International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Air Quality Prediction using Recurrent Neural Network

Shweta K. Borse and D. V. Patil

Department of Computer Engineering R.H.Sapat college of Engineering, Management studied and research, Nashik savitribai phule pune university, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Air pollution is a serious problem. Pollution affects on human health and the atmosphere, it effects on health with diseases like cancer, asthma, heart disease and so on. An airborne pollutant can be described based on the absorption of elements available in the atmosphere. If the strength of a biochemical is larger than the goal level of elements in air, it is called as an air impurity. Airborne pollution arises when risky or extreme sizes of elements with smokes (such as CO, CO2, SO2, NOx, Ch4, PM) and Organic particles are familiarized into Earth's air. Airborne pollution levels in utmost of the town areas has been a substance of thoughtful apprehension. In estimating of pollution, the soft computing methods are used. The air quality is predicted with machine learning algorithms that in-turn forecasts the AQI. AQI is a measure used to show the impurity levels over a time period. We have implemented the model to predict the AQI on previous year's data of impurity. In this paper, we have proposed to use StackLSTM and as per our knowledge it performs better as compared to available techniques. We have used machine learning techniques such as a recurrent neural network (RNN), Long short-term memory (LSTM) i.e. SimpleLSTM and StackLSTM for experimentation. It is observed that StackLSTM performs better as compared to SimpleLSTM and simpleRNN.

Keywords: *Air quality index, Simple Long short-term memory, StackLSTM, Simple Recurrent neural network.*

Introduction

An air quality index (AQI) is used by the administration to communicate to the community about the harmfulness of the air. As the AQI grows, a huge percentage of the people is likely to effects on health. Air quality index (AQI) is a numerical scale used for taking record of pollutant air increases day by day and also hourly, air value with regard to human health and the atmosphere. On an hourly base, the absorption of each impurity in the air is estimate and changed into a number starts from zero to upwards by using an ordinary index or scale. Considered number for each impurity is termed as a sub-index. The highest sub-index for any assumed period is noted as the AQI for that time. AQI is like a measure that series from 0-500. The index is a virtual scale, import, the index air quality and health issue, and so on. The absorption of each impurity varies, so, AQI standards values are sort into ranges assigned to a uniform public health warning and color code. Tsai Y. et al. [11] presented a prediction of PM2.5 concertation with recurrent neural network using Long ShortTerm Memory. They used Keras package, Keras contain highlevel neural systems are written in python, for this package need the TensorFlow support. For developing a neural network, and run a recurrent neural network with Long Short-Term Memory via TensorFlow. Freeman B. et al. [12] proposed RNN is used for air concentration forecast of input data. Air value measured the time series datasets contain important data. V.A. et al [13] presented three models as recurrent neural network, Long Short-Term Memory, gated recurrent unit. These models contain input-layer with hidden layer and then the output-layers. Data split into training data and testing data for taking accuracy.

Review of Literature

Choubin B. et al. [1] proposed air pollution and other particle are harmful to human health. Particulate matter (PM), and other gases are effect on human humanity and disease, biological, and organic system. According to the author the aforementioned actual appropriate the forecasting air quality. Some of the applications of machine learning models are designed for forecasting air quality constraint. Author used classification and regression tree and random forest. By using this model, they get the same accuracy from both models.

Fan J. et al. [2] presented three models as support vector machine with bat algorithm, particle swarm optimization, whale optimization algorithm for forecast of diffuse (Rd) to global (RS) solar radiation. These three models compare with multivariate adaptive regression spline and extreme gradient

boosting (XGBoost). In the experiment, they get SVM gives better accuracy as compare to XGBoot and multivariate adaptive regression spline (MARS). SVM-BAT model is improving the performance of Machine Learning (ML) models.

Y. C. Lin et al. [3] proposed the prediction of air quality system built on the neuro-fuzzy network method. Use the historical dataset to develop fuzzy rules. They divide the dataset into training data. Training data divided into fuzzy cluster according to their characteristics. They get fuzzy rules are of high quality and their constraints can be improved exceptionally.

Ding Y. et al. [4] presented air pollution is increases dayto-day, it is a serious issue, for air pollution control its need to predict air quality. There are many prediction methods are projected in the previous year to improve the accuracy of forecasting. In this paper, they proposed spatial-temporal Long Short-Term Memory for forecast the air quality. Model is a design built on the LSTM model.

Zhang, D., and Woo, S. S. [5] proposed determine the air quality has pointed due to growing ecological and health uncertainties in South Korea. In specific, micro dust is known to reason serious fitness problems to the public. So, determining and forecasting micro-dust is a serious problem. A typical way of calculating micro-dust is to use sensor devices from fixed spot. Forecast the air quality using sensors through three different cars and they build the application using machine learning techniques that report air quality to end-user.

Tian Y. et al. [6] presented to decrease air pollution effect by aircraft operations around the landing field of the airport. Use the classification, random forest method tracked by supervised learning algorithm. Random forest gives good accuracy results for hourly and daily datasets. In airport air quality problem solved by the machine learning method.

R. J. Kuo et al. [7] proposed air pollution increased is aserious issue not only in indoor or other places but also in the stock market, speech recognition, supply chain management, traffic problem and so on. Deep learning is also usually machine learning, its main difference is the neural network layer. Deep learning is used RNN. RNN is extracting the layers in sequence.

S. Shanthi and M. Pyingkodi [8] proposed air impurityelements are harmful to human health they are particulate matter, sulfur dioxide, nitrogen oxide. They apply the machine learning algorithm such as support vector machine (SVM), random forest (RF). They take the current year dataset from the Tamil Nadu Control Board installed air pollution monitoring station from five areas. AQI is determined on PM2.5, PM10, NO2, SO2 are display.

Dixian Zhu et al. [9] presented a machine learning approach used for air quality prediction. Forecast the hourly concentration of air impurity. By using machine learning techniques, they train models on large data with large-scale optimization algorithm. They show advanced techniques to improve the performance.

Kok I. et al. [10] proposed the concept of IoT, IoT is become a study subject in various zones such as education, industry, and commerce. They present a deep learning model to air pollution problematic issue in smart cities. Organized the network with the good hyper constraint giving to the result getting from the experiment. They used machine learning techniques such as support vector regression and LSTM model. LSTM model gives well performance than SVR model.

System Architecture / System Overview

A. Problem Statement

Air pollutant particles are come out from atmospheres different sources, the concentration of impurities in air quality depends not only on the quantities that are produced but also the capacity of the atmosphere, and also absorb or diffuse these impurities.

B. System Architecture

Figure presents steps for predicting the air quality using RNN. Information taken from different stations. 5-fold crossvalidation method is used for validation in the forecasting of the AQI level. Recurrent neural network is used for train model to forecast the AQI level using air pollution data. The recurrent neural network tries to evaluate the value of each particle. From the dataset, one hidden layer with neuron cell (GRU or LSTM) and one as output.

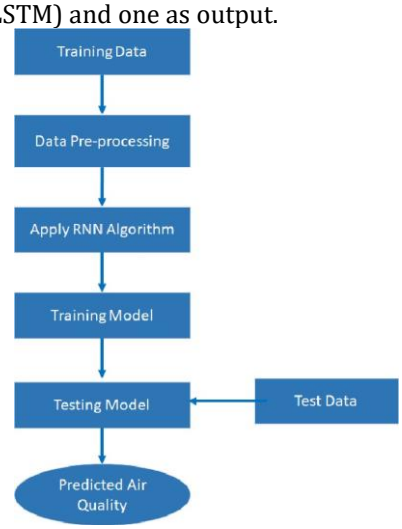

Fig. 1. Block Diagram of air quality prediction Model

C. Algorithms Used :

we have used machine learning techniques such as a recurrent neural network (RNN), Long short-term memory (LSTM) i.e. SimpleLSTM and StackLSTM for prediction.

Here we have elaborated the SimpleRNN, simpleLSTM and StackLSTM.

(a). Simple Recurrent Neural Network (RNN):

The RNN method categorize the vectors step by step. It passes the prior hidden state to the subsequent phase of the classification. The hidden state performance like the neural system. Its storage data on earlier information the system has got in the past. Old neural systems do not have any memory. So, they do not

receive into justification in earlier contribution when processing the present input. In consecutive datasets, as time-series, the data of prior period phase is classically applicable for forecasting approximately in the present step. Therefore, a state approximately the prior period phases should be maintained. The mid-air pollution at time-period t capacity be manipulated by airborne pollution in prior time stages. RNNs have an interior loop, that keep a state of prior time stages. This state is used for the forecasting in the existing time stage. The state is rearranged, a new system is existence processed. In this experimentation, we have used Simple-RNN of the Keras package. We denote an Early-Stopping call-back to break the training, there are 15 epochs without slightly development in the 'validation-loss'. The Model-Checkpoint permits us to store the weights of the greatest model.

(b). Simple LSTM:

LSTM design maintains a fixed value of the fault as it is posterior, reproduced through the structure. The slope received is less or higher as it changes posterior over the structure, it remains the equivalent. It remembers the short-term fault, across an extensive period (many regressive time stages). So, we are working to move stage by stage via the procedure of generating an RNN. We use the keras library to generate deep learning in python.

(c). Stack LSTM:

We have used Stack LSTM layers on topmost, it makes the structure of the model accomplished of higher-level knowledge from historical representations. The first two LSTMs provide their complete output systems, but the third returns the last stage in its output system, so reducing the sequential measurement. thus it is Stacking multiple LSTM layers. The model will absorb other concepts of input statistics over time period. Representing the input information not at the same time scale.

System analysis

A. Mathematical model

The BE-1-2013-2015-aggregated-timeseries dataset is used for experimentation. As mentioned earlier we have used RNN, Simple LSTM and StackLSTM for the experimentation. The algorithm used is as follows.

1) Input: Air quality data
2) FunctionS
3) Prepossess(),
4) SimpleRNN(),
5) SimpleLSTM(),
6) StackLSTM()
7) TestingPhase()

Mathematical model for RNN

1) A single time period stage of the input is supplied to the network system.
2) Estimates its current state using a set of present input state and previous state.

$$hcs = tanh(Wrn * hcs - 1 + Win * Is) \quad (1)$$

Where, h_{cs} : Current State, h_{cs-1} : Previous State, I_s : Input State

3) The current state (h_{cs}) has converted to the previous state (h_{cs-1}) for the next time step.
4) Single can go as various time stages according toward the problem and joint the data from all the previous states.
5) When all the time stages are complete then final current state is used to estimate the output.

$$Ot = Wol * hcs \quad (2)$$

where,

O_t: Output

W_{ol}: weight at output layer

6) The production is compared to the real output that is the target output and the mistake is produced.
7) The mistakes are then back spread to the system to update the weight.

Result and Discussion

We have BE-1-2013-2015-aggregated-timeseries dataset from European environment agencies site. For experimentation, we have used 5-fold cross-validation method. The data set is given as training data to the RNN "validation-loss" is observed on test data. The experiment was repeated for SimpleLSTM and stackLSTM. The results are specified in table 1.

Table I Validation Loss On Different Algorithm

Folds	SimpleRNN Validation-Loss	SimpleLSTM Validation-Loss	StackLSTM Validation-Loss
1	5.79	6.345	2.467
2	6.89	7.98	2.34
3	4.89	5.89	1.67
4	0.91	0.6292	0.3221
5	3.66	5.019	0.3781
AVg.	4.428	5.1726	1.4354

Table 1 presents the experimentation done for the prediction of AQI. Table 1 presents the validation loss obtained on 5 fold of BE-1-2013-2015-aggregation-timeseries.csv dataset. It can be observed that StackLSTM performs better as compared to Simple RNN and SimpleLSTM. Figure 2 presents the results in graphical form. Here we have proposed to use StackLSTM and as per our knowledge it performs better as compared to available techniques.

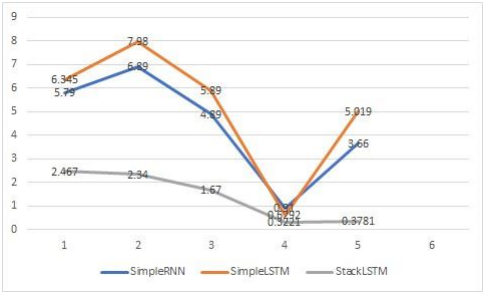

Fig. 2. Validation Loss using different method

Conclusion

Air pollution is one of the serious problems. Human health and the atmosphere is affected by the pollution. Here we have proposed to use StackLSTM and as per our knowledge it performs better as compared to available techniques. We have used RNN, SimpleLSTM, and StackLSTM to predict the air quality. As per experimentation, we conclude that StackLSTM performs better as compared to SimpleLSTM and simple RNN

Acknowledgment

I have a tremendous pleasure in presenting the project "Air Quality Prediction Using Recurrent NeuralNetwork" under the guidance of Dr. D. V. Patil and PG coordinator Prof. A. S. Vaidya. I am obligated and appreciative to Head of Department Dr. D.V. Patil for their significant direction and consolation. I might likewise want to thank the Gokhale Education Society's R. H. Sapat College of Engineering, Management Studies Research, and Nashik-5 for giving the required offices. Web get to and vital books. At last, I must express my sincere heartfelt gratitude to all the Teaching Non-teaching Staff members of Computer Department of GESRHSCOE who helped me for their important time, support, remarks, and thoughts.

References

[1]. Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., & Ghamisi, P. (2019). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Science of The Total Environment, 134474. doi:10.1016/j.scitotenv.2019.134474.

[2]. an, J., Wu, L., Ma, X., Zhou, H., & Zhang, F. (2019). Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. Renewable Energy. doi:10.1016/j.renene.2019.07.104.

[3]. Y. C. Lin, S. J. Lee, C. S. Ouyang, Chih-Hung Wu, Air quality prediction by neuro-fuzzy modeling approach, Applied Soft Computing Journal (2019) 105898, https://doi.org/10.1016/j.asoc.2019.105898.

[4]. Ding, Y., Li, Z., Zhang, C., & Ma, J. (2019). Prediction of Ambient PM2.5 Concentrations Using a Correlation Filtered SpatialTemporal Long Short-Term Memory Model. Applied Sciences, 10(1), 14. doi:10.3390/app10010014.

[5]. Zhang, D., and Woo, S. S. (2019). Predicting Air Quality using Moving Sensors. Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '19. doi:10.1145/3307334.3328647.

[6]. Tian, Y., Huang, W., Ye, B., & Yang, M. (2019). A New Air Quality Prediction Framework for Airports Developed with a Hybrid Supervised Learning Method. Discrete Dynamics in Nature and Society, 2019, 1–13. doi:10.1155/2019/1562537.

[7]. Kuo, R. J., Prasetyo, B., & Wibowo, B. S. (2019). Deep LearningBased Approach for Air Quality Forecasting by Using Recurrent Neural Network with Gaussian Process in Taiwan. 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA). doi:10.1109/iea.2019.8715113.

[8]. S. Shanthi and M. Pyingkodi, Air Quality Index Prediction using Machine Learning Algorithms, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019, DOI:10.35940/ijrte.D5326.118419.

[9]. Zhu, D., Cai, C., Yang, T., & Zhou, X. (2018). A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization. Big Data and Cognitive Computing, 2(1), 5. doi:10.3390/bdcc2010005.

[10]. Kok, I., Simsek, M. U., & Ozdemir, S. (2017). A deep learning model for air quality prediction in smart cities. 2017 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata.2017.8258144.

[11]. Tsai, Y.-T., Zeng, Y.-R., and Chang, Y.-S., Air Pollution Forecasting Using RNN with LSTM. 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). doi:10.1109/dasc/picom/datacom/cyberscitec.2018.00178.

[12]. Freeman, B. S., Taylor, G., Gharabaghi, B., & The, J. (2018).´ Forecasting air quality time series using deep learning. Journal of the Air & Waste Management Association, 68(8), 866–886. doi:10.1080/10962247.2018.1459956.

[13]. V, A., P, G., R, V., & K P, S. (2018). Deep Air Net: Applying Recurrent Networks for Air Quality Prediction. Procedia Computer Science, 132, 1394–1403. doi: 10.1016/j.procs.2018.05.068.