Expert Systems with Applications 38 (2011) 14846–14851

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction

Tsung-Sheng Chang *

Department of Information Management, National Chung Cheng University, Min-Hsiung, Chia-Yi 621, Taiwan, ROC

ARTICLE INFO

Keywords:
Artificial neural networks (ANN)
C&RT
Decision tree
Stock price forecasting

ABSTRACT

Precise prediction of stock prices is difficult chiefly because of the many intervening factors. Unpredictability is particularly notable in the aftermath of the global financial crisis. Data mining may however be used to discover highly correlated estimation models. This study looks at artificial neural networks (ANN), decision trees and the hybrid model of ANN and decision trees (hybrid model), the three common algorithm methods used for numerical analysis, to forecast stock prices. The author compared the stock price forecasting models derived from the three methods, and applied the models on 10 different stocks in 320 data sets in an empirical forecast. Average accuracy of ANN is 15.31%, the highest, in terms of match with real market stock prices, followed by decision trees, at 14.06%; hybrid model is 13.75%. The study also discovers that compared to the other two methods, ANN is a more stable method for predicting stock prices in the volatile post-crisis stock market.

1. Introduction

Since mid-2008, when the collapse of Lehman Brothers led to global economic repercussions, the stock market has been hard hit. Stock indices fell, and economies went into recession. More than one year later, stock indices have witnessed sharp fluctuations, especially in new emerging markets. In addition, globalization has made forecasting of stock prices increasingly difficult (Albuquerque, Francisco, & Marques, 2008; Stock & Watson, 2007). However, we need to know if the study samples of this period and the results of the forecasts of previous researchers are in line with our expectations. In addition, among Asian emerging markets, China's economic growth has been the engine spurring on the development of stock markets in the region; hence the greater volatility of stock prices in these markets. Thus, we need to attach greater importance to emerging stock markets (Dutta, Jha, Laha, & Mohan, 2006; Fidrmuc & Korhonen, 2009). The Taiwan stock market, which has close inter-connection with Mainland China, is another example worth observing (Lai, Fan, Huang, & Chang, 2009; Lin & Yeh, 2009).

Many algorithm methods are used to predict stock prices. Examples are the artificial neural networks (ANN) (Desai & Bharati, 2007; Kim & Shin, 2007; Pino, Parreno, Gomez, & Priore, 2008; Zhu, Wang, Xu, & Li, 2008), Fuzzy (Khashei, Hejazi, & Bijari, 2008; Lee & Kim, 2007), or other statistical or forecasting methods (Chen, Gou, Guo, & Gao, 2008; Hu & He, 2007; Ince & Trafalis, 2008). All these methods attempt to predict stock prices under different market and economic conditions, and of them, ANN has produced rather good outcomes and has been the favourite method for many.

Ou and Wang (2009), and Lai et al. (2009) believe that the decision tree (DT) method is good for forecasting stock prices. Levin and Zahavi (2001) found that problem-correlation using DT is much clearer than traditional methods. In fact, DT is a very good forecasting method. The Bayes' Theorem may be used as a basis for scientific forecasts. However, DT studies have focused on commercial activities (Aitkenhead, 2008; Reyck, Degraeve, & Vandenborre, 2008). In recent years, there has been a lack of research in the prediction of stock prices using both DT and ANN, and comparing the results of one method against the other. Most ANN studies have focused on its evolution and improvement (Ihme, Marsden, & Pitsch, 2008; Lin & Yeh, 2009; Paliwal & Kumar, 2009), and integrated fuzzy models on the forecasting of stock prices (Khashei et al., 2008; Lai et al., 2009). This study aims to fill the research gap by adopting a broader approach through in-depth empirical studies.

The study adopts a hybrid model, using ANN and DT as the foundation, to forecast stock prices. Try to find out if this model produces better forecasts of stock prices, compared to the earlier two methods. Hence, the results of forecasted stock prices using the three abovementioned methods (ANN, DT and hybrid model) are compared against each other to find out the differences. In doing so, we can see if the results of our forecasts match our expectations, and discover the most stable model.

* Tel.: +886 973 519943.
E-mail address: china@mis.ccu.edu.tw

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.05.063

2. Digital game content stocks

Taiwan began promoting the digital content industry around the year 2000. Taiwan consumers were receptive to digital games. In addition some companies were early investors in Mainland China. The absence of a language barrier allowed them to have higher overseas revenue; hence, good performance in stock prices. However, no stock price index is available for reference for this stock cluster. Instead, they are classified as "information software stocks".

A stock price index comprises prices of stocks of companies in similar businesses; hence, price movements of stocks tend to affect one another. And investors can determine fluctuations against each other of stock prices of companies in relevant clusters from the key constituent indices. However, an industry stock index may not be available for some industries, especially in emerging stock markets, because of their lack of maturity, and few listings. As a result, investors' can rely only on subjective norms and personal perception, and are unable to achieve better forecasting of stock prices (Hong, Torous, & Valkanov, 2007). Hence, to have a better grasp of digital game content stocks in Taiwan, we need to observe the business composition of a company, and based on prior expert consultations, determine if a company is indeed in the business of digital game content.

There is no precise definition for the term "digital game content stocks". But the term "digital game stocks" defined in the Taiwan stock market basically includes manufacturing of games spare parts and components, and online games. Manufacturing of games spare parts and components does not derive its main income from "game content"; instead, income is generated mainly from the production of parts and components of console games. Examples of OEM manufacturers of console games are PixArt Imaging, Delta Group, and Genius. However, digital game stocks of this nature that are mainly stocks of OEM manufacturers are not included in this study. On the other hand, as summer vacation approaches every year, most investors would think of digital content game stocks along the lines of online games, PC games or arcade games, most of which are either information software or actually contain designed game content. These are the digital game content companies whose stocks will be studied herein.

3. Literature review

The study uses the data mining method to derive actual forecasted stock prices. Data mining is concerned with the development and applications of algorithms for discovery of a priori unknown relationships. Han and Kamber (2006, p. 7) defines data mining as "the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories". Berry and Linoff (1997) point out that data mining is the exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules, and to establish effective models and rules. Data mining is currently widely used in many applications, as users attempt to discover patterns or predict the future using historical data.

The data mining methods used in this study are mainly ANN and DT in order to derive the most relevant results. As for the hybrid model, it is developed by integrating the results from ANN and DT analyses, to deduce a new model.

3.1. Artificial neural networks

Artificial neural networks, or ANN, is a popular prediction tool. It is a technique that simulates the learning process of biological neural networks, by developing models from extremely complex and non-linear formulae. By using different variables and assumed parameters, it trains a neural network to perform better analysis and predictions.

ANN has been fruitful in predicting stock prices. Antonio, Claudio, Manuel, and Nelson (1996) was 63.3% accurate in their prediction of the range of rises of San Diego's stock market, and 74.7% accurate in falls. When Steiner and Wittkemper (1997) predicted the European stock market between 1991 and 1997, the accuracy of their predictions was generally efficient. When Shachmurove and Witkowska (2000) compared results using the Ordinary Least Squares method (OLS) and ANN, they found ANN a better prediction tool. Dutta et al. (2006) predicted the Indian stock market with ANN, and found that the root mean square error (RMSE) and mean absolute error (MAE) are in line with the expected smallness of error. Zhu et al. (2008) forecasted the NASDAQ, DJIA and STI indices, and found the ANN model to have good prediction performance.

Based on previous papers written, there have been good researches and recommendations on the use of ANN as a prediction method for stock prices. Among the ANN learning models, the back-propagation neural network (BPN) is most popular and widely used. Most users of BPN would have to carry out multiple tests to arrive at a better model. Hence, there are many methods that help optimize the network and determine parameters. Multiple, Prune and *K*-means are highly accurate methods with fast responses, and are often used to train the ANN. Thus, this study will use BPN as a prediction model of stock prices.

3.2. Decision tree

Modeled after the structure of a tree, DTs are able to provide a good explanation applicable to the prediction of stock prices, and interpret problems very much according to the principles of mathematical and statistical principles (Brida & Risso, 2010). DT is a fairly mature technique which includes models such as C5.0, C&RT, CHAID, QUEST and ID3, the difference of which is in the derivations of formula, such as entropy and information gain, to determine the attributes that result in splitting. However, most decision trees consist of many nodes, which, under certain circumstances, hinder analysis or interpretation of information (Aitkenhead, 2008; Ture, Tokatli, & Kurt, 2009). The classification and regression trees (CART or C&RT) method of Breiman, Friedman, Olshen, and Stone (1984) generates binary decision trees. In the real world, the chances of biased binary outcomes are few, but the binary method allows for easy interpretation and analysis (Ture et al., 2009; Yang et al., 2003). Thus, the study uses the binary method in the DT prediction of stock prices, and attempts to deduce a correlated prediction factor.

4. Experiment design

4.1. Data collection and methods

By using digital game content stocks in Taiwan as the sample. In terms of the choice of digital game stocks, the key words "game stock" were keyed into the UDNDATA online database to search for information between 1 January 2008 and 31 May 2009, and game stocks as described in the database were extracted. Through pre-testing, we interviewed managers of securities companies and investors who traded more than US$50,000 in May 2009. Hence, every game stock was verified by a total of 30 experts and investors. We then compiled and screened the results, and selected 10 game stocks as shown in Table 1.

The 10 game stocks are traded in the OTC Exchange of Taiwan, and have fairly similar IPO backgrounds. Daily closing stock prices of each of the 10 stocks were gathered from the Gre Tai Securities Market (GTSM) database. Key information collected includes: daily