Exploring preferences on chatbot's guidance type and timing

Su Fang Yeh
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
sfy.iem07g@nctu.edu.tw

Meng-Hsin Wu
University of Toronto
Toronto, Ontario, Canada
soniawu2302@gmail.com

Yung-Ju Chang*
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
armuro@cs.nctu.edu.tw

ABSTRACT

While task-oriented chatbots have become popular recently, conversational breakdowns are still common and will often lead to unfavourable user experience. Guidance is fundamental to users' understanding in chatbot's understanding capabilities and will prepare users to better interact with them, but the underlying analysis awaits detailed characterization. Here, we explored the interplay between guidance type and guidance timing by providing various guidance to users. Participants in this study were exposed to eight different combinations of guidance types (example-based or rule-based) and timing (opening, task introduction, failure, and on-demand). This paper reports our preliminary observation of 4 participants comparing all combinations. We found that guidance provided in the early stage is favoured by most of the participants. Through analysis of participants' explanation on their preference, strengths and weaknesses of each guidance type and guidance timing were presented. We also noticed some interesting interplay effects arose from guidance timing and guidance type, which will be further investigated in our future research.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **User studies**.

KEYWORDS

chatbots, conversational agents, guidance, breakdowns , discoverability

Table 1: Guidance Timings (Initial)

Guidance Timing	Example Dialogue
Within opening message	**Bot**: Hi, I'm travel-bot. I can help you with tracking flights, search flight tickets and recommend some accommodations. *Hint* For track flight, you can say something like " Track flight which depart from Tokyo and land at Taiwan at the day after tomorrow" *For search flight tickets, you can say something like " Two economic seats for Tokyo to London (*continues...*)
Within task introduction	(opening message without *hint*) **User**: I want to find some recommended accommodations. **Bot**: *hint* For accommodation recommendations, you can say something like: " I would like to find a hostel that can be reserved for 12/1 - 12/2. And I prefer the one that provides bicycle rental." **Bot**: What kinds of accommodation are you looking for? (*continues...*)

INTRODUCTION

Chatbots have been used for various purposes such as chit-chat or for helping users to perform domain-specific tasks [22]. This research focused on task-oriented chatbots, which serve as an important tool to save time on repetitive tasks [20]. Existing chatbots usually rely on the intent-based model [24]. However, when the chatbot mis-recognizes users' intents, a conversational breakdown would occur and users cannot make progress towards their goal [12]. This may trigger users' negative emotion [19] and in some serious situations, users may even abandon the chatbot service [7, 12].

Recent studies found that people often assumed chatbots have a human-level understanding, and this may be a trigger of conversational breakdowns [7, 10, 12]. Some believe that this problem can be resolved by helping users better understand how to interact with the system [5, 7, 16]. Many researchers have sought to know how to enhance people's understanding about chatbots. For instance, Weisz et al. [23] developed a role-play tutorial to increase participants' empathy and set a relative realistic expectation toward chatbots. Ashktorab et al. [2] focused on giving explanations to repair conversation breakdowns. However, seldom did researchers discussed how to guide users with input suggestions in a conversational user interface. Our study focused on two research questions related to guidance providing, which are 1) When is the best timing to show guidance, and 2) What kinds of guidance should be provided.

Unexpected guidance often distracts users' attention from their main task and slow them down [17]. Giving guidance that is expected by the users therefore becomes a crucial task. In 2018, Jain et al.[7] interviewed 16 first-time chatbot users and found participants prefer guidance in initial stage of the interaction. Nevertheless, Kirschthaler et al. [9] found that compared to automatically provided guidance, participants preferred guidance that shows up by requested when using a voice user interface (VUI) in long-term. In contrast, it is known that in the traditional website interface, people scan content and only read system guidance when they get into trouble [14, 15]. How users perceive different timing of chatbot guidance still remains unknown. In this study, four timing were compared. For the initial stage of the interaction, we further divided it into "guidance within opening message" and "guidance within task introduction" (See Table 1). The other two timings are "guidance as a feedback for failed utterance" and "guidance on-demand" (See Table 2).

For guidance types, we found example-based guidance is the most frequently mentioned type in various kinds of intelligent user interface, perhaps due to its ability to express complex concepts[3, 7, 9, 21]. However, examples hardly tell the underlying rules [4, 18]. Therefore, we include another guidance type that is often be in comparison with example-based guidance, i.e. rule-based guidance [4, 8]. Rule-based guidance helps users to understand how a system work. It is especially useful when unexpected situations happened [13]. Nonetheless, some studies revealed that people need to take more time learning rules than learning examples[4].

Table 2: Guidance Timings (Others)

Guidance Timing	Example Dialogue
As feedback for failed utterance	(...) **User**: (*something that can not be processed by the chatbot*) **Bot**: Can you rephrase it ? I can't understand that. *Hint* For accommodation recommendations, you can say something like: " I would like to find a hostel that can be reserved for 12/1 -12/2. And I prefer the ones that provide bicycle rental." (*continues...*)
On demand	(...) **User**:How can I get accommodation recommendations ? **Bot**: *Hint* For accommodation recommendations, you can say something like: " I would like to find a hostel that can be reserved for 12/1 - 12/2. And I prefer the ones that provide bicycle rental." (*continues...*)

Table 3: Guidance Types

Example-based	Rule-based
You can say something like, "I would like to find a hostel that can be reserved for 12/1 - 12/2. And I prefer the ones that provide bicycle rental."	*Use Arabic numerals * Use "YYYY/MM/DD - YYYY/MM/DD for date format.

Both example-based and rule-based guidance have their advantages and disadvantage. It seems like whether a type is advantageous somehow depend on the timing of guidance. See Table 3 for the two guidance types we used in this study. The rules in this study were gathered by testing IBM Watson's [6] Mandarin understanding capability.

Finally, we presented eight guidance combinations that are consisted of 4 guidance timings x 2 guidance types. Scenario study and interview were conducted with four participants. Our preliminary results provide three major findings. First, while the participants generally preferred guidance that shows earlier, one participant reported that when it comes to long-term usage, she'll like on-demand guidance. Second finding is the strengths and weaknesses of each guidance timing and guidance type. For instance, we found that while examples serve as a convenient template for users to mimic, it suffers from it lengthiness. And lastly, potential interplay relationship between guidance types and timings was found. For example, one participant prefer example-based guidance within the opening message, and strongly dislikes guidance as a feedback of failure utterance. These findings show that future chatbot guidance design should both consider guidance timing and guidance types. This study is in pilot stage and we plan to recruit more participants in the future.

METHOD

This study adopted the scenario-based method that has often been used in human-robot interaction studies [2, 11] to evaluate people's perception of guidance type and timing. Four participants were recruited as pilot in this study. Three are male and one is female. They are graduate students in Taiwan who have experience using chatbots.

We developed eight scenarios in which the chatbot adopts one of the guidance combinations as mentioned above. We also included a control scenario that does not contain any guidance. User asks for accommodation information in each scenario. Inspired by insertion sort [1], this study asked participants to sort their preference in a similar manner. We first displayed one scenario for participants to read. After the participant read the scenario, another scenario will be shown. Participants then were asked to rank their preference and tell the reason behind their ranking with a short interview. One new scenario would be displayed after each ranking trial. This procedure repeat until all nine scenarios were sorted.

PRELIMINARY RESULT

Overview of the Guidance Preferences

When participants were asked to rank conversation scenarios based on their personal preferences, most of them (three out of four) had scenarios that contain early guidance (either within opening messages or within task introduction) at a higher ranking. Early guidance is deemed as helpful

because it allows participants to specifically know what to type to successfully achieve the given tasks. Interestingly, one participant (P3) proposed that automatically-provided guidance, regardless of timing and type, may be troublesome in long-term usage, which is consistent with previous findings in VUI [9]. However, timing is not the only significant factor to be considered by some participants. For instance, one participant (P1) preferred example-based guidance particularly at the early stage, and rule-based guidance when conversational failures occur. According to P1, "*Initial examples clearly demonstrate what information is needed...example may not help me to recognized errors when failure happen,*" and "*Rules is trivial...I may not carefully read it at first...it's more useful when I failed because I can recognize the problem more quickly.*" In contrast, another participant (P4) claimed that he prefers rule-based guidance particularly at the early stage, and example-based guidance when conversational failures occur. He explained, "*People have different typing habits, and if it's different from the example, you will feel that it's stipulating how you should speak...I still want a human-like conversation,* " and "*I feel annoying if such important message (rules) shows so late (when fail).*" Detailed factors that contribute to their opposite preferences, and how to provide suitable guidance combination based on those factors await further investigation in our future study.

Brief Analysis of Guidance Timing

Within Opening Message and Task-introduction. Participants (P1-2,4) reported that for users who feel uncertain about what to provide or how to structure their sentences, guiding them in the opening sentence can help them become more efficient. Not only they can avoid breakdown easily, they can also skip the process of testing chatbots' understanding capability with long back-and-forth conversations. However, P3 complained that guidance showing at this timings will become unnecessary and annoying in long-term usage since it automatically provides every time. Presently, no major differences were identified in the scenario where the guidance is shown within task introduction.

On-demand. One participant (P3) reported that this timing is more flexible, and provides a cleaner user interface as users can decide whether to let the guidance pop up, making it feel like a more natural conversation. After a few tries, users may choose not to call out the guidance. A major drawback of this timing is that users may not be able to receive the guidance if they have not actively asked for it. For example, P1 complained that " *I don't see anything wrong in my input, but the chatbot just can't understand me... It's troublesome to call for help... Why not just tell me I need to use Arabic number (rules)?*"

As Feedback for Failed Utterance. One advantage of this timing is that users could recognize the reason for conversational errors and move on. One participant (P1) said he'll only read the guidance when a breakdown occurs (in case of rule-based guidance). However, the breakdown could lead to abandonment. P2 claimed that the guidance shows quite late for him in this scenario (The reason

that sometimes chatbot will not show guidance right after failure is that misunderstanding cannot be recognized by the chatbot itself), and he would abandon the service if he fails too many time. P4 specifically addressed that the maximum number of failure he can tolerate is two. In such situation, the guidance after failure does not matter because he will give up on this service.

Brief Analysis of Guidance Type

Example-based Guidance. Three participants (P1-3) claimed that example-based guidance is more concrete, and serves as a specific template for them to type sentences that matches the tasks' goal. "*It's easier to follow examples since there are potentially infinite ways of speaking to make a request...I can imitate the sentence structure of the given example*," said P2. Such advantage is responded to be applicable even if the provided example is not in the same domain as the participants' ongoing tasks. Nevertheless, example-based guidance may suffers from its lengthiness, which makes it less readable. Some participants (P1,4) also claimed that they tends to use less words than example did when interacting with a chatbot in real life. Another major disadvantage of example-based instruction is that it is rather implicit and not everyone can understand it. For instance, a participant (P1) maintained that he couldn't see why conversational failure occurs and how to recover from it based on the example guidance.

Rule-based Guidance. Rule-based guidance provides necessary information about input formatting (P1,3-4), which was reported to be crucial because "*even my own typing format may change in many ways*,," said P3. However, participants (P2,3) also argue that rules only tell about sentence formatting, but not about how to use it properly. The information it provided is rather limited, so they will hesitate what things to type in order to make the chatbot understand. Interesting, P4 reported that a chatbot that incorporates rule-based guidance provides a more natural, human-like experience because it provided a freedom to construct a sentence, and he do not have to follow an example sentence word-by-word.

REFERENCES

[1] Khan Academy. 2021. *Insertion sort.* Retrieved January 16, 2021 from https://www.khanacademy.org/computing/computer-science/algorithms/insertion-sort/a/insertion-sort

[2] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[3] Carrie J Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 258–262.

[4] Nick Ellis. 1993. Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology* 5, 3 (1993), 289–318.

[5] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[6] IBM. 2021. *IBM Watson products and solutions.* Retrieved January 16, 2021 from https://www.ibm.com/watson/products-services

[7] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N Patel. 2018. Evaluating and informing the design of chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 895–906.

[8] F Kaptein, J Broekens, K Hindriks, and MA Neerincx. 2021. Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence* 291, 291 (2021).

[9] Philipp Kirschthaler, Martin Porcheron, and Joel E Fischer. 2020. What Can I Say? Effects of Discoverability in VUIs on Task Performance and User Experience. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–9.

[10] Knut Kvale, Olav Alexander Sell, Stig Hodnebrog, and Asbjørn Følstad. 2019. Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. In *International Workshop on Chatbot Research and Design*. Springer, 187–200.

[11] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully mitigating breakdowns in robotic services. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 203–210.

[12] Chi-Hsun Li, Su-Fang Yeh, Tang-Jie Chang, Meng-Hsuan Tsai, Ken Chen, and Yung-Ju Chang. 2020. A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[13] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2119–2128.

[14] Kate Moran. 2020. *How People Read Online: New and Old Findings.* Retrieved January 16, 2021 from https://www.nngroup.com/articles/how-people-read-online/

[15] Jakob Nielsen. 2001. *Error Message Guidelines.* Retrieved January 16, 2021 from https://www.nngroup.com/articles/error-message-guidelines/

[16] Jakob Nielsen. 2010. *Mental Models.* Retrieved January 16, 2021 from https://www.nngroup.com/articles/mental-models/

[17] Jakob Nielsen. 2011. *Workflow Expectations: Presenting Steps at the Right Time.* Retrieved January 16, 2021 from https://www.nngroup.com/articles/workflow-expectations/

[18] Alexander Renkl. 2014. Toward an instructionally oriented theory of example-based learning. *Cognitive science* 38, 1 (2014), 1–37.

[19] Ruhi Sarikaya. 2017. The technology behind personal digital assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine* 34, 1 (2017), 67–81.

[20] Agnieszka Sienkiewicz. 2021. *11 Chatbot Statistics and Trends You Need to Know in 2021.* Retrieved January 16, 2021 from https://www.tidio.com/blog/chatbot-statistics/

[21] Arjun Srinivasan, Mira Dontcheva, Eytan Adar, and Seth Walker. 2019. Discovering natural language commands in multimodal interfaces. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 661–672.

[22] Kai Sun, Seungwhan Moon, Paul Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, and Claire Cardie. 2020. Adding Chit-Chats to Enhance Task-Oriented Dialogues. *arXiv preprint arXiv:2010.12757* (2020).

[23] Justin D Weisz, Mohit Jain, Narendra Nath Joshi, James Johnson, and Ingrid Lange. 2019. BigBlueBot: teaching strategies for successful human-agent interactions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 448–459.

[24] Jason D Williams, Nobal B Niraula, Pradeep Dasigi, Aparna Lakshmiratan, Carlos Garcia Jurado Suarez, Mouni Reddy, and Geoff Zweig. 2015. Rapidly scaling dialog systems with interactive learning. In *Natural language dialog systems and intelligent assistants*. Springer, 1–13.