Available online at www.sciencedirect.com

ScienceDirect

Computer Speech & Language 54 (2019) 1–16

www.elsevier.com/locate/csl

Associative knowledge feature vector inferred on external knowledge base for dialog state tracking☆

Yukitoshi Murase*[,a], Yoshino Koichiro[a,b,c], Satoshi Nakamura[a]

[a] Nara Institute of Science and Technology, Takayamacho 8916-5, Ikoma, Nara 630-0192, Japan
[b] RIKEN AIP, Takayamacho 8916-5, Ikoma, Nara 630-0192, Japan
[c] JST PRESTO, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan

Received 12 March 2018; received in revised form 17 July 2018; accepted 6 August 2018
Available online 22 August 2018

Abstract

The dialog state tracker is one of the most important modules on task-oriented dialog systems, its accuracy strongly affects the quality of the system response. The architecture of the tracker has been changed from pipeline processing to an end-to-end approach that directly estimates a user's intention from each current utterance and a dialog history because of the growth in the use of the neural-network-based classifier. However, tracking appropriate slot-value pairs of dialog states that are not explicitly mentioned in user utterances is still a difficult problem. In this research, we propose creating feature vectors by using inference results on an external knowledge base. This inference process predicts associative entities in the knowledge base, which contribute to the dialog state tracker for unseen entities of utterances. We extracted a part of a graph structure from an external knowledge base (Wikidata). Label propagation was used for inferring associative nodes (entities) on the graph structure to produce feature vectors. We used the vectors for the input of a fully connected neural network (FCNN) based tracker. We also introduce a convolutional neural network (CNN) tracker as a state-of-the-art tracker and ensemble models of FCNN and CNN trackers. We used a common test bed, *Dialog State Tracking Challenge 4* for experiments. We confirmed the effectiveness of the associative knowledge feature vector, and one ensemble model outperformed other models.

Keywords: Dialog state tracking; Knowledge base; Knowledge graph; Associative knowledge inference

1. Introduction

Dialog state tracking (DST) is known as an important component of task-oriented dialog systems (Young et al., 2010; Thomson and Young, 2010; Williams et al., 2016). DST is a task of tracking user intentions (dialog frame) from input utterances and dialog history. Dialog state tracking challenges (DSTCs) have been held to provide a common test bed for DST (Williams et al., 2013). DSTC, DSTC2, and DSTC3 provide human-computer dialog corpora for estimating a user's dialog states (Williams et al., 2013; Henderson et al., 2014a; 2014b). DSTC4 and DSTC5

☆ This article was recommended for publication by Prof. R.K. Moore.

* Corresponding author.
E-mail address: y-murase@is.naist.jp (Y. Murase).

https://doi.org/10.1016/j.csl.2018.08.003
0885-2308/ 2018 Elsevier Ltd. All rights reserved.

provide human-human dialog corpora for estimating the states. The difficulties of estimating human-human conversation are the large intentional space and collecting enough data for covering the space.

Throughout the previous DSTCs, discriminative methods, which directly predict dialog frames, performed well (Henderson, 2015). Recurrent neural network (RNN) based approaches competitively performed well on DSTC2 (Henderson et al., 2014c). RNN approaches were also competitive to other approaches for DSTC4 and 5 (Yoshino et al., 2016; Hori et al., 2016); however, convolutional neural network (CNN) based approaches outperformed RNN-based approaches and were reported as being the state-of-the-art for DSTC5 (Shi et al., 2016b; 2016a). These approaches used distributed word representation such as word2vec or GloVe (Mikolov et al., 2013; Pennington et al., 2014) for their input features. However, the lack of information obtained from input features is still a problem from two viewpoints. The first is that it is challenging to find any unseen slot-values of dialog state frames that are not observed in a user utterance. The second is data size; the number of annotated dialog data is limited, and the number of output states is explosively large.

External knowledge such as ontology is a pivotal component for solving the problem of a lack of information in inputs. However, handcrafted ontologies are not capable of being extended without professionals. Due to the growth of the world wide web (WWW), a variety of knowledge bases (KBs) is publicly available (Bollacker et al., 2008; Vrandečić and Krötzsch, 2014). In Ma et al. (2015), a method of ontology extension was proposed uses external KBs. These KBs contain entities and properties that are transformed into a graphical model. It is possible to find associative entities by using the structure of KBs to fill a lack of input information.

In this paper, we propose a method of creating an associative knowledge feature vectors (AKFVs) highly capable of expressing the meaning of an utterance by using unobservable information in utterances. The feature vectors include information obtained from global associative entities. A fully connected neural network (FCNN) with the proposed feature vectors comparably performed the state-of-the-art CNN-based dialog state tracker. Moreover, an ensemble of the proposed method and the CNN-based tracker outperformed the CNN-based tracker and achieved the best score for neural-network-based trackers for DSTC4.

2. Related works

2.1. Dialog state tracking challenge 4

Dialog State Tracking Challenge 4 (DSTC4) is a common test bed for dialog state tracking and is aimed at achieving more human-like dialog systems by using a human-human conversation corpus for a sightseeing domain. The corpus is a collection that a total of 21 hrs of conversation between 3 tour guides and 35 tourists on Skype. It is divided into *training, development*, and *test set*, which respectively contain 14, 6, and 15 dialog sessions. Each dialog session has been manually transcribed, and dialog frames have been annotated for each sub-dialog level. A sub-dialog means any turns of the dialog session.

The total number of utterances within sub-dialog segments is 20,641. The challenge of DSTC4 is a task of tracking dialog frames for sub-dialog segments, where a frame contains a topic and slot-value pairs. An ontology is also given data, and it contains all possible slot-value pairs under each topic. The slot-values represents intention in the human conversation, and the ontology indicates the knowledge of possible human intentions for this domain.

Annotations are given to each sub-dialog segment. The topics are sub-domains under the sightseeing domain, and topics are annotated to the all sub-dialog segments. Topics are categorized into five classes: *accommodation, attraction, food, shopping*, and *transportation*. A state frame is given for all sub-dialog segments. The frames contain several slot-value pairs, which represent intentions in conversation within each sub-dialog. For example, the slot-value annotation at the *accommodation* topic frame might have a slot called "type" that represents *"What kind of accommodation style?"*, and corresponding values could be filled with *hotel, hostel*, etc.

The ontology is given to ensure estimation of slot-value pairs since it contains all topics and possible pairs in a hierarchical structure. The structure has three layers. The top layer has five topics, and the middle layer has multiple slots that are each dependent on a topic. The bottom layer has values, which depend on a slot. Therefore, a higher layer has more abstract information, and a lower layer has more specific information as shown in Table 1. The number of all slot-value pairs is 5608, so the task requires estimation within a large intentional space.