Neural Conversational Model

This research is about building a model that predicts next sequence of sentences from previous set of sequences. It is trained end-to-end and uses recurrent neural networks. It can process both domain specific and general dataset and understand it. It has good performance with general dataset compared to other models and hence it has the upper hand in cases when domain specific dataset is not available. The model is tested on IT helpdesk where the machine gives responses for technical questions. It is also tested on movie subtitles dataset for common sense reasoning.

Here, a sequence to sequence approach is chosen which works better in language translations. The model is trained on recurrent neural network to give answers based on probabilistic model and to increase the probability of the answer. Input is given in form of tokens and the output is also in tokens. Greedy approach is considered in which the output is given back to the model as input when the required output is not observed during training, so that the machine learns better. Consistency has been an important limitation of this approach.

In the IT help-desk dataset, conversations between a costumer with a technical problem and a technician are collected. 30M tokens were taken as train set and 3M tokens for validation set. Unwanted stuffs like urls, phone number and all are removed from the dataset. In the movie subtitles dataset, conversations from movie are collected in XML format and the sentences are fed consecutively not knowing if its from the same characters or different. The dataset is filtered to remove XML tags and non-conversational texts. The dataset is split into 62M sentences for train data and 26M sentences for validation data. This dataset is very huge compared to the previous one and is also very noisy. This challenge increases the scope and widens the domain for movies.

The model is observed to be remembering facts, understanding contexts and performing common sense reasoning without any complex structure. It gave good response to questions but the responses were short and simple. The model was compared with CleverBot where 4 humans were used to judge which model gave better responses. After the judging, it was clearly seen that this model performed better than CleverBot. This research shows that a simple seq2seq approach is effective in producing simple and basic conversations.