Mining for Profitable Patterns in the Stock Market

Yihua Philip Sheng
Southern Illinois University, USA

Wen-Chi Hou
Southern Illinois University, USA

Zhong Chen
Shanghai JiaoTong University, PR China

INTRODUCTION

The stock market, like other economic phenomena, is a very complex system. Many factors, such as company news, interest rates, macro economic data, and investors' hopes and fears, all affect its behavior (Pring, 1991; Sharpe, Alexander, & Bailey, 1999). Investors have longed for tools and algorithms to analyze and predict stock market movement. In this study, we combine a financial theory, the *market efficiency theory*, and a data mining technique to explore profitable trading patterns in the stock market. To observe the price oscillation of several consecutive trading days, we examine the *K-lines*, each of which represents a stock's one-day movement. We will use a data mining technique with a heuristic rating algorithm to mine for reliable patterns indicating price rise or fall in the near future.

BACKGROUND

Methods of Stock Technical Analysis

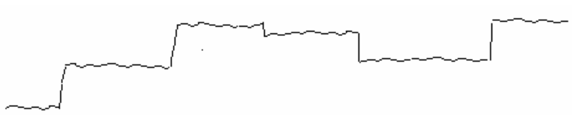
Conventional stock market technical analysis is often done through visually identifying patterns or indicators on the stock price and volume charts. Indicators like moving averages and support and resistance level are easy to implement algorithmatically. Patterns like head-and-shoulder, inverse head-and-shoulder, broadening tops and bottoms, and etcetera, are easy for human to visually identify but difficult for computers to recognize. For such patterns, methods like smoothing estimators and kernel regression can be applied to increase their machine-readability (Dawson & Steely, 2003; Lo, Mamaysky, & Wang, 2000).

The advances in data mining technology have pushed the technical analysis of stock market from simple indicators, visually recognizable patterns, and linear statistical models to more complicated nonlinear models. A great deal of research has focused on the applications of artificial intelligence (AI) algorithms, such as artificial *neural networks* (ANNs) and *genetic algorithm* (e.g., Allen & Karjalainen, 1999; Chenoweth, Obradovic, & Stephenlee, 1996; Thawornwong, Enke, & Dagli, 2003). ANNs equipped with effective learning algorithms can use different kinds of inputs, handle noisy data, and identify highly nonlinear models. Genetic algorithms constitute a class of search, adaptation, and optimization techniques to emulate the principles of natural evolution. More recent studies tend to embrace multiple AI techniques in one approach. Tsaih, Hsu, & Lai (1998) integrated the rule-based systems technique and the neural networks technique to predict the direction of daily price changes in S&P 500 stock index futures. Armano, Murru, & Roli (2002) employed ANNs with a genetic algorithm to predict the Italian stock market. *Fuzzy logic*, a relatively newer AI algorithm, has also been used in stock market prediction literature (e.g., Dourra & Siy, 2001). The increasing popularity of fuzzy logic is due to its simplicity in constructing the models and less computation load. Fuzzy algorithms provide a fairly straightforward translation of the qualitative/linguistic statements of rules.

Market Efficiency Theory

According to the market efficiency theory, a stock's price is a full reflection of market information about that stock (Fama, 1991; Malkiel, 1996). Therefore, if there is information out on the market about a stock, the stock's price will adjust accordingly. Interestingly, evidence shows that price adjustment in response to news usually does not settle down in one day; it actually takes some time for the whole market to digest the news. If the stock's price really adjusted to relevant events in a timely manner, the stock price chart would have looked more like what *Figure 1* shows.

Figure 1. Ideal stock price movement curve under the market efficiency theory



The flat periods indicate that there were no events occurring during the periods while the sharp edges indicate sudden stock price movements in response to event announcements. However, in reality, most stocks' daily price resembles the curve shown in *Figure 2*. As the figure shows, there is no obvious flat period for a stock and the stock price seemed to keep on changing. In some cases, the stock price continuously moves down or up for a relatively long period, for example, the period of May 17, 2002 to July 2, 2002, and the period of October 16, 2002 to November 6, 2002. This could be either there were negative (or positive) events for the company every day for a long period of time or the stock price adjustment to events actually spans a period of time, rather than instantly. The latter means that stock price adjustment to the event announcements is not efficient and the semi-form of the market efficiency theory does not hold. Furthermore, we think the first few days' price adjustments of the stock are crucial, and the price movements in these early days might contain enough information to predict whether the rest of price adjustment in the near future is upwards or downwards.

Knowledge Representation

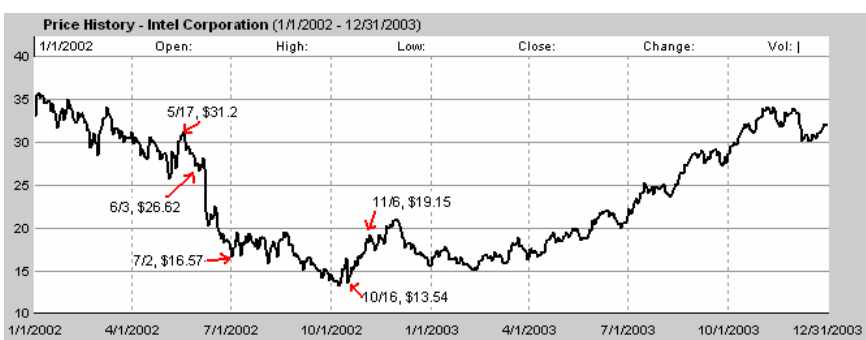
Knowledge representation holds the key to the success of data mining. A good knowledge representation should be able to include all possible phenomena of a problem domain without complicating it (Liu, 1998). Here, we use K-lines, a widely used representation method of the daily stock price in Asian stock markets, to describe the daily price change of a stock. *Figure 3* is examples of K-Lines.

Figure 3(a) is a price-up K-Line, denoted by an empty rectangle, indicating the closing price is higher than the opening price. *Figure 3(b)* is a price-down K-line, denoted by a solid rectangle, indicating the closing price is lower than the opening price. *Figure 3(c)* and *3(d)* are 3-day K-Lines. *Figure 3(c)* shows that the price was up for two consecutive days and the second day's opening price continued on the first day's closing price. This indicates that the news was very positive. The price came down a little bit on the third day, which might be due to the correction to the over-valuation of the good news in the prior two days. Actually, the long price shadow above the closing price of the second day already shows some degree of price correction. *Figure 3(d)* is the opposite of *Figure 3(c)*. When an event about a stock happens, such as rumors on merger/acquisition, or change of dividend policies, the price adjustments might last for several days till the price finally settles down. As a result, the stock's price might keep rising or falling or stay the same during the price adjustment period.

A stock has a K-Line for every trading day, but not every K-Line is of our interest. Our goal is to identify a stock's K-Line patterns that reflect investors' reactions to market events such as the releases of good or bad corporate news, major stock analysts' upgrade on the stock, and etcetera. Such market events usually can cause the stock's price to *oscillate* for a period of time. Certainly, a stock's price sometimes might change with large magnitude just for one day or two due to transient market rumors. These types of price oscillations are regarded as market noises and therefore are ignored.

Whether a stock's daily price oscillates is determined by examining if the price change on that day is greater than the average price change of the year. If a stock's price oscillates for at least three consecutive days, we regard it as a signal of the occurrence of a market event. The market's response to the event is recorded in a 3-day K-Line pattern. Then, we examine whether this pattern is followed by an up or down trend of the stock's price a few days later.

Figure 2. The daily stock price curve of Intel Corporation (NasdaqNM Symbol: INTC)



Related Content

Improving Classification Accuracy of Decision Trees for Different Abstraction Levels of Data
Mina Jeongand Doheon Lee (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 1103-1115).*
www.irma-international.org/chapter/improving-classification-accuracy-decision-trees/7689

Intelligent Querying Techniques for Sensor Data Fusion
Shi-Kuo Chang, Gennaro Costagliola, Erland Jungertand Karin Camara (2010). *Intelligent Techniques for Warehousing and Mining Sensor Network Data (pp. 213-233).*
www.irma-international.org/chapter/intelligent-querying-techniques-sensor-data/39547

Privacy in Data Mining Textbooks
James Lawlerand John C. Molluzzo (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 2872-2879).*
www.irma-international.org/chapter/privacy-data-mining-textbooks/7808

Best Practices in Data Warehousing from the Federal Perspective
Les Pang (2005). *Encyclopedia of Data Warehousing and Mining (pp. 94-99).*
www.irma-international.org/chapter/best-practices-data-warehousing-federal/10573

Use of RFID in Supply Chain Data Processing
Jan Owens, Suresh Chalasaniand Jayavel Sounderpandian (2005). *Encyclopedia of Data Warehousing and Mining (pp. 1160-1165).*
www.irma-international.org/chapter/use-rfid-supply-chain-data/10772