Forecasting Multivariate Time-Series Data Using LSTM and Mini-Batches

Chapter · January 2020

DOI: 10.1007/978-3-030-37309-2_10

4 authors, including:

Athar Khodabakhsh
Hasso Plattner Institute

Ismail Ari
Ozyegin University

Mustafa Bakır
Gebze Technical University

Forecasting Multivariate Time-Series Data using LSTM and Mini-batches *

Athar Khodabakhsh[1], Ismail Ari[1], Mustafa Bakir[2], and Serhat Murat Alagoz[2]

[1] Department of Computer Science
Ozyegin University, Istanbul, Turkey
athar.khodabakhsh@ozu.edu.tr, ismail.ari@ozyegin.edu.tr
[2] Software Development Department
TUPRAS, Kocaeli, Turkey
{mustafa.bakir,serhatmurat.alagoz}@tupras.com.tr

Abstract. Multivariate time-series data forecasting is a challenging task due to non-linear interdependencies in complex industrial systems. Thus, it is crucial to model the dependencies which can be obtained using Recurrent Neural Networks (RNNs) that are building blocks of Long Short-Term Memory (LSTM) networks. The ability of neural networks to learn features by extraction of spatial relationships was successful and taking advantage of this ability for non-spatial data is desirable. In this paper, we converted non-spatial multivariate time-series data into a time-space format and used stacked layers of LSTM for sequential analysis of multi-attribute industrial data for prediction. We compared the effect of mini-batch length and attribute numbers on prediction accuracy and learned the importance of spatio-temporal locality for detecting patterns using LSTM.

Keywords: LSTM · Multivariate time-series · RNN · Sensors · Sequence Data · Time-series.

1 Introduction

Industrial IoT (IIoT) devices collect data from complex physical devices and instruments which have time-varying and non-linear behavior. Forecasting the future is an important task that becomes possible by careful analysis of short and long-term data. The forecasting is more accurate when the dependencies between variables are better modeled [1]. In learning methods, we desire the models to learn dependencies automatically by observing the past data to predict future time-series. These methods are gaining attention for industrial applications in training non-linear models in large dimensions over fast flowing data and large historical datasets. RNNs and LSTM are now proven to be effective in processing time-series data for prediction [2].

* This research was sponsored by a grant from TUPRAS (Turkish Petroleum Refineries Inc.) R&D group.

For multivariate time-series prediction several Deep Learning architectures are used in different domains such as stock price forecasting [3], text and video processing [4], weather and extreme event forecasts [5]. In many applications, the high-dimensional data has high correlation among dimensions and these correlations are spatially located close to each other that consequently get reflected in deep neural networks for local processing [6]. For non-spatial data like time-series, the relationship and correlations among measurements can be exploited by sequence analysis which is traditionally applied by sliding-window approach. Industrial applications of these analysis can be fault detection [7], automated control, and predictive maintenance [8].

In all industries including retail and Oil & Gas, there is a need to forecast input (e.g. crude oil) supply needs, depending on the current output (e.g. gasoline, diesel, etc.) amounts and market demands. Depending on their forecasts, refineries can make future contracts and reduce their uncertainties. In these mission critical businesses, thousands of sensors are installed around physical systems. Distributed Control Systems (DCS) and Supervisory Control and Data Acquisition (SCADA) systems measure flow, pressure, temperature of turbines, pumps, and injectors. Achieving continuous safety, process efficiency, long-term durability, and planned (vs. unplanned) downtimes are among the main goals for industrial plant management. These controls and actions should be performed in real-time according to temporal patterns received from stream data.

Since most industrial systems are dynamic and the relation between features are complex, dynamic, and non-linear, the quality of models and predictions are dependent on current context of the system [9]. LSTM can be used for sequence processing tasks and directly improve the process performance, and increase profits as well as safety. In this paper, we used time-series data from the petrochemical plant of a real oil refinery with approximately 11.5 million $ton/year$ processing capacity [10].

2 Background and Related Work

Analysis of time-series data have been a subject of interest for scientific and industrial fields. They are used for knowledge extraction, classification, modeling, and prediction of time-varying systems. Depending on the context of data different linear and non-linear modeling techniques are applicable. Linear models such as Auto Regressive Moving Average (ARMA) [11] make short-term predictions. However, extracting long-term dependencies are also demanded while mining historical data. Utilizing NNs and networks with memory such as RNNs and LSTM provides ability to process temporal patterns in addition to long-term dependencies. Lai, et al., [1] proposed a novel framework called LSTNet that uses the Convolutional Neural Network (CNN) and RNN to extract short-term local dependency patterns among variables and to discover long-term patterns for time-series trends. Jiang, et al., [3] used RNNs and LSTM for time-series prediction of stock prices. Loganathan, et al., [12] used LSTM for multi-attribute sequence-to-sequence (Seq2Seq) model for anomaly detection in network traffic.

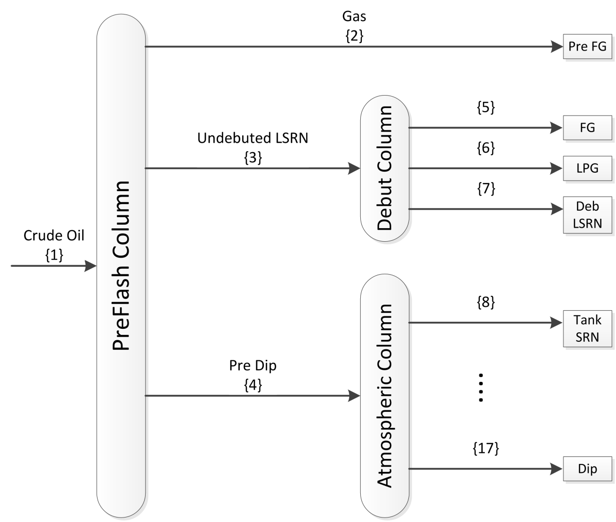

Fig. 1. Tupras petrochemical plant columns for processing crude oil.

Gross, et al, [6] interpreted time-series as space-time data for power price prediction. In our previous work [13] we used ARMA for modeling the short-term dependencies of features for error detection and in this study we investigate the effect of long-term dependencies on predictions to improve our models with another dimension for multi-mode analysis in real-time.

3 Methodology

For capturing the dependencies and extracting long-term patterns in time-series data, we used stacked LSTM networks which are designed to handle the problem of learning "long short-term" dependencies. For demonstration, we obtained time-series data from a real petrochemical plant and applied LSTM for predicting crude oil purchase amount.

As depicted in Figure 1, our simplified plant model has 17 flow sensors over 3 main branches of material flows and the corresponding sensor data streams. Crude oil columns take the oil as input and deliver several by-products such as liquid propane gas, fuel oil, kerosene, diesel, and asphalt. A preflash unit reduces the pressure and provides the first vaporization, where the vapor goes to a debutanizer for distillation and the liquid mix goes to an atmospheric column for separation. This time-series data is time-framed such that the measurements at current time t is predicted by given measurements of the by-products from prior time step. In most of the current studies the focus is on the neural network structure, whereas in this study we investigate the effect of memory size and importance of local sequence analysis on training the network and prediction

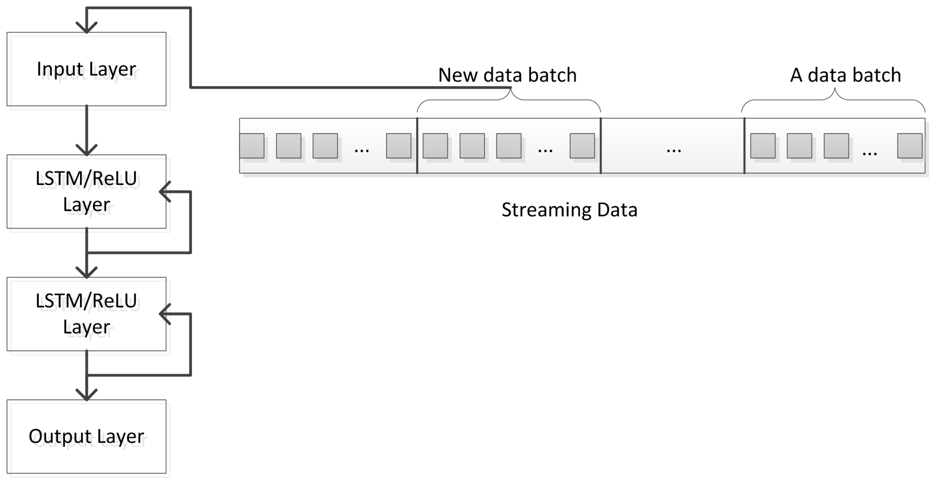

Fig. 2. Stacked architecture of LSTM networks used for supply prediction. The time-series data are transformed into spatial data in mini-batches that consist of multivariate sensor data in each box.

accuracy of future values. In our previous study [14] we managed to identify operational modes by investigating the changing patterns observed in time-varying systems. The relation between attributes change over time and it is important to react to this change to update the model. The challenge is to decide how many steps to look back into prior data.

3.1 Problem Formulation to Define and Fit LSTM

We converted non-spatial multivariate time-series sensor data into time-space frames (similar to pictures in a movie) and trained model for sequence prediction of industrial sensor data using LSTM. Each mini-batch consists of multivariate sensor data that is received consecutively. Rows of data are then transformed to a time-space by adding current data to the sequence in a given time on top of prior data building the mini-batches. The learning network consists of two LSTM layers, with $ReLU$ (Rectified Linear Unit) Activation Function and *Dense* layer as shown in Figure 2. This network is then used for unsupervised modeling which can learn long-term correlated features.

In multivariate time-series forecasting, given the series $X = \{x_1, x_2, \ldots x_{t-1}\}$, where x_i represents values at time i, the task is to predict value of x_t. For predictions we used $\{x_{t-w}, x_{t-w+1}, \ldots x_{t-1}\}$ where w is the window size. These sequences of mini-batches are then fed into a two-layer LSTM network in n epochs for training and for p step ahead predictions. The dataset must get split into training and testing sets. The network is trained with Adam backpropagation on mini-batches.

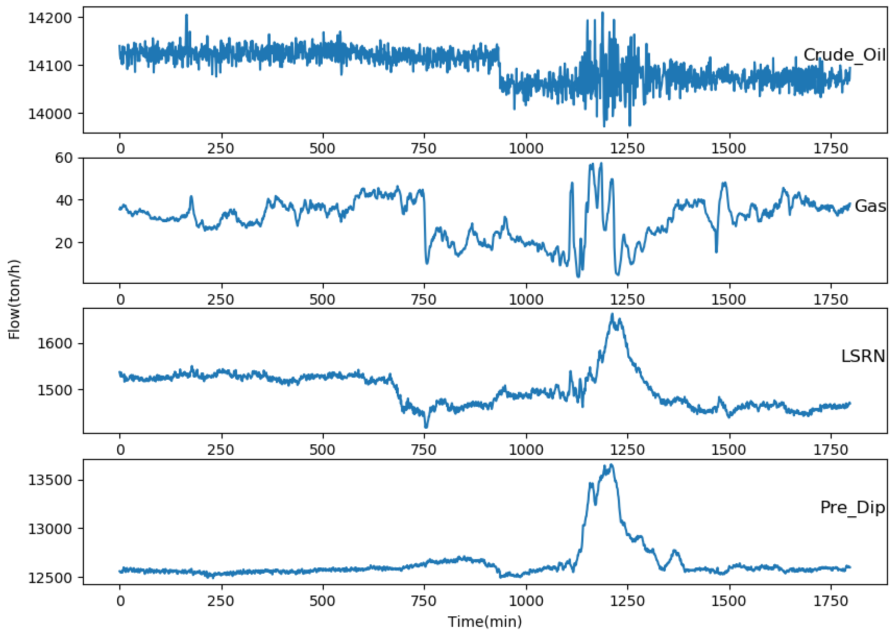

Fig. 3. Flow rates (ton/h) of Crude Oil and three main branches of by-products including Propane Gas, LSRN and, Pre Dip that show correlated and dynamic behavior of Petrochemical plant's production.

4 Results

We trained the LSTM on the multivariate data for time-series forecasting with using Tensorflow [15] in Python. The model is trained over 6 days of measurements and tested over 30 minutes of data and according to the Root Mean Squared Error (RMSE) values of every prediction the model is retained. The size of LSTM input layer is set to be 70 for sequence processing. The Mean Square Error (MSE) loss function and the efficient Adam version of stochastic gradient descent [16] is used in the LSTM model. The first LSTM layer is trained and the output of this sequence analysis are fed into second layer of our model which is another LSTM layer. The input shape is 1 time step with 2, 7, and 17 attributes. The model is applied for 50 training epochs with different batch sizes for comparison. The loss of test and train are evaluated by setting the validation data while fitting the model. After fitting the model, the forecast is obtained for test dataset. Comparing the forecast and actual values in original scale, the RMSE value of the model is calculated. In Figure 3 a fraction of crude oil data is depicted that is used for training the LSTM network. This dataset contains flow rates of crude oil measurement as input and outputs of the plants for processed by-products of 3 main branches of petrochemical plant.

We compared the effect of batch size on prediction results. The RMSE value of predictions is evaluated for 3 batch sizes of 90, 180, 360 minutes over 2, 7, and 17 attributes. As shown in Figure 4 larger number of attributes improve the prediction results whereas, smaller batch sizes result in lower RMSE values. One exception is the increase in RMSE for the smallest 90 minute batch size when the number of attributes increased from 2 to 7. This can be attributed to the increase in complexity of the system (higher dimensions) without giving the model enough data to match this complexity. The rest of the plot justifies and supports our explanation.

Although the training data is the same for all the mini-batches, the prediction results are different due to the memory of the network. Figure 4 shows trade-offs between batch size and number of features. Although, smaller batch sizes may result in smaller RMSE value, larger number of attributes improves the accuracy of prediction. This shows the importance of locality in sequential multivariate time-series forecasting problems that want to utilize LSTM networks.

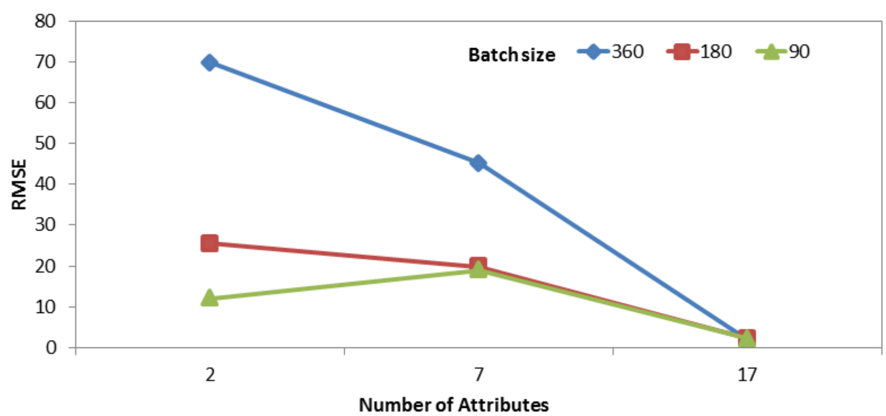

Fig. 4. Effect of batch size on RMSE prediction in LSTM network.

5 Conclusions and Future Work

In this paper we show the trade-offs between batch size and number of features affecting the prediction results of multivariate industrial sensor data analysis. We also show how a time-series dataset can be transformed to a format that is usable in LSTM time-series (i.e. deep learning) forecasting. The spatial relation between measurements values of time-series data is studied by sequence analysis using 2 layered LSTM network. The network learns features from prior raw data for predicting the future value which is critical for industrial supply forecasting. Specifically, we learned the importance of spatio-temporal locality for detecting patterns using LSTM networks. In our future work we will use LSTM predicted values for error detection and classification.

6 Acknowledgments

We would like to thank Burak Aydogan and Mehmet Aydin for collecting and providing us with TUPRAS Oil & Gas Refinery Sensor data.

References

1. Lai, G., Chang, W. C., Yang, Y., Liu, H.: Modeling long-and short-term temporal patterns with deep neural networks. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 95–104, Jun (2018). ACM.
2. Langkvist, M., Karlsson, L., Loutfi, A.: A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters, vol. 42, pp. 11–24 (2014).
3. Jiang, Q., et al.: Stock price forecast based on LSTM neural network. International Conference on Management Science and Engineering Management. Springer, pp. 393-408, (2018).
4. Varol, G., Laptev, I., Schmid, C.: Long-term temporal convolutions for action recognition. IEEE transactions on pattern analysis and machine intelligence. vol. 40(6), pp. 1510–1517, Jun (2018).
5. Laptev, N., Yosinski, J., Li, L. E., Smyl, S.: Time-series extreme event forecasting with neural networks at uber. In: International Conference on Machine Learning, No. 34, pp. 1–5, (2017).
6. Groß, W., Lange, S., Bödecker, J., Blum, M.: Predicting time series with space-time convolutional and recurrent neural networks. In: Proceeding of the 25th ESANN, pp. 71–76 (2017).
7. Lee, K.B., Cheon, S., Kim, C.O.: A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. IEEE Transactions on Semiconductor Manufacturing, 30(2), pp.135–142, (2017).
8. Troiano, L., Villa, E. M., Loia, V.: Replicating a Trading Strategy by means of LSTM for Financial Industry Applications. IEEE Transactions on Industrial Informatics, (2018).
9. Shih, S. Y., Sun, F. K., Lee, H. Y.: Temporal pattern attention for multivariate time series forecasting. arXiv preprint arXiv:1809.04206, (2018).
10. TUPRAS Refinery. [Online] Available: http://tupras.com.tr/en/rafineries. Accessed: Dec. 6 (2018).
11. Box, G. E., et al.: Time series analysis: forecasting and control. John Wiley & Sons, (2015).
12. Loganathan, G., Samarabandu, J., Wang, X.: Sequence to sequence pattern learning algorithm for real-time anomaly detection in network traffic. In IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), pp. 1–4, May (2018).
13. Khodabakhsh, A., Ari, I., Bakir, M., Ercan, A.O.: Multivariate Sensor Data Analysis for Oil Refineries and Multi-mode Identification of System Behavior in Real-time. IEEE Access vol. 6, Oct (2018).

14. Khodabakhsh, A., Ari, I., Bakir,M., Alagoz, S. M.: Stream analytics and adaptive windows for operational mode identification of time-varying industrial systems. In Proceeding IEEE International Congress Big Data (BigData Congr.), pp. 242–246, Jul. (2018).
15. Abadi, M., et al.: Tensorflow: a system for large-scale machine learning. OSDI. Vol. 16, (2016).
16. Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).