Natural Language Processing

Project Proposal

Evaluation of Language models for Sentiment Analysis

Team members

1. Anargh Viswanath
2. Protyush Kumar Das

Topic

Comparative evaluation of the performance offered by the state of art language models such as BERT and ELMo with traditional ones such as Word2Vec and GloVe for sentiment analysis.

Core components

1. Python
2. Language models based on TensorFlow and PyTorch deep learning frameworks
3. Datasets based on movie reviews and/or product reviews

Project Outline

Sentiment Analysis is the area of research concerned with understanding of human sentiments/ emotions. This is largely concerned with classification of text data by deriving the underlying implications in the text through analysis. This field has been a major focus of research traction due to the growth in applications related to market research, social media monitoring, customer service, brand monitoring etc. These applications perform analysis on user data to learn more about the user preferences.

Sentiment analysis rely on NLP and machine learning techniques to learn from large datasets. Language models are widely used for sentiment analysis. BERT and ELMo are the current state of the art language models. They have achieved tremendous success and have set benchmarks for performance.
In this project, we will evaluate these models with the traditional language models such as Word2Vec and GloVe for sentiment analysis.

The datasets should be carefully selected while covering a wide number of application cases. This will help in training, evaluation and comparison of the different language models. For our project, we will utilize datasets generated from movie reviews (IMDb, Rotten Tomatoes etc.) and/or amazon product reviews. These are free and available online.
We believe that this project is very course-oriented while providing us with hands-on experience with the SOTA and could be completed within the duration of the course.