ISSN 1870-4069

Generating Trading Strategies in the Mexican Stock Market: A Pattern Recognition Approach

David Ricardo Montalván Hernández, Ricardo Barrón Fernández, Salvador Godoy Calderón

Instituto Politécnico Nacional, Centro de Investigación en Computación, Mexico City, Mexico
{davidricardo888, barron2131, sgodoyc}@gmail.com

Abstract. With the digitalization of financial markets, namely, stock markets the development of algorithms and computational techniques in order to determine trading strategies has gained relevance as much as in academia as in the industry. This article explores the use of pattern recognition techniques (Support Vector Machines, Multilayer Perceptron and C4.5) as tools for finding trading strategies in the Mexican Stock Market. Our results show that, statistically speaking, the methods proposed here, cannot outperform the so called Efficient Market Hypothesis in its weak version. Nonetheless, this paper presents a labeling method for financial time series, which permits future investigations using other supervised learning techniques.

Keywords: trading strategies, Mexican stock market, pattern recognition, support vector machines, multilayer perceptron, C4.5.

1 Introduction

The increase in computing power, the digitalization of financial markets and the opportunity of generate big profits, have motivated to a large degree the research and development of computational algorithms whose purpose is the guidance in the investment decision process.

Even when the basic idea is: ”buy low and sell high”, given the uncertainty of financial markets, this research has used mathematical and computational techniques in order to create models that help in the trading decisions (when is the right time to buy or sell). Namely, the use of techniques in the field of artificial intelligence have gained notoriety.

The objective in this paper is the proposal of a method or algorithm in order to generate trading signals for the Mexican stock market.

We will try to find a set of trading signals that will generate profits (losses) bigger (minor) than the ones generated by the *Buy and Hold* strategy (the main benchmark in this kind of research), which is discussed in Section 3 and consists basically in the following actions:

- Fix a time period $[0, T]$.
- Buy at time 0 at price P_0.
- Sell at time T at price P_T.
- The percetange profit (loss) is given by $\dfrac{P_T - P_0}{P_0}$.

This work also compares the performance of various pattern recognition methodologies when applied to the generation of trading strategies.

Lastly, two contributions are considered. The first is the analysis of Mexican stock market from a pattern recognition perspective (to the best of our knowledge, this is the first time that this market is analyzed using such perspective), the second is the proposal of a method for labeling financial time series in order to apply supervised learning techniques.

The rest of the paper is organized as follows. A brief review of related works is presented in Section 2. Some basic financial concepts are described in Section 3. Section 4 gives the experiments and results obtained, as well as the data used. Finally, Section 5 supplies the conclusions and future work to be done.

2 State of the Art

In order to find trading strategies that consistently beat the *Buy and Hold* strategy, several artificial intelligence techniques have been explored, for example, one of the first works using such techniques is [1], in which genetic programming is used in order to create the strategies In this work the U.S. stock market is analyzed using *S&P 500* stock index as a benchmark; they consider transaction costs but do not obtain favorable results.

In [6], chart heuristics are used to detect a chart pattern called *bull flag* They beat *Buy and Hold* strategy but do not consider transaction costs.

In [9], the work from [1] is taken up, omitting transaction costs and considering another way of creating the trees. They obtain positive results in stable and bearish (trending down) markets, but not in bullish (trending up) markets.

The authors in [8] also use genetic programming in order to determine the strategies, being the main difference the use of monthly prices (not daily as is the usual practice) They consider transaction costs and beat *Buy and Hold* strategy.

[7] use *Perceptually Important Points* (PIP) and *Symbolic Aggregate Approximation* (SAX) in order to reduce the dimensionality of the data and express it using symbols. Once they have these symbols they use a genetic algorithm for obtaining the trading strategy. They obtain positive results but do not consider transaction costs.

In [4] an event-based time scale is considered and *directional changes* are defined. Using this concept they are able to generate buy and sell signals that beat *Buy and Hold* strategy even when considering transaction costs and risk adjusted performance.

The authors in [3], propose the use of biclustering mining to discover effective technical trading patterns that contain a combination of indicators from historical financial data series. A modified K nearest neighborhood method is applied to classification of trading days in the testing period. They outperform *Buy and Hold* strategy but do not consider transaction costs.

Finally in [12] , an evolutionary trend reversion model, based on an extension of the XCS (extended classifier system) algorithm, is proposed. They beat *Buy and Hold* strategy and obtain favorable results analyzing several risk-adjusted performance measures. In this work the model obtained is a set of *if..then* rules.

3 Basic Concepts in Finance

3.1 Technical Analysis

According to [5], in its basic form, technical analysis is the study of historical prices and volume from a stock series in order to determine future trends on its price. The basic assumptions for this kind of analysis are:

- Prices are uniquely determined by the interaction between supply and demand.
- Prices move following trends.
- Changes in supply and demand cause trend reversions.
- Changes in supply and demand can be detected using charts.
- Patterns in charts tend to repeat.

Technical analysis also makes the assumption that the information of all factors (including psychological factors such as greed, fear, miss information, etc...) affecting supply and demand curves, is already reflected in the stock's price.

3.2 Efficient Market Hypothesis (EMH)

This hypothesis, proposed by Nobel prize winner Eugene Fama in the 1960's, states that all the observed changes in the prices are caused only by the new available information, that is, historical data (of any kind) has no relevance when determining future trends. In particular, this hypothesis tell us that the use of technical analysis is unprofitable. There are three versions of EMH:

Weak version of EMH In its weak version, the Efficient Market Hypothesis, states that historical prices do not affect future price movements, thus, technical analysis is futile for the generation of trading strategies. This version only refers to historical prices and volume and leaves the door open for other types of data such as financial statements reports or news.

Semi-strong version of EMH In its semi-strong version, the Efficient Market Hypothesis, states that publicly available historical information (prices, financial statements reports, news, etc..) is useless when predicting future price movements. Thus, only private/classified information might be useful for predicting future trends.

Strong version of EMH Finally, in its strong version, the hypothesis states that even private/classified information cannot be used to outperform the market.

3.3 Buy and Hold strategy (BH)

This is the strategy proposed by the Efficient Market Hypothesis, and consists of the following actions:

- Fix a time period $[0, T]$.
- Buy at time 0 at price P_0.
- Sell at time T at price P_T.
- The percetange profit (loss) is given by $\frac{P_T - P_0}{P_0}$.

According to the EMH, the profit (loss) obtained by the BH strategy is the maximum (minimum) profit (loss) that one can obtain in a systematic way. Hence, this strategy will be used as a benchmark when comparing with our proposed algorithms.

3.4 Titles Referenced to Shares

According to Mexican Stock Exchange's website [1], the Mexican market has Titles Referenced to Shares (TRAC's), which are participation certificates representing equity investment trusts. Their primary objective is to replicate the behavior of the stocks or portfolio which they are referred to (underlying), that is, TRAC's are Exchange Traded Funds (ETFs).

The most important TRAC is the one that represents the Mexican Stock Market as a whole and it is called NAFTRAC.

Thus, our objective is finding trading strategies able to beat BH strategy using NAFTRAC data.

4 Experiments and Results

4.1 Datasets

We used daily price data (open price, maximum price, minimum price, adjusted close price) from Yahoo Finance [2] for NAFTRAC for a time period between February 4th 2014 up to April 5th 2018.

It is worth mentioning that this is an unlabeled dataset, hence, we first need to find a way to label it in order to use supervised pattern recognition techniques. The approach taken is based on the idea of "given the historical prices, what should one have had to do in order to make profits?"

[1] https://www.bmv.com.mx/en/markets/instruments
[2] https://finance.yahoo.com/

Training set and test set For obtaining the training and test sets, the data was divided in three-months periods, starting on the day February 4th 2014. Following a three-months training period, there is a three-months test period, which later will become the new three-months training period, that is, we use a rolling window to separate the dataset as shown in the table below.

Table 1. Training and test set separation.

Start training	End training	Start test	End test
2014-02-04	2014-04-30	2014-05-02	2014-07-31
2014-05-02	2014-07-31	2014-08-04	2014-10-31
2014-08-04	2014-10-31	2014-11-03	2015-01-30
2014-11-03	2015-01-30	2015-02-03	2015-04-30

Using the procedure above, we were able to obtain 16 training/test datasets.

4.2 Labeling Process

As mentioned before, we need to label each observation in the training datasets according to one of the three possible actions: buy, sell or hold. To achieve this we used an Estimation Distribution Algorithm (EDA, [11]), namely, we used a version of an Univariate Marginal Distribution Algorithm (UMDA).

This algorithm tries to find the best strategy for the given training period, that is, the strategy that would have generated a bigger (minor) profit (loss) compared to the BH strategy.

Each individual in the population was encoded as a vector, $\mathbf{x}$, representing a trading strategy and having length equal to the number of trading days in the training period. Each entry in the vector takes a value in the set $\{-1, 0, 1\}$, where $-1, 0, 1$; represents a sell, hold and buy signal respectively. Thus the *i-th* component is the decision taken on day i. The algorithm finds the combination maximizing the profit, which is measured as the **Excess Return** over the BH strategy, that is, the return generated by the vector $\mathbf{x}$ minus the return generated by BH.

4.3 Features

For each day, the open, minimum, maximum and adjusted close prices were used as features.

4.4 Results

For every training and test set, the following models were tested[3]:

[3] We chose these models since they are among the most popular ones used in the pattern recognition literature (see [2] and [10] for their mathematical description) .

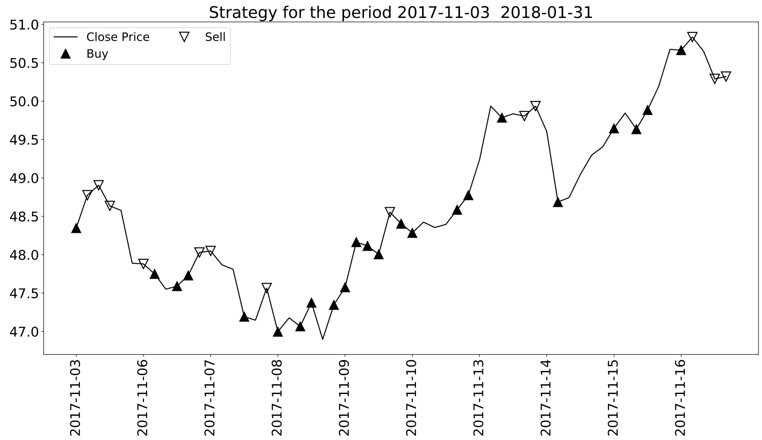

Fig. 1. Results from the labeling process.

- Support Vector Machine with gaussian kernel and assigning class weigths of $0.45, 0.10, 0.45$ for classes $-1, 0, 1$ respectively.
- Multilayer perceptron with a hidden layer with 10 neurons and $ReLu$ as activation function.
- C4.5 tree with max depth of 5.

As described above, the performance measure was **Excess Return** which is calculated using the following assumptions:

- No short-sales allowed, this means that one can only sell if a buy action occurred in the past (we cannot sell something that we do not own).
- Once a trading signal is activated, we have to wait until we see a different one, so no repetitions of the same signal are allowed. This avoids excessive buys or sells.
- Since we are using end of day data, if we have a buy or sell signal on day t, then the buy or sell price (execution price) is the average between the minimum and maximum price at day $t + 1$.
- The cost of every transaction is equal to a 0.25% over the total cost. For example, if a stock was bought at a price of \$10, then the actual monetary amount paid for it is equal to $\$10(1 + 0.0025)$; likewise if a stock was sold at a price of \$10 we end up receiving. a monetary amount of $\$10(1 - 0.0025)$.

For every test set and every model we obtained the following results.

5 Conclusions

As we can observe, among the three models used, the best results were obtained by the Support Vector Machine.

Table 2. Results obtained for every model.

	Excess Return		
Test set	SVM	MLP	C4.5
2014-05-02 / 2014-07-31	-0.013	-0.067	-0.067
2014-08-04 / 2014-10-31	0.0	-0.006	-0.006
2014-11-03 / 2015-01-30	0.0	0.0	0.0
2015-02-03 / 2015-04-30	-0.058	-0.066	-0.07
2015-05-04 / 2015-07-31	0.0	0.0	-0.039
2015-08-03 / 2015-10-30	0.02	0.0	0.025
2015-11-04 / 2016-01-29	0.113	0.054	0.091
2016-02-02 / 2016-04-28	-0.053	-0.025	-0.065
2016-05-02 / 2016-07-29	0.046	-0.025	-0.024
2016-08-01 / 2016-10-31	0.003	-0.027	-0.027
2016-11-01 / 2017-01-31	0.028	0.004	0.01
2017-02-01 / 2017-04-28	-0.008	-0.045	-0.047
2017-05-02 / 2017-07-31	-0.029	-0.029	-0.014
2017-08-01 / 2017-10-31	0.016	0.054	0.045
2017-11-03 / 2018-01-31	-0.033	0.032	-0.06
2018-02-01 / 2018-04-05	0.0	0.0	0.069
Overall sum	0.32	-0.174	-0.179
Average	0.002	-0.01	-0.011
Number of positive excess returns	6	4	5
Number of negative excess returns	6	8	10

Unfortunately, even though when in this model we obtained a positive average excess return, using a t-test for the mean (with a confidence level of 95%) we found that is not possible to reject the null hypothesis (the true mean is zero) thus, statistically speaking, we cannot conclude that this method beats the BH strategy systematically.

Nonetheless its worthwhile mentioning that thanks to the labeling method we can further explore other types of supervised learning techniques whether using a symbolic or sub-symbolic approach.

The future directions for this work might be:

- Explore symbolic approaches for supervised learning, such as Extended Classifier System (XCS).
- Analyze the inclusion of technical indicators.
- Use another set of features.
- Use other sampling frequencies.
- Use risk-adjusted performance measures.

References

1. Allen, F., Karjalainen, R.: Using genetic algorithms to find technical trading rules. Journal of Financial Economics 51(2), 245–271 (1999)

2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, 1st edn. (2006)
3. Huang, Q., Wang, T., Tao, D., Li, X.: Biclustering learning of trading rules. IEEE Transactions on Cybernetics (2015)
4. Kampouridis, M., Otero, F.E.: Evolving trading strategies using directional changes. Expert Systems with Applications 73, 145–160 (2017), http://dx.doi.org/10.1016/j.eswa.2016.12.032
5. Kirkpatrick, C.D., Dahlquist, J.R.: Technical Analysis The Complete Resource for Financial Market Technicians. Pearson Education, Inc., 2nd edn. (2011)
6. Leigh, W., Modani, N., Purvis, R., Roberts, T.: Stock market trading rule discovery using technical charting heuristics. Expert Systems with Applications 23(2), 155–159 (2002)
7. Leitao, J.: Combining rules between PIPs and SAX to identify patterns in financial markets (2016)
8. Lohpetch, D., Corne, D.: Outperforming Buy-and-Hold with Evolved Technical Trading Rules: Daily, Weekly and Monthly Trading (2010)
9. Potvin, J.Y., Soriano, P., Maxime, V.: Generating trading rules on the stock markets with genetic programming. Computers and Operations Research 31(7), 1033–1047 (2004)
10. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, 1st edn. (1993)
11. Simon., D.: Evolutionary Optimization Algorithms. Biologically Inspired and Population-Based Approaches to Computer Intelligence. John Wiley & Sons, 1st edn. (2013)
12. Zhang, X., Hu, Y., Xie, K., Zhang, W., Su, L., Liu, M.: An evolutionary trend reversion model for stock trading rule discovery. Knowledge-Based Systems (2015)