Artificial Intelligence Final Report Assignment 問題 2 (Problem 2)
レポート解答用紙 (Report Answer Sheet)

Group Leader
学生証番号 (Student ID):18521068
名前(Name):Trần Bình Luật

Group Members
学生証番号 (Student ID):18521284
名前(Name):Huỳnh Ngọc Quân

学生証番号 (Student ID):18521285
名前(Name):Nguyễn Minh Quân

問題２　(Problem 2)のレポート

I. Introduction

1. Sentiment analysis

Sentiment analysis is the process of understanding if a given text is talking positively or negatively about a given subject. Is a branch of Text Classification. In reality, Sentiment analysis is widely applied to voice of the customer(VOC) materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. Input is a sentence about review, survey, .. Output is a prediction as positve, negative, or neutral.

2. Dataset

IMDB is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. Provide a set of 25,000 highly polar movie reviews for training, and 25,000 for testing. There is additional unlabeled data for use as well. Raw text and already processed bag of words formats are provided. This dataset is created by Stanford University (ACL 2011).

Ex: I sure would like to see a resurrection of a updated Seahunt series with the tech they have today,...(Positive).

II. Method

Recurrent Neural Networks (RNNs) suffer from short-term memory. If a sequence is long enough, they'll have a hard time carrying information from earlier time steps to later ones. So if you are trying to process a paragraph of text to do predictions, RNN's may leave out important information from the beginning. Also during back propagation, recurrent neural networks suffer from the vanishing gradient problem. And we think solving above issues is the most important for helping improve the model.

Therefore, to solve limited problem from the RNN model. We propose to conduct the LSTM model which improved from the RNN model.

1. Architecture

Long Short Term Memory networks (LSTM) are a special kind of RNN, solve the learning long-term dependency problem and vanishing gradient problem which RNN suffer. They were introduced by Hochreiter & Schmidhuber (1997), and were refined and popularized by many people in following work. They work tremendously well on a large variety of problems.

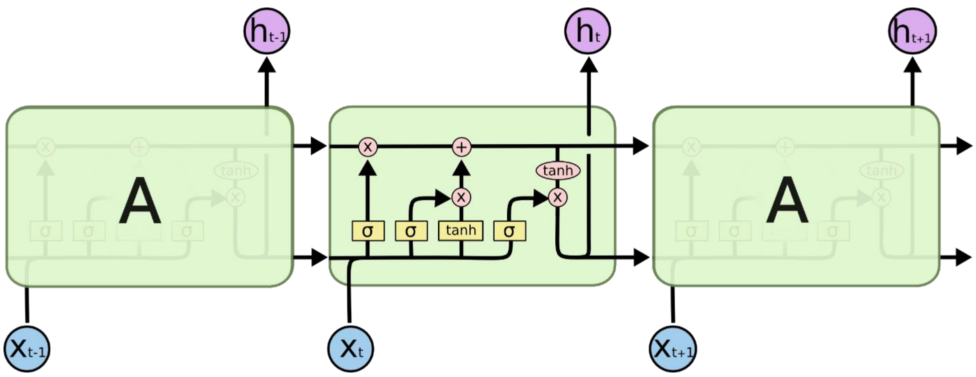
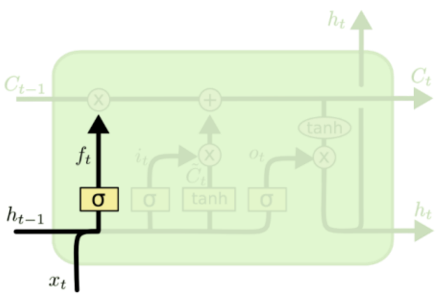


Figure 1. LSTM Architecture
(colah.github/)

Step by step in LSTM:

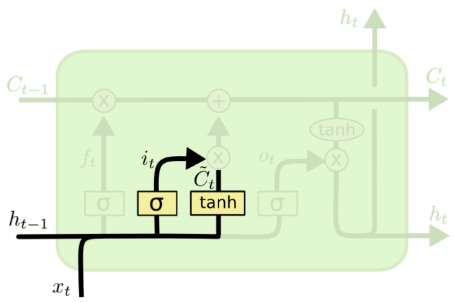
- The first step in LSTM is to decide what information are going to throw away from cell state. This step called "Forget gate".



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

Figure 2. Forget gate
(colah.github/)

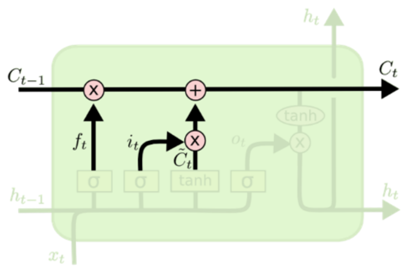
- The next step is to decide what information is going to store in the cell state. First, a sigmoid layer called the "Input gate" decides which values we'll update. And next function is to generate a new cell state for this step. Tanh is used for modulation.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

Figure 3. Input gate and next cell state
(colah.github/)

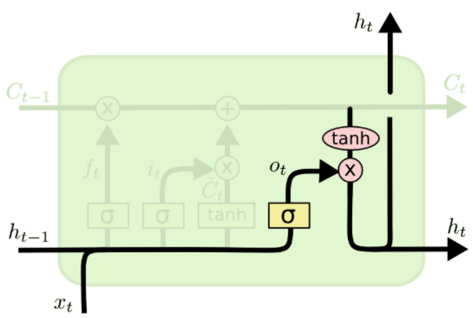
- Next, update the old Cell state C_{t-1} to C_t .



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Figure 4. Update Cell state

- Next, use a sigmoid layer that decides what parts of the cell state we're going to output called the "Output gate". Then use the tanh function to modulate the previous calculated C_t and multiply it with the Output gate to get the Output result as well as the hidden result .



$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 5. Calculate Output (Hidden)
(colah.github/)

2. Configurations

In this problem, we deploy an LSTM model with 1 layer and 2 layers. The epoch is from 10 to 30 and the learning rate is 1e-5. For the IMDB dataset we set batch size as 64.

III. Experiments

	Accuracy (%)	
Epoch	**LSTM – 1 Layer**	**LSTM – 2 Layers**
10	68.33	80.44
20	75.14	84.12
30	79.06	85.90

Table 1. Final Accuracies

We tended to transform the RNN model (in the teacher's slide) to LSTM model 1 layer and 2 layers respectively and got a higher accuracy. Then we tried to fine - tuning hyperparameters as an epoch, learning rate, drop out. We have increased epoch from 10 to 30 and give higher accuracies. We have increased and decreased the learning rate to 1e-4 and 5e-6 respectively. But didn't give higher accuracy so we think increasing and decreasing the learning rate is not a good solution. We also tried to add dropout in LSTM model but we didn't give much higher accuracies (accuracies are lower with drop out 0.5).

IV. Conclusions Future Work

In this problem, we identified several limitations associated when performing sentiment analysis with RNN model. We presented the LSTM model. Through extensive experiments and analysis, we demonstrated the effectiveness of our method with highest accuracy is 85.90% in test set. In future we will use more modern model as Bi-LSTM, Transformer and embedding as BERT, Glove to get higher accuracy in this test.