International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7, Issue-6, March 2019

FORETELL: Forecasting Environmental Data Through Enhanced LSTM and L1 Regularization

Gayathiri Kathiresan, Krishna Mohanta, Khanaa Velumailu Asari

Abstract**: *Big data analytics progressively takes over the universe, and the prediction tends to take advantage of Big data analytics to yield the incredible results. The deep learning algorithms procure higher priority than the machine learning in the prediction systems. The traditional weather prediction exploits the observations of prevailing atmospheric conditions that accumulate simultaneously either from the trained observers or the numerical prediction model. However, the weather forecasting is an arduous task due to the dynamic and uncertainty of data. The proposed system plans to use the neural network for weather forecasting to overcome these shortcomings. Recently, the Long-Short Term Memory (LSTM) network based weather forecasting has gained popularity in machine learning. It significantly reflects the superior ability to model the temporal data sequences along with the long-term dependency through the memory blocks. However, on account of memory blocks along with the loop structure leads to over fitting issues. In order to tackle this issue, this work presents FORcasting Environmental data Through Enhanced LSTM and L1 regularization (FORETELL) that extends the existing LSTM model with the two methods such as optimal neuron selection method and the regularization method. The optimal neuron selection method constitutes the FORETELL system, as it can learn the complex data sequences without longer period and overfit of data. Instead of processing the entire vast feature of the data sequence, the regularization method captures the potential features to avert the overfitting constraint that ensures the noise-free system with more excellent performance. Conclusively, the FORETELL is evaluated using the weather dataset to demonstrate the superior performance than the existing Sequence to Sequence Weather Forecasting (SSWF) method.***

Index Terms**: *Big data, prediction, deep learning, Neural network, LSTM, overfitting.***

I. INTRODUCTION

Owing to the tremendous advancement in the technology and the environment, the volume of data gets explodes that leads to big data [1]. Analysis of big data substantially imparts the benefits including, obscure correlation, market trends, new opportunities, and the decision making. Notably, it yields the best prediction over the overwhelming data in a feasible way through the different analysis tool. For instance, the weather forecasting using the numerous disordered nature of atmospheric data from the satellite, sensor, and the other computing devices. It serves as a key role to unlock the valuable perceptions buried in the data regardless of source, size, and format. Despite, it still struggles for processing the voluminous data belonging to the legacy system in predictive analytics. Thus, the big data analysis adopts the batch processing to diminish the constraint above that handles the data. In essence, machine learning methods are the core principle behind predictive analytics to uncover the feasible and statistical pattern in the dynamic environment. It brings the greatest challenge of learning the large-scale data and the prediction of the unseen data without the complexity in batches rather than individual data over the time to acquire better accuracy [2].

The batch processing deploys various machine learning algorithms to mine the enormous amount of data for the beneficial prediction results. Even though, the machine learning techniques entail an additional feature selection and retrieval to render the features of data for predicting the outcomes. Consequently, the existing works based on the machine learning methods lack behind facilitating large features belongs to the data in prediction [3]. Deep learning is a part of the machine learning algorithm, and deep learning leverages the neural network to model the high-level of data abstractions. In contrast to the machine learning, the deep learning automatically hierarchically extracts the potential features with the support of the stacked hidden layers in the neural network. Nevertheless, the deep learning networks, exceptionally forecast the results by learning the numerous potential features of data automatically. In a deep learning algorithm, each layer of the system can automatically extract the distinct features of the data for the prediction [4].

Conventionally, the weather forecasting is done by accumulating quantitative data corresponding to the prevailing situation of the atmosphere. Also, it lapses to handle the stochastic nonlinear weather data without the prediction error. The forecasting of weather is a formidable challenge due to the uncertainty of the weather and the properties mentioned above [5]. Thus, the precise forecasting is essential, even the system subjects to the nonlinear and massive sequence data. Accordingly, the weather prediction system based on the neural networks performs well, rather than the machine learning techniques. Also, it entails the attention mechanism to capture the most relevant features that contribute to the better performance of the system. In order to recognize sequences based on time, the neural network exploits the Recurrent Neural Network (RNN) [6]. The RNN is a specific kind of neural network knowing the sequential pattern.

Revised Manuscript Received on 30 March 2019.
* Correspondence Author
Gayathiri Kathiresan, Research Scholar, Bharath Institute of Higher Education and Research, Chennai, India.
Krishna Mohanta, Associate Professor, Kakatiya Institute of Technology and science for woman, Nizamabad, India.
Khanaa Velumailu Asari, Dean-Info, Bharath Institute of Higher Education and Research, Chennai, India.

The RNN possesses the ability to handle the sequence of nonlinear characteristics of atmospheric data. Further, it assists in capturing the dynamics of weather for providing promising results in weather forecasting. However, the RNN based prediction systems miserably have the pitfalls of short-term memory, vanishing gradient, and the long-term dependency issue. Accordingly, it faces the troubles while retaining the data from the earlier process which escalates the training time of the data sequences. Furthermore, it yields the inaccurate results of weather prediction. Thus, there is a demand for the weather forecasting system to obtain the accurate results.

With the aim of predicting the accurate results from the complex data representation within the appropriate training time and resolving the constraints of RNN, the deep learning enables the Long Short-Term Memory (LSTM) to explore the numerous features through the hidden unit in the hidden layers [7]. In past decades, several researchers face the challenges in the determination of hidden unit in the neural network. The inappropriate neuron selection on the hidden layer while learning induces the overfitting issue in the network that reflects the unstable prediction system. In addition, it considerably corrupts the generalization ability of the system that gives rise to the deviation in the prediction process. With the intention of suppressing these issues without the corruption of prediction accuracy, the FORETELL develops the existing LSTM model based on the optimal hidden layer selection method to provide promising weather forecasting. Moreover, it takes advantage of a regularization technique to degrade the prediction error owing to the overfitting.

1.1. Contribution

The pivotal contributions of the FORcasting Environmental data Through Enhanced LSTM and L1 regularization (FORETELL) are as follows.

- The FORETELL presents the precise weather forecasting system by employing the LSTM network that ensures the best prediction over the legacy system.
- The FORETELL involves the optimal neuron selection in the hidden layer that in turn brings the ability to handle the complex data sequences without the long training time and the overfitting of data.
- The FORETELL exploits the regularization techniques to acquire, the better generalization that assists to discard the noises and the overfitting issues while learning the training data.
- The proposed approach reduces the complexity of the prediction system by ignoring the features with the minor variance in the output based on the L1 regularization technique that contributes the minimal prediction error and the highest accuracy.
- The results of the experimental framework reflect the superior performance of the proposed FORETELL approach than the existing SSWF approach.

1.2. Paper Organization

This paper is organized in the following way. Section 2 surveys the earlier works related to the prediction system. The proposed weather predicting is presented in Section 3. Section 4 presents the experimental framework and evaluation results of the FORETELL approach. Conclusively, Section 5 describes the conclusion of the proposed method.

II. RELATED WORK

For The deep learning technique for the rainfall prediction model [8] employs the auto encoder and multilayer perception that enables the feature selection and the prediction process respectively. However, this model fails to retrieve the deep features with respect to the diversified aspects to provide the best prediction. The framework of decomposition and ensemble technique, referred to as an Ensemble Empirical Mode Decomposition (EEMD) [9] assists to predict the rainfall for Colombia. It partitions the raw data into the disparate set of components and then exploits the Feed Forward Neural Network (FNN) on each partitioned component to predict the outcome of an individual component, and then each outcome is accumulated to provide the results. In order to ensure the accurate prediction of weather, the model exploits the fusion of techniques include Artificial Neural Network (ANN) and the Particle Swarm Optimization (PSO) that imparts the best predictions with the insignificant error rate [10]. The weather forecasting based on the sliding window technique [11, 12] guarantees the prediction accuracy. Moreover, the amalgamation of the sliding window with the ID3 algorithm reflects significant the relation between the historical data to obtain better predictions that helps to eliminate the skewed forecast of weather condition. The work [13] integrates the back propagation method and the genetic algorithm for accurate weather prediction which averts the scaling issues and the long training time of the neural network. The model [14] predicts the rainfall of Korea using the ANN that enhances the predictability by the exploration of delayed effects of the climate indices and the correlation between them. However, it becomes ineffective when predicting the lower rainfall environment. In order to tackle the diversified large-scale input data and the processing complexity, the work [15] attempts to fuse the dynamic data driven system with the auto encoder neural network that explores the correlation between the recorded data to ease the self-learning process. It enhances the prediction accuracy and the speed of the weather forecasting under a dynamic environment. However, most of the conventional work on the weather forecasting intensely suffers issue owing to the prediction of the large-scale nonlinear sequence of data under the various perspectives that escalates the processing complexity and also makes the defective in the forecast. In addition, it encounters the long-term dependency issue while predicting the data based on historical data on irrespective of the recent record. With the intention of dealing with the nonlinear sequence data, the several researchers employ the recurrent neural network, a specific kind of ANN with the feedback loop.

The approach [16] exploits the Elman Recurrent Neural Network (ERNN) method to predict the specific humidity of the system with a higher accuracy compared to the back propagation model. In order to forecast the trajectory of the hurricane excluding the truncation error, the work [17] exploits the RNN in a grid manner. The RNN can model the complicated temporal hurricane using the normalized data include longitude, latitude and the wind speed that improves the accuracy of the prediction. Despite, the earlier works based on RNN intensely suffer from the issue of vanishing gradients that make the network as the unstable and inefficient network. Therefore, in order to overcome these issues, several researchers employ the Long Short-Term Memory (LSTM) networks to forecast the weather. The Convolutional LSTM (ConvLSTM) network attempts to forecast the precipitation of the region within the short period of accumulating the several ConvLSTM layers. It constructs the training model to resolve the issue of spatiotemporal sequence correlation [18]. The short-time traffic prediction model [19] explores each implicit traffic information over the large data for forecasting the traffic accurately. Accordingly, it constructs the cascaded LSTM model with the multiple layers using the memory units to capture the evolution of traffic flow which assists in improving the prediction accuracy. The forecasting model for electric load prediction [20] implemented with the LSTM network and the genetic algorithm to facilitate the load scheduling and control the electricity production based on the optimal features. It resolves the constraints of overfitting and the unstable state of forecasting. The weather prediction model based on the time series [21] exploits the multiple layer LSTM network for mapping the weather sequence having the identical length. It performs the data preprocessing techniques to deal with the noise and the inconsistent data which averts the training issue in the network.

2.1. Problem statement

The weather forecasting is essential to recognize the weather conditions of the region over the period. The early weather forecasting works mainly focus on neural networks to estimate future weather conditions. It exploits the hidden layers to analyze the features of the system. The single hidden layer that is shallow layer does not contribute the efficient analysis of nonlinear characteristics of features. Thus, the RNN based weather forecasting extensively exploited to a prediction which retains the historical data to make an efficient prediction through the loop structure. The RNN based system faces the issue of vanishing gradient that escalates the period that requires training the model. In order to prevail over this shortcoming, the LSTM based forecasting system has developed. This model able to learn the long-term dependency through the memory units of hidden layers. It comprises a forget memory gate structure model to mitigate the long-term dependency that occurs owing to the ability to remember the historical data for longer periods, which become the source of the vanishing gradient issue. Also, it explores the multiple features rather than the one or more feature through the hidden layer which eases the extraction of sufficient knowledge over the overwhelming historical data. Thus, the conventional LSTM based forecasting system exploits more than one hidden layer to the prediction. They lack in focusing on the optimal hidden layer selection over the numerous layer when dealing with the nonlinear training data sequence. As a result, there is a requirement of attention mechanism to prevent the system from the overfitting issue. Thus, the pruning the hidden units with the least variance on the output layer is necessary. Furthermore, the residence of the chain structured gate network in the LSTM allows the process in the repetitive manner which induces the error and misleads the prediction. Hence, developing the LSTM model concerning the optimal hidden layer selection method along with the regularization method has to be focused in forthcoming to enhance prediction accuracy.

III. OVERVIEW OF FORETELL METHODOLOGY

With the intention of predicting the weather condition accurately, the FORETELL method exploits the framework of LSTM model and the regularization technique. The traditional works of the weather forecasting system utilize the machine learning algorithm to predict the weather. However, it refuses to utilize the entire features of large-scale data sequences for prediction which escalates the prediction error of the forecasting system. To cope up with the above-mentioned constraints, the predictive analytics based on the LSTM neural network is implemented that explores the implicit features that belong to the nonlinearity data for prediction. Even though, it lacks to a tradeoff between the complexity and the accuracy while predicting the weather due to the disregard of the optimal selection of neurons in the hidden layer. Therefore, the proposed approach extensively develops the existing LSTM model with the optimal neuron and hidden layer selection methods. These selection methods have a greater influence on the output and the performance of the prediction system. A large amount of neurons in the layer greatly intensifies the training time of the network. Moreover, the hidden layer selection mainly relies on the complexity level of the data. Thus, the precise choice of a number of the hidden neuron in the hidden layer is essential while modeling the LSTM network. Furthermore, the FORETELL applies the feature scaling technique such as a regularization method that supports the smoother system by actively reducing the overfitting issue and the error. It incorporates the two main phases such as LSTM modeling phase and the mining phase. Figure.1 illustrates the working nature of the FORETELL.

3.1 LSTM modeling phase

The weather forecasting system intends to exploit the early captured data sequence to predict the future weather in a particular region within a short period. Accordingly, the FORETELL utilizes the most promising neural network named as LSTM that makes the in-depth knowledge about the sequence of data and their context than the other neural networks which eases the precise predictive results in weather forecasting. In the neural network, the dimension of the input specifies the complexity of the system while learning the data sequence.nAccordingly, the neural network critically suffers from the redundant and extraneous data sequences when it fed with the abundant data sequences.

In the neural network, the appropriate selection of a number of neurons and the hidden layer is essential to estimate the underlying patterns of the incoming data sequences. Conventionally, most of the neural network randomly picks the neurons in the hidden layer

which originates the overfitting issues in the prediction system. Also, the escalating number of layers in the neural network induces the progressive error rate of test data. In this sense, deciding the number of neurons with the optimal hidden layer flexible to obstruct the trouble of overfitting and the complexity. The FORETELL involves modeling of existing LSTM with the hidden neuron selection method to accelerate the learning process and resolves the issues for efficient knowledge retrieval. The proposed FORETELL approach considers the dimension of the hidden layer as two in a neural network to tackle the complexity of nonlinear data. The equation (1) computes the number of hidden units for the each layer (D) to ensure the reliable operation of the proposed extension LSTM model, whereas Ns represents the total number of data instances, and the term Avg (dev) denotes the average value of the standard deviation function that computed based on the equation(2).

$$N_H = \log(Avg(dev)) \times \left(\frac{\log N_s}{N_s} \Big/ D \right) \qquad (1)$$

In order to select the optimal number of neurons, the FORETELL computes the deviation over the data instances. It adopts the number of hidden units based on the deviation of data sequences. For example, if the deviation is high, then the FORETELL decides to allocate the additional neuron to carry out the computing process.

$$deviation = \sqrt{\frac{\sum X - \mu^2}{N}} \qquad (2)$$

In equation (2), x represents the data set value, μ denotes the mean value of the data set, and N indicates the total number of data points. The average of standard deviation judges the neuron selection process while implementing the LSTM network. Figure 2 briefly explains the neuron structure and their functions in the neural network where the disparate features of the weather data feed as input data (X_i), i= 1,2,..., n to the hidden layer. Each neuron within the hidden layer learns a various set of weights to express the diversified functions of input data, whereas the weight ranges from W_1 to W_n. Further, it captures the sum of weights and transfers it over the bias and the activation function (f). This process is distributed over the other neurons in the subsequent hidden layer, and the output layer provides the predicted results.

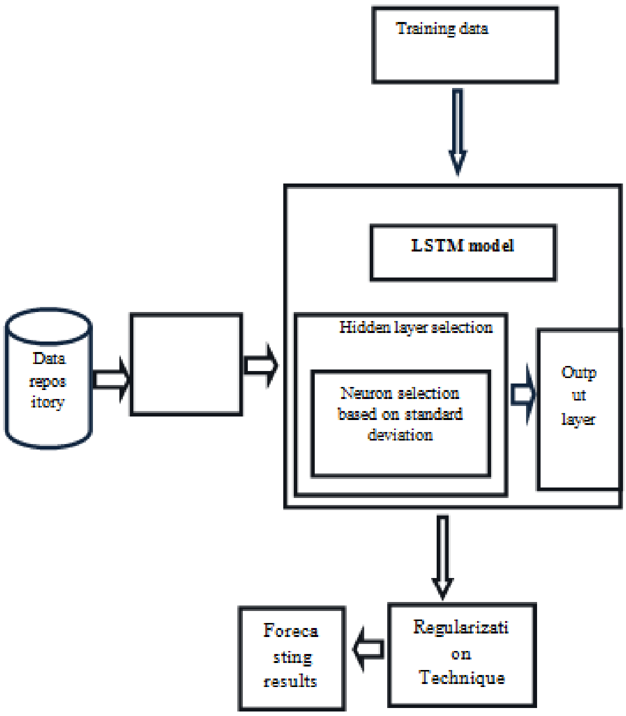

Fig.1 The FORETELL methodology

3.2 Mining Phase

The selection of optimal hidden layers and neurons is vital in big data mining. Also, generalization is essential while using the improved LSTM model to acquire the precise prediction for the forthcoming unseen record of data. At the same time, the lack of attention of learning process generalizes data as an arduous task which results in overfitting issue. Fortunately, there is a technique to decline the overfitting while modeling the LSTM network referred to as regularization that effectively constrains the learning process.

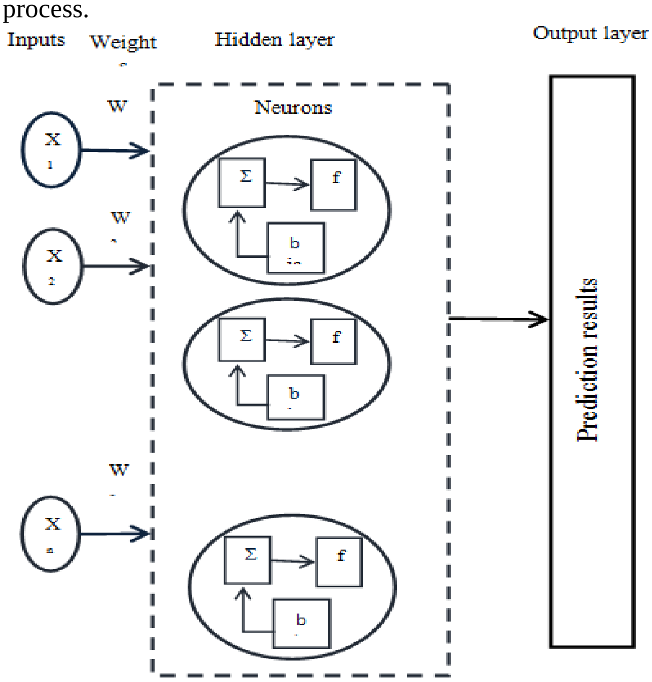

Fig.2 The function of a neuron in the neural network

Notably, among the diverse techniques of regularization, the FORETELL exploits the L1 regularization method that consonant with the more complex, noisy and unrelated features of data rather than other methods of regularization. Furthermore, the penalty term in the L1 regularization attempts to capture the potential features of the data and provides the sparse solution by considering the irrelevant features as the zeros that assist in restricting the size of the network.

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i + \lambda \sum_{w\epsilon W} |w| \qquad (3)$$

In the equation (3), a penalty parameter λ get added with the loss function (C_i), the dimension of the parameter is correlated with the dimension of weights (w) in a network. The term 'n' represents the features in the data sequence. The penalty parameter value decided by the designer whereas λ = 0 indicates there is no penalty. Notably, the lesser essential features have the no weight penalty parameter. The algorithm 1 briefly describes the entire process of the FORETELL.

Input: weather data
Output: prediction results
for each weather data **do**
{
Layer selection()
Neuron selection()
Neuron function determination()
Output layer()
}
Layer selection()
{
Data size measurement;
Compute the number of hidden layer,
If (Data==less complex)
fix hidden layer=1
else
fix hidden layer=2
endif
}
Neuron selection()
{
Compute the deviation using the equation (2);
Compute the average of deviation;
Compute number of hidden units using the equation (1);
}
Neuron function determination()
{
Retrieve the features;
Compute the activation function (f);
}
Output layer()
{
Prediction results
ifoverfitting occurs
Apply regularization method using the equation (4)
accurate predictive results
endif
}
endfor
}

Algorithm.1 The procedure of proposed FORETELL approach

IV. IMPLEMENTATION OF FORETELL USING THE HADOOP

In order to ensure the processing of overwhelming data without the defects and the time complexity in a distributed environment, the proposed FORETELL model makes use of an Apache product referred to as Hadoop. Hadoop comprises of MapReduce framework that facilitates the reliable processing of large-scale data in the distributed environment. In the Hadoop, the incoming large-scale data is partitioned into data splits and then runs it by parallel manner while processing the massive data in the map phase. Accordingly, the map phase parallelly computes the standard deviation between the incoming data sequence. Subsequently, the reduce phase of the Hadoop exploits the outcome of the map phase as input data. The reduce phase receives the standard deviation as the input. Afterward, it consolidates the deviation results to compute the highest deviation result that brings the final decision to select the number of neurons in the layer of the neural network. The MapReduce saves these results in default using the Hadoop Distributed File System (HDFS). The results of the MapReduce phase one is fed as input to the phase 2 for predicting the weather, whereas, the regularization process carried out in parallel reduce tasks to procure accurate results. The figure.3 shows entire the procedure involves when implementing the proposed FORETELL using Hadoop in a detailed manner.

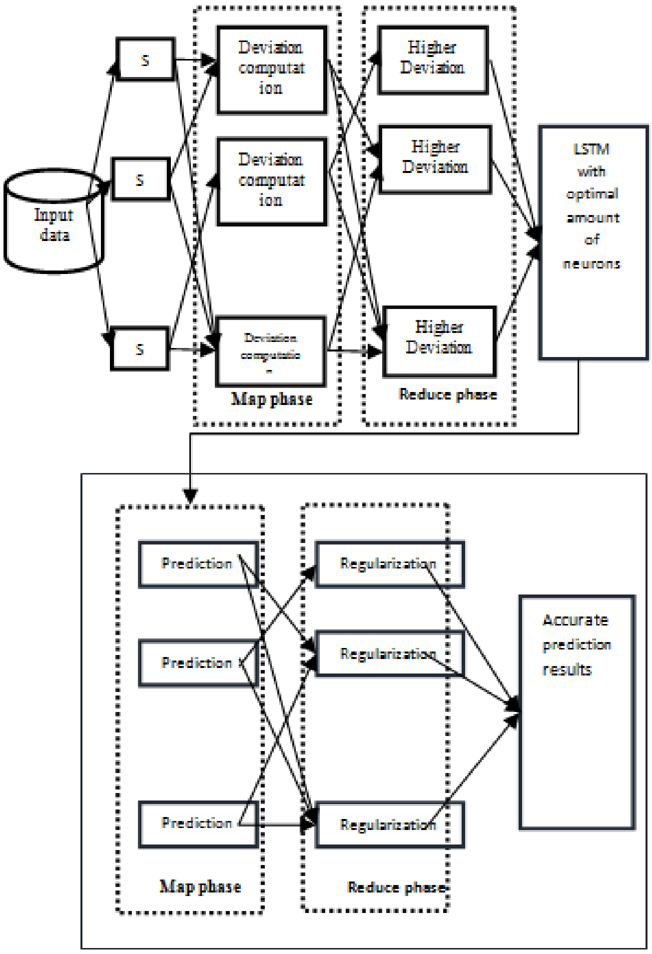

Fig.3 The FORETELL method implementation

V. EXPERIMENTAL EVALUATION

This paper contrasts the proposed FORETELL methods with the existing SSWF to demonstrate the outstanding performance of the proposed model than the SSWF model [21].

5.1 Experimental setup

The experimental work implements the experiments using the Linux Ubuntu 12.04 LTS 64-bit machine with a 2.9 GHz Intel CPU and 32 GB memory. It employs the Apache Hadoop as a processing engine, an open source data processing, distributed environment version 1.2.1 to implement the proposed algorithm. It incorporates the MapReduce framework to support the parallel processing of large-scale data. It exploits the LSTM model from the deeplearning4j library to predict the weather condition. To evaluate the FORETELL approach, the evaluation model employs the weather dataset [22]. It comprises of five years of hourly measurements of data based on the different weather features, includes air pressure, temperature, and humidity. This dataset accumulated from the 30 number of US and Canadian Cities, also 6 Israeli cities. Among these data, the proposed method takes the data set of San Francisco to make an accurate prediction of weather.

5.1.1 Evaluation metrics

Recall: It is defined as the percentage of the number of correct predictions to the number of predictions that should have been returned.

Mean Absolute Error (MAE): It measures the deviation of the prediction (x_i-x) from the actual ones. The smallest error rate ensures the accurate prediction of the model. In this equation, the term 'n' denotes the total number of data.

$$MAE = \frac{\sum |x_i - x|}{n}$$

Matthews Correlation Coefficient (MCC): It measures the quality of prediction of the model. It measures the number of true positives (TP), the number of true negatives (TN), the number of false positives (FP), and (FN) the number of false negatives to evaluate the prediction model quality.

$$MCC = \frac{\left((TP \times TN) - (FP \times FN)\right)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

F-measure: It is a measure of predictive accuracy of the model and equalizes between the precision and recall.

$$F - measure = 2 \times \frac{Precision \times recall}{Precision + recall}$$

5.2 Evaluation results

5.2.1 Number of instances Vs. Recall

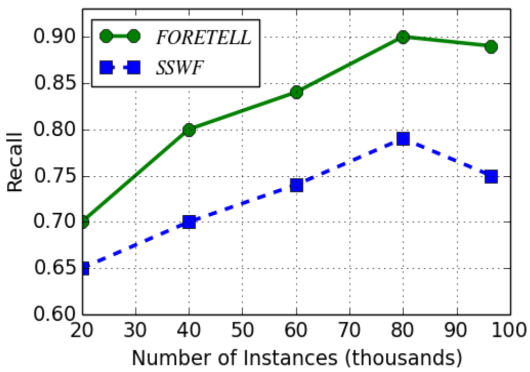

Fig.4 Number of instances Vs. Recall

Figure.4 represents the recall of both the proposed FORTELL approach and the existing SSWF approach based on the number of data instances. As a whole, the recall rate grows beyond a precise number of instances. When the system reaches a specific amount of instances of data, then the recall rate begins to degrade. The FORTELL model procures the most significant level of recall rather than the existing SSWF approach which shows the FORETELL brings the best prediction results than existing ones. Moreover, the FORETELL model earns the remarkable recall rate when increasing the number of data instances. Correspondingly, it attains 10% of greater recall rate than the existing SSWF approach when the amount of data instances is 40. It is because of, the FORETELL incorporates the optimal learning of data instances through the optimal neuron selection and the regularization technique that contribute forecasting accuracy of the FORETELL. Furthermore, even when increasing the total number of data instances beyond the value of 80, the existing approach suddenly drops the recall rate as 14% than the proposed method which is due to the lack of optimal selection methods in the SSWF method that leads the decrement of recall rate.

5.2.2 Number of instances Vs. MAE

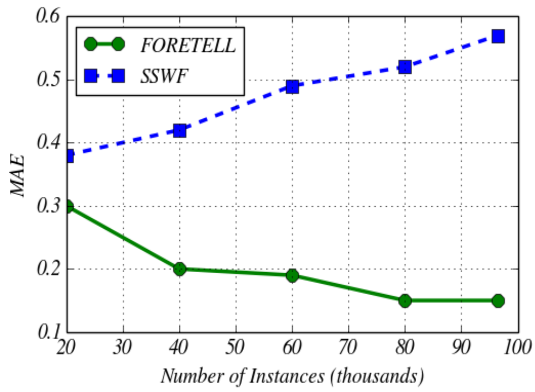

Fig.5 Number of instances Vs. MAE

The MAE error rate of the FORETELL approach and the SSWF approach is illustrated in figure5 while varying the number of instances from 20 to 100 (thousands). The minimal MAE value reflects the superior performance of the LSTM prediction model. With the increase of data instance size, the proposed method effectively decreases the error rate than the existing method. Moreover, When increasing the dimension of data instances from 40 to 80, the proposed method progressively decreases the error rate to predict the accurate condition of the weather, and then it reaches the stable state even increment the data instance beyond the value of 80. On the other hand, the existing SSWF method escalates the error rate linearly with the number of data instance. It is because of; the FORTELL exploits the regularization method in the prediction model which added the advantage of feature selection to avert the overfitting issue. The potential feature determination helps to the pruning of neurons and to reduce the complexity of the prediction system. As a result, the proposed FORETELL approach superior in performance than the SSWF approach.

5.2.3 Number of features Vs. MCC

Figure6 demonstrate the MCC value of the proposed FORTELL approach and the existing SSWF approach with the variation of the number of feature values from 3 to 7.The number of hidden units and the layer increases with the increasing the number of features in the data.

When the amount of the number of the feature is 3, the proposed approach acquires the higher value of MCC as 80%, but, the existing approach attains only 50% of MCC value. Moreover, the existing approach dramatically degrades the MCC rate when increasing the number of features due to the lack of attention of the hidden unit and the hidden layer. The escalation of feature set entails further neuron in the layer to process the data sequence, but the regardless of the selection of neuron in the layer degrades the accuracy. Accordingly, the existing approach drops the MCC rate of 52% than the proposed approach. The proposed FORETELL method moderately maintains the MCC rate even with the increase in the number of features in the data sequence owing to the advantages of optimal hidden unit selection. Furthermore, in order to ensure the accuracy, the proposed method tends to neglect the data with a smaller variance to the output layer.

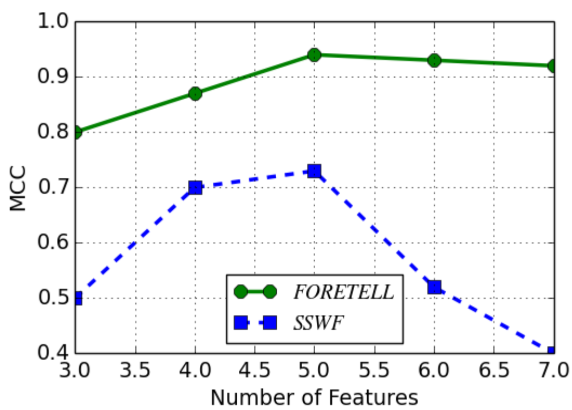

Fig.6 Number of features Vs. MCC

5.2.4 Number of neurons Vs. F-measure

The figure7 shows that the F-measure value of the proposed approach under the layer selection criteria while varying the number of neurons from the 50 to 130. The value of F-measure is best for the one layer network until it has the acceptable level of neurons that means the low-level computational complexity of the system. In case of the complexity level of the network escalates beyond the acceptable limit, the one layer network rapidly diminishes the F-measure value. At the same time, the network with a couple of layers intends to keep the f-measure value within the certain limit to enhance the accuracy even when increasing the number of neurons and the complexity of the system.

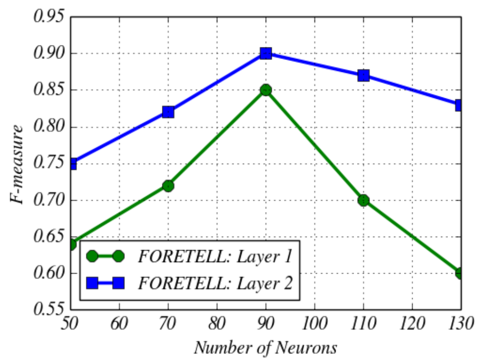

Fig.7 Number of neurons Vs. F-measure

Accordingly, With the increase of the number of neurons from 50 to 90, both of them attempts to increase the f-measure value, but, when reaching the total number of neurons value beyond the 90, the one layer network degrades the f-measure value as 0.17 compared to the two-layer network. Additionally, at this specific point, the single layer network experiences the confusion state while learning the training data owing to the inadequate number of layers to process the complex representation of a data sequence. On the other hand, the network with the 2 hidden layers effectively deals the complex representation of data without the difficulty.

VI. CONCLUSION

This work proposed the FORETELL approach, which is the enhanced LSTM neural network model for predicting the vast explosion of data. The primary concern of the proposed approach is to provide accurate weather predictions. With this intention, the FORTELL method exploits the two methods include optimal neuron determination method and the regularization method. Initially, the FORETELL attempts to select the optimal hidden unit in the layer based on the average deviation between the data sequences and the complexity of the system. It ensures the prediction model without the overfit of data and time complexity. Also, it exploits the L1 regularization method to capture the optimal feature over the legacy system to reduce the prediction error and the overfitting issue. The experimental results demonstrate that the proposed FORETELL approach acquires the superior performance rather than the existing SSWF approach in terms of accuracy and the error rate. Also, the FORETELL approach improves the recall rate of 14% than the existing SSWF method when reaching the maximum number of instances in the system.

REFERENCES

1. Khan, Nawsher, IbrarYaqoob, Ibrahim AbakerTargioHashem, ZakiraInayat, Mahmoud Ali, WaleedKamaleldin, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani, “Big data: survey, technologies, opportunities, and challenges,”The Scientific World Journal, 2014.
2. Zhang, Yunquan, Ting Cao, Shigang Li, XinhuiTian, Liang Yuan, HaipengJia, and Athanasios V. Vasilakos, “Parallel processing systems for big data: a survey,”Proceedings of the IEEE, Vol.104, No. 11, pp.2114-2136, 2016.
3. Zhou, Lina, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos,“Machine learning on big data: Opportunities and challenges,”Neurocomputing,Vol.237, pp.350-361, 2017.
4. Najafabadi, Maryam M., FlavioVillanustre, Taghi M. Khoshgoftaar, NaeemSeliya, Randall Wald, and EdinMuharemagic, “Deep learning applications and challenges in big data analytics,”Journal of Big Data,Vol.2, No.1, pp.1, 2015.
5. Jain, Himanshi, and Raksha Jain, “Big data in weather forecasting: Applications and challenges,”In IEEE International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), pp.138-142, 2017.
6. Hermans, Michiel, and Benjamin Schrauwen, “Training and analysing deep recurrent neural networks,”In Advances in neural information processing systems,pp.190-198, 2013.
7. Greff, Klaus, Rupesh K. Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber,“LSTM: A search space odyssey,”IEEE transactions on neural networks and learning systems,Vol.28, No.10, pp.2222-2232, 2017.
8. Hernández, Emilcy, Victor Sanchez-Anguix, Vicente Julian, Javier Palanca, and NéstorDuque,“Rainfall prediction: A deep learning approach,” In SpringerInternational Conference on Hybrid Artificial Intelligence Systems,pp.151-162, 2016.
9. Beltrán-Castro, Juan, Juliana Valencia-Aguirre, Mauricio Orozco-Alzate, GermánCastellanos-Domínguez, and Carlos M. Travieso-González, “Rainfall forecasting based on ensemble empirical mode decomposition and neural networks,”In Springer International Work-Conference on Artificial Neural Networks,pp.471-480, 2013.
10. Tripathy, Asis Kumar, SuvenduMohapatra, ShradhanandaBeura, and GunanidhiPradhan, “Weather Forecasting using ANN and PSO,”International J. Scientific & Eng. Res, Vol.2, pp. 1-5, 2011.
11. Kapoor, Piyush, and Sarabjeet Singh Bedi, “Weather forecasting using sliding window algorithm,” ISRN Signal Processing, 2013.
12. Kumar, Jayendra, Srinath Mishra, AnuragHansdah, and Rajiv Kumar, “Modified sliding window alogrithm for weather forecasting,”In IEEE proceddings on Technological Innovations in ICT for Agriculture and Rural Development (TIAR), pp.175-180, 2016.
13. Gill, ErJasmeen, ErBaljeet Singh, and ErShaminder Singh, “Training back propagation neural networks with genetic algorithm for weather forecasting,”8th IEEE International Symposium on Intelligent Systems and Informatics (SISY), pp.465-469, 2010.
14. Lee, Jeongwoo, Chul-Gyum Kim, Jeong Lee, Nam Kim, and Hyeonjun Kim, “Application of artificial neural networks to rainfall forecasting in the Geum River basin, Korea,”Water, Vol.10, No.10, pp.1448, 2018.
15. Lin, Szu-Yin, Chi-Chun Chiang, Jung-Bin Li, Zih-Siang Hung, and Kuo-Ming Chao, “Dynamic fine-tuning stacked auto-encoder neural network for weather forecast,”Future Generation Computer Systems,Vol.89, pp.446-454, 2018.
16. Fang, Chen, Xipeng Wang, Yi L. Murphey, David Weber, and Perry MacNeille, “Specific humidity forecasting using recurrent neural network,”In IEEE International Joint Conference on Neural Networks (IJCNN),pp.955-960, 2014.
17. Xingjian, S. H. I., Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,”In Advances in neural information processing systems, pp.802-810, 2015.
18. Xingjian, S. H. I., Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,”In Advances in neural information processing systems, pp.802-810, 2015.
19. Zhao, Zheng, Weihai Chen, Xingming Wu, Peter CY Chen, and Jingmeng Liu, “LSTM network: a deep learning approach for short-term traffic forecast,”IET Intelligent Transport Systems, Vol.11, No.2, pp.68-75, 2017.
20. Bouktif, Salah, Ali Fiaz, Ali Ouni, and Mohamed Serhani, “Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches,”Energies, Vol.11, No.7, pp.1636, 2018.
21. Zaytar, Mohamed Akram, and Chaker El Amrani, “Sequence to sequence weather forecasting with long short term memory recurrent neural networks,”Int J ComputAppl, Vol.143, No.11, 2016.
22. https://www.kaggle.com/selfishgene/historical-hourly-weather-data