Project Overview

Allan Sun, Henry Xu

Context:

With a click of mouse, people nowadays can easily retweet or repost messages on social media which gives the resurgence of fake news. Fake news on Online Social Networks like Twitter, Weibo, Wechat etc have misguided lots of people and caused enormous loss to the society. Many companies actually hire people to label fake news manually which wastes a lot of time and resources. However, with the help of deep learning, many algorithms have been designed to detect fake news in different periods of their life cycle. And many state of art results have been realized which encourage our further research into this area.

Objectives:

Most of the algorithms existed nowadays focus only on part of the fake news' life cycle. Inspired by the XGboost, we want to design a new system that ensembles different stages of detection. In another word, our senior project will focus on the whole lifespan of fake news using deep learning. First of all, we want to use CNN to develop a early stage detection algorithm which is based on detecting linguistic features. Then we want to design a new algorithms based on the spread pattern and user response to the news. The second stage can be transformed into a classification problem. The basic assumption is that fake news have a different spread pattern on the social network than a normal news. For now, a SVM is suited for solving this question. Next, we want to use Google's search result as a way to evaluation the sources' credibility. For example, if the news appears on New York Times during Google search, then it's more likely that it is not fake.
Besides all that, we also want to develop a mobile application which enables the user to login in some online social network and automatically label fake news for them.

Future Plans:

During the winter break, we have finish part of the literature review and found many interesting papers. By the end of February, we wish that we could set a reasonable benchmark to test our system. And we would spend a month design the system with another month to improve the score.