UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA MECÂNICA

JONAS MACHADO MIGUEL

Machine Learning-based Spatio-Temporal Forecasting of Wind Power Generation

São Paulo
2020

JONAS MACHADO MIGUEL

Machine Learning-based Spatio-Temporal Forecasting of Wind Power Generation

Versão original

Relatório preliminar apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do título de Bacharel em Engenharia Mecânica pelo Departamento de Engenharia Mecânica.

Área de concentração: Métodos de Aprendizado de Máquina.

Orientador: Prof. Dr. Fábio Gagliardi Cozman

Coorientador: Dr. Alexandre Cristovão Maiorano

São Paulo

2020

Ficha catalográfica

Relatório preliminar de autoria de Jonas Machado Miguel, sob o título **"Machine Learning-based Spatio-Temporal Forecasting of Wind Power Generation"**, apresentado à Escola Politécnica da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Bacharel em Engenharia Mecânica pelo Departamento de Engenharia Mecânica, aprovado em 4 de Maio de 2020 pela comissão examinadora constituída pelos doutores:

Prof. Dr. Fábio Gagliardi Cozman

Instituição: Escola Politécnica da Universidade de São Paulo

Presidente

Dr. Alexandre Cristovão Maiorano

Instituição: Itaú Unibanco

To my parents, Marcionete and Pedro, for their love and courage.

Resumo

Predizer o comportamento de sistemas regidos por correlações temporais e espaciais é uma tarefa a que se tem atribuída crescente importância em diversas áreas de aplicação, desde neurociência, epidemiologia e criminologia a logística e transporte. Neste trabalho, delineamos o estado da arte para métodos de predição espaço-temporal e implementamos uma seleção desses métodos para a predição de geração de energia eólica no nível distrital na Alemanha. Na análise, levamos em conta tanto séries temporais com resolução horária entre 2000 e 2015, como também especificações de projeto e de instalação de turbinas eólicas individuais. Os modelos são avaliados em períodos não modelados e comparados com métodos convencionais de previsão.

Palavras-chaves: Análise de Séries Temporais, Previsão Espaço-Temporal, Aprendizagem de Máquina, Redes Neurais, Energias Renováveis, Energia Eólica.

Abstract

Forecasting the behavior of systems in which both temporal and spatial dependencies play a central role has received increased attention, with applications domains including neuroscience, epidemiology, criminology and transportation. We review the state-of-the-art for spatio-temporal forecasting methods and implement selected approaches for predicting wind power generation at the district-level in Germany. Besides hourly time series for power generation in individual districts in 2000-2015, the analysis considers design and installation specifications for single wind turbines. The models are evaluated on unmodelled periods and locations and benchmarked against conventional forecasting methods.

Keywords: Time Series Analysis. Spatio-Temporal Forecasting. Machine Learning, Neural Networks, Wind Power.

Lista de figuras

Lista de tabelas

Lista de abreviaturas e siglas

AR	Auto-Regressive
ARIMA	Auto-Regressive Integrated Moving Average
CNN	Convolutional Neural Network
ES	Exponential Smoothing
ML	Machine Learning
PDF	Probability Density Function
PMF	Probability Mass Function
RNN	Recurrent Neural Network
ST	Spatio-Temporal
TS	Time Series
VARIMA	Vector Auto-Regressive Integrated Moving Average
VES	Vector Exponential Smoothing

Lista de símbolos

Λ	Lambda

Sumário

[1] De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 Introduction

Phenomena presenting high socio-economical relevance which are governed by complex dependencies of both spatial and temporal nature are found in diverse domains such as epidemiology, criminology, transportation, climate science and astrophysics (ATLURI; KARPATNE; KUMAR, 2018). Indeed, the ability to describe a system's behavior is most valuable on instances downstream in the arrow of time: forecasting (ARMSTRONG, 2001; HAGERTY, 2017). Accurate, scalable and feasible rule-based forecasting modeling, however, remains elusive in many cases. Especially as ubiquitous and continuous monitoring data become available, data-driven approaches emerge as a promising alternative.

Conventional data-driven approaches alone, however, have often shown to add limited value in spatio-temporal forecasting (MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2018). A major reason for this limitation lies on the assumptions they rely upon being typically violated in spatio-temporal settings. Seasonality and spatial independency underlie most of the approaches from time series analysis, while earlier machine learning methods assume data instances are independent and identically distributed (i.i.d.) (ATLURI; KARPATNE; KUMAR, 2018). Recently, deep learning-based approaches have shown to be able to overcome this essentially by (a) modelling both spatial and temporal dependencies and (b) considering spatial similarities in terms less obvious than geographical proximity alone (LI et al., 2017; YAO et al., 2018; GENG et al., 2019; WU et al., 2019).

In the context of renewables, accurately estimating power generation ahead of time poses a major obstacle in progressing towards carbon neutrality in power generation. Heavily conditioned on weather and climate, harvesting energy from renewable sources is characterized by intermittency. In the case of onshore wind power generation, climate change further aggravates this character, as wind speeds variability are expected to increase (MOEMKEN et al., 2018). Not accurately knowing how much wind power will be harvested in a certain time and region means power providers have to rely on unnecessarily large safety margins provided by conventional power plants for ensuring sufficient power supply. This ultimately hampers the expansion of wind farms and represents therefore a loss for the society, as part of the paid overall generated power is lost, as well as for the environment, as less environment-friendly power sources have to be relied upon (DELARUE; MORRIS, 2015). For countries committed to large-scale initiatives such as the *Energiewende* in Germany, this poses a major hindrance in decreasing overall carbon footprint in a sustainable fashion. Accuracy on wind power generation forecasting hence has significant impact on both socio-economical and environmental aspects, in both short and long terms.

For a given installed capacity, wind power generation depends primarily on local wind speeds, which heavily vary in both time and space. While the power generation can be predicted for each single region independently using historical data, we hypothesize that a significant increase in forecasting accuracy might be achieved by also considering inter-regional spatial dependencies.

The objective of this work is twofold. First, we delineate the state-of-the-art approaches for spatio-temporal forecasting in different domains. Second, we apply selected approaches for forecasting wind power generation at the district-level in Germany. By benchmarking against conventional approaches and single-regions forecasting horizons, we investigate whether more sophisticated modelling approaches add significant value in terms of accuracy in the use case of onshore wind power generation.

2 Background

In this chapter, we first present the different settings, approaches, and performance metrics used in general spatio-temporal forecasting problems. We then describe fundamental aspects of wind power generation underlying this work.

2.1 Spatio-Temporal Forecasting

In this section, we present the spatio-temporal forecasting problem, how it is approached, and how forecasting models can be assessed and compared.

2.1.1 Problem statement

In this work, we categorize spatio-temporal forecasting problems according to (1) the degree of dependency among sensors and (2) the stationarity of sensors locations. The term sensor is used here in the abstract sense of a stochastic data generation process, which could be physically represented by an actual sensor measuring a variable of interest in a particular phenomenon.

In a first regime, referred here as regime I, sensors are fixed in space, with negligible dependencies among them. The uncertainty about the state of a sensor cannot be reduced by knowing the states of its neighboring sensors. As a consequence, using a single model to represent the different sensors is expected to present no advantage over modeling every sensor independently. Characteristic of this regime is also the covariance matrix for the different sensors being both diagonal and invariant in time.

In regime II, sensors are also fixed in space, but this time with significant dependencies among them. The uncertainty about the state of a sensor can be reduced by knowing the states of its neighboring sensors. In other words, uncertainties among sensors are coupled. Modeling sensors together could be potentially beneficial in such case. Besides, the covariance matrix is expected to be non-diagonal but still invariant in time.

Finally, in regime III, dependent sensors move in space. Dependencies across sensors should hence also change over time, and a corresponding time-dependent covariance matrix is expected to follow. Again, models representing multiple sensors could make use of this and outperform models for single sensors.

2.1.2 Conventional Approaches

In this work, we refer as conventional approaches what is in the literature often referred as Time Series [1] approaches. Their formulations were motivated by forecasting problems in which time was the single independent variable. The hallmark of conventional forecasting approaches is their reliance on the well-developed theory for describing stationary random processes. There are two general ways of describing (modeling) a generic time series: (1) the Exponential Smoothing (ES) framework, and (2) the Auto-Regressive Integrated Moving Average (ARIMA) framework (BROCKWELL; DAVIS; FIENBERG, 1991).

Exponential Smoothing Framework

In the ES framework, time series are modelled as a superposition of three components: trend (m_t), seasonal (s_t), and random noise (Y_t). This is known as the Classical Decomposition (Equation 1).

$$X_t = m_t + s_t + Y_t \tag{1}$$

The underlying principle is to apply a filter to X_t that smooths out the noise component Y_t, allowing m_t and s_t to be estimated and extracted. Techniques within this framework differ by (a) the filter, (b) the assumptions on and preprocessing of X_t. In fact, the simplicity of models based on ES and their success in temporal forecasting problems have made this framework the default choice in the industry for such settings (HOLT, 2004). We describe two methods as examples: (1) the least squares, (2) the exponential smoothing method. Both assume non-seasonality of X_t (i.e.,$s_t = 0$), meaning a deseasoning of the time series is typically required as a preprocessing step.

[1] A time series is defined as a stochastic process ..., X_{t-1}, X_t, X_{t+1}, ... consisting of random variables indexed by time index t. The stochastic behavior X_t is described by $p(x_{t1}, x_{t2}, ..., x_{tm})$, i.e.,the PDF (or PMF) for all finite collections of time indexes $(t_1, t_2, ..., t_m), m < \infty$ (KEMPTHORNE C. LEE, 2013).

In the least squares method, m_t is first approximated by a parametric family of functions (e.g. $m_t = a_0 + a_1 t + a_2 t^2$). The parameters are then estimated via the the minimization of the squared errors $\Sigma_t (x_t - m_t)^2$.

In the exponential smoothing method, a pre-defined $a \in [0,1]$ is used for the estimated trend $\hat{m}_t$ by Equation 2.1.2. $\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1}, t = 1, ..., n\hat{m}_1 = X_1$

The resulting expression of $\hat{m}_t$ in terms of the past measurements $X_t, X_{t-1}, ...$, motivates the name of this method:

$$\hat{m}_t = \sum_{j=0} t - 2a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1 \hat{m}_{t-1}, \tag{2}$$

i.e.,$\hat{m}_t$ is a weighted moving average of the past measurements $X_t, X_{t-1}, ...$, with weigths decreasing exponentially.

So far, the exponential smoothing framework has been presented in its univariate version. Its multivariate version is the Vector Exponential Smoothing (VES) framework. While ES-based models can be used to properly address regime I forecasting problems, VES-based models can incorporate cross-sensors dependencies into the covariance matrix to model sensors under regime I, II or III.

ARIMA Framework

First proposed by (BOX; JENKINS, 1970) (and hence often referred as Box-Jenkins Methods), the ARIMA framework relies on differencing to achieve stationarity. Differencing operators, are recursively applied to the data x_t until the resulting observations are approximatelly stationary. (BROCKWELL; DAVIS; FIENBERG, 1991) For k recursions, the corresponding operator is referred as $\nabla^k(\cdot)$. As an instance, for k=1: $\nabla X_t = X_t - X_{t-1}$.

The framework is named after the ARIMA model, described by seção 2.1.2, with the autoregressive operator ϕ_k, moving average operator θ_k (both of order k), and innovation (white noise) at time index t a_t. x$_t = \phi_1 x_t + ... + \phi_{p+d} x_{t-p-d} - \theta_1 a_{t-1} - ... - \theta q a_{t-q} + a_t$ The first line of seção 2.1.2 corresponds to the autoregressive (AR) component of ARIMA, in which x_t is represented as a linear regression of its preceding values $x_{t-1}, ..., x_{t-p-d}$.

Important nonlinear methods are included in this framework, such as ARCH and GARCH, in which the innovation term is modelled by respectively by an AR model and by an ARMA model. The multivariate version of ARIMA is the Vector-ARIMA (VARIMA). Like its counterpart from the Exponential Smoothing, VARIMA models make use of a covariance matrix to incorporate cross-sensors dependencies for the higher coupling regimes II and III.

2.1.3 Machine Learning-based Approaches

Machine Learning approaches rely on (1) definition of relatively general architectures and (2) finding a configuration of parameter values in the given architecture that minimizes the expectation of some loss function. As the loss function represents a discrepancy between predictions and ground truth, this optimization process leads to a model that can be used to predict system behavior given a configuration for inputs values. The optimization process itself is typically performed by a gradient-based algorithm. (GOODFELLOW; BENGIO; COURVILLE, 2016)

In the context of univariate temporal forecasting (i.e.,regime I), the performance of Machine Learning algorithms was considered by some to be very limited reliability and usefulness (MAKRIDAKIS; HIBON, 2000). However, (ORESHKIN et al., 2019) recently demonstrated that a "pure"Deep Learning approach could not only (1) consistently outperform conventional ones, but also (2) be less reliant on manual tuning and (3) be made interpretable in both final and intermediate outputs. Until then, top-performing ML-based models were either a result of a combination or hybridization with conventional methods. Earlier approaches relied on ML-TS Combinations, in which outputs from statistical engines were used as features for ML algorithms. Later, TS models had their parameters optimized via gradient-descent and stacked with a Recurrent Neural Network (RNN) to form a hybrid model (SMYL, 2020).

Outside of regime I, ST forecasting problems were already successfully addressed by Deep Learning approaches, such as in forecasting traffic (LI et al., 2017), ride-hailing demand (GENG et al., 2019) and electrical power demand (TOUBEAU et al., 2018). Most of these approaches relied on RNN architectures, eventually combined @inproceedingsgeng2019spatiotemporal, title=Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting, author=Geng, Xu and Li, Yaguang and Wang, Leye and Zhang, Lingyu and Yang, Qiang and Ye, Jieping and Liu, Yan, booktitle=Proceedings of the AAAI Conference on Artificial Intelligence, volume=33, pages=3656–3663, year=2019 with a CNN architecture. More recently, approaches that model the Spatio-Temporal dependencies over a non-Euclidean space in a graph representation have been proposed and currently represent the state-of-the-art for ST forecasting problems (WU et al., 2019).

2.1.4 Forecasting Performance Metrics

Different quantities can be used for assessing models forecasting performance. Some of the most popular are MAE (Mean Absolute Error, Equation 3), $MAPE$ (Mean Absolute Percentual Error, Equation 4), $RMSE$ (Root Mean Squared Error, Equation 5). As they expose different qualities of performance, combining a reasonable number of metrics can be advisable.

$$MAE(\hat{X}^{(t+i):(t+T)};\Theta) = \frac{1}{TND}\Sigma_{i=1}^{T}\Sigma_{j=1}^{N}\Sigma_{k=1}^{D}|\hat{X}_{jk}^{(t+i)} - X_{jk}^{(t+i)}| \tag{3}$$

$$MAPE(\hat{X}^{(t+i):(t+T)};\Theta) = \frac{100}{TND}\Sigma_{i=1}^{T}\Sigma_{j=1}^{N}\Sigma_{k=1}^{D}\frac{|\hat{X}_{jk}^{(t+i)} - X_{jk}^{(t+i)}|}{|X_{jk}^{t+i}|} \tag{4}$$

$$RMSE(\hat{X}^{(t+i):(t+T)};\Theta) = \sqrt{\frac{1}{TND}\Sigma_{i=1}^{T}\Sigma_{j=1}^{N}\Sigma_{k=1}^{D}(\hat{X}_{jk}^{(t+i)} - X_{jk}^{(t+i)})^2} \tag{5}$$

2.2 Wind Power Generation

3 Use Case

3.1 Requirements

3.2 Resources

3.2.1 Dataset

Referências[1]

ARMSTRONG, J. S. *Principles of forecasting: a handbook for researchers and practitioners*. [S.l.]: Springer Science & Business Media, 2001. v. 30. Citado na página 12.

ATLURI, G.; KARPATNE, A.; KUMAR, V. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 51, n. 4, p. 1–41, 2018. Citado na página 12.

BOX, G. E.; JENKINS, G. M. *Time Series Analysis Forecasting and Control*. [S.l.], 1970. Citado na página 16.

BROCKWELL, P. J.; DAVIS, R. A.; FIENBERG, S. E. *Time series: theory and methods: theory and methods*. [S.l.]: Springer Science & Business Media, 1991. Citado 2 vezes nas páginas 15 e 16.

DELARUE, E.; MORRIS, J. *Renewables intermittency: operational limits and implications for long-term energy system models*. [S.l.], 2015. Citado na página 12.

GENG, X. et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.: s.n.], 2019. v. 33, p. 3656–3663. Citado 2 vezes nas páginas 12 e 17.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016. Citado na página 17.

HAGERTY, J. planning guide for data and analytics. *Gartner Inc*, p. 13, 2017. Citado na página 12.

HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, Elsevier, v. 20, n. 1, p. 5–10, 2004. Citado na página 15.

KEMPTHORNE C. LEE, V. S. J. X. P. *Topics in Mathematics with Applications in Finance*: 18.s096. 2013. Disponível em: ⟨https://ocw.mit.edu/courses/mathematics/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/⟩. Citado na página 15.

LI, Y. et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017. Citado 2 vezes nas páginas 12 e 17.

MAKRIDAKIS, S.; HIBON, M. The m3-competition: results, conclusions and implications. *International journal of forecasting*, Elsevier, v. 16, n. 4, p. 451–476, 2000. Citado na página 17.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, Public Library of Science, v. 13, n. 3, 2018. Citado na página 12.

MOEMKEN, J. et al. Future changes of wind speed and wind energy potentials in euro-cordex ensemble simulations. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 123, n. 12, p. 6373–6389, 2018. Citado na página 12.

[1] De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

ORESHKIN, B. N. et al. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019. Citado na página 17.

SMYL, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, Elsevier, v. 36, n. 1, p. 75–85, 2020. Citado na página 17.

TOUBEAU, J.-F. et al. Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets. *IEEE Transactions on Power Systems*, IEEE, v. 34, n. 2, p. 1203–1215, 2018. Citado na página 17.

WU, Z. et al. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019. Citado 2 vezes nas páginas 12 e 18.

YAO, H. et al. Deep multi-view spatial-temporal network for taxi demand prediction. In: *Thirty-Second AAAI Conference on Artificial Intelligence*. [S.l.: s.n.], 2018. Citado na página 12.