The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Improvement of Product Shipment Forecast based on LSTM Incorporating On-Site Knowledge as Residual Connection

Takeshi Morinibu Tomohiro Noda Shota Tanaka

Daikin Industries, Ltd. Technology and Innovation Center

It is important to predict shipments for the purpose of making a production plan of air conditioners. Although ARIMA was used for that prediction for a long time, it turned out that some products we manage had less accurate prediction score. In order to get more precise prediction, we applied LSTM to forecast shipments. Despite the complexity of LSTM, we could not get what we expected. Therefore, we further improved the accuracy by adding on-site knowledge to network structure of LSTM as residual mechanism.

1. Introduction

In order to make profit efficiently in the manufacturing industry, it is necessary to minimize the warehouse cost due to excessive inventory and the missed selling caused by sales suspension due to out of stock. Since predicted values of shipment quantity are used to make a production plan for managing that stock, better prediction accuracy is required.

ARIMA is widely used as a traditional method for time series prediction including shipment quantity. For example, there is an application for wholesale of vegetables [Shukla 13]. Even in the case of air conditioners, an ARIMA was able to make good predictions in many cases. The word "series" we describe is a group of products summarized with certain features. When we conducted a survey, however, we found that there are some series with extremely low precision among major series. We attempted to improve it.

In recent years, application of RNN to various time series data has been widespread and its superiority has been confirmed [Li 18], so we tried first to improve accuracy by using a plain LSTM. Even so, the prediction accuracy of the LSTM was not much better than the ARIMA. Hence we challenge to alter the structure of the plain LSTM to solve the problem. The idea to realize it comes from ResNet.

ResNet recorded amazing precision in the field of image recognition by using Residual block [He 16]. It ensures an information flow, leading to optimizing loss function for very deep structure efficiently. It is also used as a feature extraction before connecting to an LSTM layer [Wei 18] or applied to LSTM directly to enhance information flow [Wang 16]. There is a more challenging study creating skip connections dynamically [Gui 18].

The LSTM that is different from these LSTMs with the residual connections in that it doesn't have such a deep structure with respect to the time direction. We have incorporated that mechanism described above into the LSTM to reflect the idea of the site manager, not to prevent vanishing gradient. The improved LSTM, we equate it as "Res-LSTM" below, achieve higher accuracy than the LSTM and the ARIMA.

Contact: Takeshi Morinibu, Daikin Industries, Ltd. Technology and Innovation Center, 06-6195-7184, takeshi.morinibu@daikin.co.jp

2. Method

2.1 Dataset and previous results

The data we used is the shipping number of series per month for 2013 to 2018. As the evaluation measures, we use the average value of RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) within the prediction period. RMSE is used to compare the accuracy of the model and MAPE is used to compare series. We have been updating the ARIMA and forecasting the number of series shipments every month since last year and summarized the average accuracy of the major series over the period. From the results: Table 1, it was found that the prediction accuracy of series B was not better than the others in the major series.

Table 1: Accuracy of previous model for major series

Series	A	B	C	D	E
MAPE	11.4	17.4	8.10	13.3	12.1
RMSE	573	239	1038	301	324

According to Figure 1 and Figure 2, the strong seasonality of the year is common to both, but series B has larger variance over its cycle than A. This seems to be one of the factors that lower the prediction accuracy of the ARIMA.

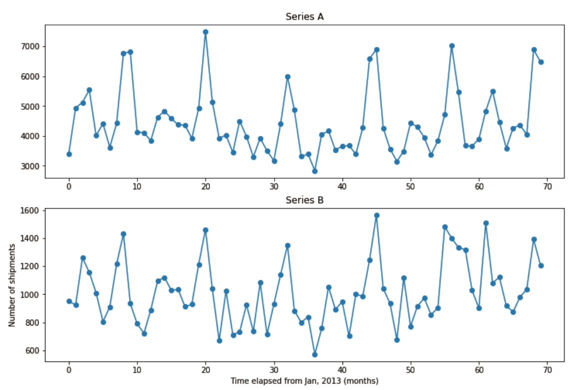

Figure 1: Comparison of observed shipments of series A and B

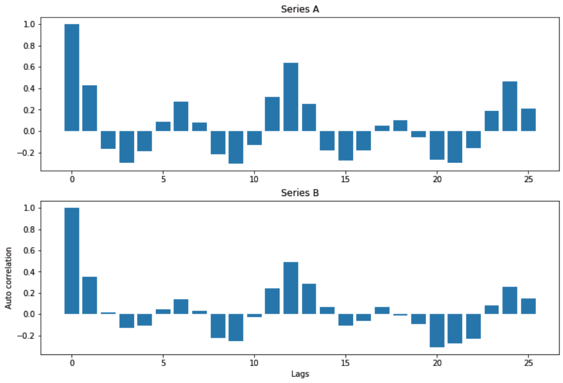

Figure 2: Comparison of autocorrelation coefficient of series A and B

According to the actual prediction result shown in Figure 3, the ARIMA can capture features well for series A but not for series B. Establishing a production plan using this forecast value can cause excessive inventory. To solve this problem, it is necessary to use a model that can deal with more complicated problems than the ARIMA conventionally. In this case, LSTM is appropriate for that model.

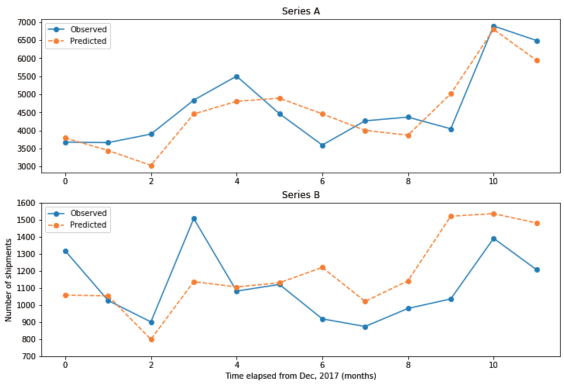

Figure 3: The difference of predicted values by ARIMA between series A and B

2.2 LSTM

LSTM is a type of RNN solves the problem of a vanishing or exploding gradient and makes it possible for network to correctly remember information far back in the sequence [Hochreiter 97]. Due to its characteristics, LSTM has resulted in a wide range of fields such as natural language processing and time series data.

Figure 4 shows the LSTM we implemented in this paper. This model predicts one-month-ahead shipments using series data of length 12 as input. The structure of this network is based on the site experience that the value of the past year is helpful. In spite of our efforts, this model did not give much better accuracy than the ARIMA as described later. Therefore we extended the network structure to make use of the additional on-site knowledge for the LSTM.

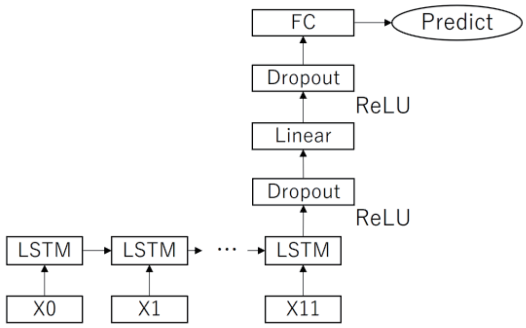

Figure 4: The structure of the plain LSTM

2.3 Res-LSTM

Product managers say that in order to predict one-month-ahead shipments, the monthly data of that just one year ago is particularly beneficial. Depending on the structure of the sequence data, the first output of the LSTM strongly reflects the data of the same month last year, accordingly we sum the first and the last output of the LSTM and connect to the next layer. Figure 5 shows its structure.

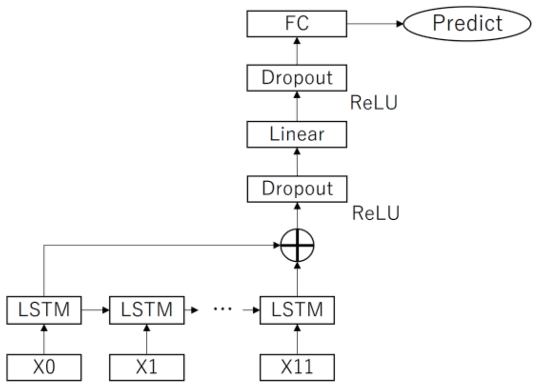

Figure 5: The structure of the Res-LSTM

2.4 Hyperparameters

We adapted Bayesian optimization instead of Grid Search owing to the fairly large space of hyperparameters to search appropriate ones. A variety of Bayesian optimization algorithms have been studied, and [Shahriari 16] introduces its features and the libraries that are implemented in some programing languages. We chose TPE (Tree-structured Parzen Estimator) as an algorithm [Bergstra 11]. It is possible for the TPE to search efficiently and apply it stably to the space containing categorical variables. The search ran 500 times within the range of the table 2.

Table 2: Space of hyperparameters

	range
Hidden1	(5, 20)
Hiddne2	(5, 20)
Optimize	(Adam, RMSprop)
Learning rate	(0.001, 0.1)
Dropout1	(0.1, 0.5)
Dropout2	(0.1, 0.5)
Epoch	(10, 100)

2.5 Experiment

We implemented both models using "PyTorch" and used "Hyperopt" library of Python for searching hyperparameters [Bergstra 15]. Predicted values over the year was made using the data up to 2016/11 for the verification and up to 2017/11 for the test. Figure 6 represents it graphically. Hyperparameters of each LSTM were optimized by Hyperopt with the verification score, and the result of testing with that value is taken as prediction accuracy. Then, we used MSE(Mean Squared Error) as the loss function and batch size as 18.

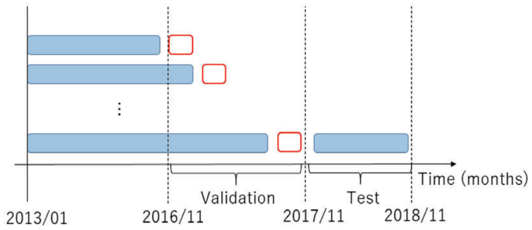

Figure 6: Division method for verification and testing

3. Results

3.1 Optimized hyperparameters

Figure 7 shows the search result of hyperparameters. Compared with the plain LSTM, the Res-LSTM shows higher verification score during training. Table 3 shows what was selected from them. Optimized hyperparameters has differences depending on the structure of the model. We will discuss it later in section 4.

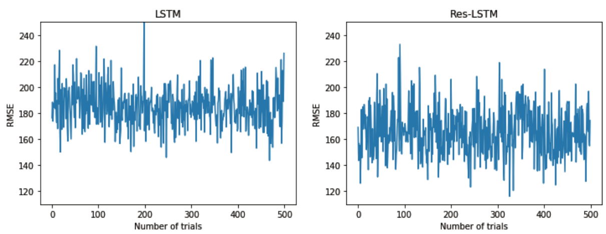

Figure 7: Result of trials while verification. The parameter of the point where the best score is recorded is used.

Table 3: Values for the respective hyperparameters in the LSTM and Res-LSTM

	LSTM	Res-LSTM
Hidden1	7	8
Hiddne2	7	9
Optimize	Adam	RMSprop
Learning rate	0.02	0.04
Dropout1	0.25	0.09
Dropout2	0.36	0.24
Epoch	39	69

3.2 Accuracy

Table 4 describes the test accuracy of each model for series B. The Res-LSTM achieved the best accuracy in terms of both RMSE and MAPE and it was better than other major series. Consequently, we managed to achieve our original objectives.

Table 4: Accuracy of each model for series B

	MAPE	RMSE
ARIMA	17.4	239
LSTM	17.3	176
Res-LSTM	11.1	146

3.3 Prediction

A comparison of predicted values of the LSTM and the Res-LSTM is shown in Figure 8, and a comparison between those of the Res-LSTM and the ARIMA is shown in Figure 9. The plain LSTM did not grasp the trend well and the ARIMA was able to grasp it, but the individual value greatly exceeded. On the other hand, the Res-LSTM catched the trend and is adaptable to each values. The value exceeding 1,500 observed at point 3 is probably Large-volume shipping. Large-volume shipping occurs rarely and it is difficult to predict from data, but in business it is often understood beforehand. Except for that, the model has made a fairly good prediction.

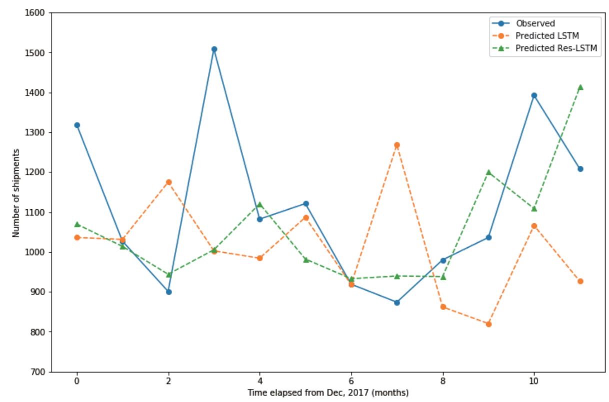

Figure 8: LSTM vs Res-LSTM

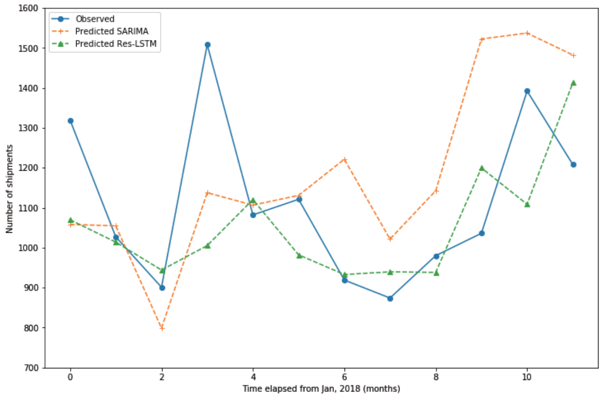

Figure 9: ARIMA vs Res-LSTM

4. Discussion

The Res-LSTM was able to significantly improve the accuracy compared with previous models. We will estimate factors that caused such results. The hyperparameters obtained by the verification: Table 3, shows that the learning rate and the epoch number of the LSTM are lower than those of the Res-LSTM. This is because the LSTM has reached the limit parameter to avoid over-fitting during the search. The over-fitting line of the Res-LSTM was higher than that of the LSTM, so it seems that it was able to acquire generalization performance even if further learning is advanced. In addition, Table 3 shows that the ResLSTM has lower Dropout rate than the LSTM. This is probably because the information that the LSTM handled stochastically was handled by the network structure of the ResLSTM.

5. Conclusion

In this paper, we showed that the Res-LSTM we proposed outperforms prediction accuracy of the plain LSTM and the ARIMA for predicting shipment quantity. Moreover, it seems that the residual mechanism is working to reflect the information of the structure of the series data rather than preventing the gradient disappearance. Although this may not be effective in all cases, it was able to solve a specific problem like this time. In the future, it is necessary to investigate whether there is effective for the others.

In this way, it is sometimes more effective to incorporate knowledge of the domain into the network structure, so we will continue to conduct research on those policies in the future.

Acknowledgment

We would like to thank Yamauchi, S. and Shiimado, T. in the SCM department of our company for providing data and knowledge about it.

References

[Bergstra 11] Bergstra, J., Bardenet,R., Bengio, Y. and Kgl, B.: Algorithms for Hyper-Parameter Optimization. In *NIPS 24*, pp. 2546-2554, (2011).

[Bergstra 15] Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., Cox, D.D.: Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. *Computational Science and Discovery*, 8(1), (2015).

[Gui 18] Gui, T., Zhang, Q., Zhao, L., Lin, Y., Peng, M., Gong, J. and Huang, X..: Long Short-Term Memory with Dynamic Skip Connections. *arXiv*, (2018).

[He 16] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 770-778, (2016).

[Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long short-term memory. *Neural Computation*, 9(8), pp. 1735-1780, (1997).

[Li 18] Li, Y. and Cao, H.: Prediction for Tourism Flow based on LSTM Neural Network. *Procedia Computer Science*, 129, pp. 277-283, (2018).

[Shahriari 16] Shahriari, B., Swersky, K., Wang, Z., Adams, R.P. and De Freitas, N.: Taking the Human Out of the Loop: A Review of Bayesian Optimization. In *Proceedings of the IEEE*, pp. 148-175, (2016).

[Shukla 13] Shukla, M., Jharkharia, S.: Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation. *International Journal of Information Systems and Supply Chain Management*, 6(3), pp. 105-119, (2013).

[Wang 16] Wang, Y. and Tian, F.: Recurrent residual learning for sequence classification. In *Proceedings of the EMNLP*, pp. 938–943, (2016).

[Wei 18] Wei, H., Zhou, H., Sankaranarayanan, J. and Sengupta, S.: Residual Convolutional LSTM for Tweet Count Prediction. In *Proceedings of the The Web Conference*, pp. 1309-1316, (2018).