**HW1: Annotated Bibliography**

CSC 540: Graduate Research Seminar

Camron Khan

1. Bengio, S., and Bengio, Y. Taking on the curse of dimensionality in joint distributions using neural networks. IEEE Transactions on Neural Networks 11, 3 (2000), 550-557.

This paper discusses how fully connected neural networks with hidden layers can be used to approximate joint distributions on high-dimensional data sets. The authors demonstrate how such highly dimensional data can be modeled with network of nodes that increases at a rate of the number of parameters squared, which is an improvement over exponential increases by other models.

1. Bengio, Y., Ducharme, R., Vincent, P., and Janvin, C. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3 (2003), 1137-1155.

This paper discusses a way to model words as vectors of real numbers in which the “meaning” of the word is distributed along the values in such vectors. This model allows the capturing of semantic and syntactic generalizations by performing vector calculations on each word in vector space.

1. Fang, A., MacDonald, C., Ounis, I., and Habel, P. Using word embedding to evaluate the coherence of topics from twitter data. In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (Pisa, Italy, Jul. 17-21). ACM, New York, 2016, pp. 1057-1060.

This paper evaluates three metrics used to identify topics in Twitter tweets: (1) Pointwise Mutual Information, (2) Latent Semantic Analysis, and (3) Word Embedding. The authors concluded that using Word Embedding metrics aligned most closely with human judgments and required order of magnitude smaller training sets that the other two approaches.

1. Vuurens, J. B. P., Larson, M., and de Vries, A. P. Exploring deep space: Learning personalized ranking in a semantic space. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (Boston, MA, Sep. 15). ACM, New York, 2016, pp. 23-28.

This paper demonstrates how users’ reviews and ratings can be leveraged along with deep learning to produce a distributed representation of movies in semantic space. This semantic space, in which movie similarity is determined by vector distance, can be paired with a user’s past preferences to create a mapping to recommendations. This method out performed many state-of-the-art movie recommenders currently in use.

1. Xu, C., Bai, Y., Bian, J., Gao, B., Wang, G., Liu, X., and Liu, T. Y. RC-NET: A general framework for incorporating knowledge into word representations. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (Shanghai, China, Nov. 03-07). ACM, New York, 2014, pp. 1219-1228.

This paper proposes a new framework for combining relational and categorical knowledge about words with the skip-gram distributed representations of those words. The authors demonstrate that this new model performs significantly better than models using solely skip-gram, relational, or categorical knowledge representations.