

COMP9417 - Assignment 2

Avinash K. Gupta, Yuyang Shu, Maria Oei

June 4, 2017

1 Introduction

With over 100 millions of monthly visitor in Quora, it is inevitable that many people are asking similar questions. This has become an issue as user has to read through responses to many questions in order to find the best answer. The aim of this project is to implement algorithms to identify 2 similar questions which can help Quora in improving user experience by finding high quality answers to questions.

The 2 approaches used in this project were perceptron learning and LSTM. In the perceptron learning, we used 3 inputs which measured semantic similarity, word order and word overlaps between 2 sentences.

In LSTM, bla bla bla bla

We will see that for our case the performance of the 2 models were quite similar. However, there are other research done on LSTM where it has been shown that LSTM has the potential to performs well with enough training and with careful selection of initial weight.

2 Methodology

2.1 Method 1 : Perceptron

- What do we implement (Theory) - How do we implement it

2.2 Method 2 : LSTM - RNN

- What do we implement (Theory) - How do we implement it

LSTM is a recurrent neural network which was first developed by Sepp Hochreiter and Jurgen Schmidhuber in 1997. It has advantage over traditional neural network model as it is able to remember information which it has seen previously. This has caused RNN to gain popularity in recent years in NLP thus making it an important model to look at.

2.2.1 Overview

The RNN which we looked at is the Long Short-Term Memory (LSTM) RNN. Each of the nodes in LSTM RNN consists of 3 main gates namely, the forget gate, the input gate and the output gate.

3 Result

How do they perform

4 Discussion

4.1 Conclusion

4.2 Limitation & Improvement