School of Informatics

Informatics Research Review
Time-Series Forecasting in Hybrid Methods

s2015708
January 2021

Abstract

Time-series forecasting as an area of machine learning, aims at predicting future information based on existing observation. Forecasting of time series task plays an important role in many fields such as weather prediction, stock prices forecasting and supply chain management. Given the increasing availability of data sources stored as time series, the research of time-series forecasting dominated by using deep learning has become a trend. Based on deep learning(DL) approach, a novel idea of combining DL with traditional statistic algorithm was proposed in 2003. In this literature review, we will provide an overall picture about hybrid model in time series forecasting task and we mostly focus on the problem we met step by step and how these challenges can be solved by more advanced algorithm and thought as well as current and future state of hybrid method.

Date: Thursday 21$^{\text{st}}$ January, 2021

Supervisor: Mohammad Tahaei

Contents

1 Introduction

With the development of big data, a large amount of data has been accumulated in many fields, among which time series data (in order of time stamp) is an important part. Since Yule[1] published the study of periodicities of sunspot numbers in disturbed series in 1927, the Use these time series data to predict the future state of has a broad application scenarios, for example, it can be applied to predict stock prices to help investors make wiser decisions; health systems with time series forecasting can analyze various indicators of a person's health to predict potential diseases; it can also be used in decision-making for company by forecasting the number of sales.[2] The effective implementation of these scenarios claim for higher requirement of time-series forecasting in the aspect of accuracy. So in this review, we will explore one state-of-the-art method of time-series prediction by reviewing relevant literature step by step.

The first focus of this review is Auto-Regressive Integrated Moving Average(ARIMA). ARIMA, as a representative of traditional statistical algorithms, has been widely applied in industry over the past decades, which is originated from the combination of autoregressive models (AR) and he moving average models (MA)[3].Thanks to the famous Box–Jenkins methodology[4] and its statistic properties, the process of ARIMA model building can be implemented easily.However,the increase of data availability and the update of state-of-the-art machine learning algorithm,ARIMA can not longer offer the best predictions due to its limitation on modelling nonlinear relationships between variables and its hypothesis[5] on the statistic distribution of data. With deep learning(DL) showing extraordinary talent in various implementation areas, increasing numbers of scholars have explored the feasibility of deep learning for timing series forecasting and numerous studies have been published on DL models with relatively better performances than classical time series forecasting techniques[6].

The second focus of this review is deep learning method called Long Short-Term Memory(LSTM). LSTM as a special case of Recurrent Neural Network (RNN) method[5], has been a great success in solving sequential forecasting problem in recent one decade year. LSTM consisting of a special structure called "gate" which is a unique technique for regulating long-term memory, can adopt the nonlinear relationship of inputs and learn long term memory,compared to traditional algorithm ARIMA which only can learn the linear relationship of data[7]. In this review, we focus not only on its particular structure but also on its superiority in deep learning properties.

In addition, many variants based on LSTM have also been proposed by scholars due to its remarkable performance on prediction task. For example,Y Zhang et al[8] proposed highway LSTM in 2015 to solve the vanishing and exploding gradient problem caused by multi-layer LSTM networks, Z Huang et al.[9] structured bidirectional LSTM network evolved from the bidirectional RNN. However,we will not focus on these variants because the study in another direction of LSTM is noteworthy and its performance is extraordinary and non-negligible in recent year, the hybrid methodology which will be the third focus of our review.

Hybrid model is not a new concept, since Zhang[10] first proposed a combined ARIMA and artifical neural hybrid model in 2003, hybrid model has shown its state-of-the-art performance on forecasting accuracy compared to single method. Therefore, we will review the universal data problem discussed by Zhang[10](2003) and his idea of combination process.Furthermore, one state-of-the-art hybrid model called ARIMA-LSTM hybrid model which is a combination of both ARIMA and LSTM methods was which will be discussed as our third focus by reviewing the effect of hybrid method on sequential information and comparing with traditional method. Moreover, hybrid model is a wide direction for forecasting task but not the only one, so at the end of this section, we will overview some mainstreams of time-series forecasting and make a evaluation of the future position of hybrid model.

The development of time-series forecasting models are complex and diverse.This review can not cover all state-of-the-art direction of time series forecasting,although some method such as traditional machine learning algorithm(support vector machine) and other deep learning algorithm(convolutional neural network and transformer) can also carry on good performance on industry and research.In this literature review,we will only focus on one advanced direction:the hybrid model and we try to provide readers with an overall logical picture of hybrid direction by analyzing standard structure as well as comparing with single method.

The remaining of this review is arranged as follows.Section 2 will start by subsection 2.1 which contains brief background about the basis features of time-series data and mathematics.Subsection 2.2 reviews traditional statistic algorithm by evaluate Auto-Regressive Integrated Moving Average(ARIMA) and its limitations.Then,In section 2.3,we will focus on Long Short-Term Memory(LSTM) method and its superiority compared to statistic algorithm.In section 2.4,we will review the basic idea of hybrid mothod and introduce a state-of-the-art hybrid model based on the combination ARIMA and LSTM as well as analyze its possible future directions.Finally,in section 3, we will give a summary and a conclusion about the literature review.

2 Literature Review

2.1 Background of time series data

Time-series forecasting is one kind of prediction task based on the observed time sequential information. Before evaluating different time series forecasting models, some features of time series information should be introduced and reviewed, which is very helpful for readers who do not have relevant background to understand ARIMA and deep learning models for time series data.

2.1.1 Components

Kirchgässner and Wolters[11]has defined the time series that a set of quantitative observations arranged in chronological orders,which means it can be mathematically considered as a set of vectors $v(t)$,where t can be continuous or discrete.Moreover,the continuous and discrete time series data can be easily transformed into each other by merging and spliting time interval[12] and this can be considered as an unique feature of time series data compared to other types of data.

In 1919,Persons [13][14] have listed four main components of time series information:trend, cyclical, seasonal and irregular components, which can be considered as the decomposed fluctuations of observed data. And these four components lay a theoretical foundation for the study of time series in the future.

- **trend variations** - the general or long-term tendency(increasing,decreasing or stagnation) of information over a long period of time, can also be termed as Secular Trend or simply. Trend[12]
- **cyclical variations** - Persons(1919)[13][14] pointed out that the cyclical variations of time series was cyclical rise and fall movement upon general and long-term trend, which means we can find the medium-term changes in time-series in cyclical variations graph[12]
- **seasonal variations** - the movements or fluctuations within the shorter period of time than the cyclical variations which can be affected by the nature of series[13][14]. Adhikari and Agrawal[12] provided examples that when we survey data variations over hundreds of years, the sales of ice-cream increase in summer, sales of woolen cloths increase in winter, and this variations affected by weather can be considered as seasonal variations of ice-cream and woolen cloths sales series.
- **irregular variations** - the time series can be controlled by some uncontrollable and unforeseen forces, such as war, earthquake.[13][14]. Irregular variations are Random and unmeasurable[12].

2.1.2 Stationarity

Time series data can be divided into two groups, stationary and non-stationary[15]. Stationarity is one of the most important features of time series data which is helpful for future forecasting especially for statistic approach, because it is an necessary condition for most statistic forecasting model[12][16].As Adhikari and Agrawal mentioned[12] in 2013 data can be considered stationary when some statistical properties such as mean and variance do not change upon time, on the contrary, the statistical properties of non-stationary time series data keep changing upon time. Furthermore, Hipel and McLeod[15] in 1994 introduced two types of Stationarity:strongly stationary and weekly stationary which two hypotheses about independence.

- **strongly stationary** - defined time series$x(t), t = 0, 1, 2, 3...$,if the joint probability distribution of $x_{t-s}, x_{t-s+1}, ...x_t, ...x_{t+s-1}, x_{t+s}$ is independent of t for all s [12][15][16][17],then the data can be considered as strongly stationary.
- **Weekly stationary** - defined time series$x(t), t = 0, 1, 2, 3...$, if the mean of time series data is constant while the covariance at any two time points only depend on time difference, then it can be considered as weekly stationary.

As Adhikari and Agrawal[12] said, some statistic time series forecasting method can only predict based on the stationary information. However,in nature, "usually time series, showing trend or seasonal patterns are non-stationary[12]"(Adhikari and Argawal,2013), the raw time series information tends to be stochastic and nonlinear. Therefore, some mathematical techniques of these traditional statistic forecasting algorithm are used to remove the trend and seasonal patterns to make the series stationary.

So far, we have learned some background about time series data, including four components (trend,cyclical,seasonal and irregular components) which can be obtained by decomposition to time series and the stationary feature of series in time sequence. In addition, according to Adhikari and Agrawal[12] statements, we have learned that trend and seasonal components are the main cause of data non-stationarity.Therefore, in the next section, we will focus on linear traditional time series algorithm called Auto-Regressive Integrated Moving Average to figure out how does ARIMA works on predicting the future information and its limitations.

2.2 Auto-Regressive Integrated Moving Average(ARIMA)

Auto-Regressive Integrated Moving Average(ARIMA) as state-of-the-art traditional statistic series forecasting model is widely used in industry, as Sima and Akbar in 2018[5] have highlighted that ARIMA model has outperformed in precision and accuracy of predicting the next lags of time series at economic and financial fields for decade years.In addition, Adhikari and Agrawal[12] demonstrated that, compared to other traditional-based algorithm, the popularity of ARIMA is mainly due to its flexibility to describe different variants series information with simplicity as well as the associated Box-Jenkins methodology[4] which is a simple model-identification method by iteratively estimating the parameters until the best predicted results.

The concept of ARIMA is quite straightforward. As Jenkins Box(1970) stated that the Auto-Regressive Integrated Moving Average(ARIMA) model is established in the process of transforming non-stationary time series into stationary time series by regressing either the dependent variable and its lag value or the present value and lag value of the random error term[4]. More specifically, Adhikari and Agrawal[12](2013) illustrated that ARIMA model can be built by three main processes: auto-regression(AR), integration and moving average(MA).Therefore, before we evaluate advantages and limitations of ARIMA model, the concepts of these three steps should be focused on.

- **Integrated(I)** - "Integrated" step can be visualized as initialization, aiming at reducing irregular fluctuations between data and transforming non-stationary information into stationary information by differentiating the observed series.

- **auto-regression(AR)** -As Sima et al.[5] illustrated in 2018 that AR is a regression process based on the dependencies between future and a number of lagged observations (p) where p is the dependency length(auto-regression order). More specifically, the expected value of target in future is equal to a linear combination of one or more lagging observations, plus a constant term, plus a random error.

- **Moving Average(MA)** -Sima et al.[5](2018) illustrated that compared to AR process, MA process focuses on the dependencies between target in future and the random error terms defined by auto-regression process. In MA process, the prediction value is also determined by the a number order(q) of previous random error. And the ARIMA equation can be defined by following formula.

$$X_t = c + \sum_{i=1}^{p} \theta_i X_{t-i} + \sum_{i=1}^{q} \phi_i \epsilon_{t-i} \tag{1}$$

Where c is constant, θ and ϕ are autocorrelation coefficients for previous observation and weights for stochastic term(ϵ)in series respectively. And p is the auto-regression order while q is moving average order which are used to determine the dependencies length.

These three steps are the cores of ARIMA, its regression-based structure which can be defined by only three internal variables(p,d,q) and initial process(change non-stationarity into stationarity by differential calculation) about non-stationary observation makes prediction simple and convenient with high accuracy and great success in many fields. However, ARIMA is not a perfect model for universe time series data due to its simplicity on structure and hypothesis. Some scholars have found various limitations in predictions which lays a foundation for those who would like to pay attention to explore the performance of deep learning in time seires forecasting. Therefore, we will review and summary some famous limitations on ARIMA in the following paragraph.

- Andreea-Cristina et al.[18](2016) found that ARIMA model can not fit the financial time series data with asymmetries, sudden outbreak at irregular time intervals very well. This is because that ARIMA assumes the differentiated financial time series data as perfect stationary with constant variance and mean, while, in nature, most differentiated financial exhibit variants in statistic and they can not be considered stationary. The assumption(the time series data or the data after difference is stationary) is too rigid and ARIMA model is limited by the stationarity of time series data because in nature the data can not always be considered as ideally stationary.
- According to the equation(1) and Jenkins Box statement[4] in 1976, ARIMA model is one special type of linear regression model which can only capture the linear feature of time series data, rather than both linear and non-linear information.
- In theory, ARIMA model can be used in both short term and long term application if time series information is stationary. However, Jenkins and Box[4](1976) has found that the non-stationary factors in trend and seasonal components of long-term series are more difficult to be eliminated, while shorter time series data can be more likely visualized as stationary. So, ARIMA model perform better based on short term forecasting than long term's typically.

Although scholars have tried to modify the ARIMA model to make it more affinity for non-stationary data, for example, seasonal ARIMA(SARIMA) model proposed by Box and Jenkins[4] in 1976 is aiming at removing non-stationarity from the series by setting seasonal difference order, the performance of ARIMA model for nonlinear data is still a struggle. The contribution of the ARIMA model to the prediction is undeniable, however, with the exploration of deep learning in the field of prediction and its out-performance shown in industry and research, complicated time series forecasting task seems can be perfectly solved. One question can be raised naturally, does deep learning will overtake statistic approach's place in time series forecasting? Zhang in 2003 [10] proposed a state-of-the-art direction of time series prediction task and he gave us a totally new perspective on the question. So before we review and evaluate this direction, it is necessary to discuss the contribution of deep learning to prediction task and start to rethink the time series data components.

2.3 Long Short-Term Memory(LSTM)

In the previous sections, we have discussed traditional statistic ARIMA method and its limitations on non-linear non-stationary time series forecasting. So in this section, we will review a popular deep learning model(Long Short-Term Memory) and their superiority for time series forecasting task compared to traditional method.

As Sima et al.[5](2018) stated that Long Short-Term Memory model is a special structure of recurrent neural network with the capability for long term memory. So in order to understand its superiority, it is necessary to have a glimpse of what a neural network looks like.

Sima et al.[5](2018) have listed a simplest neural network is made up of at least three layers, input layer, hidden layer and output layer, which consist of finite number of nodes(processing elements). This simplest structure neural network is neat but powerful, as Kihoro[19] illustrated that neural network structure can be considered as a model representing the intelligence of the human brain from the machine. In the forecasting task, the neural network structure is trying to approximate the target based on the regularities and patterns in the data, learned experience and also the known previous knowledge[12]. In addition, these nodes in one layer are connected through links with weight, which can be updated by iterative back propagation, to another nodes in upper layer. And these weight, which show the strength of links to layer can be trained by input data without any pre-request for raw information. In the hidden layer, the weighted data gotten from each hidden nodes will be applied to an activation function, which can transform linear to non-linear. And the output of hidden layer will be considered as input of output layer where a vector of probabilities for the various outputs will be generated based on hidden output and the one with minimum error rate or cost will be selected as optimum forecasting result[5].

Recurrent neural network is a specially type of neural network where the dependent hidden relationship between next step observation and previous one are connected through links(Yu et al)[20]. More specifically, the transformation in hidden layer can be mathematically expressed by: $h_t = \sigma(W_h h_{t-1} + W_x x_t + b)$ where x_t denote the input,h_{t-1} is the the recurrent information from the last time step, W_h and W_x are weights shared by hidden units and input respectively. As Yu et al[20] in 2018 said the RNN structure is like a perfect training black box for sequential data. However, many scolar have pointed that the long-term dependence problem is difficult to be solved in recurrent neural network, for example, in 2020 Zhu[21] predicted APPLE stock price using RNN model and he said that with the increase of required timesteps, the mean square error is becoming unacceptably large.Hochreiter et al.[22] in 2001 has analyzed that this problem mainly because some error signals which are back propagating through time always either blow up or vanish. So, in order to solve long-term dependence problem, a more advanced neural network structure was proposed by Hochreiter and Schmidhuber(1997)[23] called Long Short-Term Memory.

Long short-term memory model is more flexible non-linear approximator for modelling time series data with longer dependencies.Hochreiter and Schmidhuber(1997)[23] has pointed that LSTM are stuctured by special "gate" in cell which are used to improve the capacity of long term memory by adjusting the state of the "gate" and aviod the gradient vanishing problem.

- Hochreiter and Schmidhuber(1997)[23] said that forget gate in LSTM is considered as a filter used to determine whether to forget the previous cell state.

$$f(t) = \sigma(W_{fh} h_{t-1} + W_{fx} x_t + b_f) \tag{2}$$

As equation (2)[23], the output of f_t is a number between 0 and 1, where 1 illustrates

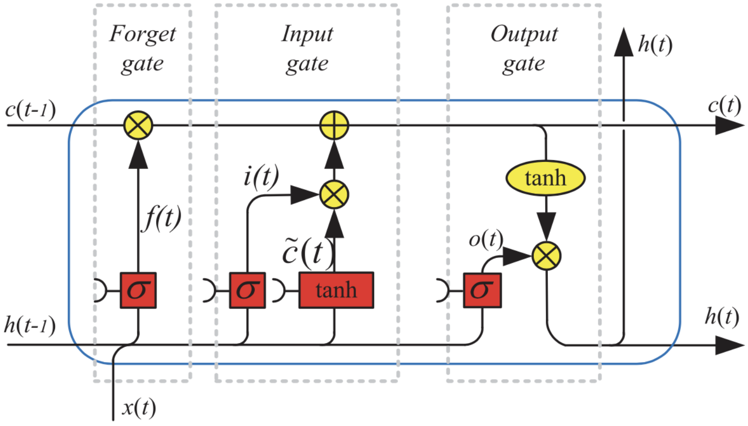

Figure 1: Architecture of LSTM[20]

completely keeping while 0 shows completely ignoring.

- The input gate in LSTM is used to update the cell state by adding previous cell state and the new cell state in this time step.

$$i(t) = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_i) \tag{3}$$

$$\tilde{C}(t) = tanh(W_{\tilde{C}h}h_{t-1} + W_{\tilde{C}x}x_t + b_{\tilde{C}}) \tag{4}$$

Where $i(t)$ can be also considered as forget gate used to determine which information in present time step and can be discarded. And the combination between $c(t-1)$,$f(t)$,$i(t)$ and $\tilde{C}(t)$ are the updated cell state output of this cell.

$$c(t) = c(t-1)f(t) + i(t)\tilde{C}(t) \tag{5}$$

- The output of output gate is the hidden state in current time step for next time forecasting based on the filtered cell state and added data.

$$o(t) = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \tag{6}$$

$$h(t) = o(t)tanh(c(t)) \tag{7}$$

In 1997, Hochreiter and Schmidhuber[23] have pointed out reasons why LSTM can solve the long-term dependencies problem mainly is the cell state which is uniquely set in LSTM and the "gate structure". Furthermore, Hochreiter and Schmidhuber(1997)[23] also illustrated that the cell state, as we can see in figure(1)[23], is more powerful to characterize the memory of a neuron that is less prone to decay, and construct long-term memory by storing the macro understanding and memory of past sequences with a smaller number of linear interactions compared to hidden state which are existed in both RNN and LStM. In addition, during the process of back propagation through time, the degree of attenuation for gradient conduction is controlled by gate structure to avoid the gradient vanishing problem, which means the gates determine the network how much the gradient can vanish. In 1998, Hochreiter[24] has worked on long time lags problems in different approaches and Hochreiter concluded that LSTM performed best compared to other approaches, involving time lags of more than 1000 steps.

2.4 Hybrid method - rethink the data components

Typically, the application scenarios of traditional prediction algorithms and deep learning prediction algorithms were segmented because most of traditional time series forecasting algorithm such as ARIMA is constrained on the hypothesis of series data(stationary hypothesis we mentioned in section 2.1 and 2.2), while deep learning limited by its high complexity and computing resources is more suitable for high accuracy requirements of the scene. However, in order to estiblish a model for universal data, in 2003, Zhang[10] proposed a novel hybrid approach to integrate the capabilities of the two types of models by rethinking the feature of time series data.

The hybrid method proposed by Zhang[10] in 2003 is straightforward without any new knowledge background. The idea of hybrid method is to rethink the components of time series data, compared to four components mentioned at section 2.1.1. Zhang(2003) has pointed out that, in real problem, it is reasonable to consider a time series in a higher level to be made up of a linear and non-linear patterns, because a time series can not be purely linear or purely nonlinear in real world problem.

$$y_t = L_t + N_t \tag{8}$$

Where L_t and N_t are denoted the linear and non-linear components respectively. In Zhang's paper[10] he chose ARIMA method to capture the linear pattern while the artificial neural network to predict the future value from non-linear components. And this process can be implemented by following:

$$e_t = y_t - \hat{L_t} \tag{9}$$

Where $\hat{L_t}$ is the linear component of time series calculated from ARIMA equation (1), while e_t is the residuals from the raw series contain only the non-linear pattern of time series which can be trained by deep learning method. Zhang(2003) also provide his evidence of the hybrid model performance. He compared three different methods using three dataset, the Wolf's sunspot data, the Canadian lynx data and the British pound=US dollar exchange rate data in his experiment and he has emphasized that the hybrid model performed better than either of single one.

Following Zhang's idea, in recent year, a state-of-the-art ARIMA-LSTM hybrid model has been proposed and few studies have tested its performance. In 2018, Choi[25] has pointed out that the performance of ARIMA-LSTM is far superior to other equivalent financial models in predicting stock price correlation coefficient. And in 2019 TEMÜR et al.[26] predicted the house sales in Turkey using ARIMA, LSTM and ARIMA-LSTM model and they found the MAPE (Mean Absolute Percentage Error) and MSE (Mean Squared Error) values obtained from hybrid model are much lower than the other two. However, Zhang(2003)[10] has raised his concerns in 2003 that traditional - deep learning hybrid method may predict a suboptimal results because the implementation of combining traditional model and neural network model tend to require subjective judgement of model order as well as the model adequacy which may result in missing the optimal solution in the prediction process.

3 Summary & Conclusion

In this literature, we provide an overall picture of hybrid methodology for time series forecasting step by step. As combination of statistic forecasting algorithm[4][12] and deep learning(DL) model[23], hybrid model takes advantages of both strength of the traditional and advanced. From the aspect of statistic, ARIMA as the mainstream of traditional algorithm is easy to implement and provide us an idea of data decomposition.However, we have reviewed[18][4] that ARIMA as a linear model can only capture the linear relationship of time series and it request strict initial assumptions for series which limit its performance on long-term memory.

These limitations from traditional algorithm can be much improved by deep learning method. As a data-driven and self-adaptive model, deep learning shows its superiority on initialization without any prior hypothesis for the distribution of data[12]. In addition, long-term problem can be successfully solved by LSTM model consisting of "gate" structure modified by RNN[23]. However, in 2003, Zhang[10] has pointed out a problem that scenarios for different models are segmented due to lack of a method that works for universal data.

To solve this problem, Zhang[10](2003) proposed a novel hybrid model combining statistic with deep learning by rethinking series data into linear and non-linear components.In current state, the-state-of-art traditional-deep learning hybrid ARIMA-LSTM has benn applied in a limited number of scenarios of time series forecasting[26][25][27] and it has achieved superior performance on accuracy and precision compared to each individual one. However, due to the lack of sufficient evidences to prove its superiority and its limitation on model selection, the mainstream of time series forecasting task is still dominated by pure deep learning[6]. From the perspective of the future, as more advanced deep learning models are proposed, the flexibility and availability of hybrid model are conductive to be considered as an auxiliary technique to improve the accuracy for these advanced deep learning algorithm in particular application scenario where further research is call for.

References

[1] Jan G. De Gooijer Rob J. Hyndman. Time series and forecasting: Brief history and future research. *Journal of the American Statistical Association*, pages 638–643.

[2] AK Palit D Popovic. Computational intelligence in time series forecasting: theory and engineering applications. *Advances in Industrial Control. London: Springer London, Limited.*, 2005.

[3] Mehdi Khashei Mehdi Bijarir. A novel hybridization of artificial neural networks and arima models for time series forecasting. *Applied Soft Computing 11, no. 2 (2011): 2664-675.*, 2010.

[4] P. Box G.M. Jenkins. Time series analysis: Forecasting and control. 1976.

[5] AKBAR SIAMI NAMIN2 SIMA SIAMI NAMIN1(2018). Forecasting economic and financial time series: Arima vs. lstm.

[6] Omer Berat Sezer Mehmet Ugur Gudelek Ahmet Murat Ozbayoglu. Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing 90 (2020): 106181.y.*

[7] Ayşe SOY TEMÜR Melek AKGÜN Günay TEMÜR. Predicting housing sales in turkey using arima, lstm and hybrid models. *Journal of Business Economics and Management 20, no. 5 (2019): 920-38.*

[8] Yu Zhang Guoguo Chen Dong Yu Kaisheng Yao Sanjeev Khudanpur James Glass1. Highway long short-term memory rnns for distant speech recognition. 2015.

[9] Zhiheng Huang Wei Xu Kai Yu. Bidirectional lstm-crf models for sequence tagging. 2015.

[10] G.P Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing (Amsterdam)*, 70:p159–175, 2013.

[11] Jürgen Kirchgässner, Gebhard Wolters. *Introduction to Modern Time Series Analysis.* Heidelberg: Springer-Verlag, 2007.

[12] Adhikari Ratnadip Agrawal R. K. An introductory study on time series modeling and forecasting. 2013.

[13] Warren Milton Persons. *Indices of Business Conditions: An Index of General Business Conditions.* Harvard University Press, 1919.

[14] Estela Dagum. Time series modelling and decomposition. *Statistica*, 50, 05 2013.

[15] K.W. Hipel A.I. McLeod. Time series modelling of water resources and environmental systems. *Amsterdam, Elsevier*, 1994.

[16] T Kantz, H. Schreiber. *Nonlinear Time Series Analysis.* Cambridge: Cambridge University Press., 2004.

[17] John H. Cochrane. Time series for macroeconomics and finance. *Graduate School of Business, University of Chicago, spring*, 1997.

[18] Andreea-Cristina PETRICĂ Stelian STANCU Alexandru TINDECHE. Limitation of arima models in financial and monetary economics. *heoretical and applied economics, XXIII(4)*, page pp.19–42., 2016.

[19] J.M. Kihoro R.O. Otieno C. Wafula. "seasonal time series forecasting: A comparative study of arima and ann models". *African Journal of Science and Technology (AJST) Science and Engineering Series Vol. 5, No. 2*, pages pp.41 − 49.

[20] Yong et al Yu. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation, 31(7)*, page pp.1235–1270.

[21] Yongqiong Zhu. Stock price prediction using the rnn model. *Journal of physics. Conference series, 1650(3),2020–10-01, Vol.1650 (3).*

[22] Stefan C Kolen, John F Kremer. Gradient flow in recurrent nets: The difficulty of learning longterm dependencies. *In A Field Guide to Dynamical Recurrent Networks. John Wiley Sons.*

[23] Sepp Hochreiter Jurgen Schmidhuber. Long short-term memory. *Neural Computation.9(8),1735:1780*, page pp.19–42., 1997.

[24] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solution. *International Journal of Uncertainty*, 1998.

[25] Hyeong Kyu Choi. Stock price correlation coefficient prediction with arima-lstm hybrid model. 2018.

[26] A. S Albayrak. Arima forecasting of primary energy production and consumption in tur- key: 1923–2006. 2010.

[27] Melek Temür Günay Temür, Ayşe Soy Akgün. Predicting housing sales in turkey using arima, lstm and hybrid models. *Journal of business economics and management, 20(5)*, 2019.