Online Tensor learning for Spatio-Temporal Financial Market Prediction

Davide Di Gioia *[1], Tao Cheng †[1]

[1]SpaceTimeLab for Big Data Analytics, Department of Civil, Environmental & Geomatic Engineering, University College London

February 14, 2021

1. Summary

Financial market prediction is significantly important for investment decision making, but it is one of the most challenging tasks as the markets are affected by multiple factors. This paper aims to predict financial market data using spatio-temporal tensor model in online fashion with different spatial weight matrices. We introduce a new spatial weight matrix to represent the association of financial variables and compared ones based on location and correlation distance. Finally, a regression model is presented for multivariate financial data prediction. Models are tested and validated using data from Reuters and Factset database.

KEYWORDS: Stocks data, tensor decomposition, prediction, online

1. Introduction

Financial time series data are more convoluted than other statistical data due to the long-term trends, cyclical variations, seasonal variations and irregular movements (Rout et al. 2017). Nevertheless, predicting future stock prices has been the goal of professionals and academics since the birth of the financial market.

Traditional statistical models have long been applied in the financial market, and more recently machine learning (ML) techniques have been adopted that are potentially more capable of capturing the dynamics of financial markets. In certain scenarios, we sequentially receive data instances, which prevents training in a batch setting (Jaeger 2002). For this reason, in recent times, incremental and online machine learning receive more and more attention especially in the context of learning from real-time data streams, such as online learning for smart healthcare platform (He et al. 2019).

In recent years, there has been an increasing number of studies treating financial series as spatio-temporal series due to the fact that the correlation of stock exchange market returns rises with reducing physical distance (e.g. Pirinsky & Wang, 2006; Barker & Loughran, 2007). In finance, however, it is not obvious how distance should be estimated (Fernandez 2011) and due to this adversity, even though there is a compressive literature on spatiotemporal models, the use of such model in finance and economics are not widespread. Several works have explored spatial weight matrix to be described as a spatial correlation function of geographical distance (Eckel et al., 2011; Zhu et al., 2013) or socio-economic distance (Abate, 2016; Asgharian et al., 2013). However, these models are not suitable for applications with multi-directional relatedness, for instance in a task that predicts multiples economics/financial variables at different locations and times.

* davide.gioia.17@ucl.ac.uk
† tao.cheng@ucl.ac.uk

A natural way to portray multivariate spatio-temporal data and a useful technique to capture inter-dependencies along multiple dimensions can be through tensors. Tensor decomposition techniques have been valuable for dealing with different problems in the field of spatio-temporal data mining, such as traffic prediction (Han & Moutarde, 2016; Tan et al., 2016), data compression (Asif et al. 2013), human mobility analysis (Sun and Axhausen 2016) and climate prediction (Bahadori et al., 2014; Li et al., 2020). However, their application for multivariate for finance is missing.

To solve abovementioned issue, we introduce a new spatial weight matrix to define the spatial correlation among financial data , and we propose to predict multivariate financial time series using Accelerated Online Low-Rank Tensor Learning for Multivariate Spatio-Temporal Streams (ALTO) (Yu, Cheng, and Liu 2015). A low-rank tensor is essentially a high-order generalization of matrix factorization (Ma, Wang, and Qin 2020). An online learning of low-rank tensors aims to dynamically update a tensor while preserving the low-rank structure. We validate our method via financial time series data provided by Reuters. The details of data representation, methodologies and case study are given as below.

2. Data description

The data used in this paper is a sample data from Reuter from May 2006 to April 2017 for eight countries in Europe (France, Germany, Switzerland, Italy, Netherlands, Spain, Sweden, and United Kingdom). Each country has around 2798 daily data for stocks per sector. The structure is shown in Figure 1. It comprises spatial and temporal information (i.e., the datetime, prices and coordinates) for each country (8 in total). The data structure consists of 8 variables (stocks per each country) per 5 different financial Sectors (Health Care, Technology and Telecommunications, Consumer Discretionary, Financial Services and Real Estate, and Basic Material Industry and Construction) and indices per country (CAC 40, DAX, SMI. FTSE MIB, AEX, IBEX, OMX, and FTSE 100). For all the datasets, each variable is normalized by removing mean and dividing by variance.

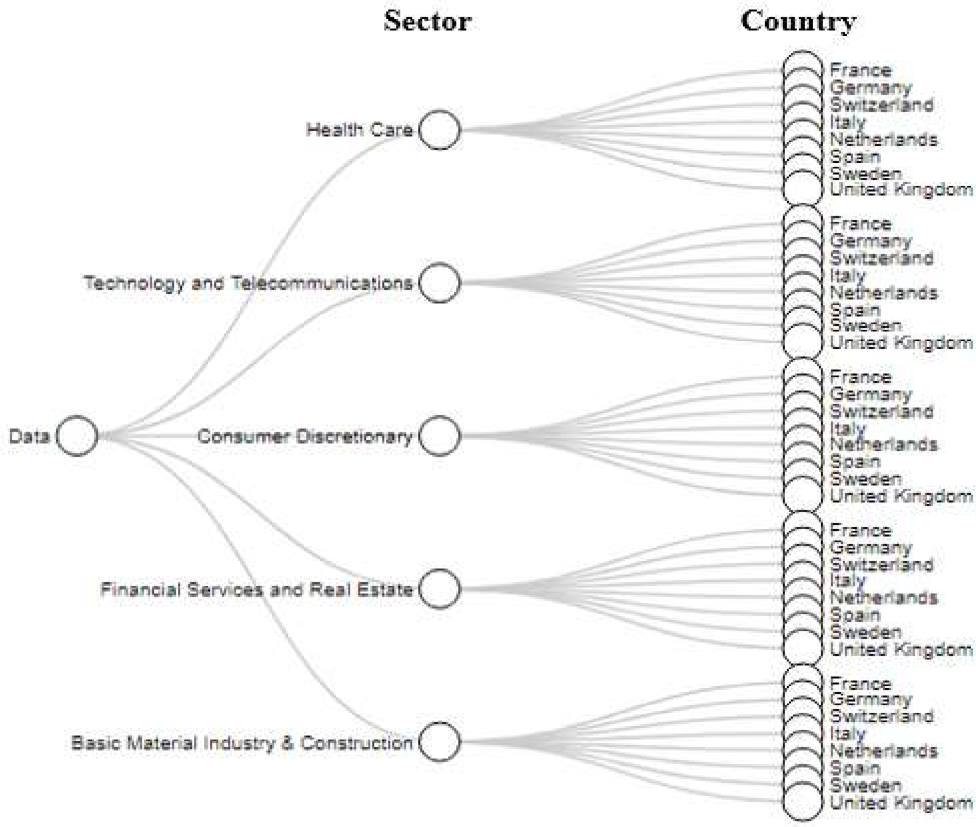

Figure 1 Data structure for sector and country

3. Methodology

3.1 Spatial Weight Matrix

In order to take advantage of a spatial dimension, we use the concept of local consistency to represent data points in term of correlation and to extract information from the dataset, three types of spatial weight matrix are explored based on Laplacian matrix. The first is based on location distance matrix by calculating the pairwise Haversine distance of locations, the second is built on Correlation Distance (CD) (Mantegna 1999) and the last one is placed on Correlation and Maximal Information Coefficient (CDMIC) calculated on the main stock exchange index that represent each country.

A limitation of the CD approach is connected with the fact that Pearson' correlation only measure linear relationships, but at the same time is the dominant tool used in financial analysis, even if it violates the solid evidence of nonlinearity of financial market (Brock, Hsieh, and LeBaron 1991), market index returns (Abhyankar, Copeland, and Wong 1995), and forex market (Hsieh, 1989; Qi & Wu, 2003). This problem has been swapping Pearson's correlation coefficient with mutual information (Baghli 2006), which permit to take into account the nonlinearity and the assumption of multivariate normality (Fiedor 2014). However, there are also a few disadvantages of using mutual information, such as to whatever the relationship is positive or negative, as an alternative, we propose similarity measure integrating Correlation based Distance and Maximal Information Coefficient (MIC) defined as:

$$CD_{MIC}(t) = \frac{1}{\sqrt{2}}\sqrt{d(_{ij})^2 + {MIC_{ij}}^2}$$

where, $d(_{ij})$ is the CD matrix and MIC is the Maximal Information Coefficient (Reshef et al. 2011).

3.2 Online Tensor for financial prediction

We reformulate the financial prediction task as a vector auto regressive (VAR) in an online tensor form and we use ALTO to predict multiple stocks integrating spatial dimension. Tensors are higher-order generalizations of vectors and matrices (Kolda and Bader 2009), which have achieved great success in the field of spatio-temporal prediction (Tang et al. 2020). The ALTO model aims to dynamically update a tensor while preserving the low-rank structure. For the multivariate spatio-temporal online learning, the model offers the solution via an unambiguous two-phase tactic: (1) solving the unconstrained optimisation problem given the new observation, (2) updating the solution with the low-rank constraint, such as projecting the solution to the space of low-rank tensor, and reaching the final goal of estimating low-rank constrained coefficient tensor .

4 Case study

The performance of prediction is evaluated by Root Mean Square Error (*RMSE*), and compared with its batch counterpart GREEDY (Bahadori et al., 2014) and VAR. Results show that the prediction accuracy is higher by including the spatial dimension and achieve in a much faster way (see Table 1).

Table 1: Prediction Evaluation with Stocks data - RMSE

Variables	**Location**		**CD Distance**		**CD-MIC Distance**		**VAR**
	Batch	*Online*	*Batch*	*Online*	*Batch*	*Online*	
Health Care	0.10544	0.14297	0. 08614	0.08496	0.08524	**0.08488**	0.320
Consum. Discr	0.27992	0.18488	0. 30195	0.21643	0.29543	**0.21462**	0.269
Basic Material & Ind.	0.18189	0.12651	0.13981	0. 1285	0.14447	**0.12678**	0.244
Financ. Services	0.2444	0.11318	0.17023	0.1014	0.16847	**0.1004**	0.279
Tech & Telecom.	0.085633	0.14747	0.06417	0.11372	0.06457	**0.11362**	0.219

5 Conclusion and Discussion

This work explores the possibility of spatio-temporal data to predict multivariate financial time series using online tensor model. Experiment results demonstrate the prediction accuracy is higher than batch methods. This research can be the baseline for more advanced prediction model. However, there is room for improvement. Currently, we train different distance matrices for stocks prediction, but we could use and develop many more. Moreover, the model is not able to deal with missing data, while financial data, for different reasons, recurrently comprehend missing values, and can introduce bias.

References

Abate, G. D. 2016. On the Link between Volatility and Growth: A Spatial Econometrics Approach. *Spatial Economic Analysis* 11 (1):27–45.

Abhyankar, A., L. S. Copeland, and W. Wong. 1995. Nonlinear Dynamics in Real-Time Equity Market Indices: Evidence from the United Kingdom. *The Economic Journal* 105 (431):864–880.

Asgharian, H., W. Hess, and L. Liu. 2013. A spatial analysis of international stock market linkages. *Journal of Banking & Finance* 37 (12):4738–4754.

Asif, M. T., S. Kannan, J. Dauwels, and P. Jaillet. 2013. Data compression techniques for urban traffic data. In *2013 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS)*, 44–49.

Baghli, M. 2006. A model-free characterization of causality. *Economics Letters* 91 (3):380–388.

Bahadori, M. T., Q. (Rose) Yu, and Y. Liu. 2014. Fast Multivariate Spatio-temporal Analysis via Low Rank Tensor Learning. In *Advances in Neural Information Processing Systems 27*, eds. Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, 3491–3499. Curran Associates, Inc. http://papers.nips.cc/paper/5429-fast-multivariate-spatio-temporal-analysis-via-low-rank-tensor-learning.pdf (last accessed 18 July 2019).

Barker, D., and T. Loughran. 2007. The Geography of S&P 500 Stock Returns. *Journal of Behavioral Finance* 8 (4):177–190.

Brock, W. A., D. A. Hsieh, and B. D. LeBaron. 1991. *Nonlinear dynamics, chaos, and instability: statistical theory and economic evidence*. Cambridge, Mass: MIT Press.

Eckel, S., G. Löffler, A. Maurer, and V. Schmidt. 2011. Measuring the effects of geographical distance on stock market correlation. *Journal of Empirical Finance* 18 (2):237–247.

Fernandez, V. 2011. Spatial linkages in international financial markets. *Quantitative Finance* 11 (2):237–245.

Fiedor, P. 2014. Networks in financial markets based on the mutual information rate. *Physical Review E* 89 (5):052801.

Han, Y., and F. Moutarde. 2016. Analysis of Large-Scale Traffic Dynamics in an Urban Transportation Network Using Non-Negative Tensor Factorization. *International Journal of Intelligent Transportation Systems Research* 14 (1):36–49.

He, Y., B. Wu, D. Wu, E. Beyazit, S. Chen, and X. Wu. 2019. Online Learning from Capricious Data Streams: A Generative Approach. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2491–2497. Macao, China: International Joint Conferences on Artificial Intelligence Organization https://www.ijcai.org/proceedings/2019/346 (last accessed 14 February 2021).

Hsieh, D. A. 1989. Testing for Nonlinear Dependence in Daily Foreign Exchange Rates. *The Journal of Business* 62 (3):339–368.

Jaeger, H. 2002. Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the echo state network approach. *GMD-Forschungszentrum Informationstechnik, 2002.* 5.

Kolda, T. G., and B. W. Bader. 2009. Tensor Decompositions and Applications. *SIAM Review* 51 (3):455–500.

Li, S., L. Wu, Y. Yang, T. Geng, W. Cai, B. Gan, Z. Chen, Z. Jing, G. Wang, and X. Ma. 2020. The Pacific Decadal Oscillation less predictable under greenhouse warming. *Nature Climate Change* 10 (1):30–34.

Ma, P., L. Wang, and J. Qin. 2020. A Low-Rank Tensor Factorization Using Implicit Similarity in Trust Relationships. *Symmetry* 12 (3):439.

Mantegna, R. N. 1999. Hierarchical structure in financial markets. *The European Physical Journal B* 11 (1):193–197.

Pirinsky, C., and Q. Wang. 2006. Does Corporate Headquarters Location Matter for Stock Returns? *The Journal of Finance* 61 (4):1991–2015.

Qi, M., and Y. Wu. 2003. Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance* 10 (5):623–640.

Reshef, D. N., Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. 2011. Detecting novel associations in large data sets. *Science (New York, N.Y.)* 334 (6062):1518–1524.

Rout, A. K., P. K. Dash, R. Dash, and R. Bisoi. 2017. Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. *Journal of King Saud University - Computer and Information Sciences* 29 (4):536–552.

Sun, L., and K. W. Axhausen. 2016. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological* 91:511–524.

Tan, H., Y. Wu, B. Shen, P. J. Jin, and B. Ran. 2016. Short-Term Traffic Prediction Based on Dynamic Tensor Completion. *IEEE Transactions on Intelligent Transportation Systems* 17 (8):2123–2133.

Tang, J., X. Wang, F. Zong, and Z. Hu. 2020. Uncovering Spatio-temporal Travel Patterns Using a Tensor-based Model from Metro Smart Card Data in Shenzhen, China. *Sustainability* 12 (4):1475.

Yu, R., D. Cheng, and Y. Liu. 2015. Accelerated Online Low-Rank Tensor Learning for Multivariate Spatio-Temporal Streams. :10.

Zhu, B., R. Füss, and N. B. Rottke. 2013. Spatial Linkages in Returns and Volatilities among U.S. Regional Housing Markets: Spatial Linkages in Returns and Volatilities among U.S. Regional Housing Markets. *Real Estate Economics* 41 (1):29–64.

Biographies

Davide Di Gioia is a part-time PhD student at SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London. His research interest includes financial prediction, online learning and tensor learning.

Tao Cheng is a Professor in GeoInformatics, and Director of SpceTimeLab for Big Data Analytics (http://www.ucl.ac.uk/spacetimelab), at University College London. Her research interests span network complexity, Geocomputation, integrated spatio-temporal analytics and big data mining (modelling, prediction, clustering, and simulation), with applications in transport, crime, health, social media, and environmental monitoring.