JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

ANALYSIS OF UNIVARIATE TIME SERIES IMPUTATION TECHNIQUES IN PRESENCE OF MISSING DATA

[1]V.Sowjanya, [2]Y. Surekha, [3]Krishnaveni Kommuri

[1]Associate Professor, Potti Sriramulu Chalavadi MallikarjunaRao College of Engineering and Technology, Vijayawada, *sowji635@gmail.com*
[2]Prasad V Potluri Siddhartha Institute of Technology, Vijayawda
[3]Koneru Lakshmaiah Education Foundation, Vaddeswararn, AP

Received: 16 March 2020 Revised and Accepted: 16 June 2020

ABSTRACT

Time series forecasts play an irreplaceable role in many applications of time-varying processes like energy markets, financial markets, etc. It is important but difficult to predict the dynamism of time-varying systems as it depends not only on the nature of the system, but also on external factors, such as environmental conditions and social and economic status. Current Neural Networks (RNNs) are a special class of neural networks with recurrent internal links that allow the dynamic nonlinear system to be modeled. The various predictive tasks were recently implemented and show that they surpass the predictive exactness relative to standard time series predictive models. However, in the presence of missing details, there is a limited study of time series forecasting using RNN. In this study, we propose a new model using the Dilated RNN(DRNN) and an amended mechanism of attention, which focuses on the question of time series prediction with missing data. With relation to benchmark data set predictability, the proposed model outperforms existing models such as the Auto Regressive Integrated Moving Average(ARIMA) and Gated Recurrent Unit(GRU).
KEYWORDS: Auto Regressive Integrated Moving Average(ARIMA), Current Neural Networks, Gated Recurrent Unit(GRU).

1. INTRODUCTION

1.1 Short-Term Load Forecasting of Electricity Demand in Time Series

The electricity as a type of energy resource has a unique characteristic that must be generated when required because no electricity can be stored[5]. The irregularity in demand for electricity rises in modern society due to an rise in population, new business openings, an increase in energy supplies, and so forth. The value of demand forecasts is thus stressed so that the resources that are temporally and quantitatively limited can be handled and distributed efficiently. Indeed, the prediction of future energy supply demand is key to the management of the limited availability of resources in an energy supply network[4]. The precise provision is a crucial factor for the planning and operation of electricity systems for the electricity industry. This results in significant savings on running and maintenance costs, improved power supply and control networks reliability and right future planning decisions[5].

Nonetheless, for some reasons load prediction is a difficult task. First, the time series has several period changes. The load for a given hour, for example, depends not only on the load of the previous hour, but also on the same time on the last day and the same hour the week before [6]. Secondly, it also depends on the exogenous variables, including the environment, climate variables, human activities, etc.[7].
This motivates research into prediction models that can boost this financial and social effect by improving only a small portion of prediction accuracy. For decades, the work has been involved in developing accurate models of load forecasts for the market for electricity. Exponential smoothing, ARIMA, neural networks, and SVM support are the most significant methods for STLF efficiency.

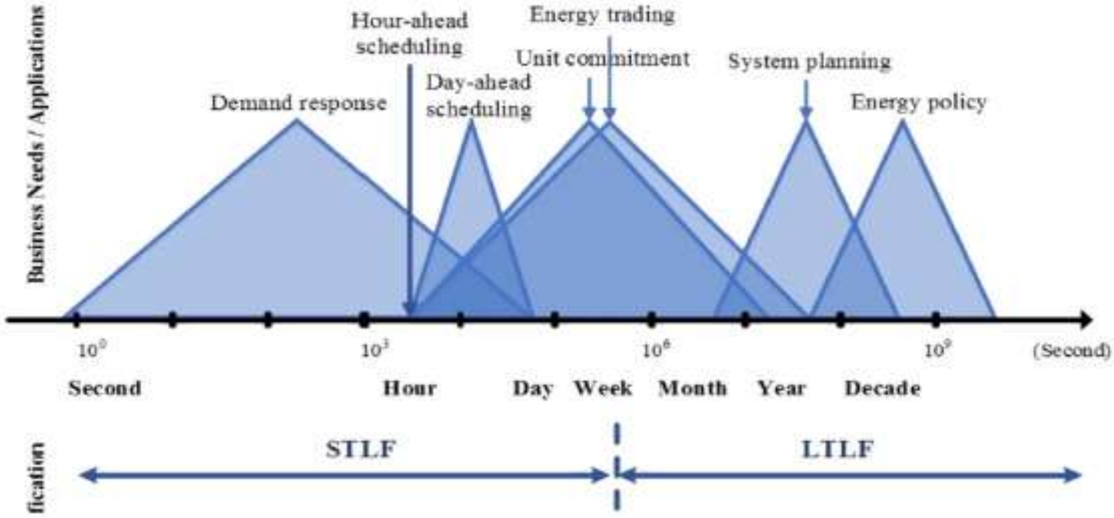

Figure 1.1: Classification of the forecasting application according to the forecast time in electric power generation.

The time series data are prevalent in our day-to-day lives, such as medical records, sales income, new user frequency and others, which have made data analysis conscious of the value of these data. The key method for data processing is typically data mining. The accuracy of the data obtained is an important factor influencing the performance of data mining. It means increased consistency of the data obtained, better performance in data mining. Nonetheless, in fact, due to insufficient data collection, noise or system failure, some time series data may be missing.

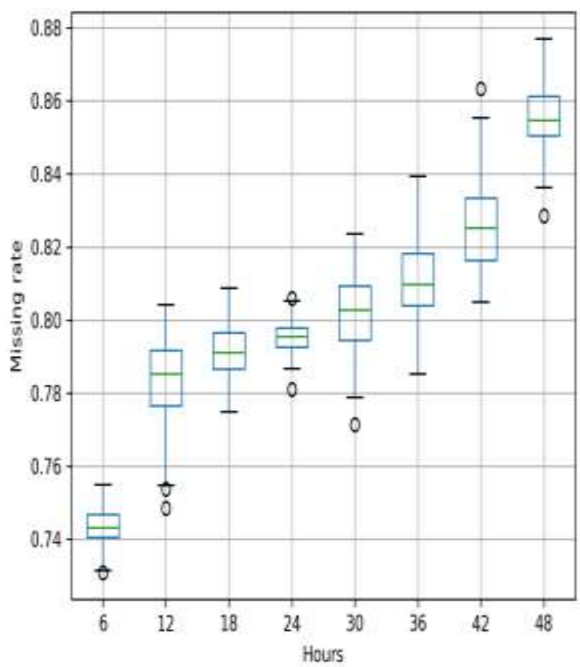

FIGURE 1.2. The boxplot stands for mean, upper quartile and lower quartile missing rates at record hour. The roughly average missing rate is 80.17 percent.

Besides, the collection time interval of time series cannot catch the requirements. Fig. 1.2 boxplot shows the missing rate of the Physionet dataset [1] in different time interval. We found that the missing average keeps a higher rate at each timestamp as time goes on. Consequently, missing information in time arrangement makes it difficult to complete any sort of time arrangement task in light of instructive missingness. These discoveries exhibit the significance of missing information ascription process in understanding a period arrangement investigation and expectation task. To process the missing qualities, there are two bearings for determination:

empty information erasure and missing worth attribution. The first is a direct methodology. It chooses to disregard missing qualities and manufactures model just dependent on the watched information [2], which cause harm to precision of the model with higher missing rates. At the end of the day, it is just proper for the very little missing rates. On the other hand, the subsequent one is a superior route for inadequate datasets, and finds appropriate qualities for the fragmented gaps. Moreover, missing worth ascription mode can be partitioned into two strategies.

2. RELATED WORKS

2.1. Time Series with the Missing Values Classification Problem

A lot of writing has been distributed on time arrangement with missing qualities. A large number of these works center around the attribution of missing qualities. The order issue can be unraveled after the attribution system utilizing customary grouping strategies, for example, piece strategy bolster vector machines and arbitrary timberland. In any case, a large portion of the conventional strategies can not straightforwardly handle multivariate time arrangement of differing length. Futoma et al. utilized perform multiple tasks Gaussian procedures to straightforwardly change the unpredictable examined time arrangement with missing qualities into an increasingly uniform portrayal and afterward do the grouping task. As of late, Mikalsen et al. proposed a portion technique called time arrangement group bit (TCK) to gain proficiency with the likenesses between multivariate time arrangement. It can straightforwardly deal with time arrangement of changing lengths and with missing information without an ascription strategy. Nonetheless, the most extreme missing rate appeared in this exploration is half, which isn't sufficient to manage gigantic missing qualities in clinical time arrangement and TCK expect that missing qualities in time arrangement happen indiscriminately. Afterward, they proposed an improved portion to misuse the educational missingness when the missing qualities happen non-arbitrarily. Other than part strategies, analysts likewise pick RNN to address the missing qualities issue in time arrangement. Yoon et al. shown that the data inside factors is as significant as the data across factors in the method of ascription. They utilized multi-directional intermittent neural systems with addition squares and ascription square to ascribe the missing qualities. Lipton et al. considered the missing imprint significant. They linked the missing imprint and the time arrangement and information them into LSTM. They utilized the forward or in reverse ascription techniques to address missing qualities. On the off chance that the variable in a period arrangement is totally missing, it will be loaded up with master experience esteems. Afterward, Lipton et al. proposed an improved technique, which is added to the mean worth, standard deviation, and the first and the last perception of every factor in a multivariate time arrangement into the model. These additional highlights describe the distinction of every factor, except they are determined physically and continue as before during the preparation procedure. Che et al. changed the system of GRU and regarded the missing imprint as another info and took care of both missing imprint and time arrangement into a GRU model, called GRU-D. Harutyunyan et al. proposed channel-wise LSTM. They input a solitary variable alongside its missing imprint into an autonomous bidirectional LSTM layer and afterward linked all the yields and took care of them into the equivalent LSTM layer. Along these lines, the data identified with univariable can be scholarly before the factors are blended in LSTM. Be that as it may, on the grounds that every factor should be prepared in a free system, it causes escalated calculations.

2.2. Clinical Time Series Analysis Using Deep Learning

In the clinical field, scoring techniques to assess the state of patients, for example, SAPS-II SOFA and APACHE have just been utilized by and by. The vast majority of these scoring strategies utilize straightforward models, for example, calculated relapse, to play out the order task, and the information for preparing are chosen physically. In any case, the exactness of these strategies has been questioned by before contemplates. With the improvement of profound learning, scientists have attempted to supplant these basic models with profound learning models to deal with clinical time arrangement issues. There are two principle issues in clinical time arrangement: missing qualities and unpredictable inspecting rate. To manage missing qualities, the main methodical investigation of clinical time arrangement issue dependent on RNN was accounted for by Lipton et al. in 2016. Choi et al. utilized RNN to play out an alternate conclusion on clinical information with determination codes, drug codes and technique codes, and they called it specialist AI. Strauman et al. just applied GRU-D to distinguish careful site disease and included a weighting plan in figuring misfortune because of the class irregularity in clinical information. Purushotham et al. proposed outfit learning strategies to join GRU and forward neural systems, which is known as the multi-modular profound learning model (MMDL) to all the while become familiar with the time-variation data and nontime-variation data. In the part of managing sporadic testing rate, Che et al.

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

proposed a profound generative model to catch transient conditions in multi-rate multivariate time arrangement. Bahadori et al. utilized an enlargement method to blend the record all the more intently divided in time from the start and put the prepared information into a neural system classifiers. Shukla et al. decided to manage missing qualities and sporadic inspecting rate issues simultaneously. They presented a two-layer insertion system to introduce multivariate time arrangement. The main layer performs univariate changes independently and the subsequent layer combines data over all measurements.

3. RECURRENT NEURAL NETWORK PROPERTIES AND TRAINING

3.1 Properties

When all is said in done, RNNs are a learning model that refreshes new state ht utilizing past state ht_1 and current information xt recursively. It can likewise be depicted by a system, which is made out of different cells associating in arrangement along the time hub. Every cell in the system registers state at a specific time step. Figure 3.1 represents forward spread of RNN in two different ways. Cyclic shape in the left is called 'collapsed diagram' while non-cyclic shape in the privilege is called 'unfurled chart' which extends cyclic shape in time. Note that boundaries in the unfurled portrayal share their qualities over the cells and the qualities are refreshed at the same time inside the enhancement technique.

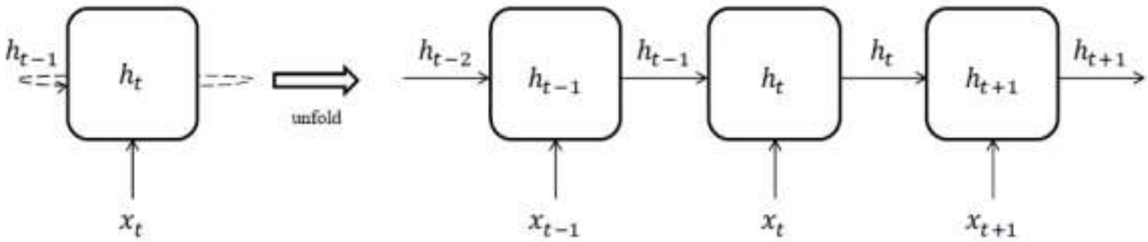

Figure 3.1: Folded graph of RNN (left) and the unfolded in time (right) during forward propagation.

Contingent upon the way to deal with update the new state ht, structures of RNN di↵er, for example, Elman RNN (ERNN), Long momentary memory (LSTM) and Gated repetitive unit (GRU). Further examination in this section will concentrate on ERNN since it has the most basic and fundamental design among them.

3.2 Training ERNN

When all is said in done, preparing implies a procedure that a model learns the ideal boundaries with a preparation set by limiting characterized blunder work which relies upon trainable boundaries. Preparing is made out of 3 stages, forward engendering, backpropagation and boundary update individually [4]. In RNN, backpropagation step is called backpropagation through time (BPTT) as the slope of the mistake must be engendered through the unfurled variant of the RNN diagram..

3.2.1 Forward propagation

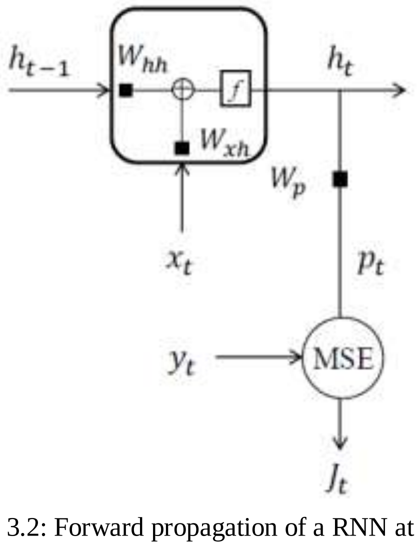

Figure 3.2: Forward propagation of a RNN at a time step t.

The design of ERNN cell, explicitly non-straight initiation work f(·) of state ht, causes disappearing or detonating angle issue when the slope backpropagates through time since it creates the result of fractional subordinate of ht_1 as for ht__1 for each retrogressive time step. Shortened BPTT(k2, k1) is one answer for this

JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

issue. Be that as it may, because of the truncation, the RNNs don't backpropagate longer than k2. It might cause another issue where the RNNs can not catch the drawn out conditions more prominent than k2 in the preparation.

LSTM and GRU both give longer range where the RNNs can learn without evaporating or detonating inclination by changing the design the RNN cell. The new methodologies both used gated structure that empowers the systems update data without non-straight actuation work.

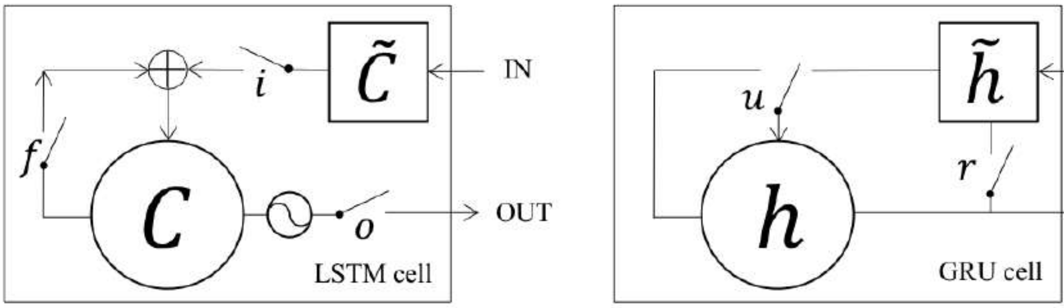

Figure 4.1: Schema of two RNN cells, LSTM (left) and GRU (right). GRU has a simpler architecture with the less number of gates than LSTM.

3.3 Long Short-term Memory (LSTM)

Cell status Ct and three door layers are introduced for LSTM, such as gate, gate input and output gate. The cell state heritages previous cell information and transfers it to the next cell. The gates determine the amount of information to be inherited, updated and transmitted to the next cell through cell interaction. The unfolded RNN graph with LSTM cells is shown in Figure 4.2.

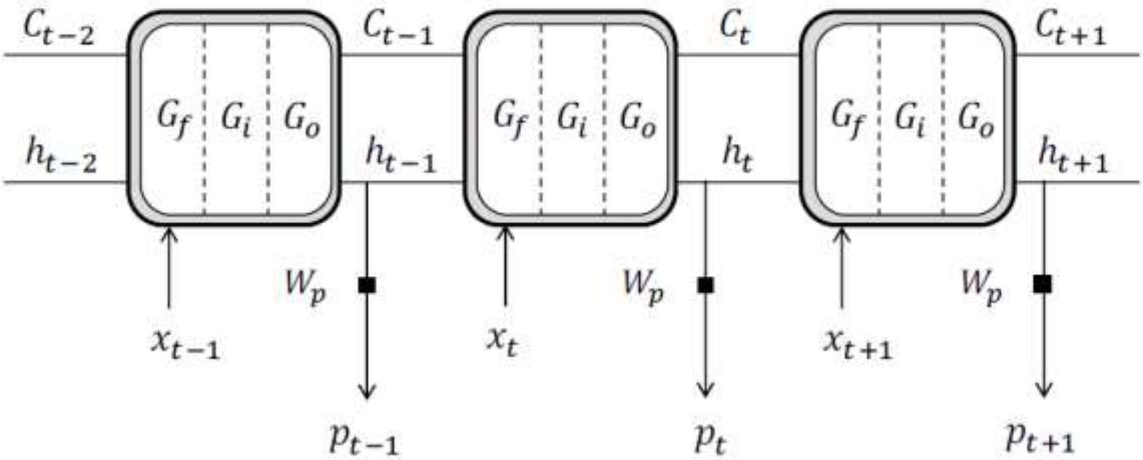

Figure 4.2: Unfolded graph of a RNN with LSTM cell that consist of three gates.

A. BRIEF REVIEW OF GRU NETWORKS

Given a time series observation with no missing T-long data, as X = {X0; X1;:::; Xt}, when the t-th value observed by Xt is indicated, a GRU network is able to render the input sequence data H = {H0;Ht}, when it is n and m , respectively, input time serial dimensions and output hidden sequences.

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

$$\begin{aligned} R_t &= \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \\ Z_t &= \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \\ \tilde{H}_t &= \tanh(X_t W_{xh} + (R_t \otimes H_{t-1}) W_{hh} + b_h) \\ H_t &= Z_t \otimes H_{t-1} + (1 - Z_t) \otimes \tilde{H} \qquad (1) \end{aligned}$$

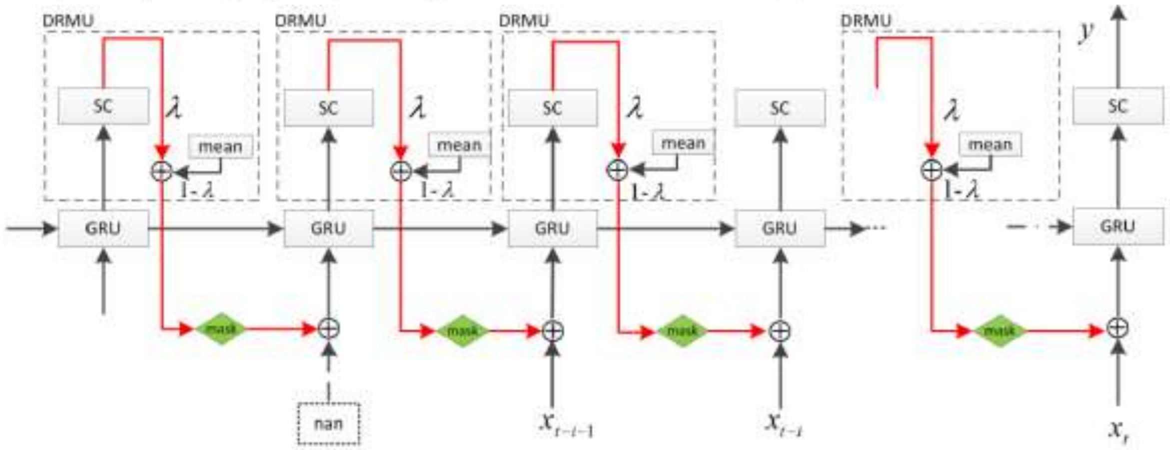

FIGURE 4.3. The architecture of DRMI-GRU. We use mean to denote the average of all observed sequential data; SC is the shortcut connection structure; mask denotes the masking matrix as a switch; nan is the missing time series in dataset.

The training cycle in the DRMI-GRU model involves estimation and imputation, and the proposed model structure is indicated in the Figure. 4.3. The evaluation process is applied if data input is detected and one is masked in the matrix, and the imputation process begins if data are missing.

4. CONCLUSION

In this paper, we first analyzed various forms of RNN for the time series prediction mission, explained their internal mechanisms, spoke about their properties and training processes. In the forecasting accuracy of the baselines, such as ARIMA and GRU, the model proposed showed substantial improvements with synthetic dataset. We found interesting phenomenons, which deserve further study, from the experiment with real world data, although we could not achieve the same degree of progress as in synthetic data, with regard to the accuracy of predictions in baseline models. The proposed models and findings presented in the thesis would contribute to a better time series forecast and eventually help to improve resource management performance. An accurate forecast of time series plays an irreplaceable role in the management and distribution of capital in industrial sectors. The DRMI-GRU model for recovering missing values in an end-to - end way from incomprehensive public data sets and real world datasets. We use DRMU and masking matrix in the GRU network, called DRMI-GRU, to control knowledge lack. This approach has been tested on different datasets, with a different value ratio.

5. REFERENCES

[1] I. Silva, ``Predicting in-hospital mortality of ICU patients: The PhysioNet/Computing in cardiology challenge 2012,'' in *Proc. Comput. Cardiol.*, Krakow, Poland, Sep. 2012, pp. 245_248.
[2] J. Kaiser, ``Dealing with missing values in data,'' *J. Syst. Integr.*, vol. 5, no. 1, pp. 42_51, 2014.
[3] P. J. García-Laencina, J.-L. Sancho-Gómez, and A. R. Figueiras-Vidal, ``Pattern classi_cation with missing data: A review,'' *Neural Comput. Appl.*, vol. 19, no. 2, pp. 263_282, 2010.
[4] L. F. Shen, Q. L. Ma, and S. Li, ``End-to-end time series imputation via residual short paths,'' in *Proc. 10th Asian Conf. Mach. Learn.*, Beijing, China, 2018, pp. 248_263.
[5] G. E. A. P. A. Batista and M. C. Monard, ``An analysis of four missing data treatment methods for supervised learning,'' *Appl. Artif. Intell.*, vol. 17, nos. 5_6, pp. 519_533, 2003.
[6] S. Hochreiter and J. Schmidhuber, ``Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735_1780, 1997.
[7] Z. Che, ``Recurrent neural networks for multivariate time series with missing values,'' *Sci. Rep.*, vol. 8, no. 1, p. 6085, Apr. 2018.

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

[8] Z. C. Lipton, D. C. Kale, and R. Wetzel, ``Directly modeling missing data in sequences with RNNs: Improved classi_cation of clinical time series," Jun. 2016, *arXiv:1606.04130v2*. [Online]. Available: https://arxiv.org/abs/1606.04130v2
[9] J. L. Schafer and J. W. Graham, ``Missing data: Our view of the state of the art," *Psychol. Methods*, vol. 7, no. 2, p. 147, Jun. 2002.
[10] Polagani Rama Devi, V.Sivaparvathi, M.Sailaja, J.Malathi "*Learning Grammar for Web Server URIs with High Accuracy using Recurrent NeuralNetwork Architecture" in* Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 04-Special Issue, Aug 2018 ISSN: 1943-023X
[11] K. Cho, ``Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. EMNLP*, Doha, Qatar, 2014, pp. 1724_1734.