Multi-domain Dialogue State Tracking as Dynamic Knowledge Graph Enhanced Question Answering

Li Zhou
Amazon Alexa Search
lizhouml@amazon.com

Kevin Small
Amazon Alexa Search
smakevin@amazon.com

Abstract

Multi-domain dialogue state tracking (DST) is a critical component for conversational AI systems. The domain ontology (i.e., specification of domains, slots, and values) of a conversational AI system is generally incomplete, making the capability for DST models to generalize to new slots, values, and domains during inference imperative. In this paper, we propose to model multi-domain DST as a question answering problem, referred to as *Dialogue State Tracking via Question Answering* (DSTQA). Within DSTQA, each turn generates a question asking for the value of a (domain, slot) pair, thus making it naturally extensible to unseen domains, slots, and values. Additionally, we use a dynamically-evolving knowledge graph to explicitly learn relationships between (domain, slot) pairs. Our model has a 5.80% and 12.21% relative improvement over the current state-of-the-art model on MultiWOZ 2.0 and MultiWOZ 2.1 datasets, respectively. Additionally, our model consistently outperforms the state-of-the-art model in domain adaptation settings.

1 Introduction

In a task-oriented dialogue system, the dialogue policy determines the next action to perform and next utterance to say based on the current dialogue state. A dialogue state defined by *frame-and-slot semantics* is a set of (key, value) pairs specified by the domain ontology (Jurafsky & Martin, 2019). A key is a (domain, slot) pair and a value is a slot value provided by the user. Figure 1 shows a dialogue and state in three domain contexts. Dialogue state tracking (DST) in multiple domains is a challenging problem. First of all, in production environments, the domain ontology is being continuously updated such that the model must generalize to new values, new slots, or even new domains during inference. Second, the number of slots and values in the training data are usually quite large. For example, the MultiWOZ $2.0/2.1$ datasets (Budzianowski et al., 2018; Eric et al., 2019) have 30 (domain, slot) pairs and more than $4,500$ values (Wu et al., 2019). As the model must understand slot and value paraphrases, it is infeasible to train each slot or value independently. Third, multi-turn inferences are often required as shown in the underlined areas of Figure 1.

Many single-domain DST algorithms have been proposed (Mrkšić et al., 2017; Ren et al., 2018; Zhong et al., 2018). For example, Zhong et al. (2018) learns a local model for each slot and a global model shared by all slots. However, single domain models are difficult to scale to multi-domain settings, leading to the development of multi-domain DST algorithms. For example, Nouri & Hosseini-Asl (2018) improves Zhong et al. (2018)'s work by removing local models and building a slot-conditioned global model to share parameters between domains and slots, thus computing a score for every (domain, slot, value) tuple. This approach remains problematic for settings with a large value set (e.g., *user phone number*). Wu et al. (2019) proposes an encoder-decoder architecture which takes dialogue contexts as source sentences and state annotations as target sentences, but does not explicitly use relationships between domains and slots. For example, if a user booked a restaurant and asks for a taxi, then the destination of the taxi is likely to be that restaurant, and if a user booked