International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT217332

Deep Learning Model for Air Quality Prediction Based on Big Data

P. Parkavi, S. Rathi
Department of Computer Science and Engineering, Government College of Technology,Coimbatore, Tamilnadu, India

Article Info
Volume 7, Issue 3
Page Number: 170-175

Publication Issue :
May-June-2021

Article History
Accepted : 10 May 2021
Published : 15 May 2021

ABSTRACT

Air pollution and its harm to human health has become a serious problem in many cities around the world. In recent years, research interests in measuring and predicting the quality of air around people has spiked. Since the Internet of things has been widely used in different domains to improve the quality for people by connecting multiple sensors. In this work an IOT based air pollution monitoring with prediction system is proposed. The internet of Things is a action interrelated computing devices that are given unique identifiers and the capability of exchange information over a system without anticipating that human to human or human to machine communication. The deep learning algorithm approach is to evaluate the accuracy for the prediction of air pollution. The main objective of the project is used to predict the air Quality. The large dataset works with LSTM for better air quality prediction. The prediction accuracy of air quality with LSTM, the evaluation indicator Root means square error is chosen to measure performance.

Keywords : Air Pollution, IOT, Deep Learning

I. INTRODUCTION

Considering the daily newspapers and any other electronic or print media, a devastating news which is spreading day by day is people is becoming sick and the climate is changing such a way that it has become miserable for living of people. From the aspect from top to bottom, every people are suffering the curse of climate change. The main reason for the climate change and people health is air pollution. It has brought changes in climate like global warming, global dimming, over raining, drought, storms, acid rain, foggy weather etc. The living things on earth and under water are suffering many problems like change in life due to lack of proper facilities of life.

Air is the most useful thing for each and every living thing. Researching on this serious issue this system's main purpose was to estimate the quality of air for people and any other living thing which exist on earth. Very important to know for our living is that how much safe we are now and how the weather and climate has changed for air pollution and it will sustain sound. This system will ease to know the answers for air quality. Four major gas sensors which are responsible for the most air pollution mostly are being used in the system to know the best result of the whole

condition of the air. CO2, CO, LPG, Humidity are declared to be the most responsible for air pollution.

II. RELATED WORK

To measure the air quality, several monitoring methods have been proposed and utilized. According to Elizabeth Bales, CitiSense system gives individuals, the real time tools they need to be able to identify when and where they are exposed to poor air. The main components: a wearable sensor board that pairs with an Android phone, a server-supported, web-based personalized daily pollution map, and a social component supported through Facebook and Twitter integration. The air-quality monitoring unit contains the following 6 sensors attached to a custom board; Carbon Monoxide, Nitrogen Dioxide, Ozone, Temperature, Barometric Pressure, Humidity. The goal of the CitiSense project is to provide individuals with a system that makes the invisible visible.

Indoor air quality analysis is of interest to understand the abnormal atmospheric phenomena and external factors that affect air quality. By recording and analyzing quality measurements, they are able to observe patterns in the measurements and predict the air quality of future. The designed a microchip made out of sensors that is capable of periodically recording measurements, and proposed a model that estimates atmospheric changes using deep learning. The dataset collected in my sql. The LSTM and GRU used for finding prediction accuracy. LSTM is best comparing with GRU.

A. Air Pollution Forecasting:

The dataset that reports on the weather and the level of pollution each hour for five years at the US embassy in Beijing, China.

The data includes the date-time, the pollution called PM2.5 concentration, and the weather information including dew point, temperature, pressure, wind direction, wind speed and the cumulative number of hours of snow and rain. The complete feature list in the raw data is as follows:

No: row number
Year: year of data in this row
Month: month of data in this row
Day: day of data in this row
Hour: hour of data in this row
PM2.5: PM2.5 concentration
DEWP: Dew Point
TEMP: Temperature
PRES: Pressure

The data and frame a forecasting problem where, given the weather conditions and pollution for prior hours, then forecast the pollution at the next hour.

B. Basic Data Preparation:

The data is not ready to use. We must prepare it first. The first step is to consolidate the date-time information into a single date-time so that we can use it as an index in Pandas.

A quick check reveals NA values for pm2.5 for the first 24 hours. We will, therefore, need to remove the first row of data. There are also a few scattered “NA” values later in the dataset; we can mark them with 0 values for now.

The script below loads the raw dataset and parses the date-time information as the Pandas Data Frame index. The “No” column is dropped and then clearer names are specified for each column. Finally, the NA values are replaced with “0” values and the first 24 hours are removed.

The “No” column is dropped and then clearer names are specified for each column. Finally, the NA values are replaced with “0” values and the first 24 hours are removed..

C. Multivariate LSTM Forecast Model:

We will fit an LSTM to the problem.

i) LSTM Data Preparation

The first step is to prepare the pollution dataset for the LSTM. This involveas framing the dataset as a supervised learning problem and normalizing the input variables. We will frame the supervised learning problem as predicting the pollution at the current hour (t) given the pollution measurement and weather conditions at the prior time step. This formulation is straightforward and just for this demonstration. Some alternate formulations you could explore include:

Predict the pollution for the next hour based on the weather conditions and pollution over the last 24 hours.

Predict the pollution for the next hour as above and given the "expected" weather conditions for the next hour.

First, the "pollution.csv" dataset is loaded. The wind direction feature is label encoded (integer encoded). This could further be one-hot encoded in the future if you are interested in exploring it. Next, all features are normalized, then the dataset is transformed into a supervised learning problem. The weather variables for the hour to be predicted (t) are then removed.

ii) Fit Model

To fit an LSTM on the multivariate input data.

Split the prepared dataset into train and test sets. To speed up the training of the model for this demonstration, we will only fit the model on the first year of data, then evaluate it on the remaining 4 years of data. If you have time, consider exploring the inverted version of this test harness.

D. DATA PROCESSING GRAPH

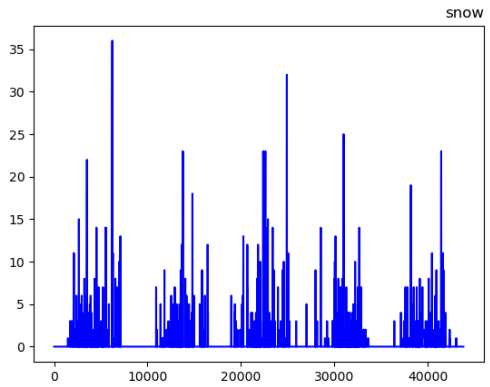

Fig.1 Snow

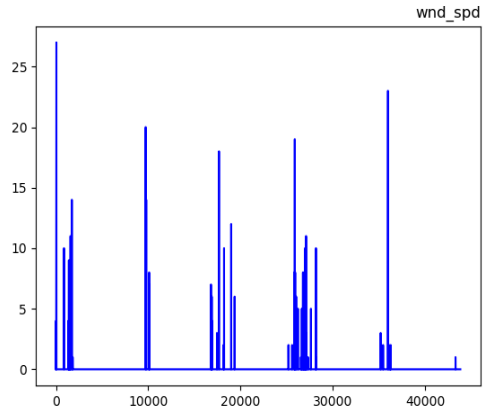

Fig.2 Wind Speed

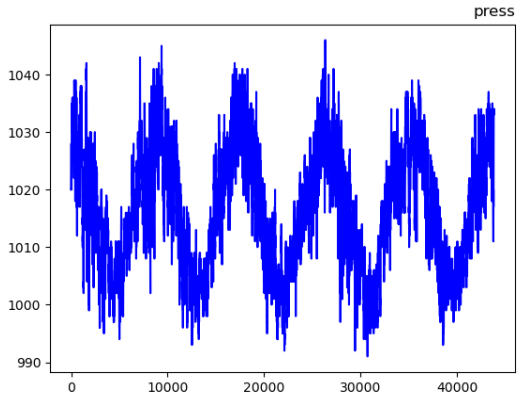

Fig. 3 Pressure

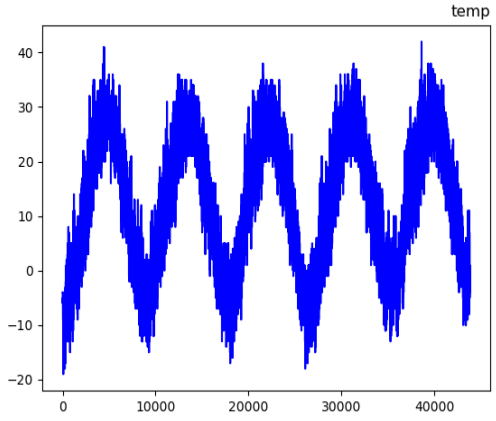

Fig.4 Temperature

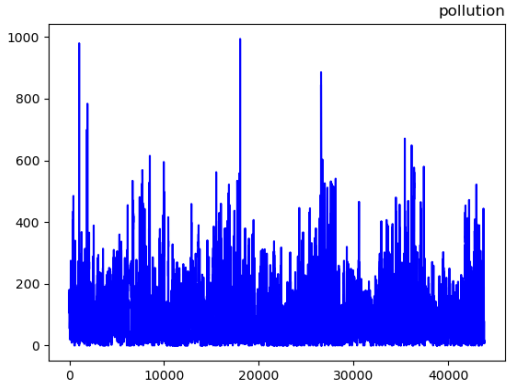

Fig. 5 Pollution

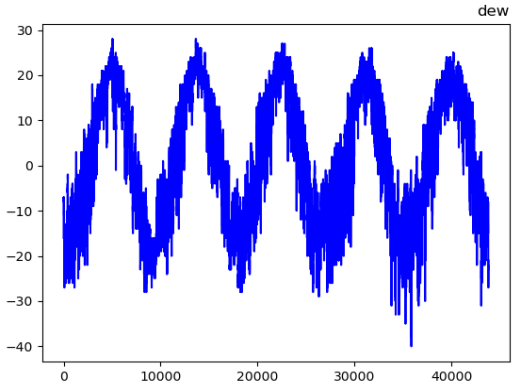

Fig.6 Dew

E. Long Short Term Memory:

LSTM is a special kind of recurrent neural network (RNN) and capable of learning long-term dependencies. It was introduced by Hochreiter and Schmidhuber in order to overcome vanishing gradient problem in 1997.

In this neural network model, a memory block takes the place of each ordinary neuron in the hidden layer of standard recurrent neural network .

The LSTM block shown in Fig. 1 has an input gate, a forget gate and an output gate which regulate the flow of information into and out of the cell. These gates, block input and block output as follows:

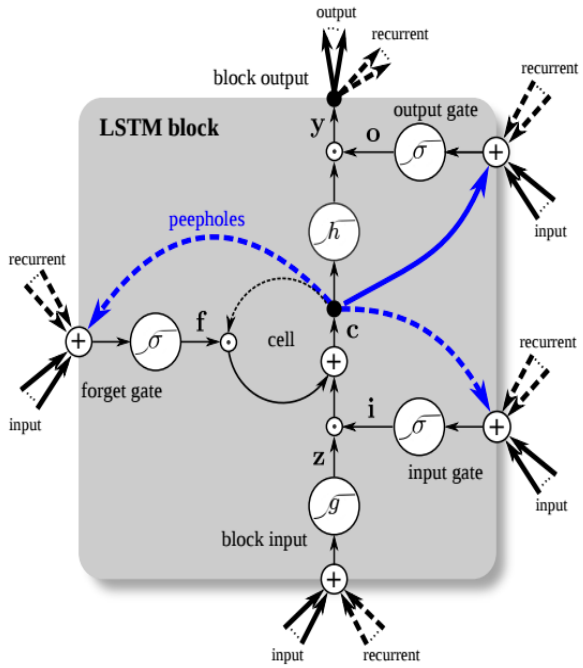

Fig.7 LONG SHORT TERM MEMORY BLOCK

block input:$\boldsymbol{zt}$ = g $\mathbf{W}$zx! + $\mathbf{R}$zy!!! + b! (1)

input gate:$\boldsymbol{it}$ = σ $\mathbf{W}$!x! + $\mathbf{R}$!y!!! + p! $\odot c$!!! + b! (2)

forget gate:$\boldsymbol{ft}$ = σ $\mathbf{W}$fx! + $\mathbf{R}$fy!!! + p! $\odot c$!!! + b! (3)

cell state:$\boldsymbol{ct}$ = $\boldsymbol{it}$ $\odot \boldsymbol{zt}$ + $\boldsymbol{ft}$ $\odot c$!!! (4)

output gate:$\boldsymbol{ot}$ = σ $\mathbf{W}$ox! + $\mathbf{R}$oy!!! + p! $\odot \boldsymbol{ct}$ + b! (5)

block output:$\boldsymbol{yt}$ = $\boldsymbol{ot}$ $\odot \boldsymbol{h}(\boldsymbol{ct})$ (6)

where $\boldsymbol{xt}$ is input vector at time t, $\boldsymbol{Wz}$,$\boldsymbol{Wi}$,$\boldsymbol{Wf}$,$\boldsymbol{Wo}$are the weights matrices connecting $\boldsymbol{xt}$ to the three gates and block input, $\boldsymbol{Rz}$,$\boldsymbol{Ri}$,$\boldsymbol{Rf}$,$\boldsymbol{Ro}$are recurrent weight matrices connecting $\boldsymbol{yt}$!$\mathbf{1}$ to the three gates and block input, b!, b!, b!, b! are the bias vectors. σ represents the

logistic sigmoid function and h represents hyperbolic tangent function. σ is used for as activation of the gates and g is used as the block input and output activation function.

The LSTM with 50 neurons in the first hidden layer and 1 neuron in the output layer for predicting pollution. The input shape will be 1 time step with 8 features.

The Mean Absolute Error (MAE) loss function and the efficient Adam version of stochastic gradient descent.

The model will be fit for 50 training epochs with a batch size of 72.The internal state of the LSTM in Keras is reset at the end of each batch.

F. TESTING AND TRAINING DATASET

The dataset are tested and trained using the lstm algorithm. The dataset are split into test set and train set into input and output variables. The testing and training of dataset loss for lstm is as shown in fig . The train and test loss are printed at end of the each training epoch.

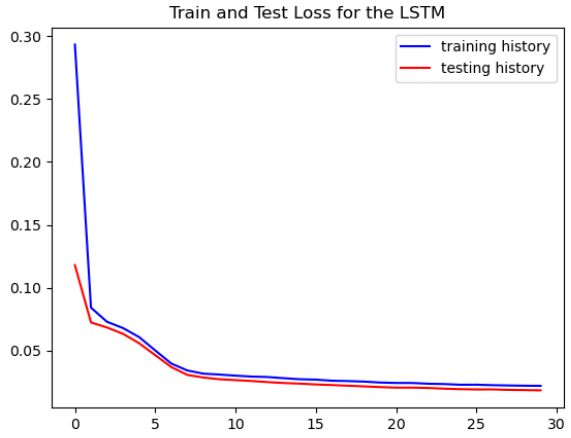

Fig8. Testing and Training

III. CONCLUSION

In this paper, long short term memory of deep learning algorithm are implemented to predict the air pollution with various environmental consideration. The algorithm showed the prediction accuracy of air quality with 30000 samples data given as input to test and train.
In future the new dataset are taken from the various sensor and work to predict the real-time air quality in the surroundings with the air quality index.

IV. REFERENCES

[1]. D. Zhang and S. S. Woo, ``Predicting air quality using moving sensors(poster)," in Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services, Jun. 2019, pp. 604-605.

[2]. DAN ZHANG 1,2 AND SIMON S. WOO3,4 “Real Time Localized Air Quality Monitoring and Prediction Through Mobile and Fixed IoT Sensing”., Stony Brook, NY 11794, USA.

[3]. J. Ahn, D. Shin, K. Kim, and J. Yang, ``Indoor air quality analysis using deep learning with sensor data," Sensors, vol. 17, no. 11, p. 2476, 2017.

[4]. I. Kok, M. U. Simsek, and S. Ozdemir, ``A deep learning model for airquality prediction in smart cities," in Proc. IEEE Int. Conf. Big Data (BigData), Dec. 2017, pp. 19831990.

[5]. Gulshan Taj Mohammed Navi Anwar et al Air quality monitoring and assessment using IoT; 2020 IOP Conf. Ser.: Mater. Sci. Eng. 955 012006

[6]. Elizabeth Bales, Nima Nikzad, Celal Ziftci, Nichole Quick, William Griswold “Personal Pollution Monitoring: Mobile Real-Time Air-Quality in Daily Life”, University of California, San Diego La Jolla, CA 92093-0404ess

[7]. M. Kampa and E. Castanas, ``Human health effects of air pollution,"Environ. Pollution., vol. 151, no. 2, pp. 362367, Jan. 2008.

[8]. Y. Jiang, L. Shang, K. Li, L. Tian, R. Piedrahita, X. Yun, O. Mansata, Q. Lv, R. P. Dick, and M. Hannigan, ``MAQS: A personalized mobile sensing system for indoor air quality monitoring,'' in Proc. 13th Int. Conf.

[9]. F. Xia, L. T. Yang, L. Wang, and A. Vinel, "Internet of Things," International Journal of Communication Systems, vol. 25, pp. 1101-1102, 2012.

[10].E. Boldo, S. Medina, A. Le Tertre, F. Hurley, H.-G. Mücke, F. Ballester,and I. Aguilera, ``Apheis: Health impact assessment of long-term exposureto PM2.5 in 23 European cities,'' Eur. J. Epidemiology, vol. 21, no. 6,pp. 449-458, Jun. 2006.

[11].S. Devarakonda, P. Sevusu, H. Liu, R. Liu, L. Iftode, and B. Nath, ``Realtime air quality monitoring through mobile sensing in metropolitan areas,''in Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput. UrbComp,2013, p. 15.

[12].B. Maag, Z. Zhou, and L. Thiele, ``W-Air: Enabling personal air pollution monitoring on wearables,'' Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol., vol. 2, no. 1, p. 24, 2018

[13].J. Lin, A. Zhang, W. Chen, and M. Lin, ``Estimates of daily PM2.5exposure in beijing using spatio-temporal kriging model,'' Sustainability,vol. 10, no. 8, p. 2772, 2018.

[14].S. Fotouhi, M. H. Shirali-Shahreza, and A. Mohammadpour, `Concentration prediction of air pollutants in tehran,'' in Proc. Int. Conf. Smart CitiesInternet Things SCIOT, 2018.

[15].A. C. Rai, P. Kumar, F. Pilla, A. N. Skouloudis, S. Di Sabatino, C. Ratti,A. Yasar, and D. Rickerby, ``End-user perspective of low-cost sensorsfor outdoor air pollution monitoring,'' Sci. Total Environ., vols. 607-608,pp. 691-705, Dec. 2017. Journal URL : http://ijsrcseit.com/CSEIT2063197

Cite this article as :

P. Parkavi, S. Rathi, "Deep Learning Model for Air Quality Prediction Based on Big Data", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 170-175, May-June 2021. Available at doi : https://doi.org/10.32628/CSEIT217332
Journal URL : https://ijsrcseit.com/CSEIT217332