

Access Path Design for Quality Assurance in Crowdsourcing

Master's Thesis Proposal

Problem Description

Crowdsourcing is a problem-solving approach that leverages the power of human computation in tackling problems that are difficult to solve with machines only. This paradigm has been successfully applied in various applications like Wikipedia, Duolingo, Stackoverflow *etc.* A central challenge in crowdsourcing is ensuring the quality of results when the crowd answers contain possible errors and uncertainty. The most common approach to this problem is to increase quality by applying *crowdsourcing redundancy*, *i.e.* asking the same question to many different individuals. Nevertheless, redundancy can be many times useless and expensive if it is not planned beforehand. For example, if we ask the same question to people who have similar opinions or that exhibit certain dependencies, the quality of results will not significantly increase.

In this context, we define the notion of *access path* as a possible but independent way of retrieving the same piece of information from the crowd. The configuration of access paths can be based on various criteria depending on the task: (i) workers' demographics (*e.g.* profession, group of interest, age) (ii) the source of information or the tool that is used to find the answer (*e.g.* phone call vs. web page, Bing vs. Google) (iii) task design (*e.g.* time of completion, user interface) (iv) task decomposition (*e.g.* part of the answers, features). Combining questions from diverse access paths provides a better quality than aggregating answers from a single access path. However, it is not always clear which is the best configuration / grouping to design access paths for a given problem. Moreover, such a structure may be hidden and has to be discovered from historical data. These are the research questions that we want to study in this Master Thesis. In summary, the goals of the project are:

- Exploring strategies for testing the efficiency of possible access path configurations.
- Discovering hidden access path structures from historical crowdsourced datasets.
- Empirically evaluating these approaches through realistic experiments in crowdsourcing platforms (*e.g.* Amazon Mechanical Turk) or online social networks (*e.g.* Twitter, Quora, Stackoverflow).

A potential candidate for this project should have some basic knowledge in statistics, machine learning and graphical models, as well as programming.

If you are interested in this project and want to discuss further, please contact Besmira Nushi (nushib@inf.ethz.ch).

The proposed master thesis will be supervised by Prof. Donald Kossmann.

