CAR SALESFORECASTING METHODS FOR ONE MODEL: CURRENTSTATUS AND CHALLENGES

Tomonari Yomono[1] and Takayuki Kataoka[2]

[1,2]Faculty of Engineering, Kindai University, Higashi-Hiroshima 7392116, Japan

Abstract

The traditional car sales forecasting methods have some problems to make production plans without depending on experience or intuition. Almost all previous papers have been focused on forecasting total sales volume for automobile markets or manufacturers. However, micro forecasting methods for one model are required in the actual car manufacturer production plans. In addition, the accuracy and stability of forecasts differ greatly depending on the car model (Sedan, SUV and Wagon, etc.). In this paper, the relationship between automobile sales data and forecasting methods is forecasted and investigated for the sales volume of one type car model. In experimental results, deep learning methods and statistical forecasting methods for sales data of two type car models are considered. As the results, statistical methods can lead high forecasting accuracy in many periods, although it shows a significant decrease in the stability of the forecasting accuracy in some periods. On the other hand, one of deep learning methods, GRU, can lead relatively stable forecasting results for small-scale data.

Keywords: Time Series, Demand Forecast, Deep Learning, Statistical Methods

1. Introduction

As the globalization of the manufacturing industry progresses, the traditional production plans based on experience or intuition are often different from the actual sales results, and improving the accuracy and stability of demand forecasts at foreign country's markets is an urgent issue. In addition, with the development of forecasting technology and computer equipment, an increasing number of companies are using big data to forecast uncertain demand.

Under such a circumstance, almost all previous papers have been documented by analyzing the impact of macroeconomic on the automobile markets and forecasting total automobile sales for markets and manufacturers [1] [2] [3] [4] [5] . However, in the actual production plans of an auto manufacturer, a micro forecasting method for each product is required.

Machine or deep learning methods in demand forecasting can lead the predicted values by the input data (the training data) observed in the past into the model and learning potential features from the input data. Therefore, a large amount of training data is needed to make forecasts with high accuracy. However, in recent years, the product life cycle in the manufacturing industry has been shortened due to diversification of consumer needs and modularization of product. In addition, every few years, the cars are replaced with newer models, making it difficult to collect sufficient training data for a single car model.

The purpose of our study is to build a model that provides the improved forecast accuracy and stability for a single car model and support the accurate production planning designs relatively.

2. Background

The Recurrent Neural Network (RNN) is one of the most prominent prediction models for time series tasks, including Natural Language Processing (NLP). In particular, the Long Short-Term Memory (LSTM) [6] and the Gated Recurrent Unit (GRU) [7], whose unit structure has been expanded, are currently being used in a wide range of fields such as machine translation and speech recognition. Hence, a new forecasting model using GRU might become the bridgehead in this study [8]. At the same time, the weekly data that divides monthly data into five weeks according to a certain rule is conducted in our experiments. The weekly data is designed to supplement a small amount of monthly data. These studies have achieved some high prediction accuracy. However, with the weekly data, it has found that the cost of data preparation is high and it is difficult to collect explanatory variables, thus the next experiments are conducted using monthly data including explanatory variables such as economic indicators and Google Trends. There is still the problem that the number of data is greatly reduced, however, the sample rate of the data has been changed from weekly to monthly, which simplifies the data handling. In previous section, we mentioned that machine learning requires a large amount of data to make highly accurate prediction. However, the deep learning methods including RNN require more data because of feature learning performed. Hence, the lack of data leads a bottleneck. Then, the prediction experiments are conducted by using statistical methods such as ARIMA [9], multiple regression analysis, and Generalized Linear Model (GLM), and machine learning such as Random Forest and Support Vector Regression (SVR). These methods do not require as much data as deep learning. In addition, an ensemble of methods [5] using these multiple prediction methods is also considered.

3. Evaluation Method

In this section, we describe the experimental settings and evaluation methods.

Comparison Methods

The methods of our comparative evaluation are as follows.

- ARIMA stands for Autoregressive Integrated Moving Average, which is one of the most prominent time series models of statistical approaches.
- Holt-Winters is an exponential smoothing method that adds trend and seasonal variation to the variation of a time series. it is also called Triple Exponential Smoothing.
- CNN is a one-dimensional Convolutional Neural Network that performs temporal convolution.
- Univariate-GRU is the RNN model using GRU cell. It is inputted with univariate training data.
- Multivariate-GRU is inputted with multivariate training data into the GRU.

RNN can easily make forecasting using many kinds of data for training (e. g., external factors) data. Several models that include a CNN component have been proposed and they can handle multivariate time series data [10] [11] [12] though they are not included in this paper. In Multivariate-GRU, in addition to the actual data of the forecasting target, other companies' sales results, nonfarm payrolls, exchange rates, and Google Trends, are used. These are determined by previous studies.

3.1 Metrics

We use two conventional evaluation metrics defined as follows:

- Root Relative Squared Error (RRSE):

$$RRSE = \sqrt{\frac{\sum_{t=1}^{n}(A_t - F_t)^2}{\sum_{t=1}^{n}(A_t - \bar{A})^2}} \# \quad (1)$$

- Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \# \quad (2)$$

where A_t, F_t are actual value and forecast value, respectively, and $\bar{A}$ is mean of A_t. RSE and MAPE are used to evaluate forecasting accuracy in regression problems, both of which indicate that lower values are superior.

3.2 Data

We use two different types of car sales data. Table 1 summarizes the data.

Table 1 Data summary

Type of car	Form	Period	Sample rate
Model A	Sedan	2013/1~2019/10	1 month
Model B	SUV	2012/2~2019/8	1 month

In order to examine the existence of long-term and/or short-term repetitive patterns in time series data, we plot the autocorrelation graph in Figure 1. Autocorrelation indicates how much the data correlates with the past history, and degree k order autocorrelation coefficient is defined below.

$$r(k) = \frac{E[(x_t - \bar{x})(x_{t+k} - \bar{x})]}{\sigma^2}$$

where x_t is the time series signals, $\bar{x}$ is mean and σ^2 is variance.

We can see in the graphs (a) and (b) of Figure 1, there are not repetitive patterns with high autocorrelation in either datum. It shows that it is difficult to forecast time series data without a certain pattern.

3.3 Experimental Details

As to the adjustable hyper parameters in each method, the hidden dimension of the Recurrent and Convolutional layer is selected from {50, 100}, and {32, 64}, respectively, and the number of epochs is {50, 100, 150}. Then, we use the loss function is Mean Squared Error (MSE), adjustment of learning rate is Adam [13] algorithm, the activation function is Linear function, and batch size is 1. In the ARIMA, the choice of parameters is determined based on AIC [14]. In the AIC criterion, the model that minimizes the AIC value estimated using the training data is adopted. Also, ARIMA orders p, d, and q (representing the order of the autoregressive, the degree of differencing, and the order of the moving-average, respectively) are chosen from the range p, d and q are 1 to 6, 1 to 3, and 1 to 6 each.

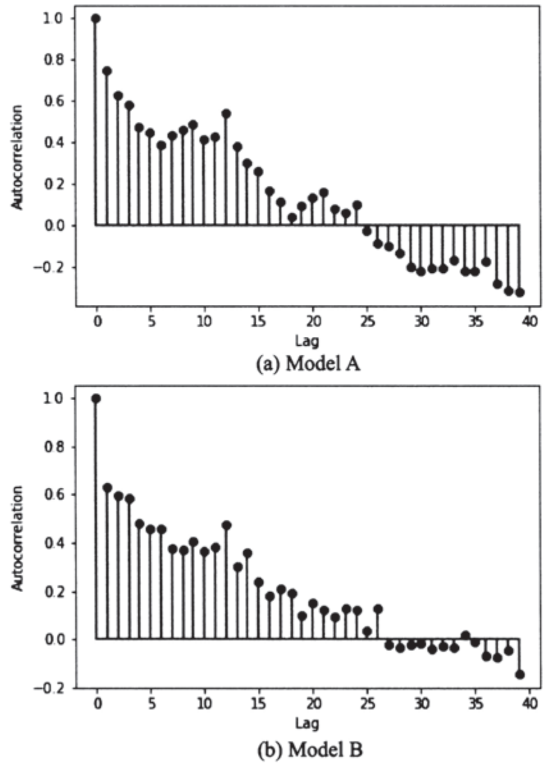

Figure 1: Autocorrelation graph of each data

4. Results

Table 2 summarizes the evaluation results of all the methods on all the test sets in all the metrics. We set all horizon = 6 (months). Also, since the forecasting results differ greatly depending on the forecasting period, we conduct model learning and prediction separately for each of the nine forecast periods. Furthermore, the forecast periods are ordered in time. Period 1 is the oldest period and period 9 is the newest period. Therefore, the number of training data for forecast period 1 is smaller than that for forecast period 9.

Table 2 also shows that good results are obtained with the statistical methods ARIMA and Holt-Winters. In particular, there is a large difference between GRU and the forecasting accuracy in the forecasting periods 1 to 4 of Model A. It is conceivable that the GRU, which requires a large amount of data, is difficult to forecast.

Table 2 Results summary (Bold face indicates the best result of each column in a particular metric)

(a) Model A

Data		Model A								
		Period								
Methods	Metrics	1	2	3	4	5	6	7	8	9
ARIMA	RRSE	1.765	**1.100**	1.517	**1.533**	1.646	1.183	**1.398**	0.953	2.574
	MAPE	18.758	**15.657**	18.061	**14.814**	19.673	25.967	**11.423**	13.262	64.571
Holt-Winters	RRSE	**0.857**	1.361	**0.988**	2.828	1.420	**0.794**	3.089	1.612	2.873
	MAPE	**9.188**	21.755	**11.508**	30.396	22.836	25.778	26.487	18.926	70.958
CNN	RRSE	1.493	2.298	1.541	1.624	1.058	1.343	4.717	**0.885**	1.111
	MAPE	16.803	38.751	21.489	15.477	15.783	24.031	43.363	**11.587**	26.231
Univariate-	RRSE	1.517	2.199	2.387	3.673	1.150	0.980	1.540	1.166	1.431
GRU	MAPE	16.255	36.755	34.196	40.260	19.038	**22.256**	13.466	15.245	35.128
Multivariate-	RRSE	1.609	2.796	1.448	3.600	**0.966**	1.156	1.908	1.759	**0.993**
GRU	MAPE	18.748	46.460	15.920	41.684	**13.916**	33.703	16.391	16.704	**16.684**

(b) Model B

Data		Model B								
		Period								
Methods	Metrics	1	2	3	4	5	6	7	8	9
ARIMA	RRSE	2.708	3.243	**0.936**	3.068	1.489	**1.274**	1.128	**1.816**	1.242
	MAPE	28.741	24.947	**13.274**	25.527	16.586	**11.460**	**8.455**	20.439	20.890
Holt-Winters	RRSE	**1.108**	2.940	1.513	1.426	1.195	1.842	1.871	1.972	1.243
	MAPE	**12.245**	20.199	22.825	**11.212**	15.162	23.257	12.136	**18.919**	20.094
CNN	RRSE	1.978	2.068	1.105	2.175	**1.129**	1.548	2.416	2.882	1.506
	MAPE	18.585	14.300	16.905	17.945	**10.842**	20.689	19.186	32.243	25.047
Univariate-	RRSE	1.875	**1.212**	0.934	**1.417**	1.303	1.886	2.330	1.957	**1.112**
GRU	MAPE	17.972	**8.673**	13.051	12.184	13.706	21.310	19.159	22.710	**19.356**
Multivariate-	RRSE	1.894	2.430	1.435	2.716	1.183	1.393	**1.114**	1.915	1.294
GRU	MAPE	19.163	18.012	16.984	21.160	14.487	14.197	9.109	20.533	20.564

Table 3 Forecasting accuracy of Model A and Model B

Data		Model A			Model B		
Methods	Metrics	Mean	Max	SD	Mean	Max	SD
ARIMA	RRSE	**1.519**	**2.574**	**0.449**	1.878	3.243	0.839
	MAPE	**22.465**	64.571	15.412	18.924	28.741	6.541
Holt-Winters	RRSE	1.758	3.089	0.868	1.679	2.940	0.535
	MAPE	26.426	70.958	17.029	17.339	23.257	4.456
CNN	RRSE	1.786	4.717	1.107	1.867	2.882	0.561
	MAPE	23.724	43.363	10.270	19.527	32.243	5.837
Univariate-GRU	RRSE	1.783	3.673	0.802	**1.559**	**2.330**	**0.442**
	MAPE	25.844	**40.260**	**9.997**	**16.458**	22.710	4.456
Multivariate-GRU	RRSE	1.804	3.600	0.828	1.708	2.716	0.536
	MAPE	24.468	46.460	11.872	17.134	**21.160**	**3.719**

for those periods. Similarly, the statistical method is almost dominant in Model B.

Table 3 summarizes the mean, maximum, and standard deviations (SD) of the metrics for all forecasting periods for each method.

In Model A, the maximal values of the ARIMA and Holt-Winters MAPE metric are 64.6% and 71.0%, respectively, showing very bad results. On the other hand, the standard deviation of MAPE metric for Univariate-GRU is the lowest, indicating that the forecast is relatively stable. Model B also shows the best stability with GRU showing the minimum value in all points of mean, max value and standard deviation.

In this study, there is no significant difference between Univariate-GRU and Multivariate-GRU. The reasons are considered as follows:

- The explanatory variables do not positively influence the forecasting.
- There is not enough data to discover the relationship between variables.

Thus, our future work will include collecting large datasets for more accurate forecasting models.

5. Conclusions

In this paper, the sales volume of one type car model is forecasted and the relationship between automobile sales data and forecasting methods is investigated. As the results, the statistical methods can lead high forecasting accuracy in many forecasting periods. However, it also shows a significant decrease in the stability of the forecasting accuracy in some periods. On the other hand, it is showed that GRU can relatively make the stable forecasting for small-scale data.

In feature, we would like to conduct an experiment using the daily data and the adding meteorological data as the explanatory variable. By changing from monthly to daily data, weather data also becomes easier to handle. Used in the field of demand forecasting, it may have a positive impact on this study.

6. Acknowledgements

We gratefully acknowledge the work of past members of our laboratory. This work was supported by JSPS KAKENHI Grant Number 20H02390 and 20K04993.

References

[1] Barber, B. M., et al., (1999). The impact of shocks to exchange rates and oil prices on U. S. sales of American and Japanese automakers.

Japan and the World Economy, 11(1), 57-93.

[2] Hülsmann, M., et al., (2012). General Sales Forecast Models for Automobile Markets and their Analysis. Transactions on Machine Learning and Data Mining, 5(2), 65-86.

[3] Fantazzini, D. and Toktamysova, Z., (2015). Forecasting German car sales using Google data and multivariate models. International Journal of Production Economics, Vol. 170, PartA, 97-135.

[4] Yang, L. and Li, B., (2016). The Combination Forecasting Model of Auto Sales Based on Seasonal Index and RBF Neural Network. International Journal of Database Theory and Application, 9(1), 67-76.

[5] Fleurke, S., (2017). Forecasting Automobile Sales using an Ensemble of Methods. WSEAS Transactions on Systems, Vol. 16, 337-345.

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

[7] Cho, K., et al., (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statical Machine Translation. Proceedings of the 2014 Conference on Empirical Methods in Neural Language Processing (EMNLP), 1724-1734.

[8] Turuoka, S. and Kataoka, T., (2019). Proposal of A Car Sales Forecast Model using Recurrent Neural Network and A Case Study. Proc. of the 63rd Domestic Conference of Japan Association of Management Systems, 104-105.

[9] Box, G. E. P. and Pierce, D. A., (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. Journal of the American Statistical Association, 65(332), 1509-1526.

[10] Börjesson, L. and Singull, M., (2020). Forecasting Financial Time Series through Causal and Dilated Convolutional Neural Networks. Entropy 2020, 22(10), 1094.

[11] Lai, G., et al., (2018). Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, 95-104.

[12] Borovykh, A., et al., (2018). Dilated Convolutional Neural Networks for Time Series Forecasting. Journal of Computational Finance. 22(4), 73-101.

[13] Kingma, D. P., and Ba, J. L., (2014). Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR).

[14] Akaike, H., (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Caski, F. (eds.), 2nd International Symposium on Information Theory, 267-281.