

Project: Sentiment Analysis of Intra-Platform Ties on Uberpeople.net
Members: Amaris Huang, Ryan Huang, Ziling Cheng

In this project, we would like to adopt the original research of Dr. Xiao Chen, and Dr. Wei Chi, on exploring the phenomenon of informal intra-platform ties, among on-demand ride-hailing drivers in North America. We base our model on data scraped from www.uberpeople.net, a widely used online platform for uber users-- drivers and customers. As an information-seeking website, we would like to detect how emotional the posts get through applying machine learning models on analyzing the wordings of individual posts. We choose to scrape data from www.uberpeople.net because it is representative of business platforms. It has lots of users and contains lots of posts for us to analyze. It demonstrates typical online interaction patterns and how informal intra-platform ties form.

The emotion keywords in each post would be the most useful for our model. We plan to first of all remove the stop words which are not relevant in the context of the data and do not contribute any deeper meaning to the phrase. Though the dataset contains various information, we would like to focus on the text of the posts only. We will process the text through tokenization, using `nltk.word_tokenize()`, through which a word list is obtained for each post on which we operate our AI model. We also need to ensure that words that look different due to casing or written another way but have the same meaning will be treated equally, so normalisation process such as lower casing the characters will be necessary.

We would try to implement multiple models in order to find the best approach. The possible choices include basic NLP models like Bert, recurrent neural network like GRU (gated recurrent unit) and LSTM, adaptive fuzzy inference neural network, convolutional neural network, and other traditional machine learning methods (Naïve Bayes, Logistic Regression). We listed those models in order of increasing performance but also increasing complexity. Since our project is essentially a classification problem, we will report confusion matrix and accuracy/precision-recall/logistic loss.

We will have a simple webpage that takes a post as input, and outputs whether it is pure informational or contains emotional components. The technology we plan to use is Flask. Flask allows us to create a simple web app using Python, which we are all familiar with.