March - April 2020
ISSN: 0193-4120 Page No. 8926 - 8932

PSO based Deep Learning Model for Forecasting $PM_{2.5}$.

S. Jeya[1],Dr. L. Sankari[2]
[1]Research Scholar, [2]Associate Professor.
[1,2]Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India.
[1]jeya.s.2000@gmail.com,[2]sankarics@srcw.ac.in

Article Info
Volume 83
Page Number: 8926 - 8932
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 09 April 2020*

Abstract

Entire eco system and all the living organisms in the earth are affected by the toxic air pollutants to a great extent day by day. It is essential to forecast air pollution in order to limit the pollutant concentration and maintain the Air Quality standards prescribed by the government. In this paper PSO optimized 1D CNN and BIGRU are applied to predict accurately the fine particulate matter (PM2.5) pollutant which triggers various death causing diseases when its level exceeds the prescribed limit. The influence of meteorological parameters on air pollution cannot be omitted when doing the prediction analysis. UCI Machine Learning Repository Beijing PM2.5 time series dataset along with meteorological attributes are taken for this analysis. Proposed PSO based CNN-BIGRU model prediction results achieves perfect prediction performance than the existing deep learning convolutional-based bidirectional gated recurrent unit short term forecast of PM2.5 model. The proposed models RMSE, MAE, SMAPE are relatively low with the existing model.

I. INTRODUCTION

Serious environment problem air pollution, particularly when exceeding the limit of PM2.5 which has adverse effect on human health due to various death causing diseases in developing countries. In order to enhance the air quality accurate forecasting is essential for the government to monitor, control and cutting down the major emission sources which increase the level of PM2.5 [1]. Pollution standard exceeds World Health Organization guiding principle in most of the cities from low and middle income countries. Though statistical models were used earlier for predicting the air pollution concentration, deep learning models provide more accurate forecasting results. It is found out through various studies that there exists a relationship or correlation between meteorological factors such as temperature, relative humidity, wind speed, surface pressure and PM2.5 concentration in terms of regional and seasonal variations [2]. Time series prediction analysis is used to explore the, effect and relationship between pollutants and weather parameters [18].

II. RELATED STUDY:

Mehdi ZamaniJoharestani et al. applied ensemble models (RF, XG Boost), and deep machine learning model approaches to forecast PM 2.5. XG Boost model obtained the best performance out of three models applied for analysis [3]. Chiou-Jye Huang et al. for forecasting, estimating, monitoring and controlling PM2.5, combining CNN and LSTM model is suggested. This deep neural network architecture shows better performance by four measurements RMSE, MAE, Pearson correlation coefficient and Index of Agreement. Along with rain, wind speed (cumulated weather data) the PM2.5 pollutant for next hour is forecasted. Feature extraction and prediction is done by CNN and LSTM [4]. Ricardo Navares et al. suggested for predicting one day ahead air quality for a group of pollution concentrations which includes

Published by: The Mattingley Publishing Co., Inc.

NO2,CO,O3,PM10, SO2 and some other air pollen concentrations from Madrid city by applying the model LSTM recurrent artificial neural networks. They proved single complete model performs better rather than multiple individual models [5]. Athira V et al. in their analysis used a combination of RNN, GRU and LSTM to forecast PM10 [6]. QING TAO et al. proposed convolutional-based bidirectional gated recurrent unit for predicting PM2.5 concentration and concluded that the error is less in the present model and performance is better than other models [7]. Brian S. Freeman et al. proposed a deep learning model consists of recurrent neural network (RNN) with long short-term memory (LSTM) to predict eight hour averaged surface ozone (O3). The forecasting is done by taking the values of hourly air quality and meteorological data and the forecast results up to 72 hours with fewer errors are produced [16]. K SrinivasaRao et al. proposed RNN-LSTM framework to predict 12 pollutants from Visakhapatnam real time data sets and achieved higher accuracy in estimating hourly based air ambience [17]. Tae-Young Kim et al. proposed PSO-based CNN-LSTM model of optimal prediction for energy prediction. Hyper parameters are automatically determined by PSO [26].

These following sections are used to describe this entire article (I) Introduction. (II) Related Study. (III) - A. Data pre-processing B. Correlation analysis of PM2.5 and meteorological data. C. Small introduction of recurrent neural network.

D. Brief coverage of convolutional neural network. E. Basic ideas of Bidirectional gated recurrent unit. F. Particle swarm optimization. G. Accuracy measures. (IV) Experimental Results. (V) Conclusion.

A. Data Pre-Processing:

For this proposed methodology UCI machine learning repository Beijing PM2.5 data set is used and meteorological data is from international airport of Beijing. Data contains hourly recorded PM2.5 concentration, and some of the weather data such as temperature, air pressure, dew point, wind speed, wind direction, snowfall and rainfall from Jan 1, 2010 to Dec 31, 2014. Wind direction has four features as NW, CV, SE and NE and is encoded with float values −10, 0, 10 and 20. For missing values of PM 2.5 concentration due to sensor errors, previous timestamp data are used for filling. The data is then normalized.

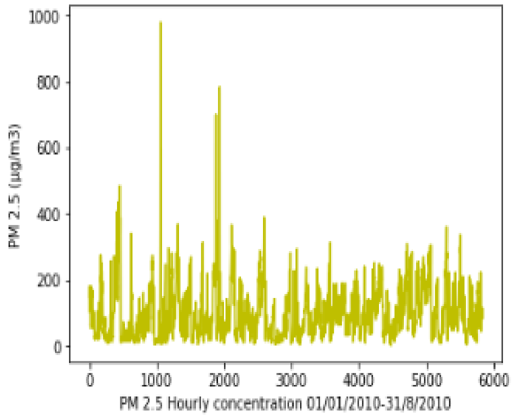

Figure 1 - Particulate Matter 2.5 Concentration over a period of time.

B. Correlation analysis:

The letter r in the below formulae is used to denote the relationship or association between the two variables and computed with a number that ranges between −1 and +1. Value zero represents no correlation, 1 denotes a perfect correlation. If there is a negative correlation then the variables are inversely related. After correlation analysis if the value is in positive range then the variables are positively correlated otherwise for negative range values it is considered as negative correlation [8]. When the correlation coefficient is −1 or +1, in a scatter plot the points will lie on a straight line, representing strong correlation between the variables [21]. In order to identify how strong the relationship between two variables below formulae is used which finds out the correlation result.

Published by: The Mattingley Publishing Co., Inc.

TEST Engineering & Management

March - April 2020
ISSN: 0193-4120 Page No. 8926 - 8932

$$r=\frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}} \quad [9] \quad (1)$$

Identifying the association is essential between PM2.5 concentration which is going to be forecasted and other various influencing meteorological factors which in turn ensure that proper input features are used for the prediction. It is found out from correlation analysis that PM 2.5 has positive correlation with dew point, wind direction and snowfall and restof the meteorological variables such as temp, pressure, wind speed and rainfall have negative correlation. From the above correlation model it is shown that air pressure, snowfall and rainfall have small correlation coefficients. Omitting those features will improve the model performance because they are not related well with PM2.5. Rest of the following variables dew point, PM2.5, temperature, wind direction and wind speed are considered as the input to the prediction model [7].

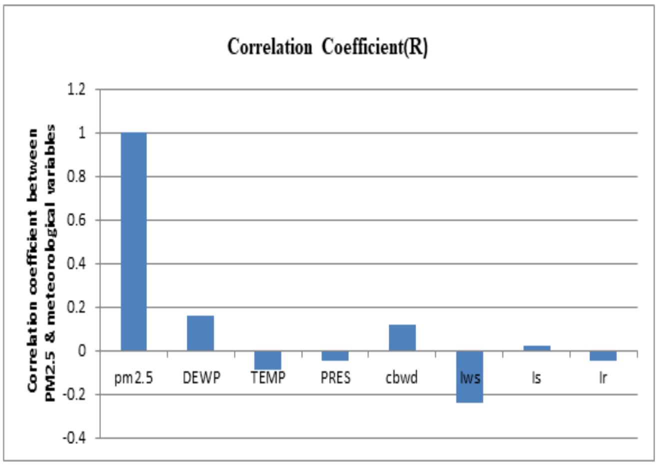

Figure 2 - Correlation analysis of PM2.5 and meteorological data

C. Recurrent Neural Network:

RNN is deep learning algorithm which is recurrent in nature and the output is based on previous computations [14]. Past or previous values are used to forecast the data especially which is time series in nature by RNN. Feedback loop is used to retain the input from previous state. [20]. Not only current input but also the previous hidden state are used by each node to produce the current hidden state and output.

$$h_t=f(W_h h_t\text{-}1+V_h x_t+b_h) \qquad (2)$$

$$o_t=f(W_o h_t+b_o) \qquad (3)$$

h_t = hidden node for time step.

From the above equation W, V weights, b bias value for hidden and output states, f is an activation function [11]

D. Convolutional neural network:

Convolutional neural network having convolutional and subsampling layers as well as fully connected layers is a deep learning model. In short the convolutional layers, pooling layers, additionally drop out layer to avoid over fitting and flatten layer activation function and fully connected layers form the architecture of CNN [15]. A convolution layer performs feature extraction. Regression and classification are done by convolutional and down sampling techniques as CNNs transform the input layer by layer. Pooling layers are used to reduce the dimensionality, obviously the parameters are reduced and the computational complexity is also reduced [13]. Local trend change features of the data set used in this experiment over time is captured [7].

E. Bidirectional gated recurrent unit:

In Bi-GRU update gate and reset gate are used to overcome the RNN vanishing gradient problem. It provides excellent results than long short term memory in some cases. It has both forward and backward states. Recurrent neural networks uses previous time sequences to learn representations, whereas bi directional gated recurrent unit networks uses future time steps to learn representations. In this case a better understanding and ambiguity elimination allows a better performance [11]. Current state prediction is based on Information from previous time steps and also from later time

TEST Engineering & Management
March - April 2020
ISSN: 0193-4120 Page No. 8926 - 8932

steps [19]. GRU's two gates are reset and update gates. The reset gate help decides to combine previous memory with new input and update gate helps to keep the needed previous memory, past information from the previous time steps for future use. The 4 components of GRU architecture – one input vector, update gate vector, one reset gate vector, one output vector. For given xI Input vector and the corresponding weight parameter matrix and vector and update gate vector zI with corresponding weight parameter matrix and vector Wz,Uz,bz and the same for reset gate vector Wr,Ur,br, output vector ht with corresponding weight parameter matrix and vector Wh,Uh,bh. The σr rectified linear unit function activation function. Operator ○ is used to represent the Hadamard product that produces element wise product.

The equation is below:

The equation is below:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \qquad (4)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \qquad (5)$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \sigma_r(W_h x_t + U_h(r_t \circ h_{t-1}) + b_h) \qquad (6)$$

The hidden unit uses for short term and long term dependencies the reset gate and update gate. They also depend on previous hidden units [12].

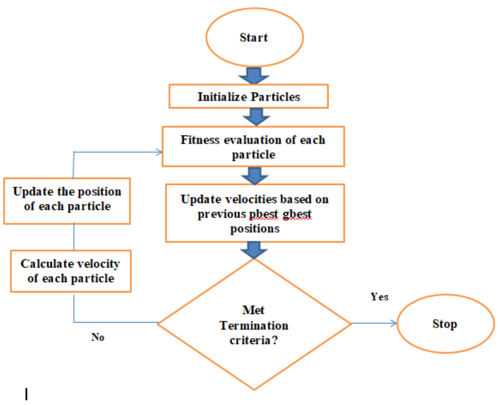

Figure 3 – Particle Swarm Optimization

F. Particle swarm optimization:

Particle swarm optimization is a population-based heuristic global optimization technology. It is a nature-inspired evolutionary computing algorithm. Dynamically altering the velocity of each particle will adjust the path of each individual in the search space, based on the flying experience of each particle and of other particles in the search space. Easy implementation and quick convergence made it very popular [22]. In PSO in the beginning particles which are nothing but random solutions are initialised and PSO has velocity, position, and fitness value which is calculated by a fitness function [24]. To attain the optimal solution, every particle moves not only in the previously known best position but also in the global best position in the swarm [23].

$$v_{id}^{k+1} = v_{id}^{k} + c_1 rand_1^k (pbest_{id}^k - x_{id}^k) + c_2 rand_2^k (gbest_d^k - x_{id}^k) \qquad (7)$$

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \qquad (8)$$

v_{id}^{k} - Velocity of dimension d– particle i and k iteration,

x_{id}^{k} - Corresponding position

$c_1$andc_2 are the accelerate parameters,

$rand_1$and $rand_2$ are considered two random numbers in [0,1] [24].

PSO is considered as the best solution for solving nonlinear optimization problem.

CNN needs large number of parameters to be tuned to achieve the best performance from particular dataset.

Automatic design methods are used for hyper parameter optimization like finding the number of layers and optimal number of neurons etc. [25]. Main contributions of this paper are, a. Optimizing 1D CNN parameters by PSO b. prediction of P.M 2.5 by CNN and BIGRU.

TEST Engineering & Management
March - April 2020
ISSN: 0193-4120 Page No. 8926 - 8932

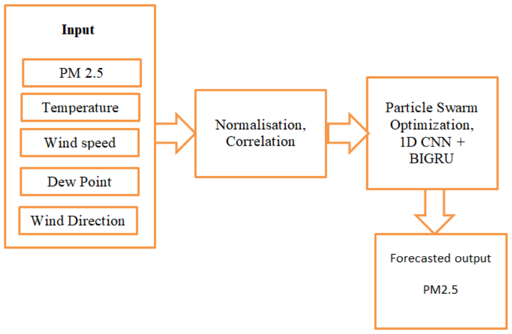

Figure 4 – Structure of the proposed model

G. Accuracy Measures:

Performance of the forecasting model is measured by statistical measurement Root mean square error (RMSE) in air quality, meteorology and climate research. Another measurement to assess the model performance is Mean Squared Error (MAE). RMSEs and MAEs are certain statistical metrics often used to assess the performance of the model.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \qquad (9)$$

$$\text{MAE} =\frac{1}{n}\sum_{i=1}^{n}|e_i| \quad [10] \qquad (10)$$

One more metric in this regard to measure the accuracy of the model performance is symmetric mean absolute percentage error (SMAPE). Both RMSE and SMAPE error metrics are used to evaluate not only the degree of change but also the accuracy of data and thus used to measure the prediction quality of the model.

$$\text{SMAPE } (y', y) =\frac{1}{n}\sum_{i=1}^{n} \frac{|y_i'-y_i|}{(y_i'+y_i)/2} (11)[7]$$

III. EXPERIMENTAL RESULTS:

In this experiment, initially optimising the 1 D CNN hyper parameters by PSO model will take place and later BIGRU is used to forecast the Pollutant. A convolution neural network does the down sampling of data to avoid complexity by reducing the size of the data and increase the learning ability, simplicity of the model. This data will again feed into RNN in order to find out the nonlinear relationship between pollutant and climatic data.

Table 1 – Parameter range of 1D CNN

Conv layer	Filter size 2,3,4,..8	Padding same, valid	Stride 1,2,3,4	Activation = 'Relu'
Pooling layer	Filter size 2,3,4,..8	Padding same, valid	Pool type – Average, Maximum	
Drop out layer	Drop out value 0.1,.....0.4			
FC layer	1 neuron, softmax activation function			

This experiment is conducted by using minimal features of 1D CNN. Two convolutional layer and in between one pooling layer is used. Encoded features for PSO in order to optimize the performance of 1D CNN are the filter size ranges from 2 to 8, stride 1,2,3,4, dropout ranging from 0.1 to 0.4. Out of that encoded values (Table-1), Rectified Linear Units are considered better though tanh and sigmoid does their job better. Small number of parameters are taken and tried on trial and error basis. For Particle swarm optimization parameters are set as Swarm size is 10. Number of iterations are 10. The activation function is fixed as ReLU for convolutional network. Drop out value is 0.2. Strides = 1, Average Pooling with pool size= 3. Padding is same. Small number of epochs is selected (20), Batch size 16. Filter size 3. In case of BIGRU two layers with 64 neurons are used.

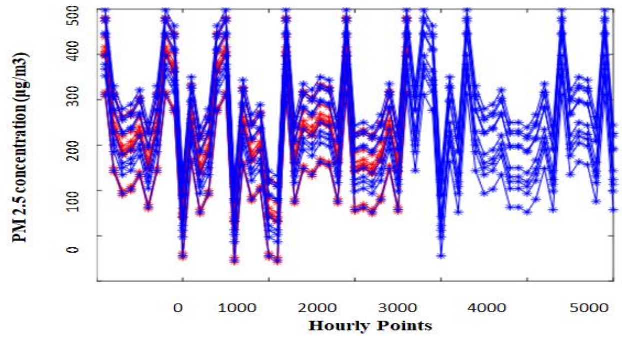

Figure 5 - Forecasting results of the proposed model

The proposed approach accuracy measures are RMSE value = 12.8855. MAE = 9.4545. SMAPE = 0.1667.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8926 - 8932

IV. CONCLUSION:

Deep learning models show superior performance than shallow machine learning models. This study applied hyper parameters of the network model 1D-CNN optimisation by PSO and forecasted PM2.5 time series by Bi-GRU. The error measures RMSE, MAE and SMAPE are less in the proposed method when compared to the existing model which helps in estimating the air pollution appropriately. Accuracy in predicting the pollution level helps to monitor and control the pollution level by the government authorities. Countermeasures can be taken appropriately due to heavy pollution in urban areas. For analysing the performance of the network, experimentation of various parameter tuning on trial and error basis by using different optimization algorithms in order to achieve better results will be extended in future research

REFERENCES

[1]. Xiaochun Yang, Qizhong Wu, Rong Zhao, Huaqiong Cheng,Huijuan He,Qian Ma,Lanning Wang, HuiLuo, "New method for evaluating winter air quality: $PM_{2.5}$ assessment using Community Multi-Scale Air Quality Modeling (CMAQ) in Xi'an", Volume 211, (2019), pp - 18-28, Atmospheric environment.

[2]. Qianqian Yang, Qiangqiang Yuan, Tongwen Li, HuanfengShen and Liangpei Zhang, "The Relationships between $PM_{2.5}$ and Meteorological Factors in China: Seasonal and Regional Variations", International Journal of Environmental Research and Public Health, (2017), *14*(12), 1510.

[3]. Mehdi ZamaniJoharestani , Chunxiang Cao, Xiliang Ni, Barjeece Bashir and SomayehTalebiesfandarani, "$PM_{2.5}$ Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data", Atmosphere (2019), 10, 373.

[4]. Chiou-Jye Huang and Ping-HuanKuo, "A Deep CNN-LSTM Model for Particulate Matter ($PM_{2.5}$) Forecasting in Smart Cities", Sensors 2018, 18, 2220.

[5]. Ricardo Navares, José L. Aznarte, "Predicting air quality with deep learning LSTM: Towards comprehensive models", Ecological Informatics - 55 (2020) 101019.

[6]. Athira V, Geetha P, Vinayakumar R, Soman K P, DeepAirNet: Applying Recurrent Networks for Air Quality Prediction, Procedia Computer Science -132 - (2018) - pp - 1394–1403

[7]. QING TAO, FANG LIU, YONG LI , AND DENIS SIDOROV "Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU" IEEE Access vol 7 - (2019) ,pp - 76690 -76698

[8]. HaldunAkoglu, "User's guide to correlation coefficients", Turkish Journal of Emergency Medicine -18 - (2018) pp - 91–93

[9]. Sergio Plata, "Anote on Fisher's correlation coefficient", Applied Mathematics Letters-19 - (2006) pp-499–502

[10]. T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature", Geoscientific Model Development, 7, (2014) - pp-1247–1250.

[11]. HTET MYET LYNN, SUNG BUM PAN, AND PANKOO KIM,"A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks", IEEE Access, -7- (2019), pp- 145395-145405.

[12]. Shi Chen, Bing Zheng, TianyongHao, "Capsule-Based Bidirectional Gated Recurrent Unit Networks for Question Target Classification", CCIR 2018. Lecture Notes in Computer Science, vol - 11168. Springer, Cham.

[13]. Rikiya Yamashita, Mizuho Nishio , Richard KinhGian Do, Kaori Togashi, "Convolutional neural networks: an overview and application in radiology", Springer, Insights into Imaging - (2018) – 9, pp-611–629.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8926 - 8932

[14]. Yuan luo," Recurrent neural networks for classifying relations in clinical notes", Journal of Biomedical Informatics - 72 (2017) pp- 85–95.

[15]. SakshiIndolia, Anil Kumar Goswami, S. P. Mishra, PoojaAsopa,"Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach", Procedia Computer Science -132 (2018) pp-679–688.

[16]. Brian S. Freeman, Graham Taylor, BahramGharabaghi& Jesse Thé, "Forecasting air quality time series using deep learning", JOURNAL OF THE AIR & WASTE MANAGEMENTASSOCIATION, 2018, VOL. 68, NO. 8, pp - 866–886.

[17]. K SrinivasaRao, G. Lavanya Devi, N. Ramesh, "Air Quality Prediction in Visakhapatnam with LSTM based Recurrent Neural Networks", International Journal of Intelligent Systems and Applications(IJISA), Vol.11, No.2, pp.18-24, 2019.

[18]. Fahimeh. Hosseinibalam and Azadeh. Hejazi, "Influence of Meteorological Parameters on Air Pollution in Isfahan", 2012 3rd International Conference on Biology, Environment and Chemistry IPCBEE vol.46 (2012).

[19]. Jingren Zhang, Fang'ai Liu, WeizhiXu andHui Yu, "Feature Fusion Text Classification Model Combining CNN and BiGRU with Multi-Attention Mechanism", Future Internet 2019, 11, 237.

[20]. Jitendra Kumar, RimshaGoomer, Ashutosh Kumar Singh, "Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters", Procedia Computer Science 125 (2018) pp - 676–682.

[21]. NikolaosPandis, "Correlation and linear regression", American Journal of Orthodontics and DentofacialOrthopedics (2016) Vol - 149, issue- 2.

[22]. Amir Robati, Gholam Abbas Barani, HosseinNezamAbadi Pour, Mohammad JavadFadaee, JavadRahimi Pour Anaraki, "Balanced fuzzy particle swarm optimization",Applied Mathematical Modelling - 36 - (2012), pp -2169–2177.

[23]. Yudong Zhang, Shuihua Wang and GenlinJi, "A Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications", Mathematical Problems in Engineering Vol - 2015, Article ID - 931256, 38 pages.

[24]. Wensheng Yi, Min Yao, and Zhiwei Jiang, "Fuzzy Particle Swarm Optimization Clustering and Its Application to Image Clustering", Advances in Multimedia Information Processing - PCM 2006, pp - 459-467.

[25]. Vishal Passricha, Rajesh Kumar Aggarwal, "PSO-based optimized CNN for Hindi ASR", International Journal of Speech Technology", **22,** 1123–1133 (2019). https://doi.org/10.1007/s10772-019-09652-3.

[26]. Tae-Young Kim, Sung-Bae Cho," Particle Swarm Optimization-based CNN-LSTM Networks for Forecasting Energy Consumption",(2019) IEEE Congress on Evolutionary Computation, DOI: 10.1109/CEC.2019.8789968.