slurk – A Lightweight Interaction Server For Dialogue Experiments and Data Collection

David Schlangen, Tim Diekmann, Nikolai Ilinykh, Sina Zarrieß
Dialogue Systems Group // CITEC // Linguistics & Literary Studies
Bielefeld University, Germany
first.last@uni-bielefeld.de

1 Introduction

Natural language processing, and artificial intelligence more generally, has seen impressive breakthroughs in recent years. An important factor in this development has been the availability of large labelled data sets such as, in NLP, the Stanford Natural Language Inference Corpus (Bowman et al., 2015) or the Stanford Question Answering Dataset (Rajpurkar et al., 2016),[1] and ImageNet (Deng et al., 2009) in language & vision research. Assembling these dataset, in turn, has been made possible by the availability of large numbers of workers who could be recruited for the annotation tasks, through so-called *crowdsourcing* platforms.

In the subfield of *dialogue modelling* or *conversational AI*, developments have been somewhat slower.[2] There are intrinsic reasons for this—as a discourse-level semantic/pragmatic phenomenon, dialogue is much more domain-specific, and so corpora may generalise less easily; as an interactive phenomenon, the space of possible dialogues is much larger than that of possible word sequences, so that even within a domain a given corpus will still fail to capture much of the possible variation—but also practical ones. One of these is that dialogue requires at least two participants between whom a connection must be established in some way, and the common crowdsourcing platforms do not offer an easy way to achieve this.

Several projects have recently built, for their own specific purposes, software that allows for pairing up of participants (*inter alia*, (Manuvinakurike and DeVault, 2015; Das et al., 2017)), and there has even been a recent effort to generalise this capability (in the "parlAI" architecture (Miller et al., 2017)). We contribute to these efforts by presenting our framework, *slurk*.[3] *slurk* is designed to be modular, to make it possible to realise various different multimodal dialogue tasks. It is available at `https://github.com/dsg-bielefeld/slurk`.

2 Overview of the System

The core of the system is a chat server implemented in Python, on top of the web framework "Flask" and an extension for using websocket connections to clients.[4] Users connect via webbrowser, to which the client application (Javascript) is then delivered. The client shows, as usual for chat tools, a chat history and an input area, but also additionally, a display area that is controlled independently from the chat area (showing an image in Figure 1).

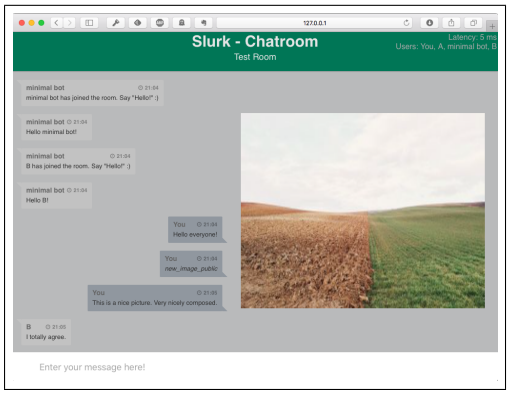

Figure 1: The Chat Client

Conceptually, individual chats happen in *rooms*. In a given room, there can be (an unlimited number of) human participants, and there can also be *bots*. If so desired, a bot

[1] To mention only two recent datasets from one site, and ignoring the role that the availability of large amounts of unannotated text corpora through the world wide web has also played.

[2] But see (Serban et al., 2018) for a recent overview of available dialogue corpora.

[3] As in "Slack™ for mechanical turk"...

[4] `http://flask.pocoo.org`; `https://flask-socketio.readthedocs.io`

can be used to control the interaction, for example by controlling who has the floor, or by controlling what is shown in the display area. The display area can be controlled on a by-user level, displaying different things to different users. (As in Figure 2.)

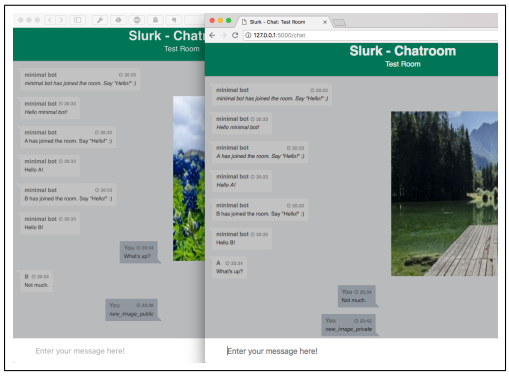

Figure 2: Different image per user

Bots can also move users to other rooms; this, together with a credential mechanism, is how we realise the interface to crowdsourcing platforms and the pairing up. Technically, bots are realised as independent processes connecting via websockets; our example bots are written in Python using the websocket / socket.io client libraries.

So far, we have used the system for a data collection in a setting where the participants play a game together (self citation; under review). They can talk to each other, but also each individually control what they see in the display area, through giving navigation commands to the bot. Their goal is to meet up, i.e., to convince themselves that they are looking at the same image. Figure 3 shows an example of an interaction in this setting, from the perspective of one player. See (Ilinykh et al., 2018) for more details.

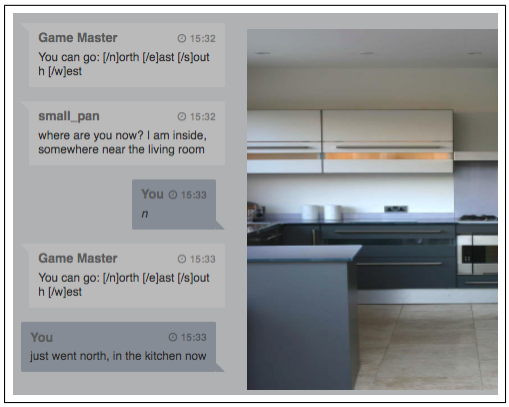

Figure 3: An example task

3 Roadmap

While the system is fully functional in the current state and can be used to collect dialogues involving discussion about (and interaction with) images, development is still ongoing and major new features are planned for the near future. Among these are a plug-in architecture for the display area, which will make it easy to insert any kind of javascript-controlled widget, for example to display a manipulable virtual environment. We are also working on capabilities for streaming audio and for inclusion of (web-based) ASR and TTS. Chat area and input area are already configurable and can be disabled; and in this way, the server will in the next version also serve as the basis for speech interaction experiments.

References

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M.F. Moura, Devi Parikh, and Dhruv Batra. 2017. Visual Dialog. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jia Deng, W. Dong, Richard Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

Nikolai Ilinykh, Sina Zarrieß, and David Schlangen. 2018. The task matters: Comparing image captioning and task-based dialogical image description. In *Proceedings of the International Conference on Natural Language Generation (INLG18)*, Tilburg, Netherlands, November.

Ramesh Manuvinakurike and David DeVault. 2015. Pair Me Up: A Web Framework for Crowd-Sourced Spoken Dialogue Collection. In *Proceedings of IWSDS 2015*, pages 1 –12, Busan, South Korea, January.

Alexander H. Miller, Will Feng, Adam Fisch, Jiasen Lu, Dhruv Batra, Antoine Bordes, Devi Parikh, and Jason Weston. 2017. Parlai: A dialog research software platform. *CoRR*, abs/1705.06476.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100, 000+ questions for machine comprehension of text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*.

Iulian Vlad Serban, Ryan Lowe, Peter Henderson, Laurent Charlin, and Joelle Pineau. 2018. A survey of available corpora for building data-driven dialogue systems: The journal version. *Dialogue and Discourse*, 9(1):1–49.