Leadership Ranking: Bringing Order to Correlated Time Series

Di Wu, Yiping Ke, Jeffrey Xu Yu, Philip S. Yu, Lei Chen

The Chinese University of Hong Kong, {dwu,ypke,yu}@se.cuhk.edu.hk

University of Illinois at Chicago, psyu@cs.uic.edu

The Hong Kong University of Science and Technology, leichen@cse.ust.hk

Keywords: Financial Applications, Time Series, Streaming Data, and Evolving Graph.

Extended Abstract

Data streams have emerged as an important research area in recent years due to its ubiquitous presence in various application domains including stock market analysis in finance. Therefore, it has become increasingly demanding to develop efficient techniques to discover useful patterns from such data streams. Among many issues, analyzing the correlation between two data streams is an important technique to study linear co-movements of two data streams [11, 9]. Yet, lagged correlation further captures lagged co-movements between time-evolving streams [4, 6], which is especially useful for forecasting and monitoring. In empirical finance, the statistics (e.g., price, return, volatility) of a security may exhibit a lead-lag effect on those of other securities [7]. This information can be used to predict the trends of the led securities. Although the lagged correlation between a pair of streams has been well studied in empirical research [3, 1] and efficient algorithms to discover lagged correlations have also been developed [8], the study on analyzing the relationships across multiple data streams is still lacking. The comprehensive relationships among multiple data streams is very helpful in understanding, monitoring and controlling the overall movement of the entity where the data streams are generated.

We study the problem of discovering the leadership relationships among a set of time series by analyzing lead-lag relations. In particular, we study this problem on one of its typical applications in financial market: finding the leaders among multiple time series of stock prices. Extracting the group of stock leaders will bring enlightening information as discussed below.

- Tracking Property: The stock leaders we exploit provide the empirical finance researchers, financial analysts and investors with insightful understanding and tracking of the market behavior. Stock leaders present a new perspective to analyze the stock market: a revealed news event introduces some changes to stock leaders' prices, whose effect then propagates to related stocks by lagged correlations. As a result, analysts only need to monitor and analyze stock leaders in order to evaluate the whole market.
- Predictive Power: The evolutionary pattern of stock leaders provides a large potential for modeling stock behavior, from which different trading strategies can be designed. For example, the price movement of stock leaders implicates the future market trend and thus has prediction power of market index (Dow Jones Index, Standard Poor 500, etc.). In other words, the market trend can be detected in an early stage. For example, traders can buy the call option of the market index when detecting that stock leaders begin a rising trend. On the other hand, index arbitrage can be implemented by trading these stock leaders, the sample of which closely mirrors the market index [5].

The problem of finding the leaders among multiple time series of stock prices poses great challenges. First, stock prices usually change rapidly over time, which implies that the leaderships among stocks may also evolve dramatically over time. In order to provide the answer to a user whenever he/she issues a request, we need to develop an efficient algorithm that is able to discover leaders in a real-time manner. Second, before computing the leaderships among multiple time series, we first need to identify the significant lagged correlations between pairs of time series. Given N time series and the maximum lag is set to be m, at each time t, the number of computations for lagged correlation is $\mathcal{O}(mN^2)$. This high complexity and dynamic updating nature makes the design of a scalable solution difficult. Finally, how to define and extract useful leaders is also a big challenge.

We proposed an efficient streaming algorithm to address the problem [10]. The main idea of our algorithm is described as follows. Given N time series, at a specific time t, we first compute the lagged correlation between each pair of time series. The existing work [8] on computing lagged correlations cannot be directly applied, since i) it tries to capture lag correlation in the whole history of streams while our objective is to obtain the local lags in the current sliding window, and ii) the approximation in their updating algorithm has accuracy preference to the points with small lags and may generate a larger error for large lag l, which is not desirable for our problem. Therefore, we propose to aggregate the effects of various lags and develop an efficient update approach to compute the aggregated lagged correlation by investigating the evolutionary characteristics of lagged correlations. Upon the pairwise lagged correlations computed, we construct an edge-weighted directed graph to analyze the lead-lag relation among the set of time series. We then detect the leaders among the time series by analyzing the leadership transmission in the graph. More specifically, we apply the PageRank algorithm [2] on the graph to compute the ranking of each time series in order to decide the importance of each time series in the graph. Finally, based on the structure of the graph and the rank values of time series, we extract the leaders by eliminating redundant leaderships.

References

- [1] R. Bhuyan. Information, alternative markets, and security price processes: A survey of literature. Finance 0211002, EconWPA, Nov. 2002.
- [2] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.
- [3] K. Chan. A further analysis of the lead-lag relationship between the cash market and stock index futures market. *Review of Financial Studies*, 5(1):123–52, 1992.
- [4] G. R. George Box, Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control.* Prentice Hall, 1994.
- [5] J. Hull. Options, Futures, and Other Derivatives. Prentice Hall, 2008.
- [6] U. A. M. O. P. R. O. Ramazan Gencay, Michel Dacorogna. An Introduction to High-Frequency Finance. Academic Press, 2001.
- [7] P. Säfvenblad. Lead-lag effects when prices reveal cross-security information. Working Paper Series in Economics and Finance 189, Stockholm School of Economics, Sept. 1997.
- [8] Y. Sakurai, S. Papadimitriou, and C. Faloutsos. Braid: Stream mining through group lag correlations. In SIGMOD Conference, pages 599–610, 2005.
- [9] D. Wu, G. P. C. Fung, J. X. Yu, and Z. Liu. Mining multiple time series co-movements. In *APWeb*, pages 572–583, 2008.
- [10] D. Wu, Y. Ke, J. X. Yu, P. S. Yu, and L. Chen. Leadership discovery when data correlatively evolve. World Wide Web, 14(1), 2011.
- [11] Y. Zhu and D. Shasha. Statistream: Statistical monitoring of thousands of data streams in real time. In *VLDB*, pages 358–369, 2002.