Machine Learning for Quality Aware Task Assignment in Crowdsourcing

Naman GOEL
(naman.goel@epfl.ch)

Crowdsourcing is a popular method of collecting labels for supervised machine learning algorithms. But the labels obtained from workers on crowdsourcing platforms are not guaranteed to be correct. Various complementary approaches exist in literature for improving the quality of data obtained from crowdsourcing. Some of them include pre-screening workers, providing monetary incentives, using aggregated response of multiple workers and cleaning the data post collection.

One particular method is to dynamically assign tasks to workers based on their proficiencies. The proficiencies of the workers are not known in advance and there is an obvious tradeoff between inferring a worker's proficiency by assigning more tasks vs getting high quality job done in minimum budget. The goal of this project is to first implement various (at least 2) algorithms for doing so, identify clearly the difference in settings of the algorithms and critically compare their strengths and weaknesses using extensive simulations. The next part of the project will be aimed at extending one of the algorithms **[1]** and improve its performance in obtaining high quality data by including the possibility of educating the workers.

References :

[1] Bragg, Jonathan, and Daniel S. Weld. "Optimal testing for crowd workers." *AAMAS*, 2016.
[2] Ho, Chien-Ju, Shahin Jabbari, and Jennifer W. Vaughan. "Adaptive task assignment for crowdsourced classification." *ICML*. 2013.
[3] Ho, Chien-Ju, and Jennifer Wortman Vaughan. "Online Task Assignment in Crowdsourcing Markets." *AAAI*. Vol. 12. 2012.
[4] Jain, Shweta, et al. "An incentive compatible multi-armed-bandit crowdsourcing mechanism with quality assurance." *arXiv preprint arXiv:1406.7157* (2014).
[5] Bhat, Satyanath, et al. "A Truthful Mechanism with Biparameter Learning for Online Crowdsourcing." *AAMAS*, 2016.
[6] Oleson, David, et al. "Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing." *Human computation* 11.11 (2011).