A Correlated Time Series Forecast System

Nicolaj Casanova Abildgaard, Casper Weiss Bang, Jonas Hansen, Tobias Lambek Jacobsen,
Thomas Højriis Knudsen, Nichlas Ørts Lisby, Chenjuan Guo, Bin Yang
Department of Computer Science, Aalborg University, Denmark
{nabild16, cbang16, jh13, tlja16, thkn16, nlisby16}@student.aau.dk; {cguo, byang}@cs.aau.dk

***Abstract*—In a cyber-physical system (CPS), different entities often interact with each other across time. With the development of various sensing technologies, the time-varying interactions among entities are often recorded as multiple, correlated time series. A typical CPS is a road transportation system, where the traffic on different road segments interact with each other. Traffic sensors are often deployed to capture travel speeds on different road segments, which results in multiple, potentially correlated, speed time series. Under this setting, an increasingly pertinent task is to forecast future speeds, which is essential in a wide variety of traffic planning scenarios. We present a system for correlated time series forecast. The system is able to employ different learning algorithms to perform correlated time series forecast, which facilities end users to choose the most appropriate algorithm for their specific service. The system is developed and integrated into *aSTEP*, a spatio-temporal data analytic platform developed by Aalborg University, and is tested using a wide variety of correlated time series data, including a user demand time series from a local mobility-as-a-service company.**

I. Introduction

With the continued development of sensing technologies, interactions among different entities over time are increasingly captured digitally, e.g., in the form of time series. For example, in a road transportation system, the speeds of different roads are captured by, e.g., loop detectors, as multiple speed time series [1]. The time series are often correlated with each other. For example, a traffic accident may influence multiple road segments' speeds. Analytics on such correlated time series have the potential to reveal holistic system dynamics of the underlying systems [2], [3]. An important analytics is forecasting, which can be used for identifying future trends and detecting outliers [4], [5].

A wide variety of learning algorithms are proposed for correlated time series forecasting [6], [7]. We demonstrate a system that aims to provide an easy-to-use interface for end users to train off-the-shelf machine learning models on different time series data sets and facilitate the end users to identify the most suitable algorithm for their specific application. In addition, the system works as a test bed which facilitates data scientists to develop new forecasting algorithms and compare them w.r.t. existing algorithms.

More specifically, the system provides multiple pre-trained machine learning models, where the users may upload their own time series data sets to make predictions. In addition, the system also provides means to facilitate users to first train a specific model based on their own time series data and then use the trained model to make predictions. The system provides a user interface to visualize the predictions with animation and provides a comprehensive evaluation interface where both the forecasting error at each timestamp and the average forecasting error over all timestamps are presented.

We test the system on a wide variety of time series, including both synthetic time series and real-world time series, e.g., traffic speeds, user demands, and bacteria activities. The system is developed base on *aSTEP*, a spatio-temporal data analytics platform developed at Aalborg University [8].

II. System overview

Figure 1 gives an overview of the system. We distinguish training and testing phases. In the training phase, a collection of multiple, correlated time series are given to the system as input. The system first normalizes the time series to make sure that all values are in the same range. Then the system provides a wide variety of forecasting algorithms, ranging from the recurrent neural network family [9] to the convolutional neural network family [10], [11], representing the two state-of-the-art algorithm families. After selecting a specific algorithm, the system trains a model based on the input time series. Finally, it outputs a trained model, which is saved as a docker image. As future work, training can be done in parallel [12].

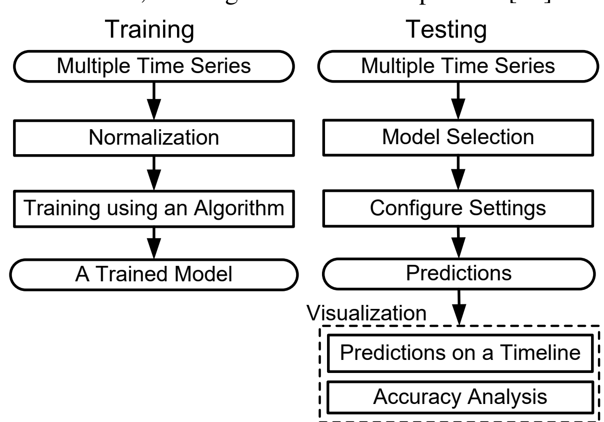

Fig. 1. System Overview

The testing phase uses a new collection of multiple, correlated time series. After selecting a trained model, the system allows a user to configure prediction settings, e.g., using historical n timestamps to predict future m timestamps. Then, the system employs the selected model to make predictions. Finally, the system visualizes the predictions as animations using a timeline to simulate that as time goes, the most recent n timestamps' data is used to make predictions for the next m timestamps. When providing the ground truth data, the system

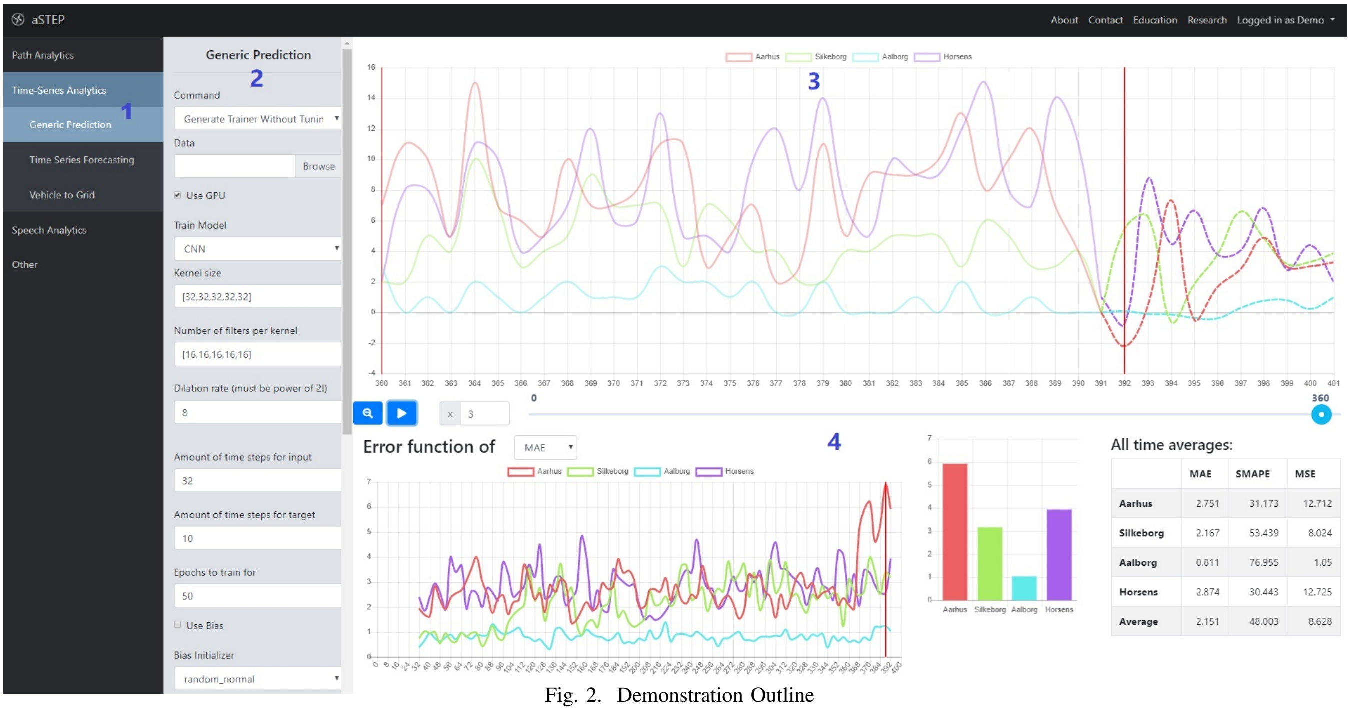

Fig. 2. Demonstration Outline

also offers a visual accuracy analysis, including an animation shows how the errors change across time using different error metrics and the statistics of the errors on different time series.

III. Demonstration Outline

We proceed to describe how users may interact with the system, which is integrated into aSTEP (https://astep.cs.aau.dk), a spatio-temporal data analytics platform developed by Aalborg University [8]. To try the time series forecasting system, in the left panel, click "Time Series Analytics" and then "Generic Prediction" (Label 1 in Figure 2). The middle panel (Label 2 in Figure 2) allows participants to configure the training or testing. More specifically, it allows participants to upload time series data, to select a specific forecasting algorithm, to configure algorithm-specific hyper-parameters, e.g., kernel sizes and dilation rates when a CNN is chosen, and to configure forecasting settings.

The right panel is split into an upper part (Label 3) and a lower part (Label 4). The upper part uses a timeline to show that as time goes, we use most recent historical data to make future predictions. The screenshot of Figure 2 shows a case where the system predicts user demands for four Danish cities, i.e., using four correlated time series, based on the data collected from a local mobility-as-a-server company. It is possible to remove specific time series from the animation so that the participants can focus on a subset of time series. The timeline can be zoomed in to focus the current prediction interval or zoomed out to show the whole time horizon. The speed of the animation can also be modified.

The lower part in the right panel provides a visual accuracy analysis. It is possible to choose different error functions such as MAE, RMSE, and MAPE, to evaluate the prediction accuracy. First, the error for each prediction at each timestamp is show. Second, the histogram show the prediction error at the current timestamp of the simulation in the upper part. Third, statistic over all predictions are summarized into a table.

Acknowledgements: This work was supported by Independent Research Fund Denmark under agreements 8022-00246B and 8048-00038B.

References

[1] J. Hu, B. Yang, C. S. Jensen, and Y. Ma, "Enabling time-dependent uncertain eco-weights for road networks," *GeoInformatica*, vol. 21, no. 1, pp. 57–88, 2017.

[2] B. Yang, C. Guo, and C. S. Jensen, "Travel cost inference from sparse, spatio-temporally correlated time series using markov models," *PVLDB*, vol. 6, no. 9, pp. 769–780, 2013.

[3] J. Hu, B. Yang, C. Guo, and C. S. Jensen, "Risk-aware path selection with time-varying, uncertain travel costs: a time series approach," *VLDB J.*, vol. 27, no. 2, pp. 179–200, 2018.

[4] T. Kieu, B. Yang, and C. S. Jensen, "Outlier detection for multidimensional time series using deep neural networks," in *MDM*, 2018, pp. 125–134.

[5] T. Kieu, B. Yang, C. Guo, and C. S. Jensen, "Outlier detection for time series with recurrent autoencoder ensembles," in *IJCAI*, 2019, pp. 2725–2732.

[6] R. Cirstea, D. Micu, G. Muresan, C. Guo, and B. Yang, "Correlated time series forecasting using multi-task deep neural networks," in *CIKM*, 2018, pp. 1527–1530.

[7] R. Cirstea, B. Yang, and C. Guo, "Graph attention recurrent neural networks for correlated time series forecasting." in *MileTS@KDD*, 2019.

[8] M. Beuchert, S. H. Jensen, O. A. Sheikh-Omar, M. B. Svendsen, and B. Yang, "aSTEP: aau's spatio-temporal data analytics platform," in *MDM*, 2018, pp. 278–279.

[9] J. Hu, C. Guo, B. Yang, C. S. Jensen, and H. Xiong, "Stochastic origin-destination matrix forecasting using dual-stage graph convolutional, recurrent neural networks," in *ICDE*, 2020, pp. 1417–1428.

[10] T. Kieu, B. Yang, C. Guo, and C. S. Jensen, "Distinguishing trajectories from different drivers using incompletely labeled trajectories," in *CIKM*, 2018, pp. 863–872.

[11] J. Hu, C. Guo, B. Yang, and C. S. Jensen, "Stochastic weight completion for road networks using graph convolutional networks," in *ICDE*, 2019, pp. 1274–1285.

[12] P. Yuan, C. Sha, X. Wang, B. Yang, and A. Zhou, "XML structural similarity search using mapreduce," in *WAIM*, 2010, pp. 169–181.