

MIDAS 2016: The 1st Workshop on MIning DAta for financial applicationS

Ilaria Bordino¹, Guido Caldarelli², Fabio Fumarola¹,
Francesco Gullo¹, and Tiziano Squartini²

¹ UniCredit, R&D Department, Italy
{ilaria.bordino, fabio.fumarola, francesco.gullo}@unicredit.eu

² IMT Institute for Advanced Studies Lucca, Italy
{guido.caldarelli, tiziano.squartini}@imtlucca.it

1 Motivation

Like the famous King Midas, popularly remembered in Greek mythology for his ability to turn everything he touched with his hand into gold, the wealth of data generated by modern technologies, with widespread presence of computers, users and media connected by Internet, is a goldmine for tackling a variety of problems in the financial domain.

Nowadays, people's interactions with technological systems provide us with gargantuan amounts of data documenting collective behavior in a previously unimaginable fashion [8, 14]. Recent research has shown that by properly modeling and analyzing these massive datasets, for instance representing them as network structures [2, 4], it is possible to gain useful insights into the evolution of the systems considered (i.e., trading [13], disease spreading [1, 12], political elections [5]). Investigating the impact of data arising from today's application domains on financial decisions may be of paramount importance. Knowledge extracted from data can help gather critical information for trading decisions, reveal early signs of impactful events (such as stock market moves), or anticipate catastrophic events (e.g., financial crises) that result from a combination of actions, and affect humans worldwide.

The importance of data-mining tasks in the financial domain has been long recognized [9]. For example, in the Web context, changes in the frequency with which users browse news or look for certain terms on search engines such as Google have been correlated with product trends [7], the level of activity in certain given industries, unemployment rates, or car and home sales [6], as well as stock-market trade volumes and price movements [3, 11, 15]. Other core application scenarios include forecasting the stock market, predicting bank bankruptcies, understanding and managing financial risk, trading futures, credit rating, loan management, bank customer profiling [9]. Despite its well-recognized relevance and some recent related efforts [10], data mining in finance is still not stably part of the main stream of data-mining conferences. This makes the topic particularly appealing for a workshop proposal, whose small, interactive, and possibly interdisciplinary context provides a unique opportunity to advance research in a stimulating but still quite unexplored field.

2 Objectives and topics

The aim of the *1st Workshop on Mining Data for financial applicationS* (MIDAS 2016), held in conjunction with the *2016 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (ECML-PKDD 2016), Riva del Garda, Italy, September 19–23, 2016, is to discuss challenges, potentialities, and applications of leveraging data-mining tasks to tackle problems in the financial domain. The workshop provides a premier forum for sharing findings, knowledge, insights, experience and lessons learned from mining data generated in various domains. The intrinsic interdisciplinary nature of the workshop promotes the interaction between computer scientists, physicists, mathematicians, economists and financial analysts, thus paving the way for an exciting and stimulating environment involving researchers and practitioners from different areas.

Topics of interest include, but are not limited to:

- Forecasting the stock market
- Trading models
- Discovering market trends
- Predictive analytics for financial services
- Network analytics in finance
- Planning investment strategies
- Portfolio management
- Understanding and managing financial risk
- Customer/investor profiling
- Identifying expert investors
- Financial modeling
- Measures of success in forecasting
- Anomaly detection in financial data
- Fraud detection
- Discovering patterns and correlations in financial data
- Text mining and NLP for financial applications
- Financial network analysis
- Time series analysis
- Pitfalls identification

3 Outcomes

MIDAS 2016 was structured as a *full-day* workshop. We encouraged submissions of regular papers (long or short), and extended abstracts. Regular papers may be up to 12 pages (long papers) or 6 pages (short papers), and report on novel, unpublished work that might not be mature enough for a conference or journal submission. Extended abstracts may be up to 2 pages long, and present work-in-progress, recently published work fitting the workshop topics, or position papers.

All submitted papers were peer-reviewed by three reviewers from the program committee, and selected on the basis of these reviews. MIDAS 2016 received 13

submissions, among which 8 papers were accepted (5 long regular papers, 1 short regular paper, 2 extended abstracts), with an acceptance rate of about 61%. The competitive acceptance rate resulted in a high-quality and exciting program.

The program was enriched by two invited speakers: Prof. Fabrizio Lillo, Scuola Normale Superiore, Pisa (Italy), who gave a talk titled “*Detection of intensity bursts using Hawkes processes: an application to high frequency financial data*”, and Dr. Marcello Paris, UniCredit, R&D Department, Rome (Italy), who gave a talk titled “*The Geometry of Financial Markets: Topological Data Analysis*”.

4 Program Committee

The scientific significance of the workshop is assured by a Program Committee which includes research scholars coming from different countries, and widely recognized as experts in the topics of interest of the workshop:

- Aris Anagnostopoulos, *Sapienza University of Rome*, Italy
- Annalisa Appice, *University of Bari*, Italy
- Xiao Bai, *Yahoo!*, USA
- Nicola Barbieri, *Tumblr*, USA
- Paolo Barucca, *Scuola Normale Superiore*, Italy
- Michele Berlingerio, *IBM Research*, Ireland
- Annalina Caputo, *University of Bari*, Italy
- Gianbiagio Curato, *Scuola Normale Superiore*, Italy
- Carlotta Domeniconi, *George Mason University*, USA
- Debora Donato, *StumbleUpon*, USA
- Andrea Ferretti, *UniCredit*, Italy
- Ruth Garcia Gavilanes, *Oxford Internet Institute*, UK
- Sara Hajian, *Eurecat*, Spain
- Roberto Interdonato, *University of Calabria*, Italy
- Andreas Kaltenbrunner, *Eurecat*, Spain
- Dragi Kocev, *Jozef Stefan Institute*, Slovenia
- Nicolas Kourtellis, *Telefonica Research*, Spain
- Iordanis Koutsopoulos, *Athens University of Economics and Business*, Greece
- Donato Malerba, *University of Bari*, Italy
- Yelena Mejova, *Qatar Computing Research Institute*, Qatar
- Davide Mottin, *Hasso Plattner Institute*, Germany
- Giuseppe Nicosia, *University of Catania*, Italy
- Marcello Paris, *UniCredit*, Italy
- Stefano Pascolutti, *UniCredit*, Italy
- Alvin Pastore, *University of Sheffield*, UK
- Giovanni Ponti, *ENEA*, Italy
- Aleksandra Rashkovska, *Jozef Stefan Institute*, Slovenia
- Giovanni Stilo, *Sapienza University of Rome*, Italy
- Antti Ukkonen, *Finnish Institute of Occupational Health*, Finland
- Edoardo Vacchi, *UniCredit*, Italy
- Tim Weninger, *University of Notre Dame*, USA
- Giovanni Zappella, *UniCredit*, Italy

References

- [1] Harshavardhan Achrekar, Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu, and Benyuan Liu. Predicting flu trends using twitter data. In *Proc. of IEEE Int. Conf. on Computer Communications Workshops (INFOCOM Workshops)*, pages 702–707, 2011.
- [2] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciché, N. Vandewalle, and R. N. Mantegna. Networks of equities in financial markets. *European Physical Journal B*, 38:363–371, 2004.
- [3] Iliaria Bordino, Stefano Battiston, Guido Caldarelli, Matthieu Cristelli, Antti Ukkonen, and Ingmar Weber. Web search queries can predict stock market volumes. *PLoS ONE*, 7(7):e40014, 2012.
- [4] G. Caldarelli. *Scale-Free Networks: Complex Webs in Nature and Technology*. Oxford University Press, 2007.
- [5] Guido Caldarelli, Alessandro Chessa, Fabio Pammolli, Gabriele Pompa, Michelangelo Puliga, Massimo Riccaboni, and Gianni Riotta. A multi-level geographical study of italian political elections from twitter data. *PLoS ONE*, 9(5):e95809, 2014.
- [6] Hyunyoung Choi and Hal Varian. Predicting the present with Google trends. Technical report, 2009.
- [7] Sharad Goel, Jake M. Hofman, Sébastien Lahaie, David M. Pennock, and Duncan J. Watts. Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences (PNAS)*, 107(41):17486–17490, 2010.
- [8] Gary King. Ensuring the data rich future of the social sciences. *Science*, 331(11 February):719–721, 2011.
- [9] Boris Kovalerchuk and Evgenii Vityaev. *Data Mining for Financial Applications*, pages 1203–1224. Springer US, 2005.
- [10] Rajasekar Krishnamurthy, Louiqa Raschid, and Shivakumar Vaithyanathan, editors. *Proceedings of the International Workshop on Data Science for Macro-Modeling, DSMM’14, Snowbird, UT, USA, June 22-27, 2014*. ACM, 2014.
- [11] T Preis, D Reith, and Stanley H E. Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society A*, 368(1933):5707?–5719, 2010.
- [12] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, Jose Ramasco, Duygu Balcan, Bruno Goncalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine*, 10(1):165+, 2012.
- [13] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences (PNAS)*, 102(30):10421–10426, 2005.
- [14] Alessandro Vespignani. Predicting the behavior of techno-social systems. *Science*, 325:425–428, 2009.
- [15] Wei-Xing Zhou, Gabriele Ranco, Iliaria Bordino, Giacomo Bormetti, Guido Caldarelli, Fabrizio Lillo, and Michele Treccani. Coupling news sentiment with web browsing data improves prediction of intra-day price dynamics. *PLoS ONE*, 11(1):e0146576, 2016.