Combining Rules between PIPs and SAX to Identify Patterns in Financial Markets

João Maria Rodrigues Leitão
Instituto Superior Técnico,
Universidade Lisboa.
joaomrleitao@gmail.com

Abstract—This paper describes a new pattern discovery approach based on the combination among rules between Perceptually Important Points (PIPs) and the Symbolic Aggregate approximation (SAX) representation optimized by Genetic Algorithm (GA). The rules and SAX are used to represent the financial time series in order to identify efficiently patterns. The GA is used to generate investment rules and find optimal solutions. We decided to call this new approach Symbolic Important Rules (SIR). The proposed approach was tested with real data from S&P500 index and all the results obtained outperform the Buy&Hold strategy. Three different case studies are presented. With this approach it was possible to obtain in the period 2011-2014 a total return of 76.7%, which outperformed the Buy&Hold strategy (61.9%).

Keywords—Pattern discovery, Rules, Perceptually Important Points (PIPs), SAX representation, Genetic Algorithm, Investment rules.

I. INTRODUCTION

Financial markets have received an increasing interest by financial people and by computational intelligence researchers. One of the main challenges is to predict the future trend of prices, in order to obtain the highest profit with the lowest risk. To achieve that it is necessary to define investment strategies that are able to process large amounts of data and consequently generate appropriate buy/sell signals. The data can be obtained from fundamental analysis [1], technical analysis [2,3] and time series. To solve this complex problem the computational intelligence area is very useful.

One way used by traders to predict the behaviour of the markets, is studying and analysing chart patterns in the historical prices of the financial assets. The visually identification of chart patterns is very complicated, because the patterns in time series are not clear and perfect as the ones in the books. In order to identify patterns, automatic systems from computational intelligence must be used.

In this paper a new approach to pattern discovery is presented based on Perceptually Important Points (PIPs) [4,5], the Symbolic Aggregate approximation (SAX) representation [6], optimized by Genetic Algorithm (GA) [7]. The identification of PIPs allows a huge dimensional reduction of the time series and at the same time, maintains the main characteristics of its data. The definition of rules between near or adjacent PIPs allows the explicit definition of relationships between time series points. The mapping between rules and characters allowed the distinction of the different types of trends between the PIPs of time series and also allowed the representation of time series by a sequence of characters, which facilitated the identification of patterns. The GA is used to optimize the type of pattern to be identified and the investment rules used in the trading simulation. The main contributions made in this paper are: 1 - the creation of the new methodology to identify patterns that combines rules between PIPs with the mapping between those rules and different characters in the SAX representation; 2 - the combination of multiple exit/sell methods namely time, price and pattern; 3 - the use of a GA adaptive approach able to automatically identify multiple patterns and generate trading rules. The approach was tested with real data from S&P500 index and the results outperformed the Buy&Hold investment strategy, which is used as reference in the efficient markets theory [8].

This paper is organized as follows; in Section 2 the related work is discussed. Section 3 describes the method to represent time series and to create investment rules with GA. Section 4 describes the case studies and results. Section 5 draws the conclusions.

II. RELATED WORK

Analyzing the historical prices of a financial asset it is possible to observe some similar geometric shapes over the time. Those geometric figures represent the behavior of the traders in the market. Knowing that history repeats, the identification of geometric figures allow the analysts to predict with some confidence the behavior of the traders and consequently the future trend of prices. The chart patterns, according to [9], can be divided in 2 types: continuation patterns and reversal patterns. The continuation patterns generally are faster to form than the reversal patterns. In order to be more certain of the future direction of prices, the volume indicator can be used to confirm the formation of chart patterns.

In recent years, many studies were developed with the aim of predict financial markets, based on the identification of several well-known chart patterns [10] in time series. In order to realise that several methodologies combined with optimization techniques like Genetic Algorithms [7], [11] were used and obtained good results, Table 1. To create an efficient method of search patterns, the main challenge is to reduce the data dimensionality of the original time series.

Initially several studies used a methodology based on templates to detect the Bull Flag pattern, where time series and the pattern were represented by matrices in [12,13]. The template of the pattern is represented by a 10x10 matrix, where each cell could have a value between -2.5 and +1.0. The sum of all the values of each column it is always equal to 0. Then each time series is represented by a 10x10 matrix as well, where each cell has a value between 0 and 1 dependent in the number of days that are mapped in it. After that, it is used a fit function (1) that multiplies the two matrices,
pattern’s matrix “T” and time series matrix “I”, in order to obtain a value that indicates the level of similarity between them. Thus, highest values will occur when the matrix “I” is in highest conformance with matrix “T”.

\[
\text{Fit}_{i,j} = \sum_{i=1}^{10} \sum_{j=1}^{10} (T(i,j) \cdot I(i,j))
\]  

Some studies applied optimization techniques like genetic algorithms and neural networks with this methodology to create investment rules [14]. After that, many studies like [15,16,17], based on this methodology tested the detection of other patterns and also a new type of template to represent them.

Other methodology used was the representation of time series and patterns by its most relevant points, denominated Perceptually Important Points (PIPs) [4,5]. These points are the most relevant because are the ones who characterize the time series and the patterns. After the identification of PIPs in time series the detection of the patterns is made by two different techniques, one based on templates and other based on rules. In the rule-based technique the structure of the patterns is defined visually which allows comparison point-to-point between the time series and the patterns. In the rule-based technique each pattern is defined by a set of rules to describe its shape, where these rules are created according to the relations between PIPs. Applying the rule-based pattern matching allows explicit definition of the relationships between the most relevant points. Those time series, represented by its PIPs that can validate all the rules of a certain pattern are identified as one.

To measure the maximum distance between one point and its two adjacents PIPs, in [4] are presented 3 methods:

- Euclidean distance (ED)
- Perpendicular distance (PD)
- Vertical distance (VD)

### Euclidean distance (ED)

Calculates the sum of the ED of the test point \( p_j \) to its adjacent PIPs \( p_i \) and \( p_2 \).

\[
\text{ED}(p_j,p_i,p_2) = \sqrt{(x_2-x_j)^2 + (y_2-y_j)^2}
\]  

### Perpendicular distance (PD)

Calculates the PD between the test point \( p_j \) and the line connecting the two adjacent PIPs \( p_i \) and \( p_2 \).

\[
\text{Slope}(p_i,p_2) = \frac{y_2 - y_i}{x_2 - x_i}
\]

\[
x_c = x_i + \frac{s \cdot (y_2 - y_i)}{y_2 - y_i}
\]

\[
y_c = y_i + s \cdot (x_2 - x_i)
\]

\[
\text{PD}(p_j,p_i,p_2) = \sqrt{(x_c - x_j)^2 + (y_c - y_j)^2}
\]

### Vertical distance (VD)

Calculates the VD between the test point \( p_j \) and the line connecting the two adjacent PIPs \( p_i \) and \( p_2 \).

\[
\text{VD}(p_j,p_i,p_2) = |y_i - y_j| - \left| \left( y_i + \frac{x_i \cdot s}{y_2 - y_i} \right) - y_j \right|
\]

The three distance methods were tested, using 2500 points of data from Hang Seng Index (HSI), the vertical distance method proved to be the best in capturing the shapes of patterns. More recently, some studies used other methodology to reduce the data dimensionality, which was the Symbolic Aggregate approXimation (SAX) representation [6], which is based on PAA [18]. This method [19], begins by dividing the time series in windows then each window in several segments, which are then represented by the arithmetic mean of its points, and finally each segment is converted to a symbol according to the result of its arithmetic mean. This process allows the representation of time series by a sequence of symbols therefore to search patterns those sequences must be compared with each other in order to identify the similarity between them.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Year</th>
<th>Method Used Data</th>
<th>Period</th>
<th>Financial Market</th>
<th>Algorithm Performance</th>
<th>Buy-and-hold Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>[13]</td>
<td>2008</td>
<td>Bull Flag of Matrix Template</td>
<td>Spx price</td>
<td>NYSE Composite Index</td>
<td>4.56% (Transaction average over the period)</td>
<td>4.56% (Transaction average over the period)</td>
</tr>
<tr>
<td>[17]</td>
<td>2007</td>
<td>Bull Flag of Matrix Template</td>
<td>Spx price</td>
<td>NASDAQ &amp; TWI</td>
<td>4.38% (Transaction average over the period)</td>
<td>3.27% (Transaction average over the period)</td>
</tr>
<tr>
<td>[14]</td>
<td>2002</td>
<td>Hybrid Neural Networks with Pattern detection</td>
<td>Spx price</td>
<td>NASDAQ &amp; TWI</td>
<td>69% (Average return)</td>
<td>69% (Average return)</td>
</tr>
<tr>
<td>[16]</td>
<td>2010</td>
<td>Bull Flag of Matrix Template</td>
<td>Spx price</td>
<td>Dow Jones Industrial Average Index</td>
<td>13% (Average return)</td>
<td>N/A</td>
</tr>
<tr>
<td>[19]</td>
<td>2007</td>
<td>Bull Flag of Matrix Template</td>
<td>Spx price</td>
<td>NASDAQ &amp; TWI</td>
<td>28.52% (Total return)</td>
<td>-4.65% (Total return)</td>
</tr>
<tr>
<td>[1]</td>
<td>2007</td>
<td>Template-Based</td>
<td>Spx price</td>
<td>News on pattern identification</td>
<td>56% (News on pattern identification)</td>
<td>N/A</td>
</tr>
<tr>
<td>[2]</td>
<td>2007</td>
<td>Rule-Based</td>
<td>Spx price</td>
<td>News on pattern identification</td>
<td>39% (News on pattern identification)</td>
<td>N/A</td>
</tr>
<tr>
<td>[4]</td>
<td>2007</td>
<td>PAA</td>
<td>Spx price</td>
<td>News on pattern identification</td>
<td>82% (News on pattern identification)</td>
<td>N/A</td>
</tr>
<tr>
<td>[19]</td>
<td>2010</td>
<td>SAX + GA</td>
<td>Spx price</td>
<td>NASDAQ &amp; TWI</td>
<td>10.25% (Average annual return)</td>
<td>7.75% (Average annual return)</td>
</tr>
</tbody>
</table>

Table 1 - Algorithms results.
III. SYMBOLIC IMPORTANT RULES METHODOLOGY

The objective of this research is to develop a pattern discovery algorithm that can combine ideas from how humans identify patterns and automatic classification of patterns. The method uses points that normally a human would consider important, and then creates rules to describe the relationship between them. Then using GA and SAX makes an automatic search for the relevant patterns.

A. Time series representation

The proposed method is divided in 4 steps, represented in Fig. 1. These steps are described here shortly, and next each one in detail. Firstly, the historical prices of a financial asset are divided into smaller time series all with the same size in order to identify patterns with the same time length, Fig. 1 a). After this, it is possible to identify patterns in the time series but the dimension of data is too high, making this process very expensive in time and computational resources. Secondly in order to reduce the dimension of the data, each time series is represented by its most relevant points, denominated Perceptually Important Points (PIPs), Fig 1. b). Thirdly, rules are created that identify the relationship between two PIPs. The two PIPs do not need to be consecutive it is possible to have rules between two PIPs that are apart to each other more than one unit. Finally the fourth step where each different rule created is transformed in a different symbol in order to represent each time series by a sequence of symbols. We call our approach Symbolic Important Rules (SIR).

PIPs are points that a human looking at a time series would consider important to identify the pattern. The method used to identify PIPs is based on [5] that starts by defining the first and the last point of a time series as the first two PIPs. Then the third PIP is the point of the time series with maximum vertical distance to the line between the first two PIPs. The next PIP will be the point with maximum vertical distance to its two adjacent PIPs, i.e., between either the first and the second PIPs or the second and the last PIPs. This process continues until the limit of PIPs to identify in the time series is reached, as represented in Fig. 1 b).

In the third step, rules are created according to the PIPs identified and the relations between them see Fig. 1 c). These rules are defined based on the percentage difference between two adjacent or non-adjacent PIPs, allowing the definition of different rules between one PIP and others. With these rules the normalization of data between time series is done because is used the percentage difference between two points and not the absolute value difference of those points, an example is a rise from 55 to 105 and a rise from 100 to 205, where both have the same percentage difference (100%) but not the same absolute value difference (5 and 100).

In order to create 5 different types of rules, two variables x and y are defined, where each represents a percentage and the percentage y is higher than the percentage x. The 5 different types of rules presented in Fig. 2, can be described as:

1. Percentage difference between two PPIs higher than \( x\) - strong increase of price
2. Percentage difference between two PPIs higher than \( x\) and lower than \( y\) - slight increase of price
3. Percentage difference between two PPIs higher than \(-x\) and lower than \(-y\) - sideways movement of price
4. Percentage difference between two PPIs lower than \(-x\) and higher than \(-y\) - slight decrease of price
5. Percentage difference between two PPIs lower than \(-y\) - strong decrease of price

The rules definition algorithm receives as input the PIPs of a time series and a maximum limit of relations between PIPs, which defines the maximum number of rules that can be defined between one PIP and the others. In Fig. 3 are represented two examples with five PIPs and different limits of relations between PIPs. In the first example (left) the limit of relations is one therefore the rules defined are only between adjacent PIPs (green arrows), i.e., between the first and the second PIPs, the second and the third PIPs, and so on. In the second example (right) the limit of relations between PIPs is 2, which means that each PIP can be related with its two following PIPs when possible, consequently the rules defined are the same of the first example (green arrows) plus the rules between one PIP and the next PIP to its adjacent PIP (yellow arrows).

The proposed algorithm, Fig 4, begins by calculating the percentage difference between the first PIP and second PIP and according to the result assigns one of the five rules in Fig. 2. If the limit of relations between PIPs had not been reached the next rule assigned will be defined by the result of the percentage difference between the first and the third PIPs and so on until the limit is reached. After that, the process repeats with the second PIP, until the limit is reached and after that with the others PIPs until the rule related with the percentage difference between the penultimate and the last PIPs is assigned, which terminates the algorithm.
Procedure RulesDefinition(P, l, x, y)

\[ P[1...m] = \text{set of PIPs in the time series} \]
\[ l = \text{limit of relations} \]
\[ x = \text{lower percentage}, \quad y = \text{higher percentage} \]

For \( i = 1 \) until \( \text{size}(P) \)

For \( j = 1 \) until \( l \)

If \( \text{Diff}%(P[i], P[j + 1]) > y \)

Rule\[i, j\] = 1

If \( \text{Diff}%(P[i], P[j + 1]) < x \) AND \( \text{Diff}%(P[i], P[j + 1]) < y \)

Rule\[i, j\] = 2

If \( \text{Diff}%(P[i], P[j + 1]) > x \) AND \( \text{Diff}%(P[i], P[j + 1]) < y \)

Rule\[i, j\] = 3

If \( \text{Diff}%(P[i], P[j + 1]) < x \) AND \( \text{Diff}%(P[i], P[j + 1]) > y \)

Rule\[i, j\] = 4

If \( \text{Diff}%(P[i], P[j + 1]) > y \)

Rule\[i, j\] = 5

End

Fig. 4 – Pseudo code of the rules definition process.

The advantage of defining these rules in time series is to obtain an explicitly definition of the relationships between the points, in terms of price movements. Many well-known patterns are defined by a specific set of rules between its points, as an example the Head-and-Shoulders pattern where the two shoulders in the pattern must have a null or almost null percentage difference between them (rule 3 Fig. 2) and both must be lower than the head of the pattern.

In the fourth step, all the rules defined are converted into characters, allowing the representation of time series by a sequence of characters (string). To do that, each of the five different rules is mapped to one different character in order to distinguish precisely the different trends of price represented by the different rules. The alphabet chosen and the mapping between the characters and the rules are represented in Fig. 5.

To find new patterns the sequences of characters must be compared with each other or with a known sequence of characters to find some wanted pattern. In order to identify the matching between sequences of characters, it is used (8) to calculate the distance between two sequences and identify the level of similarity between them, through the ASCII code of each character, Fig. 6. Lower values mean more similarity between sequences and higher values mean the opposite.

\[
\text{DIST}(T, P) = \sqrt{\frac{1}{w} \sum_{i=1}^{w} (T_i - P_i)^2} \tag{8}
\]

where \( w \) is the size of the sequences, \( T_i \) is the character \( i \) of the time series \( T \) and \( P_i \) is the character \( i \) of the pattern \( P \).

The characters used for each rule where carefully chosen to improve the performance of the algorithm. Each character of the alphabet has a different ASCII code, where “C” = 67, “H” = 72, “M” = 77, “R” = 82 and “W” = 87, allowing the distinction of the different trends of price defined by the different rules. Rules more identical in terms of trend have lower difference in ASCII code of the characters and rules less identical have higher difference in ASCII code of the characters. The most contrasting rules, i.e., the strong increase and the strong decrease are mapped with the first and last character, “C” and “W” respectively, of the alphabet in order to hold the biggest difference (20) in ASCII code between all the characters. The sideways movements is mapped with the character “M” which is the character with the same distance to the characters that represent the opposite rules. From 1 to 5 the consecutive rules have smaller difference (5) between each other due to its higher similarity.

**B. Investment rules**

The goal of this work was not only to identify patterns in time series but also to create investment rules based on the patterns identified. The algorithm analyses stocks historical prices with the help of a sliding window of variable length and converts each time series in a sequences of characters. Every time a pattern is identified in the time series a buying order is generated. After that in order to evaluate the return of the operation an exit point should be defined and can be by 3 different methods or combinations between them:

1. **By time:** where is defined a variable holding period of days until the sell order is generated and the operation is closed.

2. **By price:** where is defined a variable take profit and a variable stop loss, which will limit both the profit and the loss of each operation, respectively. When one of the limits is reached the operation is closed with loss or profit. The gain at the take profit level is often greater than the loss at the stop loss level so that the total profit depends on the success rate of operations. These variables are defined based on a positive and negative percentage over the stock buying price.

3. **By pattern:** where is defined an uptrend pattern with the goal of identify it in each time series subsequent to the buying order. The operation is closed only when the uptrend pattern is not identified in one of those time series, which means that the prices stopped increasing. In order to identify the uptrend pattern, the time series following the buying order is represented by a sequence of characters using the method in section 3.1. Then the uptrend pattern is represented by the sequence of characters “C”, which means consecutive strong increases of price. After that the time series’ sequence of characters is compared with the uptrend pattern’s sequence of characters using the distance method (1) that is used to generate buying orders, where the only difference is that in this case the distance, in terms of ASCII code, between characters “H” and “C” is 0 instead of 5 due to the fact that “H” represents also a increase of prices, which is what is supposed to happen to prices, so that the operation is not closed. Then, if the result of the comparison is higher than a threshold a sell...
order is generated, if not it means the pattern is identified in that time series so the process is repeated with the next time series until the pattern is not identified in some time series.

In Fig. 7 it is possible to observe an example, where the 3 methods described before are used simultaneously. After opening the position the exit by time defines a holding period of 35 days to close it (blue dashed line), the exit by price defines take profit (green line) and stop loss (red line) based on the buying price and the exit by pattern begins by represent the time series from day 47 to day 94 in a sequence of characters, according to the method of section 3.1 and then compare it with the sequence ["C", "C", "C"] which represents the uptrend pattern, using the method (8). In this example, the position is closed by price because the price reached the take profit level before the other methods generate sell orders.

C. Genetic Algorithms (GA)

To optimize the parameters related to the investment rules is used the Genetic Algorithm (GA). The chromosome used to create the population is represented in Fig. 8.

The chromosome is divided in 2 major parts. In the first one (first 8 genes) are the parameters related to the buy and sell decisions, where:

- Method to sell: defines what is the method or the combination of methods that are used to close the operations.
- Number of stocks: defines the maximum number of different stocks that can be in the portfolio at the same time.
- Lower and Higher Percentage: define the percentages used to create the 5 different types of rules, x and y in Fig. 2 respectively.
- Window size: represents the size of the sliding window that is used to divide the historical prices in smaller time series.
- Distance SAX: defines the maximum distance that identifies a pattern in the time series, therefore it is used to generate buying orders and to exit by pattern.
- PIPs: the number of relevant points to identify in each time series, in order to reduce the data dimensionality and at the same time maintain the main characteristics of the time series. Limit of Relations: maximum limit of relations between the PIPs that are identified in each time series. Defines the number of rules that will be created for each PIP with the others PIPs because each relation corresponds to a rule.

In the second part (L1,…,Lx) are the characters that represent the pattern sequence. Each gene represents a character of the alphabet described in Fig. 5 ("H" or "M" or "R" or "W"). This sequence of characters represents the pattern that is used to identify in each time series of the historical price of stocks.

The chromosomes could have different sizes due to the length of the sequence of characters, which is dependent on the sequence of PIPs and relations values in each chromosome. Also the number of PIPs is dependent on the window size value, the limit of relations between PIPs is dependent on the number of PIPs and the distance to buy is dependent on the number of PIPs and the limit of relations. These 4 genes (yellow block in Fig. 8) that depend on each other must be always together in the chromosome, otherwise the crossover method is not performed for those chromosomes where these 4 genes are always swapped together or none of them is swapped between chromosomes. When the two chromosomes have different sizes, only the first four genes can be swapped because the others are responsible for the size of each chromosome. When the chromosomes have equal size all the genes can be swapped but the 4 dependent genes have to be swapped always together as one.

Each chromosome corresponds to a different investment strategy, which is tested by the program where the fitness function that the GA optimizes is the Return On Investment (ROI) of each investment strategy.

After testing all chromosomes in all the historical prices from 2010 to 2014, the program generates a buying order. This process is repeated to all the stocks, in order to get a set of buying orders for the day after each time series identified as a pattern. After that, the set of buying orders is ordered according to the distance to the pattern, where the lowest distance will be in the top and the highest distance will be in the bottom of the set, aiming to buy the stocks whose time series are more identical to the pattern. In this order, the stocks are bought until the limit of stocks that can be opened at the same time defined by the gene ‘Number of Stocks’ is reached or there are no more to buy. For each stock/company, the amount of money invested is the same so the number of shares of each company that is bought is different because it depends on its share price at that moment. To determine the number of shares to buy, the balance is divided by the number of stocks/companies that can be bought (portfolio size – nº stocks already bought) which represents the amount of money that can be invested in each stock. After that, this amount is divided by the share price of each company, which results in the number of shares that will be bought for each company. Lastly, the operations are closed according to the exit method or the combination of exit methods defined in the gene ‘Method to Sell’. After testing all chromosomes in all the historical prices, a new population is created with the best chromosomes of the last population and with new chromosomes resulting of the crossover and mutation operations. The process described before is repeated with the new population until the number of evolutions is reached or there are no improvements of the fitness function in consecutive generations.

IV. EXPERIMENTS AND RESULTS

In this section three case studies are presented, where all were tested with 422 stocks of the S&P500 index. The program was tested with real market conditions with the close prices of the stocks. The daily close prices were obtained in Yahoo finance platform from 2010 to 2014,
which is a period defined as a bull market where prices rose sharply. In each case study the training and testing periods were different in order to test the program in different situations.

A. Case study nº1

In this case study, the goal was to simulate a real life scenario, where the solutions were trained in one year and then the best ones were tested in the next year, in order to assure that the solutions that are tested don’t know the market behaviour in that year and also to assure that the algorithm is tested in distinct periods of time. All the operations that were closed in the end of each year without being by the exit method of its investment strategy start the next year opened in order to be closed by its exit method. Of the 4 testing years, the only where this did not happen was 2011 because it is the first and there is no year before.

The GA parameters used were an initial population of 128 individuals and 50 generations as stop criteria. The tests were repeated for 5 runs in each year and the chromosome used was the one in Fig. 8.

The results are represented in Table 2, which are the average of the 5 runs in each year of testing. These results were compared with the Buy&Hold strategy return for the same period, Fig. 9.

<table>
<thead>
<tr>
<th>Year</th>
<th>Nº operations</th>
<th>SR</th>
<th>D</th>
<th>SR/GA Return (%)</th>
<th>B&amp;H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Nº operations</td>
<td>SR</td>
<td>D</td>
<td>SR/GA Return (%)</td>
<td>B&amp;H</td>
</tr>
<tr>
<td>2011</td>
<td>34</td>
<td>61.76</td>
<td>39</td>
<td>0.73</td>
<td>5.91</td>
</tr>
<tr>
<td>2012</td>
<td>34</td>
<td>61.27</td>
<td>43</td>
<td>3.18</td>
<td>15.19</td>
</tr>
<tr>
<td>2013</td>
<td>18</td>
<td>73.40</td>
<td>78</td>
<td>22.15</td>
<td>29.05</td>
</tr>
<tr>
<td>2014</td>
<td>12</td>
<td>70.31</td>
<td>103</td>
<td>3.83</td>
<td>17.88</td>
</tr>
</tbody>
</table>

As can be seen in Table 2 the average return in each year was higher than the return of the Buy&Hold strategy. The years with the highest difference, in average return to the Buy&Hold strategy were 2011 and 2012, 5.91% and 5.49% respectively, and the year with the lowest difference was 2012, 1,78%. The success rate of operations was always higher than 60%, where the highest percentage was 73.40% in 2013, and the lowest was 61.27% in 2012. The average total return for the 4 years was 72.16%, which outperformed the return of the B&H strategy (61.9%).

The investment strategy, which obtained the best result in 2012, 31.47% of return, is represented in Fig. 11. The exit method was defined by a combination of price and pattern. In the case of the exit by price the operations were closed with profit when the price reached 9.62% over the buying price and with loss when the price reached -10.47% over the buying price. In the case of the exit by pattern the time series after the buy order that were compared with the uptrend pattern, had a length of 51 days. The operations were closed 37 times by price and 17 times by pattern with this investment strategy.

The pattern used in this strategy is represented in Fig. 12. This pattern represents an upward trend of prices where each successive peak (points 3, 5 and 7) and trough (points 4, 6 and 8) is higher.
The pattern used in this strategy is similar to the well-known Double Bottom pattern where the point 2 defines the first bottom and the point 5 the second bottom. This pattern normally signals a reversal of the downtrend into an uptrend.

In the last year the best investment strategy obtained a return of 27.76%, Fig. 14. The time series of 40 days were represented by 10 PIPs and the rules were created only between adjacent PIPs. The exit method used was a combination between time and price, where the holding period until the selling order was 68 days and the positive limit, take profit, was much higher than the negative limit, stop loss, which allows the operations to be closed or with high profits or with small losses. The operations were closed 11 times by time and 6 times by price.

In Fig. 15, is represented the pattern used in this strategy which starts with an uptrend (points 1, 2 and 3) that is followed by a downward trend (from point 5 to point 10), where each successive peak and trough is lower. Although this pattern is characterized by a downtrend instead of an uptrend like the patterns of the previous years, it allowed finding bottoms that were followed by an increase of prices, which originate the good result of this strategy.

B. Case study nº2

In this case study, the goal was to simulate the previous case study with all the same conditions but the exit method used in all the investment strategies would be only the exit by pattern in order to prove the capability and strength of this type of exit method in a bull market.

In Table 3 are represented the average results of the 5 investment strategies in each year. The average results and the best results in each year were compared with the B&H strategy in terms of total return, Fig. 16.

<table>
<thead>
<tr>
<th>Year</th>
<th>Nº</th>
<th>SR</th>
<th>D</th>
<th>SIR/GA Return (%)</th>
<th>B&amp;H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Worst</td>
<td>Avg</td>
</tr>
<tr>
<td>2011</td>
<td>28</td>
<td>62.14</td>
<td>40</td>
<td>8.76</td>
<td>11.78</td>
</tr>
<tr>
<td>2012</td>
<td>23</td>
<td>62.93</td>
<td>65</td>
<td>9.27</td>
<td>19.06</td>
</tr>
<tr>
<td>2013</td>
<td>8</td>
<td>85.29</td>
<td>137</td>
<td>22.76</td>
<td>27.52</td>
</tr>
<tr>
<td>2014</td>
<td>11</td>
<td>66.07</td>
<td>105</td>
<td>8.98</td>
<td>17.92</td>
</tr>
</tbody>
</table>

As can be seen in Table 3, the average return in each year outperformed the return of the B&H strategy, where the first year (2011) was by far and away the highest in difference between the average and the B&H return (11.78%) and the 2013 year was the lowest in difference (1.13%). Comparing the average return of this case study with the average return of the previous case in each year, Table 2, it’s possible to observe that the results were better in all years, except in 2013 which proves the capacity of the exit by pattern to obtain good results in a bull market. The average total return of this strategy for the 4 years was 76.7%, which outperformed the return of the B&H strategy (61.9%).
In 2012 the best investment strategy, which obtained 44.16%, Fig. 18, used 53 days as sliding window to identify the uptrend pattern in the time series and 50 days to identify the pattern of Fig. 18 in order to generate buying orders. This pattern is defined with two successive bottoms (points 2 and 3; points 4, 5 and 6), where the second is higher than the first, which represents an upward trend.

In 2013 the best strategy obtained a total return of 34.10% and is represented in Fig. 19. In this year the number of operations was the lowest due to the high length of days in the sliding window of the investment strategies, which in this case was 66 days and consequently in the time series that were used to identify with the uptrend pattern, which in this case was 94 days. The pattern used in this investment strategy is almost the reverse of the pattern used in 2011 (Fig. 17), which contains a trough (point 2) followed by a peak (point 6) and then a decrease of prices.

In the last year, the best strategy, which obtained 25.23% in total return, is represented in Fig. 20. In this case the length of the time series used in the exit by pattern was 63 days and they were represented by 8 PIPs and a limit of 2 relations between PIPs. The pattern identified in the time series is characterized by a slight increase between points 1 to 8, as illustrated in Fig. 20.

C. Case study n°3

In the previous case studies, the period of training and testing was always 1 year, which is a small period of time to obtain investment strategies that can outperform constantly the market in longer periods of time. So, in this case study the goal was to expand the period of training of the previous case study in order to obtain more robust solutions and test them in consecutive years. For that reason, the period of training was 2010-2011 (2 years) and the best solutions were tested in 2012-2014 (3 years). The tests were repeated 5 runs in the training period and then each of them was tested in the period of test.

The GA parameters were the same of the previous case study, 128 individuals and 50 generations as stop criteria. The parameters optimized by the GA were the ones of the chromosome of Fig. 8.

The results are represented in Table 4 which are the average of the 5 runs for the period. The results like in the previous case studies were compared with the Buy&Hold strategy. The total average return of the
investment strategies was higher than the return of Buy&Hold strategy in 17.23%, where the best strategy almost obtained the double (111.75%) of the Buy&Hold total return (61.91%).

<table>
<thead>
<tr>
<th>Period</th>
<th>N°</th>
<th>SR</th>
<th>D</th>
<th>SIR/GA Return (%)</th>
<th>B&amp;H</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-2014</td>
<td>86</td>
<td>68.84</td>
<td>59</td>
<td>59.74</td>
<td>78.45</td>
</tr>
</tbody>
</table>

Table 4 – Results of the average investment strategies for the period.

In Table 5, are represented the individual results of each strategy. The total return of the first strategy was by far and away the highest with a result of 111.75%. The worst strategy, in terms of total return, was the third with 59.74% but with the second highest success rate of operations. The lowest success rate was 50% of the fifth strategy, which obtained the second best total return and the highest average days in the market. The second strategy was the highest in number of operations (188) and at the same time the highest in success rate with 72.87%.

In Fig. 21 it is possible to observe the total return over the period 2012-2014 of the best SIR/GA strategy and the average of the SIR/GA strategies that are compared with the Buy&Hold strategy. As can be seen in Fig. 21 the investment strategies, the best and the average, start to increase more significantly than the B&H strategy in the middle of the period (around August 2013) and continue to outperform the B&H until the end of the period.

In Fig. 22 is represented the best strategy and its pattern, which obtained a total return of 111.75%. The exit method used by this investment strategy was by pattern, which proves again the capacity of this exit method to obtain good results in bull markets. The time series that were compared with the uptrend pattern had a length of 26 days. The time series were represented by 6 PIPs and the maximum limit of relations was 3. The pattern of this investment strategy is very similar to the Double Top pattern, which is very curious due to the fact that the Double Top is a bearish pattern but in this investment strategy the pattern was successfully used in a bull market because it was used to find bottoms that were followed by upward movements.

Some ideas to future work are:

- Test the program in other markets like European indexes (Euro Stoxx 50, DAX-30, etc.) to create a more robust program
- Include several technical indicators like OBV, RSI, etc. to support the decision of buy and sell in the investment strategies
- Add to the investment strategies the short operation in order to make the program much more completed for the real life scenarios and to perform in bear markets too
- Add an option to find some wanted and well-known patterns like the Double Bottom and Top, Head-and-Shoulders, etc.

ACKNOWLEDGMENT

This work was supported in part by Fundação para a Ciência e a Tecnologia (Project UID/EEA/50008/2013).

REFERENCES


