COMP9444 Assignment2 Report
Recurrent Networks and Sentiment Classification

z5092923 Jintao Wang
z5104857 Xiaoyun Shi

1. Text preprocess

(1). Changing all the upper cases to lower cases.
(2). Replace punctuations by space, for instance, “#$%&()*+,-./:;<=>@[\]^_`{|}~. However, warning sign(!) and asking sign(?) are kept because I think these two punctuations are related to some kind of sentiment.
(3). Remove all the linking words and linking phrases like 'based on','in order to', 'given that', 'due to', 'all of a sudden’.
(4). Replace deny phrases and deny words by “not”, for instance, "other than", "no longer", "in no way”, "can't", "cant", “didn’t”,”doesn't", “don’t".
(5). Strengthen samples by repeating the 3 words twice after link verbs because I think most of them are adjective words and they might be very important.
(6) Remove stop words, abbreviation and some other noise.
(7) Remove words randomly to keep the number of words in reviews same as the argument MAX_WORDS_IN_REVIEW.

2. RNN model

For this run model, I constructed a RNN model of 2 layers, for each layer, there are 32 GRU cells. Batch size is 32 and max words in a review is 150. All the weights are initialised by tf.truncated_normal function and all the bias are initialised as constant 0.1. Softmax cross entropy is used to compute the loss and Adam optimizer is applied to minimise the loss with default learning rate.

3. Result

The final accuracy of evaluation is about 85%. And here are the graphs of loss and training rate: