**CS 482/682 Final Project Midterm Report**

**Cong Gao, Zhaohao Fu, Liujiang Yan**

We plan to design a deep neural network to classify topics from text. The input of the model includes news headlines, news articles, tweet comments and other public news or social networking related text files, which can be of different length, but limit to English language. These texts will be labeled to finite set of classes. Our goal is to compare the classification performances under several datasets.

At this midterm stage, we implement a character-level convolutional networks (character-level ConvNets)1 and a recurrent convolutional neural network (RCNN)2 for the text categorization task. We test our model on data set of Trump and Clinton tweets. The dataset contains 6444 tweets, all of which is tweeted by Trump or Clinton. Our task is to classifier who the tweet is posted by.

For the character-level ConvNets approach, we design the following model, as shown in Figure 1. We put character quantization on top of network. The character quantization is an encoding process which transforms the natural languages into a sequence of *m* sized vectors with fixed length *l0*1. After that, we add two convolution layers. Followed by two linear layer with the first coupled with ReLu and Dropout, we use softmax (logit) as our output layer. Compared to the original network, which contains 6 convolutional layers, we reduce the depth of our network. The reason for doing this is to avoid under fitting, considering the small size of our dataset at this stage. As a matter of fact, we tried using the 6-layer network, of which most of the weights turns out to be zero.

Figure 1: Network architecture for initial attempt

The second model is RCNN approach contains three parts: recurrent structure, max-polling layer and output layer. In the model, the recurrent structure, which is a bidirectional recurrent neural network, is used to capture the word’s context. The recurrent structure composes of long short-term memory (LSTM) units. After the recurrent structure is a pooling layer, which transfers the information into a fixed length vector. And in the end we use softmax as our output layer.

Due to the time limit, we only trained our model on the small tweet data, as described before. The two models achieve 0.979 and 0.911 accuracy respectively in the training set. More network architectures and dataset will be explored, as promised.

References

[1] Zhang, X., Zhao, J. and LeCun, Y., 2015. Character-level convolutional networks for text classification. In Advances in neural information processing systems (pp. 649-657).

[2] Lai, S., Xu, L., Liu, K. and Zhao, J., 2015, January. Recurrent Convolutional Neural Networks for Text Classification. In AAAI (Vol. 333, pp. 2267-2273).