Using a Deep Learning Dialogue Research Toolkit in a Multilingual Multidomain Practical Application

Graham Wilcock
CDM Interact, Helsinki, Finland
graham.wilcock@cdminteract.com

Abstract

The demo shows a practical application of an open-source research toolkit developed by University of Cambridge. The toolkit (PyDial) supports research on deep reinforcement learning for multi-domain dialogues. The application (CityTalk) is a spoken dialogue system for robots that give information to tourists about local hotels and restaurants. We had a very positive experience using the toolkit, but in a few areas we decided to do things our own way.

1 Introduction

Recently many academic and industrial research groups have made their research frameworks available as open source, creating opportunities for others to apply these research tools to develop practical applications. This demo presents such an application, which uses PyDial [Ultes *et al.*, 2017], a research toolkit developed at University of Cambridge. PyDial aims to facilitate research on deep reinforcement learning and other deep learning techniques for multi-domain dialogues.

CityTalk [Wilcock, 2018b] is a spoken dialogue system for Nao robots from SoftBank Robotics. The robots give information to tourists about local places of interest such as hotels and restaurants. One feature of CityTalk that is enabled by using PyDial is that the robots can switch domains smoothly during the dialogue. For example, the user may ask about hotels and then continue by asking about restaurants.

Robots are already in everyday use in Japan in department stores and shopping centres. As well as welcoming customers with greetings, the robots can provide information about the departments and shops. However, they cannot easily switch from one topic to another. The most popular robot is Pepper from SoftBank Robotics, which has a tablet computer around its neck. Customers select applications from a touch-screen menu on the tablet, and the robot performs the relevant interaction. After that, if the user wants to know about something else, another application must be selected from the menu.

With CityTalk, the user can simply ask about another topic, and the robot immediately switches to the new topic if it is one of the available domains, without needing a touch-screen menu like the Pepper robots. If the system is unsure which domain the user needs, it tells which domains are available and asks for clarification in a meta-dialogue.

Dialogue systems that can handle multiple topic domains are challenging. They can be based either on encyclopaedic knowledge bases like Wikipedia or on large corpora of dialogue interactions that enable machine learning of QA pairs. For instance WikiTalk [Wilcock, 2012; Jokinen and Wilcock, 2014] uses Wikipedia texts to enable robots to talk fluently about thousands of topics in three languages, and a Japanese chat system [Higashinaka *et al.*, 2016] generates responses based on a large corpus of Twitter sentences.

2 The CityTalk Application

The CityTalk application combines the speech processing modules of the robot with the dialogue processing modules of PyDial. Example interactions can be seen in demo videos at www.cdminteract.com.

The oldest video [Wilcock, 2017] uses two domains from the PyDial toolkit about Cambridge hotels and restaurants. The user first asks about hotels and the robot starts talking using a hotel database. When the user asks about restaurants the robot switches to a restaurant database. The robot asks clarification questions that are appropriate for each domain, for example *Would you like the place to have parking?* for the hotel domain, and *What kind of food would you like?* for the restaurant domain.

Figure 1: Video of an interaction about the new Tokyo domains.

Figure 1 shows a more recent video [Wilcock, 2018a] that uses new domains made by CDM Interact about hotels and restaurants in Tokyo Waterfront. The robot gives information about three hotels in different price ranges. When the user asks about restaurants the robot switches domains and gives

information about several restaurants in different price ranges and serving different types of food.

3 The PyDial Research Toolkit

PyDial is a research toolkit for statistical dialogue systems, developed by Cambridge University Engineering Department (CUED) and made freely available as open source.

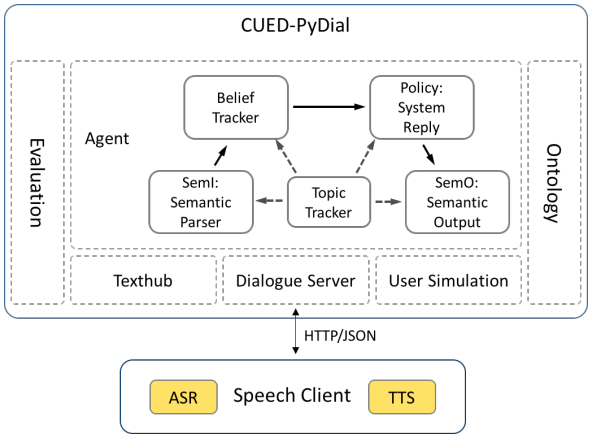

Figure 2: Architecture of the PyDial dialogue research toolkit.

PyDial (Figure 2) includes a dialogue pipeline with four modules that all use deep learning methods. Semantic Parser uses deep neural models for spoken language understanding [Rojas Barahona *et al.*, 2016]. Neural Belief Tracker uses RNNs to update the dialogue state [Mrkšić *et al.*, 2015; 2017]. Policy module uses a committee approach to decide dialogue acts for system replies [Gasic *et al.*, 2015]. Semantic Output uses LSTMs for natural language generation [Wen *et al.*, 2015; 2016]. These four modules provide base classes with functionality for their roles in the dialogue pipeline.

In addition, a Topic Tracker module monitors user utterances to detect when a domain-switch is required. Domain-switching is effected by loading domain-specific instances of the base classes of the four dialogue modules.

The PyDial toolkit provides some example domains such as Cambridge hotels and restaurants, used in the first video [Wilcock, 2017]. New domains can be added by providing an ontology and a database for each new domain. CityTalk has added new domains including Tokyo hotels and restaurants, used in the second video [Wilcock, 2018a].

4 Experiences in Applying a Research Toolkit

Overall, we had a very positive experience using this research toolkit in a practical application. In this section we mention a few areas where we decided to do things in our own way.

For example, to help in creating new domains PyDial provides an ontology tool that automates creating ontology rules from domain databases. However, in practice we found it better to copy existing example ontology files and edit them appropriately, instead of using the ontology tool.

4.1 Implementation Details

The PyDial toolkit supports cutting-edge research, such as GP-SARSA algorithms for deep reinforcement learning [Casanueva *et al.*, 2017]. However, it also provides a baseline set of rule-based methods such as regular expression-based language analysis and template-based language generation. These older components are not the focus of research, and their implementation details have caused a few odd results when used in CityTalk.

For example, in the first video [Wilcock, 2017] the speech recognizer hears *phone number* correctly, but the dialogue system does not reply with the hotel phone number. Later it successfully gives the restaurant phone number. This problem also applies to hotel addresses. It was caused by the regular expressions used for language analysis, and once identified it was easily fixed. In the second video [Wilcock, 2018a] the robot gives hotel addresses and phone numbers.

4.2 Domain Vocabulary for Speech Recognition

PyDial is text-based and does not include speech processing. The architecture in Figure 2 assumes that speech recognition and synthesis are performed by external cloud services connected to PyDial via an HTTP/JSON interface.

Several cloud services do provide open-vocabulary speech recognition, but CityTalk does not use them because open-vocabulary speech recognition services have not yet reached sufficient accuracy for use in practical applications. CityTalk uses the robot's speech recognition and synthesis modules, which allow the speech recognition vocabulary to be changed on-the-fly when domain-switching occurs. This approach was previously used successfully in the WikiTalk system [Wilcock, 2012; Jokinen and Wilcock, 2014].

4.3 Multilingual Localizations

When robots talk with tourists from many different countries, they should be able to speak many languages. Nao robots can now speak about 20 languages, and multilingual localizations of CityTalk are under development using methods similar to those described by Laxström et al. [2017].

One problem in using research toolkits is that they often work only in English. PyDial supports research on new RNN models for language generation, but its older template-based language generator does not use Unicode and assumes there are spaces between words. That works for English but not for Japanese, which does not have spaces between words.

Following experience making the Japanese localization of WikiTalk, which was demonstrated by Wilcock et al. [2016], we have replaced the PyDial template-based generator with a new language generator for CityTalk that is intended to be more suitable for multilingual generation. The new generator is used in the second video [Wilcock, 2018a].

Acknowledgments

The PyDial toolkit was developed by Cambridge University Engineering Department. We thank the Dialogue Systems Group for making it available as open source. We also thank Kristiina Jokinen of AI Research Center, AIST Tokyo Waterfront, Japan for valuable discussions.

References

[Casanueva *et al.*, 2017] Iñigo Casanueva, Paweł Budzianowski, Pei-Hao Su, Nikola Mrkšić, Tsung-Hsien Wen, Stefan Ultes, Lina M. Rojas Barahona, Steve Young, and Milica Gasic. A benchmarking environment for reinforcement learning based task oriented dialogue management. [arXiv], 2017. https://arxiv.org/abs/1711.11023.

[Gasic *et al.*, 2015] Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve Young. Policy committee for adaptation in multi-domain spoken dialogue systems. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Scottsdale, AZ, 2015. IEEE.

[Higashinaka *et al.*, 2016] Ryuichiro Higashinaka, Nozomi Kobayashi, Toru Hirano, Chiaki Miyazaki, Toyomi Meguro, Toshiro Makino, and Yoshihiro Matsuo. Syntactic filtering and content-based retrieval of twitter sentences for the generation of system utterances in dialogue systems. In Alexander Rudnicky, Antoine Raux, Ian Lane, and Teruhisa Misu, editors, *Situated Dialog in Speech-Based Human-Computer Interaction*, pages 15–26. Springer, 2016.

[Jokinen and Wilcock, 2014] Kristiina Jokinen and Graham Wilcock. Multimodal open-domain conversations with the Nao robot. In Joseph Mariani, Sophie Rosset, Martine Garnier-Rizet, and Laurence Devillers, editors, *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialogue Systems into Practice*, pages 213–224. Springer, 2014.

[Laxström *et al.*, 2017] Niklas Laxström, Graham Wilcock, and Kristiina Jokinen. Internationalisation and localisation of spoken dialogue systems. In Kristiina Jokinen and Graham Wilcock, editors, *Dialogues with Social Robots: Enablements, Analyses, and Evaluation*, pages 207–219. Springer, 2017.

[Mrkšić *et al.*, 2015] Nikola Mrkšić, Diarmuid Ó Seaghdha, Blaise Thomson, Milica Gasic, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve Young. Multi-domain dialog state tracking using recurrent neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 794–799, Beijing, China, July 2015. Association for Computational Linguistics.

[Mrkšić *et al.*, 2017] Nikola Mrkšić, Diarmuid Ó Seaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young. Neural belief tracker: Data-driven dialogue state tracking. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 1777–1788, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[Rojas Barahona *et al.*, 2016] Lina M. Rojas Barahona, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, Stefan Ultes, Tsung-Hsien Wen, and Steve Young. Exploiting sentence and context representations in deep neural models for spoken language understanding. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 258–267, Osaka, Japan, 2016. Association for Computational Linguistics.

[Ultes *et al.*, 2017] Stefan Ultes, Lina M. Rojas Barahona, Pei-Hao Su, David Vandyke, Dongho Kim, Iñigo Casanueva, Paweł Budzianowski, Nikola Mrkšić, Tsung-Hsien Wen, Milica Gasic, and Steve Young. PyDial: A Multi-domain Statistical Dialogue System Toolkit. In *Proceedings of ACL 2017, System Demonstrations*, pages 73–78, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[Wen *et al.*, 2015] Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young. Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1711–1721, Lisbon, Portugal, 2015. Association for Computational Linguistics.

[Wen *et al.*, 2016] Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Lina M. Rojas Barahona, Pei-Hao Su, David Vandyke, and Steve Young. Multi-domain neural network language generation for spoken dialogue systems. In *Proceedings of NAACL-HLT 2016*, pages 120–129, San Diego, California, 2016. Association for Computational Linguistics.

[Wilcock *et al.*, 2016] Graham Wilcock, Kristiina Jokinen, and Seiichi Yamamoto. What topic do you want to hear about? A bilingual talking robot using English and Japanese Wikipedias. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, Osaka, Japan, 2016. Association for Computational Linguistics.

[Wilcock, 2012] Graham Wilcock. WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. In *Proceedings of the COLING 2012 Workshop on Question Answering for Complex Domains*, pages 57–69, Mumbai, India, 2012. Association for Computational Linguistics.

[Wilcock, 2017] Graham Wilcock. CDM CityTalk: The robot gives information about hotels and restaurants in Cambridge. [YouTube video], 2017. https://youtu.be/zWdd7kv5sX8.

[Wilcock, 2018a] Graham Wilcock. CDM CityTalk: The robot gives information about several hotels and restaurants in Tokyo Waterfront. [YouTube video], 2018. https://youtu.be/OhjIJp8XBEA.

[Wilcock, 2018b] Graham Wilcock. CityTalk: Robots that talk to tourists and can switch domains during the dialogue. In *Ninth International Workshop on Spoken Dialogue Systems (IWSDS 2018)*, Singapore, 2018.