Leveraging the Crowd to Support the Conversation Design Process

Yoonseo Choi
School of Computing
KAIST
Daejeon, South Korea
yoonseo.choi@kaist.ac.kr

Nyoungwoo Lee
School of Computing
KAIST
Daejeon, South Korea
leenw2@kaist.ac.kr

Hyungyu Shin
School of Computing
KAIST
Daejeon, South Korea
hyungyu.sh@kaist.ac.kr

Jeongeon Park
School of Computing
KAIST
Daejeon, South Korea
patricia021912@kaist.ac.kr

Toni-Jan Keith Monserrat
Institute of Computer Science
University of the Philippines Los
Baños
Los Baños, Laguna, Philippines
tpmonserrat@up.edu.ph

Juho Kim
School of Computing
KAIST
Daejeon, South Korea
juhokim@kaist.ac.kr

Author Keywords

Iterative design; Conversation design; Chatbot design process; Early-stage design support; Crowdsourcing

CCS Concepts

•**Human-centered computing → Systems and tools for interaction design; Empirical studies in interaction design;** *User interface design;*

Abstract

Building a chatbot with human-like conversation capabilities is essential for users to feel more natural in task completion. Many designers try to collect human conversation data and apply them into a chatbot conversation, aiming that it could work like a human conversation. To support conversation design, we propose the idea of inviting the crowd into the design process, where crowd workers contribute to improving the designed conversation. To explore this idea, we developed ProtoChat, a prototype system that supports a conversation design process by (1) allowing the crowd to actively suggest new utterances based on designers' pre-written design and (2) visually representing crowdsourced conversation data so that designers can analyze and improve their conversation design. Results of an exploratory study indicated that the crowd is helpful in providing insights and ideas as designers explore the design space.

Introduction

When building a chatbot, designers aim to embed human-like conversation capabilities so that users feel more natural in completing their tasks with a chatbot. When designing a conversation, many designers attempt to collect from small to large scale human conversation data so that their chatbot models a desired human conversation. Existing approaches either collect conversation data from humans or formulate the conversation by analyzing existing data sources. Ex-

amples include Wizard-of-Oz [7, 4], workshops [5], Twitter conversation data [3, 8], mail threads of DBpedia [1], and existing chatbot logs [9].

To understand the challenges designers face during the conversation design process, we conducted semi-structured interviews with two professional conversation designers (at least 1 year of experience) and seven amateur designers with prior experiences in conversation design. Results show that designers find it difficult to discover possible conversations in the design process, especially when they are not familiar with the chat domain. Also, it is overwhelming to rapidly iterate the conversation design as the iterative process requires not only the design of the conversation but also prototyping and testing the working chatbot, which can be built using frameworks such as Google Dialogflow [1], BotKit [2] and Chatfuel [3]. Supporting the iterative design of conversation helps designers to get a sense of how potential users might follow the current design of conversation. Yet, there has not been much investigation on how designers iterate the conversation design, challenges within the process, and how to design a system to support the process.

To address these challenges, we suggest crowdsourcing as a solution to supporting iterations in the design process. Inviting crowd workers in the design process potentially lessens the burden of designers to quickly test and improve the design. There have been studies that incorporated crowdsourcing in real-time chat scenarios. Chorus [6] demonstrated how the crowd could come up with not only a diverse set of responses but also a diverse set of variations of descriptions on a given topic, where they expected

crowdsourcing as a potential approach to explore diverse conversations in the chat domain. Extending this line of work, we aim to utilize the crowd during the *conversation design process*, not during the chat session. We believe that the crowd could empower the iterative design process by enabling rapidly testing intermediate designs. By testing with the crowd, designers can easily test their design in a lightweight manner compared to Wizard-of-Oz, a workshop, or a lab study.

To explore the idea of using crowdsourcing in the conversation design process, we developed *ProtoChat*, a prototype system for supporting designers by (1) allowing the crowd to actively suggest new utterances and turn-taking conversation based on designers' scenarios and (2) representing crowdsourced data in effective ways to assist designers analyze and improve the conversation. We decided to set the role of the crowd as the active suggester of conversation design so that designers could get enough evidence in decision making. We first designed the crowd-testing interface to test the designer's draft conversation with the crowd. Crowd workers are provided with the interface to either follow the designed conversation by responding or suggesting new messages between the conversation. Other than the crowd-testing interface, we designed a separate interface for designers where they can design the conversation flow with a set of chatbot utterances and their corresponding topics. Furthermore, after crowd-testing, designers could review and analyze the collected responses from the crowd with the designer interface to iteratively improve their design.

Through our exploratory study with four conversation designers, we found that ProtoChat enables designers to quickly design and test their ideas about a conversation with the crowd and analyze the data provided by the crowd
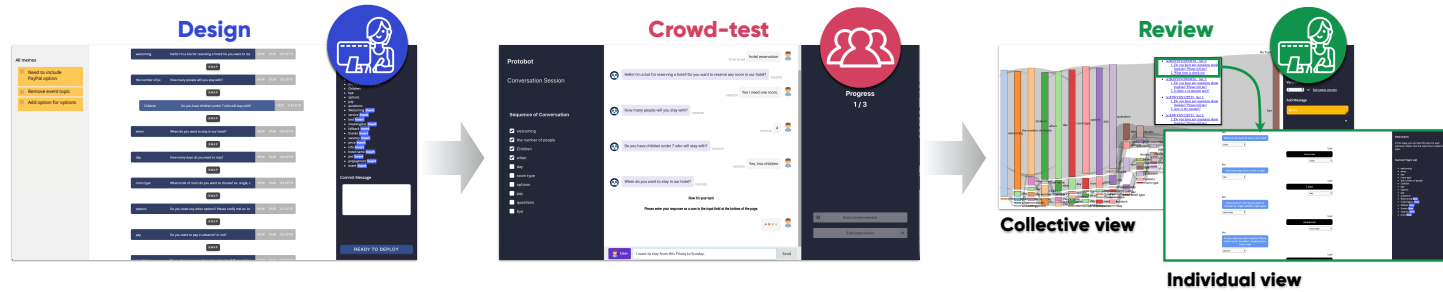

Figure 1: Conversation design process with ProtoChat. Designers can design a conversation, test with the crowd, and browse and analyze the crowdsourced conversation.

Figure 2: Chabot's turn on the crowd interface. The crowd could either (a) proceed conversation with the designed scenario, (b) insert a new scenario, or (c) follow new scenarios that the crowd created.

so that they could improve the design. The crowd enabled need-finding of the chat domain which is important in the early stage of conversation design. Moreover, the crowd provided evidence for decision making regarding later versions of the design. This research aims to help chatbot designers with crowd power, from collecting conversation scenarios by crowd contribution to effectively representing the crowdsourced output, which enables designers to explore potential scenarios, important in understanding the needs and use cases of real users.

This work is based on our ongoing research [2], in which we introduce the ProtoChat system and its preliminary evaluation. In this workshop paper, we briefly summarize the system and our findings so far, and discuss the use of crowdsourcing as a method for helping the conversation design process. Through our work, we expect to promote a discussion in the community of CUI (Conversational User Interface) researchers about applying crowdsourcing for data collection, decision making support, and design space exploration and confirmation in a broad set of CUI application scenarios.

ProtoChat: Crowd + Conversation Design
ProtoChat supports a conversation design process by (1) allowing crowds to actively suggesting the design based on

designers' scenarios and (2) providing designers effective representation of crowdsourced data to assist designers analyze and improve the conversation. The system consists of two interfaces: the designer interface and the crowd-testing interface.

The designer interface supports designers to design low-fidelity conversations, plan how to test the conversation design, review the crowdsourced data, and refer to the previous version design and analysis as well. Designers can design a conversation with 'topic' and 'utterance', two basic building blocks of designing a conversational flow. Topic refers to what kind of questions need to be asked in the domain (e.g. Payment) and utterance refers to how the topic is addressed within the conversation (e.g. "How would you like to pay?"). Three main features are provided in designer interface: *Draft*, *Review* (See Figure 1-1, 3), and *History*. In the Draft page, designers can create a conversation by defining a set of topics and utterances (See Figure 1-1). Designer could also configure parameters and launch customized tests. In the Review page, designers can browse and review the crowdsourced conversations with collective view and individual view (See Figure 1-3). Additionally, ProtoChat provides the History page, where designers can review their previous design versions and test settings.
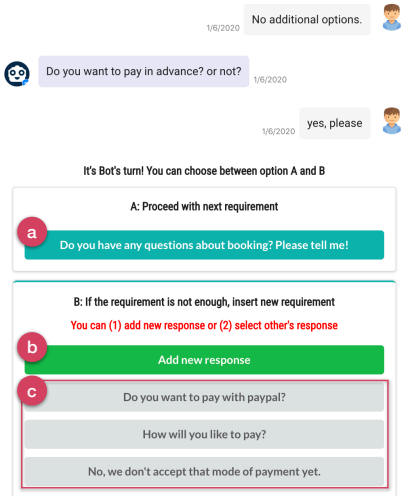
Figure 3: The procedure of the 3-day experiment.

Design

(Day 1) Create a conversation flow to successfully guide users to finish the task, in the form of chatbot.
(Day 2, 3) Revise their previous version of conversation design after reviewing and analyzing crowd data in the Review phase.

Crowd-test

Configure test settings after designing the overall conversation flow and run the test. The testing method was fixed to MTurk.

Review

Browse and analyze the data crowdsourced by the Crowd-testing interface. Participants could move back-and-forth between the individual and the collective view.

Export

Export the design into a hi-fidelity design prototype each day. By requiring participants to export their conversational flow into a design prototype that supports end-to-end scenarios, they were able to recognize whether their ongoing design needs more editing or iterations.

Figure 4: Detailed explanation of each task performed in the experiment.
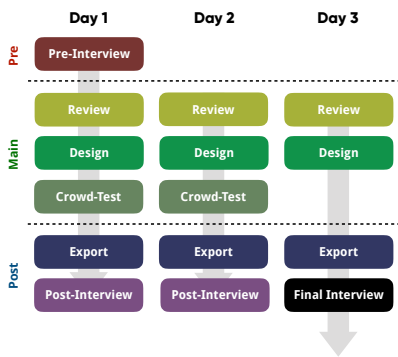
The crowd-testing interface is a separate auxiliary interface to quickly test the designed conversation. Crowds are guided to proceed the conversation with the designers' conversation flow in the form of chatbot, with pre-instructions provided. The crowd-testing interface(See Figure 1-2) has a unique feature that enables the crowd to either proceed conversation with responding to pre-written utterances of a chatbot or insert new conversations within the existing conversation flow (See Figure 2) to elaborate on the conversation. If the crowd chooses to insert a new conversation, it will be performed as a self-dialogue.

Evaluation

Experiment procedure
The experiment focused on exploring the role and contributions of the crowd in conversation design. To examine ProtoChat's role during the overall design process, we conducted a 3-day long experiment. We recruited 4 designers (3 female, 1 male) who work on research topics involving conversation design and have prior experience of chatbot conversation design. Participants received $45 for the whole experiment. During the experiment, we asked participants to mainly work on three tasks: 1) *Design*, 2) *Crowd-test*, and 3) *Review* the conversation for each day so that they work on three iterations. The overall procedure of the 3-day experiment is shown in Figure 3. Figure 4 shows detailed explanation of each task performed in the overall experiment.

Result
Participants chose different domains for their conversation design (P1: movie reservation, P2: ice cream order, P3: hotel reservation, P4: house fixing). We found that several interactions supporting conversation design were enabled by crowd power. With the current version of crowd-testing interface, we were able to observe two types of crowd contribution toward the design of chatbot conversation.

Need-finding of a new conversation
With crowd-testing, participants were able to observe both the desired flow of conversation by reviewing the collective view and the needs from the crowd by analyzing individual chat logs. Specifically, participants were able to do need-finding within the context of a conversation. Need-finding includes collecting responses about specific topics/questions or collecting new questions that can be asked for a more natural conversation. They agreed that the responses collected from the crowd helped them discover user needs and make internal decisions for next step designs. For example, P1 and P4 mentioned that based on the responses from open-ended questions, they were able to make UI decisions in the Export step such as deciding between button UI or free-formed question, or even the content of answer format. P2 realized that they missed out on the option between choosing a cup or cone, after seeing the new crowd-added utterance − "A cup or a cone?", which they felt was an essential topic that needs to be dealt during an ice cream order.

Design confirmation with majority voting
When participants tried to make decisions about their design, they checked the path that the majority of crowds followed in the Sankey diagram during the Review phase and verified whether their conversation flow makes sense and is easy to follow. Based on the the majority responses of the crowd in the chatbot's each utterance, participants were able to refine their utterances during the iterative process, not the overall sequence or the order of topics. Detailed revisions were made such as changing the tone of the bot utterances (e.g. format of the questions), but the overall sequence of topics was similar for multiple iterations.
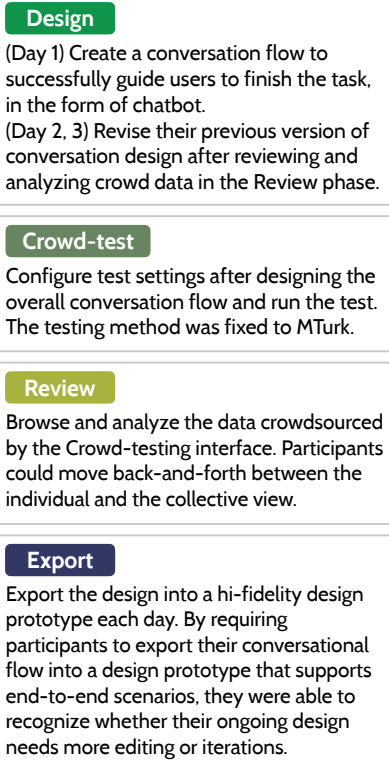
Discussion and Future Work

In the workshop, we want to discuss ways in which crowd-sourcing could be applied in CUI research and ways to make it more effective.

Further investigation toward realistic conversation design
In a real-world setting, it is hard to run iterations in early-stage chatbot design process, as it involves prototyping and testing a working chatbot with potential users. P2 mentioned that "*This tool helps in collecting feedback for my own design with a large number of crowds in a lightweight manner, which enables experiencing a quick iterative process.*". We believe that the design space of supporting iterative design of conversation needs to be further investigated, especially by focusing on how to support more diverse forms of conversations. For example, our system did not support branching interactions in a conversation, which is one of the fundamental building blocks in chatbots. Also, it is important to investigate not only the iteration of the conversation design but also the whole iteration of chatbot building process that incorporates the iterative conversation design.

Supporting branching interactions with the crowd
Designers pointed out that branching out the conversation threads based on user input would be necessary to support a more realistic conversation design than the current version. For example, if a chatbot asks the user "Would you like to order snacks?" then an adaptive conversation flow can respond differently based on user input between "yes" or "no". With increased flexibility of the role of crowd workers such as providing adaptive responses with specific intents, we envision the crowd-driven expansion of the scenario tree with moderate designer intervention. The branching support can increase the the level of complexity in conversation design that our system can handle.

Collecting alternative expressions
By observing the ability to collect diverse topics of conversation within each chat domain, we wish to improve the system to collect diverse alternatives for a specific expression. Collecting various alternatives with the same meaning is crucial because those expressions will be able to provide a more dynamic, natural conversation experience to users, by displaying utterances of the same meaning in different ways. For instance, designers can build a chatbot that gives different responses every time even if the user asks similar questions. Plus, the collected conversation can be used as training data to build a machine learning model that powers a human-like chatbot. One possible way of collecting diverse expressions would be to improve the prompting mechanism in the crowd-testing interface, where the crowd is provided with questions that ask them to paraphrase the sentence that they just provided.

All in all, we believe that exploring how crowdsourcing can be applied in CUI design can lead to constructive discussions and novel research directions. In the workshop, we plan to further discuss challenges of using the crowd in CUI design and effective ways to address them.

Acknowledgments

REFERENCES

[1] Ram G. Athreya, Axel-Cyrille Ngonga Ngomo, and Ricardo Usbeck. 2018. Enhancing Community Interactions with Data-Driven Chatbots–The DBpedia Chatbot. In *Companion Proceedings of the The Web Conference 2018 (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 143–146. DOI:

http://dx.doi.org/10.1145/3184558.3186964

[2] Yoonseo Choi, Hyungyu Shin, Toni-Jan Keith Palma Monserrat, Nyoungwoo Lee, Jeongeon Park, and Juho Kim. 2020. Supporting an Iterative Conversation Design Process (To appear). In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, Article Paper LBW1458, 6 pages.

[3] Tianran Hu, Anbang Xu, Zhe Liu, Quanzeng You, Yufan Guo, Vibha Sinha, Jiebo Luo, and Rama Akkiraju. 2018. Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 415, 12 pages. DOI: http://dx.doi.org/10.1145/3173574.3173989

[4] Meng-Chieh Ko and Zih-Hong Lin. 2018. CardBot: A Chatbot for Business Card Management. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion (IUI '18 Companion)*. ACM, New York, NY, USA, Article 5, 2 pages. DOI: http://dx.doi.org/10.1145/3180308.3180313

[5] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 70

(July 2018), 26 pages. DOI: http://dx.doi.org/10.1145/3214273

[6] Walter S. Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F. Allen, and Jeffrey P. Bigham. 2013. Chorus: A Crowd-Powered Conversational Assistant. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. Association for Computing Machinery, New York, NY, USA, 151–162. DOI: http://dx.doi.org/10.1145/2501988.2502057

[7] Archana Prasad, Sean Blagsvedt, Tej Pochiraju, and Indrani Medhi Thies. 2019. Dara: A Chatbot to Help Indian Artists and Designers Discover International Opportunities. In *Proceedings of the 2019 on Creativity and Cognition (C&C '19)*. ACM, New York, NY, USA, 626–632. DOI: http://dx.doi.org/10.1145/3325480.3326577

[8] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A New Chatbot for Customer Service on Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3506–3510. DOI: http://dx.doi.org/10.1145/3025453.3025496

[9] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2018. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. *CoRR* abs/1812.08989 (2018). http://arxiv.org/abs/1812.08989