A Wizard of Oz Data Collection Framework for Internet of Things Dialogues

Carla Gordon **Volodymyr Yanov** **David Traum** **Kallirroi Georgila**
Institute for Creative Technologies, University of Southern California
12015 Waterfront Drive, Los Angeles, CA 90094-2536, USA
{cgordon|yanov|traum|kgeorgila}@ict.usc.edu

Abstract

We describe a novel Wizard of Oz dialogue data collection framework in the Internet of Things domain. Our tool is designed for collecting dialogues between a human user, and 8 different system profiles, each with a different communication strategy. We then describe the data collection conducted with this tool, as well as the dialogue corpus that was generated.

1 Introduction

The Internet of Things (IoT) refers to a network of physical devices which are connected to the Internet, and can perform services and provide information to satisfy remote requests. We describe a novel Wizard of Oz (WOz) tool that can be used to investigate several questions relating to how users could communicate in natural language with a Virtual Home Assistant (VHA) that is connected to IoT devices. The tool is designed to address several issues for this kind of dialogue, including relaying aspects of the environmental context to the user and testing different communication styles.

When interacting with a VHA, a user will typically be inside a home and will know what room they are in, what devices exist, and may be able to see or hear changes in a device's state if they are in the same room. Therefore, in order to generate realistic dialogues in this domain, there needs to be some environmental context provided to the user. Our WOz tool allows us to provide this context (see Figure 1).

Additionally, previous analysis of observer ratings of IoT dialogues authored by linguists (Georgila et al., 2018; Gordon et al., 2018) suggested several features of VHA interaction that may affect user satisfaction. The tool allows us to define system profiles each with a different communication style (with arbitrary system names shown to the user).

2 System Profiles

We examine 3 binary system behavior features: Register (Direct, e.g.,"Thank you.", or Conversational, e.g.,"Thanks, it's my pleasure to help."), Explicitness (Explicit, e.g., "I've turned on the light in the kitchen.", or Implicit, e.g., "Your request has been taken care of."), Errors (misunderstandings exist or not). Combining these 3 features leads to 8 different system profiles. The wizard interface allows the wizard to toggle between profiles, each with a different set of utterances that conform to the system behavior features.

3 The WOz Tool

The WOz tool includes 3 different views: one for the user, and two for the wizard. The user view can be seen in Figure 1. It shows 3 rooms (Bedroom, Kitchen, Living Room) and indicates what room the participant is currently in. In each room all devices in that room are displayed, along with information about the current device state (on/off), as well as other relevant information for that device (e.g., current channel, temperature, volume).

Below the display of rooms and devices is the text chat window in which users can enter their commands to the system. A log of all system and user utterances is available to the user at all times when interacting with the interface. The system profile identities are displayed to the user in the upper right corner of the user interface and by the name in the log (the monkey in Figure 1), but no explicit description is given of the features associated with the profile. The wizard interface includes one view containing buttons to communicate with the user and change the state of the devices, and a second view with information on the state of all the devices.

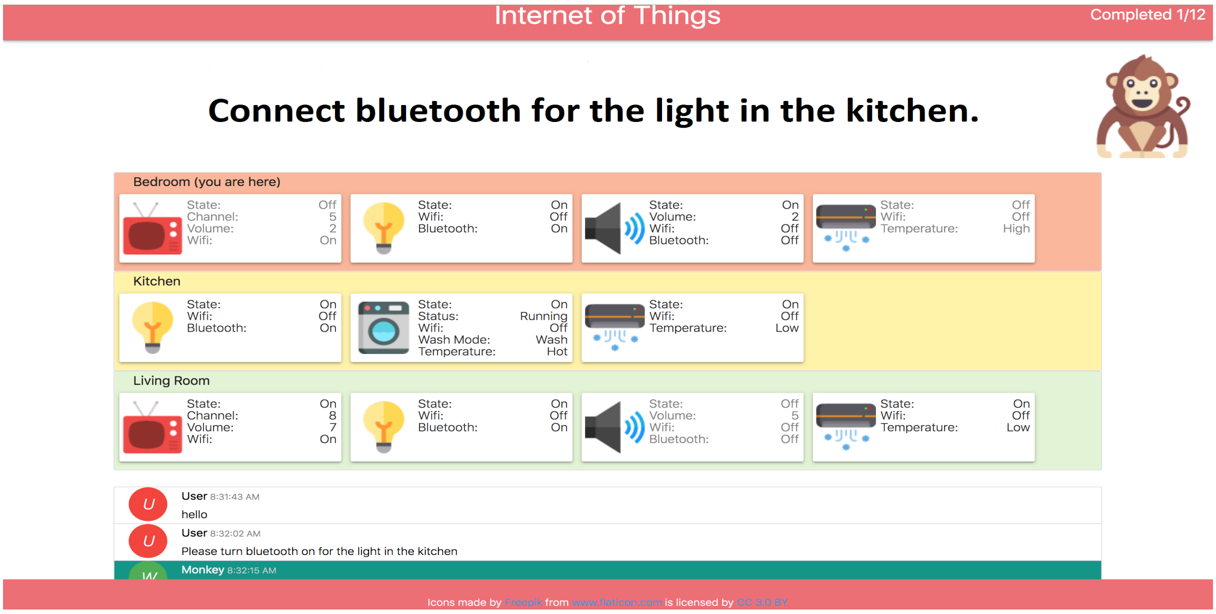

Figure 1: The user interface view

4 Data Collection

Users were recruited via Craigslist (`https://www.craigslist.org`), and interaction with our WOz tool took place on a Macbook Pro laptop, in a room without any experimenters present. After filling out a short survey, participants typed commands in text to accomplish 12 different tasks such as turning on a light. Each user interacted with only 4 of the 8 system profiles (3 dialogues per profile). Users had to accomplish their task before they could move on to another task; if a user tried to advance to the next task before accomplishing the current one, the wizard would send a message that the task was not complete.

The WOz tool provided the environmental context necessary for users to understand whether a task had been completed correctly, even without explicit confirmation from the system, as exemplified by Table 1. The system initially turns on the AC in the wrong room, but because of the context provided to the user they were able to see the AC in the kitchen was not turned on like they asked, and repeated their command to complete the task.

Users were asked to fill out a post-task Likert-scale survey designed to measure the level of user satisfaction, as well as whether the user perceived the system as friendly, natural, intelligent, or as having a personality. There was also a post-experiment survey administered once the user had completed all tasks, in which users were asked to rank the 4 systems they interacted with from best to worst.

This data collection yielded a corpus of 216 human-system dialogues (18 users, 12 dialogues per user, 27 dialogues per system profile), each with accompanying Likert-scale survey data. This corpus will be used to further investigate the interaction between these subjective features, the system behaviors, and the user's overall satisfaction.

Actor	Text	Action
User	Please turn on the air conditioner in the kitchen.	
System	Ok, I will do that.	AC on in Bedroom
System	It has been done.	
User	Please turn on the air conditioner in the kitchen.	
System	I'm sorry, I will.	AC on in Kitchen
User	Thank you.	
System	You're welcome.	

Table 1: Environmental context example

Acknowledgments

This work was supported by Samsung Electronics Co., Ltd. and the U.S. Army. Any statements or opinions expressed in this material are those of the authors and do not necessarily reflect the policy of the U.S. Government, and no official endorsement should be inferred.

References

Kallirroi Georgila, Carla Gordon, Hyungtak Choi, Jill Boberg, Heesik Jeon, and David Traum. 2018. Toward low-cost automated evaluation metrics for Internet of Things dialogues. In *Proceedings of the 9th International Workshop on Spoken Dialogue Systems Technology (IWSDS)*, Singapore.

Carla Gordon, Kallirroi Georgila, Hyungtak Choi, Jill Boberg, and David Traum. 2018. Evaluating subjective feedback for Internet of Things dialogues. In *Proceedings of the 22nd Workshop on the Semantics and Pragmatics of Dialogue (SemDial:AixDial)*, Aix-en-Provence, France.