# Sentiment Analysis for Reviews (Project Proposal)

**Shuhong Zheng**
Department of Computer Science
University of Illinois Urbana-Champaign
szheng36@illinois.edu

## 1 Basic Information

Shuhong Zheng (NetID: szheng36) is the only team member of the team and undoubtedly I am the captain.

## 2 Topic Selection

I choose the free topic theme and my project is called "Sentiment Analysis for Reviews".

- **What is the task?** In detail, the task is about how to determine the sentiment expressed in reviews. The goal is to classify the sentiment of reviews into categories of positive, negative or neutral based on the language used in the reviews.
- **Why is it important or interesting?** Sentiment analysis for reviews is crucial for companies to obtain feedbacks from users, in order to further improve their products and make important business decisions. It will also help customers in purchasing products because they can use reviews from other people as references.
- **What is your planned approach?** My planned approach is to use deep neural networks for sentiment analysis. In detail, I will first use a tokenizer to tokenize the sentences in the review. Then, I will map them into latent embedding vectors in the vector space, where texts of similar meanings will get closer in the vector space. Afterwards, I will use certain networks in the natural language processing to process the vectors of the tokens to get the final classification of the sentiment.
- **What tools, systems or datasets are involved?** Some machine learning frameworks like Pytorch and some scientific computing packages like Numpy, Scipy would be involved. For the datasets, I would use prevalently used datasets like the IMDb movie review dataset [1].
- **What is the expected outcome?** The system is expected to achieve satisfactory performance on the testing split of the dataset used for training. Also, it is expected to generalize to in-the-wild review texts that might be somewhat different from the training dataset.
- **How are you going to evaluate your work?** On the testing split of the dataset used for training, we could perform quantitative evaluation that use accuracy as the evaluation metric. For the in-the-wild analysis, we could perform qualitative evaluation to manually see whether the sentiment analysis results match the exact sentiment conveyed in the reviews.

## 3 Programming Language

I will mainly use Python as my programming language.

## 4 Justification of the Workload

The whole workload includes getting familiar with the topic, processing the dataset, training the network, evaluation, experimental analysis, and writing report. Since I am the only team member in the team, I need to perform all the above works which I think would take at least 20 hours.

## References

[1] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts. Learning word vectors for sentiment analysis. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2011.