

# Spoken Dialogue Systems for Medication Management

Joan Zheng<sup>1</sup>, Raymond Finzel<sup>2</sup>, Serguei Pakhomov<sup>2</sup>[0000-0001-8113-4788], and Maria Gini<sup>1</sup>[0000-0001-8841-1055]

<sup>1</sup> Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455

<sup>2</sup> Department of Pharmaceutical Care & Health Systems, University of Minnesota, Minneapolis, MN 55455

**Abstract.** The interest towards spoken dialogue systems has been rapidly growing in the last few years, including in the field of health care. There is a growing need for automated systems that can do more than order airline and movie tickets, find restaurants and hotels, or find information on the internet. Eliciting information from patients about their current health and medications using natural language at the point of care is a task currently performed by skilled nurses during the intake interview in both inpatient and outpatient settings. This routine task lends itself well to automation and a well-crafted dialogue system with state management can enable standardized yet individually tailored interactions with the patient using natural language. The need for extensive domain knowledge (e.g. medications, dosages, disorders, symptoms, etc.) in order to achieve broad coverage makes this task particularly challenging. In this project, we explore the use of the PyDial framework and a medication-oriented knowledge base containing information from RxNorm to create a dialogue system capable of eliciting medication history information from patients.

## 1 Introduction

Spoken dialogue systems (SDSs), systems that users can interact with through conversation, have been rapidly increasing in popularity within the past few years. Commercial SDSs, such as the Amazon Alexa, Apple’s Siri, and Google Home, can now be found in millions of households. Their popularity is due in part to the recent advancements of voice-activated technology, namely their high success rate at understanding spoken input and responding appropriately in natural language. Most commercial dialogue systems are robust enough to allow users to perform a variety of complex tasks using only voice commands, e.g. telephone banking, travel information retrieval, music management, and remote control of other smart devices.

Alongside these uses, voice assistant technology provides an opportunity for new advances in healthcare. Digital assistants capable of helping in a medical context are currently in high demand, especially considering the future of health

and lifestyle trends. For example, the number of senior citizens in the United States is expected to nearly double by 2050 [6], which will place a significant burden on the already overburdened healthcare system. SDS technology can provide aid with performing tasks and gathering information crucial to providing healthcare yet do not require supervision from healthcare professionals. For those individuals experiencing decline with their visual or motor skills, the ability to obtain care without the constant need of a healthcare worker beside them may also provide an additional opportunity to care for themselves independently.

The long-term objective of our work is to develop an adaptable dialogue system capable of interacting with a range of other medical devices and other existing technologies on an as-needed basis. The work reported in this paper focuses on building a dialogue system capable of communication information relating to medication data. Here, we target the conversation domain of medication reconciliation, the process of gathering information from patients prior to their visit with a healthcare provider [2]. This is a routine task that is currently performed by trained nursing staff. Automating this task can not only save time and lessen the cost of healthcare delivery but can also allow healthcare providers to become more effective by focusing on less routine and more complex tasks. Automation of this type can also result in the standardization of the collection of medication information, reduce confusion between different healthcare providers, and improve patient safety and satisfaction. This task requires that the SDS is knowledgeable of a large number of medications and their associated information, such as available dosages, formulations, and routes of administration, and can effectively use this information to update an individual patients' medical records.

To accomplish this, this current work uses an open-source dialogue system framework (PyDial) and a knowledge base of medications currently prescribed in the U.S., developed and maintained by the National Library of Medicine (RxNorm)

**Linguistic Context.** Many commercial SDSs do not recognize linguistic context while executing a particular task. The SDSs available for use today implement a turn-based dialogue management framework, which keeps track of a series of turns in a conversation in order to accomplish a specific task. However, if the user were to suddenly switch conversation topics while the system is trying to complete a task, the system may not recognize the switch in conversation and will (unsuccessfully) try to accomplish the task it is processing, leading to a negative experience for the user [3]. Our target SDS would take changes of topic into consideration when processing information and switch tasks if needed.

## 2 Proposed Approach

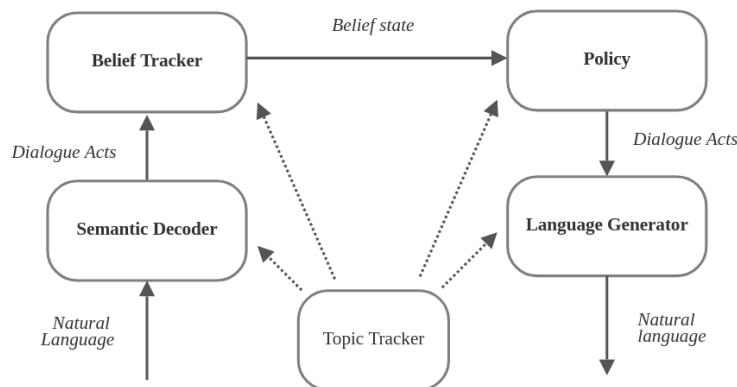
Our initial pilot application uses a SDS to perform medication reconciliation, by interviewing a patient to obtain a record of current drugs and prescription information directly from them. The task of medication reconciliation allows

us to create an interface to collect medication information from the user and compare it to the standardized data in RxNorm, yet presents the challenge of receiving and handling open-ended responses.

For this use case, there are two key types of information for each medication to consider: drug product information, and prescription information. Drug product information (e.g., Prinivil 10mg oral tablet) is specific to each variation of a medication: different variants (e.g., differing brands or dosage amounts) of a medication will have unique identifiers that are key to note in a patient’s medical record. Prescription information (e.g., take twice a day with meals) is defined by the healthcare provider, and is specific to the patient. The latter can be obtained directly from the electronic health records or pharmacy systems, if the SDS could connect to them; however, directly interfacing with clinical systems is subject to protected health information constraints. There is variability in how individual patients implement their doctors’ recommendations and treatment plans in their everyday lives. Our proposed approach is to create a conversational agent capable of eliciting the details of individual use of medications to aiding patients and caregivers with medication reconciliation.

## 2.1 Application Components

**PyDial.** PyDial [8] is a multi-domain statistical spoken dialog system toolkit that provides a framework for building a modular dialogue system. It has been created by the Dialogue Systems group at the University of Cambridge. Each module in the dialogue system has pre-implemented statistical and non-statistical approaches to process data. The main focus of PyDial is to perform task-oriented dialogue, in which a user can search for an entity in a domain that matches some number of constraints.



**Fig. 1.** An overview of the PyDial architecture. The Topic Tracker plays a role in each module, ensuring that the actions generated by the system are appropriate given the current conversation domain.

PyDial provides modules for input processing (e.g., semantic decoding), dialogue management (belief tracking and policy management), and output processing (e.g., language generation), as shown in Figure 1. All of its modules are capable of processing dialogue spanning multiple domains of conversation. While the current domain of conversation plays a role in understanding, processing, and generating natural language, PyDial’s domain-related functionality is independent from its dialogue modules, and thus the same language processing modules can be used across multiple domains of conversation. Furthermore, each module can be customized and replaced with modules specific to one’s needs – a pre-implemented module can be replaced with a customized module as long as the required signatures of the module’s function match.

PyDial’s framework was chosen as the framework of our SDS because of its ability to recognize multiple domains of conversation and switch between tasks related to separate domains. Because of PyDial’s modular nature, custom modules (namely, an ontology containing RxNorm data) were added to allow the SDS to perform tasks related to medication. The framework was also designed to perform task-oriented dialogue, allowing us to design modules to lookup information in the RxNorm knowledge base.

**ULMS RxNorm.** RxNorm [5] is a dataset created and managed by the U.S. Library of Medicine (ULMS) designed to allow different computer systems to communicate drug-related information effectively and unambiguously. This dataset provides normalized names for different variants of clinical drugs. Each medication variant is assigned a unique identifier, called a RxCUI, that differentiates between other variants of the same medication. The scope of this database includes all prescription and over-the-counter medications available in the United States, including both the generic and branded variant of every clinical drug, as well as drug interactions between different drug classes. Data containing newly-approved drug information is added once a week, which is to be used in conjunction with the full RxNorm dataset, which is updated once a month. ULMS’s frequent updates to RxNorm ensure that our SDS has the most up-to-date medication information.

RxNorm was chosen as the foundation of the ontology behind our dialogue system due to its regulated, high-quality information on medications. Ensuring that the dialogue system knows the exact medication patients are on will also aid the caregivers’ role in managing their medication history. Furthermore, using the nationally standardized information ensures that the information recorded by our SDS will be recognized and understood by other healthcare providers.

## 2.2 Application Implementation

**Dialogue System.** We created a medication ontology based on RxNorm with PyDial as the framework of the dialogue system (see Figure 3). A sample patient record was created as an example using a small subset of RxNorm records to test the performance of the SDS. When creating a speech domain with the PyDial

framework, a SQL database file containing the knowledge behind the ontology must be included with a configuration file that defines what types of information the user and system are allowed to inquire and provide.

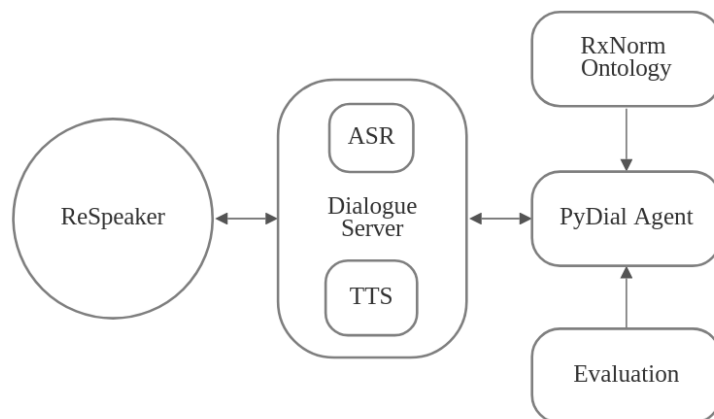
```
Prompt > Hello! I'm here to collect information about your medications.
What medications are you currently taking?
User > I'm taking metformin and prinivil
Prompt > Thanks! Let's talk about prinivil. What dose are you taking?
User > 20 milligrams
Prompt > OK. How frequently do you take it?
User > I take it once a day
Prompt > Got it. Let me confirm: You take 20 milligrams of prinivil daily,
is this correct?
User > Yes
Prompt > Thanks! Let's talk about metformin. What brand of medication do
you take?
User > I have glucophage
Prompt > OK. What dose are you taking?
User > I take 750 milligrams, once a day
Prompt > Got it. Let me confirm: You take 750 milligrams of glucophage
daily, is this correct?
User > Is glucophage metformin
Prompt > Glucophage is a brand of metformin
User > Yes that is correct
Prompt > Thanks! Are you currently taking any other medications, clinical
or over-the-counter?
User > No that's all
Prompt > Thank you! I have collected the information I needed
User > Ok goodbye
```

**Fig. 2.** An example conversation using the RxNorm ontology with handcrafted PyDial modules.

For the purpose of medication reconciliation, the user is allowed to provide the brand name, primary ingredient, dosage, and frequency of their medications. From the provided information, the system determines the concept unique identifier (CUI) of the medication the user is taking. The system is allowed to inquire about the brand name, primary ingredient, and dosage of the medication. The patient's medication record is kept as a separate ontology in PyDial that contains the drug information as well as the dosage frequency of each medication the user is taking.

Handcrafted rules for natural language generation and understanding were created to match information inquiries for the patient record. Our first iteration of this project uses regex-based handcrafted rules in both the natural language generation and natural language understanding modules of PyDial. Of course, writing handcrafted rules dealing with all possible types of medication-related

inquires for every type of information that RxNorm offers is not a realistic task. For the sake of this pilot application, our ontology was limited to a few types of information that RxNorm provided (e.g., drug interactions were left out) so that handcrafted modules could be used easily. Despite this limitation, writing handcrafted rules for our subset of RxNorm information proved robust enough for straightforward medication reconciliation tasks, as shown in Figure 2.



**Fig. 3.** The architecture of the pilot application. The ReSpeaker hardware acts as a speech client. The dialogue server can receive natural language as input and output. Text is sent and received from the dialogue system.

**Hardware.** The ReSpeaker kit, a microphone and speaker expansion board for Raspberry Pi, served as a speech client for the application. The Dialogue Server in Figure 3 consists of two RESTful services implemented using Flask technology to a) convert speech received from the ReSpeaker client into text (automatic speech recognition or ASR); and b) synthesize audio from text messages received from the PyDial Agent (text to speech or TTS). Both the ASR and TTS components of the Dialogue Server are implemented using deep learning methods. The ASR component was created using the Deep Speech 2 architecture based on Baidus Warp-CTC implementation of the connectionist temporal classification function [1]. The ASR system was trained on approximately 1500 hours of spontaneous speech using deepspeech.pytorch toolkit (<https://github.com/SeanNaren/deepspeech.pytorch>) and deployed on a GPU-enabled server. For decoding, we use beam search with a language model constructed from the transcriptions of telephone conversations collected as part of the Switchboard project [4]. The TTS component is based on the Mozilla Common Voice TTS (<https://github.com/mozilla/TTS>) project with the TTS model constructed from the LJ Speech (<https://keithito.com/LJ-Speech-Dataset/>) dataset.

### 3 Conclusions and Future Work

In the future, we plan to explore the use of PyDial’s built-in language understanding module that uses support vector machines to classify input into a set of semantic concepts. We will also explore the use of recurrent neural networks and/or long short term memory networks for the natural language generation component of PyDial. This approach can generate natural language from previous dialogue acts, allowing a greater variability of responses from the dialogue system while also taking linguistic context into consideration. It will also be important to develop an interface between PyDial and the APIs of multiple medical vocabularies such as RxNorm in order to leverage their extensive resources for information on side effects and drug-drug interactions.

#### 3.1 Language Understanding

PyDial offers a language understanding module that uses support vector machines to classify input onto a set of semantic concepts. This module maps input onto a high-dimensional feature space, allowing data to be linearly separable. The classifier must be trained using corpora data annotated with semantic data.

However, annotated corpora containing conversations relating to medication are currently not available. To collect such data, we must first obtain transcripts of patients conversing about their medications, and then annotate the scripts with dialogue intent information. Due to its sensitive nature, there is a lack of actual user dialogue done for medication reconciliation and medical record taking. As a starting point, we have taken initial steps to collect transcriptions from medication history and medication reconciliation videos from YouTube designed to train medical staff. These transcriptions will be manually annotated to create a small corpus we plan to use to bootstrap the training of statistical components of the conversational agent.

A challenging issue for the ASR component of the system that will need to be addressed in future development is to improve the recognition of a wide range of medication names and medication-related information. We plan to address this by implementing Cold Fusion methods for introducing target domain language patterns during training of an out-of-domain ASR model [7].

#### 3.2 Language Generation

Creating a hand-crafted policy spanning all of the RxNorm data would be a much more expensive task than the small sub-domains used in the pilot application. This task can be avoided by implementing a language generation module that uses recurrent neural networks [9] and/or long short term memory networks [10]. This approach, which generates natural language from previous dialogue acts, would allow a greater variability of responses from the dialogue system while also taking linguistic context into consideration. Overall, this would lead to a more natural flow of conversation.

The pronunciation of much of the medical vocabulary and medication names in particular is highly idiosyncratic, leading to erroneous audio synthesis by text to speech systems. For example, the system that we currently use trained on the LJ Speech data pronounces the word "aspirin" as [ax s p ay ax r ih n] rather than [ae s p r ih n]. We plan to address this issue by adding pronunciations of medication names to the training data.

**Acknowledgements.** Work supported in part by CRA-W Distributed Research Experiences for Undergraduates program.

## References

1. Amodei, D., Anubhai, R., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Chen, J., Chrzanowski, M., Coates, A., et. al.: Deep speech 2: End-to-end speech recognition in english and mandarin. CoRR **abs/1512.02595** (2015), <http://arxiv.org/abs/1512.02595>
2. Aronson, J.: Medication reconciliation. *BMJ* **356** (2017)
3. Chowdhury, S.A., Stepanov, E.A., Riccardi, G.: Predicting user satisfaction from turn-taking in spoken conversations. Proc. Annual Conference Int'l Speech Communication Assoc. (INTERSPEECH) **08-12-Sept**, 2910–2914 (2016)
4. Godfrey, J.J., Holliman, E.C., McDaniel, J.: Switchboard: telephone speech corpus for research and development. In: [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing. vol. 1, pp. 517–520 vol.1 (March 1992). <https://doi.org/10.1109/ICASSP.1992.225858>
5. Liu, S., Ma, W., Moore, R., Ganesan, V., Nelson, S.: Rxnorm: prescription for electronic drug information exchange. *IT professional* **7**(5), 17–23 (2005)
6. Ortman, J.M., Velkoff, V.A., Hogan, H., et al.: An aging nation: the older population in the United States. US Census Bureau, Economics and Statistics Administration, US Department of Commerce (2014)
7. Sriram, A., Jun, H., Satheesh, S., Coates, A.: Cold fusion: Training seq2seq models together with language models. CoRR **abs/1708.06426** (2017), <http://arxiv.org/abs/1708.06426>
8. Ultes, S., Rojas Barahona, L.M., Su, P.H., et al.: Pydial: A multi-domain statistical dialogue system toolkit. In: Proc. ACL 2017, System Demonstrations. pp. 73–78 (2017), <http://www.aclweb.org/anthology/P17-4013>
9. Wen, T.H., Gašić, M., Mrkšić, N., M. Rojas-Barahona, L., Su, P.H., Vandyke, D., Young, S.: Multi-domain neural network language generation for spoken dialogue systems. In: Proc. 2016 Conference on North American Chapter of the Association for Computational Linguistics (NAACL) (June 2016)
10. Wen, T.H., Gašić, M., Mrkšić, N., Su, P.H., Vandyke, D., Young, S.: Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In: Proc. 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP) (September 2015)