

# CS410 Technology Review Proposal

Xiaolong Yang

My project is to build the recommendation system on Wikipedia data, based on Topic Models & Neural language Model. In this case, I would compare the difference between the LDA and Word2vec, model of Topic Models and Neural Language Model.

The main difference between these models is the contextual information they use: LSA and topic models use *documents* as contexts, and Neural language models and distributional semantic models instead use *words* as contexts. These different contextual representations capture different types of semantic similarity; the document-based models capture semantic relatedness (e.g. “boat” – “water”) while the word-based models capture semantic similarity (e.g. “boat” – “ship”). [1]

I would use LDA with Metapy [2], and use word2vec with Tensorflow [3] to code and run, try to compare the accuracy.

I would finish the project by myself.

## Reference:

[1] <https://www.gavagai.se/blog/2015/09/30/a-brief-history-of-word-embeddings/>

[2] <https://github.com/meta-toolkit/metapy/blob/8716109742d1f0763be61e37482f4bbca5a1ff82/tutorials/5-topic-modeling.ipynb>

[3] <https://www.tensorflow.org/tutorials/word2vec>