International Journal of Information Processing, 7(2), 90-100, 2013
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

Forecasting Stock Time-Series using Data Approximation and Pattern Sequence Similarity

Vishwanath R H[a], Leena S[a], Srikantaiah K C[a], K Shreekrishna Kumar[b], P Deepa Shenoy[a], Venugopal K R[a], S S Iyengar[c] and L M Patnaik[d]

[a]Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. Contact: vishwa.gld@gmail.com

[b]Director, All India Council for Technical Education, SWRO, Bangalore, India

[c]Director and Ryder Professor, Florida International University, USA

[d]Honorary Professor, Indian Institute of Science, Bangalore 560 001, India

Time series analysis is the process of building a model using statistical techniques to represent characteristics of time series data. Processing and forecasting huge time series data is a challenging task. This paper presents Approximation and Prediction of Stock Time-series data ($APST$), which is a two step approach to predict the direction of change of stock price indices. First, performs data approximation by using the technique called Multilevel Segment Mean (MSM). In second phase, prediction is performed for the approximated data using Euclidian distance and Nearest-Neighbour technique. The computational cost of data approximation is $O(n * n_i)$ and computational cost of prediction task is $O(m * |NN|)$. Thus, the accuracy and the time required for prediction in the proposed method is comparatively efficient than the existing Label Based Forecasting (LBF) method [1].

Keywords: Data Approximation, Nearest Neighbour, Pattern Sequence, Stock Time-Series.

1. INTRODUCTION

Data mining is the process of extracting knowledge, by dredging the data from huge database. Sequence database consists of sequence of ordered events with or without notion of time. Time series data is a sequence database which consists of sequences of values or events obtained over repeated measurements of time, which can be used in prediction of any future events for user applications. Forecasting is the prediction of forth coming events based on historical events. The recurring intervals for forecasting is based on the duration observed, *i.e.*, it requires many years for long term prediction, a year or more for medium term prediction and weeks or days for short term prediction.

1.1. Motivation

The main motivation behind this work is that, it is very much crucial for the stock market investors to estimate the behavior or trend of the stock market prices as precisely as possible in order to reach the best trading decisions for their investments. On the other hand, the complexity of many financial market is based on the nonlinearity and nonparametric nature of the variables influencing the index movement directions including human psychology and political events. The unpredictable volatile market index makes it a highly challenging task to accurately forecast its path of movement. In this context, it is required to build an efficient forcasting model, so that the investor can utilize the most accurate time series forecasting model to maximize the profit or to minimize the risk.

1.2. Methodologies

In this paper, we are using sliding window model to analyze stock time-series data. The

Table 2
The Prediction Errors by LBF and $APST$ methods on the TAIEX dataset for the financial year 2010

Month	MER(LBF)	MER($APST$)	MAE(LBF)	MAE($APST$)
April	5.02	4.22	0.53	0.43
May	8.30	7.22	0.56	0.46
June	6.89	5.59	0.51	0.41
July	7.41	6.21	0.47	0.37
Aug	8.37	7.57	0.47	0.37
Sep	7.30	6.40	0.45	0.35
Oct	4.62	3.68	0.47	0.37
Nov	7.26	6.28	0.44	0.34
Dec	6.88	5.88	0.43	0.35
Jan	7.20	8.26	0.44	0.36
Feb	6.26	4.26	0.44	0.35
Mar	7.26	5.26	0.44	0.38

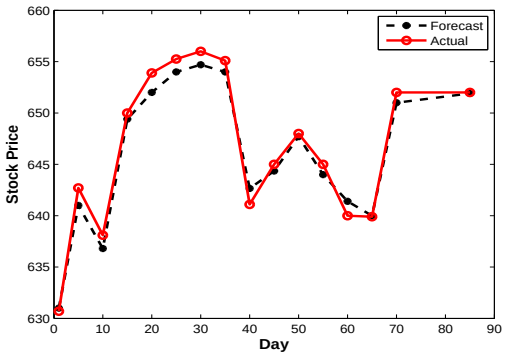

Figure 5. Comparison between Actual value and Forecasted value in $APST$ method

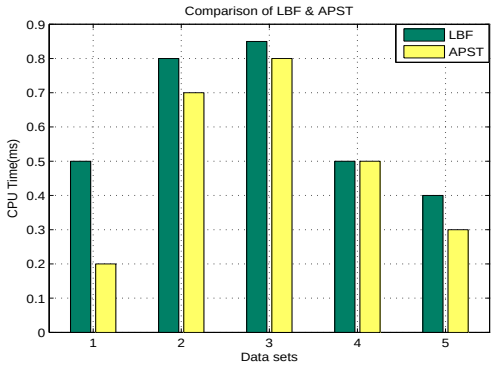

Figure 6. Comparison of Forecasting Time between LBF and $APST$ methods

average MER is 6.89%, average MAE is 0.47% in the existing LBF method, whereas in the proposed method, the average MER is 5.90% and average MAE is 0.37%. Thus, the proposed method is $\approx$ 1% more efficient with respect to MER and 0.1 % more efficient for MAE compared to existing LBF method.

Also, the average CPU time required for existing LBF method is 0.61 miliseconds, whereas in the proposed method, it is 0.5 miliseconds. Thus, proposed method is 0.11% more efficient than the existing method. Future enhancement can be focused on selecting the window size dynamically and fine tune the matching sequence.

REFERENCES

1. F Martnez-Alvarez, A Troncoso, J C Riquelme and J S Aguilar Ruiz. LBF: A Labeled-Based Forecasting Algorithm and Its Application to Electricity Price Time Series, *in Proceedings of Eighth IEEE Intl Conf. Data Mining*, pages 453–461, 2008.
2. Xiang Lian, Lei Chen, Jeffrey Xu Yu, Jinsong Han, Jian Ma. Multiscale Representations for

Fast Pattern Matching in Stream Time Series, *in IEEE Transaction on Knowledge and Data Engineering*, 21(4):568–581, 2009.

3. A Troncoso, J C Riquelme, J M Riquelme, J L Martnez and A Gomez. Electricity Market Price Forecasting Based on Weighted Nearest Neighbours Techniques, *in IEEE Transaction on Power Systems*, 22(3):1294–1301, 2007.
4. K J Kim. Financial Time Series Forecasting using Support Vector Machines, *in Neuro-Computing*, 55:307–319, 2003.
5. Y Radhika and M Shashi. Atmospheric Temperature Prediction using Support Vector Machines, *in International Journal of Computer Theory and Engineering*, 1:55–58, 2009.
6. R Nayak and te Braak. Temporal Pattern Matching for the Prediction of Stock Prices, *in Proceedings of 2nd International Workshop on Integrating Artificial Intelligence and Data Mining (AIDM2007)*, pages 99–107, 2007.
7. A J Conejo, M A Plazas, R Espnola and B Molina. Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and *ARIMA* Models, *in IEEE Transaction on Power Systems*, 20:1035–1042, 2005.
8. Akinwale Adio T, Arogundade O T and Adekoya Adebayo F. Translated Nigeria Stock Market Price using Artificial Neural Network for Effective Prediction, *in Journal of Theoretical and Applied Information Technology*, 2009.
9. Kuang Yu Huang and Chuen-Jiuan Jane. A Hybrid Model Stock Market Forecasting and Portfolio Selection Based on ARX, Grey System and RS Theories, *in Expert Systems with Applications*, pages 5387–5392, 2009.
10. M Suresh babu, N Geethanjali and B Sathyanarayana. Forecasting of Indian Stock Market Index Using Data Mining and Artificial Neural Nework, *in International journal of Advance Engineering and Application*, 2011.
11. F M lvarez, A Troncoso, J C Riquelme and J S A Ruiz. Energy Time Series Forecasting Based on Pattern Sequence Similarity, *in IEEE Transcation on Knowledge and Data Engineering*, 23(8):1230–1243, 2011.
12. Vishwanath R H, Leena S V, Srikantaiah K C, K Shreekrishna Kumar, P Deepa Shenoy, Venugopal K R, S S Iyengar and L M Patnaik. *APST*: Approximation and Prediction of Stock Time-Series Data using Pattern Sequence, *in Proceedings 8th International Multi Conference on Information Processing (ICIP2013)*, 2013.

Vishwanath R Hulipalled is an Assistant Professor in the Department of Computer Science and Engineering at Sambhram Institute of Technology, Bangalore, India. He received his Bachelors degree in Computer Science and Engineering from Karnataka University and Master of Engineering from UVCE, Bangalore University, Bangalore. He is presently pursuing his Ph.D in the area of Data Mining in JNTU Hyderabad. His research interest includes Time Series Mining and Data Analysis.

Leena is pursuing B.E in Department of Computer Science and Engineering, University Visveswaraya College of Engineering, Bangalore. Her research interest is in the area of Data Mining and Time Series Mining.

Srikantaiah K C is an Associate Professor in the Department of Computer Science and Engineering at S J B Institute of Technology, Bangalore, India. He obtained his B.E and M.E degrees in Computer Science and Engineering from Bangalore University, Bangalore. He is presently pursuing his Ph.D programme in the area of Web Mining in Bangalore University. His research interest is in the area of Data Mining, Web Mining and Semantic Web.

K Shreekrishna Kumar is currently the Director of All India Council for Technical Education, SWRO, Bangalore. He obtained his Master of Science from Bhopal University. He received his Masters degree in Information Technology from Punjab University. He was awarded Ph.D in Physics (Glass Technology) from Ma-

hatma Gandhi University. He was the member of the Jury Panel, Indian Journal of Pure and Applied Physics, CSIR (New Delhi), He was the Collaborative researcher, Nuclear Science Centre, New Delhi.

P Deepa Shenoy was born in India, on May 9, 1961. She graduated form UVCE, completed her M.E. from UVCE., has done her MS(Systems and information) from BITS., Pilani, and has obtained her Ph.D in CSE from Bangalore University. She is presently employed as a Professor in department of CSE at UVCE. Her research interests include Computer Networks, Wireless Sensor Networks, Parallel and Distributed Systems, Digital Signal Processing and Data Mining.

Venugopal K R is currently the Principal, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. He obtained his Bachelor of Engineering from University Visvesvaraya College of Engineering. He received his Masters degree in Computer Science and Automation from Indian Institute of Science Bangalore. He was awarded Ph.D in Economics from Bangalore University and Ph.D in Computer Science from Indian Institute of Technology, Madras. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science and Journalism. He has authored and edited 39 books on Computer Science and Economics, which include Petrodollar and the World Economy, C Aptitude, Mastering C, Microprocessor Programming, Mastering C++ and Digital Circuits and Systems *etc.*. During his three decades of service at UVCE he has over 350 research papers to his credit. His research interests include Computer Networks, Wireless Sensor Networks, Parallel and Distributed Systems, Digital Signal Processing and Data Mining.

S S Iyengar is currently the Roy Paul Daniels Professor and Chairman of the Computer Science Department at Louisiana State University. He heads the Wireless Sensor Networks Laboratory and the Robotics Research Laboratory at LSU. He has been involved with research in High Performance Algorithms, Data Structures, Sensor Fusion and Intelligent Systems, since receiving his Ph.D degree in 1974 from MSU, USA. He is Fellow of IEEE and ACM. He has directed over 40 Ph.D students and 100 Post Graduate students, many of whom are faculty at Major Universities worldwide or Scientists or Engineers at National Labs/Industries around the world. He has published more than 500 research papers and has authored/co-authored 6 books and edited 7 books. His books are published by John Wiley & Sons, CRC Press, Prentice Hall, Springer Verlag, IEEE Computer Society Press *etc.*. One of his books titled Introduction to Parallel Algorithms has been translated to Chinese.

L M Patnaik is currently Honorary Professor, Indian Institute of Science, Bangalore, India. He was a Vice Chancellor, Defense Institute of Advanced Technology, Pune, India and was a Professor since 1986 with the Department of Computer Science and Automation, Indian Institute of Science, Bangalore. During the past 35 years of his service at the Institute he has over 700 research publications in refereed International Journals and refereed International Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academies in India; Fellow of the IEEE and the Academy of Science for the Developing World. He has received twenty national and international awards; notable among them is the IEEE Technical Achievement Award for his significant contributions to High Performance Computing and Soft Computing. His areas of research interest have been Parallel and Distributed Computing, Mobile Computing, CAD for VLSI circuits, Soft Computing and Computational Neuroscience.