

# Research into the Modeling of Complex Systems

## Predicting the Stock Market

John Sherwood  
Period 2

### Abstract

My project is involved with using data mining techniques on the internet in order to gather enough information for the use of a genetic algorithm in trend analysis of a complex system; e.g., the stock market. To do this, I use a data mining algorithm in order to gather information on various stocks, and then use different algorithms to make decisions about buying and selling stocks.

### Phase 1

The early version of my program used data gathered by "Albert", the alpha data miner that used information from Yahoo! RSS stock news feeds. The program set up various subroutines, referred to as bots, that analyzed the data and used it in order to determine what stocks should be bought and sold. The bots were, on average, capable of turning a 7% return in less than a month, and Bill, the second bot I wrote, even managed to turn a 26% return in that period (a picture of his data is below). However, the market was doing quite well during that timeframe, and although such returns were far above the average increase in the market, it appeared that a better standard was needed to determine a successful prediction methodology.

### Phase 2

After the early stage where the Yahoo! RSS feeds and the stockbots were used, I decided to begin writing algorithms to calculate discrete price change in stocks over fixed intervals. To model the changes, I chose to analyze different pieces of news and determine what category they fell into, then quantitatively rank them as positive or negative news, and using all the data, create a system of equations based on determining the effect of the news based on the time since the news broke, the 'goodness/badness' of the news, and the type of news. The equations are in the format  $K \cdot C^t$ , where K is the value determined based on the 'goodness/badness'

score of the news, C is the constant associated with that type of news, and t is time passed. The equations are refined using a genetic algorithm to check forecasts at various points in the past against the actual prices. For example, the genetic algorithm might calculate the price one might expect for the previous Thursday based on data gathered up to the previous Wednesday, then see how accurate that prediction would have been and accordingly edit the equations

### Methodology

For the new data miner to replace Albert, I decided that I needed an algorithm capable of parsing XML/HTML files, and thus began to write an iterative parser. Once the parser is complete, I intend to begin using the Google search engine to search for data, and thus store much more relevant information than the previous data found by Albert. For the classification of the news, I have already written an algorithm that, given a sampling of already categorized news, determines combinations of discriminants to effectively categorize other news.

## SherwoodForge

HOME

DOCUMENTS

INFORMAT

Current liquid assets: \$8.651799999994

### Portfolio

Stock	Shares	Current Price	Total Value
TRFC	7616	5.56	42344.96
AOB	5779	7.5977	43907.1083
WNI	6533	6.18	40373.94

Total value of Stocks: **\$126626.0083**

Total assets: **\$126634.6601**

### Expected Results

In doing this project, I expect to at the very least have a very useful genetic algorithm, that given a list of independant and dependant data, can generate equations to create a tentative correlation. While the extremely chaotic nature of the specific application may prevent quantitative success in this instance, I do expect to have success on general ter