

Stock Trend Analysis and Trading Strategy

Hongxing He¹

Jie Chen¹

Huidong Jin¹

Shuheng Chen²

¹CSIRO Mathematical and Information Sciences, GPO Box 664, Canberra, ACT 2601, Australia

²AIECON Research Center, Department of Economics, National Chengchi University, Taipei, Taiwan 11623

Abstract

This paper outlines a data mining approach to analysis and prediction of the trend of stock prices. The approach consists of three steps, namely partitioning, analysis and prediction. A modification of the commonly used k -means clustering algorithm is used to partition stock price time series data. After data partition, linear regression is used to analyse the trend within each cluster. The results of the linear regression are then used for trend prediction for windowed time series data. The approach is efficient and effective at predicting forward trends of stock prices. Using our trend prediction methodology, we propose a trading strategy TTP (Trading based on Trend Prediction). Some preliminary results of applying TTP to stock trading are reported.

Keywords: Data Mining, Clustering, k -means, Time Series, Stock Trading

1 Introduction

Trend analysis and prediction play a vital role in practical stock trading. Experienced stock traders can often predict the future trend of a stock's price based on their observations of the performance of the stock in the past. An early sign of a familiar pattern may alert a domain expert to what is likely to happen in the near future. They can then formulate their trading strategy accordingly.

The search for and matching of similar patterns have been studied extensively on time series analysis [1, 2]. Patterns in long time series data repeat themselves due to seasonality or other unknown underlying reasons. Early detection of patterns similar to those that have occurred in the past can readily provide information on what will follow. This informa-

tion will be able to help decision-making on the trading strategy in stock market trading practice. In this paper, we report an approach to predict the trend of the stock prices and apply it to stock trading practice.

2 Time Series Data Preparation

We create training data by sliding a fixed-length time window from time t_b to t_e . The following $N = t_e - t_b$ time series are created with a given window length w_{tr} .

$$s_1 : p_1, p_2, \dots, p_{w_{tr}}$$

$$s_2 : p_2, p_3, \dots, p_{w_{tr}+1}$$

...

$$s_N : p_N, p_{N+1}, \dots, p_{w_{tr}+N-1}$$

where $p_i (i = 1, 2, \dots, w_{tr} + N - 1)$ are stock prices at time i . We therefore create an N by w_{tr} matrix or a data set with N data records and w_{tr} attributes. Note that all attributes take continuous values and conventional data mining methods can be applied directly [3, 4].

For the test data, we use another window of length $w_{te} < w_{tr}$. Each training windowed series is then divided into two parts. The first part has the same length as the test data. The second part of length $w_{im} = w_{tr} - w_{te}$ is used to decide the classification of a cluster. All windowed time series are properly normalised. Figure 1 gives a schematic view of a windowed time series.

3 Methodology for Trend Analysis

Our data mining approach consists of the following steps.

pare our trading strategies with other existing strategies, we test them on one time period, namely for year 1999-2000. The corresponding training period is 1989-1998 (ten years). The comparison is made with [7]. To facilitate the comparison, stock indexes from five countries are used in the paper.

Tables 1 lists the return from NT, TTP, GP (Genetic Programming) and twenty one practical trading strategies for selected countries in the test time period. The values listed are the investment returns as fraction (for example, 0.1778 in Table 1 means that the return is 17.78%). For more details please refer to [7]. The B&H refers to buy and hold strategy.

We have the following observations based on the results presented in Table 1.

1. TTP's performance exceeds NT's performance in most countries. This clearly indicates that the trend prediction is able to find the correct trend in some cases. The trading strategy considering the price trend does improve the trading performance.
2. As shown in Table 1 for the time period 1999–2000, TTP has the best performance for US and Singapore in comparison with GPs, i.e., GP 1 and 2, and the twenty one practical trading strategies. For UK, NT, which is slightly better than TTP, performs the best. While all the twenty one practical trading strategies get negative or a slight positive return, TTP is able to produce significant positive returns for the time period 1999–2000. For Canada, GPs perform best, which is followed by B&H. TTP gives a slight positive return while most of the twenty one practical strategies get negative returns. For Taiwan, the GPs perform much better than all the other trading strategies. However TTP is able to exceed B&H and most of the twenty one practical strategies.

6 Conclusions and Future Work

We have applied a data mining approach to analyse and predict the trend of the stock price and applied it in real stock trading practice. Results have shown that the proposed methodology improves the trading performance over some existing strategies in some cases. While the methodology developed in the work can correctly predict the trend of stock prices for some

countries, it is not able to predict well for all. The stock price is very volatile in nature. The proposed trend prediction approach certainly has its limitations. The following future work may improve the performance of the method.

1. A simple decision on classification of clusters is made using the linear regression model in the present work. We can further improve the accuracy of the trend prediction by using fuzzy or probabilistic decision systems in the future.
2. Improve the computation efficiency by using sophisticated and scalable clustering techniques, such as [4, 8].
3. Introducing scale change to pattern matching can discover similar patterns with different time scales.
4. Combine our method with other techniques, such as GP, for better and more sophisticated trading strategies.

Acknowledgements

The authors acknowledge Damien McAullay and Arun Vishwanath for their assistance in the preparation of the paper.

References

- [1] X. Ge. Pattern matching financial time series data. Project Report ICS 278, UC Irvine, 1998.
- [2] E. Keogh and P. Smyth. A probabilistic approach to fast pattern matching in time series databases. In *Proceedings of KDD'97*, pages 24–30, Newport beach, CA, USA, 1997.
- [3] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 2001.
- [4] H.-D. Jin, M.-L. Wong, and K.-S. Leung. Scalable model-based clustering for large databases based on data summarization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1710–1719, Nov. 2005.

Table 1: The Total Return of Stock Trading for 1999–2000 in comparison with GP and 21 practical trading strategies

Rule	USA	UK	Canada	Taiwan	Singapore
B & H	0.0636	0.0478	0.3495	-0.2366	0.3625
GP 1	0.0655	0.0459	0.3660	0.1620	0.1461
GP 2	0.0685	0.0444	0.3414	0.5265	0.1620
TTP	0.1778	0.1524	0.0541	-0.22	0.4654
NT	0.0786	0.1560	0.0207	-0.1480	0.0524
1	-1.1173	-1.2855	-1.8943	-1.5102	-1.0679
2	0.0292	-0.5265	-0.9935	-0.8737	-0.8182
3	-0.1640	-0.6941	-0.2494	-0.3338	-0.7028
4	-0.9865	-0.8252	-0.1182	-0.7371	-0.5123
5	-0.0896	-0.3062	-0.9872	-0.2571	-0.6288
6	-0.7176	-0.6335	-0.0440	0.0048	-0.7599
7	-1.1736	-1.7050	-2.1544	-1.1646	-1.9132
8	-1.2402	-1.3594	-2.1444	-0.7130	-0.8391
9	-1.3883	-1.0738	-1.6657	-1.0748	-0.7450
10	-1.6532	-1.4603	-1.5322	-1.0678	-0.4226
11	-1.0941	-0.5934	-1.4946	-0.3628	-0.9329
12	-1.4735	-1.2046	-2.6474	-1.5254	-1.6464
13	-0.9116	-0.7762	-0.1522	-0.6863	-0.3210
14	-0.2477	-0.2666	-0.9692	-0.2258	-0.5817
15	-0.6658	-0.5571	0.0019	0.0218	-0.7405
16	-0.7576	-0.9016	-0.1671	-0.4350	-0.0302
17	-0.1607	0.0126	-1.0631	0.3375	-0.5044
18	-0.4397	-0.6185	-0.0055	0.1213	-0.4336
19	-0.4240	-0.7951	-0.0942	-0.1480	-0.1412
20	0.1419	-0.0474	-1.0680	-0.5793	-0.5628
21	-0.4195	-0.6143	0.0827	0.2087	-0.5644

- [5] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, Berkeley, University of California, 1967.
- [6] J. M. Chambers and T.J. Hastie, editors. *Statistical Models in S*, chapter Linear Models. Wadsworth & Brooks/Cole, 1992.
- [7] S. H. Chen, T. W. Kuo, and K. M. Hsu. *Handbook of Financial Engineering*, chapter Genetic Programming and Financial Trading: How Much about “What we Know”? Kluwer Academic Publishers, 2006.
- [8] H.-D. Jin, K.-S. Leung, M.-L. Wong, and Z.-B. Xu. Scalable model-based cluster analysis using clustering features. *Pattern Recognition*, 38(5):637–649, May 2005.