Trading in Financial Markets using Pattern Recognition Optimized by Genetic Algorithms

Paulo Parracho Rui Neves Nuno Horta

Instituto de Telecomunicações
Instituto Superior Técnico
Torre Norte, AV. Rodisco Pais, 1
1049-001 Lisboa, Portugal

paulo.parracho@ist.utl.pt; rui.neves@ist.utl.pt; nuno.horta@lx.it.pt

ABSTRACT

In financial markets there is a theory called *Random Walk*, which is supported by academics in economics and finance. This theorem claims that it is impossible to predict the future evolution of any financial asset, since its intrinsic value is already reflected on itself. This means the asset's price already reflects all the information available in the market. Therefore it is customary to argue that you cannot beat the market, since all the actions performed in it are in perfect balance, and any research that is made in order to get higher gains than those that come from a *Buy & Hold* strategy will be completely futile.

Technical analysis in financial markets literature documents the existence of certain chart formations, which once completed advocate the future trend of the underlying asset, with a high percentage of accuracy.

Thus, the aim of this work involves the application of chart pattern and trends detection techniques on financial markets, using historical prices of the underlying financial assets. On a sideline we intend to use some additional technical indicators that allow us to sustain the decision process and confirm/estimate the future evolution of a particular listed company, given previously by the analysis of chart patterns. That is, we will use other technical indicators, in addition to the discovery of the aforementioned chart patterns, to maximize the profitability obtained and to maximize our investment strategy in the face of others, like *Buy & Hold*.

Categories and Subject Descriptors

I.2.M [**Artificial Intelligence**]: Miscellaneous.

General Terms

Algorithms, Economics.

Keywords

Graphic patterns, Trends, Financial Market, Stock, Profitability, Technical Indicators.

1. INTRODUCTION

Financial theory considers the hypothesis of Random Walk, which determines the evolution of prices in the capital markets according to a "random walk" and argues that the price quoted can not be predicted. Historically, this theory is accepted by economists, since they have several tests and assumptions that support the idea that the evolution of the stock price is completely random, due to efficient markets. They argue that there are opportunities for above-normal profits, and that all the effort spent on research, analysis and trading in the stock market is a complete waste of energy and resources.

In this paper we intend to contradict this hypothesis and discuss the core idea that anyone, not just a professional in the field of financial markets, since it has the appropriate knowledge and relevant information, you can safely earn on the stock exchange, making use of a strategy averse to risk, based on technical analysis or fundamental.

There are always many questions asked by investors in their decision process, being crucial to obtain a reply in relation to when and what to buy or sell. It is this direction that we focus our senses, ie to determine the time.

It becomes therefore crucial to define a simple strategy and winning. It is on this basis that we will conduct this study. We seek to achieve an automated trading system, neglecting to do so what you buy and emphasizing the timing of purchases and sales, based on heuristics that rely solely on historical share price and its volume.

In technical analysis recourse is usually to identify patterns and trends in the historical price and volume to achieve the anticipated future behavior in the evolution of prices in financial markets. This type of analysis is called for stock charting. What this means is that depending on the chart formations in the past we can predict with some probability, to rise or fall of the capital market in near future.

The problem is that the analysis and pattern identification tasks are very time consuming if done only visually, since there are thousands of assets listed on financial markets, trading histories that can reach some tens of years. The aim of this work is to build a tool that allows the automation of search and identification of chart patterns, the history of financial assets in order to assist investors in decision-making process. That is, the tool will have the ability to place purchase orders and sales automatically, depending on the outcome of its analysis thereof.

In the perspective of optimizing the search space of this major problem, which is to analysis and detection of chart patterns and trends, we have the broad objective of our work in integrating the concept of genetic algorithms and evolutionary computation, in order to overcome the solutions usually encountered by human agents using traditional approaches to computing (mainly due to limited processing and storage). Evolutionary computation will thus find solutions to problems of demand and thorough optimization.

In addition, we also combine other technical indicators that will support and help to ensure that buy/sell decision making have been the most appropriate in probabilistic terms.

This paper is organized as follows; section 2 provides an overview over the methodologies we used to develop the system. Section 3 validates the proposed system solution against a Buy & Hold approach. Finally, section 4 summarizes the work that was developed and section 5 points toward future directions.

2. RELATED WORK

In recent years the area of financial investments has attracted huge interest within the scientific community. And within the theme of chart patterns have arisen several methodologies for their detection in the history of assets in financial markets.

So in the following sections highlight different approaches to solving this problem, proceed to their respective analysis. We consider three main techniques used to detect patterns. One is based on the application of templates, one for detection of perceptually important points, and other positional rules in applying for these spots detected previously.

The plan of this work consists mainly in the choice of one of these strategies, thus seeking to converge to the spectrum of the graphic pattern detection methodology that we use to make buying and selling shares in the financial markets.

2.1 TEMPLATES HEURISTIC

In the approach in [8] [9] and [10] considers the application of an algorithm based on a template, the detection of Bull Flag pattern [3], which signals a rise in prices in the near future.

The main objective is to confirm that technical analysts can predict future stock prices, considering only the past of these same assets, and thereby gain higher than average return on financial markets.

The detector is created together with some simulated trading rules on the NYSE index, the last 35 years, considering all the closing prices of each day.

The results of this work are systematic and consistent, and help to achieve this objective, ie, greater than the gains from a completely random approach.

-1	-1	-1	-1	-1	-1	-0,5	0	0	1
-0,5	-1	-1	-1	-1	-0,5	0	0	1	1
-0,5	-0,5	-0,5	-0,5	-0,5	0	0	1	1	1
-0,5	-0,5	0	0	0	0	1	1	0,5	0
0	0	0	0	0	1	1	0,5	0	0
0	0	0	0	1	1	0,5	0	0	0
0,5	0,5	0,5	1	1	0,5	0	0	-0,5	-0,5
0,5	0,5	1	1	0,5	0	-0,5	-1	-0,5	-0,5
0,5	1	1	0,5	0	-0,5	-1	-1	-1	-1
1	1	0	0	-0,5	-1	-1	-1	-1	-1

Figure 1. Matrix Template for the Uptrend Pattern.

This methodology seeks to detect the pattern in Figure 1, which usually means the formation of a consolidation period characterized by fluctuations within a narrow band. Technical analysts interpret this sign as a temporary interruption of a strong upward trend, in which investors consolidate their gains before the action resume its positive sense.

As confirmed in Figure 1, we have a matrix of 10 x 10 which I shall call "T", whose cells have values between -2.5 and 1.0. Cells without content equivalent to the value 0.0.

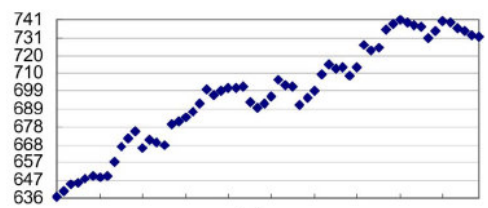
The seven columns on the left define the consolidation movement while the other three define the upward movement (called "Breakout"). The sum of the weights in each of the columns makes 0, and these are assigned to each cell of the array so that all the patterns that deviate from this template contribute to a negative factor for the fitness function.

Considered in the study were also created images reflecting the history of prices, with a fixed size of 10 x 10 and we'll call the "I". These new arrays are based on the concept of sliding window. That is, each of these matrices "I" will include in your image the temporal extension of the historic prices that we intend to, and which may be several days. Thus we find the same pattern in time windows with different lengths.

Within each time window, we will remove the noise present, replacing the closing prices that exceed the boundary value from the average of prices considered.

Then subdivide the time window in 10 equal groups, and each of these groups will be mapped in one of the columns of "I". This process will compress the time frame considered in a matrix of 10 x 10, which will maintain the essential features of the original pattern. Each cell in the new matrix is subjected to a factorization function, so that the field values are between 0.0 and 1.0.

The following is an example of a time window of 60 days (a) and its matrix "I" (b):



(a)

0.	0.	0.	0.	0.	0.	0.	0.33	0.83	1.
0.	0.	0.	0.	0.	0.	0.	0.5	0.17	0.
0.	0.	0.	0.	0.	0.	0.5	0.17	0.	0.
0.	0.	0.	0.17	0.5	0.5	0.5	0.	0.	0.
0.	0.	0.	0.5	0.5	0.5	0.	0.	0.	0.
0.	0.	0.33	0.33	0.	0.	0.	0.	0.	0.
0.	0.33	0.33	0.	0.	0.	0.	0.	0.	0.
0.	0.17	0.33	0.	0.	0.	0.	0.	0.	0.
0.33	0.5	0.	0.	0.	0.	0.	0.	0.	0.
0.67	0.	0.	0.	0.	0.	0.	0.	0.	0.

(b)

Figure 2. (a) Temporal window of 60 days; (b) Matrix "I".

Once matrix "I" is obtained, a function of Fit is computed, based on the multiplication of this matrix and the matrix T, i.e.,

$$\text{Fit}_k = \sum_{i=1}^{10} \sum_{j=1}^{10} (T(i,j) \cdot I_k(i,j))$$

Thus, the higher values of the function Fit will occur when the image in the array "I" is more in keeping with the image matrix template "T".

However, for each time window, we get yet its height of the window as follows:

$$\text{Height}_k = \text{range}_k / p_k$$

In the end, these values obtained in Fit and Height are used in the parameterization of trading rules, identifying which days are most appropriate to buy or sell (in the case of a reverse pattern).

2.2 PERCEPTUAL IMPORTANT POINTS

As is known, the historical price action is constructed from a given sequence of data points. In turn, the amplitude of each of these points will have a different influence on the perception of the shape of the series. That is, each point will have its relevance. That is, we both have a point that contributes significantly to the overall shape of the chart pattern, how can we have another point to exercise little influence on the series, and therefore can be discarded.

Therefore, it becomes that the chart patterns, technical analysis, are typically characterized by only a few salient points. For example we can consider the standard H & S (Head and Shoulders), which consists of a point on the head, two shoulders and two more on the neckline.

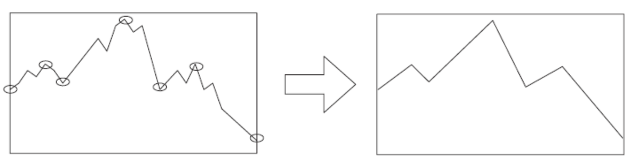


Figure 3. 7 Identification of PPIs in the H & S pattern.

These points are perceptually important in the human visual identification. Are more relevant than all the other data points within the same series, and we designate them by the Perceptually Important Points (PPIs). [16]

The process to identify these points is described as follows:

Given the price history of a particular financial asset P, all data points, p1, ..., pm, at P will be subjected to the identification of PPIs according to the method defined in [17].

Thus, the first two matches PPI are respectively the first and last points within P. The PPI will be the next point P with the greatest distance from the first two PPIs. The PPI will then be the fourth point P with the largest distance to any two adjacent PPI, or will be located between the first and second PPI or between the second and final PPI. This process continues until PPIs location of all points of P are connected to a list.

To calculate the longest distance from one point on two adjacent PPIs, we propose the following method:

2.2.1 VERTICAL DISTANCE

This methodology, identified in Figure 5, computes the DV between the test point P3 and the line connecting the two adjacent PPIs, i.e.,

$$\mathrm{DV}(p_3, p_c) = |y_c - y_3| = \left| \left(y_1 + (y_2 - y_1)\frac{x_c - x_1}{x_2 - x_1} \right) - y_3 \right|,$$

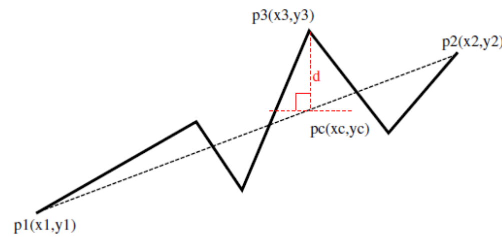


Figure 4 - Vertical Distance in identifying the farthest point.

In the above figure we consider xc = x3. This methodology aims to capture the fluctuation in the wake of the PPIs between adjacent points. In turn, these points will be captured themselves considered new PPIs.

2.3 PATTERN IDENTIFICATION BASED ON RULES

This methodology is defined rules that describe the desired graphics standards. One drawback of the previous methodology is that the relationship between the PPIs is difficult to define explicitly using templates. And it easily surpasses specifying the location of points that make up the pattern.

For example, a standard Head-And-Shoulders, the two points of the shoulders should be at a scale more or less similar (with a height difference of less than 15%) and below the point of the head.

Although it is possible to describe this pattern into the template according to the desired requirements, we can not guarantee the reliability of that identification process in time sequence, since similar patterns may exist that violate the rules defined.

For example, assuming that each of the patterns in Figure 15 PPI has 7 of the SP7 sp1, we can illustrate a set of rules for identifying the Head-And-Shoulders:

- sp4> sp2 and sp6;
- sp2> sp1 and sp3;
- sp6> sp5 and SP7;
- sp3> sp1;
- sp5> SP7;
- diff (sp2, sp6) <15%
- diff (sp3, sp5) <15%

By setting these rules so we can evaluate the temporal sequences of patterns specified in the search. Namely, we first identify the number of PPIs, which in this case is 7, and then validate the time series and consider only those who obey the rules defined.

Although this technique results disappointing overall, it shows its excellence in detecting particular patterns such as Head-And-Shoulders Top and Triple, with a percentage of hits of 100%. [19]

3. PROPOSED SYSTEM SOLUTION

The proposed solution consists on a genetic algorithm that is coupled with an uptrend pattern detection technique. The main goal of this evolutionary computation lies in the finding of an optimal configuration for all the main parameters involved in the matching process of the desired pattern, like noise removal, sliding window size, fit buying value and fit selling value.

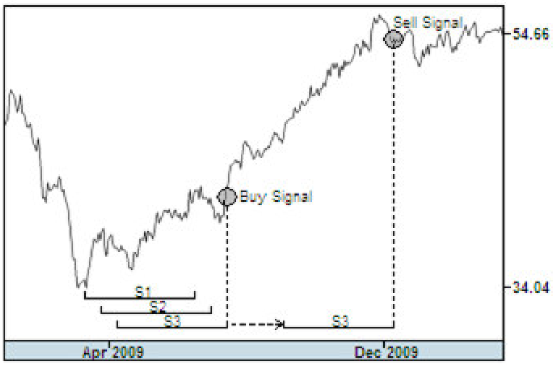


Figure 5. Uptrend found on Novartis' Historical Prices.

3.1 Pattern Detection

As Figure 5 illustrates, the application tries to find the uptrend pattern. It gives the indication of an upward trend, which develops because the demand for shares is larger than the supply of shares. So the algorithm is based on the assumption that once this pattern is detected, the upward trend will continue to follow until at some point a reversal is detected **Erro! A origem da referência não foi encontrada.**. This is the premise we are starting from. So, once this pattern is detected a buying signal is generated, and when the uptrend eventually breaks the software will sell this position.

The algorithm basically consists on the cross multiplication between a 10x10 template matrix that represents the pattern we are looking for (see Figure 2), and whose cell values go from -1 to +1.0, and another 10x10 image matrix that represents the historic prices found on a certain number of trading days, as indicated by the sliding window size (*S1*, *S2* and *S3* in Figure 1). Part of the detection heuristic can be found in **Erro! A origem da referência não foi encontrada.** and **Erro! A origem da referência não foi encontrada.Erro! A origem da referência não foi encontrada.**, although the author used it for a different graphical pattern. For the template matrix creation refer to **Erro! A origem da referência não foi encontrada.**.

By summing the product of the cross-multiplication between the previous described matrixes it is obtained a fitting value that ranges between -10 (for no relation at all with the uptrend pattern) and +10 (for a total relation).

Since the goal is to find the uptrend pattern, which may be defined for different slopes, there are several templates with positive slopes that can range from 22º to 60º. As an example, the template in Figure 2 has an uptrend with a slope of 45º.

3.2 Genetic Optimization Kernel I

A genetic optimization kernel is applied in order to optimize the 4 main parameters, namely, the *Sliding Window Size*, the *Noise Removal* rate, the *Fit Buy* and the *Fit Sell*, which are used together in the detection of the uptrend pattern. *Sliding Window Size* is the number of trading days we are considering each time to compress under the 10x10 image matrix. This value can go from 10 to 100 days. *Noise Removal* is a value defined between 0% and 50%, which designates de amount of noise it will be removed before compressing the time series under the image matrix. *Fit Buy* represents the result from the cross matrix multiplication and indicates the presence of the uptrend pattern. It has a range between +4 and +10. *Fit Sell* it is the opposite of the Fit Buy value and it will indicate the trading point by which we will no longer be in the presence of the uptrend pattern. *Fit Sell* will have to be less than *Fit Buy* and cannot be inferior to -1. Finally, the trading algorithm will acquire a buying position as long as the Fit value is higher than *Fit Buy* and it will sell its position if the calculated Fit value is lower than *Fit Sell*.

The implementation of the algorithm is carried out as follows:

1 - Sliding Window advances one day at a time on the historical financial asset in question;

2 - The noise is removed in each instance of the Sliding Window;

3 - Fitting The value is obtained by multiplying between the Matrix and the Matrix Template image;

4 - If the value obtained Fitting is higher than what is defined in Fit Buy, then the system will make buying a financial asset in question;

5 - If on the other hand the value of Fitting is lower than that defined in the Fit Sell, then the system will sell its stake in the financial asset.

3.2.1 Experimental Results

The application was tested in real market conditions. Data was retrieved from 1998 to 2010 for three main stock indexes: S&P500, Dow Jones Industrial Average and NYSE Composite Index. The GA was trained from 1998 to 2004 and then it was tested from 2005 to 2010. In Figure 3 we have a histogram which demonstrates the profitability distribution for 50 different executions of the GA. The results show the algorithm is capable

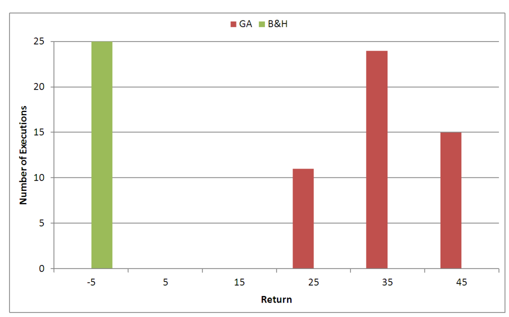


Figure 3. GA approach vs. B&H on S&P500.

of producing consistent and far better results than the B&H strategy.

In Figure 4 it is clearly shown that the algorithm also has the capability of detecting bear markets and staying out of them. The best GA sells its position around January 2008 and only reenters the market in July 2009. As for the average GA, it leaves the market on September 2007, only to reenter temporarily from April to July 2008 and then finally from June 2009 onward.

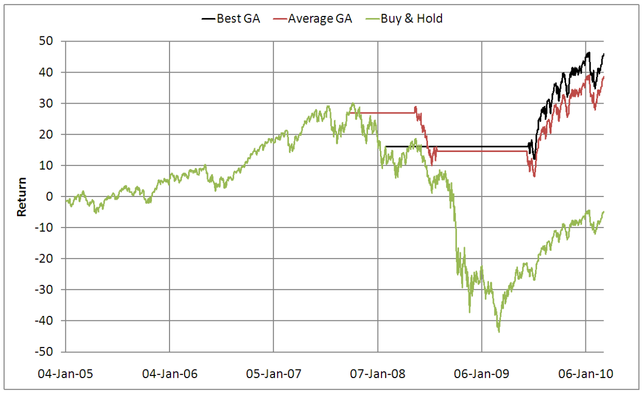


Figure 4. Profitability Growth for three Strategies on S&P500.

Finally, Table 1 depicts the results obtained for three stock indexes. The robustness and superiority of the genetic algorithm is evident by the returns presented.

Table 1. Profitability for Different Stock Indexes.

Measure	S&P500	Dow Jones	NYSE
Average GA	36.92%	16.33%	10.02%
Best GA run	46.02%	30.54%	14.57%
Buy and Hold	-4.69%	0.117%	-1.41%

3.3 Genetic Optimization Kernel II

After the tests performed on indexes, which have produced very positive results, we moved to implement the algorithm in stocks, namely financial assets of listed companies. The returns were disappointing when compared with those obtained by the methodology Buy & Hold. The index essentially captures the movements of a stock exchange and includes the average value of a particular group of shares. They are considered the most significant assets in the total market movement or grassroots enterprises in certain sectors of the economy and its variation reflects the trend of scholarship - which may be high or low.

In previous tests we have only considered indexes which reflect a very low volatility and it facilitated the entry into the asset at the time of detection of the upward trend and the exit in the absence of it.

Analyzing financial assets with an increased volatility seen that the algorithm took on erratic behavior in decision making buy / sell. The genetic algorithm typically chose the training period for the values and fitbuy fitsell that were not consistent with the methodology that was intended to apply to investment in stocks, i.e., buy in the presence of an upward trend and sell it when not in the presence of it.

Thus, we tried to make the genetic algorithm more robust and to achieve this goal we began by observing how it was that the yield obtained was determined using different settings for the values of fitbuy and fitsell.

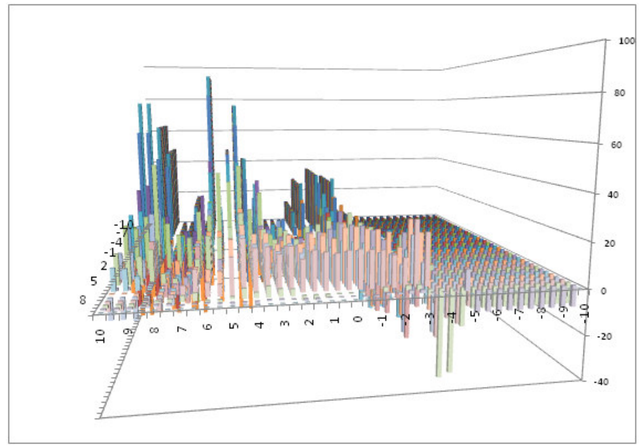


Figure 6: Profitability obtained for the S&P500 for different configurations of FitBuy and FitSell.

Figure 6 profitability is highlighted on the vertical axis and the different values used for FitBuy FitSell and horizontal axes are indicated in.

Thus, remained fixed values for the size of sliding window (60 days) and the value of noise removal (20%). In relation to the values of FitBuy and FitSell we obtained all possible combinations between -10 and +10, with variations of 0.5.

That is, several chromosomes were tested with different characteristics during the execution of the algorithm.

Thus, there were 1681 chromosomes found. For example we obtained the following yields for some possible configurations:

Cromossoma	1	2	....	1039
Janela Deslizante	60	60	....	60
Remoção de Ruído	20%	20%	....	20%
FitBuy	-10	-10	....	7
FitSell	-10	-9,5	....	3
Rentabilidade	-4,69%	-4,69%	....	17.44%

Table 1: Chromossomes results

Given these results, and with the support of Figure 6, it becomes apparent that there are certain areas that are less stable and have a characteristic of randomness in the production of higher yields.

Although these areas have more volatile chromosomes can generate a higher return, only a small range in FitBuy FitSell and to achieve a return of negative value.

Therefore, the objective is to seek an area of the figure which gives more robustness to the algorithm. After analyzing the results of various financial assets and market indexes, we concluded that restricting the range of FitBuy between +6 to +8.5, and the value of FitSell between -1.5 to +3.5, we would obtain a more robust algorithm to generate a more reliable investment strategy and risk averse.

We also considered some additional chart patterns and technical indicators that supported the decision to purchase / sale of the algorithm.

Noteworthy are the following additional chart patterns:

0.655	0,162	-0,428	-1	-1,38	-1,496	-1,415	-1,224	-1,5	1
1	0,584	0,048	-0,5	-0,908	-1,08	-1,07	-0,946	-0,875	1
1	1	0,524	0	-0,428	-0,674	-0,72	-0,668	-0,25	1
0.655	1	1	0,5	0,048	-0,248	-0,38	-0,39	0,375	1
0.31	0,584	1	1	0,524	0,162	-0,035	-0,112	1	0.524
-0.035	0,168	0,524	1	1	0,584	0,31	0,166	1	0.048
-0.38	-0,248	0,048	0,5	1	1	0,655	0,444	1	-0.428
-0.72	-0,674	-0,428	0	0,524	1	1	0,73	0,375	-0.904
-1.07	-1,08	-0,908	-0,5	0,048	0,584	1	1	-0,25	-1.38
-1.415	-1,496	-1,38	-1	-0,428	0,168	0,655	1	-0,875	-1.856

Table 2: Bull Flag Pattern

The Bull Flag formation is a continuation of short-term mark a small consolidation before the continuation of the previous move. These formations are usually preceded by a sharp advance with strong volume, and mark the midpoint of the movement.

To be considered a continuation pattern there must be evidence of a prior trend well marked, that this case will be the trend. This move represents just the first step in a significant increase being the flag a mere pause it.

-1.224	-1,224	-1,224	-1,224	-1,224	-1,224	-1,224	-1,496	-0,875	1
-0.946	-0,946	-0,946	-0,946	-0,946	-0,946	-0,946	-1,08	-0,25	1
-0.668	-0,668	-0,668	-0,668	-0,668	-0,668	-0,668	-0,674	0,375	1
-0.39	-0,39	-0,39	-0,39	-0,39	-0,39	-0,39	-0,248	1	1
-0.112	-0,112	-0,112	-0,112	-0,112	-0,112	-0,112	0,162	1	0.524
0.166	0,166	0,166	0,166	0,166	0,166	0,166	0,584	1	0.048
0.444	0,444	0,444	0,444	0,444	0,444	0,444	1	0,375	-0.428
0.73	0,73	0,73	0,73	0,73	0,73	0,73	1	-0,25	-0.904
1	1	1	1	1	1	1	0,584	-0,875	-1.38
1	1	1	1	1	1	1	0,168	-1,5	-1.856

Table 2: Breakout Pattern

The Breakout is a pattern of lateral movement that has a well defined resistance, in which the price fluctuates. In case of breach of this, we may be facing a substantial projection of the share prices to positive territory, which also is normally associated with a significant increase in volume.

3.3.1 Methodology

These new chart patterns and technical indicators introduced in the algorithm were used in a clustered and parallel to the existing technique which consists in detecting the upward trend. The results from these additions to the complexity of the algorithm are depicted in the following case studies.

For this we used 100 stocks from the S & P500. Each should have an associated history that extended the period of 1998. That is, we were only interested in selecting actions that they understood a history of trading between 1 January 1998 and April 21, 2010 (most current date on the day of testing). This would allow the computer to perform a workout with a duration of seven years (between January 1, 1998 and December 31, 2004), and a run with a duration of five years and four months (between January 1, 2005 and 21 April 2010).

Selected so the first list of 100 stocks in the S & P500 which met this requirement. This was the only criterion of choice, and therefore neglected here as the risk factors associated with financial assets, the volatility, the average volume of business and its weight in the index.

Thus, by performing a simulation of trading on each of these financial assets without an associated pre-selection, we are recreating a applicational behavior for the algorithm is closer to reality.

In the training phase the algorithm will create a maximum of 15 generations that does not undergo repeated changes in the profitability of the best chromosome. That is, if the end of 15 generations during the training phase, which covers the period between 1 January 1998 and December 31, 2004, there appears an improvement in the value of return achieved, then the algorithm completes the procedure of creating new generations and runs the best chromosome obtained during the testing phase running from 1 January 2005 and April 21, 2010.

The whole cycle will be executed 10 times for each financial asset. This will take some interesting information with regard to these executions.

For the training phase we focus on obtaining the following parameters, which result in better performance from the chromosome that has a return closer to the average of 10 plays:

- Value obtained Fit to Buy;
- Value obtained Fit to Sell;
- Value obtained for the Removal of Noise;
- The value obtained for the size of the Sliding Window;
- Return on Buy & Hold during the training period;
- Return to the Best Chromosome obtained during the training period;

In the test phase we extract the following results obtained by applying the best chromosome found during the training phase of the algorithm:

- Profitability obtained for the Buy & Hold during the test period;
- Profitability obtained from 10 to Media Runs during the test period;

Once it becomes too exhausting to characterize the results for the 100 stocks tested, we chose to select only 10 in each case study, which will be representative of all of them in proportional terms, ie, as a whole will have a proportion equivalent terms of total positive and negative results obtained.

Padrão	**Fit B**	**FitS**	**Remoção de Ruído (%)**	**Janela (dias)**	**Média de 10 Runs no Teste**
Tendencia Ascendente (1)	5,45	1,28	34,10%	54,4	35,57%

Tendencia Ascendente + Bull Flag (2)	5,62	0,53	32,36%	49,4	42,36%
Tendencia Ascendente + Breakout (3)	5,41	0,34	34,89%	50,1	45,89%
Tendencia Ascendente + Breakout Slim (4)	5,41	0,41	35,64%	50,5	41,96%
Tendencia Ascendente + MM30 (5)	5,65	0,94	33,09%	43,9	9,94%
Tendencia Ascendente + MM30v2 (6)	5,48	0,27	34,06%	50,5	38,18%
Tendencia Ascendente + Cruzamento de MM (7)	6,06	1,26	25,28%	49,7	28,44%
Tendencia Ascendente + Cruzamento de MMv2 (8)	6,53	0,80	26,48%	49,6	50,59%

Table 3: Different Trading Techniques.

(1) The algorithm by applying standard graph related to the detection of trend. The system will acquire a position where you're buying before the trend and sell when it ceases to exist.

(2) In this version of the algorithm include the detection of other standard features bullish. In this case we are referring to the Bull Flag. This update to the kernel will make the purchase option is exercised when we stand before the upward trend or the Bull Flag.

(3) Here we seek to test the behavior of the algorithm with the standard chart breakout. Like the previous one, the system acquires a position where it is against this standard and the upward trend.

(4) The difference to the previous methodology is used to detect the accuracy of the standard Breakout, more precisely in your template. So the template that defines this graphic in a matrix formation 10x10 has a broad spectrum less than the previous technique.

(5) The algorithm uses only the training graphics featuring the upward trend (with different slopes) and an additional technical indicator, in this case the moving average of 30 days. Will acquire a position in the market only if the upward trend is in force and the moving average to a higher level than the online price. The position will only be sold if the upward trend is not already present or if the moving average price of crossing the line in the negative sense.

(6) In this variation of the technique before moving average is only used during the execution of the algorithm and only for the purchase. That is, the training period optimization is performed only using the identification of the trend. Once we entered the period of implementation the algorithm takes only one position in the action that is before you consider whether the upward trend and moving average of 30 days to overcome the price range. Already for sale, it will just do without the upward slope.

(7) In this case we used both to detect the upward trend and the crossing of moving averages (30 and 150 days) for our algorithm trading. Whenever we are faced with an upward trend and short-term average is situated above the long term then the algorithm acquires a position to purchase. On the other hand will only sell if the upward trend no longer exists or if the average short-term is less than the long term.

(8) In relation to prior art the only difference lies in the timing of the sale, which will only happen when the short-term moving average is less than the long term.

As can be seen on the results of the previous table, the settings that result from the optimization of genetic algorithm are generally characterized by excellent results in the period associated with their training.

However, the lack of correspondence of these results with the implementation period is remarkable. Given the Buy & Hold, the genetic algorithm can not achieve the results in the period of training in terms of profitability, regardless of the technical indicator used and the chart pattern chosen for the optimization of the parameters of GA.

Note that the same parameters obtained for the different methodologies of algorithm (fitbuy, fitsell, noise and sliding window) are quite similar, and this leads us to believe that this may be an area to explore in order to make our strategy more robust investment and risk-averse. If we consider the technique described in (8) in the previous table we notice that although we have achieved a return lower than the previous techniques in the training period, the implementation period is characterized by an above-average profitability and what was expected. This may mean that this version of the kernel optimization is more consistent in producing results and more appropriate for a less risky investment profile. We also emphasize the value of fitbuy that is superior to previous techniques, and determining in more detail to detect the trend.

3.4 Genetic Optimization Kernel II

Since the main objective of our algorithm is to overcome the methodology of Buy & Hold, it is necessary to explore additional techniques and adding new restrictions and modifications to the kernel of the GA in order to ensure we can produce more satisfactory results.

In the previous optimizations were found to be some problems with decision making purchase / sale of financial assets, and this was mainly due to the additional need for more genes on chromosome genetic algorithm to characterize the lack of upward trend.

Typically the algorithm could detect with a high probability of accuracy the presence of this trend, and consequently acquired a position of always buying when it happened. Later the system would just sell your position when she stopped to stand before the rising trend. However, nothing in ensuring that the training did not have graphical features bullish force. This caused outflows of financial markets too early historical periods that were determined by a positive slope in the price range. Moreover, due to the use of only a sliding window, the output of financial markets was caused sometimes so late.

Thus, we have modified the chromosome that characterized the value of fit for sale (fitsell) tried to add more templates and graphics and a new gene into the chromosome of our optimization:

FitSell - This value will be associated with the Fit templates featuring chart formations with various slopes downward trend. Thus, the algorithm will be more accurate in detecting this chart pattern and ensure that the exit will only happen in the situation we are facing a downward slope patently obvious.

Sliding Window Descent - Previously only resorted to using a single sliding window. This encapsulated the history of prices considered a 10x10 matrix and used it as a comparison with other matrix template 10x10, which in turn characterized the training that was intended to find graphic.

Considering that the chart patterns are formed in different time periods, we decided to give the genetic algorithm with a sliding window that will further be associated with parallel detection of other chart formations.

1	0,655	0,162	-0,428	-1	-1,38	-1,496	-1,415	-1,224	-0.998
1	1	0,584	0,048	-0,5	-0,908	-1,08	-1,07	-0,946	-0.776
0.723	1	1	0,524	0	-0,428	-0,674	-0,72	-0,668	-0.564
0.446	0,655	1	1	0,5	0,048	-0,248	-0,38	-0,39	-0.332
0.161	0,31	0,584	1	1	0,524	0,162	-0,035	-0,112	-0.11
-0.108	-0,035	0,168	0,524	1	1	0,584	0,31	0,166	0.112
-0.385	-0,38	-0,248	0,048	0,5	1	1	0,655	0,444	0.334
-0.682	-0,72	-0,674	-0,428	0	0,524	1	1	0,73	0.556
-0.939	-1,07	-1,08	-0,908	-0,5	0,048	0,584	1	1	0.778
-1.216	-1,415	-1,496	-1,38	-1	-0,428	0,168	0,655	1	1

Table 4: Inversion Pattern.

The reversal of trend is a graphic training that will serve to determine the moment when we cease to be a prevailing trend towards the rise or horizontal consolidation. Once it detects this pattern emerges a break with the previous trend, if it's a trend. This training will also serve to keep the algorithm out of the market if it has already detected a downward trend earlier.

With these changes we want to obtain an algorithm more precise, accurate and robust.

Padrão	Fit B	FitD	Remoção de Ruído (%)	Janela (dias)	Rent. B&H em Treino	Rent. Melhor Cromossoma em Treino	B&H no Teste	Média de 10 Runs no Teste
Tendência Ascendent e + Inversão de Tendência (1)	5,45	6,96	27,70%	39,7/22, 8	154,21 %	554,68%	51,37 %	35,08 %
Tendência Ascendent e + Inversão de Tendência v2 (2)	5,22	6,92	26,41%	52,1/23, 5	154,21 %	588,04%	51,37 %	33,43 %

Table 5: Ascending Tendency + Inversion Pattern.

(1) In this methodology, the algorithm takes a position when it detects the upward trend of at least one of the four arrays that are used with several positive inclinations. On the other hand sells when it detects the downward trend with a value fit to be greater than 4. The size of the window associated with the detection of the downward trend should be equal or less than the sliding window of the upward trend.

(2) This version has only one variation from the previous methodology. In this case, the window size associated with the downward trend should always be less than the sliding window of the upward trend, except in the case that the window of the upward trend has a length of 20 days, since in this case the window of the downward trend also may have a size associated with 20 days.

As illustrated in the table above the return achieved for the period of training the best chromosome is higher than the yields obtained in the previous table that describes the different methodologies used in the kernel optimization II.

However, average yields obtained in the test period (execution) does not reach the expected values, and given the values recorded for the size of two sliding windows, we conclude that the algorithm will react too quickly to small fluctuations in the market.

Thus, we conducted a study on the influence of the size of sliding windows in the results of the algorithm.

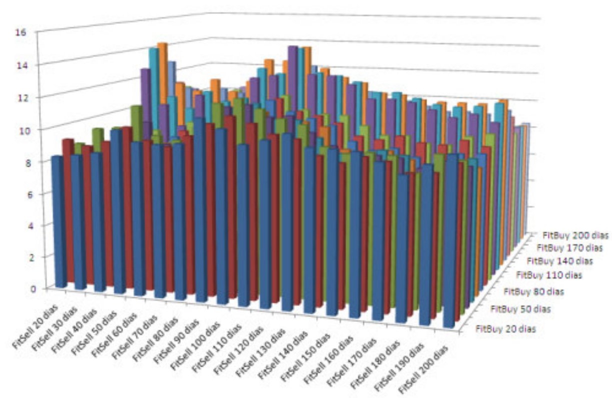


Figure 7: Average Annual Return ona anual obtida on S&P500 for different window size configurations.

As you can see the figure above, higher profitability and more obvious are those that come with a sliding window size and higher than expected. Typically we can find good results in the area where the fitbuy fitsell and are determined by an upper window of 100 days.

To study the real impact of the variation in the size of sliding window on the profitability of the algorithm, we set the following parameters of genetic chromosome:

Parâmetros do GA	Valores
FitBuy	4.5
FitSell	7
Ruído (%)	44.8

Tabela 6: Parâmetros resultantes para o GA encontrado no S&P500.

The values were obtained based on the best chromosome achieved during the training period the S & P500, which runs between 1998 and 2004.
Since the 100 stocks chosen for analysis are an integral part of this index, we felt it would make sense to use the values obtained for these parameters during the execution of the algorithm, for each 100 shares, between the years 2005 and 2010.

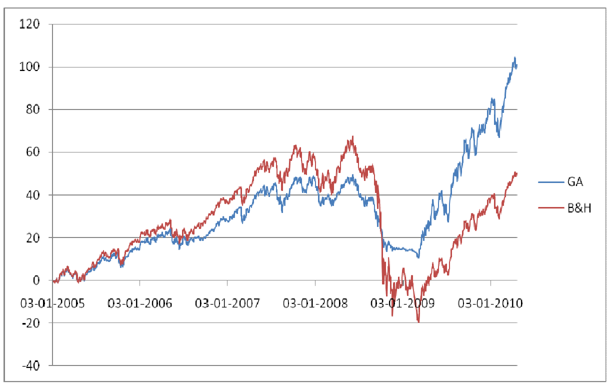


Figure 8 – Return of best GA and B&H for all the 100 stocks.

In this figure it becomes apparent the time period in which the genetic algorithm detects the downward trend, and exit the market in late 2008. Thereafter the algorithm returns to reenter the market when a new upward trend.

4. CONCLUSIONS

This document presents an investment strategy outlined in the financial markets. It focuses mainly on detection of chart patterns and prediction of market behavior that underlies these, but also takes into account other technical indicators that can support the decision to buy / sell / maintain the financial instrument.

To implement the algorithm to consider various strategies presented in papers and journals in the area of financial applications. All information and working inside this field is very preliminary, so this work has focused on the combination, adaptation and innovation in one application, the various methods of detection of chart patterns and trends, since each of these different methods will have their own advantages and disadvantages. Working together we can take a greater accuracy in identifying the desired chart formations and superior profitability in the trading period under consideration.

The results are very motivating and encouraging. The algorithm can be easily extended and customized with new features, since there is much to explore within the field of detection of graphics standards.

5. FUTURE WORK

We suggest some future changes to the work under this Master Thesis:

• Include new strategies for detecting patterns through the discovery of underlying technical chart formations using perceptually important points;

• Provide the mechanism of short selling in order to prevent the algorithm remains very long periods out of financial markets;

• Include additional chart patterns, which were not only studied and incorporated into the final optimization algorithm. It will be necessary to explore the contribution of positive / negative impact of each standard added to the system.

• Add new filtering mechanisms for the set of actions to be considered in order to eliminate undesirable assets that can cause erratic behavior in the execution of the algorithm.

• Test the behavior of the algorithm with training periods and different execution, which may be shorter or prolonged.

• Include technical indicators also support the analysis of chart patterns. It will be important to endow the algorithm with other indicators that support the decision making process in buying and selling shares.

6. REFERENCES

[1]- David E. Goldberg, "Genetic and Evolutionary Algorithms Come of Age", ACM, New York, NY, USA, 1994

[2]- António Gorgulho, Rui Neves and Nuno Horta, "Using GAs to Balance Technical Indicators on Stock Picking for Financial Portfolio Composition", Genetic and Evolutionary Computation Conference - Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, Montreal, Québec, Canada, pp. 2041-2046, July 2009

[3] - Charles D. Kirkpatrick and Julie R. Dahlquist, "Technical Analysis: The Complete Resource for Financial Market Technicians", August 27, 2006

[4] - Fernando Braga de Matos, "Ganhar em Bolsa", Publicações Dom Quixote, 2ª edição, Junho de 2007, pp. 195-203

[5] - Fernando Braga de Matos, "Ganhar em Bolsa", Publicações Dom Quixote, 2ª edição, Junho de 2007, pp. 288-292

[6] - Millenium BCP Investimento, "Estudos e Histórias – Principais Formações Gráficas de Análise Técnica", August 2006.

[7] - Millenium BCP Investimento, "Estudos e Histórias – Principais Indicadores de Análise Técnica II", August 2006.

[8] - Leigh W., Frohlich C.J., Hornik S., Purvis R.L. and Roberts T.L., "Trading With a Stock Chart Heuristic", Systems, Man and

Cybernetics, Part A, IEEE Transactions on Volume 38, Issue 1, Jan. 2008 pp. 93 – 104

[9] - William Leigh, Naval Modani, Russell Purvis and Tom Roberts, "Stock market trading rule discovery using technical charting heuristics", Expert Systems with Applications, Volume 23, Issue 2, August 2002, pp. 155-159

[10] - William Leigh, Noemi Paz and Russell Purvis, "Market timing: a test of a charting heuristic", Economics Letters, Volume 77, Issue 1, September 2002, pp. 55-63

[11] - Leigh W., Paz M. and Purvis R., "An analysis of a hybrid neural network and pattern recognition technique for predicting short-term increases in the NYSE composite index", Omega, vol. 30, pp. 69-76, 2002

[12] - William Leigh, Russell Purvis and James M. Ragusa, "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support", Decision Support Systems, Volume 32, Issue 4, March 2002, pp. 361-377

[13] - R. Edwards and J. Magee, "Technical Analysis of Stock Trends", New York: Amacom, 1997

[14] - Jar-Long Wang, Shu-Hui Chan, "Trading rule discovery in the US stock market: An empirical study", Expert Systems with Applications: An International Journal, vol. 36, Issue 3, pp. 5450-5455, April 2009

[15] - Jar-Long Wang and Shu-Hui Chan, "Stock market trading rule discovery using pattern recognition and technical analysis", Expert Systems with Applications, Volume 33, Issue 2, August 2007, pp. 304-315

[16] - T.C. Fu, F.L. Chung, K.Y. Kwok and C.M. Ng, "Stock Time Series Visualization Based on Data Point Importance", Engineering Applications of Artificial Intelligence, Vol. 21, Issue 8, pp. 1217-1232, December 2008

[17] - T.C. Fu, F.L. Chung, R. Luk and C.M. Ng, "Representing Financial Time Series Based on Data Point Importance", Engineering Applications of Artificial Intelligence, Vol. 21, Issue 2, pp. 277-300, March 2008

[18] - Keogh E., Chakrabarti K., Pazzani M. and Mehrotra S., "Dimensionality reduction for fast similarity search in large time series databases", Journal of Knowledge and Information Systems 3 (3), pp. 263–286, 2000

[19] - T.C. Fu, F.L. Chung, R. Luk and C.M. Ng, "Stock Time Series Pattern Matching: Template-based vs. Rule-based Approaches", Engineering Applications of Artificial Intelligence, Volume 20, Issue 3, April 2007, Pages 347-364

[20] - T.C. Fu, F.L. Chung, R. Luk and C.M. Ng, "Preventing Meaningless Stock Time Series Pattern Discovery by Changing Perceptually Important Point Detection", The 2nd International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2005), pp. 1171-1174, Changsha, China, 27-29 August 2005

[21] - F.L. Chung, T.C. Fu, Vincent Ng, and Robert Luk, "An Evolutionary Approach to Pattern-based Time Series Segmentation", IEEE Transactions on Evolutionary Computation, Vol. 8, pp. 471-489, October 2004.

[22] - E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series", in Proc. IEEE Int. Conf. Data Mining, 2001, pp. 289–296.

[23] - James N.K. Liu, Raymond W.M. Kwong, "Automatic extraction and identification of chart patterns towards financial forecast", Applied Soft Computing, Volume 7, Issue 4, August 2007, Pages 1197-1208.

[24] - Chang-shing Perng, Haixun Wang, Sylvia R. Zhang and D. Stott Parker, "Landmarks: A New Model for Similarity-Based Pattern Querying in Time Series Databases", ICDE, Proceedings of the 16th International Conference on Data Engineering, page 33, 2000.

[25] – R. Agrawal, C. Faloutsos and A. N. Swami, "Efficient similarity search in sequence databases", in FODO, 1993.

[26] - P.L. Bernstein, "A new look at the efficient market hypothesis", The Journal of Portfolio Management, vol. 25, no. 2, pp. 1-2, Winter 1999.