

Multi-domain Task completion dialogue system

Sanchit Nevgi

College of Information and Computer Science
University of Massachusetts, Amherst

March 26, 2020

Introduction

Dialogue systems can be broadly classified based on three use-cases — Task-Oriented, Question & Answering, and Social chatbots. Traditionally, hand-crafted features were used to build dialogue systems. With advancements in deep learning, neural models have been trained in an end-to-end manner to decrease reliance on manual feature-engineering. These models have demonstrated ability to adapt to user *utterance paraphrasing* and handle Out-of-Vocabulary (OOV) terms. There is an increasing interest in building complex bots that span multiple sub-domains to accomplish a complex user goal such as travel planning Li et al. 2020

Problem outline

The Dialogue State Tracking challenge (DSTC) Kim et al. 2019 is an annual challenge aimed to push the frontier in dialogue systems. In the selected task — "Multi-Domain Task-Completion Dialog Challenge", we have to build an **end-to-end multi-domain** dialog system for tourist information desk setting.

Evaluation: Although the challenge makes use of both automatic and human evaluation metric, we forego the human evaluation. Two metrics are used for automatic evaluation; *Task success rate* and *precision/recall* score for slots prediction. We follow a similar methodology.

Methodology

Baseline: A simple baseline would be to predict a constant for all utterances. For a stronger baseline, we follow the proposed baseline model in the challenge Li et al. 2020, which is a modular pipeline approach consisting of multi-intent language understanding model (MILU), a rule-based dialog policy and template-based NLG module.

Approach: We reuse the rule-based DST and Dialogue Policy from the supplied ConvLab package. As a preliminary experiment, we propose using contextualized BERT embeddings in the NLU module of the pipeline. Further, we experiment with attention mechanisms over the sentences which have shown to handle domain switch. Another approach to consider is to move from template-based NLG methods to neural based language models. We analyse the viability of each approach to select one for further investigation.

Datasets

DSTC-8 motivates employin the MULTIWOZ 2.1 dataset Budzianowski et al. 2018 as the corpus. MULTIWOZ is a multi-domain dialog dataset, where dialog agents interact with tourists to satisfy their demands, such as making hotel reservation. It covers 7 domains in tourist information setting including *Attracting*, *Hospital*, *Police*, *Hotel*, *Restaurant*, *Taxi*, and *Train*. It consits of 10,438 dialogs with 1,000 dialogs used for validation and test respectively. No additional human annotation is needed.

Software and Frameworks

ConvLab: ConvLab Lee et al. 2019 is a tool introduced to enable researchers to compare different approaches, perform evaluation, etc. It includes pre-trained models for dialog system components such as NLU, NLG, RL-policies along with user simulators. ConvLab suggests using `nltk` library for tokenization.

PyTorch: For building neural models, we will make use of the current version of PyTorch.

References

- Budzianowski, Paweł et al. (2018). *MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling*. arXiv: 1810.00278 [cs.CL].
- Kim, Seokhwan et al. (2019). *The Eighth Dialog System Technology Challenge*. arXiv: 1911.06394 [cs.CL].
- Lee, Sungjin et al. (2019). “ConvLab: Multi-Domain End-to-End Dialog System Platform”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.
- Li, Jinchao et al. (Feb. 2020). “Results of the Multi-Domain Task-Completion Dialog Challenge”. In: *Proceedings of the 34th AAAI Conference on Artificial Intelligence, Eighth Dialog System Technology Challenge Workshop*. URL: <https://www.microsoft.com/en-us/research/publication/results-of-the-multi-domain-task-completion-dialog-challenge/>.