Combined Learning of Dialog State Tracking and Dialogue Management

The aim of this project is to investigate the feasibility of optimizing a dialog manager on sequences of output from Natural Language Understanding (NLU) inputs in task-oriented dialog systems with deep learning approaches tailored to sequenced data (e.g. NNs with recurrence, convolutions or seq-seq architecture).

Background

Task-oriented dialog systems typically consist of the following pipeline:
1. Speech Recognition: speech-to-text
2. Natural Language Understanding: intent recognition
3. Dialog State Tracking (DST): maintain summary of conversation so far with user desires
4. Dialog Management: determine next dialog action / what to say
5. Language Generation
6. Speech Generation: text-to-speech

Although approaches to domain-independent dialog state tracking exist, they are tailored to the task at hand (recommendation via slot-filling). Furthermore, results (DSTC2) indicate that ensembles of different dialog state trackers are better than individual DST models. Finally and most importantly, tuning / evaluating a DST algorithm for a particular task+domain requires data that includes the true dialog state (or the true user goal, which is typically used a proxy). This may be expensive and does not leverage any existing data collected from real users.

Research Goal

The goal of this research is to investigate the feasibility of learning dialog state representations and optimal dialog management simultaneously.

Possible Approaches

Various architectures are to be implemented in a Deep Learning framework and tested on existing benchmarks. Frameworks for working with dialog such as Parl.AI and Pydial enable testing various approach in simulation.