

# DialPort: A General Framework for Aggregating Dialog Systems

**Tiancheng Zhao and Maxine Eskenazi**

Language Technologies Institute  
Carnegie Mellon University  
{tianchez, max+}@cs.cmu.edu

**Kyusong Lee**

Pohang University of  
Science and Technology  
kyusonglee@postech.ac.kr

## Abstract

This paper describes a new spoken dialog portal that connects systems produced by the spoken dialog research community and gives them access to real users. We introduce a prototype dialog framework that affords easy integration with various remote dialog agents as well as external knowledge resources. To date, the DialPort portal has successfully connected to two dialog systems and several public knowledge APIs. We present current progress and envision our future plan.

## 1 Introduction

Much fundamental research in the spoken dialog domain remains to be done, including adaption for user modeling and management of complex dialogs. In recent years, there has been increasing interest in applying deep learning to modeling the process of human-computer conversation (Vinyals and Le, 2015; Serban et al., 2015; Wen et al., 2016; Williams and Zweig, 2016; Zhao and Eskenazi, 2016). One of the prerequisites for the success of these methods is having a large conversation corpus to train on. In order to advance the research in these uphill areas of study with the state-of-the-art data-driven methods, large corpora of multi-type real user dialogs are needed. At present, few existing large corpora cover a wide set of research domains. It is also extremely difficult for any one group to devote time to collecting and curating a significant amount of real user data. The users must be found and kept interested, and the interface must be created and maintained.

Our proposed solution is DialPort, a data gathering portal that groups various types of dialog systems, gives potential users a variety of interesting applications, and shares the collected data amongst all participating research groups. The connected dialog systems are not simply listed on a website. They are fully integrated into a single virtual agent. From the user’s perspective, DialPort is a dialog system that can provide information in many domains and it becomes increasingly more attractive as new research groups join and resulting more functionalities to discover.

## 2 Challenges

Besides creating new corpora for advanced dialog research, DialPort encounters new research challenges.

- *Advanced Dialog State Representation Learning*: Traditional dialog states are represented as sets of symbolic variables that are related to domain-specific ontology and are tracked by statistical methods (Williams et al., 2013). Such an approach soon becomes intractable if we want to capture all the essential dialog state features within nested multi-domain conversations, such as modeling user preferences and tracking discourse features. DialPort must address this challenge if it is to effectively serve as a portal to many systems.
- *Dialog Policy that Combines Various Types of Agents*: DialPort is powered by multiple dialog agents from research labs around the world. It is different from the traditional sin-

gle dialog agent and requires new methods to develop decision-making algorithms to judiciously switch amongst various systems while creating a homogenous users experience.

- *Dialog System Evaluation with Real Users*: Evaluation has always be challenging for dialog systems because inexpensive methods, (e.g. user simulator or recruited users) are often not accurate. The best evaluation, real users, is costly. DialPort will create streams of real user data, which opens the possibility of developing a principled evaluation framework for dialog systems.

### 3 Proposed Approach

The prototype DialPort system includes the user interface, remote agents/resources and the master agent.

#### 3.1 User Interface

The user interface is the public front-end<sup>1</sup>. The audio interface uses the web-based ASR/TTS to recognize the user’s speech and generate DialPort’s speech output. The visual representation is a virtual agent that has animated embodiments powered by the Unity 3D Engine<sup>2</sup>.

#### 3.2 Remote Agents and Resources

A *Remote agent* is a turn-based dialog system, which inputs the ASR text output of the latest turn and returns the next system response. Every external dialog system connecting to DialPort is treated as a *remote agent*. DialPort also deals with *remote resources*, which can be any external knowledge resource, such as a database of bus schedules. DialPort is in charge of all of the dialog processing and uses the remote resources as knowledge backends in the same way as a traditional goal-oriented SDS (Raux et al., 2005).

#### 3.3 The Master Agent

The *master agent* operates on a set of *remote agents*  $U$ , and a set of *remote resources*  $R$ . In order to serve information in  $R$ , the *master agent* has a set of primitive actions  $P$ , such as *request* or *inform*.

<sup>1</sup><https://skylar.speech.cs.cmu.edu>

<sup>2</sup>[unity3d.com/](http://unity3d.com/)

Together  $P \cup U$  composes the available action set  $A$  for the master agent. The dialog state  $S$  is made up of the entire history of system output and user input and distributions over possible slot values. Given the new inputs from the user interface, the master agent updates its dialog state and generates the next system response based on its policy,  $\pi : S \rightarrow A$ , that will choose the action  $a$  that is the most appropriate. One key note is that for  $a \in U$ , it takes more than one turn to finish a session, i.e. a *remote agent* usually will span several turns with the users, while  $a \in P$  is primitive action that only spans for one turn. Therefore, we formulate the problem as a sequential decision making problem for Semi-Markov Decision Process (SMDP) (Sutton et al., 1999), so  $a \in U$  is equivalent to a macro action. Therefore, when DialPort hands over control to a *remote agent*, the user input is directly forwarded to the remote system until the session is finished by the remote side. Core research on DialPort is about how to construct an efficient representation for  $S$ , and how to learn a good policy  $\pi$ .

#### 3.4 Current Status

To date, the DialPort has connected to two *remote agents*, the dialog system at Cambridge University (Gasic et al., 2015) and a chatbot, and to two *remote resources*: Yelp food API and NOAA (National Oceanic and Atmospheric Administration) Weather API.

### 4 Evaluation

Given the data collected by DialPort, assessment has several aspects. In order to create labeled data for the first two challenges mentioned in Section 2, we developed an annotation toolkit to label the correct system responses and state variables of the dialogs. The labeled data can then be used for new models for advanced dialog state tracking and multi-agent dialog policy learning. We will also solicit subjective feedback from users after a session with the system.

### 5 Travel Funding

Two authors are respectively PhD and postdoctoral students that need travel funding.

## References

- M Gasic, Dongho Kim, Pirros Tsiakoulis, and Steve Young. 2015. Distributed dialogue policies for multi-domain statistical dialogue management. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 5371–5375. IEEE.
- Antoine Raux, Brian Langner, Dan Bohus, Alan W Black, and Maxine Eskenazi. 2005. Lets go public! taking a spoken dialog system to the real world. In *in Proc. of Interspeech 2005*. Citeseer.
- Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2015. Building end-to-end dialogue systems using generative hierarchical neural network models. *arXiv preprint arXiv:1507.04808*.
- Richard S Sutton, Doina Precup, and Satinder Singh. 1999. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1):181–211.
- Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.
- Tsung-Hsien Wen, David Vandyke, Nikola Mrkšić, Milica Gašić, Lina M. Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. 2016. A network-based end-to-end trainable task-oriented dialogue system. *arXiv preprint: 1604.04562*, April.
- Jason D Williams and Geoffrey Zweig. 2016. End-to-end lstm-based dialog control optimized with supervised and reinforcement learning. *arXiv preprint arXiv:1606.01269*.
- Jason Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. 2013. The dialog state tracking challenge. In *Proceedings of the SIGDIAL 2013 Conference*, pages 404–413.
- Tiancheng Zhao and Maxine Eskenazi. 2016. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. *arXiv preprint arXiv:1606.02560*.