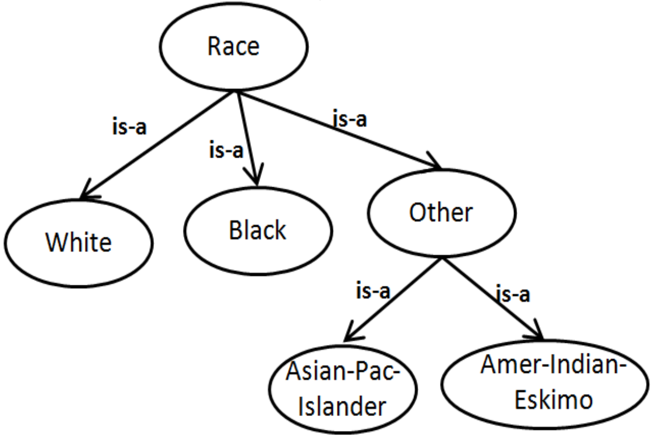

Figure 3: Class Hierarchy for attribute "Race"

4. RESULTS AND ANALYSIS

This section presents the experimental evaluation for the proposed discrimination prevention system using ontologies. The algorithms were implemented using Java programming language. The ontology and hierarchy graphs have been created using protégé 4.3.0 tool, which is a collaborative development effort between Stanford University and University of Manchester. The tests were performed on a 2 GHz Intel Core i7 machine, equipped with 4 GB of RAM, and running under 64 bit Windows 8 Operating System.

The proposed method for Discrimination Prevention was implemented and evaluated in terms of utility measures. Table 3 shows the results of direct and indirect discrimination prevention for 5% confidence and 10% support at three different levels of α. Since the number of rules generated is considerably low in the case of ARIPSO framework, as depicted by Figure 4, the computational cost proportionally decreases. Table 4 shows the comparison between the direct and indirect discrimination prevention method [1] here after referred as method-1 and the proposed method here after referred as method-2, which happens to be a modified evolution of method-1. These results are based on DI_s = {Foreign worker=Yes, Race=Black, Gender=Female} for rule protection, and all the attributes listed in Table 2 for rule generalization. In these tables "n.a." implies that the respective metrics are not applicable for that method.

Table 3: Utility Measures at Support = 5% and Confidence = 10% on Adult Data Set

α	No. of Rules	No. of α-discriminatory rules	No. of redlining rules	DDPD	DDPP	IDPD	IDPP	MC	GC
α =1	238	38	23	92.2	n.a.	88.3	n.a.	9.4	n.a.
α =1.5	193	29	17	94.5	n.a.	91.1	n.a.	22	n.a.
α =2	167	22	9	95.1	n.a.	92.8	n.a.	27.3	n.a.

Table 4: Utility Measures at Support=5%, Confidence=10% and α =2 on Adult Data Set

Method	No. of Rules	No. of α-discriminatory rules	No. of redlining rules	DDPD	DDPP	IDPD	IDPP	MC	GC
Method-1	204	31	15	93.47	100	~93	100	15.24	4.7
Method-2	167	22	9	95.1	n.a.	92.8	n.a.	27.3	n.a.

The comparison of both the methods against all the considered utility measures is summarized in table 4.

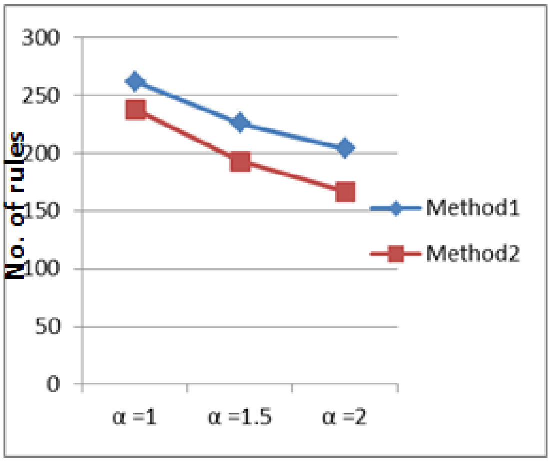

Figure 4: Comparison – No. of rules

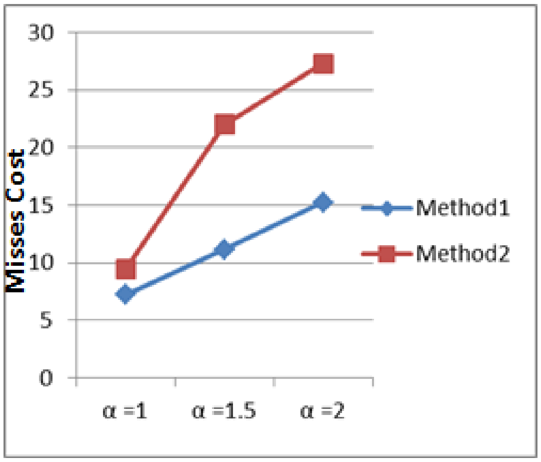

Figure 5: Comparison - MC

Figure 5 shows a comparison of Misses cost for method-1 and mehod-2. The Misses Cost is high in the case of method-2. This can be justified and is acceptable due to the interestingness measure that is considered in the ARIPSO framework, during filtering. The discrimination removal effectiveness of both the methods is nearly identical. This in effect proves that, usage of ontologies in data-mining in general and discrimination prevention in particular is a constructive move, which enhances not only performance, but also the relevance of the mined rules to the context. Similar is Figure 6 which shows the comparison of α- discriminatory rules (direct and indirect) generated out of the two methods. Figure 7 denotes the number of Red-Lining rules generated out of the two methods. From all the analysis performed during the comparison of the two methods, it can be conferred that usage of Ontologies and the additional measures aid to a more efficient filtering of rules, in turn leading to a better discrimination removal methodology.

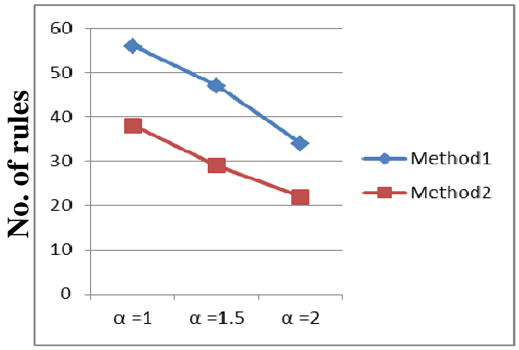

Figure 6: No. of α-discriminatory rules

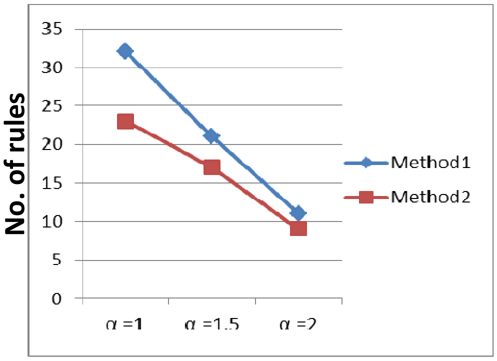

Figure 7: No. of redlining rules

5. Conclusion And Future Work

Due to the adoption of ARIPSO framework in the proposed system, the interestingness of the rules are preserved, and only those which evidently contribute to the decision making are retained in the resulting set of rules. This proves to be the advantage of the proposed system over the existing discrimination prevention methods. But much remains to be done in this arena to fine tune the proposed method, and some of the enhancements that are noteworthy are-

- Weighted lift and elift measures should be considered instead of flat measures for the attributes of the data set. By doing so, each attribute is assigned a value of importance, which might yield in more efficient method of discrimination prevention.

- Present real case studies for discrimination discovery and prevention using ontologies in data mining.

- Extend the existing approaches and algorithms to a variety of data mining tasks and multiple types of input data. Study and analyse the problem of discrimination prevention in run time, in the case of on-line transaction systems. This calls for attention due to the fact that the discrimination prevention algorithms should cater to the instant of service request and not on a repository of historical data.

- Extend concepts and methods to the analysis of discrimination in social network data. This provides an important case study because of the huge amounts of data that is present in the social networking sites, and their behavioural aspects pertaining to each user.

ACKNOWLEDGEMENTS

We express our gratitude to Dr.Shylaja S S, Professor and Head of Department of ISE, PES Institute of Technology, Bengaluru, India for facilitating during the tenure of the course. We are greatly thankful to all the researchers who have authored numerous literature works which has been a rich plethora of knowledge and forms the basis of this study.

REFERENCES

[1] Sara Hajian, 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] D. Pedreschi, S. Ruggieri, and F. Turini, “Discrimination-Aware Data Mining”, Proc. 14th ACM Int’l Conf. Knowledge Discovery andData Mining (KDD ’08),p. 560, 2008.

[3] Claudia Marinica and Fabrice Guillet,“Knowledge-Based Interactive Postmining of Association Rules Using Ontologies”,IEEE Transactions on Knowledge And Data Engineering, vol. 22, No. 6, June 2010

[4] T. Dalenius. The invasion of privacy problem and statistics production: an overview. Statistik Tidskrift, 12:213-225, 1974.

[5] P. Samarati and L. Sweeney. Generalizing data to provide anonymity when disclosing information. In Proc. of the 17th ACM SIGACTSIGMOD-SIGART Symposium on Principles of Database Systems (PODS 98), Seattle, WA, June 1998, p. 188.

[6] F. Kamiran and T. Calders. Data preprocessing techniques for classification without discrimination. Knowledge Information Systems, 33(1): 1-33, 2011.

[7] T. Calders and I. I. Zliobaite, “Why unbiased computational processes can lead to discriminative decision procedures. In 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, p. 4357. Springer, 2013.

[8] F. Kamiran and T. Calders, “Classification with no Discrimination by Preferential Sampling” , Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

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

[10] T. Calders and S. Verwer, “Three Naive Bayes Approaches for Discrimination-Free Classification,” Data Mining and Knowledge Discovery, vol. 21, no. 2, p. 277, 2010.

[11] D. Pedreschi, S. Ruggieri, and F. Turini, “Measuring Discrimination in Socially-Sensitive Decision Records,” Proc. Ninth SIAM Data Mining Conf. (SDM ’09), p. 581, 2009.

[12] D. Pedreschi, S. Ruggieri and F. Turini. Integrating induction and deduction for finding evidence of discrimination. In ICAIL 2009, p. 157. ACM, 2009.

[13] B. Liu, W. Hsu, K. Wang, and S. Chen, “Visually Aided Exploration of Interesting Association Rules,” Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining (PAKDD), p. 380, 1999.

[14] T.R. Gruber, “A Translation Approach to Portable Ontology Specifications,” Knowledge Acquisition, vol. 5, p. 199, 1993.

[15] H. Nigro, S.G. Cisaro, and D. Xodo, Data Mining with Ontologies: Implementations, Findings and Frameworks. Idea Group, Inc., 2007.

[16] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, “Discovering Frequent Closed Itemsets for Association Rules,” Proc. Seventh Int’l Conf. Database Theory (ICDT ’99), p. 398, 1999.

[17] M. Zaki, “Mining Non-Redundant Association Rules,” Data Mining and Knowledge Discovery, vol. 9, p. 223, 2004.

[18] A. Maedche and S. Staab, “Ontology Learning for the Semantic Web,” IEEE Intelligent Systems,vol. 16, no. 2, p. 72, Mar. 2001.

[19] D. Pedreschi, S. Ruggieri and F. Turini, “Measuring discrimination in socially sensitive decision records”, Proc. of the 9th SIAM Data Mining Conference (SDM 2009), p. 581. SIAM, 2009

[20] P. Samarati, “Protecting respondents' identities in microdata release” IEEE Transactions on Knowledge and Data Engineering, 13(6):1010-1027, 2001.

[21] L. Sweeney. “k-Anonymity:a model for protecting privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557-570, 2002.

[22] B. C. M. Fung, K. Wang, and P. S. Yu. Top-Down Specialization for Information and Privacy Preservation. In ICDE 2005, p. 205. IEEE, 2005