Deep Reinforcement Learning for Dialogue System

Zvengin
zvengin@nii.ac.jp

January 3, 2018

1 General Introduction

1.1 Problem

Based only on neural network, the dialogue system generates short-sighted response and predictes response one at a time ignoring their influence on future outcome. Therefore, response generated by neural network based system usually seems dull and lasts for short turns.

1.2 Methods

Author applies reinforcement learning methods, which can optimize long-turn rewards designed by system-developer, to dialogue systems. Maximzing long-turn rewards ensure that current repsonse will have maximum positive impact on future outcome and keep conversation going on as long as possible. Sequence-to-Sequence model is the backbone of author's dialogue system which is used to define a policy of generating utterance. Optimization process actually is a process of searching for optimal policy. In a word, this paper combines the strength of reinforcement learning in optimizing for long-turn goals accross conversation and the power of seq2seq model to learn compositional semantic meaning of utterance.

2 Details

2.1 Reward Definition

Three kinds of rewards are provided in author's dialogue system.
1. Easy Responding
Author thinks that a good repsonse should be easy to answer and in this way, conversation can go on as long as possible. Therefore, author proposes to measure the ease of answering a generated turn by using the negative log likihood of responding to that utterance with a dull response

2.Information Flow
In order to avoid entering a cycle or repetition, similarity of responses from two turns is measured by calculating negative log cosine of the two responses.

3.Semantic Coherence
In order to avoid situations in which the generated replies are highly rewarded but ungrammatical

or not coherent. author proposes to use mutual information between current response and the previous response in dialogue history.