A REVIEW ON LONG TERM SHORT MEMORY (LSTM)

Abhishek Patel
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India

Shabnam
School of Computer Science and Engineering
Lovely Professional University,
Phagwara, Punjab, India.

Abstract—Long Short-Term Memory (LSTM) network is widely applied in research areas having complex solution having sequential data as text, video format. LSTM handles the long-term dependencies problems in its cell structure (sigma or logic cell). These paper consist review of different RNN-LSTM long-term dependencies applicable areas along with applied problems and its effective model of implementation which has implemented and its future scope. In addition, it is seen as RNN-LSTM network performs well than other machine, deep learning algorithm while predicting the results.

Keywords: Long Short-Term Memory, Recurrent Neural Network, Deep Learning, Natural Language Processing.

I. INTRODUCTION

LSTM (Long Short-term memory) is a complex and enhancement form of deep learning and falls under the recurrent neural network. Its architecture and features like feedforward connection are identical as recurrent neural network (RNN). LSTM form of neural network process data in a sequence order rather than single data flow. So, LSTM from of neural network are mostly used in recognition filed as hand, speech, videos recognition and many more areas of network traffic detection or intrusion detection.

The architecture of the LSTM network consists of cell, input, output and forget gate. The function of cell is to remember the value after every sequence of time-interval and three different gates helps regulates information in and out of the cell. Input gate takes the new value, the forget gate manages the value to which extend it remain in cell and output gate handles the value used for activation of LSTM network. The activation function can be anymore but logistic sigmoid function are commonly used.

LSTM networks performs well in time series data, as there are many lags between the duration of events in real time scenario. LSTMs were created to manage vanishing gradient problems with exploding which can be experience while training RNNs. Relative intensity helps in disparity length which is main assets of LSTMs than RNNs.
Let us assume an example of predicting stock prices. Current stock price can calculate on basis of previous prices which can be in form of uptrend, downtrend or both at a same day. New stock factors depend on any huge policy change of company which is not accepted globally or change in senior administration member of company. So in predicting the todays prices we have to consider all the factors the previous, the hidden and the current scenario. We LSTMs is used as it has a cell to remember all factors and predict out the output in more accurate way.

The RNN with LSTM network is trained in a supervised way with a set of optimization algorithm such as backpropagation combined with a gradient descent which optimizes the computation by changing the weight of LSTM network in proportion to Output layer (error) with its corresponding weights. These entire scenario is known as Gradient vanish. LSTM are trained connectionist temporal classification (CTC).

II. ARCHITECTURE OF LONG SHORT TERM MEMORY

A. WORKING OF LSTM NETWORK

In LSTM firstly, we decide which information to keep in or out of the cell which is done by sigmoid function layer knowns as a forget gate. Secondly, we decide which new data we will store in cell state. Storing of data categories in two state. First one is sigmoid layer (known as input layer), which concludes the value which should be updated, second one is tanh layer (a vector) of new values which can be added to the state. Now, both sigmoid and tanh layer will combine to decide a next state of update.

Now, while adding some new value to the cell the old value will replace by new one by forgetting old cell state. New value will be decided by previous state decision. After replacing and computing the final output will be decided. The final output will be filtered version generated on basis of previous and current cell state. Firstly, the sigmoid layer will decided which state of the cell will be output. Then these will go for a tanh (it margins the value between -1 to 1) and multiply by sigmoid gate which is the output. We multiply with sigmoid layer to get the output which we have decided. There exits a variation in forget and input gate.
Instead of making separate decision regarding the things to forget. There must be new data added to those cell state only where we want to add rest will be same.

As for example, In language model while predicting out the next word the output we want should be relevant to many grammatical forms such as singular, plural, verb and many more. So, previous and new predicting output from a cell state should match for correct prediction.

### III  Literature Review

F. Adeeba and S. Hussain in 2019 [1] proposed a native language identification through LSTM. They use spectrogram and cochleagram for extraction of speech with bidirectional LSTM (BLSTM). An experiment is set up with network architecture and accuracy was find out on basis of validation dataset with help of Mel-frequency cepstral coefficient (MFCC) and Gammatone frequency cepstral coefficient (GFCC). Both MFCC and MFCC are combined with BLSTM network to increase the accuracy. Additionally, search of network architecture was made after merging BLSTM models and observed that accuracy of these network increases while using the hidden layer above one and the best accuracy was obtained on the use of three hidden layer used in GFCC feature BLSTM network. The two hidden layers in GFCC feature BSTM network on validation dataset gives highest accuracy.

Jonathan Mackenzie et al. in 2019 [2] evaluates HTM (hierarchical temporal memory) and LSTM (long short-term memory) in predicting traffic flow. Big data has huge amount of data available, and its system is capable of processing those large data and machine leaning algorithm process them to predict results and combination helps in predicting traffic flow. Here, HTM and LSTM networks with batch and online learning are evaluated on real time scenario of Sydney in Southern Australia. HTM works better than LSTM and all other deep learning technique. It showed that LSTM shifted from motor and highway to aerial road domain. HTM performance might degrade while increasing a greater number of hierarchy but it is best for predicting short term traffic. Many possible future works can be performed as optimizing HTM parameter, investigating single LSTM, adding additional parameters and investing in anomalous traffic.

Albert Yaw Appiah in 2019 [3] used LSTM network in feature extraction as photovoltaic array and detection of fault. Photovoltaic diagnosis helps to decrease the losses of energy, reduces electric shocks and fire hazard. The PV array technique used for diagnosis is capable of extracting features automatically from huge and raw data itself and helps in classification using deep learning and LSTM network. Extracted feature is feed in SoftMax Regression algorithm for diagnosing a fault and compared with other techniques available and found that proposed technique is obtaining high accuracies on all kind of noiseless as well as noise data. Manually, feed of features on PV array for fault diagnosis consume time and is expensive so these papers has automatically featured extraction PV array technique to avoid all anomalies present in manual feeding technique.

YongZhi Zhang et al. in 2018[4] purpose a LSTM network for predicting remaining life of lithium batteries. Remaining useful life (RUL) increases battery reliability to determine failure and mitigate of battery. Existing technique for RUL is inefficient for long-term dependencies in terms of capacity degradation. The LSTM recurrent neural network learn the long-term dependencies in lithium batteries. These network adaptively optimized resilient mean square with backpropagation and drop-out technique which address the
overfitting problem and increases predicting ability of LSTM RNN model. LSTM RNN network dependencies degrades capacities and it explicitly construct capacity-oriented RUL predictor. RUL predictor helps to increase long-term learning with help of SVM model, particle filter model and RNN model. Monte Carlo simulation generates probabilistic RUL prediction. Data collected from many lithium batteries are used in constructing a model, verification and compared result for predicting battery life when offline data is also present.

Baowei Wang et al. in 2019 [5] with acceleration of urbanization there is tremendous effect on environment as air quality degrades with change in smog concentration which effect on mental and physical health. IOT technique with LSTM is used to monitor and process the data and prediction of next data using neural network. Existing prediction technique have limitation which is overcome by LSTM and Gated Recurrent Unit (GRU). Double layer of RNN was set for prediction of PM2.5 value. The use of both LSTM & GRU use historical data to predict the future air pollution status. Experimental is conducted from data available at national environment protection department and some data are IOT monitored data. First sample data of four cities for 96 consecutive hours were chosen and experiment is performed and observed experimental values are close to true value. Daily smog data of 74 cities for consecutive year from 2014-2018 are chosen as training and test set in ratio (70::30) and observed that model predict the better result. In future, the enhancement in IOT technology and neural Network optimization technique for error correction can improve the data collection accuracy and prediction of future smog data with more accuracy.

Yu Wang et al. in 2018 [6] integrated lasso and LSTM for solar intensity forecasting. The internet of things (IOT) and smart grid where energy management is difficult for utilizing bestreset of renewable energy source such as wind and solar. This paper has absolute shrinkage, LASSO and LSTM integrated together for forecasting solar intensity depend on meteorological data. Basically, it is the combination of time-series model, clustered data, statistical model with machine learning technique. Each cluster has formed using a K means++ and each of them has separate forecasting model which is formed by LSTM model. Nonlinear relationship is established between cluster model and LASSO was applied to establish linear relationship between the data. This trace-driven simulation is used to validate entire model under many timescale and benchmark.

Yang Sun et al. in 2019 [7] proposed a model for solving monaural domain separation with LSTM network. LSTM with DNN (Deep Neural Network) is used for monaural speech differentiation is separated into three different aspect as signal approximation (SA), masking and mapping technique. However, traditional method on performance is not enough robust due to real world changing scenario. Vanilla DNN method we can’t utilize information fully, so LSTM network is proposed for complex signal approximation which utilizes phase information of required signal to increase the separation performance. The experiment was conducted based upon the cSA LSTM RNN methods to measures the performance based upon the objectives and compare with oSA-based method. Here cSA method is used for both cleaning magnitude and phase information.

Yanjun Qin et al. in 2019[8] proposed a transportation recognition model (TRM) using a convolutional learning along with a LSTM RNN network. Increase in sensor-rich devices has attracted much attention in
transportation mode which help in improving traffic management, urban and journey planning. Early machine learning technique can’t give a reasonable and effective feature. So, these papers has CL-TRANSMODE, which is algorithm of deep learning for recognition transportation. Here, firstly CNN model was used to train appropriate and robust mode for TRM. Then LSTM learns further dependence characteristics which is a vector (CNN output). Further enhancement is done by extracting peak features, artificial segment. Combining of these both CNN and handcrafted features generates more appropriate result with differentiating eight different transport mode. These entire models are experimented both SHL dataset (barometric data) and HTC dataset.

Zisis I. Petrou and Yingli Tian in 2019 [9] proposed a Convolutional LSTM model for predicting the motion of sea ice which will help in navigating ship, fisheries, gas exploration and climate predicting model. Deep learning model are used for prediction as it has a data of several days of motion of ice. The proposed technique has encoder-decoder layer of deep network with convolutional LSTM. Optical flow data is obtained from satellite microwave and scatterometer which covers a entire arctic area. The convolutional network proves long-time dependency with time series spatial correlation among neighbouring motion vector. Unsupervised end-to-end technique is used which required manual annotation. The model proposed in these papers can predict motion of ice in sea up to 10 days. Training is done by AMSR-E image. The AMSR2 and ASCAT data are not used for training but also model gave accurate result in these approaches also.

Hamid Palangi et al. in 2016 [10] use a LSTM network for sentence embedding for information retrieval. This paper has a model address sentences using RNN with LSTM cells. Model pick up each word from sentence, extract its features and embedded in semantic vector. LSTM network capture ability reaches till last word and hidden layer provides semantic representation of entire sentence. The LSTM_RNN network was trained in a supervised manner where data are logged from a commercial search engine. Visualization are performed on basic of embedded work process where model automatically attenuate unnecessary words with detection of salient keywords. Further, these salient keywords to activate different cells of LSTM model and RNN is used to activate the cell where words belong to similar topics. Semantic representation helps to embedded vector which can be used furthermore. These combine model of LSTM network enables network to achieve document retrieval, complex language processing function where similarity can be calculated between query and entire sentence. This model performs well on web based documentation retrieval task with noise robust with semantic vector representation in an embedded method and important for language processing techniques.

Yonggang Liu et al. in 2019 [11] has done a case study on identification of intention driving in shifting strategy. As identification of intention driving is important references given to intelligent control system of vehicle. The collected data from vehicle are marked by K-means clustering technique. The LSTM network is embedded to detect the online longitudinal intensity with high precision. Driving intentions of running vehicle on straight, flat road is categories in five type such as vehicle speed, acceleration is processed to label test data of road. LSTM classification technique(model) identifies driving intension with opening angle of accelerator pedal, vehicle pedal and pedal brake force. The result of identifying reveal that highest
accuracy of model (proposed algorithm) attain to be 95.36% which is nearly 20% more than previous back propagation algorithm (BPA). Finally, gear shifting technique gains to generation recognition algorithm and simulation result with effective accuracy and less shifts as well as with better fuel economy.

Xiaodong Li et al. in 2018 [12] proposed a used of long short-term memory (LSTM) network for training novel algorithm. Due to effective potential of LSTM network different training algorithms are developed with a help of Kalman filter (EKF) and particle filter (PF) based training algorithm. We have to also considered the system having high nonlinear, the linearization in Kalman filter cause instability and particle filter might suffers from particle degeneracy. Therefore, poor optimum can find on training PF algorithm to solve these unscented Kalman filter (UKF) was found were deterministic sampling technique is used with no linearization in it to remove degeneracy algorithm. The computational complexity of UKF based algorithm is similar to EKF algorithm. These complexities can decrease further with use of minimum non UKF (MN-UKF). It generates a good trade-off among performance and complexity.

Olga Krestinskaya et al. in 2019 [13] proposed a neural network architecture for memristive learning using an analog backpropagation. They proposed that implementing one chip with a learning algorithm can speed up the training of the neural network having crossbar array. The entire circuit uses a back-propagation algorithm with a gradient decent operation for architecture of neuron. This architecture consists of three different neural network layers of DNN, BNN, MNN, along with a HTM SP and LSTM network. This paper consists of analog back-propagation technique for many memristive learning architectures as deep learning neural networks, binary and multi neural networks, hierarchical networks and LSTM networks. The circuit architecture and its verification are done by TSMC 180-nm CMOS model and TiO2-based memristor model and the entire validation of these application is done using XOR problem, MNIST dataset and a database of Yale face images is also taken. We can investigate on issue like signal integrity when memristor technology is suitable as a fabricating large-scale reliable array.

Farah Adeeba and Sarmad Hussian in 2019 [14] proposed a model for identifying the native language. These model works by identifying first language of user on the basis of speech or written text in second language. This paper has use of spectrogram and cochleagram for extracting the features from short speech utterances (average of 0.8) to infer urdu speaker native language. The entire set of experiment was performed, and accuracy was evaluated on the basis of validation dataset and observed that overall accuracy was 74.81% and using Mel-frequency cepstral coefficient (MFCC) accuracy was 71.61% and there was the use of Gammatone frequency cepstral coefficient (GFCC). Further, enhancement model was merged as MFCC with BLSTM and GFCC feature with BLSTM network to gain the advantage of both network and accuracy was calculated and observed that in short duration over 50% accuracy was obtained. In future, prosodic based feature can investigate with hierarchical classification to increase the performance of model.

Fazle Karim et al. in 2017 [15] proposed a fully connected neural (FCN) network to gain the state of art performance based on time series pattern. FCN was proposed with LSTM-RNN network sub module for classifying time series. Proposed module enhances the performance of FCN with slight increase in size of model and minimal processing in dataset. The proposed module of LSTM-FCN gains state-of-the-art performance while comparing with other modules present. Attention mechanism where also improved for
time series classification with attention fully connected LSTM network (ALSTM-FCN). These mechanism of attention helps in visualizing the LSTM cell decision process. Further, there are refinement method which increases performance of training modules. An overall, performance of the model is compared with existing techniques. Due to generalize input state to this model, it has large range of applicability in different modelling tasks as analysing the text, music and voice recognition and due to its small size and efficiency it can be deployed in real time scenario.

Jinlei Zhang et al. in 2019 [16] proposed a cluster-based LSTM module and innovative STPFF URT module for short-term forecasting passenger flow in urban rail transit. Urban rail transit (URT) is one of the important components for forecasting passenger flow. Therefore, it is essential to gain high prediction with enhancement of URT. Firstly, they proposed a K-Means clustering to obtain the passenger flow but also trends ridership volume characteristics. The assessment predictability model was developed to suggest a reasonable granularity time interval model known as CB-LSTM who’s proposed is to forecast short-term passenger flow. The experiment is conducted on subway station in form of cluster for prediction of short-term passenger prediction using limited dataset and result was critical insight as transport policymakers and subway operators.

III. CONCLUSION & FUTURE WORK

These paper demonstrated the different development phases of an LSTM neural network, its development complexity, fault tolerant capacity and the different models used for smooth traffic management, ariable collected data processing. These paper consist of review of LSTM network and its uses in different field within an enhancement of LSTM network with RNN its future work related to different field and modification along with the different optimization technique.

IV. REFERENCES


