Meta-Learning Framework with Applications to Zero-Shot Time-Series Forecasting

Boris N. Oreshkin[1], **Dmitri Carpov**[1], **Nicolas Chapados**[1], **Yoshua Bengio**[2]
[1]Element AI, [2]Mila
boris.oreshkin@gmail.com

Abstract

Can meta-learning discover generic ways of processing time series (TS) from a diverse dataset so as to greatly improve generalization on new TS coming from different datasets? This work provides positive evidence to this using a broad meta-learning framework which we show subsumes many existing meta-learning algorithms. Our theoretical analysis suggests that residual connections act as a meta-learning adaptation mechanism, generating a subset of task-specific parameters based on a given TS input, thus gradually expanding the expressive power of the architecture on-the-fly. The same mechanism is shown via linearization analysis to have the interpretation of a sequential update of the final linear layer. Our empirical results on a wide range of data emphasize the importance of the identified meta-learning mechanisms for successful zero-shot univariate forecasting, suggesting that it is viable to train a neural network on a source TS dataset and deploy it on a different target TS dataset without retraining, resulting in performance that is at least as good as that of state-of-practice univariate forecasting models.

1 Introduction

Time series (TS) forecasting is both a fundamental scientific problem and one of great practical importance. It is central to the actions of intelligent agents: the ability to plan and control as well as to appropriately react to manifestations of complex partially or completely unknown systems often relies on the ability to forecast relevant observations based on past history. Moreover, for most utility-maximizing agents, gains in forecasting accuracy broadly translate into utility gains; as such, improvements in forecasting technology can have wide impacts. Unsurprisingly, forecasting methods have a long history that can be traced back to the very origins of human civilization (Neale 1985), modern science (Gauss 1809) and have consistently attracted considerable research attention (Yule 1927; Walker 1931; Holt 1957; Winters 1960; Engle 1982; Sezer, Gudelek, and Ozbayoglu 2019). The applications of forecasting span a variety of fields, including high-frequency control (e.g. vehicle and robot control (Tang and Salakhutdinov 2019), data center optimization (Gao 2014)), business planning (supply chain management (Leung 1995), workforce and call center management (Chapados et al. 2014; Ibrahim et al. 2016), as well as such critically important areas as precision agriculture (Rodrigues Jr et al. 2019). In business specifically, improved forecasting translates in better production planning (leading to less waste) and less transportation (reducing CO_2 emissions) (Kahn 2003; Kerkkänen, Korpela, and Huiskonen 2009; Nguyen, Ni, and Rossetti 2010). The progress made in univariate forecasting in the past four decades is well reflected in the results and methods considered in associated competitions over that period (Makridakis et al. 1982, 1993; Makridakis and Hibon 2000; Athanasopoulos et al. 2011; Makridakis, Spiliotis, and Assimakopoulos 2018a). Recently, growing evidence has started to emerge suggesting that machine learning approaches could improve on classical forecasting methods, in contrast to some earlier assessments (Makridakis, Spiliotis, and Assimakopoulos 2018b). For example, the winner of the 2018 M4 competition (Makridakis, Spiliotis, and Assimakopoulos 2018a) was a neural network designed by Smyl (2020).

On the practical side, the deployment of deep neural time-series models is challenged by the cold start problem. Before a *tabula rasa* deep neural network provides a useful forecasting output, it should be trained on a large problem-specific time-series dataset. For early adopters, this often implies data collection efforts, changing data handling practices and even changing the existing IT infrastructures on a large scale. In contrast, advanced statistical models can be deployed with significantly less effort as they estimate their parameters on single time series at a time. In this paper we address the problem of reducing the entry cost of deep neural networks in the industrial practice of TS forecasting. We show that it is viable to train a neural network model on a diversified source dataset and deploy it on a target dataset in a *zero-shot regime*, i.e. without explicit retraining on that target data, resulting in performance that is at least as good as that of advanced statistical models tailored to the target dataset. We would like to clarify that we use the term "zero-shot" in our work in the sense that the number of history samples available for the target time series is so small that it makes training a deep learning model on this time series infeasible.

Addressing this practical problem provides clues to fundamental questions. Can we learn something general about forecasting and transfer this knowledge across datasets? If so, what kind of mechanisms could facilitate this? The ability to learn and transfer representations across tasks via