Project Name: Smart Email Composer

Members: Yuhua He
Jia Ma
Ying Liu
Yuanzhe Li
Samuel Yang

Problem Statement:

1. Describe your problem and its significance

Email is the most popular media for communication among the groups. With the rapid growing of emails from the inbox, it is very time consuming to type and reply them all individually. In our project, we are trying to build a deep learning model which can efficiently provide suggestions based on user preference or the history of the emails in order to help individuals to write and reply email in a more efficient way.

2. Briefly provide some background for your problem

While developing the model, it is also very critical for us to think about the model in certain aspects such as minimize the response latency since the we would like to provide the suggestion on each keystroke while user is typing, also we need to make sure we will not expose sensitive PII data. Those two aspects will play significant role for us to build the model and will also be the most challenging part for us.

3. Provide details about the objective function and constraints in your problem. For example: The aim of this project is to recognize voice gender using Multilayer Perceptron (MLP) deep learning model.

Apply sequence to sequence model to find the most relevant content based on the existing user input.

4. Give an overview of how you propose to solve your problem. Examples: An MLP deep learning algorithm has been applied to detect gender- specific traits.

To resolve our problem, we proposed a model called sequence-to-sequence(seq2seq), where the source sequence is the concatenation of the subject and the previous email body (if there is one), and the target sequence is the current email the user is composing.

The input is the tokens of the original message(encoder RNN), and the output is the conditional probability distribution of the sequence of response tokens given the input(decoder RNN). First, the sequence of original message tokens, including a special end-of-message tokens are read in, Then, given this hidden state, a softmax output is computed and interpreted as the probability distribution for the first response token. As response tokens are fed in, the softmax at each timestep is interpreted as the probability distribution for the first response token. Given the factorization above, these softmaxes can be used to compute conditional probability distribution.

5. Explain why studying your problem is important. What will a solution to your problem enable you to do.

Because the traditional feed forward neural network can not work very well on random sequenced tokens, we need to compute probability of future sequence based both on current email writing sequence and previous emails. Seq2Seq solves this problem in a similar way to how backpropagation-through-time solves the task of training cyclic networks. With backpropagation-through-time, we took atemporal self-referential network and changed it to a spatial non-self-referential network. Here, with seq2seq, we reinterpret a spatial problem (a variable-length sequence of tokens) as a temporal one (tokens generated over time).

Literature review

Paper Reference:

https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45189.pdf
https://github.com/harvardnlp/seq2seq-attn
https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf
https://ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html

Datasets: Enron email dataset:

https://www.kaggle.com/wcukierski/enron-email-dataset

1. Describe previous work related to your problem.

Google has released an Artificial intelligent tool called Smart Compose, which uses deep neural network techniques to suggest successive wordings during email composition. This tool is based on Smart Reply technology, which suggests responses to a short messages from Android messaging apps. This technology uses hierarchical learning model to process words all at once, along with LSTM (long-short-term-memory) recurrent neural network.

2. Start with the articles finding other related work by looking at the references in these articles. [Google Scholar (Links to an external site.)Links to an external site.](#)is a good place to start.w

Character-based Neural Machine Translation

https://aclweb.org/anthology/P/P16/P16-2058.pdf

Sequence to Sequence Learning with Neural Networks
https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf

Smart Reply: Automated Response Suggestion for Email
https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45189.pdf

Recurrent Neural Networks for Language Understanding
https://www.isca-speech.org/archive/archive_papers/interspeech_2013/i13_2524.pdf

3. After perusing through the related work you find, select a subset that is most related to your problem.

We are planning to research more on the BoW (bag of words) model with an RNN-LM, and compare this mixed model with sequence to sequence deep learning model and LSTM (long-short-term-technology). Based on the literature review, this mixed model should have advantage over seq-to-seq or LSTM.

4. Tell a coherent story of what the authors of these works did, and how they are related to your project.

The authors of Recurrent Neural Networks for Language Understanding uses RNN to understand different languages. Working with Elman Architecture and Apriori method, they furthermore makes observations of future words. They used ATIS dataset to come up with both lexical and non-lexical features as well as the BoW (Bag of Words) effect. For all the these they do are really related to our project because our main goal is to make predictions on future words. The algorithms they used are also really good tools for us to start our project with.

In particular, for each previous work, you should answer the following questions:

6. What problem did the authors study? How is it similar/different from your project problem?

For all the problems above are quite related to the RNN algorithm and also seq to seq corresponding to the LSTM(Long Short Term Memory). Applying heavy text datas to the model and make predictions on future words are the similar part to our project. As well as the email research, the data they used are also similar to the data we are going to use in our project. The different part will be that some of their research are focusing on recognizing characters and words and we focus more on making predictions on future words.

7. How did the authors solve their problem? How is their approach similar/different from your proposed approach?

At first one the authors propose a neural MT(Machine Translation) system using character-based embeddings in combination with convolutional and highway layers to replace the standard lookup-based word representations.

Similarity: follows an encoder-decoder architecture with attention, and introduce elements from the character based neural language model.

For the second one, The idea is to use one LSTM to read the input sequence, one time step at a time, to obtain large fixed dimensional vector representation, and then to use another LSTM to extract the output sequence
from that vector. The second LSTM is essentially a recurrent neural network language model except that it is conditioned on the input sequence.

Similarity: Both are facing sequence to sequence problem, and trying to find a best way to solve it,

Different: The topic in the paper is focus on machine translation, we are looking for frequent sequence.

The third one is most related to our project, since its a auto reply mechanism. it has been successfully deployed in Inbox, currently. Response selection, Response set generation, Diversity, Triggering model. The combination of these components is a novel end-to end method for generating short, complete responses to emails, going beyond previous works. For response selection it exploits state-of-the-art deep learning models trained on billions of words, and for response set generation it introduces a new semi-supervised method for semantic understanding of user-generated content.

8. For an example, take a look at paragraphs 2-4 in the "Introduction" section from the related papers.

1. The translation unit continues to be the word, and we continue using word embeddings related to each word as an input vector to the bidirectional recurrent neural network

2. A useful property of the LSTM is that it learns to map an input sentence of variable length into a fixed-dimensional vector representation. Given that translations tend to be paraphrases of the

source sentences, the translation objective encourages the LSTM to find sentence representations that capture their meaning, as sentences with similar meanings are close to each other while different 2 sentences meanings will be far

3. To address this problem, we leverage the sequence-to sequence learning framework, which uses long short term memory networks (LSTMs) to predict sequences of text. Consistent with the approach of the Neural Conversation Model, our input sequence is an incoming message and our output distribution is over the space of possible replies.