Time Series Financial Data Mining

Chiu-Che Tseng, Ching-Tsai Kang

Department of Computer Science and Information Systems
Texas A&M University – Commerce, TX 75429, U.S.A.

Abstract

This study examines the stock price effects of cross-listings ADRs by 8 Taiwanese companies during the period 1996 to 2003. After analyzing the numerical information, the result is going to be compared with those in 2004 to estimate the accuracy of prediction and sees if there is any positive co-relation between the stock prices in those two countries. In the study, we use decision tree and rule base system which is different from the traditional statistical methodology, which has been used in a fairly extensive empirical researches, to examine stock price information.

Keywords: Decision Tree, Neural Networks, Hybrid model.

1. Introduction

There has been a dramatic increase in the trading of foreign stocks as investors recognize the need for international diversification and as foreign companies seek to broaden their shareholder base and raise capital. As a result, the number of American depositary receipts (ADR) listings on U.S. exchanges has also risen sharply. Though corporations view cross-listings as value enhancing, the changes in liquidity and volatility, and the cost of training associated with order flow migration following cross-listing may affect the quality of the domestic equity market.

For those, there is an essential issue in this international portfolio investment that comes up , which is that is there any correlation between those ADRs in the international stock markets. Are those international stock markets interdependent?

Neural network, which is modeled on the human brain in which there are extensively interconnected units (neurouns) that make up a vast network capable of complex patter recognition. As such, it is composed of many computational elements operating in parallel and arranged in patterns reminiscent of biological neural nets. The purpose of such emulation is to produce artificial systems capable of intelligent computation and pattern recognition similar to those that the human brain routinely performs.

The decision tree approach in this study is based on the C5.0 implementation of SPSS' Clementine[8]. The C5.0 decision tree learning algorithm is a commercial decision tree and rule induction engine developed by Ross Quinlan[17,20]. It is the state-of-the-art successor of the widely used C4.5 decision tree algorithm[20]. In contrast to other decision tree algorithms such as CART[3], C5.0 is able to generate trees with a varying number of branches per node. Decision trees based on C5.0 algorithm provide a clear indication of which attributes are important for the classification task at hand.

Since the trading hours of US markets do not coincide with Taiwanese markets, in this study, we apply neural network, decision tree and rule-based to analyze the stock price variances of ADRs in the US and those in Taiwanese market and see if the ADR listed in the US market really reflect the real-time information that became available while the US market was open right after the Taiwanese market was closed.

2. Relative Works

Among numerous empirical researches, the co-relationship between international stock prices has always been discussed. Jayaraman et al. [10] show ADR listing to be associated with both positive abnormal returns on the listing day and an increase in the volatility of returns to the underlying stock. Foerster and Karolyi find that their sample of non-US firms cross-listing on US exchanges, over the period 1976 to 1992, experienced average excess returns of 19% during the year before listing, 1.2% the listing week, and -- 14% the year following listing.

Moreover, Jiang [11] uses weekly data, over the sample period January 1980 to September 1994, on ADRs and market indices to conduct co-integration tests and to estimate EC and multifactor models. The study's findings shows that, most of the time, ADRs and the home markets are interrelated and do influence each other. As a result, the inter-relationship among international markets does exist.

Even so, there are not many empirical

researches which discuss the issue of whether the cross-listed ADR has any influence on the stock issued in the home country. Whether there is any positive movement between the return of ADRs and that issued in the home country. In the following, we are going to apply decision tree and do adverse analysis to make contrasts with above researches.

3. Training Methods

3.1 Neural Network

Neural networks are software models inspired by biological neural networks. Neural networks are inherently nonlinear. They estimate nonlinear functions well and extract residual nonlinear elements. The networks can at least partially transform the input data if needed. These overcome the limitations of models such as linear regression. Neural network consist of basic units, termed neurons, whose design is suggested by their biological counterparts. These artificial neurons have input paths just as biological neurons have dendrites; they have output paths just as biological neurons have axons. Both artificial and biological neurons also have predispositions which effect the strength of their output. The neuron combines the inputs signals. In both real and artificial neurons, learning occurs and alters the neurons.

In neural network's, the neuron input path i has a signal on it (X_i), and the strength of the path is characterized by a weight (w_i). The neuron forms a sum of the path weight times the input signal over all paths and the node bias (Q). The output (Y) is usually a sigmoid shaped logistic function of the sum. Mathematically, the sum is expressed as:

$$\text{sum} = \Sigma w_i x_i + Q$$

And it is transformed into the output Y with the sigmoid-shaped logistic function:

$$Y = 1/(1 + e^{-\text{sum}})$$

Sigmoid-shaped curve approaches a minimum and maximum value at the asymptotes. It is common for this curve to be called a sigmoid when it ranges between 0 and 1, and a hyperbolic tangent when it ranges between -1 and 1. Mathematically, the exciting feature of these curves is that both the function and its derivatives are continuous. This option works fairly well and is often the transfer function of choice.

By the learning process, random valued parameters (weights and bias) of a neural network are adapted through a continuous process of simulation by the environment in which network is embedded.

A neural network is a powerful modeling tool that is able to capture and represent complex relationships. It can compare existing stock-trading patterns with previous situations, analyze all kind of indicators and eventually "learn" what works and what doesn't as the program digests more data. The true power and advantage of neural networks lies in their ability to digest a huge amount of data, find both linear and non-linear relationships from trading patterns, and make deep analytics that can never be accomplished by human analysts at the same time.

3.2 Decision Tree Algorithm

We chose to use decision trees because they provide a comprehensible representation of their classification decisions. Although techniques such as boosting [5, 19] or support vector machines might obtain slightly higher classification accuracy, they require more computation during classification and they further obscure the decision making process.

We chose to use the C5.0 decision tree algorithm[17] a widely used and tested implementation. For details regarding the specifics of C5.0 the reader is referred to[17, 18].

4. Training Model

Our study is designed to estimate the accuracy of the prediction. We first normalize the data used in the study and describe the supervised learning algorithm we chose. Then we train the data by BrainMaker neural network and C 5.0 decision tree classifier and rulesets classifier. After that, we use the data of 2004 to test and estimate the accuracy of prediction.

We used daily data during June of 1996 to 2004 8 Taiwanese companies which have cross-listings of their stocks in the US stock market as American Deposit Receipt (ADR) is being applied and these Taiwanese companies daily data of Taiwanese stock market as sample. The data during 1996 to 2003 will be trained and constructed by C 5.0 decision tree classifier and rulesets classifier. Then they will use the decision tree and rulesets to run the data of 2004. The result would be used to compare with existed data of Taiwanese stock variation or ADR variation in 2004.

5. Summary and Conclusions

The predicting which ranges from January 2, 2004 to June 16, 2004 is based on the stock

price variation of Taiwanese's stocks and their ADRs. Each company has 115 trading days. The accuracy rate shows the moving trend between the predicted data and the real data.

When the network is trained to the satisfaction, the current input data and the network will make a prediction of output (the variation of stock price). The following is the accuracy rate of prediction acquired from training.

5.1 The Compare Of The Three Methods

In this section, we want to compare the result of the three artificial intelligence methods that we used in this paper.

TW predicts ADR	Rule Base	Decision Tree	Neural Network
MXICY	55.26%	60.53%	62.28%
ASTSF	50.43%	49.57%	47.83%
TSM	54.39%	47.37%	53.51%
ASX	62.28%	61.40%	38.60%
UMC	56.14%	53.51%	55.26%
SPIL	50.00%	50.88%	57.89%
AUO	64.35%	54.78%	60.87%
CHT	61.74%	60.00%	51.75%

Table 1 The result is acquired by training the data using rule base, decision tree and neural network of each company.

Table 1 shows the predicting accuracy rate by using each company's historical data.

ADR predicts TW	Rule Base	Decision Tree	Neural Network
MXICY	48.62%	44.95%	44.74%
ASTSF	47.71%	50.46%	43.86%
TSM	44.04%	45.87%	48.25%
ASX	54.13%	49.54%	39.47%
UMC	47.22%	37.96%	34.21%
SPIL	43.12%	47.71%	50.00%
AUO	47.71%	46.79%	40.53%
CHT	39.45%	44.04%	39.47%

Table 2 The result is acquired by training the data using rule base, decision tree and neural network of each company.

Table 2 shows the predicting accuracy rate by using each company's historical data.

5.2 Combining Prediction

	TW predicts ADR	ADR predicts TW
MXICY	41.74%	22.02%
ASTSF	25.56%	27.52%
TSM	25.43%	25.69%
ASX	20.18%	18.35%
UMC	35.09%	8.26%
SPIL	23.68%	29.36%
AUO	33.04%	33.95%
CHT	27.83%	32.11%

Table 3 The result is acquired by combining the prediction of decision tree and neural network.

Table 3 shows the predicting accuracy rate by using each eight companies' historical data and combining the prediction of decision tree algorithm and neural network algorithm.

5.3 Conclusion

As the result, the combined hybrid approach does not work in this data set. It produced a lower accuracy than any single algorithm we used in this study. Overall, we believe that the ADRs co-related with stock prices, especially using Taiwanese stock price to predict ADR's.

6. References

[1] Anant K., S., Dennis E., L. (1st Qtr., 1996). *Valuation Effects of Foreign Company Listings on U.S. Exchanges*. Journal of International Business Studies, 27(1), p.67-88.

[2] Bennett K. and Campbell C. (2000). *Support Vector Machines:Hype or Hallelujah?* SIGKDD Explorations, 2:1–13, 2000.

[3] Berry, M.J. and Linoff, G(1997*). Data Mining Techniques For Marketing, Sales and Customer*

Support, John Wiley & Sons, Inc., New York, 1997.
[4] David E., Mehdi, S. (2001). *American depositary receipts: An analysis of international stock price movements*. International Review of Financial Analysis, 10, p.323-363.
[5] Freund Y. (1995). *Boosting a Weak Learning Algorithm by Majority*. Information and Computation, 121(2):256–285, 1995.
[6] Geng, C., Man, L., W. *Implementing neural networks for decision support in direct marketing.*International Journal of Market Research. 46(2).
[7] Granzow M.., Berrar D.., Dubitzky W.., Schuster A.., Azuaje F.J.., Eils R. (2001) *Tumor Classification by Gene Expression Profiling:Comparison and Validation of Five Clustering Methods*, ACM SIGBIO Newsletter,Volume 21 , Issue 1 (April 2001), 16 - 22
[8] Huntsberger, T.L. and Aijimarangsee P(1992). *Parallel selforganising feature maps for unsupervised pattern recognition*. Bezdek J.C. and Pal N.R, Fuzzy models for pattern recognition, 483-495. IEEE Press, New York. 1992
[9] James P. Early, Carla E. Brodley, Catherine Rosenberg(2003), *Behavioral Authentication of Server Flows*, 19th Annual Computer Security Applications Conference, Las Vegas, Nevada, December 8-12, 2003
[10] Jayaraman, Shastri and Tandon (1993*). The impact of international cross-listings on risk and return: the evidence from ADRs.* Journal of Banking and Finance. 91-103.
[11] Jiang C. (1998)*"Diversification with ADRs: The Dynamics and the Pricing Factors."* Journal of Business Finance and Accounting, 25, (1998), pp. 683–700.
[12] Karolyi (1996). *What happens to stocks that list shares abroad?* A Survey of the evidence and its managerial implications. University of Western Ontario Working paper.
[13] Karolyi, G. Andrew, and René M. Stulz, (1996), *"Why do Markets Move Together?, An Investigation of U.S.-Japan Stock Return Co-movements,"* Journal of Finance 51, 951-986.
[14] Lan D., Jack G., Ananth M(1998*)., International Cross-Listing and Order Flow Migration : Evidence From An Emerging Market*, The Journal of Finance.vol LIII,No. 6, Dec 1998
[15] Mark E., W., William J., Crowder. (1998). *COINTEGRATION, FORECASTING AND INTERNATIONAL STOCK PRICES*. Global Finance Journal. 9(2).
[16] Niklas, A., Jan, A. (2002). *Testing for cointegration between international stock prices*. Applied Financial Economics. 12, p.851-861.
[17] Quinlan. J. R. *C4.5: Programs for Machine Learning*.Morgan Kaufmann, San Mateo, CA, 1993.
[18] Ross Quinlan. Data Mining Tools See5 and C5.0. URL `http://www.rulequest.com/see5-info.html`.
[19] Robert E. Schapire. *A Brief Introduction to Boosting*. In IJCAI, pages 1401–1406, 1999. URL citeseer.
nj.nec.com/schapire99brief.html.
[20] Witten, I.H., and Frank, E(1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Pub., San Francisco, 1999.