Forecasting Big Time Series: Old and New

Christos Faloutsos
CMU and Amazon Research
christos@cs.cmu.edu

Jan Gasthaus
Amazon AI Labs
gasthaus@amazon.com

Tim Januschowski
Amazon AI Labs
tjnsch@amazon.com

Yuyang Wang
Amazon AI Labs
yuyawang@amazon.com

ABSTRACT

Time series forecasting is a key ingredient in the automation and optimization of business processes: in retail, deciding which products to order and where to store them depends on the forecasts of future demand in different regions; in cloud computing, the estimated future usage of services and infrastructure components guides capacity planning; and workforce scheduling in warehouses, call centers, factories requires forecasts of the future workload. Recent years have witnessed a paradigm shift in forecasting techniques and applications, from computer-assisted model- and assumption-based to data-driven and fully-automated. This shift can be attributed to the availability of large, rich, and diverse time series data sources, posing unprecedented challenges to traditional time series forecasting methods. As such, how can we build statistical models to efficiently and effectively learn to forecast from large and diverse data sources? How can we leverage the statistical power of "similar" time series to improve forecasts in the case of limited observations? What are the implications for building forecasting systems that can handle large data volumes?

The objective of this tutorial is to provide a concise and intuitive overview of the most important methods and tools available for solving large-scale forecasting problems. We review the state of the art in three related fields: (1) classical modeling of time series, (2) scalable tensor methods, and (3) deep learning for forecasting. Further, we share lessons learned from building scalable forecasting systems. While our focus is on providing an intuitive overview of the methods and practical issues, we also present technical details underlying these powerful tools.

PVLDB Reference Format:
Christos Faloutsos, Jan Gasthaus, Tim Januschowski, Yuyang Wang. Forecasting Big Time Series: Old and New. *PVLDB*, 11 (12): 2102-2105, 2018.
DOI: https://doi.org/10.14778/3229863.3229878

Proceedings of the VLDB Endowment, Vol. 11, No. 12

DOI: https://doi.org/10.14778/3229863.3229878

1. INTRODUCTION

Time series data occur naturally in countless domains including medical analysis [24], sensor network monitoring [27], financial analysis [43], social activity mining [23, 22] and database systems [35, 19]. Of all the time series related data mining tasks, forecasting is one of the most sought-after applications (and arguably the most difficult one) due to its importance in industrial, social and scientific applications. For example, forecasting plays a key role in automating and optimizing operational processes in most businesses and enables data driven decision making. Forecasts of product supply and demand can be used for optimal inventory management, staff scheduling and topology planning, and are more generally a crucial technology for most aspects of supply chain optimization. Outside of the retail use-case, the increasing volume of online, time-stamped activities represents a vital new opportunity for data scientists and analysts to measure the collective behavior of social, economic, and other important evolutions [14].

Time series forecasting is a well-known topic that has attracted interest from many research communities including statistics, machine learning, econometrics, operations research, databases, data mining for several decades. Each community has focused on different aspects of the problem. In the statistics and econometrics communities, the prevalent forecasting methods in use today have been developed in the setting of forecasting individual or small groups of time series with complex models designed and tuned by domain experts. On the other hand, data mining and database researchers have been focusing on finding patterns in thousands or millions of related time series. Examples include forecasting the energy consumption of individual households, forecasting the load for servers in a data center, or forecasting the demand for all products that a large retailer offers. In these scenarios, a substantial amount of data on past behavior of similar, related time series can be leveraged for making a forecast for an individual time series. Using data from related time series not only allows fitting more complex (and hence potentially more accurate) models without overfitting, it can also alleviate the time and labor intensive human selection and preparation of co-variates and model selection steps required by classical techniques. Recent studies has revealed some new directions for research on large scale time series forecasting, including:

- **Scalable state-space models:** Classical state-space models serve as a reliable workhorse for forecasting, with appealing properties such as interpretability, robustness, and theoretical guarantees. Modern extensions include scalable implementations, as well as the support for missing data and multiple data types.

- **Large-scale tensor analysis:** Time series data can be modeled as tensors, and tensor analysis is an important data mining tool that has various applications including sensor streams, hyperlinks, medical records and social networks over time.
- **Deep Learning for forecasting:** With its dominance in machine learning applications such as image recognition and machine translation, deep learning has recently also received revived interests in the field of time series forecasting. Modern deep learning techniques not only improve the state-of-art forecasting performance but also, from a systems perspective, greatly reduce the complexity of the forecasting pipeline, and therefore increase maintainability.

This tutorial aims to bring together classical forecasting techniques, time series data mining techniques, and deep learning based-forecasting methods through a concise and intuitive overview of the most important tools and techniques that we can use to help us understand and forecast time series. We will provide a comprehensive overview of proven and current directions for time series forecasting, and deal specifically with the following key topics: (1) classical linear modeling of time series, (2) scalable tensor methods, (3) deep learning for forecasting, and (4) lessons learned developing, running and maintaining large scale forecasting systems. We supply Jupyter notebooks to illustrate commonly used forecasting techniques covered in this tutorial.

Who should attend. The target audience consists of database and data mining researchers who wish to familiarize themselves with the major techniques and recent developments in time series forecasting. Additionally, this tutorial is suitable for practitioners who want a concise, intuitive overview of the state of the art.

Prerequisites. The tutorial assumes familiarity with basic linear algebra, calculus, and discrete math, as well as with fundamentals of machine learning.

Related tutorials and how the current proposal differs. Related tutorials have been presented, e.g., (a) *Mining and Forecasting of Big Time Series Data*, by Yasuhi Sakurai, Yasuko Matsubara, and Christos Faloutsos, SIGMOD 2015, WWW 2016, KDD 2017, (b) *Indexing and Mining Streams*, by Christos Faloutsos, SIGMOD 2004, (c) *Indexing and Mining Time Sequences*, by Christos Faloutsos and Lei Li, SIGKDD 2010, and (d) *Mining Shape and Time Series Databases with Symbolic Representations*, by Eamonn Keogh, SIGKDD 2007.

The proposed tutorial differs from the existing ones in several aspects: (1) we exclusively focus on the *forecasting* part of the time series analysis; (2) we bring together the *statistical and econometric* aspects with *data mining and management* of large scale time series forecasting, and (3), our major addition is the inclusion of recent *deep learning models for forecasting*, which have become increasingly popular in various domains and frequently shown superior performance compared to classical methods.

2. OUTLINE

1. **Introduction to Forecasting and Classical Models (20 minutes)**
   - Basic (explanatory) analysis and decomposition of time series, i.e., trend, level, seasonality, etc.
   - Linear models and exponential smoothing
     - Generalized Linear Models (GLM)
     - Exponential smoothing (ES), Holt-Winters, and general Innovation state space models (ISSM)
2. **Modern Methods: Tensor Analysis and Deep Learning (60 minutes)**
   - Scalable Tensor Analysis
   - Deep learning for forecasting
     - Multi-layer perceptron (feedforward neural networks)
     - Recurrent neural networks (RNN)s: vanilla, Seq2seq and other architectures
     - Others structures: Convolution, WaveNet, and all that
3. **Conclusions and Lessons learned (10 minutes)**
   - Building large scale forecasting systems
   - Developing Deep Autoregressive Network (DeepAR) in AWS Sagemaker

1. Introduction to Forecasting and Classical Methods. In the opening chapter of the tutorial, we introduce the basic forecasting concepts and terminology. The classical time series analysis tools such as time series decomposition, lag plots, autocorrelations, etc. are also introduced [3, 2]. We discuss how to evaluate the accuracy of a forecast with metrics such as mean absolute percentage error (MAPE), quantile losses [12], as well as metrics to evaluate hierarchical forecasts [29]. Then, we cover the classic linear methods for forecasting, including linear regression, autoregressive and moving average models (ARIMA), and useful tools in data management system such as MUSCLES [39] and AWSOM [26]. We also introduce linear dynamical systems (LDS), Kalman filter (KF) and their variants [16, 15, 34]. In particular, we focus on the exponential smoothing models (Innovation state space models (ISSM)) [11] and structural time series models [8, 5, 30] along with their Bayesian counterparts [32, 31]. We show how to incorporate trend, seasonality factors, external signals such as promotional and other types of events and missing (or partially missing) observations [31]. We also cover how to model different types of time series observations, e.g. real, positive, integer. We close this part by discussing the topic of scalable implementations of (Bayesian) state-space models [32, 31].

2. Modern methods: Tensor Analysis and Deep Learning. In this part, we introduce contemporary methods based on tensors and deep learning models. In contrast to the classical methods, the modern approaches learn across multiple related time series globally.

2.1 We present large-scale studies of complex time-stamped events and big sparse tensors. We introduce classic matrix factorizations (MF), tensor factorizations (Tucker, CP, HOSVD), and forecasting methods that are based on MF [40]. We describe several methods to automatically mining complex time-stamped tensor, including TriMine [22], TensorCast [4] and others [24, 20, 21, 33].
2.2 Next, we introduce deep learning/neural network models for forecasting. In the 90s, Feedforward NNs were popular among forecasters [42] with applications in electrical load [28, 18], financial time series [7], and others [9]. Recent ground-breaking successes of deep neural network in other areas of machine learning have brought revived interests in applying deep learning techniques, especially recurrent neural networks and their variants [10], to time series forecasting. In this part, we first introduce the multi-layer perceptron (feedforward NN) as an extension of the linear regression models introduced in part 1. Then we review different types of Recurrent Neural Networks (RNNs), which capture the sequential nature of time series data. Different RNN forecasters are introduced [37, 6, 41, 17, 25] and we explain the intuitions behind different structures and demonstrate their performances on a variety types of time series. We shall discuss convolutional NNs [38], WaveNet [36] and illustrate how they can be used for forecasting. Finally, we discuss new directions for deep generative models for forecasting, in particular, with models that combines the strengths of both RNNs and classical probabilistic graphical models.

3. Conclusions and Lessons learned. We conclude the tutorial with a summary of the previous parts and share the lessons learned building the scalable forecasting system for retail and cloud resource within Amazon [1] and developing cloud-based deep learning forecasting algorithms such as DeepAR [6] in AWS SageMaker [13].

3. PRESENTERS' SHORT BIOGRAPHY

Christos Faloutsos is a Professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, the SIGKDD Innovations Award (2010), twenty "best paper" awards (including two test of time awards), and four teaching awards. Five of his advisees have attracted KDD or SCS dissertation awards. He is an ACM Fellow, he has served as a member of the executive committee of SIGKDD; he has published over 300 refereed articles, 17 book chapters, and two monographs. He holds eight patents and has given over 40 tutorials and over 20 invited distinguished lectures. His research interests include data mining for graphs and streams, fractals, database performance, and indexing for multimedia and bioinformatics data.

Jan Gasthaus is a Senior Machine Learning Scientist in the Amazon AI Labs, working mainly on time series forecasting and large-scale probabilistic machine learning. He is passionate about developing novel machine learning solutions for addressing challenging business problems with scalable machine learning systems, all the way from scientific ideation to productization. Prior to joining Amazon, Jan obtained a BS in Cognitive Science from the University of Osnabrueck, an MS in Intelligent Systems from UCL, and pursued a PhD at the Gatsby Unit, UCL, focusing on Nonparametric Bayesian methods for sequence data.

Tim Januschowski is a Machine Learning Science Manager in Amazon AI Labs. He has worked on forecasting since starting his professional career. At Amazon, he has produced end-to-end solutions for a wide variety of forecasting problems, from demand forecasting to server capacity forecasting. Tim's personal interests in forecasting span applications, system, algorithm and modeling aspects and the downstream mathematical programming problems. He studied Mathematics at TU Berlin, IMPA, Rio de Janeiro, and Zuse-Institute Berlin and holds a PhD from University College Cork.

Yuyang Wang is a Senior Machine Learning Scientist in Amazon AI Labs, working mainly on large-scale probabilistic machine learning with its application in Forecasting. He received his PhD in Computer Science from Tufts University, MA, US and he holds an MS from the Department of Computer Science at Tsinghua University, Beijing, China. His research interests span statistical machine learning, numerical linear algebra, and random matrix theory. In forecasting, Yuyang has worked on all aspects ranging from practical applications to theoretical foundations.

4. REFERENCES

[1] J.-H. Böse, V. Flunkert, J. Gasthaus, T. Januschowski, D. Lange, D. Salinas, S. Schelter, M. Seeger, and Y. Wang. Probabilistic demand forecasting at scale. *PVLDB*, 10(12):1694–1705, 2017.

[2] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[3] P. J. Brockwell and R. A. Davis. Time series: Theory and method. Springer-Verlag, 1991.

[4] M. R. de Araujo, P. M. P. Ribeiro, and C. Faloutsos. Tensorcast: Forecasting with context using coupled tensors. In *Data Mining (ICDM), 2017 IEEE International Conference on*, pages 71–80. IEEE, 2017.

[5] J. Durbin and S. J. Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.

[6] V. Flunkert, D. Salinas, and J. Gasthaus. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *arXiv preprint arXiv:1704.04110*, 2017.

[7] E. Gately. *Neural networks for financial forecasting*. John Wiley & Sons, Inc., 1995.

[8] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

[9] T. Hill, L. Marquez, M. O'Connor, and W. Remus. Artificial neural network models for forecasting and decision making. *International journal of forecasting*, 10(1):5–15, 1994.

[10] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[11] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.

[12] R. J. Hyndman and G. Athanasopoulos. Forecasting: Principles and practice. otexts; 2014. *www. otexts. org/fpp.*, 987507109, 2017.

[13] T. Januschowski, D. Arpin, D. Salinas, V. Flunkert, J. Gasthaus, L. Stella, and P. Vazquez. Now available in amazon sagemaker: Deepar algorithm for more accurate time series forecasting. *https://aws.amazon.com/blogs/machine-learning/now-available-in-amazon-sagemaker-deepar-algorithm-for-more-accurate-time-series-forecasting/*, 2018.

[14] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. *ACM SIGKDD*, 2009.

[15] L. Li, J. McCann, N. Pollard, and C. Faloutsos. Dynammo: mining and summarization of coevolving sequences with missing values. In *ACM SIGKDD*, pages 507–516, 2009.

[16] L. Li, B. A. Prakash, and C. Faloutsos. Parsimonious linear fingerprinting for time series. *PVLDB*, 3(1-2):385–396, 2010.

[17] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Graph convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

[18] C.-N. Lu, H.-T. Wu, and S. Vemuri. Neural network based short term load forecasting. *IEEE Transactions on Power Systems*, 8(1):336–342, 1993.

[19] L. Ma, D. Van Aken, A. Hefny, G. Mezerhane, A. Pavlo, and G. J. Gordon. Query-based workload forecasting for self-driving database management systems. In *SIGMOD*, pages 631–645. ACM, 2018.

[20] Y. Matsubara, Y. Sakurai, and C. Faloutsos. The web as a jungle: Non-linear dynamical systems for co-evolving online activities. In *WWW*, pages 721–731. ACM, 2015.

[21] Y. Matsubara, Y. Sakurai, and C. Faloutsos. Non-linear mining of competing local activities. In *WWW*, pages 737–747. ACM, 2016.

[22] Y. Matsubara, Y. Sakurai, C. Faloutsos, T. Iwata, and M. Yoshikawa. Fast mining and forecasting of complex time-stamped events. In *KDD*, pages 271–279. ACM, 2012.

[23] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14. ACM, 2012.

[24] Y. Matsubara, Y. Sakurai, W. G. van Panhuis, and C. Faloutsos. Funnel: automatic mining of spatially coevolving epidemics. In *KDD*, pages 105–114. ACM, 2014.

[25] S. Mukherjee, D. Shankar, A. Ghosh, N. Tathawadekar, P. Kompalli, S. Sarawagi, and K. Chaudhury. Armdn: Associative and recurrent mixture density networks for eretail demand forecasting. *arXiv preprint arXiv:1803.03800*, 2018.

[26] S. Papadimitriou, A. Brockwell, and C. Faloutsos. Adaptive, hands-off stream mining. In *VLDB*, pages 560–571. Morgan Kaufmann, 2003.

[27] S. Papadimitriou and P. S. Yu. Optimal multi-scale patterns in time series streams. In *SIGMOD Conference*, pages 647–658, 2006.

[28] D. C. Park, M. El-Sharkawi, R. Marks, L. Atlas, and M. Damborg. Electric load forecasting using an artificial neural network. *IEEE transactions on Power Systems*, 6(2):442–449, 1991.

[29] L. Roberts, L. Razoumov, L. Su, and Y. Wang. Gini regularized optimal transport with an application to spatio-temporal forecasting. *NIPS Workshop on Optimal Transport*, 2017.

[30] S. L. Scott and H. R. Varian. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2):4–23, 2014.

[31] M. Seeger, S. Rangapuram, Y. Wang, D. Salinas, J. Gasthaus, T. Januschowski, and V. Flunkert. Approximate bayesian inference in linear state space models for intermittent demand forecasting at scale. *arXiv preprint arXiv:1709.07638*, 2017.

[32] M. W. Seeger, D. Salinas, and V. Flunkert. Bayesian intermittent demand forecasting for large inventories. In *NIPS*, pages 4646–4654, 2016.

[33] K. Takeuchi, H. Kashima, and N. Ueda. Autoregressive tensor factorization for spatio-temporal predictions. In *ICDM*, pages 1105–1110. IEEE, 2017.

[34] Y. Tao, C. Faloutsos, D. Papadias, and B. Liu. Prediction and indexing of moving objects with unknown motion patterns. In *SIGMOD Conference*, pages 611–622. ACM, 2004.

[35] D. Van Aken, A. Pavlo, G. J. Gordon, and B. Zhang. Automatic database management system tuning through large-scale machine learning. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pages 1009–1024. ACM, 2017.

[36] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. In *SSW*, page 125, 2016.

[37] R. Wen, K. Torkkola, and B. Narayanaswamy. A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*, 2017.

[38] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *NIPS*, pages 802–810, 2015.

[39] B. Yi, N. Sidiropoulos, T. Johnson, H. Jagadish, C. Faloutsos, and A. Biliris. Online data mining for co-evolving time sequences. In *Data Engineering, 2000. Proceedings. 16th International Conference on*, pages 13–22. IEEE, 2000.

[40] H.-F. Yu, N. Rao, and I. S. Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. In *NIPS*, pages 847–855, 2016.

[41] R. Yu, S. Zheng, A. Anandkumar, and Y. Yue. Long-term forecasting using tensor-train rnns. *arXiv preprint arXiv:1711.00073*, 2017.

[42] G. Zhang, B. E. Patuwo, and M. Y. Hu. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.

[43] Y. Zhu and D. Shasha. Statstream: Statistical monitoring of thousands of data streams in real time. *PVLDB*, pages 358–369, 2002.