Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition

António Gorgulho, Rui Neves, Nuno Horta

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

Abstract

The management of financial portfolios or funds constitutes a widely known problematic in financial markets which normally requires a rigorous analysis in order to select the most profitable assets. The presented paper proposes a new approach, based on Intelligent Computation, in particular genetic algorithms, which aims to manage a financial portfolio by using technical analysis indicators (EMA, HMA, ROC, RSI, MACD, TSI, OBV). In order to validate the developed solution an extensive evaluation was performed, comparing the designed strategy against the market itself and several other investment methodologies, such as Buy and Hold and a purely random strategy. The time span (2003–2009) employed to test the approach allowed the performance evaluation under distinct market conditions, culminating with the most recent financial crash. The results are promising since the approach clearly beats the remaining approaches during the recent market crash.

1. Introduction

The fast technology evolution together with the massive evolvement of financial markets in modern societies leads, nowadays, to an increasing interest to the field of computational finance.

This field is becoming popular among computer scientists, especially to computational intelligence specialists who try to combine elements of learning, evolution and adaptation in order to create intelligent software. In particular, subjects such as neural networks, swarm intelligence, fuzzy systems and evolutionary computation are becoming extremely notorious on market’s domain. The mentioned techniques can be applied to financial markets in a variety of ways: to predict the future movement of a stock’s price, or to optimize a collection of investment assets, such as a fund or a portfolio. This innovation is of special importance due to the high volume of securities (financial instruments) involved, normally, it is very hard for a simple investor to optimize his profits without requiring the skills of financial markets specialists. The goal of this work is to provide an application which tries to partially replace those specialists in order to help an investor or an investment company to achieve a significant profit on buying and selling (trading) financial instruments. In order to apply such procedures we must accept that the historical data related to stocks and markets gives appropriate signals about the market future performance. This premise constitutes the basis of technical analysis which simply tries to analyze the securities past performance in order to evaluate investments at the present time. This philosophy relies on three bases (Murphy, 1999); the fact that market action discounts everything, the fact that price moves in trends, and that history tends to repeat itself. These considerations allow, through the study of charts and financial data, the recognition of which way the market is more likely to go. Despite the fact that technical analysis is becoming widely used, there are still some criticisms to this perception on the market evolution. For instance, Burton Malkiel (Malkiel, 1973) stated that the “past movement or direction of the price of a stock, or overall market cannot be used to predict its future movement”. His findings become popular, leading to a new investment theory called The Random Walk Theory where the author stipulates that if we cannot beat the market, then the best investment strategy we can apply is Buy and Hold in which an investor buys stocks and holds them for a long period of time, regardless of market fluctuations. For the technical community, this idea of purely random movements of prices is totally rejected, and more recent studies (Lo & MacKinlay, 2001; Park & Irwin, 2007) try to evidence their beliefs. For instance, in (Lo & MacKinlay, 2001) the author demonstrated the validity of technical analysis using more than seventy technical indicators which showed that market movements can be predicted at a certain degree. Also, if we consider the price movement as unpredictable, how can we explain that price moves in trends? If we observe several stock charts considering a predefined period we can easily detect an uptrend or a downtrend.
The presented paper provides a detailed discussion on a new approach for intelligent portfolio management. The paper is structured as follows: Section 2 addresses the theory behind the developed work, namely the concepts of financial portfolio, portfolio management, and technical analysis. Also, in this section, it is given a brief overview about different methodologies which can be used to address the portfolio management problematic. Section 3 illustrates the system architecture. Section 4 proposes the validation procedure used to evaluate the developed strategy. Section 5 summarizes the provided document and supplies the respective conclusion.

2. Related work

To get a better understanding about the underlined problem and the existing solutions, some of the fundamental concepts and tools related to the financial portfolio composition are explained in the following subsections.

2.1. Financial portfolio

A financial portfolio (Maginn, Tuttle, McLeavey, & Pinto, 2007) consists in a group of financial assets, also called securities or investments, such as stocks, bonds, futures, contract for difference (CFDs), or groups of these investment vehicles known as exchange-traded-funds (ETFs). In order to construct a portfolio, it is capital to define investment objectives that should focus on a certain and accepted degree of risk, i.e. the chance of incurring in a loss.

The core of this work is related to portfolio management, the act of deciding which assets need to be included in the portfolio, how much capital should be allocated to each kind of security and when to remove a specific investment from the holding portfolio. During this process, it is required to take into account the investor's preferences since some investors are more willing to accept a specific degree of risk than others, hoping that way to achieve better returns.

2.2. Portfolio management

As it was already mentioned, the goal of this work is concentrated on the automatic management of a portfolio. So, it is important to understand that we can apply two forms of management (Maginn et al., 2007):

Passive management in which the investor concentrates his objective on tracking a market index. This is related to the idea that it is not possible to beat the market index, as stated by the Random Walk Theory (Malkiel, 1973). More concretely, a passive strategy aims only at establishing a well diversified portfolio without trying to find under or overvalued stocks.

Active management in which the main goal of the investor consists on outperforming an investment benchmark index, buying undervalued stocks and selling overvalued ones.

In the case of the work here described, the purpose is to adopt an active management approach by using technical analysis indicators and evolutionary computation techniques.

2.3. Technical analysis

When defining a financial fund or portfolio the goal is to pick the best potential assets within the market in order to minimize losses and maximize returns. There are several ways to perform a reasonable evaluation of the market to select potential profitable securities. Usually, investment analysts perform a fundamental or a technical analysis of the market. In this work, a pure technical analysis (Murphy, 1999) methodology was employed. A technical analyst believes that market action, namely, the volume of transactions and the securities prices include all the fundamentals that can possibly affect market's price; political, economical, or psychological. The applied strategies based on technical analysis normally embody a set of technical indicators which try to give a future perspective of market development according to what is visible on price charts. A technical indicator consists in a formula that is normally applied to stock's prices and volumes. The resulting values are plotted and then analyzed in order to offer a perspective on price evolution. More specifically, a technical indicator tries to capture the behavior and investment psychology in order to determine if a stock is under or overvalued. In Section 3, several technical indicators will be discussed and illustrated.

2.4. Automatic portfolio management approaches – overview

In respect to the solutions already developed to address the portfolio management problem, most of them focus on a passive management approach by using the Mean–Variance model (Markowitz, 1972) proposed by Harry Markowitz. The author is pioneer in the Modern Portfolio Theory (MPT) after analyzing the effects related with risk, correlation and diversification over the expected returns of investment portfolios. After completing his study, Markowitz concluded that rational investors should diversify their investments, in order to reduce the respective risk and increase the expected returns. The author’s assumption focus on the basis that for a well diversified portfolio, the risk which is assumed as the average deviation, from the mean, has a minor contribution to the overall portfolio risk. Instead, it is the difference (covariance) between individual investment’s levels of risk that determines the global risk. Based on this assumption, Markowitz provided a mathematical model which can be easily solved by metaheuristics such as Simulated Annealing (SA), Tabu Search (TS) or genetic algorithms (GA).

Generally, solutions (Chang, Meade, Beasley, & Sharaiha, 2007; Cura, 2009; Schaefer, 2002), based on this model, focus their goal on optimizing a single-objective; the risk inherent to the portfolio, in order to determine the optimal portfolio composition and the weights assigned to each of the chosen stocks. Besides this single-objective formulation, other approaches (Branke, Scheckenbach, Stein, Deb, & Schmeck, 2009; Streichert, Ulmer, & Zell, 2003) try to optimize simultaneously two conflicting objectives, the global risk and the expected returns of the securities within the portfolio.

Besides the referred works which mainly generate a diversified portfolio maintaining it for a specific set of time, Aranha and Hitoshi (Aranha & Hitoshi, 2008; Aranha & Iba, 2007, 2009) provided a very interesting active management approach, by coupling the Markowitz’s model with a modeling cost mechanism, responsible for rebalancing the portfolio trough time while, at the same time, minimizing the transaction costs. In their works, a completely different portfolio representation is used, based on a tree structure, which allowed them to obtain very interesting results.

Although Markowitz’s model is widely used to design the portfolio optimization problem, other models can also be considered. For instance, Black and Litterman (Black & Litterman, 1992) suggested a new formulation, the Black–Litterman model. In their work they propose means of estimating expected returns to achieve better-behaved portfolio models. The designed model is very similar to the Markowitz’s one, the main difference is concentrated on the calculation of the expected returns which generates portfolios considerably different when using the original model. According to the authors their new design tries to rectify some of the flaws presented by Markowitz’s model.
In addition to this passive approach which only tries to maintain a well diversified portfolio, recurring to the Markowitz’s model for picking the assets from the market and assigning the respective weight within the portfolio. It is also possible to adopt an active strategy which tries to find under or overvalued stocks in order to achieve a significant profit with price's rise or fall. For instance, Wagman (Wagman, 2003) provided a simple framework based on genetic programming (GP) which tries to find an optimal portfolio with recurrence to a simple technical analysis indicator, the moving average (MA). The provided solution starts by generating a set of random portfolios (population), and the GP algorithm tries to converge in an optimal portfolio by using an evaluation function which considers the weight of each asset within the portfolio and the respective degree of satisfaction against the MA indicator, using different period parameters.

Another solution, based on the same kind of analysis, was provided by Yan, Clack et al. (Hassan & Clack, 2009; Yan, Sewell, & Clack, 2008). Their solution is based on a genetic programming approach which tries to find an optimal model to classify the stocks within the market. The top stocks adopt long positions, the bottom ones, short positions. This approach is very interesting since it is capable to get a very realistic experience on financial portfolio management, besides being very robust. Their model is based on the employment of a fundamental analysis approach which consists on studying the underlying forces of the economy to forecast the market development.

Table 1 summarizes some of the most relevant existing solutions to approach the portfolio problematic, specified according to several parameters.

### 3. Proposed approach

#### 3.1. System architecture

The system’s architecture can be structured, as illustrated in Fig. 1, on a traditional layer architecture composed by three distinct layers:

Each layer is associated with several modules, represented by the oval shapes. The presented modules correspond to distinct units of implementation with a specific functional responsibility within the system.

#### 3.2. Data flow

In respect to the data flow within the application, as illustrated in Fig. 2, the system starts to ask distinct inputs from the user, executes the optimization algorithm, and then provides the recommended portfolio. More specifically, the complete process is performed as follows:

The user starts by specifying the desired parameters, depending on its role, which can be normal or advanced, according to its knowledge on optimization techniques. Afterwards, the system applies a set of technical indicators in order to calculate the values given by those indicators on the available data prices. After this process, the GA starts its execution by defining several random individuals, which correspond to different models for classifying the market’s assets. These different models, called Classifier Equations take into account the set of data calculated in the previous step. In order to evaluate each individual, an Investment Simulator is necessary to rank each stock within the market and subse-
quently, picking the best stocks for defining a financial portfolio. Afterwards, the portfolio is updated and evaluated during the training period in order to classify the attractiveness of the current classifier equation in terms of its performance on the end of the considered time period.

When the GA converges in a final solution, the system executes again the investment simulator system, but to the current date period, in order to provide the recommended portfolio taking into account today’s date.

Every week the Investment Simulator is again executed to update the current portfolio, adding new positions or closing former ones. From time to time, the GA process is repeated so that a new classifier equation is determined considering the most recent data.

3.3. Data layer

The data layer is responsible for managing financial data. Its behavior is decomposed in two distinct modules, the financial data processing module and the technical rules module.

3.3.1. Financial data processing module

This module is accountable for processing all the financial data which is of primary use on the developed application. In order to provide to the system the ability of generating real-life portfolios, it is necessary to first download a complete history of all the available data on distinct markets. The process of retrieving all the historical data was performed just once. Afterwards, it is only necessary to update the database with new available information.

In respect to the considered data, the Dow Jones index was used:

The DJI, Dow Jones Industrial Average Index (Dow Jones Indexes, 2010), which contains the stock prices of 30 of the largest held companies in the United States.

All the financial data relative to the former index was downloaded through the Yahoo Finance Database (Yahoo! Inc, 2010). The complete retrieving process can be described as following:

- Specify the desired index. Each index is identified with a unique keyword. For instance, the Dow Jones Industrial Average is tagged with the acronym DJI.
- After defining the target index, the download process is executed and a single file containing the tickers (specific group of letters representing a particular security) of all companies composing the previously defined index, is stored. The second process consists on downloading all the historical data, from a specific date until today's date for each of the previously acquired companies. The designer has the possibility of indicating the desired data period through a single parameter; daily, weekly, or monthly. Within this download process, the storage functionality is executed, responsible for defining csv files with the desired financial data. Each record within these stored files has the following configuration:

<table>
<thead>
<tr>
<th>Date</th>
<th>Open</th>
<th>High</th>
<th>Low</th>
<th>Close</th>
<th>Volume</th>
<th>Adj. Close</th>
</tr>
</thead>
</table>

where:

- **Date**: The date record, using the format “dd-mm-yyyy”.
- **Open**: The opening price in which the security was traded during a specific date.
- **High**: The highest price in which the stock was traded during a specific date.
- **Low**: The lowest price in which the equity was traded during a specific date.
- **Close**: The closing price in which the asset was traded during a specific date.
scores were used: week, or month) in the data set, a score was assigned. Four distinct

**Adj. Close:** The adjusted closing price in which the stock was traded during a specific date.

- After the complete historical data has been downloaded, when the application is again executed, the update action is invoked, which corresponds to a specific method accountable for processing all companies’ files and for each one identifying the last record, in particular the date of the last available record. After this processing phase, each company’s file is updated; the new necessary records are inserted. Notice that each data file is ordered from the oldest date to the most recent one.

### 3.3.2. Technical rules module

One of the major problems faced on portfolio management is the right choice of assets. However, technical indicators can be used to give a future perspective on its behavior in order to determine the best choice. So, in order to classify each asset within the market, a set of rules based on technical indicators was employed.

Nevertheless, there are several problems that can show up when using technical indicators. First, there’s not a better indicator, the indicators should be combined in order to offer us different perspectives. Sometimes a technical indicator gives false signals, so our best option is to combine different technical indicators. Second, a technical indicator always needs to be applied to a specific time span, it can be 10 days, 50 days, more or less. Determining the best time period is a complex choice, in this case the selected time window considered was the one proposed by the technical analysis experts, for each of the used indicators.

Regarding the GA aspects, the algorithm can be applied in several ways, as to determine the best time span, for instance, Fernández-Blanco et al. (Bodas-Sagi, Fernández, Hidalgo, Soltero, & Risco-Martín, 2009; Fernández-Blanco, Bodas-Sagi, & Hidalgo, 2008) applied an EA to determine the best settings for the MACD and RSI indicators. However, in this work the algorithm is applied in the context of obtaining the best model to classify the assets, an optimal balance between different technical indicators. Since only one indicator cannot possibly be enough, we try to find which were the best indicators to use in the past to form a basket of securities and subsequently, pick the most attractive assets.

In this work several technical indicators were applied to find attractive stocks in the market. The indicators were chosen in order to build a basket of different types of technical indicators such as momentum oscillators and trend following devices:

A trend following indicator tries to identify trends in the market. A trend represents a consistent change in prices, the investors’ expectations.

A momentum based indicator tries to measure the velocity of directional price movement in order to identify the speed/strength of a price movement and the enthusiasm of buyers and sellers involved in the price development.

For each technical indicator calculated for each period (day, week, or month) in the data set, a score was assigned. Four distinct scores were used:

**Very Low Score:** Assigns – 1.0 points, indicates a strong sell/short signal.

**Low Score:** Assigns – 0.5 points, indicates an underperformed signal, potentially to sell or to go short.

**High Score:** Assigns – 0.5 points, indicates a reasonable buy signal.

**Very High Score:** Assigns – 1.0 points, indicates a strong buy signal.

Taking into account all the historical data, for each period a specific score was assigned taking into account the following technical indicators and defined rules.

#### 3.3.2.1. Extensibility and technical rules module implementation

As stated before, the intent was to mix different kinds of technical indicators: oscillators and trend following mechanisms. In order to respect that guideline, the indicators employed were the *Exponential Moving Average* (EMA), the *Hull Moving Average* (HMA), the *Rate of Change* (ROC), the Double Crossover method, the *Relative Strength Index* (RSI), the *Moving Average Convergence Divergence* (MACD), the *True Strength Index* (TSI), and the *On Balance Volume* (OBV). Notice however, that is possible to easily extend the developed solution with more technical indicators. The algorithm is adapted for any technical indicator, the only requirement is to implement the desired indicators and define the respective rules. On adding more indicators, the confirmation of a possible buy or sell signal is possibly more accurate, improving the results of the designed solution. The parameters assigned to each technical indicator can also be altered to any value desired by the designer.

In the following paragraphs, the reader can have an insight on the selected indicators as well as the respective classification rules.

#### 3.3.2.2. Exponential Moving Average (EMA)

The *Exponential Moving Average* (EMA) (Murphy, 1999) is a trend following indicator. The goal of this device is to identify that a trend has begun, or it is finishing its cycle. In order to accomplish it, the EMA averages the price data, in order to produce a smooth line which can be easily perceived, in contrast to the irregular curve signaling the prices. There are several kinds of moving averages; the exponential one assigns more weight to the most recent data in order to give more importance to it. Its formula can be defined as following:

\[
EMA_t(n) = EMA_{t-1}(n) \times \left(1 - \frac{2}{n + 1}\right) + X_t \left(1 + \frac{2}{n + 1}\right)
\]

Where:

- \( n \) is the length of the moving average.
- \( X \) corresponds to the stock’s price.
- \( t \) defines the considered period (day, week, or month).

Based on this indicator, the rules, presented in Table 2, were defined.

Fig. 3 provides an example of the EMA line. As can be seen, it defines a smoothing curve which can be easily analyzed, in contrast to the zigzag performed by the stock’s prices.

#### 3.3.2.3. Hull Moving Average (HMA)

Like the EMA, the *Hull Moving Average* (HMA) (Hull, Hull Moving Average, 2004), illustrated in Fig. 4, tries to identify the prevailing market trend. However, it can define a smoother curve and can follow the price graph much more closely, reducing the lag present on its predecessor moving average. Its formula is calculated as following:

\[
HMA_t(n) = WMA_t\left(\text{floor}\left(\sqrt{n}\right)\right)
\]

\[
\text{of} \left(2 \times WMA_t\left(\text{floor}\left(\sqrt{n} \div 2\right)\right) - WMA_t(n)\right)
\]

Where:

- \( n \) is the length of the moving average.
- \( WMA \) corresponds to the weighted moving average.
- \( t \) defines the considered period (day, week, or month).

What this formula says is that the Hull moving average is the weighted moving average of length square root \( n \) of the difference
between two weighted moving averages one of length \( n/2 \) and the other \( n \). In this difference the weighted moving average of length \( n/2 \) has 2 times the value of the weighted moving average of \( n \).

For the HMA indicator, the rules, presented in Table 3, were defined.

3.3.2.4. Double Crossover. The Double Crossover (Murphy, 1999), illustrated in Fig. 5, method is characterized by using two distinct moving averages to generate market signals. Normally, it is made a couple between a shorter moving average (more sensible to the market signal and consequently faster, although it can produce false signals) and a longer moving average which has a longer lag, although it can produce better trend signals. In this work, the couple was made between an exponential moving average of five weeks and one with twenty weeks.

For the Double Crossover method, the rules, presented in Table 4, were defined.

The picture below demonstrates how we can apply the Double Crossover procedure.

3.3.2.5. Rate of Change (ROC). The Rate of Change (ROC) (Murphy, 1999), illustrated in Fig. 6, ratio presents the percentage difference
between the current closing price and the price \( n \) time periods ago. On doing so it allows us to measure how rapidly the price of a specific stock is moving. If the price is rising or falling too quickly it will probably indicate overbought or oversold conditions. Its formula can be specified as following:

\[
\text{ROC}_t(n) = \frac{X_t - X_{t-n}}{X_{t-n}} \times 100
\]

Where:

- \( n \) is the number of periods considered.
- \( X_t \) corresponds to the stock's price on period \( t \).

For the ROC indicator, the rules, presented in Table 5, were defined.

3.3.2.6. Relative Strength Index (RSI). The Relative Strength Index (RSI) (Murphy, 1999), illustrated in Fig. 7, indicator is a momentum oscillator used to compare the magnitude of a stock's recent gains to the magnitude of its recent losses, in order to determine overbought or oversold conditions. The formula used on its calculation is:

\[
\text{RSI}_t(n) = 100 - \frac{100}{1 + \text{RS}(n)}
\]

Where:

- \( \text{RS} = \frac{\text{Average gains}}{\text{Average Losses}} \).
- \( t \) defines the considered period (day, week, or month).

When calculated, the RSI line forms a signal between 0 and 100, which specifies determined overbought or oversold conditions.
when its value is above or below specific levels. For the RSI indicator the rules, presented in Table 6, were defined.

3.3.2.7. Moving Average Convergence Divergence (MACD). The Moving Average Convergence Divergence (MACD) (Murphy, 1999) indicator constitutes one of the most reliable indicators within the market. It is a trend following momentum indicator that exhibits the relation between two distinct moving averages. Essentially, it defines two lines; the MACD line which corresponds to the difference between a 26-week and 12-week EMA and a trigger line which corresponds to an EMA of the MACD line. The difference between the former lines allows us to obtain a histogram which can be easily analyzed and offering us perspectives on price evolution.

\[ MACD_t(s,l) = EMA(s)_t - EMA(l)_t \]  
\[ Trigger_t(n) = EMA_t(n) \text{ of } MACD_t(s,l) \]  
\[ Hist_t = MACD_t(s,l) - Trigger_t(n) \]

Where:
- \( n \) is the number of periods considered for the trigger signal.
- \( s \) corresponds to the number of periods considered for the shorter MA.
- \( l \) corresponds to the number of periods considered for the longer MA.

For the MACD indicator, the rules, presented in Table 7, were defined.

Fig. 8 exemplifies the application of the MACD histogram.

3.3.2.8. On Balance Volume (OBV). The On Balance Volume (OBV) (Murphy, 1999), illustrated in Fig. 9, indicator is a momentum indicator that relates volume with price change. It tries to show if volume is flowing into or out of a security, assuming that volume changes precede price changes. For instance, a rising volume can indicate the presence of smart money flowing into a security preceding its rise on price. The OBV line can be calculated as following:

\[ IFX_t < X_{t-1} \rightarrow OBV_t = OBV_{t-1} - V_t \]  
\[ IFX_t > X_{t-1} \rightarrow OBV_t = OBV_{t-1} + V_t \]  
\[ X_t \text{ corresponds to the stock's price on period } t. \]
\[ V_t \text{ is the volume referent to period } t. \]

The OBV always takes a direction, a rising OBV line indicates that the volume is heavier on up days, confirming a possible up trend.

For the OBV indicator, the rules, presented in Table 8, were defined.

3.3.2.9. True Strength Index (TSI). The True Strength Index (TSI), illustrated in Fig. 10, is a momentum-based indicator which tries to determine both trend and overbought or oversold conditions. In order to accomplish these features, the TSI corresponds to a one-day/week/month momentum which is double smoothed with two moving averages to show the trend and specifying, at the same time, the overbought and oversold conditions. Its formula can be defined as following:

\[ MNT = X_t - X_{t-1} \]  
\[ TSI_t(r,s) = 100 * \frac{EMA(s) \text{ of } (EMA(r) \text{ of } MNT)}{EMA(s) \text{ of } (EMA(r) \text{ of } |MNT|)} \]  
\[ Trigger_t(n) = SMA_t(n) \text{ of } TSI_t(r,s) \]

Where:
- \( X_t \) corresponds to the stock’s price on period \( t \).
- \( MNT \) corresponds to the momentum line which calculates the difference between the current price and the price observed on the previous period.
- \( r \) corresponds to the number of periods considered for the first EMA.
- \( s \) corresponds to the number of periods considered for the second EMA.
corresponds to the number of periods considered for the trigger line.

The rules, presented in Table 9, were defined for the former indicator.

3.4. Business logic layer

This layer is accountable for defining the optimizer techniques and correspondent representation in order to result on a classifier system capable of defining models to score the different assets within the market. The layer is structured on two distinct modules: Optimization Module, responsible for the optimization process, and the Investment Simulator Module, accountable for simulate the portfolio management during the training/validation.

3.4.1. GA Optimization Module

Since a GA is composed by several components we will start to describe how each component of the algorithm was defined.

3.4.1.1. Chromosome representation. Starting with the chromosome representation, an individual in the population is represented by a real valued array structure where each element corresponds to the weight, importance given to a specific technical rule within the classifier equation. Besides the described weights, assigned to each technical rule, four bound values are also employed to define the necessary score that an asset needs...
to obtain so it can adopt a long or a short position within the portfolio, or to close the former position. In order to get a better understanding on the considered representation, presented in Table 10.

As we can observe, from the previous table, each rule has a specific weight within the classifier model. The classifier is given by the following equation:

\[
X_N^i = \sum_{i=0}^{N} W_i \cdot \text{Score}(X, i) 
\]

(14)

\[
0 \leq W_i \leq 1
\]

(15)

\[
0 \leq \sum_{i=0}^{N} W_i \leq 1
\]

(16)

Where:

- \( W_i \) is the weight/importance assigned to the technical rule \( i \).
- \( \text{Score}(X, i) \) corresponds to the score given by the technical rule \( i \) to stock \( X \).

After the optimization performed by the algorithm, resulting on a classifier equation, where a set of technical indicators are correctly balanced, all the assets within the market are classified, as illustrated in Fig. 11. The stocks whose classification is higher than
the value given by the Buy Limit field adopt long positions. The ones whose classification is below the Short Limit adopt short positions. The last two bound values; Close Buy Position and Close Short Position determine the necessary score to achieve so a specific position in the portfolio can be closed. Notice, however, that more conditions need to be fulfilled so a specific position within the portfolio can be closed.

Supposing the exemplified chromosome and financial market, represented on the previous figure:

**Week 1:** Given the Score \((X, i)\) which represents the score given to X on a specific period, by the indicator \(i\), and the chromosome values, we can easily calculate the punctuation given to the contemplated assets. Considering the obtained performance, only the securities A and C present a higher classification than the value given by the Buy Limit Gene, defining this way the portfolio composition;

**Week 2:** Given the previous explanation; securities B and D are included on the basket. Since the position C presents a classification lower than the Close Buy Gene, in opposition with the security A; position C its closed and A is maintained.

**Week 3:** Considering the former processes, stocks B and D compose the final portfolio.

Notice, however, that the proposed example only represents a short view on the considered strategy. The developed investment process considers several technical indicators and applies stop techniques in order to avoid sudden losses not predicted by the applied algorithm.

3.4.1.2. Selection. After defining the encoded representation it is necessary to specify how the algorithm will choose the individuals that will generate offsprings for the next generation. This process is performed via a Truncation Selection methodology which mainly consists on sorting the population according to their fitness, and subsequently, selecting the best individuals for reproduction. From the set of best individuals a roulette procedure is applied, in order to choose the breeders. The number of considered parents is given by the Trunc Threshold parameter, which is set to be half of the population, by default.

3.4.1.3. Mutation. In respect to the mutation procedure, a new random value is generated for each variable selected for mutation. The number of variables to be mutated depends on the value given to the Mutation Rate parameter, the chromosome size, and the number of population individuals as you can see below:

\[
\text{Mutations} = \text{Mut Rate} \times \text{Cr Size} \times (\text{Pop Size} - 1)
\]
As you can observe from the previous equation, the number of mutations largely depends on the number of total variables considered by the algorithm. Notice, however, that one single individual was discarded, as you can see from the minus one within the equation. The purpose of this restriction is to maintain the best individual in the current population, in each generation of the algorithm. This technique is normally referred as Elitism.

3.4.1.4. Crossover. Considering the crossover operator, different types of crossover operators were implemented, in particular, the Single Arithmetic Recombination, the Whole Arithmetic Recombination, and the One-Cut Point Crossover method, contemplating the generation of two offsprings. After performing a rigorous testing on the algorithm convergence, it was concluded that the one-cut point methodology allowed us to obtain the best results for the represented chromosome.

3.4.1.5. Constraints. One of the major problems presented by the defined chromosome concentrates on the restrictions over the different weights assigned to the stipulated technical rules.

A trivial way on handling an inequality constraint such as the former one consists on applying a death penalty function (Coello, 1999), discarding infeasible individuals within the population. Although it seems an extremely basic approach, this methodology has as major problem the fact of not exploring any information from the infeasible individuals, in order to guide the search more effectively. To surpass this problem, we have employed a simulated artificial immune system (Coello & Cortés, 2001) which provides an efficient way of guiding the search, taking into account the information generated by the infeasible individuals. Besides the fact of being easy to implement, this strategy is also very effective on the proposed goal of exploring information gathered by the non feasible genes. Very generically, the algorithm maintains in each generation a population of infeasible individuals designated as antibodies which suffers the same kind of evolution of the main population. However, the evaluation function is much easier which allows us to rapidly execute the convergence process within this smaller population. This convergence procedure corresponds to the process of executing a genetic algorithm inside the main genetic algorithm. The principle behind this algorithm corresponds to the Negative Selection Model which tries to capture the behavior of the human immune system on knowing what is really part of the human system, and what is not.

3.4.1.6. Evaluation function. In order to evaluate each individual within the population, so the algorithm can pick the best ones for reproduction, and consequently, converge on an optimal solution, the Return On Investment (ROI) function was applied. The ROI is used to evaluate the efficiency of different investments during a specific period of time.

As you can see, a simple objective was considered for evaluating each solution, i.e., the goal of the algorithm is to maximize the ROI. However, the solution could be easily extended with a multi-objective consideration, where the goal was to optimize simultaneously two conflicting objectives; the ROI and the risk involved, which could be measured by the volatility of returns, for instance.

3.4.2. Investment simulator module

In order to evaluate each individual, an investment simulator (IS) is necessary for generating a portfolio according to the classifier equation, and managing it through time. This management module is used by the genetic algorithm, in order to classify each chromosome and performing test/real-life simulations. There are several specifications that need to be concretely defined over this Investment Simulator module. As already stated the IS will use a specific equation to classify the assets within the market.

The complete management process is the following:

- The first step consists on applying 50% of the available budget on generating the initial portfolio using the equation given by the algorithm.

- In each new week, during the period of validation or training, the portfolio is updated using the following rules:
  o If there are positions in the portfolio presenting a loss of 10% or higher, the current position is immediately closed. This condition is an insurance to avoid an unexpected crash on the company.
  o If there is a position which presents a score indicating a possible close and it has already given profit, the position is closed.
  o If there are stocks in the market which present a classification possible to add, and the portfolio has not achieved its maximum size, new positions are formed within the portfolio. The remaining 50% of the budget is used for considering these new positions.

3.5. Presentation layer

The presentation layer is responsible for the application interface. Since a detail explanation of this layer is out of scope for this paper, just a brief description on the necessary inputs, to be specified by the average user, is given:

**Budget**: The capital available to invest.

**Max Size**: The maximum number of assets included on the desired portfolio.

**Short Selling**: This parameter is used for specifying if short selling is allowed or the user just want to adopt long positions.

**Transaction Costs**: Used for the consideration of transaction costs. This parameter is used to include the commission costs involved on buying or selling shares.

4. System validation

In this section the validation approach for evaluating the developed system is described.

In order to validate the designed application during its development, a Backtesting (Ni & Zhang, 2005) strategy was employed.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>All stocks from DJI</td>
</tr>
<tr>
<td>Period</td>
<td>01/01/03 – 31/06/09</td>
</tr>
<tr>
<td>Budget</td>
<td>100 000 USD</td>
</tr>
<tr>
<td>Maximum size portfolio</td>
<td>10</td>
</tr>
<tr>
<td>Short selling?</td>
<td>TRUE</td>
</tr>
<tr>
<td>Commissions</td>
<td>0.02 / Share Minimum Fee: 14.00 USD</td>
</tr>
<tr>
<td>Number of executions</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>64</td>
</tr>
<tr>
<td>Mutation rate (%)</td>
<td>10</td>
</tr>
<tr>
<td>Generations</td>
<td>350</td>
</tr>
<tr>
<td>Trunc. Threshold (%)</td>
<td>50</td>
</tr>
<tr>
<td>Sliding window</td>
<td>6 months/6 months</td>
</tr>
</tbody>
</table>
4.1. Case study – description

In order to validate the designed solution, a critical time interval including both bullish and bearish trends, the DJI between 2003 and 2009, was considered. Moreover, the proposed approach is compared against the B&H and a random investment strategy.

**Buy and Hold**: According to some theories, prices are independent to each other, meaning that we cannot use past data to forecast market development, so the best strategy we can employ is Buy and Hold on which we maintain a specific set of assets regardless of market fluctuations. In this case the market index was considered as reference for the B&H strategy.

**Random**: The random strategy implemented has a purely random behavior; each new week the portfolio is updated by closing random positions, and picking new random assets from the market, to add to the already existent portfolio. Both long and short positions are considered.

4.2. Case study – parameters setup

For the designed case study, the configuration described in Table 1 was applied, particularly, all stocks from the DJI were considered for performance analysis and for picking up the most promising for the portfolio composition, the maximum portfolio size is parameterized and was here set to 10, additionally, both long and short selling was allowed including transaction costs and, finally, the process was executed for 100 times in order to allow a proper.
In respect to the evolutionary strategy, the GA kernel was set with the parameters described in Table 12. Here, Sliding Window refers to the training/testing period combination employed during the evaluation period. For instance, if the validation starts on January 2003, then the previous six months are used to train the algorithm. Then, the process repeats for each six months period until the end of the evaluation period.

4.3. Case study – performance analysis

The graph, illustrated by Fig. 12, exhibits the return on investment (ROI) obtained for the considered strategies within the years of 2003 to 2009, for the B&H, the Random, the proposed approach (GA) and the Best GA iteration. The GA and particularly the Best GA clearly outperform the B&H and Random strategies. This, together with the histogram of Fig. 13, were the results from the 100 runs are outlined, shows, first, that applying a Random strategy for portfolio selection is clearly worse than any of the other approaches in a long run, moreover, the distribution of the results from proposed strategy (GA) show that a large majority of the 100 runs lie above the B&H giving a high confidence level for the proposed strategy.

A detailed statistical analysis on the performance of each strategy is presented on Table 13. The intervals here presented correspond to a confidence degree of 95% determined by 100 runs. The ROI gives a return interval for the proposed approach, referred as GA, which clearly is above the competing strategies outlining their superiority. Moreover, the number of positions with positive return exceeds 80%, for the GA, confirming again the high confidence level of the proposed approach.

5. Conclusions

This work proposes capable new approach to automatically manage a portfolio by using a GA conjugated with technical analysis rules. As observed under the previous sections, the system shows a good adaptive degree to different market trends achieving outstanding return rates. Although, several management rules were defined to increase the system performance, evolutionary computation plays a fundamental role to provide a correct balance between several types of technical indicators in order to pick the most promising stocks for portfolio composition.

Table 13

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Random</th>
<th>GA</th>
<th>Best GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI (%)</td>
<td>[–21.97, –14.77]</td>
<td>[16.68, 25.29]</td>
<td>62.95</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>–0.93</td>
<td>0.21</td>
<td>0.67</td>
</tr>
<tr>
<td>Sortino Ratio</td>
<td>−1.25</td>
<td>0.40</td>
<td>21.03</td>
</tr>
<tr>
<td>Positions</td>
<td>[1371, 1389]</td>
<td>[151, 159]</td>
<td>156</td>
</tr>
<tr>
<td>Profitable Positions (%)</td>
<td>[46.83, 47.55]</td>
<td>[80.24, 81.50]</td>
<td>88.46</td>
</tr>
<tr>
<td>Non Profitable Positions (%)</td>
<td>[52.45, 53.17]</td>
<td>[18.50, 19.76]</td>
<td>11.54</td>
</tr>
<tr>
<td>Avg. Profit Per Position (%)</td>
<td>[−0.16, −0.07]</td>
<td>[1.93, 2.53]</td>
<td>4.00</td>
</tr>
<tr>
<td>Max. Profit (%)</td>
<td>[63.04, 78.21]</td>
<td>[104.69, 136.57]</td>
<td>59.66</td>
</tr>
<tr>
<td>Min. Profit (%)</td>
<td>[−42.96, −39.03]</td>
<td>[−36.46, −34.94]</td>
<td>−30.28</td>
</tr>
</tbody>
</table>

References


