



# Crowd Learning with Candidate Labeling: an EM-based Solution\*

\***Note:** The full contents of this paper have been published in the volume *Lecture Notes in Artificial Intelligence 11160* (LNAI 11160)

Iker Beñaran-Muñoz  
Basque Center for Applied Mathematics  
Bilbao, Spain  
ibenaran@bcamath.org

Jerónimo Hernández-González  
University of the Basque Country  
UPV/EHU  
Donostia, Spain  
jeronimo.hernandez@ehu.eus

Aritz Pérez  
Basque Center for Applied Mathematics  
Bilbao, Spain  
aperez@bcamath.org

**Abstract**—Crowdsourcing is widely used nowadays in machine learning for data labeling. Although in the traditional case annotators are asked to provide a single label for each instance, novel approaches allow annotators, in case of doubt, to choose a subset of labels as a way to extract more information from them. In both the traditional and these novel approaches, the reliability of the labelers can be modeled based on the collections of labels that they provide. In this paper, we propose an Expectation-Maximization-based method for crowdsourced data with candidate sets. Iteratively the likelihood of the parameters that model the reliability of the labelers is maximized, while the ground truth is estimated. The experimental results suggest that the proposed method performs better than the baseline aggregation schemes in terms of estimated accuracy.

**Index Terms**—Supervised classification, Crowdsourced labels, Weak supervision, Candidate labeling, Expectation-Maximization based method