Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Computer Speech & Language xxx (2018) xxx-xxx
www.elsevier.com/locate/csl

End-to-end task dependent recurrent entity network for goal-oriented dialog learning☆

Q1 Chang-Uk Shin[a], Jeong-Won Cha[b]*

[a] Department of Eco-friendly Offshore plant FEED Engineering, Changwon National University, 20 Changwondaehak-ro, Changwon-si, Gyungsangnam-do, Republic of Korea

[b] Department of Computer Engineering, Changwon National University, 20 Changwondaehak-ro, Changwon-si, Gyungsangnam-do, Republic of Korea

Received 12 March 2018; received in revised form 30 April 2018; accepted 25 June 2018
Available online xxx

Abstract

In this paper, we introduce the Task Dependent Recurrent Entity Network (TDREN) to solve Dialogue System Technology Challenges 6 (DSTC 6) track 1. Traditionally, there have been methods such as collecting the intent of the user in a conversation directly using rules. We design an end-to-end structure that properly models the restaurant pre-related user preferences that appear in the dialogue and gives appropriate responses. We perform experiments on the TDREN and achieved 97.7% at precision 1. We propose a new artificial neural network structure and recurrent cell for modeling user preference information. Then, we show that task-oriented dialogue modeling experiment results using the structure and the recurrent cell.

1. Introduction

The goal-oriented dialogue system is one of the dialogue systems which conducts dialogue with the user about specific domains and goals. To achieve the user's desired goal, the system has to be able to identify the intent of the user and the preference information of the user during conversation. It should also be able to ask the user for insufficient information to obtain the appropriate information.

Prior to the deep-learning study, dialogue modeling using rules and 'Partially Observable Markov Decision Process (POMDP)' has been studied (Young et al., 2013). After the deep learning has begun to be studied actively, dialogue modeling has been attempted using neural network architectures such as sequence-to-sequence (Vinyals and Le, 2015), Hybrid Code Network (HCN, (Williams et al., 2017; Ham et al., 2017)), and memory networks (Sakai et al., 2017; Kim et al., 2017).

This paper is a study of neural network structure for task-oriented dialogue modeling. We propose our novel architecture named Task Dependent Recurrent Entity Network (TDREN) and a new recurrent cell to properly understand the intent and the preference of the user. Our model can be seen as a variant of the Dynamic Memory Network (DMN, (Kumar et al., 2016)), Recurrent Entity Network (REN, (Henaff et al., 2016)), or Question Dependent Recurrent Entity Network (QDREN, (Madotto and Attardi, 2017)) for task-oriented dialogue.

☆ This paper has been recommended for acceptance by Roger Moore.

* Corresponding author.
E-mail address: jcha@changwon.ac.kr (J.-W. Cha).

https://doi.org/10.1016/j.csl.2018.06.004
0885-2308/ 2018 Elsevier Ltd. All rights reserved.

Our structure focused on the entities what we want to manage in the task. Therefore, we have set up several encoders to manage different entities. In the remaining part of the paper, we describe the proposed structure and recurrent cell, and show experimental results of using it. We also share some insights into the errors in the experiment.

2. Related research

As research on natural language processing using artificial neural networks is activated, there have been cases of end-to-end modeling of dialogue using artificial neural network structures. Some dialogue systems based on artificial neural networks have been proposed that adopt RNN (Vinyals and Le, 2015; Kumar et al., 2016), bag-of-words (BOW) (Williams et al., 2017) or simple matrix multiplication approaches (Henaff et al., 2016; Madotto and Attardi, 2017; Sukhbaatar et al., 2015) to acquire distributed representations of utterances, and separate modules for language understanding and generating responses.

There was an approach to balance traditional approach and end-to-end one. Wen et al, 2017) used a database to get clear attribute information and adopted two modules to catch user's intent, an intent network and a belief tracker. After the information was all gathered, the policy network summed the information and transmitted it to the generation network. Finally, the generation network generated the response.

Vinyals and Le (2015) studied dialog modeling system using sequence-to-sequence architecture. First they encoded sentence uttered by user using Long-short term Memory (LSTM) encoder, and then generated system response using LSTM decoder. They evaluated the system performance in both IT help-desk domain and daily conversation domain. The experimental results show that the purely data-driven sequence-to-sequence approach can produce answers according to the user's utterance.

Williams et al. (2017) proposed the Hybrid Code Network (HCN), which is another end-to-end structure for task-oriented dialogue modeling. HCN significantly reduced the complexity of dialogue modeling by processing the named entity information that had difficulty managing in traditional artificial neural network architectures. In addition, novel method named action template has been used to reduce the complexity of the dialog modeling, again. By separating the concerns of the end-to-end model for named entity in this way, dialog modeling was successfully done with even less training data.

There was also some dialogue system research using Memory Network (Sakai et al., 2017; Kim et al., 2017). Memory Network points out that existing Recurrent Neural Network (RNN)-based models cannot capture long-term dependencies and that real problems can be solved by modeling this long-term dependency. Therefore, instead of using the conventional method of using the RNN internal cell as a memory, the memory network has an external memory component and learns it properly to achieve high performance.

The Memory Network newly defines four modules. The first module is the input module. The input module processes input sentences and converts them into distributed representations. The generalization module selects a specific slot in memory and holds an input representation in the memory there. The memory goes from a state where no information was input (empty state) to a state where all of slots are configured (full state), and if the memory becomes full, a forgetting procedure can be performed. The output module retrieves the information from the existing memory and response module generates the final response (Weston et al., 2015).

Sukhbaatar et al. (2015) has modified some operations of the memory network structure. Also, it is shown that performance improvement is possible by increasing the number of hops in the memory module. This work has been applied to several task-oriented dialogue modeling studies.

2.1. Dynamic memory network

Dynamic Memory Network (DMN) is the artificial neural network structure proposed in Kumar et al. (2016) for question-answering (QA) task that give an appropriate response to a question. DMN consists of an input module, a question module, a memory module and an answer module. The input module and the question module respectively receive information and the memory module performs a role of understanding the information. And finally, the answer module generates the system response using the memory representation from the memory module and the question representation from the question module.