

Abstract

Crowdsourcing has emerged as a convenient mechanism to collect human judgments on a variety of tasks, ranging from document and image classification to scientific experimentation. However, in recent times crowdsourcing has evolved from solving simpler tasks, like recognizing objects in images, to more complex tasks such as collaborative journalism, language translation, product designing etc. Unlike simpler micro-tasks performed by a single worker, these complex tasks require a group of workers and greater resources. In such scenarios, where groups of participants are the atomic units, it is a non-trivial task to distinguish workers (who contribute positively) from idlers (who do not contribute to group task) among the participants using only group's performance. The first part of this thesis studies the problem of distinguishing workers from idlers, without assuming any prior knowledge of individual skills and considers "groups" as the smallest observable unit for evaluation. We draw upon literature from group-testing and give bounds over minimum number of groups required to identify quality of subsets of individuals with high confidence. We validate our theory experimentally and report insights for the number of workers and idlers that can be identified for a given number of group-tasks with significant probability.

In most crowdsourcing applications, there exist dependencies among the pool of Human Intelligence Tasks (HITs) and often in practical scenarios there are far too many HITs available than what can realistically be covered by limited available budget. Estimating the accuracy of automatically constructed Knowledge Graphs (KG) is one such important application. Automatic construction of large knowledge graphs has gained wide popularity in recent times. These KGs, such as NELL, Google Knowledge Vault, etc., consist of thousands of predicate-relations (e.g., is Person, is Mayor Of) and millions of their instances (e.g., (Bill de Blasio, is Mayor Of, New York City)). Estimating accuracy of such KGs is a challenging problem due to their size and diversity. In the second part of this study, we show that standard single-task crowdsourcing is sub-optimal and very expensive as it ignores dependencies among various predicates and instances. We propose Relational Crowdsourcing (RelCrowd) to overcome this challenge, where the tasks are created while taking dependencies among predicates and instances into account. We apply this framework in the context of large-scale Knowledge Graph Evaluation (KGEval) and demonstrate its effectiveness through extensive experiments on real-world datasets.