Candlestick Stock Analysis with Genetic Algorithms

Peter Belford
Symantec Corporation
peter_belford@symantec.com

ABSTRACT
Candlestick analysis, a form of stock market technical analysis, is well suited for use with a genetic search algorithm. This paper explores an implementation of marrying these two techniques by creating agents that attempt to identify stocks that will change in price. The best of run individuals, produced by the genetic algorithm, performed statistically better than an agent that makes random investment decisions.

Categories and Subject Descriptors
I.2.8 [**Artificial Intelligence**]: Problem Solving Control Methods, and Search – *heuristic methods.*

General Terms
Algorithms, Economics.

Keywords
Artificial Intelligence, Genetic Algorithms, Stock Analysis, Candlestick Method.

1. Introduction
One of the first steps in making profitable investment decisions is to correctly identify the direction a stock will move [1]. A stock may move in three directions: up, down and sideways. Stocks that will be moving sideways should be avoided as profit potential can often be less than transaction costs. Investors profit from an up trend by opening a long position, and a short position is opened when the prediction is a down trend. A position is then closed when it is determined that the trend will no longer continue. The trend may reverse or may start moving sideways.

For the purposes of this paper candlestick analysis (CA) graphs a single stock's price data for a single day in a picture that somewhat resembles a candle [2]. For more information on the diagram see Morris, 1992, or conduct an internet search. These "candle lines" may then form into patterns that indicate a bullish (up) or bearish (down) trend. This pattern predicts the trend will reverse to an up trend.

These candle patterns are normally formed by three candles, or three days of trading. In order for a pattern match to occur each of the candles must have certain attributes. One pattern that was used in this project is known as *three white soldiers* (TWS). In order for a match to occur the stock must be in a down trend, there needs to be three *long* days all with *short shadows* as well as some additional restrictions.

GECCO'06, July 8–12, 2006, Seattle, Washington, USA.
ACM 1-59593-186-4/06/0007.

Although CA is good at recognizing the start of a profitable trend it cannot always be used to determine when the trend will cease [2]. CA is not exhaustive so it may not recognize when a trend will reverse, and there are no indicators when a trend will go sideways. For these reasons CA is not suitable for determining when a position should be closed. The purpose of this project was to examine the effectiveness of CA and genetic algorithms (GA); therefore, developing an agent that determined when to close an investment position was beyond the scope.

2. Methodology
A *long* day is one that is defined as a large difference between a stock's opening and closing price [2]. A day with *short shadows* is one in which there is little difference between the day's high price and the closing price, and little difference between the day's low price and the opening price. No text on CA can give a deterministic definition for either one of these two candle attributes.

In order to recognize this genre of "soft" attributes, functions were written that returned an integer. The greater the value returned the more likely the candle possessed that attribute. For example the function *isLongDay* might return a 2 if there was a small difference between the opening and closing prices, and an 8 if there was a large difference between the two prices.

The agents that were tasked to look for an occurrence of a TWS pattern were given a threshold value for each of the soft attributes that made up the pattern. One instance of a TWS agent might require at least a five for the length of the first day, where another might need a three or greater. All of the threshold values were encoded into the GA string.

Stock price data was gathered on one hundred diverse companies for a six month time period. Two separate time periods were gathered. One time period would be used as a training data set, the other for evaluation. For the GA run individuals were exposed to the training data set. When a pattern match was made a simulated investment position was opened. The fitness of an individual was the sum of the profit from each investment opened minus some transaction cost. In short, the more money an individual made the more fit the individual.

In order to determine the profitability from an open investment position it must first be closed. This project assumed that there existed a highly efficient "when to close an open position" agent. After a position was opened, the next fifteen trading sessions were examined. The position was then closed with the most beneficial price for the transaction. Although this may seem a bit artificial all transactions were closed in this manner. This includes those opened by the random agent.

To provide a bench mark an agent was developed that made random investment decisions. The random agent was run against the evaluation data set several times. This agent had an average profit of 39.9 per transaction with a 95% confidence interval (CI) of 24.14 to 55.66. In comparison the best in class TWS agent had an average profit per transaction 127 for the evaluation set, well outside the CI of the random agent. While much more work remains this methodology has so far proven successful.

3. REFERENCES

[1] Murphy, J. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications.* New York Institute of Finance, New York, NY, 1999.

[2] Morris, G. *Candlestick Charting Explained*. McGraw-Hill, New York, NY, 1992.