CS 410 Text Information System: Final Project Proposal

TEAM PYTHON

Team members:

- Chuanyue Shen (NetID: cs11, team leader)
- Jianjia Zhang (NetID: jianjia2)
- Runpeng Nie (NetID: runpeng3)

Kaggle competition name:
Toxic Comment Classification Challenge

Kaggle competition link:
https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/overview

Competition description:
Our final project is going to build a model for Toxic Comment Classification. In some public online forums, meaningless, offensive and embarrassing comments can make people who really want to communicate stop expressing their ideas. Research teams like Google and Jigsaw were trying to build different models in order to use text classification to better identify toxic comments. We will build a multi-headed model that is capable of detecting different types of toxicity like threats, obscenity, insults, and identity-based hate better than Perspective's current models.

We will be using a dataset of comments from Wikipedia's talk page edits. In this dataset, there are six types of toxicity: toxic, severe_toxic, obscene, threat, insulti and dentity_hate. The data size is around 50MB. The data contains around 150,000 lines of information about the sentence id and labels. The data type is numerical, and all the data are stored in a table format.

Project will be evaluated on the mean column-wise ROC AUC. In other words, the score is the average of the individual AUCs of each predicted column. Our model will predict a probability for each of the six possible types of comment toxicity in order (toxic, severe_toxic, obscene, threat, insult, identity_hate).

Methodology:
This project is a classification problem. We will explore and implement various supervised machine learning approaches and possibly an ensemble of them. The potential algorithms that can be used include but not limited to:

- SVM
- LSTM
- Deep learning framework

The main packages we will use include but not limited to:

- Numpy
- Pandas
- Scikit-learn
- Keras
- Pytorch

Programming language: Python

Relevant prior experience with the project:
The team members are familiar with Python language and have some experience with mini machine learning projects but not much. We chose this project to gain some hands on in text classification and to explore machine learning approaches.

Feasibility:
This project will be fairly feasible since we can use code and papers published by other researchers as reference, to gain some ideas when building the model architecture. We will implement some popular methods and then improve the model performance based on our own design. In addition, we will leverage existing packages and tools such as pandas, sklearn, and keras to facilitate the data manipulation and classification modeling.

Timeline:

- Model buildup: week 9 - 12
- Training and testing: week 12-14
- Model improvement: week 14-15
- Final Report: week 15-16