

Mohamed Kharaev

43121144

CS 175 HW 2

The topic I chose to explore is the Long-Short Term Memory (LSTM) architecture. In the first article, LSTM memory cells were implemented into a neural network being used for language modeling. These memory cells add intermediate steps when calculating gradients, which helps to avoid the vanishing gradient problem in recurrent neural networks. In this first experiment, the combination of an LSTM network with existing clustering techniques caused gains in speed of training and testing times with low loss in performance. The second article discusses combining convolutional neural network (CNN) with LSTM to create a new C-LSTM architecture. This new architecture was being used in the context of text classification. The CNN is used to extract higher-level sequences of word features and the LSTM to capture long-term dependencies over window feature sequences respectively. The resulting semantic sentence representations were evaluated on sentiment classification and question type classification, which yielded a 94.6 accuracy.

References

1. Sundermeyer, Martin, et al. "INTERSPEECH 2012." ISCA Archive, *ISCA's 13th Annual Conference*, 2012, pp. 194–197.
2. Zhou, Chunting, et al. "A C-LSTM Neural Network for Text Classification." *ArXiv:1511.08630*, 30 Nov. 2015.