Project Proposal: Sentiment Analysis on Stock Price Return

Problem Statement and Background

Most financial models assume that all investors are rational, which means for those investors following the same principle, their reactions to the same specific event should be the same optimal one. However, it is often not the case in the real world. Studies have shown that stocks' returns do have some relationships between investors' behaviors. [i] [ii] [iii] Due to human nature, even a relatively rational investor can waver under consecutive bad news, regardless the credibility of the news. This project aims to find the relationship between the investors' behaviors and common stocks returns as well as the relationship between the number of different types of news and investors' behaviors; also, to determine potential leading indicators or key triggers, which in short can be interpreted as signs for future stock movements and are meaningful in investment strategy design.

Main Reference

We observed that people have already used sentiment-based prediction on Cryptocurrency price fluctuations[iv], which contains hourly data from tweets classifying as positive, neutral and negative. Our main idea is to apply the same approach on equity market, concentrating on the stock return.

Methods and Algorithms

The project will be implemented mainly using XGBoost[1] (Extreme Gradient Boosting), which is a supervised learning algorithm often produces easily interpretable results and is also the basis of our main reference. Besides, for comparison purpose, other algorithms such as Bayes classifier and SVM will also be implemented.

Data Set

The main data comes from a user behavior data set of one of the most popular stock application in China[2] with daily data ranging from 2013 to 2017 containing the following information. Also, the corresponding stock returns data are obtained directly from Yahoo Finance.

Variable	Type	Description
Positive & Negative News Numbers	Int	The number of positive and negative news, respectively
Favorite Percentage	Float	(number of users who add stock A to their favorite list)/(number of users)
Click	Int	number of users who click stock A for information
Click Percentage	Float	Click/(Total click number)
News Click Percentage	Float	(Click number of news on stock A)/(total click number of news)
Favorite Percentage Group by Asset	Float	Similar to Favorite Percentage but are grouped by client type
Favorite Percentage Group by Experience	float	Similar to Favorite Percentage but are grouped by client experience

Team Member and Assignments

Yuankang Xiong, Yifei Lu: Data screening and model implementation

Yinyuan, Rui Xiao: Model evaluation and essay and report writing

At the end of the project, we would figure out the following results: the relationship between investors' behaviors and returns of the common stocks they are interested in or invested; the relationship between the number of different types of news and investors' behaviors; finding the leading indicator or some key triggers.

---

[1] A brief introduction can be found at https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/ and

[2] The data set is not open to public, hence we only obtained the 2013-2017 data, which is available on https://github.com/RickYuankangHung/Quantitative-Finance/blob/master/PosAndNegNewsNumber.xlsx and due to some reason the application name is classified.

Project Proposal: Sentiment Analysis on Stock Price Return

Reference

i T. Hai, N. Kiyoaki, Sentiment Analysis on Social Median for Stock Movement Prediction, Expert Systems with Applications Volume 42, Issue 24, Pages 9603-9611

ii N. Oliveira, P. Cortez, N. Areal, The Impact of Microblogging Data for Stock Market Prediction: Using Twitter to Predict Returns, Volatility, Trading Volume and Survey Sentiment Indices, Expert Systems with Applications Volume 73, Pages 125-144

iii L. Xiaodong, X. Haoran, C. Li, News Impact on Stock Price Return via Sentiment Analysis, Knowledge-Based Systems Volume 69, Pages 14-23

iv Tianyu R.L., Anup S.C., Xander R.F., Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model https://arxiv.org/pdf/1805.00558.pdf