Evolving Optimal Stock Market Predictions Through Use of Multi-Objective Genetic Algorithms

Farhan Hormasji Bonnie Reiff

I. Background and Motivation

In the area of economics and financial markets, there is currently no leading philosophy when it comes to predicting market trends. Various approaches include random prediction, buy and hold (in which one should buy if the market is positively trending and hold until it goes down), and lastly, using mathematically derived predictors. As more research has been done to prove technical approach's superiority, people have started investigating how evolutionary algorithms and genetic programming can be applied to financial trading and analysis. These algorithms aim to determine good indicators of trends and predict the mean stock/index price, as well as maximize profit while minimizing risk.

II. Project Goal

Recently, methods such as feature extraction using a basic genetic algorithm [2] and multi-objective genetic algorithms (MOGAs) for parameter optimization [3] have been tested for stock market analysis. The goal of this project is to use MOGAs to optimize the technical indicators of when to buy and sell stock. The weighted optimization of parameters will be evaluated by how well they meet the objectives chosen. Initially we will focus on optimizing parameters to meet two basic objectives, profit maximization and risk to return ratio. There has been research into optimizing these parameters, and one of the most common algorithms used and the one we will implement is NSGA-II.

The next step is to add objectives to optimize in order to improve prediction analysis. Since NSGA-II does not handle more than two objectives well, we will implement an NSGA-III algorithm. We have found no research into this algorithm being applied to financial market prediction.

III. Basic Method Details

a. Data

Financial data for stocks and indices is widely available online via informational websites such as "Yahoo! Finance", data set repositories such as UC Irvine's Machine Learning Repository, and data mining competition websites such as KDD and Kaggle.

One data set that the team looked at is Dow Jones Index Data Set from the UCI Machine Learning Repository [1]. There are 750 data instances in the set (where each record is data for a week), each containing 16 attributes. These attributes include information that will be needed for the prediction of market trends and well as information against which those predictions can be checked.

b. Evolutionary Algorithm

In order to achieve optimization of the technical indicators involved in the prediction of stock market trends, a multi-objective evolutionary algorithm (MOEA) needs to be applied. The NSGA family of algorithms was due to the fact that multiple versions exist – using these different implementations, we can perform testing to determine which variation works best for the given objective.

There are several different technical indicators (with just two objectives) which are commonly used by financial analysts that can be optimized. For NSGA-III, either new objectives to optimize will be added to the existing NSGA-II problem or different indicators that have at least three objectives will be chosen.

The following are initial technical indicators which will be used in the first NSGA-II pass [3], [6]:

- Relative Strength Index (RSI) – a momentum indicator (between 0 and 100) that compares market recent gains against recent losses
$$RSI = 100 - \frac{100}{1 + RS},$$
where RS is the ratio of ups to down over n days.
- Exponential Moving Average (EMA) – average value of data over a specified time period
$$EMA(i) = W_1 * price(i) + W_2 * EMA(i-1)$$
$$W_1 = \frac{2}{n+1}, W_2 = 1 - W_1,$$
where i is the day, and n is the number of days for calculating the average.
- Moving Average Convergence/Divergence (MACD) – used to tell whether the stock/index is moving towards or away the 'signal line', which determines whether you should buy or sell. It is the difference of the long term and short term EMA.

i. Chromosome

The chromosome representation for the MOEA contains one section for each parameter in the technical indicators that we need to optimize.

ii. Fitness Function

The objectives of the prediction algorithm are to maximize profit and minimize risk. There are several fitness functions to measure this, including [3],[5]:

- Maximum return = average profit return
- Minimum Variance of the Results versus Return with risk:
$$Average\ Profit\ Return - Risk\ Aversion\ Factor * \sigma,$$
where σ is the standard deviation
- Risk-adjusted return (Sharpe Ratio) = $\frac{\mu}{\sigma_p}$, where μ is the mean value of all returns and σ_p is the standard deviation of returns
- Minimize Maximum Drawdown (a financial indicator of risk)
- Annual return:
$$\left[\left(\frac{Return}{Captial}\right)^{\frac{1}{n}} - 1\right] * 100,$$
where n is the number of trading days / 250.

iii. Parameters

As with any evolutionary algorithm, there are several parameters that need to be defined and tuned for optimal algorithm performance. These include initial population size, how the population is generated, crossover and mutation rates and types, and selection mechanisms for which parents should be used to create offspring and which offspring survive into the next generation.

A paper found on the use of multi-objective optimization on stock market data suggests the following as parameters to the authors' chosen algorithm [3]. These values may provide a good starting point for the algorithm trials for this project:

- Initial Population Size: 100
- Selection Technique: Rank-based Selection
- Crossover Probability: 0.8
- Mutation Probability: 0.05

c. Software Implementation

The plan for this project is to use existing implementations of the NSGA family of algorithms. The two implementations chosen are Matlab for NSGA-II and C++ and Java for NSGA-III.

IV. Team Collaboration

The project team plans to use a Github for the storage and version control of all technical data for this project. Technical data here refers to files containing parameters for the genetic algorithm, source code, and results files.

V. References

[1] "UCI Machine Learning Repository: Dow Jones Index Data Set." *UCI Machine Learning Repository*.

[2] He, Yuqinq, Kamaladdin Fataliyev, and Lipo Wang. "Feature Selection for Stock Market Analysis." *Neural Information Processing Lecture Notes in Computer Science* (2013): 737-44. Web.

[3] Fayek, Magda B., Hatem M. El-Boghdadi, and Sherin M. Omran. "Multi-objective Optimization of Technical Stock Market Indicators Using GAs." International Journal of Computer Applications, Apr. 2013. Web.

[4] Yuan, Yuan, Hua Xu, and Bo Wang. "An Improved NSGA-III Procedure for Evolutionary Many-Objective Optimization." *GECCO Proceedings* (2014).

[5] Pinto, J., and R. Neves. "Maximize Profit While Minimizing Risk - A Fitness Function Evaluation Study Using Genetic Algorithms." May 2012.

[6] Hegazy, Omar, Omar S. Soliman, and Mustafa Abdul Salam. "LSSVM-ABC Algorithm for Stock Price Prediction." (2014): n. pag. International Journal of Computer Trends and Technology. Web.