iClarify – A Tool to Help Requesters Iteratively Improve Task Descriptions in Crowdsourcing

Zahra Nouri, [1] Nikhil Prakash, Ujwal Gadiraju, [2] Henning Wachsmuth [1]

[1]Department of Computer Science, Paderborn University
Paderborn, Germany
[2] Web Information Systems, Delft University of Technology
Delft, Netherlands
znouri@mail.upb.de, nikhil07prakash@gmail.com, u.k.gadiraju@tudelft.nl, henningw@upb.de

Abstract

Quality control and assurance are among the most important challenges in crowdsourcing. Low quality and sub-optimal responses from crowdworkers have been found to often result from unclear or incomplete task descriptions, especially from novice or inexperienced task requesters. Creating clear task descriptions with adequate information however, is a complex task for requesters in crowdsourcing marketplaces. To meet this challenge, we present *iClarify*, a tool that enables requesters to iteratively discover and revise eight common clarity flaws in their task description before deployment on the platform. A requester can use *iClarify* to formulate a task description from scratch or also to evaluate the clarity of prepared descriptions. The tool employs support vector regression models based on various feature types that were trained on 1332 annotated real-world task descriptions. Using these models, it scores the task description with respect to the eight flaws, and the requester can iteratively edit and evaluate the description until the scores shown by the tool reach a satisfactory level of clarity. We are currently conducting a usability study with both requesters and crowdworkers to assess to which extent the tool is effective in improving task clarity.

Introduction

Crowdsourcing marketplaces provide access to a diverse set of cost-effective solutions and services on-demand by leveraging the wisdom, abilities, and creativity of a large and diverse pool of workers for problems that require human input or intelligence (Demartini et al. 2017).

The quality of solutions provided by crowdworkers has been the focus of a plethora of prior research in crowdsourcing (Kittur et al. 2013). Low-quality results are known to be the dominant challenge to harness the full potential of crowdsourcing (Weld, Lin, and Bragg 2015). Among several factors that have been shown to shape the quality of crowdwork, unclear task design has been highlighted as one of the most critical (Gadiraju, Yang, and Bozzon 2017). Poor task design can lead to disappointment and frustration among crowdworkers due to a misalignment of expectations and unwarranted rejection of work (Kittur, Chi, and Suh 2008).

Writing a clear task description is thus vital for an effective task design. Usually, the combination of a task title along with a body containing specific instructions deliver the whole description of tasks. In general, a task description should be easy to understand and follow, and should describe sufficiently what is expected to be done by workers and how this should be done (Alonso and Baeza-Yates 2011).

A considerable number of workflows, models, methods, and tools has been developed to tackle the unclear task design problem (Nouri et al. 2021), including WingIt (Manam and Quinn 2018), SPRUOT (Bragg, Weld et al. 2018), and Daemo (Gaikwad et al. 2017). These tools mainly rely on involving the workers in the process of task clarity improvement, which makes it time and cost intensive and which requires good requester-workers communication. Moreover, they thereby make the quality of submissions depend on the workers which may endanger their effectiveness (Nouri, Wachsmuth, and Engels 2020).

In contrast, we introduce *iClarify*,[1] a tool that automatically detects clarity flaws in task descriptions by means of natural language processing. We argue that such an approach is not only more efficient in terms of time and cost, but it also leads to more reliable effectiveness, as the task improvement process is neither influenced by the various challenges emerging from workers' and platforms' involvement nor by the complications that workers face with requesters in the process (Nouri, Wachsmuth, and Engels 2020).

iClarify helps requesters to find the clarity flaws in their initial task description in an interactive and iterative process. Our tool has two pivotal purposes: (1) It introduces inexperienced requesters to the clarity dimensions that need to be covered for minimal task ambiguity. (2) It scores the clarity of a given task description according to these eight dimensions. Using our tool, revising and evaluating the clarity of a task description can be repeated until the scores suggest an acceptable level of clarity and completeness. Currently, we are evaluating the effectiveness of the tool through a usability study with both requesters and crowdworkers.

Models

In prior work (Nouri et al. 2021), we established the eight main clarity dimensions of task descriptions including *easy wording*, *important terms*, *desired solution*, *desired format*

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

[1]The "i" in *iClarify* stands for *iteratively*.

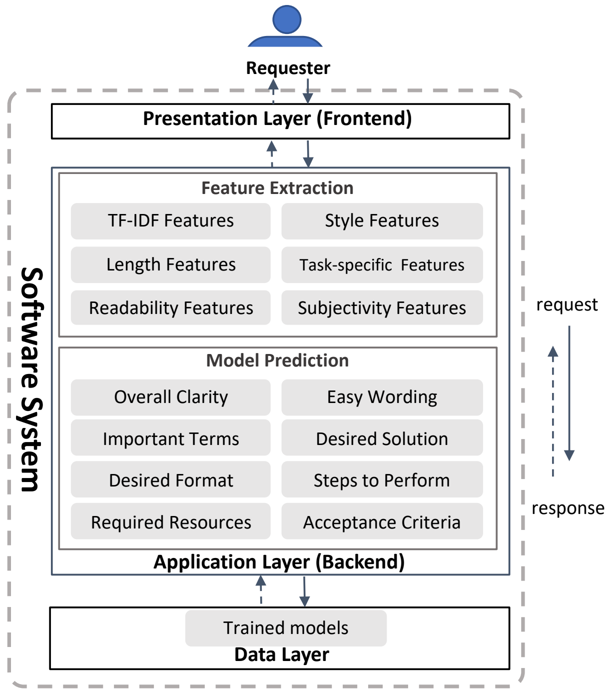

Figure 1: The three-layered architecture of *iClarify*.

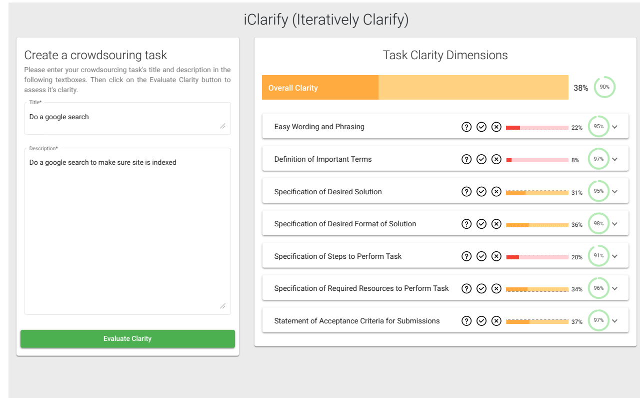

Figure 2: The Frontend (User Interface) of *iClarify*.

of solution, *steps to perform*, *required resources*, and *overall clarity*, based on which *iClarify* assesses the clarity and completeness of a given task description. We used the complete dataset, containing 1332 task descriptions, introduced in (Nouri et al. 2021) to train three scoring models for each of the eight clarity dimensions using support vector regression on various feature types capturing content, style, length, readability, subjectivity, and task specific concepts. In cross-validation on the dataset, we optimized the scoring performance of each model in terms on mean squared error and identified three sets of k best features. These feature sets along with the corresponding optimized hyperparameters were used to train the models for each dimension.

Tool Overview

iClarify is a web-based software tool which is developed using web technologies. The underlying system is constructed based on a three-layered architecture shown in Figure 1. The architecture consists of (a) the presentation layer (frontend) through which requesters interact with the system, (b) the application layer (backend) which handles the computation of clarity scores, and (c) the data layer which stores the pre-trained models.

Figure 2 depicts the system's user interface,[2] which primarily consists of two sections: (a) the input section through which user can feed their task description to the system (b) the evaluation section which represents task clarity dimensions, their corresponding scores, and a confidence value on a scale from 1 to 100. In addition, this section contains a brief description, a good example, and a bad example for each dimension, which can be viewed by clicking the ?⃝, ✓⃝, and ⊗ icons respectively. This information not only helps users to understand the meaning of each dimension, but it also guides them to create less vague task descriptions.

An initial version of a task description (title and body) can be either written within or outside the tool. By clicking the button *Evaluate Clarity*, the presentation layer sends the task description to the application layer. In the application layer, the task description is passed through all the feature type modules to compute their corresponding feature values. Afterwards, the description's feature values are fed into the pre-trained models which are fetched from the data layer. The dimensionality scores are computed by the best-performing model of each dimension. The standard deviation of the predictions of the three trained models is used to compute the confidence score of each dimension. The dimensionality and confidence scores are sent to the presentation layer and finally shown on the user interface.

Conclusion & Future Work

Unclear task descriptions written by task requesters has been identified as one of the primary issues leading to low quality crowdwork, as a result of potential misunderstanding and misinterpretation of the tasks. In this paper, we introduce *iClarify*, a tool which automatically assists requesters to find clarity flaws in their task descriptions. In an interactive process, the tool helps requesters in iteratively revising the descriptions until a satisfactory level of clarity is reached. The tool follows an automated process where no worker intervention is required, making it potentially more efficient and effective in comparison to prior solutions. We are currently evaluating the effectiveness of our tool in two main steps: (a) we are first asking requesters to create and improve task descriptions using our tool, and then (b) we are then asking crowdworkers to compare the initial and the final versions of tasks created as a result of using *iClarify*. In the imminent future, we will evaluate the usefulness of the tool with both experienced and inexperienced task requesters.

[2]A demo video presenting *iClarify* can be viewed here: https://www.youtube.com/watch?v=kA9if9X8nTw

References

Alonso, O.; and Baeza-Yates, R. 2011. Design and implementation of relevance assessments using crowdsourcing. In *European Conference on Information Retrieval*, 153–164. Springer.

Bragg, J.; Weld, D. S.; et al. 2018. Sprout: Crowd-powered task design for crowdsourcing. In *The 31st Annual ACM Symposium on User Interface Software and Technology*, 165–176. ACM.

Demartini, G.; Difallah, D. E.; Gadiraju, U.; and Catasta, M. 2017. An introduction to hybrid human-machine information systems. *Foundations and Trends in Web Science* 7(1): 1–87.

Gadiraju, U.; Yang, J.; and Bozzon, A. 2017. Clarity is a worthwhile quality: On the role of task clarity in microtask crowdsourcing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, 5–14. ACM.

Gaikwad, S. N. S.; Whiting, M. E.; Gamage, D.; Mullings, C. A.; Majeti, D.; Goyal, S.; Gilbee, A.; Chhibber, N.; Ginzberg, A.; Richmond-Fuller, A.; et al. 2017. The daemo crowdsourcing marketplace. In *Companion of the 2017 ACM conference on computer supported cooperative work and social computing*, 1–4.

Kittur, A.; Chi, E. H.; and Suh, B. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 453–456.

Kittur, A.; Nickerson, J. V.; Bernstein, M.; Gerber, E.; Shaw, A.; Zimmerman, J.; Lease, M.; and Horton, J. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, 1301–1318.

Manam, V. C.; and Quinn, A. J. 2018. Wingit: Efficient refinement of unclear task instructions. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

Nouri, Z.; Gadiraju, U.; Engels, G.; and Wachsmuth, H. 2021. What is Unclear? Computational Assessment of Task Clarity in Crowdsourcing. In *Submitted for publication*.

Nouri, Z.; Wachsmuth, H.; and Engels, G. 2020. Mining Crowdsourcing Problems from Discussion Forums of Workers. In *Proceedings of the 28th International Conference on Computational Linguistics*, 6264–6276. Barcelona, Spain (Online): International Committee on Computational Linguistics. doi:10.18653/v1/2020.coling-main.551. URL https://www.aclweb.org/anthology/2020.coling-main.551.

Weld, D. S.; Lin, C. H.; and Bragg, J. 2015. Artificial intelligence and collective intelligence. *Handbook of Collective Intelligence* 89–114.