

ERICA: Expert Guidance in Validating Crowd Answers

Nguyen Quoc Viet Hung, Duong Chi Thang, Matthias Weidlich[†], Karl Aberer
École Polytechnique Fédérale de Lausanne and [†]Humboldt-Universität zu Berlin

ABSTRACT

Crowdsourcing became an essential tool for a broad range of Web applications. Yet, the wide-ranging levels of expertise of crowd workers as well as the presence of faulty workers call for quality control of the crowdsourcing result. To this end, many crowdsourcing platforms feature a post-processing phase, in which crowd answers are validated by experts. This approach incurs high costs though, since expert input is a scarce resource. To support the expert in the validation process, we present a tool for *ExpeRt guidance In validating Crowd Answers (ERICA)*. It allows us to guide the expert's work by collecting input on the most problematic cases, thereby achieving a set of high quality answers even if the expert does not validate the complete answer set. The tool also supports the task requester in selecting the most cost-efficient allocation of the budget between the expert and the crowd.

Categories and Subject Descriptors: H.1.2 [User/Machine Systems]: Human information processing

Keywords: crowdsourcing; validation; guiding user feedback

1. INTRODUCTION

Crowdsourcing has been established as an efficient and scalable approach to overcome problems that are computationally expensive or unsolvable for machines, but rather trivial for humans [6]. In essence, crowdsourcing enables users to post tasks in the form of questions, which are then answered by crowd workers for financial rewards. Crowd workers, however, have different backgrounds and wide-ranging levels of expertise and motivation [3], so that the obtained answers are not necessarily correct. Even aggregation of answers from multiple workers to a single question cannot guarantee result correctness. Platforms for crowdsourcing, such as Amazon Mechanical Turk (AMT), aim at including the answer trustworthiness by including a validation phase. That is, the obtained crowd answer are checked against expert answers that are supposedly correct. Involving an expert incurs high costs, though. Hence, only a fraction of the obtained crowd answers can be considered in the validation phase, which raises the question of how to involve experts effectively and efficiently.

This demo presents ERICA: a tool for *ExpeRt guidance In validating Crowd Answers*. Its functionalities are summarized as follows:

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author(s). Copyright is held by the owner/author(s).

SIGIR'15, August 09-13, 2015, Santiago, Chile.

ACM 978-1-4503-3621-5/15/08.

<http://dx.doi.org/10.1145/2766462.2767866>.

- (1) *Minimizing expert efforts:* ERICA guides experts during the whole validation process. By exploiting the fact that some validations are more beneficial than others, it reduces the cost of expert validation.
- (2) *Handling faulty workers:* Faulty workers, such as spammers or sloppy workers, can increase the cost of acquiring correct answers. By detecting them, ERICA improves the quality of the crowdsourcing result.
- (3) *Estimating quality:* During the validation process, ERICA maintains an *aggregated answers set* which are the answers considered to be correct together with various statistics related to this set. These statistics support the expert in deciding when to end the validation.
- (4) *Optimizing overall cost:* ERICA enables crowdsourcing users to determine the best budget allocation between a validating expert and the crowd w.r.t a budget and a crowdsourcing setup.
- (5) *Detecting expert mistakes:* In practice, expert input may contain mistakes that are not caused by the lack of knowledge, but stem from the interaction as part of the validation [5]. ERICA supports the detection of such mistakes, so that they can be fixed by the expert themselves.

To the best of our knowledge, ERICA is the first system to provide such comprehensive support for expert-based validation of crowdsourcing results, thereby providing a valuable add-on to existing crowdsourcing services. In contrast to related approaches to the detection of faulty workers, quality estimation or cost optimization, all our techniques are grounded in expert feedback, which results in higher quality and better performance [5]. In the remainder, we first discuss the implementation of these functionalities in Section 2. Next, Section 3 demonstrates the ERICA by means of an example scenario, before Section 4 closes the paper with a summary.

2. IMPLEMENTATION

This section discusses how ERICA handles large amounts of crowd answers and implements the above functionalities. The underlying theory of the respective techniques is discussed in detail in our earlier work [5].

Functionality

Minimizing expert efforts. To guide the expert validation, ERICA selects and ranks questions for which expert feedback should be sought based on their expected benefit. The expected benefit of a validation question is measured using an information theoretic model. Then, the question with the highest information gain is shown first for expert feedback elicitation [5].

Handling faulty workers. To detect faulty workers, ERICA computes the likelihood of a worker to be spammer (*spammer score*) [5]. To this end, confusion matrices are constructed based on the ex-

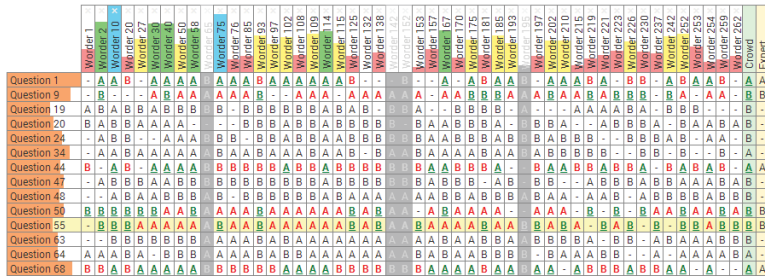


Figure 1: An answer matrix

pert feedback obtained so far. Answers from workers with a high likelihood to be a spammer are removed by ERICA.

Estimating quality. The quality of the aggregated answers is measured using k-fold cross validation [4]. That is, ERICA arbitrarily partitions the expert feedback into test and training sets. Then, the training set is used to compute the aggregated answers for the test set and the similarity between two aggregated answers sets (one created using only the training set, one using the whole feedback set) is measured. Repeating this procedure k times, ERICA estimates the quality of the aggregated answers.

Optimizing cost. ERICA helps to find the optimal distribution of a pre-defined budget between an expert and the crowd based on characteristics of the crowd. Since these characteristics are not known beforehand, it samples the crowd with a few test questions and simulates the crowd to decide on the best budget distribution [5].

Detecting expert mistakes. ERICA also uses k-fold cross validation to detect expert mistakes. Here, k is fixed to 1, though, in order to detect incorrect feedback. Then, the aggregated answer for the question in the test set is compared with the feedback by the expert. If they are different, the expert is encouraged to revisit the respective question and reconsider the answer.

Scalability

To achieve an efficient implementation for large-scale data, ERICA employs two techniques: *parallelization* and *partitioning*. First, since the computations of the information gain and the spammer score are independent for different questions, they are executed in parallel concurrently for all questions. Second, common answer matrices are sparse [1]. Therefore, ERICA uses sparse matrix partitioning [2] to divide a large answer matrix into smaller dense ones that fit for human interactions and can be handled more efficiently.

Technicality

ERICA is implemented as a web service. The back-end is written in Python and Java. The front-end is an HTML5 website communicating with the server using Javascript and jQuery.

3. DEMONSTRATION

To demonstrate the application of ERICA, we first present its user interface before we turn to an example scenario.

User interface. ERICA offers a rich user interface (see Figure 1,2), which consists of 2 main components:

- *Answer matrix:* contains all the questions and crowd answers. Each row is a question, each column is a worker and a cell denotes the respective answer. The workers and the questions are color-coded to reflect their reliability and difficulty, respectively. The aggregated answers, considered to be correct according to the crowd answers and the expert feedback, are shown in a separate column.
- *Expert feedback view:* contains three panels that present additional information during the validation. Panel A shows

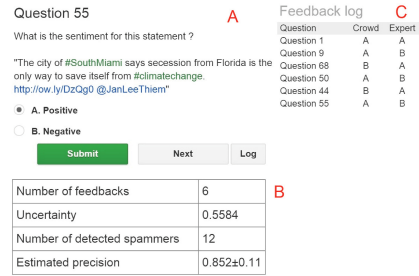


Figure 2: Expert feedback view

the content of the currently-selected question that requires expert input. Various statistics are provided in Panel B for the expert’s consideration. Further, the history of expert feedback is visualized in Panel C.

Demonstration scenario. We use the example from Figure 1 to demonstrate how to use ERICA to analyze and validate the answer matrix (a screencast of the demonstration is publicly available ¹).

Before posting a task to a crowdsourcing platform, a user may compute the budget allocation between an expert and the crowd. First, a user posts a set of test questions to study the characteristics of the crowd. Based on the answers, ERICA will simulate the crowd and compute the best budget allocation. Then, the user decides how many questions are posted to the crowd.

After all answers have been collected for the crowdsourcing task, the expert starts the validation. First, ERICA will select the question, for which the feedback is most beneficial. The question and its possible answers are shown in Panel A. Once the expert has submitted their answer, ERICA will recompute the aggregated answers and the reliability of the workers. After several feedback iterations, the expert can remove the answers of a worker, if the estimated reliability of the worker is too low. ERICA will measure the effect of this removal and recompute the result accordingly. The expert can stop the validation process when the estimated precision shown in Panel B is high or the aggregated answers and the feedback are similar for a consecutive number of times, as shown in Panel C.

4. SUMMARY

This paper presented ERICA: a tool for expert guidance in validating crowd answer. Unlike traditional tools for crowdsourcing, ERICA is the first to streamline the validation process by an expert for crowd answers. The expert feedback provides a reliable means to detect and remove faulty workers, which can improve the quality of the result significantly. ERICA helps crowdsourcing users by computing the best budget allocation between the expert and the crowd. As such, ERICA serves not only as a tool to reduce the validation effort, but also acts as a decision-support system to achieve a reduction of validation cost and an improvement of result quality.

Acknowledgment. The research has received funding from the EU-FP7 EINS project (grant 288021) and the ScienceWise project.

5. REFERENCES

- [1] H. J. Jung et al. Improving quality of crowdsourced labels via probabilistic matrix factorization. In *HCOMP*, 2012.
- [2] G. Karypis et al. Metis-unstructured graph partitioning and sparse matrix ordering system. *Technical Report*, 1995.
- [3] G. Kazai et al. Worker types and personality traits in crowdsourcing relevance labels. In *CIKM*, 2011.
- [4] B. Mozafari et al. Scaling up crowd-sourcing to very large datasets: A case for active learning. In *VLDB*, 2014.
- [5] Q. V. H. Nguyen et al. Minimizing efforts in validating crowd answers. In *SIGMOD*, 2015.
- [6] A. J. Quinn et al. Human computation: a survey and taxonomy of a growing field. In *CHI*, 2011.

¹<https://code.google.com/p/crowdvalidator/>