**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.

In this midterm stage, following Kim Kim's Convolutional Neural Networks for Sentence Classification paper1, we implement a Computational Neural Network (CNN) on top of pre-trained word vectors for the text categorization task. We test our model on a data set of Trump and Clinton tweets. The dataset has binary category: from Trump / from Clinton.

We design a following model, as shown in Figure 1. The word vectors are obtained from an unsupervised neural language model, which is a popular method to improve performance in the absence of a large supervised training set2. On top of that, we add two convolution layers. 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 test dataset. As a matter of fact, we tried implementing the 6-layer network, of which most of weights turns out to be zero. Followed by two linear layer with the first coupled with ReLu and Dropout, we use softmax (logit) as our output layer.

Figure 1: Network architecture for initial attempt

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

References

[1] Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).

[2] Collobert, Ronan, et al. "Natural language processing (almost) from scratch." Journal of Machine Learning Research 12.Aug (2011): 2493-2537.