creative commons

COMMONS DEED

저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

[UCI]I804:11048-000000167918

이화여자대학교 대학원
2019학년도
석사학위 청구논문

A Hybrid Model Based on Fourier-ARIMA and LSTM for Seasonal Time Series Forecasting

통 계 학 과
Namkyeong Lee
2020

dCollection @ ewha

A Hybrid Model Based on Fourier−ARIMA and LSTM for Seasonal Time Series Forecasting

이 논문을 석사학위 논문으로 제출함

2020 년 7 월

이화여자대학교 대학원

통 계 학 과 Namkyeong Lee

dCollection @ ewha

이남경의 석사학위 논문을 인준함

지도교수 _______신 동 완________

심사위원 _______신 동 완________

_______송 종 우_______

_______안 재 윤_______

이화여자대학교 대학원

dCollection @ ewha

Table of Contents

dCollection @ ewha

List of Tables

dCollection @ ewha

List of Figures

dCollection @ ewha

Summary

This study proposes a hybrid model combining Fourier autoregressive integrated moving average (Fourier-ARIMA) and long short term memory (LSTM) neural network for forecasting seasonal time series. ARIMA is a linear forecasting model based on past observations and random shocks. The prediction accuracy of ARIMA model for seasonal data can be enhanced by adding external regressors in the form of Fourier terms. Meanwhile, neural networks such as LSTM are powerful tools to capture the nonlinear dependencies in time series. Because real world data is generally mixed with linear and nonlinear patterns, forecasting real world data using each model is a particularly difficult task. In this thesis, a hybrid methodology combining the forecasts from Fourier-ARIMA and LSTM by weight is proposed. The experimental results show that the hybrid model provides a promising alternative to separate Fourier-ARIMA model or LSTM model.

dCollection @ ewha

I. Introduction

Forecasting time series finds a lot of applications in many fields. With the advent of the big data era, a massive amount of data is stored and processed in the server. We can use this data to estimate future time points. There are many techniques to model a time series. The suitable model can help us make effective decisions, which in return can reduce risk and increase return.

The autoregressive integrated moving average (ARIMA) model is a statistical method of time series forecasting. The model explains a given time series based on historical observations and the lagged errors. The linear combination of its past values can be used to forecast future values. A linear relationship is assumed among the time series values, and therefore no nonlinear patterns can be captured by the ARIMA model. That traditional model firstly proposed by Box and Jenkins (1976) has been widely used due to its simplicity and its ability to generalize for non-stationary series. However, the approximation of linear models does not necessarily reflect complex real world data and lacks predictive efficiency.

Thus, a wide range of models that suggest different mathematical representations of the non-linearity has been studied. Recently, long short term memory (LSTM) model has become an important method for time series forecasting due to their flexible nonlinear modeling capability. LSTM models use the history of a sequence of data and extract features, so that correctly predict what the future elements of the sequence are going to be. Even though it was shown that an LSTM approach can model complex nonlinear feature interactions, the LSTM may not be sufficient for modeling many empirical datasets that has both linear and nonlinear patterns equally well.

This paper proposes a hybrid method combining the forecasts from linear and nonlinear model. The method employs Fourier-ARIMA and LSTM. Fourier terms are the external regressors added to the ARIMA model in order to deal with multiple seasonality.

The remaining of the thesis is organized as follows. Section 2 outlines the related works

dCollection @ ewha

about hybrid systems. In Section 3, the proposed hybrid model that combines the Fourier-ARIMA and the LSTM models is described. Section 4 reports evaluation results of the proposed model. Section 5 contains concluding remarks.

dCollection @ ewha

II. Related Works

To predict the outcome of an event to be observed at a certain point in the future, we may use several predictive models or expert views. At this point, each predictive model or expert has a different theoretical background and assumption. Thus, the values predicted by more than one model or expert will reflect different information, and the combination prediction that combines them appropriately may be better than individual predictions.

For a time series y_t, Zhang (2003) introduced the linear combination of the linear (L_t) and nonlinear (N_t) patterns as follows

$$y_t = L_t + N_t \,.$$

ARIMA produces linear forecasts ($\hat{L}_t$), and then the error series (E_t) is obtained through the difference between the original series (y_t) and the linear forecasts ($\hat{L}_t$) as

$$e_t = y_t - \hat{L}_t \,.$$

The residual series should not be correlated according to Box & Jenkins methodology of identifying ARIMA models. If the linear model is well specified, there is no linear pattern in the error series. Thus, the error series (e_t) can be used to predict nonlinear component (N_t)

$$\widehat{N}_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \epsilon_t \,.$$

An Artificial Neural Network (ANN) performs *f*(.) as a nonlinear model and the forecasts ($\widehat{N}_t$) are made without predicting the random error (ϵ_t).

In this way, the final forecast ($\hat{y}_t$) is obtained by the sum of $\hat{L}_t$ and $\widehat{N}_t$ as follows

$$\hat{y}_t = \hat{L}_t + \widehat{N}_t \,.$$

Meanwhile, Yuwei Chen and Kaizhi Wang (2019) proposed linear combination system to forecast satellite time series, but their method is different from the work of Zhang (2003) in the way the components are combined. In the first step of this approach, L_t and N_t are predicted by ARIMA and LSTM model from observations, respectively. Then, L_t and N_t are

dCollection @ ewha

combined by a certain weight balance to forecast final value y_t,

$$\hat{y}_t = w_1\hat{L}_t + w_2\hat{N}_t \text{ ,}$$

where w_1 and w_2 are the weights of each model.

In addition to linear combining method, there are also the nonlinear combination systems. Khashei and Bijari (2011) proposed a nonlinear methodology. There is nonlinear relationship *f*(.) between linear (L_t) and nonlinear patterns (N_t) as

$$y_t = f(L_t,\ N_t)\ .$$

First, ARIMA model predicts L_t as in the work of Zhang (2003). Then, the residuals ($e_{t-1}, e_{t-2}, \dots , e_{t-q}$), the linear foreacsts ($\hat{L}_t$) and the past observations ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) are used as input for the ANN to forecast final value y_t as follows.

$$\hat{y}_t = f\left(e_{t-1}, e_{t-2}, \dots , e_{t-q}, \hat{L}_t, y_{t-1}, y_{t-2}, \dots, y_{t-p}\right) \text{ ,}$$

where p and q are the size of the temporal window of observation y_t and the residual, respectively.

dCollection @ ewha

III. Model Description

In this section, the structures of the ARIMA with Fourier terms and the LSTM model are reviewed. Then, we present details about the hybrid model obtained by combining these ARIMA with Fourier terms and LSTM models in a rational way.

A. Fourier-ARIMA

1) ARIMA model

The Autoregressive Integrated Moving Average (ARIMA) is a traditional forecasting model used for linear data. This is a mathematical model designed to forecast data based on past observations and random shocks. It is expressed generally as ARIMA(p, d, q) where the p, q determines the definite orders of AR and MA parts, whereas d specifies the number of iterations for differencing required to make the input data y_t stationary. The ARIMA is formulated as follows

$$z_t = c + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

where $z_t = \Delta y_t$, $\Delta y_t = y_t - y_{t-1}$ and ε_t are the observation value and random error at time t, respectively, and φ_i's and θ_i's are the coefficients of ARIMA model.

2) Harmonic regression

If the time series is relatively short and there is only one seasonal feature, the seasonality can be dealt with error-trend-seasonal (ETS) model or seasonal ARIMA. However, high-frequency time series often shows more complex seasonal patterns. In this case, harmonic regression is effective. A regression model that contains Fourier terms is referred to as harmonic regression and expressed as

dCollection @ ewha

$$y_t = \alpha + \beta t + \sum_{j=1}^{K} a_j \cos(j\omega t) + \sum_{j=1}^{K} b_j \sin(j\omega t) + \varepsilon_t \,,$$

where y_t is the observation series, K is the order of the Fourier polynomial, β is the value that represents the trend of data, and a_j and b_j are the adjustable Fourier coefficients, α is an intercept, j is the order of harmonic component and ω is the frequency of the signal.

3) Fourier-ARIMA model

The residual series of ARIMA often becomes smaller when Fourier terms were incorporated into ARIMA model. Forecasting time series has been proved to be significantly improved by adding Fourier terms as considered by C.M. Eze et al. (2020). The Fourier-ARIMA model can be written in the form

$$z_t = c + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$
$$+ \sum_{j=1}^{K} a_j \cos(j\omega t) + \sum_{j=1}^{K} b_j \sin(j\omega t) + \varepsilon_t \,.$$

C.M. Eze et al (2020) demonstrated that the predictability of ARIMA model can be significantly improved by adding the Fourier terms.

B. LSTM

Long short term memory (LSTM) model is a special type of recurrent neural network (RNN). It has the ability to perform learning that requires a long period of dependence. LSTM was introduced by Hochreiter & Schmidhuber (1997), after which it continued to develop and became widely used through several further studies.

Cell state is a key part of LSTM. The cell state passes through the entire chain like a conveyor belt. This structure allows information to continue to be delivered to the next stage without major changes. LSTM utilizes elements called gates with a carefully refined structure to add or

dCollection @ ewha

remove information. Gates are devices that selectively allow information to flow in. The gates consist of a sigmoid neural net layer and a point-by-point operation as shown in Figure 1.

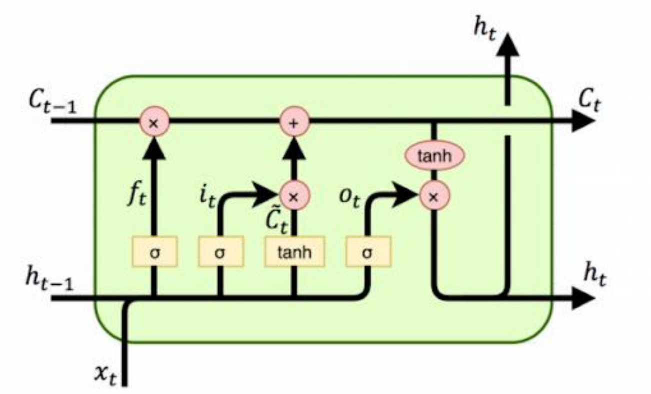


Figure 1. The structure of LSTM

The first step of LSTM is the forget gate (f_t), which determines how much information it will forget about the past. The value of the past hidden layer (h_{t-1}) and the current information (x_t) is accepted as input value and output value between 0 and 1 is obtained through the sigmoid activation function. The next step is to determine whether new information will be stored in a cell state. The candidate ($\tilde{C}_t$) is recorded in the cell considering the input gate (i_t). The third step is to change the old information (C_t) to new candidate ($\tilde{C}_t$). Finally, output gate (o_t) determines the hidden layer value at this point (h_t). Mathematical expressions for this structure are:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right),$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i),$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C),$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t,$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o).$$

$$h_t = o_t * \tanh(C_t)$$

dCollection @ ewha

In this way, LSTM solves the problem of long distance dependence and delivers the information to the next point.

C. Proposed hybrid method

This thesis proposes a hybrid method based on previous studies. Yuwei Chen and Kaizhi Wang (2019) suggested a forecasting model combining linear and nonlinear by weight. This thesis also uses the weighted linear combination in this way. The linear component is modelled using the Fourier-ARIMA, in which the model coefficients are estimated using the Box & Jenkins method and harmonic regression. The best-suited Fourier-ARIMA order p, q, K are chosen by the BIC. Also, the LSTM model performs a nonlinear fitting to the data. From these two models, one-step forecasts, say Fourier-ARIMA forecast $\hat{y}_t^L$ and LSTM forecast $\hat{y}_t^N$, can be obtained using the available past data values and the model coefficients. The predictions obtained from both the Fourier-ARIMA model and the LSTM model are combined by weight as $\hat{y}_t = w\hat{y}_t^L + (1-\mathrm{w})\hat{y}_t^N$. The best suited model is achieved by changing the weight w until the result shows best accuracy.

dCollection @ ewha

IV. Experimental Evaluation

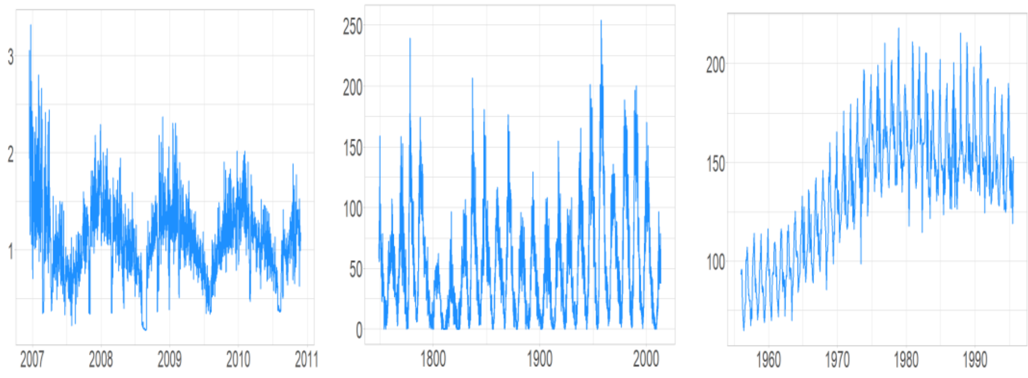
A. Data

The hybrid methodology in Section 3-C is applied to all the 3 datasets in Table 1. The electric power consumption data consists of the daily averaged electric power consumption in one household over a period of almost 4 years. The sunspot series is the monthly records of incidence of spots on the sun surface between January 1749 and September 2013. The beer production time series is composed of the monthly beer production in Australia from 1956 to 1995. The electric power consumption and sunspot data sets have been split into training set (the first 70%) and test set (the remaining 30%), while the last twelve observations of the beer production have been used as the test data of beer production series. Time series plots of the three datasets are displayed in Figure 2.

Data	Period	Unit	Size		
			n	Training set (M)	Test set (N)
Electric Power Consumption	2006/12/16-2010/11/26	Daily	1442	1009(70%)	433(30%)
Sunspot	1749/01-2013/09	Monthly	3177	2224(70%)	953(30%)
Beer Production	1956/01-1995/08	Monthly	476	464(97.5%)	12(2.5%)

Table 1. Basic features of data sets used in the experiments.

Collection @ ewha



Electric Power Consumption Sunspot Beer Production

Figure 2. Time series plots of the datasets.

B. Experimental Details

The settings for the various parameters used in the proposed hybrid system are listed in Table 2. An automatic stepwise approach with BIC is employed for the Fourier-ARIMA model.

		Electric Power Consumption	Sunspot	Beer Production
Fourier-ARIMA	(p,d,q)	(2,1,1)	(1,0,2)	(2,1,3)
	K, ω	1, 365	1, 132	5, 12
LSTM	Number of layers	3	3	4
	Epochs	20	10	20
	Batch size	1	1	1
	Loss function	MSE	MSE	MSE
	Optimizer	Adam	Adam	Adam

Table 2. Parameter settings for the models

dCollection @ ewha

C. Metrics

The result analysis is performed using three conventional evaluation metrics defined as

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_{M+t} - \hat{y}_{M+t})^2}$$

$$\text{Mean Absolute Error (MAE)} = \frac{1}{N}\sum_{t=1}^{N}|y_{M+t} - \hat{y}_{M+t}|$$

$$\text{Mean Absolute Percentage Error (MAPE)} = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{y_{M+t} - \hat{y}_{M+t}}{y_{M+t}}\right|$$

where M is the series length of modeling data, N is the series length of test data (see Table 1) and $\hat{y}_{M+t}$ is the one–step ahead forecast obtained from an estimated model constructed from observations up to time $(M + t - 1)$. The prediction technique that minimizes MAE produces a median of the predicted values, while the prediction technique that minimizes RMSE averages the forecasts. RMSE is widely used although more difficult to interpret than MAE. MAPE is a useful method to compare predictive performance because it is a unit-free measure. For these metrics, smaller values represent better accuracy.

D. Results

Table 3 summarizes the evaluation results on all the test sets. The electric power consumption series has a minimum value at different weights for each metric. The power consumption series has the smallest RMSE at the Fourier-ARIMA weight 0.8. On the other hand, the MAPE value is the smallest at the weights of (0.9, 0.1) among the MAPE values of the electric power consumption series. There is little difference between the weights of (0.8, 0.2) and (0.9, 0.1) in the MAE values. For the sunspot time series, both RMSE and MAE have the minimum value at the Fourier-ARIMA weight 0.8. Sunspot series has values of MAPE that are close to infinity, and therefore they are not listed in the table. In the beer production series,

dCollection @ ewha

the metrics of MAE and MAPE have the minimum value at the weights of (0.6, 0.4), while RMSE has the smallest value at the weights of (0.7, 0.3). It can be seen from the table that hybrid model performs better than the single models on all the datasets given that the weights of the Fourier-ARIMA and the LSTM are neither (0,1) nor (1,0).

Weight (Fourier-ARIMA)	Weight (LSTM)	Electric Power Consumption			Sunspot		Beer Production		
		RMSE	MAE	MAPE	RMSE	MAE	RMSE	MAE	MAPE
0	1	0.2700	0.2053	20.460	18.948	14.001	12.202	9.638	6.208
0.1	0.9	0.2650	0.2012	20.021	18.701	13.787	11.725	9.384	6.095
0.2	0.8	0.2607	0.1973	19.596	18.482	13.593	11.308	9.140	5.990
0.3	0.7	0.2569	0.1941	19.252	18.294	13.426	10.959	8.896	5.885
0.4	0.6	0.2538	0.1914	18.953	18.137	13.291	10.684	8.690	5.804
0.5	0.5	0.2514	0.1894	18.717	18.011	13.187	10.489	8.568	5.779
0.6	0.4	0.2497	0.1878	18.519	17.917	13.110	10.378	**8.446**	**5.754**
0.7	0.3	0.2486	0.1866	18.352	17.857	13.064	**10.354**	8.506	5.828
0.8	0.2	**0.2482**	**0.1862**	18.257	**17.829**	**13.054**	10.417	8.691	5.972
0.9	0.1	0.2486	**0.1862**	**18.201**	17.835	13.073	10.567	8.959	6.168
1	0	0.2497	0.1867	18.230	17.874	13.113	10.798	9.343	6.435

Table 3. Results summary of hybrid method on three datasets (RMSE, MAE and MAPE of the test set)

dCollection @ ewha

V. Conclusion

This paper presents a hybrid system which combines the linear forecasts and the nonlinear forecasts. By analyzing the advantages and disadvantages of these two algorithm models, the two models are combined according to the weight. An experimental evaluation using three evaluation metrics is performed with the electric power consumption, sunspot and beer production time series. The high precision and applicability of the hybrid model in time series data prediction is proved by comparing the experimental results of Fourier-ARIMA model and LSTM model. The experimental results show that this model has higher accuracy with the stability of the ARIMA model and the flexibility of the LSTM model. In the future work, the main goal is to further optimize the combination of the two models at non-monotonous interval. In addition, other nonlinear models that can be used for the nonlinear components should be investigated.

dCollection @ ewha

Bibliography

C.M. Eze, O.C. Asogwa, C.U. Onwuamaeze, N.M. Eze, C.I. Okonkwo (2020). On the Fourier Residual Modification of Arima Models in Modeling Malaria Incidence Rates among Pregnant Women. *American Journal of Theoretical and Applied Statistics*, 9(1), 1-7

C. Narendra Babu, B. Eswara Reddy (2014). A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data, *Applied Soft Computing*, 23, 27-38

Domingos S., Joao F., Paulo S. (2019). An intelligent hybridization of ARIMA with machine learning models for time series forecasting. *Knowledge-Based Systems*, 175, 72-86

G.P. Zhang (2003). Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 50, 159-175

Li Wang, Haofel Zou, Jia Su, Ling Li, Sohail Chaudhry (2013). An ARIMA-ANN Hybrid Model for Time Series Forecasting. *Special Issue: Systems Science in Industrial Sectors*, 30, 211-411

M. Khashei, M. Bijari (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Applied Soft Computing*, 11, 2664-2675

Yuwei Chen, Kaizhi Wang (2019). Prediction of Satellite Time Series Data Based on Long Short Term Memory-Autoregressive Integrated Moving Average Model (LSTM-ARIMA), *IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, 308-312

dCollection @ ewha

국문초록

이남경

통계학과

이화여자대학교 대학원

본 연구는 시계열 예측을 위해 Fourier-ARIMA 모델과 LSTM 신경망을 결합한 하이브리드 모델을 제안한다. ARIMA는 과거 관측치와 예측 오차에 기초한 선형 예측 모델이다. ARIMA 모델의 예측 정확도는 푸리에 항을 이에 추가함으로써 개선될 수 있다. 한편 LSTM과 같은 신경망은 시계열의 비선형 의존성을 포착하는 강력한 도구다. 일반적으로 실제 데이터는 선형 및 비선형 패턴이 결합되어 나타나기 때문에, 각 모델을 사용하여 실제 데이터를 예측하는 것은 특히 어려운 작업이다. 본 논문에서는 Fourier-ARIMA와 LSTM 모델을 가중치를 두어 결합하는 하이브리드 방법론을 제안한다. 실험 결과는 하이브리드 모델이 시계열 예측에 대한 유망한 대안을 제공한다는 것을 보여준다.

dCollection @ ewha