World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:15, No:12, 2021

Semi-Supervised Learning Using Pseudo F Measure

Authors : Mahesh Balan U, Rohith Srinivaas Mohanakrishnan, Venkat Subramanian

Abstract : Positive and unlabeled learning (PU) has gained more attention in both academic and industry research literature recently because of its relevance to existing business problems today. Yet, there still seems to be some existing challenges in terms of validating the performance of PU learning, as the actual truth of unlabeled data points is still unknown in contrast to a binary classification where we know the truth. In this study, we propose a novel PU learning technique based on the Pseudo-F measure, where we address this research gap. In this approach, we train the PU model to discriminate the probability distribution of the positive and unlabeled in the validation and spy data. The predicted probabilities of the PU model have a two-fold validation – (a) the predicted probabilities of reliable positives and predicted positives should be from the same distribution; (b) the predicted probabilities of predicted positives and predicted unlabeled should be from a different distribution. We experimented with this approach on a credit marketing case study in one of the world's biggest fintech platforms and found evidence for benchmarking performance and backtested using historical data. This study contributes to the existing literature on semi-supervised learning.

Keywords : PU learning, semi-supervised learning, pseudo f measure, classification
Conference Title : ICDSML 2021 : International Conference on Data Science and Machine Learning
Conference Location : Tokyo, Japan
Conference Dates : December 02-03, 2021