A Deep Learning Powered Chatbot

Trinity College
HARTFORD CONNECTICUT

Wang Shufan ’18 | Advised by Prof Takunari Miyazaki
B.S. Mathematics, Computer Science | Department of Computer Science, Trinity College

Abstract

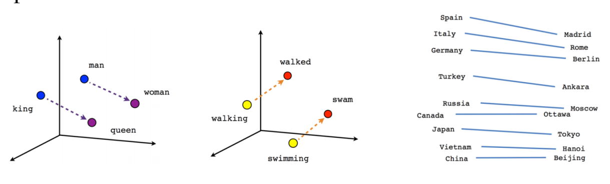
Natural dialogue generation is a key application area of machine learning, which many consider to be the holy grail of artificial intelligence. As of today, it is still an unsolved open problem and a hot area of research. Recent development in deep learning has encouraged new waves of attempts to design and train the Chatbots (conversational agent). We follow this direction of research and explore employing deep generative models to learn natural language and held conversations with people. Specifically, we adopt the “sequence to sequence” encoder decoder architecture. By using a large dataset of movie dialogues, we train a recurrent neural network autoencoder, which learns to generate responses to input messages. We demonstrate our result both quantitatively by showing the decreased cost function value during training and qualitatively by talking to our trained Chatbot.

Word2Vec

In processing text as data, the word2vec model [1] is often used to convert word tokens into vectors and **each sentence will therefore be processed as a sequence of word tokens.** One way to obtain embedding vectors is to train training a language model such as the Continuous-Bag-Of-Word (CBOW). The objective function used for the training process is follows.

$$\begin{aligned} J &= -\log P(w_c \mid w_{c-m}, \ldots, w_{c-1}, w_{c+1}, \ldots, w_{c+m}) \\ &= -\log P(u_c \mid \hat{v}) \\ &= -\log \frac{\exp(u_c^T \hat{v})}{\sum_{j=1}^{|V|} \exp(u_j^T \hat{v})} \\ &= -u_c^T \hat{v} + \log \sum_{j=1}^{|V|} \exp(u_j^T \hat{v}) \end{aligned}$$

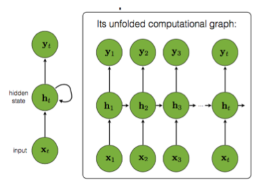
The resulted embedding vectors should therefore capture the semantic relationship between words. For example, the word “*King*” is to “*Queen*” as “*Men*” is to “*Women*” in the embedding space:



Male-Female | Verb tense | Country-Capital

Recurrent Neural Network

To process sequential data (sentences), we will use the state-of-the-art deep learning model, the **Recurrent Neural Network (RNN)** and its variation, the Long-Short-Term-Memory (LSTM). A simple illustration of the RNN is as follows:

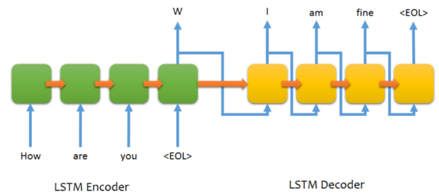


$$y_t = g(h_t; \theta_y) = W_{hy} h_t$$
$$\theta_y = \{W_{hy}\}$$
$$h_t = f(h_{t-1}, x_t; \theta_h) = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$
$$\theta_h = \{W_{hh}, W_{xh}\}$$

As shown above, RNN takes in each token of the sequential data in each time-step and returns an output, as well as a hidden state that is passed on to the next time step (as a memory). Therefore, it retains a memory of the overall sequence.

Seq2seq model

Our model architecture follows that of a **sequence to sequence autoencoder** [2]. It consists two parts: an encoder and decoder, each a Recurrent Neural Network, to process and output sequential data respectively. A graphical representation is as follows, where the encoder encodes information about the input sentence into its final hidden state and then pass that to the decoder.



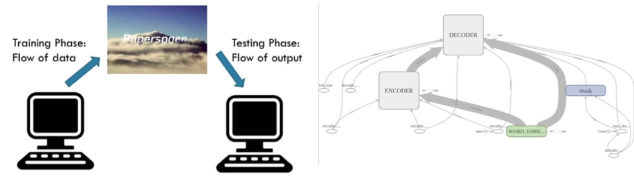
Implementation

Python libray NLTK is used for construcint word tokens and Tensorflow is used to construct the RNN and seq2seq model.

TensorFlow™ | Natural Language Analyses with NLTK

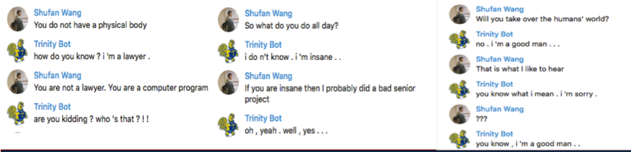
Model Training

We use the Cornell Movie Dialogue Dataset [3] and gather 150000 pairs of conversations from over 600 movies as training data. Due to the large model size, we deployed the model on to the cloud for storage and training.



Result

With 100 epochs training on the GPU, the cost function value dropped by 90% from 10.67 to 1.09, showing that our model can fit the data. Additionally, we present some sample conversations with our chatbot, showing its ability to generate natural language.



Future Research

Although the deep learning approach is promising, the natural language processing community still have a along way towards generating natural human dialogues. Particularly, models can be improved on area such as incorporating common sense, developing context awareness and using creative language.

Acknowledgements

I would like to thank Professors Takunari Miyazaki, Peter Yoon, and Ewa Syta of the Trinity College Computer Science Department for their guidance and support in this project, and also Wang Yichun ’20, Wang Yinghuan ’18 and Zhang Weixi ’18 for many helpful discussions and their support.

Works Cited

1. Mikolov et al. *Efficient Estimation of Word Representations in Vector Space*. ICML 2016.
2. Cho et al. *Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation*. EMNLP 2014.
3. Danescu-Niculescu-Mizil. *Cornell Movie-Dialogue Corpus*. ACL 2011.