Discrimination-Aware Classifiers for Student Performance Prediction

Ling Luo
School of Information Technologies,
The University of Sydney, Australia
National ICT Australia
ling.luo@sydney.edu.au

Irena Koprinska
School of Information Technologies,
The University of Sydney, Australia
irena.koprinska@sydney.edu.au

Wei Liu
Faculty of Engineering & IT, University
of Technology Sydney, Australia
National ICT Australia
wei.liu@uts.edu.au

ABSTRACT

In this paper we consider discrimination-aware classification of educational data. Mining and using rules that distinguish groups of students based on sensitive attributes such as gender and nationality may lead to discrimination. It is desirable to keep the sensitive attributes during the training of a classifier to avoid information loss but decrease the undesirable correlation between the sensitive attributes and the class attribute when building the classifier. We illustrate, motivate, and solve the problem, and present a case study for predicting student exam performance based on enrolment information and assessment results during the semester. We evaluate the performance of two discrimination-aware classifiers and compare them with their non-discrimination-aware counterparts. The results show that the discrimination-aware classifiers are able to reduce discrimination with trivial loss in accuracy. The proposed method can help teachers to predict student performance accurately without discrimination.

Keywords

Predicting student performance; association rule mining; decision tree; discrimination-aware classification

1. INTRODUCTION

Educational data often contains sensitive attributes such as age, gender and nationality. Mining such data may generate discriminating rules. For example, if our goal is to predict the exam mark of current students, and in the historic dataset used for training of the prediction algorithm, males have achieved significantly higher exam marks than females, a prediction rule using the attribute gender may be generated. It may produce high accuracy but we cannot use it for providing feedback to students or other decision making, as it can be seen as discriminating based on gender, which is unethical and also against the law. Sensitive attributes such as gender should be used as an information carrier and not as distinguishing factors [1]. In this paper we consider building discrimination-aware classification models for predicting student performance.

The task of discrimination-aware classification can be defined as follows [2; 3]: given a labelled dataset and an attribute S, find a classifier with high accuracy that does not discriminate on the basis of S. There are two approaches to deal with this problem: 1) not using the sensitive attribute to build the classifier and 2)

modifying the classification algorithm by integrating a discrimination-aware mechanism to reduce discrimination. The first approach, simply removing the sensitive attribute from the training data, results in information loss and also typically doesn't solve the problem as other attributes are correlated with the sensitive attribute, and will discriminate indirectly. In this paper, we develop and apply methods from the second group which incorporate discrimination awareness during the building of the classifier and use information from the sensitive attribute without causing discrimination.

There are two important aspects that need to be considered when applying discrimination-aware classifiers in educational settings. Firstly, adjusting the classifier to reduce discrimination typically leads to lower predictive accuracy. Given this trade-off between accuracy and discrimination, our aim is to build a classifier with lower discrimination without significant loss in accuracy. Secondly, the output of the classifier should be easy to understand and use by teachers and students. Therefore, we consider classifiers based on decision tree and association rules, which generate sets of rules to guide prediction and decision making.

Our contribution can be summarized as follows:

- We illustrate and motivate the problem of discrimination-aware classification for mining educational data, and show its importance and challenges in educational data mining. Discrimination-aware classification has not been studied for educational data mining and our main goal is to raise the awareness of the community to this problem.

- We introduce our recently proposed classification method Discrimination-aware Association Rule classifier (DAAR) [4]. DAAR uses the novel Discrimination Correlation Indicator (DCI) to measure the discrimination severity of an association rule and select non-discriminatory rules.

- We consider the task of predicting the student exam performance in a first year computer programming course. We apply two discrimination-aware classifiers: our method DAAR and the state-of-the-art Discrimination-Aware Decision Tree (DADT) [3], and compare their performance with standard non-discrimination-aware association rules and decision tree. We show that both DAAR and DADT are able to produce non-discriminatory rules with minimum loss in accuracy.

2. RELATED WORK

Mining educational data to predict student performance has gained increasing popularity. Romero et al. [5] predicted the final student mark based on the Moodle usage data such as the number of passed and failed quizzes, number of completed assignments, number of sent and read messages on the discussion board and the time spent on the assignments, quizzes and discussion board. In their subsequent work [6], the same group studied predicting the

student grade (pass or fail) based on the student participation in a discussion forum, using a number of machine learning algorithms, in the middle and at the end of the semester. Kotsiantis et al. [7] applied an ensemble of classifiers to predict the exam grade (pass or fail) from assessment data during the semester in an online informatics course. Lykourentzou et al. [8] predicted dropouts and completers in e-learning courses on computer networks and web design, using demographic and assessment data.

The discrimination-aware classification problem was introduced in by Pedreshi et al. [2] and Kamiran and Calders [9]. Discrimination-aware naïve Bayes approaches were proposed in [1] and discrimination-aware decision trees were developed in [3].

In this paper, we investigate discrimination-aware classifiers for mining of educational data. We apply our recently proposed discrimination-aware classifier based on association rules and also a discrimination-aware decision tree. We show how these algorithms can be applied for predicting student performance in a first year programming course, discuss the results, and raise the awareness of the Educational Data Mining community to the importance of discrimination-free classification.

3. METHODOLOGY
In this section we describe the main principles of the two discrimination-aware classifiers: our method DAAR and the state-of-the-art DADT. Both classifiers are designed to decrease the discrimination of the predictive model with minimal impact on the accuracy. They are based on the popular and successful association rule classifiers and decision trees, which produce rules that can be easily understood and directly applied by teachers and students.

3.1 Association Rule Classifiers and DAAR
Association analysis discovers relationships among items in a dataset. An association rule takes the form $X \rightarrow Y$, where X and Y are disjoint item sets [10]. Two measures, *support* and *confidence*, are used to evaluate the quality of an association rule. Given a dataset containing N instances and an association rule $X \rightarrow Y$, the support and confidence of this rule are defined as:

$$\text{Support } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}, \qquad \text{Confidence } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$$

where $\sigma(\cdot)$ is the frequency of an item set $(\cdot)$. High-quality rules have high support and confidence.

Classification Based on Association (CBA) [10] uses association rules to solve classification problems. In a standard association rule, any attribute which is not included in X, can appear in Y while in CBA only class attributes can appear in Y.

3.1.1 DCI Measure
To measure the degree of discrimination for an association rule, we propose a new measure called DCI. Given a rule $X \rightarrow y$ and a sensitive attribute S, DCI is defined as:

$$\text{DCI} = \begin{cases} \dfrac{|P(C = y| S = S_{\text{rule}}) - P(C = y| S = S_{\text{others}})|}{\left(P(C = y| S = S_{\text{rule}}) + P(C = y| S = S_{\text{others}})\right)} \\ 0 \qquad \text{if either of the above } P(\cdot) \text{ is } 0 \end{cases}$$

where $P(C = y|S = S_{\text{rule}})$ is the probability of the class to be y given the value of the sensitive attribute S is S_{rule}.

When S is a binary or multi-valued attribute, the specific S value in the rule is considered as S_{rule}, and the S_{others} includes the set of all attribute values except the one which appears in the rule. For example, if the rule is "*gender = female, degree = CS → assessment = low*", where *gender* is the sensitive attribute, then

S_{rule} refers to *female*, and S_{others} refers to *male*. The DCI for this rule will be:

$$\frac{|P(C = \text{low}|\text{gender} = \text{female}) - P(C = \text{low}|\text{gender} = \text{male})|}{P(C = \text{low}|\text{gender} = \text{female}) + P(C = \text{low}|\text{gender} = \text{male})}$$

When the sensitive attribute does not appear in that rule, we define DCI to be 0.

Therefore, DCI has a range of [0, 1) and its interpretation is the following:

- If DCI is 0, the rule is free of discrimination. DCI is 0 when the probability of the class value to be y is the same for different values of the sensitive attribute S.
- If DCI is not 0, the higher the value, the more discriminatory the rule is with respect to the sensitive attribute S. Thus, the DCI value is monotonically increasing with the discriminatory severity of a rule.

3.1.2 DAAR
DAAR uses DCI together with minimum confidence and support to efficiently select non-discriminatory rules. DAAR's algorithm is shown in Figure 1.
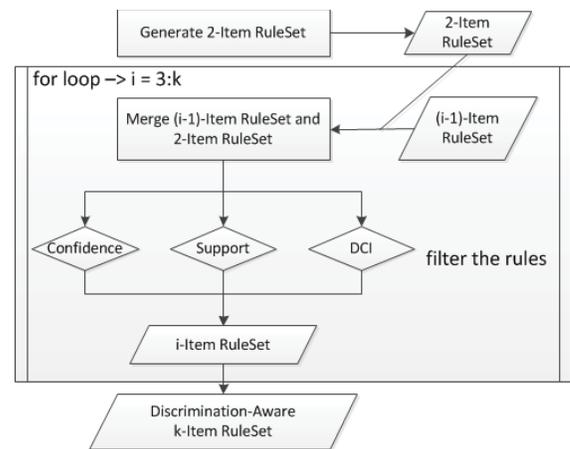

Figure 1. DAAR's Algorithm

DAAR starts from the set of 2-item rules (i.e. the rules with one attribute value and the class attribute), which is the base case, and merges with other 2-item rules iteratively until it gets the k-item rules, where k is the upper bound for the number of items in the rule. In each iteration, the rules are filtered by confidence, support and DCI. To classify new instances, DAAR uses majority voting based on the number of rules that predict the same class. If the vote is tied, the DCI sum for all rules for each class is compared and the class with lower sum (i.e. less discrimination) is selected.

3.2 Decision Tree and DADT
Decision Trees (DTs) are one of the most popular machine learning algorithms. The standard DT algorithm uses information gain to select the best attribute at each step as a root of the tree/subtree, until all examples in the subset belong to the same class, in which case it creates a leaf node labelled with this class. DTs can be seen as generating a set of mutually exclusive rules – each path from the root of the tree to a leaf node is one rule, and each rule is a conjunction of attribute tests. DADT is a discrimination-aware version of DT introduced by Faisal et al. in [3]. The tree is constructed in two phases. In the first phase, it generates a tree by using a new splitting criterion: IGC-IGS. IGC is the standard information gain (Information Gain regarding the

Class label) and IGS is Information Gain regarding the Sensitive attribute, defined as:

$$IGS = H_S(D) - \sum_{j=1}^{k} \frac{|D_j|}{|D|} H_S(D_j)$$

where S is the sensitive attribute, $H_S(D) = -\sum_{i=1}^{n} P_i * \log_2 P_i$ is the entropy of set D with respect to S and P_i is the proportion of items with the ith value of the sensitive attribute.

As the aim is to have higher IGC but lower IGS, the difference IGC-IGS is an appropriate criterion. In the second phase, the leaves are relabeled to decrease the discrimination severity to less than ε (where ε is a non-discriminatory constraint), while sacrificing as little accuracy as possible. Experiments on census income datasets showed that DADT can produce a tree with a lower discrimination while maintaining accuracy [3].

4. EXPERIMENTS AND RESULTS

We consider the task of predicting exam performance in a fist year programming course. We compare the performance of the discrimination-aware classifiers DAAR and DADT with their standard non-discrimination-aware counterparts CBA (standard AR) and C4.5 (standard DT).

4.1 Dataset and Experimental Setup

Learning computer programming is difficult as it requires a lot of practice with feedback, and a very precise way of thinking. It is easy for students to fall behind, especially since introductory computer programming courses have a large number of students. Predicting students at risk of failing or not performing well is highly desirable.

Our evaluation is conducted using data from a first year computer programming course at an Australian University with 220 students. Our goal is to predict the exam performance, *high* or *low*, based on the student grades on the assessments during the semester and some enrolment attributes such as country of residence, degree name and if the student is local or international. A description of the attributes and their values is given in Table 1.

Table 1. Description of Attributes

Attribute	Description	Number of Attribute Values
Country	Country of permanent residence: {Australia, Brazil, China, …}	26
Degree	Name of the degree the student is enrolled into: {Bachelor of Science, Bachelor of Engineering,…}	27
Local	Indicates if the student is Australian or not: {Local, International}	2
a1_grade	The grade of assessment 1 during semester: {HD, D, CR, P, F}	5
a2_grade	The grade of assessment 2 during semester: {HD, D, CR, P, F}	5
a3_grade	The grade of assessment 3 during semester: {HD, D, CR, P, F}	5
a4_grade	The grade of assessment 4 during semester: {HD, D, CR, P, F}	5
a5_grade	The grade of assessment 5 during semester: {HD, D, CR, P, F}	5
Exam	Exam performance during examination period: {high, low}	2

The grades for the 5 assessments during the semester are the standard grades used at the university defined as follows: *HD* (High Distinction, mark of [85, 100]), *D* (Distinction, mark of [75, 84]), *CR* (Credit, mark of [65, 74]), *P* (Pass, mark of [50, 64]) and

F (Fail, mark below 50). The exam performance is defined as *high* if the exam mark is 65 or higher (i.e. HD, D or CR), and *low* if it is below 65 (i.e. P or F). There were 105 students in the *high* group and 115 in the *low* group.

We selected the exam grade as a variable to predict rather than the final grade in the course, as the exam is the major assessment component (worth 50% and covering all topics) and it is also independent of the assessment components during the semester, while these components contribute to calculating the final grade for the course.

Among the 8 predictors, we consider *country* as the sensitive attribute, which means that we would like to avoid discrimination based on the student nationality. Originally, this attribute had 26 different values, with 5 or less number of students for most of the countries, so we aggregated these values into three groups: *Australia*, *China* and *Others*. The number of students in each group was 127, 54 and 39, respectively.

4.2 Results and Discussion

To evaluate the performance of the classification methods, we use 10-fold cross validation in all experiments. We report both the average value and the standard deviation for the 10 folds. As predictive accuracy measures, we use both classification accuracy and F-measure.

To assess the discrimination severity of the classifier, we calculate a discrimination score. In [1] a discrimination score for a binary sensitive attribute S with values S_1 and S_2, and class values C_+ and C_- is defined as:

$$Score = |P(C = C_+| S = S_1) - P(C = C_+| S = S_2)|$$

As our sensitive attribute has three values, we extend this definition to multi-valued attribute with m ($m>2$) values. We compute the score for each value S_i and then average the m scores:

$$Score = \frac{1}{m} * \left(\sum_{i=1}^{m} |P(C = C_+| S = S_i) - P(C = C_+| S = S_{others})| \right)$$

where S_{others} represents all the attribute values other than S_i.

If the score is 0, there is no discrimination. Otherwise, a higher score corresponds to a higher discrimination severity.

4.2.1 DAAR

Table 2 presents the accuracy results and discrimination score for the standard AR and DAAR. We can see that DAAR was able to decrease the discrimination score of AR from 0.2831 to 0.2653. The trade-off was a slightly lower accuracy - DAAR achieved 73.92% accuracy, which is 4.72% lower than AR's accuracy.

Table 2. Results for Standard AR and DAAR

	Standard AR		DAAR	
	Mean	Std.	Mean	Std.
Accuracy	78.64%	0.0037	73.92%	0.0128
F-measure	0.7863	0.0037	0.7389	0.0131
Disc. score	0.2831	0.0109	0.2653	0.0163

Table 3 shows some representative and interesting rules produced by DAAR with their confidence, support and DCI. These rules are very compact, easy to understand and apply by teachers.

Table 4 shows the rules with high confidence and support that were filtered out by DAAR, as they were discriminatory with respect to the sensitive attribute *country*.

Table 3. Sample Rules Produced by DAAR

Rules	Conf.	Sup.	DCI
a1_grade=CR → exam = low	1.0	0.01	0
degree= Bachelor of Commerce, a4_grade=HD → exam = high	1.0	0.01	0
a4_grade=F → exam = low	1.0	0.19	0
degree= Bachelor of Engineering & Bachelor of Science, a5_grade=HD → exam = high	0.84	0.08	0

Table 4. Discriminatory Rules Removed by DAAR

Rules	Conf.	Sup.	DCI
country=Other → exam = high	0.62	0.12	0.17
country=CH → exam = low	0.77	0.18	0.26
country=Others, a5_grade=HD, a4_grade=HD → exam = high	0.83	0.08	0.17

4.2.2 DADT

The trees produced by the standard DT and DADT are shown in Figure 2 and Figure 3, respectively. The standard DT achieved accuracy of 83.46% but it used the sensitive attribute *country* and its discrimination score was 0.2298. DADT achieved a slightly lower accuracy of 82.73% without using the sensitive attribute. Thus, DADT is able to avoid discrimination with a minimum loss in accuracy. Both DTs included the attribute *a4_grade* as a root of the tree, which shows the importance of this attribute for predicting exam performance.
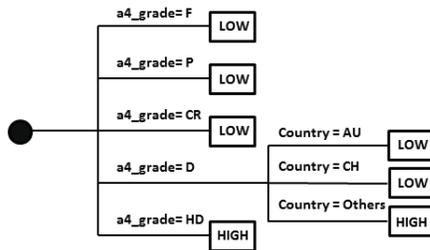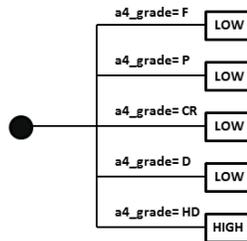

Figure 2. Tree Produced by the Standard DT

Figure 3. Tree Produced by DADT

4.2.3 Discussion

In terms of overall performance, all four methods had reasonable accuracy, from 73.92% to 83.46%, with the DT-based classifiers outperforming the AR-based classifiers. All classifiers generated a small set of rules that are easy to understand and use by teachers. The AR classifiers used more attributes in the rules which, for our case study, provided additional insights about the important attributes in predicting student performance and providing feedback to students.

In terms of discrimination, we can see that both DAAR and DADT decreased the severity of the discrimination compared to their standard counterparts, with trivial loss in accuracy.

Specifically, DAAR removed the rules with higher DCI values and reduced the discrimination score, and DADT using IGC-IGS as an attribute selection criterion, built a DT without using the sensitive attribute *country*.

5. CONCLUSIONS

Educational data often contains sensitive attributes, which should only be used as information carriers rather than factors to distinguish students and potentially discriminate them. We investigated discrimination-aware classification for mining of educational data, with a case study in predicting student exam performance based on enrolment information and assessment marks during the semester, in the context of a computer programming course. We applied our discrimination-aware method DAAR, which is based on association rules, and also DADT, a discrimination-aware decision tree method, and compared DAAR and DADT with their non-discrimination-aware alternatives. The experiment results showed that both DAAR and DADT decreased the discimination with minor impact on the predictive accuracy. Both classifiers generated a small set of rules that are easy to understand and use by teachers and students. The discrimination-aware classifiers can be used for any classification tasks in educational settings, such as identifying students at risk, to provide timely feedback and intervention.

6. REFERENCES

[1] Calders, T. and Verwer, S., 2010. Three naive Bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery 21*, 2, 277-292.

[2] Pedreshi, D., Ruggieri, S., and Turini, F., 2008. Discrimination-aware data mining. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* ACM, 560-568.

[3] Kamiran, F., Calders, T., and Pechenizkiy, M., 2010. Discrimination aware decision tree learning. In *Proceedings of the 10th IEEE International Conference on Data Mining* IEEE, 869-874.

[4] Luo, L., Liu, W., Koprinska, I., and Chen, F., 2015. Discrimination-Aware Association Rule Mining for Unbiased Data Analytics. *TR700*, School of Information Technologies, The University of Sydney.

[5] Romero, C., Ventura, S., Espejo, P.G., and Hervás, C., 2008. Data Mining Algorithms to Classify Students. In *Proceedings of the International Conference on Educational Data Mining (EDM)*, 8-17.

[6] Romero, C., López, M.-I., Luna, J.-M., and Ventura, S., 2013. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education 68*, 458-472.

[7] Kotsiantis, S., Patriarcheas, K., and Xenos, M., 2010. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems 23*, 6, 529-535.

[8] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., and Loumos, V., 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education 53*, 3, 950-965.

[9] Kamiran, F. and Calders, T., 2009. Classifying without discriminating. In *International Conference on Computer, Control and Communication* IEEE, 1-6.

[10] Ma, Y., Liu, B., and Yiming, W.H., 1998. Integrating classification and association rule mining. In *Proceedings of the 4th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 80-86.