

Abstract

In supervised learning - for instance in image classification - modern massive datasets are commonly labeled by a crowd of workers. The obtained labels in this crowdsourcing setting are then aggregated for training. The aggregation step generally leverages a per-worker trust score. Yet, such worker-centric approaches discard each task's ambiguity. Some intrinsically ambiguous tasks might even fool expert workers, which could eventually be harmful to the learning step. In a standard supervised learning setting - with one label per task - the Area Under the Margin (AUM) is tailored to identify mislabeled data. We adapt the AUM to identify ambiguous tasks in crowdsourced learning scenarios, introducing the Weighted AUM (WAUM). The WAUM is an average of AUMs weighted by task-dependent scores. We show that the WAUM can help discard ambiguous tasks from the training set, leading to better generalization or calibration performance.

Keywords: crowdsourcing, task ambiguity, data cleaning