Content-aware Crowdsourcing Vote Aggregation

Piyush Bansal
International Institute
of Information Technology
Hyderabad, India
piyush.bansal@research.iiit.ac.in

Andrey Bârsan, Martin Davtyan,
Carsten Eickhoff
Dept. of Computer Science
ETH Zurich, Switzerland
{firstname.lastname}@inf.ethz.ch

Crowdsourcing has been able to leverage the potential of a globally distributed and diverse workforce to efficiently create and enrich academic datasets. This has led to a noticeable decrease in the overall time, and monetary cost involved in corpus creation. While, traditionally, a group of domain experts were employed for a task, crowdsourcing involves multiple workers who may not hold significant domain expertise. The recurring challenge of quality control (QC), which arises due to an untrained workforce. While most workers attempt to truthfully complete tasks, there are frequent reports of workers showing sloppy or random judging behaviour in order to increase their time efficiency [5]. A widely accepted way of overcoming this obstacle is to present the same HIT to multiple workers and subsequently aggregate their submissions, *e.g.* , in the form of unbiased majority votes. For many tasks, including document relevance assessment, this practice has effects that go beyond mere filtering of spam submissions but can also account for subjective differences in judgments across workers.

Instead of uniformly merging raw votes, much work has been dedicated into estimating worker reliability based on their past accuracy, judging behaviour, or topic affinity [6]. Subsequently, this form of worker information can be used to bias the aggregation process towards the most reliable workers or to empower active learning schemes in which the most suitable worker for each task is to be selected [4]. There is, however, another, largely untapped source of information, the document's content. For example, one could exploit similarities between documents to aggregate worker votes in an efficient manner.

Consider a cosine similarity between the tf–idf representations of documents. Figure 1 shows the distribution of similarities a) between relevant documents (inner similarities) and b) between relevant and irrelevant documents (outer similarities) across all topics of the TREC Crowdsourcing Track 2011 [7]. We can clearly observe how relevant documents share much stronger commonalities than irrelevant ones. In this paper, we exploit this well-known *Clustering Hypothesis* by propagating relevance assessments to "*nearby*" neighbours for the purpose of vote aggregation.

To this end, we propose a Gaussian Process model that spatially smoothes relevance votes in a space induced by document similarity information. On the basis of this content-aware vote aggregation method, we devise two alternative active learning schemes that address highly budget-constrained crowdsourcing scenarios in which only a subset of all possible query document pairs can be explicitly voted on by workers. While the first method is based on a graph-theoretical information diffusion model, the second approach relies on information theoretic estimates of label variance and mutual information. A series of experiments on historic TREC data as well as live crowdsourcing experiments confirms the competitive performance of the proposed methods.

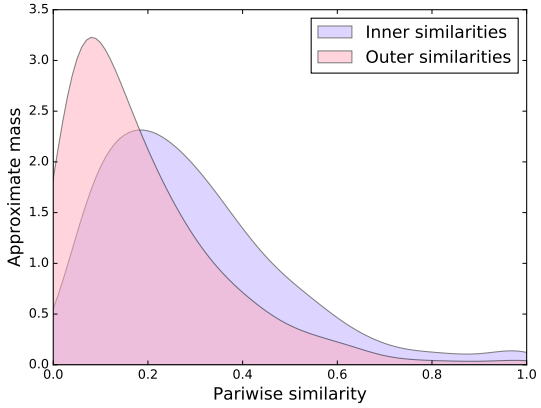

Figure 1: Kernel density estimate of the inner and outer similarities of the documents aggregated across all TREC 2011 Crowdsourcing Track topics.

The line of work summarized in this talk originally appeared in a series of accepted full paper submissions to CIKM 2015 [3], CIKM 2016 [1] and a current submission to WSDM 2017 [2].

1. REFERENCES

[1] Piyush Bansal, Carsten Eickhoff, and Thomas Hofmann. Active content-based crowdsourcing task selection. In *CIKM 2016*. ACM.

[2] Andrey Bârsan and Carsten Eickhoff. An information diffusion model of crowdsourcing vote aggregation and sampling. In *WSDM 2017*. ACM.

[3] Martin Davtyan, Carsten Eickhoff, and Thomas Hofmann. Exploiting document content for efficient aggregation of crowdsourcing votes. In *CIKM 2015*. ACM.

[4] Djellel Eddine Difallah, Gianluca Demartini, and Philippe Cudré-Mauroux. Pick-a-crowd: tell me what you like, and i'll tell you what to do. In *WWW 2013*. ACM.

[5] Carsten Eickhoff and Arjen P de Vries. Increasing cheat robustness of crowdsourcing tasks. *Information retrieval*, 16(2), 2013.

[6] Gabriella Kazai, Jaap Kamps, and Natasa Milic-Frayling. Worker types and personality traits in crowdsourcing relevance labels. In *CIKM 2011*. ACM.

[7] Matthew Lease and Gabriella Kazai. Overview of the trec 2011 crowdsourcing track. In *TREC 2011*.