

Influence of user personality on dialogue task performance: A case study using a rule-based dialogue system

Ao Guo¹, Atsumoto Ohashi¹, Ryu Hirai²
Yuya Chiba³, Yuiko Tsunomori¹, and Ryuichiro Higashinaka¹

¹Graduate School of Informatics, Nagoya University

²School of Informatics, Nagoya University, ³NTT Communication Science Laboratories
{guo.ao@a, ohashi.atsumoto@g, hirai.ryu@a, tsunomori.yuiko@b}
.mbox.nagoya-u.ac.jp, yuuya.chiba.ax@hco.ntt.co.jp
higashinaka@i.nagoya-u.ac.jp

Abstract

Endowing a task-oriented dialogue system with adaptiveness to user personality can greatly help improve the performance of a dialogue task. However, such a dialogue system can be practically challenging to implement, because it is unclear how user personality influences dialogue task performance. To explore the relationship between user personality and dialogue task performance, we enrolled participants via crowdsourcing to first answer specified personality questionnaires and then chat with a dialogue system to accomplish assigned tasks. A rule-based dialogue system on the prevalent Multi-Domain Wizard-of-Oz (MultiWOZ) task was used. A total of 211 participants' personalities and their 633 dialogues were collected and analyzed. The results revealed that sociable and extroverted people tended to fail the task, whereas neurotic people were more likely to succeed. We extracted features related to user dialogue behaviors and performed further analysis to determine which kind of behavior influences task performance. As a result, we identified that average utterance length and slots per utterance are the key features of dialogue behavior that are highly correlated with both task performance and user personality.

1 Introduction

Personality is the characteristics of behavior, cognition, and emotional patterns that evolve from biological and environmental factors (Corr and Matthews, 2020). As a fundamental difference between individuals, personality has been found to influence the performance of task-oriented dialogue systems, such as the user satisfaction to dialogue systems (Cassell and Bickmore, 2003). Endowing a task-oriented dialogue system with adaptiveness to user personality is therefore essential to achieve better dialogue task performance.

In a previous study, Mairesse and Walker analyzed the differences between introverts and extro-

verts in dialogue behavior, such as extroverts being more talkative and using fewer words per utterance than introverts (Mairesse and Walker, 2007). They further designed a parameterized natural language generation system named "PERSONAGE." Increasing applications of personality have been put forward on a dialogue system that is able to generate natural, coherent, and personalized utterances (Zheng et al., 2020). However, two problems remain that hinder the dialogue system adapting to user personality to achieve better task performance: (1) what kind of personality traits are correlated with task performance and (2) how the personality influences the task performance.

In this work, we performed a case study analysis of user personality and task performance using a rule-based dialogue system on the prevalent Multi-Domain Wizard-of-Oz (MultiWOZ) task (Budzianowski et al., 2018; Eric et al., 2020). Specifically, we enrolled participants via crowdsourcing to first take several personality questionnaires and then interact with the dialogue system via text chat to accomplish assigned tasks. Participants' personality and dialogue data were collected. We then analyzed the correlation between personality and task performance to identify the personalities highly correlated with task performance. Considering that the task performance largely depends on the user's dialogue behavior, we further examined the main features of dialogue behavior influenced by the identified personalities.

For data collection, we selected four mainstream personality questionnaires that could possibly influence the dialogue task performance and use a rule-based dialogue system for the MultiWOZ task (Zhu et al., 2020). In total, 211 participants were enrolled, and their personalities and 633 dialogues were collected and analyzed. Five task performance metrics are used.

Our contributions are two-fold:

- The correlation of user personality and task

performance is clarified by performing a large-scale data collection experiment and analysis.

- By revealing the correlation between personality and dialogue behavior, the influence of user personality on task performance is clarified.

The remainder of this paper is organized as follows. The related work on personality and dialogue systems is presented in the next section. Section 3 describes the research approach and analysis procedures. The data collection and data analysis are presented in Sections 4 and 5. Conclusions and future work are addressed in the last section.

2 Related Work

Related studies can be classified into: (1) the exploration of correlations between personality and dialogue system performance, (2) the application of personality to a dialogue system, and (3) user-adaptive dialogue system.

In the past decades, [Nass and Lee \(2000\)](#) explored the differences between extroverts and introverts in speaking, such as pitch and volume. [Cassell and Bickmore \(2003\)](#) further identified that extroverts feel extroverted voices are generally more credible, whereas introverts feel the opposite. Conscientiousness (the tendency to act in an organized or thoughtful way) and agreeableness (the tendency to be compassionate and cooperative toward others) have been found to be important factors influencing the dialogue in customer service ([Blignaut et al., 2014](#); [Sackett and Walmsley, 2014](#)). [Qiu et al. \(2017\)](#) examined the consistency of the correlation between personality and word usage in different languages and found that such correlation is independent of language. [Wang et al. \(2020\)](#) recently found that people with different personalities use different persuasion strategies, which can potentially be used by a personalized persuasion dialogue system. Although numerous studies have found that personality is more or less related to dialogue behaviors, few studies have comprehensively investigated the influences of user personality on dialogue task performance.

As for applying personality to a dialogue system, [Mairesse and Walker \(2007\)](#) applied extraversion as the parameter of a natural language generation system, so as to generate personalized and natural dialogue. Specifically, they used an extraversion parameter to control the communication content

and the template sentence with different syntactic styles. [Qian et al. \(2018\)](#) applied a profile as personality into an encoder-decoder dialogue system to increase the coherence of generated utterances. [Zhang et al. \(2018\)](#) proposed a dataset named “Persona-Chat” containing a corpus with a corresponding persona description to help build an end-to-end chat dialogue system with consistent personality and coherent utterance generation.

Towards a more user-adaptive dialogue system, several researchers have investigated user-adaptive dialogue systems. [Komatani et al. \(2003\)](#) focused on a system that generates the responses to users’ information queries in accordance with their skill with regard to the system, domain-related knowledge, and the degree of hastiness. [Dohsaka et al. \(2010\)](#) developed a dialogue system that adapts to users’ behavior regarding pause duration preceding system utterance and gaze to reduce user’s discomfort during conversation. To increase user engagement, [Liang et al. \(2020\)](#) proposed a system that is able to select a chat topic in accordance with user interests.

The aim of our research is to find out the influences of personality on task performance to benefit further development of a dialogue system that can adapt to user personality to achieve better dialogue task performance.

3 Approach

To explore the correlation between user personality and dialogue task performance, a dataset is needed that contains a variety of user personality traits (e.g., extraversion), dialogue data, and the corresponding dialogue task performances (e.g., task success rate).

The data to be analyzed is collected via crowdsourcing, and various features of dialogue behavior related to task feature, utterance feature, and action/utterance intent are extracted from the dialogue data.

After collecting a sufficient amount of data, user personality and task performance are explored in three steps: (1) correlation analysis of user personality and dialogue task performance to identify the personality traits highly correlated with task performance; (2) correlation analysis of behavior feature and dialogue task performance to determine their relevance; and (3) analysis of highly correlated personalities, behaviors, and task performances to figure out how a person’s personality influences

task performance through dialogue behavior.

4 Data Collection

We designed a data collection task to obtain user personality and dialogue data on the Amazon Mechanical Turk (AMT), a Human-Intelligence Task (HIT) crowdsourcing platform. The task has been approved by the ethical review of the relevant organization.

4.1 Description of HIT

Our HIT consisted of two main parts: personality questionnaire and conversation with a dialogue system. Each enrolled participant (Turker) followed a three-step procedure to chat with the dialogue system:

1. The Turker carefully reads the description of the assigned task before starting the conversation.
2. The Turker chats with the dialogue system within 30 turns and tries his/her best to finish the assigned task; once the Turker has covered all the items and accomplished the tasks, he/she sends a “success” message to the dialogue system to finish the current conversation.
3. The Turker rates the performance of the dialogue system on the metrics of language understanding, response appropriateness, and satisfaction with the dialogue system.

The eligible Turkers should have (1) a HIT accomplishment number greater than 100, (2) a HIT approval rate greater than 95%, and (3) have passed our designed qualification test. The qualification test contains two common-sense questions written in English aiming at selecting English native speakers. Each Turker who finished our HIT received \$10 as a reward.

4.2 Description of Personality Questionnaire

Some personality recognition tools can be adopted to measure the user personality directly from their dialogue text, e.g., IBM Watson’s personality insights (Mostafa et al., 2016), but these tools can only measure a few dimensions of personality with lower accuracy than personality questionnaires. Therefore, all users’ personalities were measured through personality questionnaires in this research.

Category	Item
Gender	• Male • Female • Other
Age	• ~19 • 30~39 • 50~59 • 70~ • 20~29 • 40~49 • 60~69
Education	• Less than four-year college • Four-year college • Postgraduate
Employment status	• Out of work • Employed • Homemaker • Student • Unable to work • Retired
How often use a dialogue system	• Never • Rarely • Sometimes • Often • Very often • Always

Table 1: Demographic categories

To cover a wide spectrum of user personality dimensions, we prepared four prevalent personality questionnaires considering their potential correlations with task performance.

- Inclusion of Others in the Self (IOS) Scale: IOS corresponds to how close the respondent feels to another person or group (Aron et al., 1992). The IOS scales from 1 to 7 in terms of closeness.
- Big Five Personality Traits (BF-44): BF-44 contains 44 questions to identify five common parameters of human personality: extraversion, neuroticism, agreeableness, conscientiousness, and openness to experience (Raad, 2000).
- Kikuchi’s Scale of Social Skills (KISS-18): KISS-18 contains 18 questions to identify six social skills (Takahashi et al., 2013): basic skill, advanced skill, emotional management skill, offense management skill, stress management skill, and planning skill.
- Adult Temperament Questionnaire (ATQ): ATQ contains 77 questions to measure 13 user temperaments with regard to four dimensions: negative affect, extraversion, effortful control, and orienting sensitivity (Evans and Rothbart, 2007).

User’s demographic information regarding age, gender, final education, occupation, and dialogue system usage frequency were also collected, as presented in Table 1.

Domain	Info Slot	Request Slot
Restaurant	Area : South	Address Phone
	Food : Indian Price Range : Expensive	
Hotel	Name : Gonville Hotel	Area Postcode
		Address Phone
Attraction	Area : Centre	Entrance Fee Address Phone
	Type : Theatre	
Taxi	Departure : Hotel	Phone Car Type
	Destination : Attraction Leave At : 12:30	

Table 2: An example task with very hard difficulty. Note that these pieces of information are presented to users in the form of a textual scenario so as to elicit natural interaction from users.

4.3 Description of Dialogue System

We utilized ConvLab-2, an open-source toolkit by [Zhu et al. \(2020\)](#), in order to build the system for data collection. The best performing rule-based dialogue system built by the ConvLab-2 on MultiWOZ 2.1 task ([Eric et al., 2020](#)) was selected. The rule-based dialogue system consists of four modules in a pipeline structure: joint-BERT as natural language understanding (NLU) ([Zhang et al., 2019](#)), rule-based dialogue state tracking, rule-based dialogue policy, and template-based natural language generation (NLG). For the system to smoothly chat with the users, we further (1) limited the utterance to contain a maximum of three sentences and (2) revised the grammar mistakes and a few inappropriate texts of the NLG module’s template utterances.

The MultiWOZ task requires the Turkers to play the role of a tourist searching for information about Cambridge, UK, and the dialogue system acts as an information desk clerk to help the tourist plan their tour. Four domains of MultiWOZ were selected to generate the dialogue tasks, i.e., hotel, restaurant, train, and attraction. Each domain contains five slots regarding user request. Below is an example task description on the restaurant domain:

“You are looking for a restaurant called ‘Mahal of Cambridge’. Once you find the restaurant, make sure you get its address and phone number. You want to book a table for 19:00 on Tuesday.”

The difficulty of the task should be balanced to ensure a task success rate per dialogue of about 50% because the influences of user personality on task performance will be revealed less if the task is too easy or too hard for the majority of Turkers to accomplish. To balance the difficulty to ensure

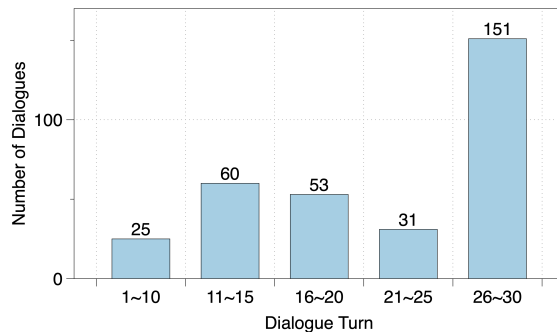


Figure 1: Distribution of successful dialogue turns

Dataset Statistics	
Participants	211
Dialogues	633
Avg. questionnaire time	905s
Avg. dialogue time	583s
Avg. task time	2654s
Avg. slots per dialogue	14.8
Avg. words per utterance	7.9
Avg. time per utterance	21s
Success rate per dialogue	51%
Avg. satisfaction score	2.3/5
Avg. understanding score	2.6/5
Avg. appropriateness score	2.7/5
Total unique tokens	2281

Table 3: Statistics of collected data

a 50% success rate per dialogue, we performed a small scale test with four levels of difficulty by changing the number of domains and corresponding slots: easy (1 domain), normal (2 domains), hard (3 domains), and very hard (4 domains). As a result, we selected the task with the very hard level, which consists of four domains with 20 slots in total. We designed three very hard tasks, one of which is shown in Table 2. We use the three tasks in the following data collection. We use this fixed set of tasks to examine behaviors/task performance with the same stimuli.

4.4 Data Statistics

The statistics of the collected data are presented in Table 3. In total, 211 participants were enrolled in our tasks, and 633 dialogues were collected. We observed that the average task (including four personality questionnaires and three rounds of conversation with the same level of difficulty) took about 44 minutes, and the success rate per dialogue was about 51%. We present the distribution of dialogue turns only from successful dialogues, as shown in Figure 1. We observed that there are two types of users: one kind of user finished the dialogue tasks rapidly (within 11 to 15 turns), whereas the

	Task Success Rate	Slot Filling Rate	Satisfaction	Understanding	Appropriateness
Education	.01	.06	-.07	-.13*	-.10
Usage Frequency	-.14*	-.10	.02	.00	-.10
IOS	-.07	-.09	.16*	.15*	.08
Extraversion	-.10	-.13*	.07	.03	-.02
Sociability	-.08	-.10	.06	.06	.01
Neuroticism	.06	.12*	-.01	.04	.02
Sadness	.06	.14*	.07	.11	.11
Neutral Perceptual Sensitivity	.00	.13*	.00	.06	.08

* $p < .05$

Table 4: Correlation between user personality and task performance (selected by $r \geq 0.1$)

other kind finished the task slowly (within 25 to 30 turns).

5 Data Analysis

The Pearson correlation coefficient r (range from -1 to 1) is used for analysis. We first extracted a variety of dialogue behaviors and task performances to be analyzed. We then analyzed the correlation between user personalities and task performances. We further clarified the influences of user personalities on task performances through dialogue behaviors that are highly correlated with personalities and task performance.

5.1 Data Processing

The user dialogue behavior and task performance parameter are necessary for our analysis. The dialogue behaviors extracted from the task data were categorized into three aspects:

- **Task Feature:** Task feature contains questionnaire time, dialogue time, average utterance time, and average slots covered per utterance.
- **Utterance Feature:** Utterance feature contains average utterance length, the richness of word token variability (reflecting vocabulary), and part-of-speech tags.
- **Action Intent:** Action intent refers to user intent that is recognized by the dialogue system from user utterances. In the rule-based dialogue system, the action intents are classified into greet, inform, request, recommend, request more, thank, and say goodbye.

The values to represent task performance are categorized into two aspects:

- **Objective Task Performance:** Objective task performance covers the task success rate and slot filling rate $rate_{slot}$ as formulated below:

$$rate_{slot} = \frac{\text{num}_{filled_slot}}{\text{num}_{total_slot}} \quad (1)$$

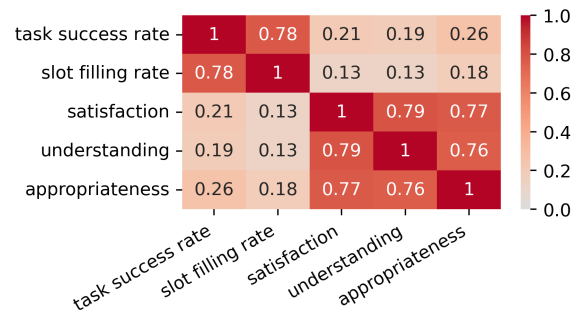


Figure 2: Correlation heatmap of task performances

where num_{filled_slot} denotes the slots user filled in three rounds of tasks, and num_{total_slot} implies the total slots of three rounds of tasks (60 slots in our tasks).

- **Subjective Task Performance:** Subjective task performance contains the values of the evaluation metrics of dialogue system language understanding, appropriateness of the generated utterances, and user satisfaction with the dialogue system. The values scale from 1 (strongly disagree) to 5 (strongly agree).

To compare different task performance parameters, a heatmap is provided in Figure 2, in which we can see that there is a high correlation between task success rate and slot filling rate, and a high correlation between satisfaction, understanding, and appropriateness scores. However, the objective and subjective task performance values do not show a high correlation.

5.2 Analysis between User Personality and Task Performance

To identify the user personalities highly correlated with task performance, we performed the correlation analysis between user personality and five task performances, as presented in Table 4. The five task performances are task success rate, slot filling

	Task Success Rate	Slot Filling Rate	Satisfaction	Understanding	Appropriateness
Slots per Utt.	.70**	.72**	.29**	.30**	.33**
Avg Utt. Length	.28**	.41**	.01	.12	.11
User Vocabulary Request	-.14*	-.02	-.18**	-.16*	-.21**
Recommend	.58**	.58**	.23**	.17*	.19**
	-.16*	-.35**	.21**	.16*	.16*

* $p < .05$, ** $p < .01$

Table 5: Correlation between dialogue behavior and task performance (selected by $r \geq 0.1$)

rate, and average task performance over three dialogues with regard to satisfaction, understanding, and appropriateness. Only those user personalities correlated to task performances with correlation value r no less than 0.1 are shown in Table 4. User personality showed a weak correlation with task performance; the highest correlation was determined between the IOS score and the satisfaction with the dialogue system, reaching .16 (significance level $p < .05$).

Table 4 shows that: (1) the higher a person’s final education level, the lower he/she rates the response appropriateness of the dialogue system; (2) the more frequently a person uses the dialogue system, the less likely he/she is to succeed in the task; (3) extroverted and sociable people tend to fail the task; and (4) neurotic and sensitive people who often feel sad are more likely to succeed in the task. The correlation could be categorized into intuitive correlations (e.g., highly educated people are unsatisfied with the dialogue system) and counter-intuitive correlations (e.g., extroverted and sociable people tend to fail the task). User personality and task performance are further analyzed in view of dialogue behavior in the next subsections to better understand the intuitive and counter-intuitive correlations.

5.3 Analysis between Dialogue Behavior and Task Performance

As shown in Table 5, the selected dialogue behavior exhibited a moderate or strong correlation with the task performance. The correlation coefficient r between slots per utterance and filled slots reached .72 ($p < .01$). The request intent (e.g., “What is the price for a ticket?”) and recommend intent (e.g., “How about Lensfield hotel?”) recognized by the dialogue system are highly correlated with filled slots.

To better illustrate the correlation between dialogue behavior and task performance, comparisons among average utterance time, slots per utterance, and user utterance length depending on the number

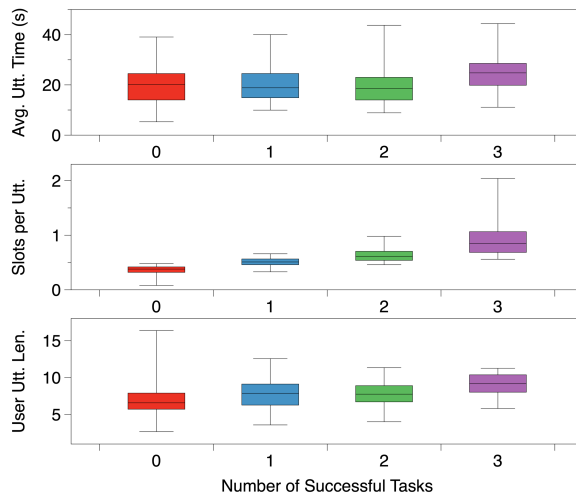


Figure 3: Average utterance time, slots per utterance and user utterance length on number of successful tasks

of successful tasks are plotted in Figure 3. The plot-box shows the maximum, 75 percentile, average, 25 percentile, and the minimum value. Apparently, the users who succeeded in more tasks have an upward trend in the parameters of the plot-box, which shows that their average utterance time and user utterance length are longer, and their slots per utterance are larger.

5.4 Analysis of User Personality and Task Performance through Dialogue Behavior

Since a set of dialogue behaviors was identified to be highly correlated with dialogue performance, we further explored the influences of user personality on task performance through dialogue behavior. The dialogue behaviors highly correlated with those personalities ($r \geq 0.1$) and task performances ($r \geq 0.2$) in Table 4 are selected in the subsequent analysis.

The selected personality, dialogue behavior, task performance, and correlations of $corr_{ps,be}$, and $corr_{be,pf}$ are shown in Table 6. $corr_{ps,be}$ and $corr_{be,pf}$ are correlations of dialogue behavior respectively with personality and task performance. The table shows that: (1) people with higher final education levels use a more enriched vocabulary

Personality	<i>CORT_{ps.be}</i>	Behavior	<i>CORT_{be.pf}</i>	Performance
education	.13*	vocabulary	-.21**	appropriateness
usage frequency	-.10	slots per utt	.42**	task success rate
sociability	-.13	slots per utt	.44**	slot filling rate
extraversion	.15*	recommend	-.35**	slot filling rate
neuroticism	.11	average utt length	.41**	slot filling rate
sadness	.14*	average utt length	.41**	slot filling rate
neutral perceptual sensitivity	.13*	average utt length	.41**	slot filling rate
neutral perceptual sensitivity	.10	slots per utt	.44**	slot filling rate

* $p < .05$, ** $p < .01$

Table 6: Associations among personality, dialogue behavior and task performance

and thereby tend to negatively evaluate the appropriateness of the dialogue system generating utterances by fixed templates; (2) people who usually use a dialogue system or are social or insensitive to external stimuli use utterances containing fewer slots in our designed task, which thus results in failure of the task; (3) extroverted people are more likely to use sentences like “*How about*” (with an underlying recommendation intent); and (4) neurotic and sensitive people who often feel sad tend to use longer utterances that are likely to contain more slots.

6 Conclusion and Future Work

We presented a case study analysis using a rule-based system on the Multi-Domain Wizard-of-Oz (MultiWOZ) task to explore the correlation between user personality and task performance. We identified the user personalities highly correlated with task performances and found that sociable and extroverted people tended to fail the task, while neurotic people were more likely to succeed. We further selected user personalities, dialogue behaviors, and task performance having high correlations with each other to understand the correlation of personality and task performance in view of dialogue behavior. As a result, we clarified that average utterance length and slots per utterance are the two key dialogue behaviors influenced by the personality, affecting the task performance in the case of our data collection using MultiWOZ.

Our work is an initial step to enable a dialogue system to adapt to user personality. Since the analyzed correlation results are based on a rule-based system, more dialogue systems, e.g., end-to-end dialogue systems such as SimpleTOD (Hosseini-Asl et al., 2020) and UBAR (Yang et al., 2020), and different task domains will need to be considered for the analysis to derive more general correlations. The effect of personality and task performance should also be qualitatively explored to better inter-

pret the correlations we observed in the experiment. Furthermore, the dialogue system adapting to user personality should be implemented to ascertain its dialogue task performance.

7 Acknowledgments

Funding was provided by a Grant-in-Aid for Scientific Research (Grant No. JP19H05692).

References

- Arthur Aron, Elaine N Aron, and Danny Smollan. 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, 63:596–612.
- Linda Blignaut, Leona M Ungerer, and Helene Muller. 2014. Personality as predictor of customer service centre agent performance in the banking industry: An exploratory study. *South African Journal of Human Resource Management*, 12:16.
- Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Inigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. MultiWOZ: A large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. *arXiv preprint arXiv:1810.00278*.
- Justine Cassell and Timothy Bickmore. 2003. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modelling and User-Adapted Interaction*, 13:89–132.
- Philip J Corr and Gerald Matthews. 2020. *The Cambridge handbook of personality psychology*. Cambridge University Press.
- Kohji Dohsaka, Atsushi Kanemoto, Ryuichiro Higashinaka, Yasuhiro Minami, and Eisaku Maeda. 2010. User-adaptive coordination of agent communicative behavior in spoken dialogue. In *Proceedings of the SIGDIAL 2010 Conference*, pages 314–321.
- Mihail Eric, Rahul Goel, Shachi Paul, Adarsh Kumar, Abhishek Sethi, Anuj Kumar Goyal, Peter Ku, Sanchit Agarwal, Shuyang Gao, and Dilek Hakkani-Tür. 2020. MultiWOZ 2.1: A consolidated multi-domain dialogue dataset with state corrections and

- state tracking baselines. In *Proceedings of the 12th International Conference on Language Resources and Evaluation*, pages 422–428.
- David E. Evans and Mary K. Rothbart. 2007. Developing a model for adult temperament. *Journal of Research in Personality*, 41:868–888.
- Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. 2020. A simple language model for task-oriented dialogue. *arXiv preprint arXiv:2005.00796*.
- Kazunori Komatani, Shinichi Ueno, Tatsuya Kawahara, and Hiroshi G Okuno. 2003. Flexible guidance generation using user model in spoken dialogue systems. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 256–263.
- Kaihui Liang, Austin Chau, Yu Li, Xueyuan Lu, Dian Yu, Mingyang Zhou, Ishan Jain, Sam Davidson, Josh Arnold, Minh Nguyen, et al. 2020. Gunrock 2.0: A user adaptive social conversational system. *arXiv preprint arXiv:2011.08906*.
- François Mairesse and Marilyn Walker. 2007. Personage: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 496–503.
- Mohamed Mostafa, Tom Crick, Ana C Calderon, and Giles Oatley. 2016. Incorporating emotion and personality-based analysis in user-centered modelling. In *Proceedings of the International conference on innovative techniques and applications of artificial intelligence*, pages 383–389.
- Clifford Nass and Kwan Min Lee. 2000. Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 329–336.
- Qiao Qian, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Assigning personality/profile to a chatting machine for coherent conversation generation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 4279–4285.
- Lin Qiu, Jiahui Lu, Jonathan Ramsay, Shanshan Yang, Weina Qu, and Tingshao Zhu. 2017. Personality expression in Chinese language use. *International Journal of Psychology*, 52(6):463–472.
- Boele De Raad. 2000. *The big five personality factors: The psycholexical approach to personality*. Hogrefe & Huber Publishers.
- Paul R Sackett and Philip T Walmsley. 2014. Which personality attributes are most important in the workplace? *Perspectives on Psychological Science*, 9:538–551.
- Junichi Takahashi, Koju Tamaki, Nozomi Yamawaki, et al. 2013. Autism spectrum, attachment styles, and social skills in university student. *Creative Education*, 4:514–520.
- Xuwei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2020. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 5635–5649.
- Yunyi Yang, Yunhao Li, and Xiaojun Quan. 2020. UBAR: Towards fully end-to-end task-oriented dialog systems with GPT-2. *arXiv preprint arXiv:2012.03539*.
- Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, page 2204–2213.
- Zhichang Zhang, Zhenwen Zhang, Haoyuan Chen, and Zhiman Zhang. 2019. A joint learning framework with BERT for spoken language understanding. *IEEE Access*, 7:168849–168858.
- Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. 2020. A pre-training based personalized dialogue generation model with persona-sparse data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9693–9700.
- Qi Zhu, Zheng Zhang, Yan Fang, Xiang Li, Ryuichi Akanobu, Jinchao Li, Baolin Peng, Jianfeng Gao, Xiaoyan Zhu, and Minlie Huang. 2020. ConvLab-2: An open-source toolkit for building, evaluating, and diagnosing dialogue systems. *arXiv preprint arXiv:2002.04793*, pages 142–149.