Project Proposal

Stock categorization by price fluctuation

creating a subset of stocks that rise and fall in a similar pattern

Name: Sun Wu Choi

**[1] Who is on the team? Is the team size rational?**

Team size is one person, me. I am planning to create a Deep Neural Network system to find the relationship between companies. The project might require group of two, but I decided to solely focus on myself and accept the challenge. The team size is rational.

**[2] Problem we are trying to resolve, the formulation, what is the input and output?**

I am trying to give a better solution for the investors in two ways. First, categorize the stocks that shows a similar fluctuation pattern. Second, show how much each company is related to each other with some number between -1 to 1.

The formulation

Initial State: there is no pattern found in all the data, raw initial big data

Actions: One epoch will try to learn the optimal way to categorize the data

Transition Model: One epoch will result in slightly better or worse categorization method

Goal Test: It will be able to predict a certain stock based on the other stocks it is related to, providing the category and the stocks it is related to with a parameter between -1 to 1

Path Cost: Epochs will consume time and computing power based on the input size, number of neural networks per layer, and number of layers. Also, the learning rate, drop rate, and convolutional layers

**[3] Ideal outcome of the project, what do you expect to show**

The ideal outcome of this project is whenever a random stock is selected, it shows the related stocks and its relation rate. A stock can affect another stock in one way, or both ways, or not at all. Ideal model is to show the relation rate between stocks with directions such as A affects B, B affects A, or A and B affects each other. The sum of the related stock parameters will be able to predict a stocks fluctuation with 50% or higher. For example, Tesla stock price affects LG Chemical and Panasonic, the main manufacturer for Tesla battery, and ideally our model will show that Tesla stock affects LG Chemical and Panasonic with a positive one way or both ways. Because if Tesla sale increases, logically LG Chemical and Panasonic will make more money by selling more batteries.

**[4] What algorithm expected to use**

I am expecting to develop this model using Python TensorFlow and Keras. Mostly based on Deep Neural Networks, but since this is not a simple input to output system, I might implement Hebbian learning method.

**[5] What topics / libraries / platforms will you have to learn?**

Economics, machine learning, algorithms, neuroscience will be needed to develop this model.

I am planning to develop in Python using PyCharm, using TensorFlow, sci-kit, pandas and many more libraries that might be necessary.

[6] Where will the dataset come from? Do you have the data set resource?

The dataset will come from RapidAPI.com Bloomberg financial API and Morningstar API. I can request any data from these API and that means I do not own the dataset, but I have full access to them at any moment.

**[7] Define the milestone for the project ( due on 11/3 and 11/24 )**

Due on 11/3: Setup a Deep Neural Network for this model and try to teach the pattern from the data. The priority for this milestone is to get used to the environment and logically approach the problem with a correct method. Optimal milestone is to generate some sort of relationship map of Tesla, which is the most controversial company right now.

Due on 11/24: Provide a relationship map of Tesla and tweak the parameters to increase the relationship prediction rate. Think outside of the box to fundamentally solve the errors of this problem, especially think this relationship map like figuring out the connectivity of a brain. Optimal milestone is to have 30% or higher relationship prediction rate and figure out the relationship of companies that is related to Tesla, which can be LG Chemical. Figure out if Tesla and LG Chemical is a one-way relation, or both way relation. By achieving this relationship map between Tesla and LG Chemical, it can be recursively applied to any dataset and show a feedback map like mapping connectivity between neurons in a brain.

**[8] Week by week plan**

10/21 – 10/27: Setup the dataset, preprocess the data and think about the methodology

10/28 – 11/03: Start to apply DNN to the dataset and see the outcome, check for any progress

11/04 – 11/10: Debug the DNN and develop a GUI for the map

11/11 – 11/17: Keep improving the model and keep the methodology in mind

11/18 – 11/24: Check if the first milestone is achieved, check if Hebbian learning is available

11/25 – 12/01: If milestone is done, try to apply to other stocks recursively, prepare the report

12/02 – 12/08: Clean up the model and the GUI, polish the presentation and the report