2010 IEEE International Conference on Data Mining

Discrimination Aware Decision Tree Learning

Faisal Kamiran, Toon Calders and Mykola Pechenizkiy
Email: {*f.kamiran,t.calders,m.pechenizkiy*}*@tue.nl*
Eindhoven University of Technology, The Netherlands

***Abstract*—Recently, the following discrimination aware classification problem was introduced: given a labeled dataset and an attribute B, find a classifier with high predictive accuracy that at the same time does not discriminate on the basis of the given attribute B. This problem is motivated by the fact that often available historic data is biased due to discrimination, e.g., when B denotes ethnicity. Using the standard learners on this data may lead to wrongfully biased classifiers, even if the attribute B is removed from training data. Existing solutions for this problem consist in "cleaning away" the discrimination from the dataset before a classifier is learned. In this paper we study an alternative approach in which the non-discrimination constraint is pushed deeply into a decision tree learner by changing its splitting criterion and pruning strategy. Experimental evaluation shows that the proposed approach advances the state-of-the-art in the sense that the learned decision trees have a lower discrimination than models provided by previous methods, with little loss in accuracy.**

I. Introduction and Motivation

Discrimination is a sociological term that refers to the unfair and unequal treatment of individuals of a certain group based solely on their affiliation to that particular group, category or class. Such discriminatory attitude deprives the members of one group from the benefits and opportunities which are accessible to other groups. Different forms of discrimination in employment, income, education, finance and in many other social activities may be based on age, gender, skin color, religion, race, language, culture, marital status, economic condition etc. Many anti-discrimination laws, e.g., the *Australian Sex Discrimination Act 1984*, the *US Equal Pay Act of 1963* and the *US Equal Credit opportunity act* have been enacted to eradicate the discrimination and prejudices, imposing severe penalties for the act of discriminating.

In this paper we consider the case where we plan to use data mining for decision making, but we suspect that our available historical data contains discrimination. Applying the traditional classification techniques on this data will produce biased models. Due to the above mentioned laws or simply due to ethical concerns the straightforward use of classification techniques is not acceptable. The solution is to develop new techniques which we call *discrimination aware* – we want to learn a classification model from the potentially biased historical data such that it generates accurate predictions for future decision making, yet does not discriminate with respect to a given discriminatory attribute. It can be argued that in many real-life cases discrimination can be explained; e.g., it may very well be that females in an employment dataset overall have less years of working experience, justifying a correlation between the gender and the class label. Nevertheless, in this paper we assume this not to be the case. We assume that the data is already divided up into strata based on acceptable explanatory attributes. Within a stratum, gender discrimination can no longer be justified.

As shown in previous works [7], [3], simply removing the sensitive attribute from the training data does not work, as other attributes may be correlated with the suppressed attribute. It was observed that classifiers tend to pick up these relations and discriminate indirectly. Therefore, in these works preprocessing methods that, prior to learning, cleanse away the discrimination were proposed. In this paper we explore another solution based on the integration of discrimination awareness into the model induction process of a decision tree. Particularly, we introduce the following two techniques for incorporating discrimination awareness into the decision tree construction process:

- **Dependency-Aware Tree Construction.** When evaluating the splitting criterion for a tree node, not only its contribution to the accuracy, but also the level of discrimination caused by this split is evaluated.
- **Leaf Relabeling.** Normally, in a decision tree, the label of a leaf is determined by the majority class of the tuples that belong to this node in the training set. In leaf relabeling we change the label of selected leaves in such a way that discrimination is lowered with a minimal loss in accuracy.

The results of an experimental study show what generalization performance we can achieve while trying to have as little discrimination as possible. The results show that the introduced discrimination-aware classification approach for decision tree learning improves upon previous methods that are based on dataset cleaning (or so-called Massaging) [3].

II. Related Work

The topic of discrimination in data mining recently received quite some attention. The authors of [10], [11] concentrate mainly on identifying the discriminatory rules that are present in a dataset, and the specific subset of the data where they hold, rather than on learning a discrimination aware classifier for future predictions. Discrimination-aware classification and its extension to independence constraints,

1550-4786/10 $26.00 © 2010 IEEE
DOI 10.1109/ICDM.2010.50

was first introduced in [7], [3], [8] where the problem of discrimination is handled by "cleaning away" the discrimination from the dataset before applying the traditional classification algorithms: *Massaging* changes the class labels of selected objects in the training data in order to obtain a discrimination free dataset while *Reweighing* selects a biased sample to neutralize the impact of discrimination. Next to the preprocessing techniques, also model selection techniques exist. [4] propose three approaches for making the naive Bayes classifiers discrimination-free.

III. Problem Statement

We assume a set of attributes $\{A_1, \ldots, A_n\}$ and their respective domains $dom(A_i)$, $i = 1 \ldots n$ have been given. A *tuple* over the schema $S = (A_1, \ldots, A_n)$ is an element of $dom(A_1) \times \ldots \times dom(A_n)$. We denote the component that corresponds to an attribute A of a tuple x by $x.A$. A dataset over the schema $S = (A_1, \ldots, A_n)$ is a finite set of tuples over S and a labeled dataset is a finite set of tuples over the schema $(A_1, \ldots, A_n, Class)$. $dom(Class) = \{+, -\}$.

As usual, a classifier C is a function from $\prod_{i=1}^{n} dom(A_i)$ to $\{+, -\}$. Let B be a binary attribute with domain $dom(B) = \{0, 1\}$. The *discrimination of C w.r.t. B in dataset D*, denoted $disc_B(C, D)$ is defined as :

$$disc_B(C, D) := \frac{|\{x \in D \mid x.B = 0, C(x) = +\}|}{|\{x \in D \mid x.B = 0\}|} - \frac{|\{x \in D \mid x.B = 1, C(x) = +\}|}{|\{x \in D \mid x.B = 1\}|} .$$

(When clear from the context we will omit B and D from the notation.) A higher discrimination means that tuples with $B = 1$ are less likely to be classified as positive by the classifier C than others. The discrimination of D w.r.t. B, denoted $disc_B(D)$, is defined as the difference

$$disc_B(D) := \frac{|\{x \in D \mid x.B = 0, x.Class = +\}|}{|\{x \in D \mid x.B = 0\}|} - \frac{|\{x \in D \mid x.B = 1, x.Class = +\}|}{|\{x \in D \mid x.B = 1\}|} .$$

For $\epsilon \in [0, 1]$, the formula $disc_B(C, D) \leq \epsilon$ is called a *non-discriminatory constraint*.

Problem 1 (Discrimination aware classification)**.** *Let a labeled dataset D and a sensitive attribute B be given. The discrimination aware classification problem is to learn a classifier such that (a) The accuracy of C is high, and (b) the discrimination of C w.r.t. B is low. (Both accuracy and discrimination are to be computed with respect to an unaltered test set).*

The formulation of the problem statement is rather informally requiring "high" accuracy and "low" discrimination. This ambiguity is not arbitrary, but due to the trade-off which exists between the accuracy and the resulting discrimination of a classifier. In general, lowering the discrimination will result in lowering the accuracy as well and vice versa.

In the remainder of the paper we make the following three assumptions:

(A) There is only one non-discriminatory constraint. The sensitive attribute is B and $dom(B) = \{0, 1\}$.

(B) The prime intention is learning the most accurate decision tree for which the discrimination is close to 0. Essentially we envision a scenario in which a maximally allowable discrimination ϵ is specified.

(C) As it is assumed that the discrimination on B is an artifact, the learned classifier should not use the attribute B at prediction time. Only at learning time we can use the attribute B.

IV. Solutions

In this section we propose two solutions to construct decision trees without discrimination. The first solution is based on the adaptation of splitting criterion for tree construction to build a discrimination-aware decision tree. The second approach is post-processing of decision tree with discrimination-aware pruning and relabeling of tree leaves.

A. Discrimination-Aware Tree Construction

Traditionally, when constructing a decision tree, we iteratively refine a tree by iteratively splitting its leaves until a desired objective is achieved.The optimization criteria used are usually locally optimizing the overall accuracy of the tree, e.g., based on the so-called *information gain*. Suppose that a certain split divides the data D into $D_1, \ldots, D_k$. Then, the information gain is defined as $IGC := H_{Class}(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} H_{Class}(D_i)$, where H_{Class} denotes the entropy w.r.t. the class label, and D_i, $i = 1 \ldots k$ are the partitions induced by the splitting criterion under evaluation. From all splitting criteria being considered, the one that (locally) optimizes the information gain is chosen.

In this paper, however, we are not only concerned with accuracy, but also with discrimination. Therefore, we will change the iterative refinement process by also taking into account the influence of the splits under evaluation on the discrimination of the resulting tree. To measure the influence of the introduction of a split on the discrimination, we will use the same notion of information gain, but now also w.r.t. the sensitive attribute B instead of only w.r.t. the class $Class$. This gain in sensitivity to B will be denoted IGS; i.e.,

$$IGS := H_B(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} H_B(D_i) ,$$

with H_B the entropy w.r.t. sensitive attribute. Based on these two measures IGC and IGS, we introduce three alternative criteria for determining the best split:

IGC-IGS: We only allow for a split if it is non-discriminatory, i.e., we select an attribute which is homogenous w.r.t. class attribute but heterogenous w.r.t. sensitive

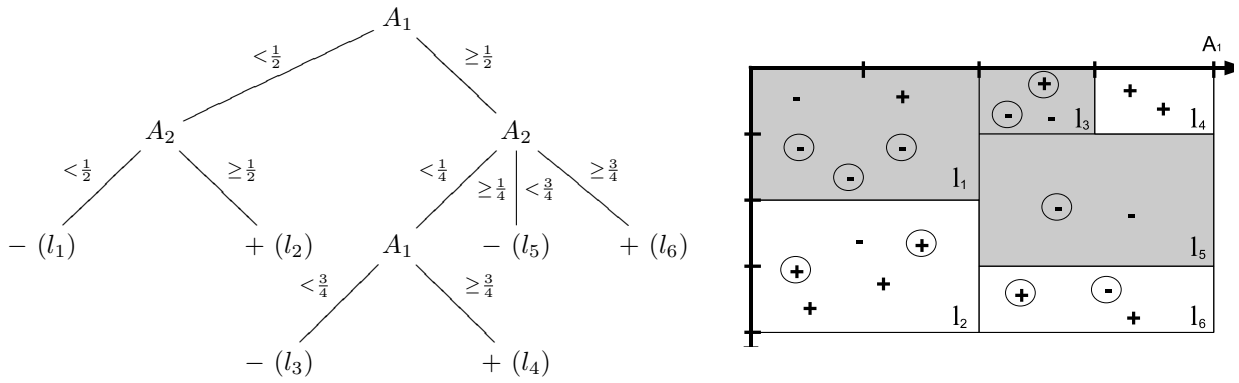


Figure 1. Decision tree with the partitioning induced by it. The $+$ and $-$ symbols in the partitioning denote the examples that were used to learn the tree. Encircled examples have $B = 1$. The grey background denotes regions where the majority class is $-$

attribute. We subtract the gain in discrimination from the gain in accuracy to make the tree homogenous w.r.t. class attribute and heterogenous w.r.t. sensitive attribute.

IGC/IGS: We make a trade-off between accuracy and discrimination by dividing the gain in accuracy by gain in discrimination.

IGC+IGS: We add up the accuracy gain and the discrimination gain. This favors splits that result in a homogenous tree w.r.t. both accuracy and the sensitive attribute. Even though this measure in isolation does not make sense as it favors more discrimination, it will lead to good results in combination with the relabeling technique we show next.

Many other measures could have been tried as well, yet as the experiments will show, this technique was not very successful.

B. Relabeling

For the relabeling approach we assume that a tree is already given and the goal is to reduce the discrimination of the tree by changing the class labels of some of the leaves. Let T be a decision tree with n leaves. Such a decision tree partitions the example space into n non-overlapping regions. See Figure 1 for an example; in this figure (left) a decision tree with 6 leaves is given, labeled l_1 to l_6. The right part of the figure shows the partitioning induced by the decision tree. When a new example needs to be classified by a decision tree, it is given the majority class label of the region it falls into; i.e., the leaves are labeled with the majority class of their corresponding region.

The *relabeling* technique, however, will now change this strategy of assigning the label of the majority class. Instead, we try to relabel the leaves of the decision tree in such a way that the discrimination decreases while trading in as little accuracy as possible. We can compute the influence of relabeling a leaf on the accuracy and discrimination of the tree on a dataset D as follows. Let the joint distributions of the class attribute C and the sensitive attribute B for respectively the whole dataset and for the region corresponding to the leaf be given by the following contingency table (For the dataset additionally the frequencies have been split up according to the predicted labels by the tree):

Dataset

Class →	$-$	$+$	
Pred. →	$-/+$	$-/+$	
$B=1$	U_1/U_2	V_1/V_2	b
$B=0$	W_1/W_2	X_1/X_2	$\bar{b}$
	N_1/N_2	P_1/P_2	1

Leaf l

	$-$	$+$	
$B=1$	u	v	b
$B=0$	w	x	$\bar{b}$
	n	p	a

Hence, e.g., a fraction a of the examples end up in the leaf we are considering for change, of which n are in the negative class and p in the positive. Notice that for the leaf we do not need to split up u, v, w, and x since all examples in a leaf are assigned to the same class by the tree.

With these tables it is now easy to get the following formulas for the accuracy and discrimination of the decision tree *before* the label of the leaf l is changed:

$$\begin{aligned} acc_T &= N_1 + P_2 \\ disc_T &= \frac{W_2 + X_2}{\bar{b}} - \frac{U_2 + V_2}{b} \end{aligned}$$

The effect of relabeling the leaf now depends on the majority class of the leaf; on the one hand, if $p > n$, the label of the leaf changes from $+$ to $-$ and the effect on accuracy and discrimination is expressed by:

$$\begin{aligned} \Delta acc_l &= n - p \\ \Delta disc_l &= \frac{u + v}{b} - \frac{w + x}{\bar{b}} \end{aligned}$$

on the other hand, if $p < n$, the label of the leaf changes from $-$ to $+$ and the effect on accuracy and discrimination is expressed by:

$$\begin{aligned} \Delta acc_l &= p - n \\ \Delta disc_l &= -\frac{u + v}{b} + \frac{w + x}{\bar{b}} \end{aligned}$$

Algorithm 1: *Relabel*

Input Tree T with leaves $\mathcal{L}$, $\Delta acc(l), \Delta disc(l)$ for every $l \in \mathcal{L}, \epsilon \in [0,1]$
Output Set of leaves L to relabel

$\mathcal{I} := \{\ l \in \mathcal{L} \mid \Delta disc_l < 0\ \}$
$L := \{\}$
while $rem\_disc(L) > \epsilon$ **do**
  $best\_l := \arg\max_{l \in \mathcal{I} \setminus L}(disc_l / acc_l)$
  $L := L \cup \{l\}$
end while
return L

Notice that relabeling leaf l does not influence the effect of the other leaves and that Δacc_l is always negative.

Example 1. *Consider the dataset and tree given in Figure 1. The contingency tables for the dataset and leaf l_3 are as follows:*

Dataset

Class →	−	+	
Pred. →	−/+	−/+	
$B=1$	$\frac{5}{20}/\frac{1}{20}$	$\frac{1}{20}/\frac{3}{20}$	$\frac{1}{2}$
$B=0$	$\frac{3}{20}/\frac{1}{20}$	$\frac{1}{20}/\frac{5}{20}$	$\frac{1}{2}$
	$\frac{8}{20}/\frac{2}{20}$	$\frac{2}{20}/\frac{8}{20}$	1

Leaf l_3

	−	+	
$B=1$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
$B=0$	$\frac{1}{20}$	0	$\frac{1}{20}$
	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{3}{20}$

The effect of changing the label of node l_3 from $-$ to $+$ hence is: $\Delta acc_l = -\frac{1}{20}$ and $\Delta disc_l = -\frac{1}{10}$.

The central problem now is to select exactly this set of leaves that is optimal w.r.t. reducing the discrimination with minimal loss in accuracy, as expressed in the following *Optimal relabeling problem* (RELAB):

Problem 2 (RELAB)**.** *Given a decision tree T, a bound $\epsilon \in [0,1]$, and for every leaf l of T, Δacc_l and $\Delta disc_l$, find a subset L of the set of all leaves $\mathcal{L}$ satisfying*

$$rem\_disc(L) \ := \ disc_T + \sum_{l \in L} \Delta disc_l \leq \epsilon$$

that minimizes

$$lost\_acc(L) \ := \ -\sum_{l \in L} \Delta acc_l \ .$$

The RELAB problem can be reduced to the well-known combinatorial optimization problem KNAPSACK [2]:

Problem 3 (KNAPSACK [2])**.** *Let a set of items $\mathcal{I}$, a weight $w(i)$ and a profit $p(i)$, both positive integers, for every item $i \in \mathcal{I}$, and an integer bound K be given. Find a subset $I \subseteq \mathcal{I}$ subject to $\sum_{i \in I} w(i) \leq K$ that maximizes $\sum_{i \in I} p(i)$.*

The following theorem makes the connection between the two problems explicit.

Theorem 1. *Let T be a decision tree, and $\epsilon \in [0,1]$ and for every leaf l of T, Δacc_l and $\Delta disc_l$ have been given.*

The RELAB problem with this input is equivalent to the KNAPSACK problem with the following inputs:

- $\mathcal{I} = \{\ l \in \mathcal{L} \mid \Delta disc_l < 0\ \}$
- $w(l) = -\alpha \Delta disc_l \qquad$ *for all* $l \in \mathcal{I}$
- $p(l) = -\alpha \Delta acc_l \qquad$ *for all* $l \in \mathcal{I}$
- $K = \alpha \left(\sum_{l \in \mathcal{I}} disc_l - disc_T + \epsilon \right)$

Where α is the smallest number such that all $w(l), p(l)$, and K are integers. Any optimal solution L to the RELAB problem corresponds to a solution $I = \mathcal{I} \setminus L$ for the KNAPSACK problem and vice versa.

Based on the connection with the KNAPSACK problem, the greedy Algorithm 1 is proposed for approximating the most optimal relabeling. The proofs of theorem, the NP-completeness of RELAB and a bound on the greedy algorithm can be found in [9].

V. Experiments

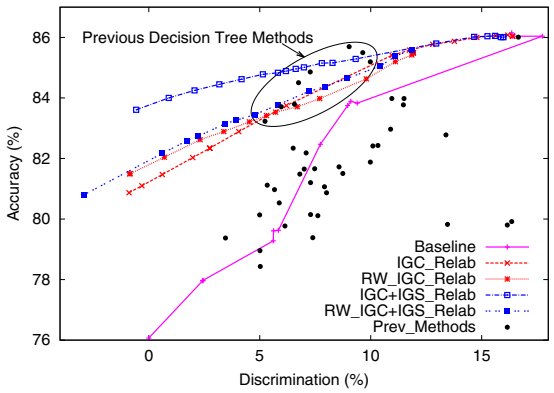
All datasets and the source code of all implementations reported upon in this section are available at http://www.win.tue.nl/~fkamiran/code.

In this section we show the results of experiments with the new discrimination-aware splitting criteria and the leaf relabeling for decision trees. As we observe that the discrimination-aware splitting criteria by themselves do not lead to significant improvements w.r.t. lowering discrimination, we have omitted them from the experimental validation. However, the new splitting criteria IGC+IGS is an exception: sometimes, when used in combination with leaf relabeling, it outperforms the leaf relabeling with original decision tree split criterion IGC. IGC+IGS in combination with relabeling outperforms other splitting criteria because this criterion tries to make tree leaves homogenous w.r.t. both class attribute and sensitive attribute. The more homogenous w.r.t. the sensitive attribute the leaves are, the less number of leaves we will have to relabel to remove the discrimination from the decision tree. For the relabeling approach, however, the results are very encouraging, even when the relabeling is applied with normal splitting criterion IGC. We compare the following techniques (between brackets their short name):

(1) The baseline solutions (Baseline) that consist of removing B and its k most correlated attributes from the training dataset before learning a decision tree, for $k = 0, 1, \ldots, n$. In the graphs this baseline will be represented by a continuous line connecting the performance figures for increasing k.

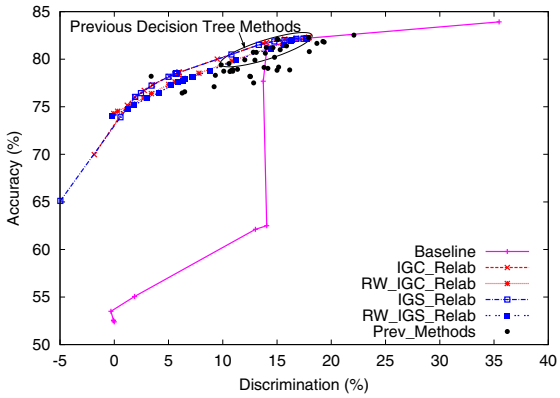
(2) We also present a comparison to the previous state-of-the-art (Prev_Methods) techniques, shown in Figure 2, which includes the pre-processing methods *Massaging* and *Reweighing* [7], [3] that are based on cleaning away the discrimination from the input data before a traditional learner is

(a) Census Income Data



Baseline Disc=19.3 Acc=76.3

(b) Dutch Census 2001 Data



Baseline Disc=29.85 Acc=52.39

Figure 2. Accuracy-discrimination trade-off for different values of epsilon $\epsilon \in [0, 1]$ is plotted. We change the value of epsilon from the baseline discrimination in the dataset (top right points of lines) to the zero level (bottom left points of these lines).

applied and discrimination aware *naive Bayesian* approaches [4].

(3) From the proposed methodes we show the relabeling approach in combination with normal decision tree splitting criteria (IGC_Relab) and with new splitting criteria IGC+IGS (IGC+IGS_Relab).

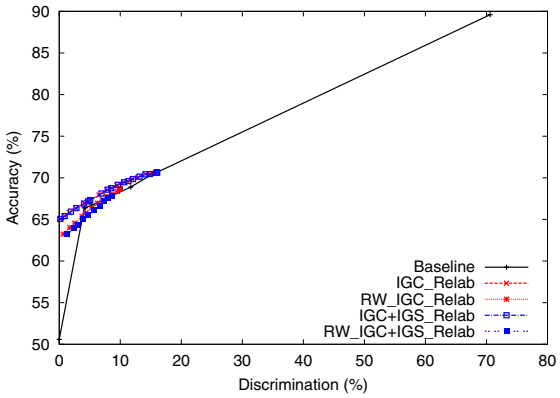
(4) Finally we also show some hybrid combinations of the old and new methods; we present the results of experiments where we first applied the *Reweighing* technique of [3] on the training data to learn a tree with low discrimination (either with the normal or the new splitting criterion). On this tree we then apply relabeling to remove the last bit of discrimination from it (RW_IGC_Relab and RW_IGC+IGS_Relab). The other combinations led to similar results and are omitted from the comparison.

We apply our proposed solutions on the Census Income dataset [1], and two Dutch census datasets of 1971 and 2001 [5], [6].

Testing the Proposed Solutions. The reported figures are the averages of a 10-fold cross-validation experiment. Every point represents the performance of one decision tree on original test data for which the sensitive attribute was not used during prediction. Every point in the graphs corresponds to the discrimination (horizontal axis) and the accuracy (vertical axis) of a classifier produced by one combination of a splitting criteriun (IGC, IGC-IGS, IGC/IGS, or IGS+IGS) and a relabeling approach (RELAB). Ideally, points should be close to the top-left corner. The comparisons show clearly that relabeling succeeds in lowering the discrimination much further than the baseline and previous state-of-the-art approaches. Figure 2 shows a comparison of our discrimination aware techniques with the baseline approach over two different datasets. We observe that the discrimination goes down by removing the sensitive attribute and its correlated attribute but its impact on the accuracy is very severe. On the other hand the discrimination aware methods classify the unseen data objects with minimum discrimination and high accuracy for all values of ϵ. Figure 2 shows that our proposed methods outperform the current state-of-the-art methods w.r.t. both accuracy and discrimination.

It is very important to notice that we measure the accuracy scores over discriminatory data. Ideally we would have non-discriminatory test data at our disposition. If our test set would be non-discriminatory, we expect our discrimination aware methods to outperform the traditional method w.r.t. both accuracy and discrimination. In our experiments, we mimic this scenario by using the Dutch 1971 Census data as a training set and the Dutch 2001 Census dataset as a test set. We use the attribute *economic status* as class attribute because this attribute uses similar codes for both 1971 and 2001 dataset. The use of *occupation* (used as class attribute in the experiments of Figure 2 (b)) as class attribute was not possible in these experiments because its coding is different in both datasets. The attribute *economic status* determines whether a person has some job or not, i.e., is economically active or not. We remove some attributes like *current economic activity* and *occupation* from these experiments to make both datasets consistent w.r.t. codings. In Dutch 1971 Census data, there is more discrimination toward female and their percentage of unemployment is higher than in the Dutch 2001 Census data. Now if we learn a traditional classifier over 1971 data and test it over the same dataset using 10-fold cross validation method, it will give excellent performance as shown in Figure 3 (a). When we apply this classifier to 2001 data without taking the discrimination aspect into account, it performs very poorly and accuracy level goes down from 89.6% (when tested on 71 data; Figure 3 (a)) to 73.09% (when tested on 2001 data; Figure 3 (b)). Figure 3 makes it very obvious that our discrimination aware technique not only classify the

(a) Dutch 1971 Census data is as test set.



(b) Dutch 2001 Census data is used as test set.

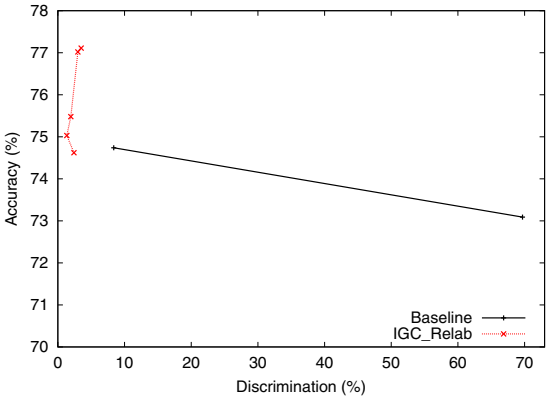


Figure 3. The results of experiments when Dutch 1971 Census dataset is used as train set while the test set is different for both plots.

future data without discrimination but they also work more accurately than the traditional classification methods when tested over non-discriminatory data. In Figure 3 (b), we only show the results of IGC_Relab because other proposed methods also give similar results. Figure 3 (b) shows that if we change the value of ϵ from 0 to 0.04 the accuracy level increases significantly from 74.62% to 77.11%. We get the maximum accuracy at $\epsilon = 0.04$ because the Dutch 2001 Census data is not completely discrimination free.

From the results of our experiments we draw the following conclusions: (1) Our proposed methods give high accuracy and low discrimination scores when applied to non-discriminatory test data. In this scenario, our methods are the best choice, even if we are only concerned with accuracy. (2) The improvement in discrimination reduction with the relabeling method is very satisfying (reduces discrimination to almost 0 at $\epsilon = 0$). (3) The relabeling methods outperform the baseline in almost all cases. (4) Our methods significantly improve the current stat-of-the-art techniques w.r.t. accuracy-discrimination trade off.

VI. Conclusions

In this paper we presented the construction of a decision tree classifier without discrimination. This is a different approach of addressing the discrimination-aware classification problem. Most of the previously introduced approaches were focused on "removing" undesired dependencies from the training data and thus can be considered as "preprocessors". In this paper on the contrary, we propose the construction of decision trees with non-discriminatory constraints. Especially relabeling, for which an algorithm based on the KNAPSACK problem was proposed, showed promising results in an experimental evaluation. It was shown to outperform the other discrimination aware techniques by giving much lower discrimination scores and maintaining the accuracy high. Moreover, it is shown that if we are only concerned with accuracy, our method is the best choice when training set is discriminatory and test set is non-discriminatory.

Acknowledgments: This work was supported by funding from the Netherlands Organization for Scientific Research (NWO) and Higher Education Commission (HEC) of Pakistan.

References

[1] A. Asuncion and D. Newman. UCI machine learning repository. 2007.

[2] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Prosati. *Complexity and Approximation. Combinatorial Optimization Problems and Their Approximability Properties*. Springer, 2003.

[3] T. Calders, F. Kamiran, and M. Pechenizkiy. Building classifiers with independency constraints. In *IEEE ICDM Workshop on Domain Driven Data Mining*. IEEE press., 2009.

[4] T. Calders and S. Verwer. Three naive bayes approaches for discrimination-free classification (accepted for publication). In *Proc. ECML/PKDD*, 2010.

[5] Dutch Central Bureau for Statistics. Volkstelling, 1971.

[6] Dutch Central Bureau for Statistics. Volkstelling, 2001.

[7] F. Kamiran and T. Calders. Classifying without discriminating. In *Proc. IC409*. IEEE press.

[8] F. Kamiran and T. Calders. Classification with no discrimination by preferential sampling. In *Proc. BENELEARN*, 2010.

[9] F. Kamiran, T. Calders, and M. Pechenizkiy. Discrimination aware tree learning. Technical report, Dept. Comp. Science, TU/e, 2010.

[10] D. Pedreschi, S. Ruggieri, and F. Turini. Discrimination-aware data mining. In *Proc. ACM SIGKDD*, 2008.

[11] D. Pedreschi, S. Ruggieri, and F. Turini. Measuring discrimination in socially-sensitive decision records. In *Proc. SIAM DM*, 2009.