AML Final Project Proposal

Xiangjing Dong, Yingying Cao

Image Analysis for COVID-19 Diagnosis

In the early days of COVID-19 pandemic, fast COVID-19 PT-PCR test had not been developed. Doctors needed to analyze chest X-ray images to give a diagnosis. Thus many machine learning based algorithms are proposed to assist the diagnosis, which can give a diagnosis much faster than a radiologist at the same time with high sensitivity. We believe these algorithms have a broader prospect since they can not only be used for COVID-19 diagnosis but also for other diseases such as lung cancer.
We are going to replicate several papers with different image analysis algorithms and compare them on the same dataset.

Relevent Works and Methodology

Rachna Jain and Linda Wang used convolutional neural networks(CNN) to classify different types of patient. [1-2] However, Yu-Huan Wu pointed out that CNN models lack the explainable transparency to assist the doctors during the medical diagnosis.[3] So they used joint classification and segmentation(JCS) which could achieve a very high sensitivity and give explainable diagnosis.[3]

Dataset

We have different datasets to chose. COVID-CS is given by Yu-Huan Wu. And there are also availible datasets on Kaggle.

Potential Difficulties

One of the potential difficulties is to replicate the algorithms and I think the most difficult part might be about computation resources. Training CNN models with a training set of thousands of images could be time-consuming. We will make use of Google Colab and the GPU resources provided by our department to accomplish this final project.

References

[1] Jain, R. , Gupta, M. , Taneja, S. , & Hemanth, D. J. . (2021). Deep learning based detection and analysis of covid-19 on chest x-ray images. Applied Intelligence, 51(3), 1-11.
[2] Wang, L. , Lin, Z. Q. , & Wong, A. . (2020). Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports, 10(1).
[3] Wu, Y. H. , Gao, S. H. , Mei, J. , Xu, J. , & Cheng, M. M. . (2021). Jcs: an explainable covid-19 diagnosis system by joint classification and segmentation. IEEE Transactions on Image Processing, PP(99), 1-1.

Stock Prediction from Tweets

Hypothesis

"Twitter dropped 6% in market value after the company decided to permanently suspended President Trump's account" [3]. If the short-term stock price, in essence, is a reflection of public's attitudes towards

the market then it is very reasonable to predict the stock trend based on social media content such as tweets.

Relevant Works

[2] They approach this question by first measuring people's mood and investi- gate the correlation between Twitter moods and stock price through Granger causality analysis. Then they fed the sentiment features with past stock values into the model to predict the future stock prices. One of pros is that the data is relatively easy to acquire which is essential for convincing results. Second, in- stead of using only one mood measurement method, two were applied and they also validated the sentiment analysis results by comparing dominant moods in special event periods to ensure the reliability. They did not, however, account for the impact variations caused by different numbers of followers.
[4] Their research is based on Bollen's work and more algorithms were exam- ined, including Linear, Logistic Regression, SVMs and SOFNN. Additionally, they proposed a sequential k-fold cross validation since it is meaningless to train on future values while evaluate on previous stock data. Finally, they tested the model on a more general time period rather than on a rare time slot. The good side of their study is that the adjustments they made appeared to make results more convincing, however, they only extract sentiment information from tweets containing certain words like "feel", "makes me", "I'm" which may miss other valuable information.
Regarding the above, we will build from the second paper since the adjust- ments they made are more reasonable to us and it is also beneficial for our study to try different models.

Methodology

The answer to our question mainly lies in two factors, feature engineering and suitable algorithm. In terms of feature engineering, In addition to creating mood measurements, we are considering adding extra features e.g. stock-related hashtags information. We will also weight the influence of tweets, which could be done by, for instance, taking the amount of followers into account. For the prediction task, a binary (up/down) classification in our case, we will start with logistic regression, tree-based, and eventually move onto neural networks, e.g. LSTMs are good for a time-series analysis or/and SOFNN proposed in the paper.

Data Source

The data source from [1] contain 79 days around a million tweets mentioning any NASDAQ 100 companies. The corresponding stock price in that period can be crawled from Yahoo Finance. Although the data has not been processed nor cleaned, the size should be large enough to test out our hypothesis. Given the amount of the data, GPUs are possibly required, and we will utilize Google Colab to do this. If computation resource is still insufficient, further truncation of the data, such as limiting down the number of days or/and companies, will be considered.

Computation Resources

Given the amount of the data, a GPU may be required, and we will utilize Google Colab to do this. If computation resource is still insufficient, further truncation of the data, such as limiting down the number of days or/and companies, will be considered.

Potential Difficulties

The difficulty begins with feature construction, such as extracting enough valu- able information from the massive data and forming meaningful features. Sec- ond, the sentiment tags and ranges we choose will directly affect the reliability of the mood measurement and last will be finding the most effective model.

References

[1] One hundred nasdaq 100 companies - free twitter datasets - followthehash- tag.
[2] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1–8, Mar 2011.
[3] Jessica Bursztynsky. Twitter shares close down more than 6 percent first trading day after trump ban, Jan 2021.
[4] Anshul Mittal and Arpit Goel. Stock Prediction Using Twitter Sentiment Analysis. 2011.