Crowdsourcing Multi-Label Classification

Jonathan Bragg **Mausam** **Daniel S. Weld**
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195
{jbragg, mausam, weld}@cs.washington.edu

Abstract

Machine learning and decision theoretic techniques offer the potential for dramatically reducing the amount of human labor required in crowdsourcing applications. A large amount of recent work has addressed solving "consensus tasks" such as answering multiple-choice questions using responses from noisy workers (Dai et al. 2013; Kamar, Hacker, and Horvitz 2012; Raykar et al. 2010; Sheng, Provost, and Ipeirotis 2008; Wauthier and Jordan 2011; Welinder et al. 2010; Whitehill et al. 2009). However, another important class of problems that has not been studied in this setting is multi-label classification tasks, which require matching dataset items to subsets of labels. A straightforward way to solve these problems is to treat each item-label pair as a binary consensus task. Since consensus tasks are frequently solved by threshold voting (which includes majority voting as a special case), we first present easy-to-implement improvements to these methods that reduce the amount of labor required with little or no loss in performance. Next, we present a decision-theoretic approach with two components: (1) a noisy worker model that allows a worker to be probabilistically accurate and estimates the true value of item-label relationships based on probabilistic inference, and (2) a controller that chooses, for each item, which questions provide the maximum value of information toward a joint categorization. We demonstrate that an inference model that exploits learned dependencies between labels achieves statistically significant improvement over models that assume labels are independent. Our results extend to the practical situation where one is interested in generating a worker task which includes a batch of related questions (labels), as is common practice in online labor marketplaces.

Multi-label classification has many applications, one of which is as a subroutine in the CASCADE taxonomy creation algorithm (Chilton et al. 2013). While CASCADE needs only unskilled labor and produces taxonomies whose quality approaches that of human experts, it uses significantly more labor than experts, largely concentrated in solving a multi-label classification task. Substituting our optimized classification approach, therefore, offers potential savings. We present live experiments on Amazon Mechanical Turk showing that our best combination of policies matches the accuracies achieved by CASCADE's threshold voting method using less than 10% as much labor.

Keywords. Human computation; multi-label classification; machine learning; decision-theoretic optimization

Note. This abstract summarizes (Bragg, Mausam, and Weld 2013), to appear at HCOMP 2013.

References

Bragg, J.; Mausam; and Weld, D. S. 2013. Crowdsourcing multi-label classification for taxonomy creation. In *Proceedings of the First Conference on Human Computation and Crowdsourcing (HCOMP '13). To appear.*

Chilton, L. B.; Little, G.; Edge, D.; Weld, D. S.; and Landay, J. A. 2013. Cascade: Crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*.

Dai, P.; Lin, C. H.; Mausam; and Weld, D. S. 2013. POMDP-based control of workflows for crowdsourcing. *Artificial Intelligence* 202(0):52–85.

Kamar, E.; Hacker, S.; and Horvitz, E. 2012. Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS '12)*.

Raykar, V. C.; Yu, S.; Zhao, L. H.; and Valadez, G. 2010. Learning from crowds. *Journal of Machine Learning Research* 11:1297–1322.

Sheng, V. S.; Provost, F.; and Ipeirotis, P. G. 2008. Get another label? Improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*.

Wauthier, F., and Jordan, M. 2011. Bayesian bias mitigation for crowdsourcing. In *Advances in Neural Information Processing Systems 24 (NIPS '11)*. 1800–1808.

Welinder, P.; Branson, S.; Belongie, S.; and Perona, P. 2010. The multidimensional wisdom of crowds. In *Advances in Neural Information Processing Systems 23 (NIPS '10)*. 2424–2432.

Whitehill, J.; Ruvolo, P.; Wu, T.; Bergsma, J.; and Movellan, J. 2009. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems 22 (NIPS '09)*. 2035–2043.