DSTC6 Challenge Task Proposal

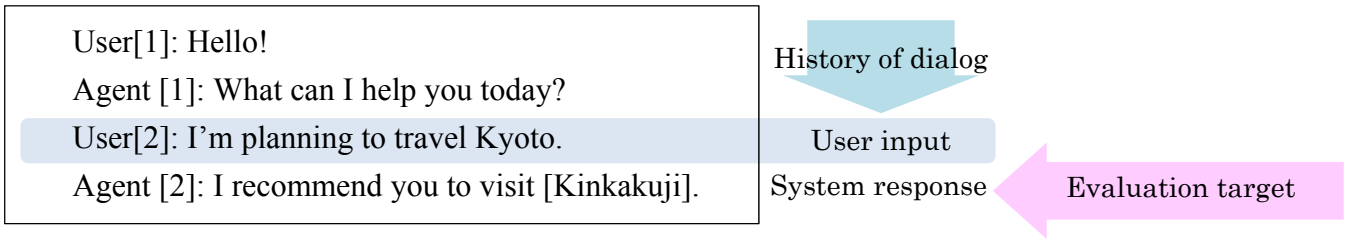
Chiori HORI
chori@merl.com
Mitsubishi Electric Research Laboratories

1. Introduction

End-to-end training of neural networks is a promising approach to generate responses of dialog systems automatically using human-to-human dialog corpus. Google tested the performance using OpenSubtitles in 2015 (https://arxiv.org/pdf/1506.05869.pdf). Montreal and McGill Universities tested using Ubuntu Dialogue Corpus in 2016 (https://arxiv.org/pdf/1506.08909.pdf). The goal of end-to-end dialog systems can accomplish task oriented dialogs to provide information properly with natural communications. Microsoft tested their knowledge grounded neural conversation model using Twitter in 2017 (https://arxiv.org/pdf/1702.01932.pdf). This research is aiming at combining conversational dialogs with task-oriented knowledge using unstructured data such as Twitter data for conversation and Foursquare data for external knowledge. We would like to propose an end-to-end dialog task, in which a system plays a role of human agent and generates natural and informative sentences responsive to user's questions or comments given dialog context and external knowledge.

2. Task specification

The system has to generate sentence(s) responsive to a user input in a given dialog history, where it can use external knowledge from web.



The quality of automatically generated sentences is evaluated with objective and subjective measures to judge whether or not the generated sentences are natural and informative for the user.

3. Data set

Scenes of dialog where users ask concrete information to the agent will be extracted from dialog corpora such as OpenSubtitles, Twitter, Switchboard. Scenes where users don't request information explicitly but the agent provides some concrete information will also be included. Full or part of the training data will be used to train conversation models. Any open data, e.g. from web, are available as external knowledge to generate informative sentences. But they should not overlap with the training, validation and test data provided by organizers.

Training data will be part of OpenSubtitles/Twitter/Switchboard, consisting of a few million dialogs.
Test and validation data will be extracted from social media: 1000-2000 dialogs (up to the number of system submissions and budgets)

4. Baseline system

   LSTM-based sentence-to-sentence generation system and a pre-trained model will be provided.

5. Evaluation framework

   Objective measure: Perplexity, Bleu, etc.

   Subjective measure: Human rating using crowd source.

6. Sample data

   From Twitter @AskTSA

   Highlighted sentences are expected to be generated by the system.

Agent	Travel Tip: It's spring break; you can expect lots of company in lines for tickets, baggage & security screening. Pls arrive 2hrs early.
User	Even if I have 'pre-chek' status?
Agent	Our recommendation is based on delays that may occur from curb to gate. TSA Precheck lanes are generally less than 10 mins.

Agent	Do you have questions or comments to share? TSA wants to hear from you! Tweet to us for live assistance with your travel-related issues.
User	Travelling Toronto YYZ --> Chicago - can i bring a personal meal with fruits and vegetables???
Agent	Fruits and vegetables are allowed in your carry-on, but may undergo additional screening. If you meal includes liquids...1/2
User	do I need to declare it if it's meal size in my carry on?
Agent	You can find out more on our website here: http://ow.ly/WgKig
User	Thanks for the information!