Deep Reinforcement Learning with Hierarchical Recurrent Encoder-Decoder for Conversation

Heejing Jeong and **Xiao Ling**
[heejinj, lingxiao]@seas.upenn.edu

Abstract

This paper investigates the efficacy of combining deep reinforcement learning with a hierarchical recurrent encoder decoder for conversational dialogue systems.

1 Introduction

In recent years, various neural network based methods have demonstrated state of the art performance in automatic dialogue generation. Vinyals showed that a simple sequence to sequence (SEQ2SEQ) model mapping input utterances to output repsonses could learn to converse fluently in restricted domains (Vinyals and Le, 2015). However this model often fails to capture long-term conversation histories, and does not consider influences of current responses on future outcomes. Therefore it often fails to maintain a coherent topic of converation and outputs short-sighted responses. Li showed that integrating SEQ2SEQ and reinforcement learning (DRL-SEQ2SEQ model) generated more interesting responses and thereby encouraged a longer dialogue session (Li et al., 2016). Serban improved topic coherence using a hierarchical recurrent encoder-decoder architecture (HRED), which incorporated conversation history to bias local probability of words within an utterance (Serban et al., 2015). In this paper, we propose a dialogue generation model using deep reinforcement learning (DRL) and HRED, thus incorporating the advantages of both models. We train the model on the CALLHOME American English Speech corpus (LDC97S42), consisting of 120 30-minute phone conversations between native English speakers.

2 HRED

HRED is a generative model that learns a probability distribution over the set of all possible dialogues of arbitrary lengths. Specifically, HRED assumes the probability of the current word is a function of both previous words in the utterance, and previous utterances in the session. This multiscale long term dependency is captured by an utterance level RNN that encodes sentences into hidden vectors, and a dialogue level RNN that updates this vector throughout the session.

3 DRL with HRED

Our model differs from DRL-SEQ2SEQ model in three aspects. First, instead of using two previous dialogue utterances to define a state, we define a state s_t at time t as a hidden state c_t of the dialogue level RNN at the time, $s_t = c_t = f(c_{t-1}, h_T^{(t)})$, where $h_T^{(t)}$ is the last hidden state of the tth utterance and f is a parametrized non-linear function. We use the same definition for actions - generating a dialogue utterance. Thus, a RL policy is defined as $\pi(a_t|s_t) = p_{RL}(u_{t+1}|c_t)$. Second, we pre-train a model using HRED to initialize the RL policy π. Finally, we replace the pre-trained SEQ2SEQ model with the pre-trained HRED model in reward functions of RL.

References

Jiwei Li et al. 2016. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 1192–1202.

Iulian Serban et al. 2015. Building end-to-end dialogue systems using generative hierarchical neural network models. https://arxiv.org/abs/1506.05869.

Oriol Vinyals and Quoc V. Le. 2015. A neural conversational model. In *Proceedings of the 31st International Conference on Machine Learning*. https://arxiv.org/abs/1506.05869.