Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 5 (2022) 855 - 859 ISSN Media Electronic: 2580-0760

Intermittent Demand Forecasting Using LSTM With Single and Multiple Aggregation

Fityan Azizi[1], Wahyu Catur Wibowo[2]
[1,2] Fakultas Ilmu Komputer, Universitas Indonesia
[1]fityan.azizi@ui.ac.id, [2]wibowo@cs.ui.ac.id

Abstract

Intermittent demand data is data with infrequent demand with varying number of demand sizes. Intermittent demand forecasting is useful for providing inventory control decisions. It is very important to produce accurate forecasts. Based on previous research, deep learning models, especially MLP and RNN-based architectures, have not been able to provide better intermittent data forecasting results compared to traditional methods. This research will focus on analyzing the results of intermittent data forecasting using deep learning with several levels of aggregation and a combination of several levels of aggregation. In this research, the LSTM model is implemented into two traditional models that use aggregation techniques and are specifically used for intermittent data forecasting, namely ADIDA and MAPA. The result, based on tests on the six predetermined data, the LSTM model with aggregation and disaggregation is able to provide better test results than the LSTM model without aggregation and disaggregation.

Keywords: *Intermittent Demand, LSTM, ADIDA, MAPA*

1. Introduction

Intermittent demand data has infrequent occurrences of demand and the size of the demand varies when a request occurs[1]. This demand data occurs on a discontinuous basis. That is, at one time it will be worth more than one if there is a demand and is zero if there is no demand. Intermittent demand patterns can be found in the Stock Keeping Unit (SKU) application, generally found in the aviation, automotive, military, and information technology sectors[2]. Intermittent demand forecasting is useful for providing inventory control decisions. So, a more accurate demand forecast is needed.

There are several traditional methods developed specifically to solve the intermittent demand forecasting problem, the first to appear is known as the Croston method[3], followed by the emergence of other methods such as Syntetos-Boylan Approximation (SBA)[4], and Aggregate-Disaggregate Intermittent Demand Approach (ADIDA)[2].

Furthermore, in the last few years, several studies have been carried out related to intermittent demand forecasting using the Deep Learning (DL) approach, specifically using Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), and Long Short Term Memory (LSTM). MLP is a feed forward artificial neural network which consists of input, output, and several hidden layers. RNN is a DL architecture specifically designed to handle time series data and sequential data. The difference between RNN and MLP is that each unit in the RNN receives input data not only from the output in the previous layer, but also receives output from the same neuron at the previous time[5]. LSTM is an extension of RNN which has a technique to overcome the vanishing gradient problem. The LSTM unit consists of a cell, input gate, forget gate, and output gate. The cell remembers values in some time interval and the other three gates regulate the information flow associated with the cell[6]. In the research that has been done related to intermittent demand forecasting using DL, Kiefer et al. [5] implements MLP and LSTM methods and compares with some traditional methods. As a result, the forecasts generated by deep learning are not able to give better results than the Croston method. Furthermore, Muhaimin et al. [7] used MLP and RNN to perform intermittent demand forecasting and got no better results when compared to the traditional method. However, if 7 days of aggregation of the data is performed before forecasting, the two deep learning architectures get slightly better results than traditional methods. Both studies use the same dataset, namely using the M5 competition dataset, which is a time series

Accepted: 19-08-2022 | Received in revised: 16-10-2022 | Published: 31-10-2022

data forecasting competition and focuses on intermittent data series [8]. Figure 1 is an example of demand data for one of the products in the M5 dataset. The image displayed shows the demand data for the first 500 days. The M5 dataset contains sales data for 30490 products with a span of 1941 days.

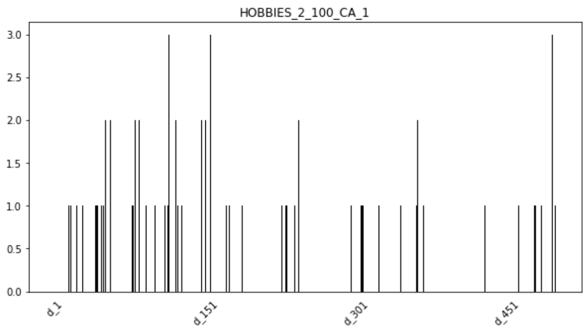

Figure 1. Example of demand product in M5 Dataset

Based on previous research, DL models, especially MLP and RNN-based architectures, have not been able to provide better intermittent data forecasting results than the Croston method. However, research conducted by Muhaimin et al. [7] shows that using aggregation can improve the accuracy of the DL model on the results of intermittent data forecasting.

Aggregating data before forecasting is one of the steps in the ADIDA method[2]. In this method, the data that has been disaggregated and forecasted is disaggregated to return the value in the actual time range. In developing the ADIDA method, a method called Multiple Aggregation Prediction Algorithm (MAPA) has been developed. MAPA combines forecast results from several different levels of aggregation, by utilizing information from a number of aggregation levels, this method is considered to be able to improve the accuracy of the forecasting model[9].

In this study, intermittent demand forecasting will be carried out using LSTM with multiple aggregation by implementing the MAPA method and single aggregation by implementing the ADIDA method.

2. Research Methods

2.1 Data Collecting and Data Preprocessing

The dataset used in this study uses the M5 dataset taken from kaggle. This dataset contains data on sales of Walmart products in 3 regions in the United States which are divided into 3 categories (hobbies, food, and household appliances) and contains sales records of 30490 products throughout 1941 days. The data was created to be used in the M5 competition, which is a forecasting competition that focuses on intermittent data[8]. In this study, the data used as observation data were taken at random as much as two for each category, so that there were six data used for observation.

Table 1. Observation Data

ID	Product Name
12560	HOBBIES_1_373
9682	HOBBIES_2_120
13209	HOUSEHOLD_1_458
16426	HOUSEHOLD_2_085
7712	FOODS_1_003
14324	FOODS_2_302

Table 1 shows the product name and product id used for observation. Data selection is done randomly.

Furthermore, the aggregation levels used in this study were 7, 14, and 28 days. The three levels will be used for both LSTMs with single aggregation and LSTMs with multiple aggregations.

Figure 2 is an illustrative example of performing level three aggregation. At the top of the figure, there is data with nine timesteps. In performing level three aggregation, we add up the data every three timesteps, so that in the bottom image, we get data with three timesteps containing the sum of every three timesteps in the data at the top of the image. From the aggregation results, we get data that contains no zero values.

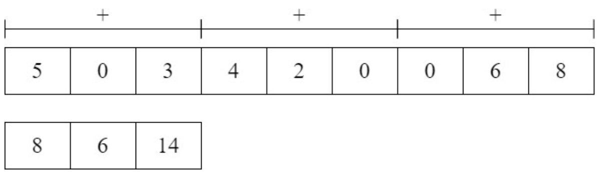

Figure 2. An example of aggregating time series data using three levels of aggregation..

2.2 LSTM

To achieve good performance and results, There are parameters that need to be set for the LSTM model, in this study, we use two hidden layers and 64 neurons in one layer.

Table 2. Hyperparameters used for Train model

ID	Product Name
Epoch	100
Learning Rate	0.001
Loss Function	Mean Squared Error
Optimizer	Adam

Table 2 shows the hyperparameters used to train the model both for LSTM models without aggregation, with single aggregation, and with multiple aggregations. To train the LSTM model, the data needs to be transformed into a lag or sliding window form. Each value in the lag represents one timestep. In this study, the lag length for the LSTM model without aggregation is 28 and the lag length for the LSTM model with aggregation is 20.

As an example of transforming time series data into a sliding window or lag, Figure 3 is an illustration of the application of a sliding window with a length of three to generate training data for LSTM. In the illustration,

DOI: https://doi.org/10.29207/resti.v6i5.4435
Creative Commons Attribution 4.0 International License (CC BY 4.0)

there are time series data with six timesteps $(x_1, x_2, x_3, x_4, x_5, x_6)$.

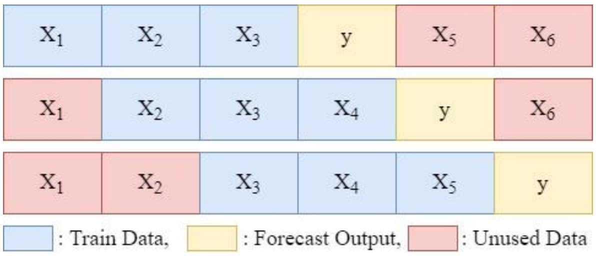

Figure 3. An example of transforming time series data into a sliding window.

Based on these data, if we look at the image at the top, we get the first data for training, namely $X = (x_1, x_2, x_3)$ with a blue mark, y (output data) with a yellow mark, obtained from x_4, and the data marked in red is not used. For the second data, we get $X = (x_2, x_3, x_4)$ where y is taken from x_5. For the third data, $X = (x_3, x_4, x_5)$ and y are obtained from x_6.

2.3. LSTM With Single Aggregation

In implementing the LSTM with a single aggregation, all data with the specified aggregation level is used as training data for the LSTM model. To perform tests or perform forecasting, the data generated by the model is disaggregated or returned to its original frequency according to the specified aggregation level.

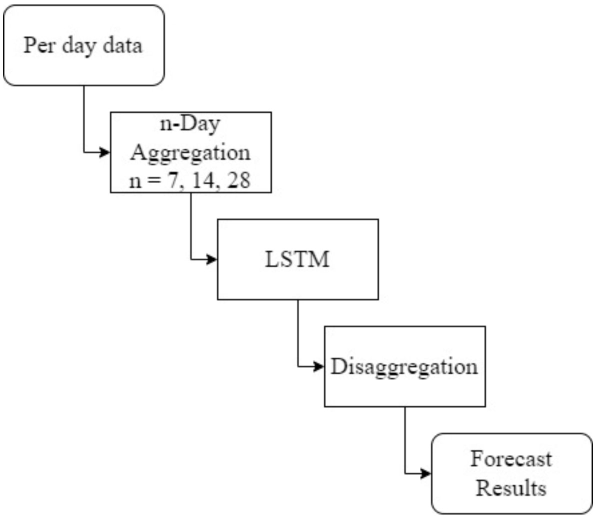

Figure 4. LSTM with a single aggregation, implements LSTM into the ADIDA method.

Figure 4 is an illustration of an LSTM with a single aggregation. By implementing LSTM to ADIDA, first we have the original time series data (data per day), then the data is aggregated with predetermined levels (7, 14, and 28 days). The data that has been aggregated is then transformed into a sliding window in order to obtain training data for the LSTM, to get the forecasting results from the LSTM, the forecasting results are disaggregated by dividing the data at each timestep with a predetermined level of aggregation. For example, if level seven aggregation is performed, then the disaggregation is done by dividing each value in the timestep by seven.

2.4 LSTM With Multiple Aggregation

In contrast to single aggregation, LSTMs with multiple aggregations implements LSTM into MAPA. The model get forecasting results based on a combination of forecasting results from several levels of aggregation and results without aggregation. In this study, the level of aggregation used was the same as the level of aggregation determined in LSTM with single aggregation, which are 7, 14, and 28 days. Thus, LSTM with multiple aggregations combines four forecasting results, which are results without aggregation, 7-day aggregation results, 14-day aggregation results, and 28-day aggregation results. The combination is done by calculating the average value of the four forecasting results.

Figure 5 is an illustration of the implementation of LSTM with multiple aggregation. By implementing LSTM to MAPA, we combine the results of four LSTM models with different aggregation values. The first model is an LSTM model without aggregation, the second is an LSTM with level seven aggregation, the third is an LSTM with level 14 aggregation, and the last is an LSTM with level 28. To get forecasting results from an LSTM with aggregation, the same method is used with LSTM with a single aggregation. as in 2.3, where data is aggregated according to the specified level and transformed into a sliding window, then LSTM training is carried out and the forecasting results are disaggregated into the original form. Then, the four forecasting results that have been obtained are then combined by calculating the average value of the forecasting results of the four models.

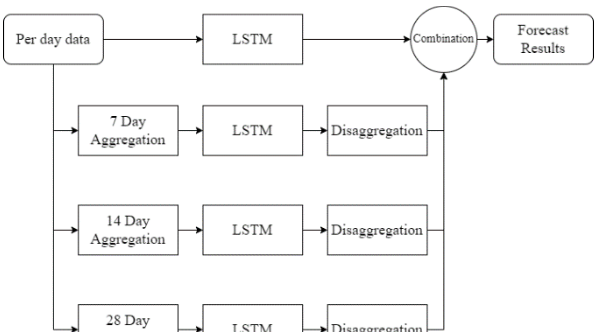

Figure 5. LSTM with multiple aggregation, implements LSTM into the MAPA method.

2.5 Evaluation Metrics

To evaluate, the model will be measured using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Stock-keeping-oriented Prediction Error Costs (SPEC). MAE will calculate the absolute average

DOI: https://doi.org/10.29207/resti.v6i5.4435
Creative Commons Attribution 4.0 International License (CC BY 4.0)

value of the errors obtained based on the results of forecasting the test data and has the following equation:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (1)$$

Where y_i is the forecast result, x_i is the true value, and n is the total of the data. RMSE will calculate the root value of the average square of the errors obtained and has following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \quad (2)$$

The notation is the same as MAE[10]. SPEC is a new measurement that is specifically created to evaluate the forecasting results of intermittent data. SPEC has the following equation:

$$SPEC_{\alpha_1\alpha_2} = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{t}\left(\max\left[0; \min\left[y_i; \sum_{k=1}^{i} y_k - \sum_{j=1}^{t} f_j\right] \cdot \alpha_1; \min\left[f_i; \sum_{k=1}^{i} f_k - \sum_{j=1}^{t} y_j\right] \cdot \alpha_2\right] \cdot (t - i + 1)\right) \quad (3)$$

With n is the length of time series, y_t is the demand at time t, f_t is the forecast result, α_1 is the opportunity cost, and α_2 is the stock-keeping cost. α_1 and α_2 are hyperparameters that need to be set. We use $\alpha_1 = 0.75$ and $\alpha_2 = 0.25$ because Martin et al. shows that these values are effectively used to evaluate demand forecast results[11].

Furthermore, the model was also compared with the baseline Croston model and the LSTM without aggregation.

3. Results and Discussions

In this study, the implementation of the LSTM model with single aggregation and multiple aggregation was carried out on the three data specified in Table 2. In evaluating the implementation of this study, the measurements used to evaluate forecasting results are MAE, RMSE, and SPEC. Furthermore, the two models were compared with other models, namely the LSTM model without aggregation and the Croston method. Tables 3 to 8 show the results of the implementation of the LSTM model with single aggregation and multiple aggregation and each model is named LSTM-SA-n for single aggregation (n is the aggregation level) and LSTM-MA for multiple aggregation.

Table 3. Performance Comparison on ID 12560.

Model Name	MAE	RMSE	SPEC
Croston	0.8092	0.9013	4.6210
LSTM	0.7579	0.8522	2.5762
LSTM-SA-7	0.7034	0.8271	0.5847
LSTM-SA-14	**0.6955**	**0.8254**	0.7293
LSTM-SA-28	0.7022	0.8265	0.5820
LSTM-MA	0.7020	0.8262	**0.5812**

Table 4. Performance Comparison on ID 9682.

Model Name	MAE	RMSE	SPEC
Croston	0.3481	0.3786	3.0691
LSTM	0.3045	0.3606	1.4492
LSTM-SA-7	0.3245	0.3673	2.0961
LSTM-SA-14	0.3057	0.3601	1.5111
LSTM-SA-28	**0.3010**	**0.3586**	**1.3534**
LSTM-MA	0.3051	0.3604	1.4802

Table 5. Performance Comparison on ID 13209.

Model Name	MAE	RMSE	SPEC
Croston	0.8162	1.0057	3.3345
LSTM	0.7682	0.9999	0.5236
LSTM-SA-7	0.7188	0.9591	0.6815
LSTM-SA-14	**0.7162**	0.9632	0.9371
LSTM-SA-28	0.7230	**0.9565**	**0.4643**
LSTM-MA	0.7222	0.9610	0.6271

Table 6. Performance Comparison on ID 16426.

Model Name	MAE	RMSE	SPEC
Croston	0.3360	**0.5185**	**5.5343**
LSTM	0.3408	0.5245	6.0672
LSTM-SA-7	0.3325	0.5201	6.0358
LSTM-SA-14	0.3327	0.5198	5.8939
LSTM-SA-28	0.3276	0.5215	6.6685
LSTM-MA	**0.3317**	0.5207	6.2025

Table 7. Performance Comparison on ID 7712.

Model Name	MAE	RMSE	SPEC
Croston	1.0244	**1.2545**	1.3397
LSTM	1.0385	1.2716	1.1374
LSTM-SA-7	1.0008	1.2620	1.9527
LSTM-SA-14	**1.0006**	1.2546	0.9766
LSTM-SA-28	1.0335	1.2554	1.5154
LSTM-MA	1.0052	1.2604	**0.9623**

Table 8. Performance Comparison on ID 14324.

Model Name	MAE	RMSE	SPEC
Croston	0.7714	0.8901	6.7788
LSTM	0.7104	0.8172	2.5748
LSTM-SA-7	0.6930	0.8111	2.5141
LSTM-SA-14	**0.6808**	**0.7737**	**1.3136**
LSTM-SA-28	0.6786	0.7968	2.5060
LSTM-MA	0.6826	0.7899	2.1548

Table 3 shows the comparison of forecasting results on data 12560, with the LSTM-SA-7 model providing the lowest MAE and RMSE values. The LSTM-MA model gives the lowest SPEC value followed by LSTM-SA-28 with the difference between the two results of the two models not being too significant. In Table 4, the forecasting results on data 9682 show that the LSTM-SA-28 model gives the best results on the overall evaluation measurement. Furthermore, the forecasting results on data 13209 in Table 5, show that LSTM-SA-28 gives the lowest results on RMSE and SPEC measurements, and LSTM-SA-14 gives the lowest MAE results. Table 6 shows the forecasting results of data 16426, the best results on the RMSE and SPEC measurements were produced by the Croston method, with LSTM-MA giving the best MAE results.

DOI: https://doi.org/10.29207/resti.v6i5.4435
Creative Commons Attribution 4.0 International License (CC BY 4.0)

Forecasting results on data 7712 in Table 7 show that the best results for each measurement are produced by different models. The best results on MAE were produced by LSTM-SA-14, the best RMSE was produced by the Croston method, and the best SPEC results were produced by LSTM-MA. Finally, Table 8 shows the forecasting results on data 14324, with LSTM-SA-14 giving the best results on all evaluation measurements.

The results of forecasting on the three data indicate that doing aggregation and disaggregation in forecasting using LSTM is able to give better results than forecasting using LSTM without aggregation and disaggregation. This is shown in Tables 3 to 8, where the LSTM without aggregation did not produce any of the best results in the evaluation measurement on the six data. However, even though using aggregation and disaggregation can give better results, it cannot be concluded that the best type and level of aggregation among the four models with aggregation and disaggregation, using either single aggregation or multiple aggregation, as well as the level of aggregation that has been set. The best results on the three data were generated by different types and levels of aggregation.

4. Conclusion

In this study, intermittent demand forecasting was carried out using the implementation of single aggregation and multiple aggregation on LSTM model based on the ADIDA and MAPA methods. Based on the tests on the six data that have been determined, the LSTM model with aggregation and disaggregation is able to provide better test results than the LSTM model without aggregation and disaggregation. Furthermore, it is not possible to determine the best aggregation level in LSTM with single aggregation. This is because the best results in the third test of data were produced by different levels of aggregation.

Reference

[1] B. Pochiraju and S. Seshadri, Essentials of Business Analytics, Springer, 2019

[2] K. Nikolopoulos, A. A. Syntetos, J. E. Boylan, F. Petropoulos and V. Assimakopoulos, "An aggregate–disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis," *Journal of the Operational Research Society*, vol. 62, p. 544–554, 2011.

[3] J. Croston, "Forecasting and Stock Control for Intermittent Demands," *Operational Research Quarterly*, vol. 23, pp. 289-303, 1972.

[4] A. Syntetos and J. Boylan, "On the bias of intermittent demand estimates," *International Journal of Production Economics*, vol. 71, pp. 457-466, 2001.

[5] D. Kiefer, F. Grimm, M. Bauer and C. V. Dinther, "Demand Forecasting Intermittent and Lumpy Time Series: Comparing Statistical, Machine Learning and Deep Learning Methods," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, Hawaii, 2021.

[6] O. B. Sezer, M. U. Gudelek and A. M. Ozbayoglu, "Financial time series forecasting with deep learning : A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, 2020.

[7] A. Muhaimin, D. D. Prastyo and H. H.-S. Lu, "Forecasting with Recurrent Neural Network inIntermittent Demand Data Forecasting with Recurrent Neural Network in Intermittent Demand Data," in *11th International Conference on Cloud Computing, Data Science and Engineering*, Noida, 2021.

[8] S. Makridakis, E. Spiliotis and V. Assimakopoulos, "The M5 competition: Background, organization, and implementation," *International Journal of Forecasting*, 2021.

[9] N. Kourentzes, F. Petropoulos and J. R. Trapero, "Improving forecasting by estimating time series structural components across multiple frequencies," *International Journal of Forecasting*, vol. 30, pp. 291-302, 2014.

[10] J.M. González-Sopeña, V. Pakrashi and B. Ghosh, "An overview of performance evaluation metrics for short-term statistical wind power forecasting," Renewable and Sustainable Energy Reviews, vol. 138, 2021.

[11] D. Martin, P. Spitzer and N. Kuhl, "A New Metric for Lumpy and Intermittent Demand Forecasts: Stock-keeping-oriented Prediction Error Costs," *CoRR*, vol. abs/2004.10537, 2020.

DOI: https://doi.org/10.29207/resti.v6i5.4435
Creative Commons Attribution 4.0 International License (CC BY 4.0)