

Fairness-aware online learning under class-imbalance

Fairness-aware online learning has become an evolving field during the last few years. Fairness-aware online learning goal is to maintain a classifier that performs well and does not discriminate over the course of the stream. Some initial works have been proposed to tackle discriminatory outcomes from online classification [1, 2]; however, these methods do not take into consideration the uneven class distribution over the course of the stream. If the imbalance problem is not tackled, the learner mainly learns the majority class and strongly misclassifies/rejects the minority. Such methods might appear to be fair for certain fairness definitions that rely on parity in the predictions between the protected and non-protected groups. In reality though the low discrimination scores are just an artifact of the low prediction rates for the minority class.

In this master thesis, we want to investigate the combined problem of class-imbalance and fairness-aware learning in the online setup. We focus on Naive Bayes classifier which has been extensively studied in the context of fairness but in the static setting. In this work, we plan to extend these models to the online setting taking into account the imbalance of the population under different fairness notions such as statistical parity[3], equal opportunity [4], and equalized odds [4].

An ideal candidate should be:

- a self motivated and independent learner
- knowledgeable about machine learning (good grades in Data Mining I, Data Mining II)
- experienced with python or java

Interested students are encouraged to email to Eirini Ntoutsis at ntoutsis@l3s.de and/or Vasileios Iosifidis at iosifidis@l3s.de for scheduling an appointment. CV and transcript of records must be sent beforehand.

References

1. V. Iosifidis, H. Tran, E. Ntoutsis, "Fairness-enhancing interventions in stream classification", 30th International Conference on Databases and Expert Systems Applications (DEXA), 2019.
2. W. Zhang, E. Ntoutsis, "An Adaptive Fairness-aware Decision Tree Classifier", International Joint Conference on Artificial Intelligence (IJCAI), 2019.
3. Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33(1), 1-33.
4. Hardt, M., Price, E. and Srebro, N., 2016. Equality of opportunity in supervised learning. In *Advances in neural information processing systems* (pp. 3315-3323).