Advanced Big Data Analytics
Proposal

AI Trader in the US

Qing Gao: qg2175

Wenfeng Lyu: wl2733

Date: 2/14/2020

Abstract

The topic of predicting returns, stock price or market movements is very crucial for investors, as an accurate forecast will allow them to make appropriate trading decisions and realize an attractive return. Using traditional asset pricing models to estimate returns is time-consuming, AI Trader thus far has been developed as a product of the technology improvement. AI Trader uses machine learning models and different algorithms to train the past market data and recognize stock market trends and patterns to help making trading decisions. The goal of the project is finding the best strategy for each investor based on each investor's preference and budget. Instead of choosing individual stocks, investors can choose their investment goals and risk thresholds, and the Robo-advisory program will automatically create investment portfolios for them.

I. Background

Stock market forecasting and developing profitable trading models have always attracted researchers and practitioners. However, it's very challenging to come up with a stable model that works reliably under different market conditions. Computational intelligence models started emerging as viable alternatives to the traditional stock trading decision support. Consequently, AI Trader has been developed over the past few years. AI Trader uses machine learning models and different techniques to train large past trading data and recognize market trends and patterns to make trading decisions. The advantage of AI Trader is being able to train a huge dataset and find trends and patterns of the market, as well as consider parameters that are humanly

impossible to guess. Moreover, unlike those trading decisions made by humans, AI Trader can make calculated, accurate and unbiased decisions, and it offers a better and more accurate predictability tool that will allow investors to realize abnormal return. In the meantime, the rise of computational intelligence such as neural networks has illustrated good performance achievements over the past few years. In particular, as implementing machine learning models using big data is becoming mainstream, the majority of transactions that are executed by automated trading agents in the New York Stock Exchange.

Currently, analysts are using different kinds of algorithms and models to analyze market performance in order to predict the stock trend. However, due to the randomness in the stock market, analyzing the past stock price cannot produce a reliable portfolio with a stable return.

II. Introduction to the project

Nowadays, machine intelligence is playing a crucial role in the US stock market. This project will train an artificial intelligence model as an automated trading agent with a specification in the US stock market. The goal of this project is using the trained automated trading agent to find the best strategies for each investor based on each investor's preference and budget. Instead of choosing individual stocks, investors can choose their investment goals and risk thresholds, and the Robo-advisory program will automatically create a portfolio for them to maximize their market returns with a return-risk balance. One of the methods we will use to achieve our final goal is Long Short-Term Memory Networks (LSTM). They were introduced by Hochreiter & Schmidhuber (1997), and were refined and popularized by many people later on. LSTM is a special type of Recurrent Neural Network that can be useful in sequential data such as time series. It is advantageous to use in our project because it overcomes a vanishing gradient problem in a recurrent neural network (RNN) to learn long-term dependencies in time series data using memory cells and gates. Moreover, we will use the SVM model as our second method. By using SVM model, we can classify the stocks into different groups based on their daily returns and improve the model by training with different classifiers.

However, there are few drawbacks to the AI trader. One drawback is that it does not have the capability to handle extreme scenarios. Many economists claim that the stock prices already reflect all public information, therefore it is impossible to predict the future stock price. Many current stock prediction models are able to have a good fit on the historical data, but they have low accuracy in predicting future trends. For example, the recent coronavirus outbreak brought a lot of variation and randomness to the stock market in the short term both in the US and China stock market. These variations and random events are not able to be predicted by the past stock price.

Therefore, in this project, we will use different approaches to validate and improve the accuracy of the models. We will train different models to find one which produces a stable return. To validate our models, we might measure the prediction accuracy by loss function and AUC, and fitting the models with data in different time periods.

III. Introduction to the Dataset

The dataset would be the data of the S&P 500 companies' historical stock price, including the open price, close price, and volume on a daily basis. The data would cover the stock price from 2008 to 2020. Also, we need the company's financial data to evaluate the company's performance. The stock prices and company financial statement data would be collected from Yahoo Finance or Google finance. We might use web scraping to collect the most up-to-date data.

The challenge of data collection is that the dataset would be very large and it would be time-consuming to do data cleaning and data prepossessing. After preprocessing the data, the challenge is ingesting and interpreting the data to determine which data is useful, finding the signal in this sea of information. Also, the size of the datasets would influence the time of testing the models.

IV. Plan

Timeline	Item
Milestone1	Read relevant research papers Data gathering Data cleaning
Milestone2	Build up models & portfolio Train models Generate prediction function Feature selection for our initial model
Milestone3	Test the accuracy of our model using historical trading data from 2008 onwards Improve model prediction accuracy as needed
Final	Select the stable model and perform its return Check for the robustness of the results given different network configuration Preparing the final project report and presentation

V. Reference

Alexandra Zavadskaya. "Artificial Intelligence in Finance: Forecasting Stock Market Returns Using Artificial Neural Networks." *Hanken School of Economics*, 2017, https://pdfs.semanticscholar.org/28ce/a2121c67331973d8ed7b6aa7291305aba974.pdf

Christopher Olah. "Understanding LSTM Networks." *colah's blog.* github page, 27 Aug 2015, https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Kim, Taewook, and Ha Young Kim. "Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data." PLOS ONE, Public Library of Science, 15 Feb. 2019, journals.plos.org/plosone/article?id=10.1371/journal.pone.0212320.

Ponduru, Aneesh. "Artificial Intelligence: A New Age in the Stock Market." *Medium*, Data Driven Investor, 26 Oct. 2019, medium.com/datadriveninvestor/artificial-intelligence-a-new-age-in-the-stock-market-991ac08aa11e.

Roshan Adusumilli. "Machine Learning to Predict Stock Prices." Towards Data Science, 26 Dec 2019, https://towardsdatascience.com/predicting-stock-prices-using-a-keras-lstm-model-4225457f0233.

Sezer, Ozbayoglu and Dogdu (2017) . "An Artificial Neural Network-based Stock Trading System Using Technical Analysis and Big Data Framework". Available at: https://arxiv.org/pdf/1712.09592.pdf