

Understanding Conversational Style in Conversational Microtask Crowdsourcing

Sihang Qiu,¹ Ujwal Gadiraju,² Alessandro Bozzon,¹

¹Delft University of Technology

²L3S Research Center, Leibniz Universität Hannover
s.qiu-1@tudelft.nl, gadiraju@L3S.de, a.bozzon@tudelft.nl

Abstract

Crowdsourcing marketplaces have provided a large number of opportunities for online workers to make a living. To improve satisfaction and engagement of such workers, recent works have used conversational interfaces to support crowdsourcing task execution. The rationale behind using conversational interfaces stems from the potential engagement that conversation can stimulate. Prior works in psychology have also shown that ‘*conversational styles*’ can play an important role in communication. In this study, we investigate the role of conversational styles in conversational microtask crowdsourcing. To this end, we design a conversational interface which supports the task execution, and we propose ways to estimate the conversational style of a worker. We describe an experimental setup to empirically understand how estimating conversational styles of workers can help in improving quality-related outcomes. Our findings can have important implications on task design with respect to improving worker engagement unintrusively in microtask crowdsourcing.

Introduction

Crowdsourcing has become a primary means to gather human input for a variety of purposes: to build groundtruth, create datasets, to evaluate systems, and run human-centered experiments (Demartini et al. 2017; Zhang et al. 2019) among others. Currently, most online crowdsourcing tasks are executed on web platforms, like Amazon Mechanical Turk and Figure Eight, where workers usually view and complete tasks using HTML-based interfaces. However, it is quite common that large batches of tasks suffer from worker drop-outs during the course of task execution (Han et al. 2019). Many factors can cause task abandonment, and make workers less engaged.

To tackle this problem, researchers have introduced conversational agents into the realm of crowdsourcing. A conversational agent can provide online workers with a natural way to interact with crowdsourcing systems (Lasecki et al. 2013; Huang, Chang, and Bigham 2018; Bradeško et al. 2017; Jonell et al. 2018). In our recent work, we found that conversational interfaces can be used as an alternative to assist workers in task execution, producing similar output quality and task execution time, compared to traditional

web interfaces (Mavridis et al. 2019). Previous works in the field of psychology have studied how conversational styles can affect inter-human communication (Lakoff 1979; Tannen 1987; 2005). Researchers have also attempted to make computer systems emulate specific conversational styles to make systems more user-friendly (Shamekhi et al. 2016; Kim, Lee, and Gweon 2019). However, to the best of our knowledge, current conversational agents (particularly for crowdsourcing) have not exploited the conversation styles to improve the overall effectiveness of the crowdsourcing paradigm. Understanding the role of conversational styles in human computation can help us better adapt strategies to improve output quality and worker engagement, or better assist and guide workers in the training process.

In this study, we will delve into the following:

- How can we reliably estimate the conversational style of a crowd worker?
- How do different workers perceive conversational agents emulating distinct conversational styles?
- To what extent does the conversational style of an agent affect quality related outcomes, and worker engagement in different types of tasks?

We designed and implemented a conversational interface to estimate the conversational style of workers. We plan to conduct experiments to analyze the impact of conversational style on the worker performance and engagement.

Conversational Microtask Crowdsourcing

Microtask crowdsourcing is a process where workers interact with a crowdsourcing system to execute an online task (such as data annotation, image labeling, etc). The crowdsourcing system is responsible for worker selection, microtask generation, microtask assignment and answer aggregation. Workers, who are recruited from the Internet, are responsible for executing microtasks assigned to them.

Text-based Conversational Interface

To enable a conversational interface that can support crowdsourced task completion in the same manner that traditional web interfaces can do, the conversational interface should consist of the following parts.

Greetings. To simulate the essential characteristics of conversation, the conversational agent begins with greetings.

The goal of this phase is to let workers familiarize themselves with the conversational interface, and to help them build a sense of ‘conversing with the system’.

Task Instructions. In contrast to the traditional web interface, the conversational interface lets workers understand how to execute tasks by introducing the task using dialogue.

Questions & Answers. The conversational interface asks questions to workers, and workers can respond to these questions by either typing answers or by using the UI elements provided. The conversational flow of this phase needs to be specifically designed according to the task type.

Review. On the traditional web interface, a worker can easily navigate back to a question and edit the previous answer. To preserve this affordance in the conversational interface, workers are provided with an opportunity to check and edit any of their previous answers during this phase.

Task Types and UI Elements

The traditional web interfaces of popular crowdsourcing platforms are designed and developed based on HTML/CSS/Javascript. To make sure workers from these crowdsourcing platforms can easily use the conversational interface on their task pages without redirecting to another chatting or messaging application, the conversational interface is also implemented based on such languages, so that it can be readily embedded on the default task page.

The conversational interface can support any data types that are supported by HTML5, including text, image, audio and video. Thus, most common task types such as image classification, sentiment analysis, information finding, object recognition, audio transcription can be easily executed on the conversational interface. Furthermore, UI elements for the conversational interface (such as bubble-like buttons, customized keyboard) are designed to map different types of web elements (like buttons, checkboxes). UI elements from traditional web interfaces can also be simply ported on conversational interfaces. All these features make the conversational interface suitable to execute most online tasks.

Methodology

Existing works have not explored the potential benefits of understanding the role of conversational styles of agents and how they interact with that of workers in crowdsourcing systems. Emulating particular conversational styles suitable to given contexts, or aligning the conversational style of an agent to the preferred style of workers, may help to improve worker engagement, satisfaction and even output quality.

Features of Conversational Style

According to Tannen’s theory, conversational style can be classified into two categories: High Involvement and High Considerateness (Tannen 2005). The conversational style of High Involvement has the following important features with regard to topic, pacing, narrative strategy and expressive paralinguistics. 1) In terms of the **topic** of the conversation, people of High-Involvement conversational style prefer personal topics, shift topics abruptly, introduce topics without

hesitation, and keep reintroducing the topic. 2) **Pacing** is the most easy-to-observe feature. The High-Involvement style implies a faster rate of speech, faster turn taking, and fewer inter-turn pauses. 3) Furthermore, **narrative strategies** are commonly used during the conversation. Telling more stories, telling them in rounds and internal evaluation characterize the High-Involvement conversational style. 4) **Expressive paralinguistics** features are also important. However, most paralinguistic features are out of the scope of our work, since this paper only focuses on text-based conversation.

Style Classification

Our proposed method to automatically estimate the conversational style of a worker, is to design the conversational interface so as to facilitate a “natural” conversation with the worker prior to the task execution. During this phase, we can measure the relevant features from the conversation, and use a supervised machine learning model to estimate whether the style is of High Involvement or High Considerateness. Since not all features discussed earlier can be easily measured from a text-based conversation, we select the following features: frequency of pronouns used, rate of typing, length of pause, the number of interruptions and length/complexity of the message. These features will be normalized and used as an input to the classification model. Furthermore, a coding scheme for manually labeling ground-truth data required to train and validate the supervised model will be designed according to Tannen’s definition of the conversational style and corresponding linguistic devices.

Experiments and Evaluation

RQ1: How can we reliably estimate the conversational style of a crowd worker?

We will design for conversational interactions before the actual task execution begins. We will build the context around controversial topics to facilitate active conversation. The rationale behind this is that a controversial topic can better stimulate the desire of expressing thoughts in general. We will then label workers’ styles using the coding scheme, and train a model for conversational style estimation.

RQ2: How do different workers perceive conversational agents emulating distinct conversational styles?

We will collect self-reported worker preferences of conversational styles by leveraging Tannen’s characterization of the features. Next, we will analyze the relationships between workers’ own conversational styles, their preferred style, and that of the conversational agent.

RQ3: To what extent does the conversational style of a conversational agent affect quality related outcomes, and worker engagement in different types of tasks?

We will deploy several crowdsourcing tasks with different data types (text and image) and input types (free text and multiple choices). To observe the impact of the conversational style, we will measure the accuracy of responses, worker retention, and use UES-SF (O’Brien, Cairns, and Hall 2018) and the NASA-TLX instrument to analyze user engagement and the perceived cognitive task load respectively.

References

- Bradeško, L.; Witbrock, M.; Starc, J.; Herga, Z.; Grobelnik, M.; and Mladenić, D. 2017. Curious cat–mobile, context-aware conversational crowdsourcing knowledge acquisition. *ACM Transactions on Information Systems (TOIS)* 35(4):33.
- Demartini, G.; Difallah, D. E.; Gadiraju, U.; Catasta, M.; et al. 2017. An introduction to hybrid human-machine information systems. *Foundations and Trends® in Web Science* 7(1):1–87.
- Han, L.; Roitero, K.; Gadiraju, U.; Sarasua, C.; Checco, A.; Maddalena, E.; and Demartini, G. 2019. All those wasted hours: On task abandonment in crowdsourcing. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 321–329. ACM.
- Huang, T.-H. K.; Chang, J. C.; and Bigham, J. P. 2018. Evorus: A crowd-powered conversational assistant built to automate itself over time. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 295. ACM.
- Jonell, P.; Bystedt, M.; Dogan, F. I.; Fallgren, P.; Ivarsson, J.; Slukova, M.; Ulme Wennberg, J. L.; Boye, J.; and Skantze, G. 2018. Fantom: A crowdsourced social chatbot using an evolving dialog graph. *Proc. Alexa Prize*.
- Kim, S.; Lee, J.; and Gweon, G. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 86:1–86:12. New York, NY, USA: ACM.
- Lakoff, R. T. 1979. Stylistic strategies within a grammar of style. *Annals of the New York Academy of Sciences* 327(1):53–78.
- Lasecki, W. S.; Wesley, R.; Nichols, J.; Kulkarni, A.; Allen, J. F.; and Bigham, J. P. 2013. Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, 151–162. ACM.
- Mavridis, P.; Huang, O.; Qiu, S.; Gadiraju, U.; and Bozzon, A. 2019. Chatterbox: Conversational interfaces for micro-task crowdsourcing. In *International conference on user modeling, adaptation, and personalization*. Springer.
- O'Brien, H. L.; Cairns, P.; and Hall, M. 2018. A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies* 112:28–39.
- Shamekhi, A.; Czerwinski, M.; Mark, G.; Novotny, M.; and Bennett, G. A. 2016. An exploratory study toward the preferred conversational style for compatible virtual agents. In *International Conference on Intelligent Virtual Agents*, 40–50. Springer.
- Tannen, D. 1987. Conversational style. *Psycholinguistic models of production* 251–267.
- Tannen, D. 2005. *Conversational style: Analyzing talk among friends*. Oxford University Press.
- Zhang, Z.; Singh, J.; Gadiraju, U.; and Anand, A. 2019. Dissonance between human and machine understanding. *Proc. ACM Hum.-Comput. Interact.* 3 CSCW 56:26.