Group Project

December 4, 2023

Abstract

1 Introduction

2 Related work

3 Proposed algorithm: Instance Dependent Cost Online Classification

In this section, we design heuristics instance-dependent cost in online learning. To simplify cost matrix, we only consider the misclassification cost for each class(degree of freedom: $k^2 - k$ down to $k - 1$) and the misclassification cost for each instance (degree of freedom: $n \times (k^2 - k)$ down to $n \times (k - 1)$) , where k is number of classes, n is number of instances.

3.1 The original structure of the proposed algorithm

For each prediction, if our classifier predict correctly, we continue to train the classifier by fitting this instance by the class-dependent cost; Otherwise, i.e. the current classifier can not predict this instance correctly, the instance is difficult to the classifier that it can not handle, our current classifier should pay more attention to this instance. We continue to train the classifier by fitting this instance by the heuristics instance-dependent cost.

Note that in algorithm 1 we take the prediction error into consideration when calculating the heuristics instance-dependent cost. The prediction error term focus more on the instance that p nears to 0.5 but is misclassified as the opposite class. For such instances, our current classifier is almost able to predict correctly, so we give them a higher prediction error cost. However, for the instance that p nears to 0 or 1 but is misclassified as the opposite class, there are two possible reasons: 1) our current classifier is still weak 2) the instance is a noise. For such instance, we give them a lower prediction error cost to prevent overfitting or learning from noise. By receiving the feedback from prediction, the model is trained to strengthen the ability to correctly classify samples that are easy to classify incorrectly, thereby improving the overall performance of the classifier step by step.

```
Input: Input data, Input Labels
Output: Prediction
while Have more samples do
    datum, label <- next sample
    prediction label ŷ <- Predict by current classifier
    y <- the true label of this sample
    current ratio of Class y: CRC_y <- (# of currentsamples with label y) / (# of all current samples)
    Class-dependent Cost: CDC_y <- 1 / CRC_y
    if Prediction is correct then
        train the classifier by fitting this instance with the
          class-dependent cost
    else
        prediction error <- αe^(β(1−y(1−p)−(1−y)p))
        Instance-dependent Cost: IDC <- CDC_y + prediction error
        train the classifier by fitting this instance with the heuristics
          instance-dependent cost
    end
end
```

Algorithm 1: Heuristics Instance-dependent Cost Online classification

3.2 Change the structure of the proposed algorithm

We have modified the structure of the original algorithm in two aspects, one is the cost of using feedback based on instance prediction results each timestep, and the other one is the modification in the definition of prediction error term.

For the first aspect, we do not distinguish between correctly classified and incorrectly classified instance, using the same definition of instance-dependent cost as the cost for each instance. Since we found that, if the classifier correctly classifies each instance, is will degenerate into class-dependent cost classifier, which shows no obvious difference between class-dependent cost classifier and instance-dependent cost classifier.

For the second aspect, the prediction error is redefined as:

$$\text{prediction error} = (1 - p_t)^{\alpha}, \tag{1}$$

where p_t is the classification probability as the true label, defined as:

$$p_t = \begin{cases} p, \text{ if y=1} \\ 1 - p, \text{ if y=0} \end{cases} \tag{2}$$

In the above y $\in \{0, 1\}$ specifies the ground-truth class and p $\in$ [0, 1] is the model's estimated probability for the class with label y = 1.

Then the instance-dependent cost is redefined as

$$\text{instance-dependent cost} = (1 + \text{prediction error}) \times \text{class-dependent cost} \quad (3)$$

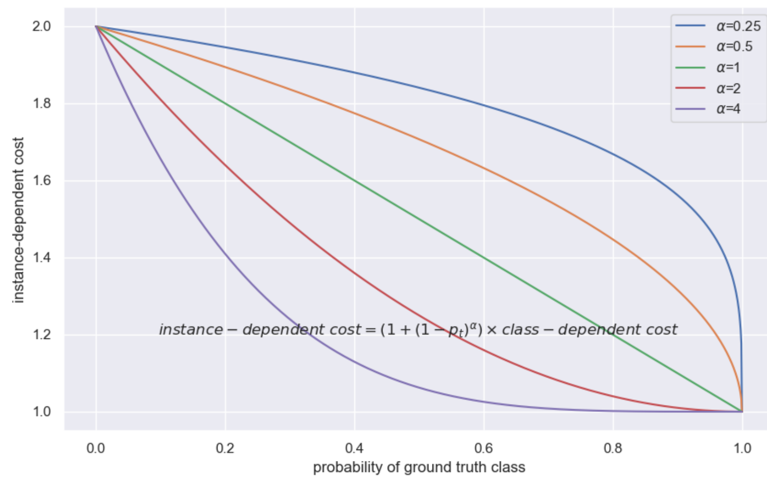

Figure 1: instance-dependent cost with different value of α

The instance-dependent cost is visualized for several values of $\alpha \in [0.25, 4]$ in Figure 1. We note two properties of the instance-dependent cost. (1) When an instance is misclassified and p_t is small, the prediction error term nears to 1, which contributes a large instance-dependent cost. As p_t nears to 1, the prediction error goes to 0 and the instance-dependent cost for well-classified instance is down-weighted to class-dependent cost. (2) The hypothesis parameter α smoothly adjusts the rate at which hard instances are focused on more.

4 Experiments

5 Conclusion

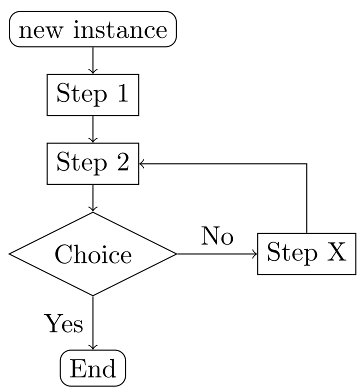

Figure 2: Flow chart