

COMPUTER SCIENCE
DEFENSE

Honors Defense

Yilin Dong

Abstract: Classification algorithms build models that can classify new observations. While they require a training set of samples' features and labels for training, in reality, many unstructured datasets do not meet the requirement. Since having experts to give out manual labels has a high cost, many industries adopted crowdsourcing, which enables a group of people to contribute to the same labeling task. However, multiple annotations cannot apply to classification algorithms because they assume that labels are single and consensus. In this paper, we use truth inference methods to estimate single labels given different annotations from multiple annotators. While the Expectation-Maximization method provides the best accuracy, our empirical results suggest that better predictive performance can be achieved by accounting for disagreements. Thus, we propose Medaboost, a new predictive model, that considers the degree of disagreements between annotators to improve predictive performance. Medaboost outperforms AdaBoost on both synthetic dataset and MIMIC-III dataset under different sets of simulated nurses.

Monday, March 30, 2020, 3:00 pm
<https://emory.zoom.us/j/785410306>

COMPUTER SCIENCE
EMORY UNIVERSITY