Division of Computing, Engineering and Mathematical Sciences
Department of Computing, University of Kent *in Canterbury*

Fairness-Aware Classification Algorithms

Written By

Lukasz R G Tomaszewski

lrgt2@kent.ac.uk

Computer Science (Artificial Intelligence) MSc

Module CO832: Data Mining and Knowledge Discovery

LaTeX Word Count: 905

March 2022

Contents

1 Background Research

1.1 Two Naïves Bayes Models

Fairness-aware algorithms are designed to analyze data without being too bias on the result. The model algorithm described in [1] is altered to lose dependence on A_s so that the data-set will be processed differently without losing accuracy in the number of attributes. Normally the data-set will process S from A_s through the data-set and produce the result S, to lose the results dependence on A_s the model splits into two (M_+ and M_-). M_+ takes the favour values of S_+ whereas M_- takes the discriminated values S_- to which the final classifier will be dependent on S which is shown in fig. 1 where All the values in the tree (array) are associated with S, this is applicable for both M_+ and M_- as both share an identical Naïve Bayes model, the 2 Naïve models related to the M_+ and M_- models.

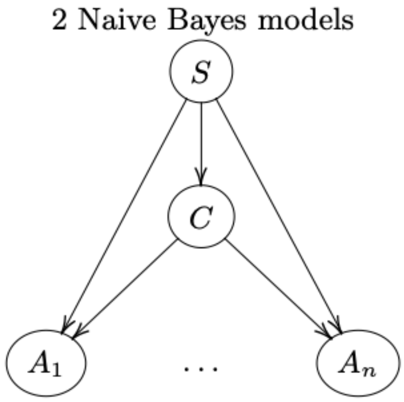

Figure 1: Illustration of the 2 Naïves Bayes models where S connects to each attribute. [1]

In this model specifically as stated takes all the discriminated values and assigns them to the M_- model based off of the S_- values. The algorithms in eq. (1) is a base Naïve model algorithm which depends on a class C, instead of this class its modified into eq. (2) where the S values are used. The S_- discriminated values are used in eq. (2) that removes this discrimination from the Naïve Bayes algorithm.

$$P(C, S, A1, ..., An) = P(C)P(S|C)P(A1|C)...P(An|C) \tag{1}$$

$$P(C, S, A1, ..., An) = P(S)P(C|S)P(A1|C)...P(An|C) \tag{2}$$

1.2 Decision Tree Algorithms

In the second paper [2], a discrimination-aware tree was proposed by "adapting" the splitting criterion for the tree's branches, however multiple adaptations were formed and tested:

$$IGC = H_{Class}(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} H_{Class}(D_i) \tag{3}$$

$$IGS = H_B(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} H_B(D_i) \tag{4}$$

A formula is used to calculate the information gain (IGC) in eq. (3) the data is split into D_i per individual data point under evaluation and H_{Class} is the entropy of the class, so as the tree continues to split, each split of (D_i) is based off of the splitting criterion. However a second formula is conceived eq. (4) to state the influence of gain (labelled as B) of the individual splits of the trees branches labeled as the influence gain (IGS), "the gain sensitivity to B" [2] where H_B is the entropy of B.

From these two formulae, three other formulae are formed (IGC-IGS, IGC/IGS and IGC+IGS). All three equations serve a purpose, the first allows the split only if no discrimination, the second is "a trade-off between accuracy and discrimination" [2] and the third is a sum total of the gains. Another method in this paper associated with the decision trees algorithm is relabelling, "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." [2], the relabelling method would later show that in terms of the IGC, IGC-IGS, IGC/IGS and IGC+IGS with relabelling applied, all showed increased accuracy and slight improvement to discrimination.

2 Discussion

Examining the 2 Naïve Bayes models, the model would work well with small data as it can only split the data set into favoured values and discriminated values however the values in M_- contain all discriminated values, it doesn't specify the types of discrimination i.e. age, gender, race.... however it can save run time by not having to filter through all these types or classes of discrimination. An interesting negative to evaluated is that the base Naïve Bayes algorithm doesn't have a dependence of A on S but in this case it must as the data is split and balanced to form a total of S. The results of testing of the 2 Naïve Bayes in [1] proved that the it "it achieves high accuracy scores with zero discrimination, and has the smallest dependency on S." [1]. The utilisation of splitting the data and learning two separate models showed insightful as S was calculated not from C but from the attributes A_i.

For the algorithm trees, a formula must be chosen from the five formulae to which truly represents the data. When dealing with large amounts of data, trees that implement a simple splitting criteria (IGC) offered a lack of improvement to accuracy, the IGC+IGS only improves the system when relying on an external variable, this was in the form on a relabelling system in [2]. The decision tree algorithm stated in the results that "From the results of our experiments we draw the fol- lowing 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]. This states that this method is effective when the training data is discriminatory and the test data is non-discriminatory, is shows with the high accuracy and low discrimination scores implying that this algorithm suites large data sets that are pre-complied into bias data-sets.

3 Conclusion

Though both the fairness-aware algorithms are similar in the output goal, in which they process data into non discriminatory. However they are vastly different with the decision tree algorithm allowing for further usage and on larger more complex data-sets than that of the proposed 2 Naïve Bayes algorithm in [1]. Such examples would include that the Naïve Bayes model is useful for gender discrimination whereas the decision tree would look at age as it will be able to create new branches base off of the splitting criterion. Gender seeks only male or female, the data is split into two models as M_+ and M_-, where age seeks not only up to 100 values but also groups of age brackets, to which can be branched off the tree where the entropy of the splitting criterion will justify the next node.

4 References

[1] T. Calders and S. Verwer, *Three naive bayes approaches for discrimination-free classification*, Data Min. Knowl. Discov., 21 (2010), pp. 277–292.

[2] F. Kamiran, T. Calders, and M. Pechenizkiy, *Discrimination aware decision tree learning*, 2010 IEEE International Conference on Data Mining, (2010), pp. 869–874.