Evaluating Semantic Search - Improving Crowdsourcing through Malicious Activity Detection

Ujwal Gadiraju

L3S Research Center, Leibniz Universität Hannover, Germany
{gadiraju}@L3S.de

Host : Dr. Gianluca Demartini
Information School,
University of Sheffield,
United Kingdom

Period : 01.08.2015 - 30.08.2015

1 Purpose of the STSM

Evaluating methods and algorithms in the semantic search and information retrieval fields requires reliable ground truths, which are costly to generate. In this context, crowdsourcing has gained significant popularity through the last decade, to aid various tasks pertaining to structured data sources. Researchers and practitioners have turned to crowdsourcing in order to generate ground truths for evaluation, validate quality of knowledge bases, and for relevance assessment in semantic search or ad-hoc object retrieval scenarios. However, one of the main concerns about crowdsourced data and results is the concomitant quality and reliability. Crowdsourcing microtasks are often plagued by malicious activity that is prevalent in the crowd as we have shown in our recent work [1]. In this STSM, we aim to address the following research questions.

- **RQ#1**: How can task requesters in crowdsourcing platforms, benefit from the knowledge of worker types?
- **RQ#2**: Can worker behavioral traces be used to classify workers automatically into distinct types?
- **RQ#3**: What is the impact of *task complexity* and *microtask type* on the behavior of crowd workers?

2 Description of the work carried out during the STSM

2.1 Crowd Worker Typology

We typecast crowd workers into distinct categories based on the following dimensions: (i) worker behavior , (ii) quality of work produced, and (iii) worker motivation to participate in crowdsourced tasks. The behavior of workers in crowdsourced tasks has been

priorly studied [1, 2]. We present a holistic approach towards an elaborate worker typology, by building on these works. Crowd worker behavior is influenced by several aspects, some of which are inherent to the worker (such as trustworthiness of a worker) and others that are induced by the nature of tasks (such as *task complexity*). Workers can be categorized based on the quality of their work. Some categories proposed by [1] and [2] include elite workers (those who perform with an accuracy of 100%), competent workers, incompetent workers, and so forth.

2.2 Modeling Task Complexity

Task complexity in the crowdsourcing paradigm is a complicated aspect that depends on several factors. One can model task complexity from a worker's point of view, where worker competence for example, could play a role in determining how complex a given task is. This is logically sound, since one worker can find a given task to be difficult while another can find the same task to be simple. However, including inherent worker traits in task complexity modeling would make it subjective. In order to define task complexity from a purely objective standpoint, we consider the characteristics of the task alone. Herein, we model task complexity as a function of (i) the objective difficulty-level of task and (ii) the length of the task.

2.3 Experimental Setup

We deployed 9 tasks of the content creation type, with varying combinations of length (20, 30, 40 units) and objective difficulty-levels. For each task, we gathered responses from 100 distinct workers. We deployed tasks of the same difficulty-level concurrently, in order to avoid potential learning biases.

2.4 Crowd Worker Behavioral Traces

We implemented JavaScript based mouse-tracking for tracking and logging worker activity in crowdsourced tasks. Through these behavioral traces we identify different worker traits. We took additional measures to distinguish between workers that use a mouse and those who use a touchpad. We also distinguish between worker mannerisms with respect to scrolling behavior; use of scrollbar as opposed to the mousewheel. In this way, we gathered worker activity data from each of the experimental tasks deployed on CrowdFlower.

2.5 Predicting Worker Accuracy Based on Behavioral Traces

Using features designed on top of the raw data collected during the work performed on HITs published on the crowdsourcing platforms, it is possible to predict work accuracy. In this section we present the results of applying machine learning models to predict three levels of worker quality. Using the following high-level behavioral features it is possible to train a model that classify workers into quality levels: Speed (i.e., time to transcribe one image), Tab keystrokes, window focus changes, copy and paste frequency, backspace keystrokes, right and left click counts, scroll up and down counts, times the mouse cursor enters and leaves the task window. Performing a 10-fold cross

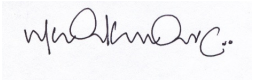
validation of a Random Forest model over 900 workers who participated in our study it is possible to predict their final accuracy level using a C4.5 decision tree model with an accuracy of 71%. Feature selection performed by information gain shows that the most indicative features to predict worker quality level are number of left mouse clicks, scroll down activity, and backspace keystrokes (indicating special attention to errors).

3 Main Results and Future Collaboration

- We have proposed a two-level worker typology based on high-level dimensions as well as low-level behavioral traces.
- We are currently working on a full-paper submission to the top-tier conference, WWW 2016 in Montreal, Canada.
- Our fruitful collaboration in this STSM has led to several research ideas that we plan to work on following the completion of this work.

References

1. U. Gadiraju, R. Kawase, S. Dietze, and G. Demartini. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Republic of Korea, April 18-23, 2015*, pages 1631–1640, 2015.
2. G. Kazai, J. Kamps, and N. Milic-Frayling. Worker types and personality traits in crowdsourcing relevance labels. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1941–1944. ACM, 2011.



22.08.2015