Expert Systems with Applications 37 (2010) 6885–6890

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A method for automatic stock trading combining technical analysis and nearest neighbor classification

Lamartine Almeida Teixeira [a], Adriano Lorena Inácio de Oliveira [b,*]

[a] Department of Computing Systems, University of Pernambuco, Recife, Brazil
[b] Informatics Center, Federal University of Pernambuco, Recife, Brazil

ARTICLE INFO

Keywords:
Stock trend prediction
Financial forecasting
Machine learning
Nearest neighbor prediction

ABSTRACT

In this paper we propose and analyze a novel method for automatic stock trading which combines technical analysis and the nearest neighbor classification. Our first and foremost objective is to study the feasibility of the practical use of an intelligent prediction system exclusively based on the history of daily stock closing prices and volumes. To this end we propose a technique that consists of a combination of a nearest neighbor classifier and some well known tools of technical analysis, namely, stop loss, stop gain and RSI filter. For assessing the potential use of the proposed method in practice we compared the results obtained to the results that would be obtained by adopting a buy-and-hold strategy. The key performance measure in this comparison was profitability. The proposed method was shown to generate considerable higher profits than buy-and-hold for most of the companies, with few buy operations generated and, consequently, minimizing the risk of market exposure.

1. Introduction

Prediction of price movements in the stock market is generally believed to be a very difficult task. A well known hypothesis amongst academics, the Efficient Market Hypothesis (Fama, 1970), suggests that prices immediately reflect all the available information and the only thing that causes security prices to change is new information. Therefore, as the arrival of new information is unpredictable, then prices in the market appear to be randomly generated. Consequently, it would not be possible to consistently earn excess returns using any available information. According to this hypothesis, the best choice in the market would be to follow the buy-and-hold strategy, which consists of not trying to predict the price movements, i.e. buying and never sell the securities.

Several researches debate about the credibility of the Efficient Market Hypothesis. We can cite the study performed by Haugen (1999) and Los (2000). Haugen presents a revision about the deficiencies of this hypothesis, and Los performed experiments considering the six main Asian markets. Los stated that none of these markets presented the efficient behavior suggested by that hypothesis. Besides these works, the practical experience in trading suggests the existence of price trends and that attempts to predict future price changes can generate good returns.

In practice, two approaches are commonly used to analyze price movements, and consequently to predict price trends of securities. The first is the Fundamental Analysis, which utilizes economic factors to estimate the intrinsic values of the securities. The second approach, known as Technical Analysis, is based on the principles of the Dow Theory (Murphy, 1999) and uses the history of prices to predict future movements. The approach used in technical analysis can be formulated as a pattern recognition problem, where the inputs are derived from the history of prices and the output is an estimate of the price or an estimate of the prices trend. Several alternatives to approach this type of problem have been proposed, which range from traditional statistical modeling to methods based on computational intelligence.

Since the 1980s decade many attempts have been made to predict the stock markets. Vanstone and Tan (2003) surveyed the works in this area and classified them in the following topics: time series (Cao & Tay, 2003; Nagarajan, Wu, Liu, & Wang, 2005), pattern recognition and classification (Bao & Yang, 2008; Guo, Liang, & Li, 2007; Leigh, Frohlich, Hornik, Purvis, & Roberts, 2008; Nanni, 2006; Saad, Prokhorov, & Wunsch, 1998; Sai & Yuan, 2007), optimization (Chang & Hsu, 2007; Tan, 1994) and hybrid methods (Afolabi & Olude, 2007; Kim & Shin, 2007; Kwon & Moon, 2007; Mandziuk & Jaruszewicz, 2007). The first work we found in this area was published in 1988 (White, 1988) and used feedforward neural networks to detect regularities in prices time series. Another interesting work, which we consider our main reference for this paper, was the recent research performed by Kwon and Moon (2007), which uses a hybrid approach based on the context of the

* Corresponding author.
E-mail addresses: lat@dsc.upe.br (L.A. Teixeira), alio@cin.ufpe.br (A.L.I. de Oliveira).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.03.033

test day. The main goal of that work was to predict future price changes by using technical indicators as inputs. This prediction was based on regression with recurrent neural networks, whose weights were optimized by a combination of genetic algorithms and the traditional backpropagation. The method was tested with 36 stocks, considering a period of 13 years, and was capable of generating considerable higher profits than the buy-and-hold strategy. One drawback of the method by Kwon and Moon is that the classifier employed by them is complex and is trained by genetic algorithms and backpropagation, which are known to be slow methods for training neural networks (Kwon & Moon, 2007). To use their methods in practice for a large number of stocks would require large computational power.

The researches in this area are highly diverse regarding the approaches used to model the problem and the metrics used to evaluate the techniques. However, for most of these works it is not clear if the techniques proposed would be capable of generating profits in practice. Therefore our general purpose in this paper is to present a price trend prediction technique and evaluate it according to its capability to generate profits. In this paper, our focus is predicting price trends in the stock market by using some common tools of technical analysis and the well known *k*-NN algorithm (Cover & Hart, 1967). Another aim of the research described here was to propose a method that uses a classifier simpler than that of Kwon and Moon (2007). This would enable the system to be used to analyze more stocks for trading in a given day. It is important to emphasize that these common tools of technical analysis, like technical indicators, stop loss, stop gain and RSI filters, are commonly used by traders in practice, but they are usually ignored in academic research. This has motivated us to propose a method that puts together a well known classifier and some tools that are commonly used in practical trading, with the purpose of investigating the feasibility of using an intelligent trading system exclusively based on data derived from security prices and volume of trades. To evaluate the capability of the proposed system in generating profits we compare its results to the profits that would be obtained by using a buy-and-hold strategy, which is the best alternative according to the Efficient Market Hypothesis.

There are a number of classifiers that are known to achieve better results in terms of generalization than *k*-NN for many real problems. However, in this first moment we wanted a simple and fast algorithm, since our purpose now is to analyze the influence in the results of some technical analysis tools, like the stops and the RSI filter.

The paper is organized as follows. The basic concepts of technical analysis are presented in Section 2, while Section 3 presents the proposed method, including a brief revision of the *k*-NN algorithm, the data sets and the input features, as well as the trading model considered. In Section 4 we perform some experiments in order to assess the economic significance of the method. In Section 5 we provide some concluding remarks.

2. Basic concepts of technical analysis

According to Murphy (1999), technical analysis is the study of market action, primarily using charts, aiming to forecast future price trends. The technician observes the market action via two main variables, the prices and the volume of trades. The most important premise of this type of analysis is that the market action discounts everything. It means the technician believes that anything that can possibly affect the market is already reflected in the prices, as well as that all the new information will be immediately reflected in those prices. As a consequence, all the technician needs is to analyze the history of prices.

The main tools of the technical analysis are the volume and price charts. Based on the information of prices and volume the technical indicators (Murphy, 1999) are built. Technical indicators are mathematical formulas that are applied to the price or volume data of a security for modeling some aspect of the movement of those prices. Some indicators, for example, use only closing prices, while others incorporate the volume of trades into their formulas. The most used indicators are those based on moving averages and the oscillators. Those based on moving averages are used to smooth the price movements and, consequently, to make the identification of trends easier. On the other hand, the oscillators are indicators usually employed for identifying momentum.

Initially we have selected four indicators to use in this work. Moving averages, Relative Strength Index (RSI) and stochastics are also considered in the work in Kwon and Moon (2007), while the Bollinger bands are not considered. There is a high diversity of technical indicators in the market. The choice of the technical indicators used in our work was made considering: (i) selecting some of the most widely used by traders in practice and (ii) selecting indicators of different categories, some based on moving averages and some oscillators. Thus it would be interesting to analyze their relevance in practice.

2.1. Moving averages

We used in this work the Simple Moving Average (SMA), which is formed by computing the average price of a security over a specified number of periods. The calculation is repeated for each day on the chart and then joined to form a smooth curve, which makes easier to identify the direction of the trend. The SMA formula for a period of n is presented in (1)

$$SMA(t) = \frac{1}{n}\sum_{i=t-n}^{t} x(i) \tag{1}$$

2.2. Relative strength index (RSI)

The RSI is one of the most popular momentum oscillators. It compares the magnitude of the recent gains to the magnitude of the recent losses and generates a number that ranges from 0 to 100. The technicians say that generally when the RSI rises above 30 it is considered a bullish signal, i.e. it is considered a signal that prices trend to rise. Conversely, when it falls below 70 it is considered a bearish signal, i.e. a signal that prices are likely to fall. The RSI is computed using (2)–(4)

$$RSI = 100 - \frac{100}{1+RS}, \quad \text{where } RS = \frac{avgGain}{avgLost} \tag{2}$$

$$avgGain = (\text{total of gains during past } n \text{ periods})/n \tag{3}$$

$$avgLost = (\text{total of losses during past } n \text{ periods})/n \tag{4}$$

2.3. Stochastics

The stochastics oscillator compares the location of the current close price of a security relative to its price range over a certain number of periods. This indicator is composed by two lines, the fast (*%K*) and the slow (*%D*), which are calculated according to (5) and (6). According to Murphy (1999), closing levels that are consistently near the top of the range indicate accumulation (buying pressure) and those near the bottom of the range indicate distribution (selling pressure)