OELP 2023
(ID3802)

REPORT
(An efficient AI based weather forecasting system)

NAME: OM TIWARI 132001027
PRIYANSHU GUPTA 112001033

SUBMISSION DATE: 19.05.23

BACHELOR OF TECHNOLOGY

IIT PALAKKAD, Kerala

INDIAN INSTITUTE OF TECHNOLOGY
PALAKKAD

MAY 2023

INDIAN INSTITUTE OF TECHNOLOGY
(Autonomous Institution)
PALAKKAD – 678557

AN EFFICIENT AI BASED WEATHER FORECASTING SYSTEM

Abstract

Every aspect of industry and existence is directly correlated with the daily weather, and social growth depends greatly on accurate weather forecasts. A deep learning network was created to forecast urban weather conditions based on the data features of urban weather conditions, and its viability was demonstrated through trials. Empirical mode decomposition (EMD) was used to perform the stationary processing for the daily humidity, minimum pressure, maximum pressure, average pressure, maximum temperature, minimum temperature, average temperature, and wind speed due to the non-stationary and seasonal fluctuation of the time series of daily weather conditions in Shenzhen from 2015 to 2019. According to the level of importance, the deconstructed components, residual sequence, and original sequence were recreated.

In order to increase the precision of the forecast of Shenzhen weather features, a long short-term memory (LSTM) neural network was employed for the daily weather forecast in Shenzhen. This model took advantage of the LSTM model's advantages in time-series data processing. The experimental findings demonstrate that our EMD-LSTM model design outperforms conventional models in terms of forecasting precision and efficiency, which offers fresh approaches to weather forecasting.

Introduction

The weather forecast has a big impact on people's productivity and standard of living in today's society, and daily travel, agricultural output, natural catastrophe avoidance, and other fields are all crucial to the smooth running of contemporary society.

Typhoons and flood disasters have recently affected Shenzhen; however, proper weather forecasting can stop flood disasters. The outcomes of weather forecasts are utilized to support weather warning systems as well as offer reasonable information for emergency response and contingency planning. The daily weather forecast for Shenzhen is therefore very important for avoiding natural disasters.

Automatic and intelligent technologies started to play a significant part in weather forecasting as a result of the growing size of meteorological data, or "big data," on the market. As a result, the public's need for increased weather forecast accuracy is growing daily. Since research on weather conditions has shown that most have seasonal trends and the weather forecast inaccuracy of conventional time-series models, one should fully utilize the variety of meteorological history data to improve the accuracy of forecasting weather conditions.

EMD was utilized to address nonlinear sequence issues and demonstrated superior data processing performance over conventional approaches. The output from each model was combined to get the final predicted outcome. It was established that the outcomes of the prediction model following EMD decomposition are more accurate.

Data and Methods

Data :

We found that Shenzhen data has much more variations than the other datasets we saw so we looked forward to continue training our model on shenzhen dataset as achieving high accuracy on shenzhen data can guarantee it to work it on other datasets of trivandrum and palakkad.

EMD Method :

Each signal is broken down into stationary sequences of various feature scales using this method, and each stationary sequence is an intrinsic modal function (IMF) that contains various feature scales from the original signal. The average of the upper and lower envelopes of the original signal must always be zero, and each IMF component must satisfy both of the following requirements simultaneously: the number of original signal extreme points crossing zero must be equal to or different by one. Any original time-series signal that needs to be processed goes through the following EMD decomposition steps:

The upper and lower envelope lines are created using cubic spline interpolation and the upper and lower envelope lines' mean values are calculated in accordance with the upper and lower extreme value points of the original signal.

The original signal is subtracted from the mean envelope to get the intermediate signal.

Repeat the process several times to obtain the IMF signals, and then use the Pearson correlation coefficient to identify the IMF feature with the highest value as the best IMF feature.

LSTM Method

LSTM network structure is a neural network model of recurrent neural network (RNN) formed by adding different gating units in the hidden layer, which has longer short-term memory, stronger memory ability and better processing of long serial number signal data. It enables the recurrent neural network to effectively utilize the training data in a wide range, thus improving the performance of the model.

We used adam optimizer and 3 hidden layers of 50 neurons each which came out to be the best figure as the results kept going worse as we increased or decreased the neurons.

Experimental Design

Construction of EMD-LSTM Combined Model

As we found out that all features are dependent on each other so we are taking all features in order to find one feature and then EMD that feature into IMF. Further we find the best correlation IMF using Pearson coefficient. We give input of all features along with the IMF and train the model on that to give that feature value as the output.

Design of Forecast Model

After training 8 different models for all of the features we will ask for which feature to compute and then give the prediction for next 7 days. For achieving this we are taking the last 7 days data and predicting the next 1 day of data. Next we use a sliding window and update that feature in the data and find the next day's prediction. By this we can achieve prediction of many days of one feature.

Experiment and Analysis

Data Preprocessing

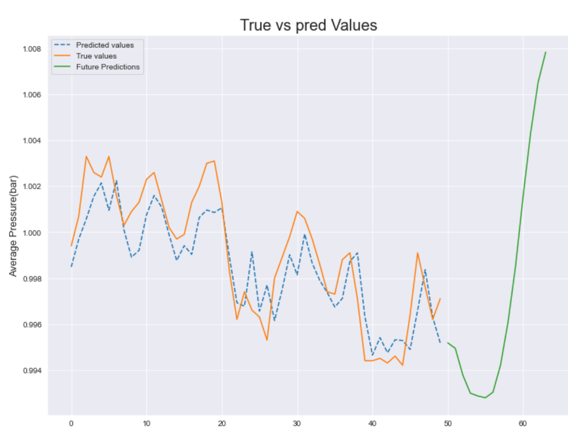
Variable correlation is determined by using the KMO Test. The closer the KMO test value is to 1, the stronger the correlation between variables is.

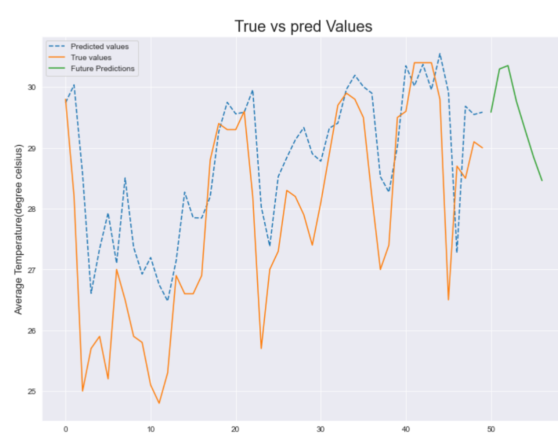
Each component of seven features was obtained through EMD Decomposition and correlation coefficient analysis was carried out between the decomposed feature and all components decomposed.

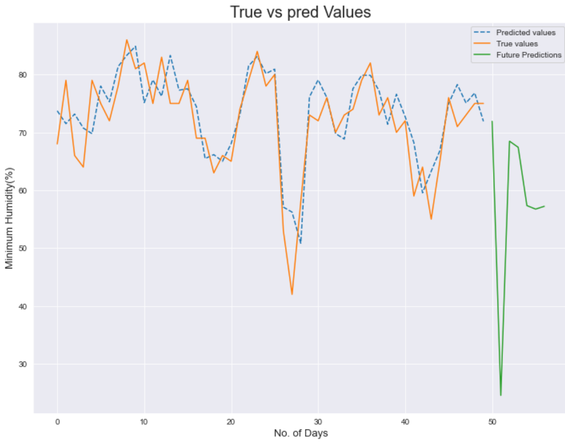
Decomposed components and the minimum temperature can be used to compute the Pearson coefficient of each component and the minimum temperature of the original sequence, and the variables that have a high correlation with the minimum temperature can be eliminated.

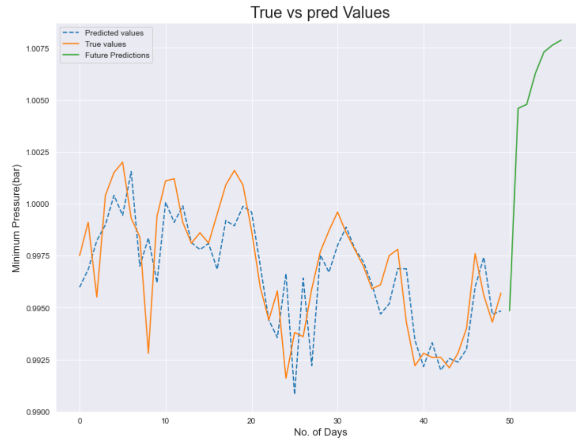
Effect of Model

Training and Validation on Shenzhen Dataset

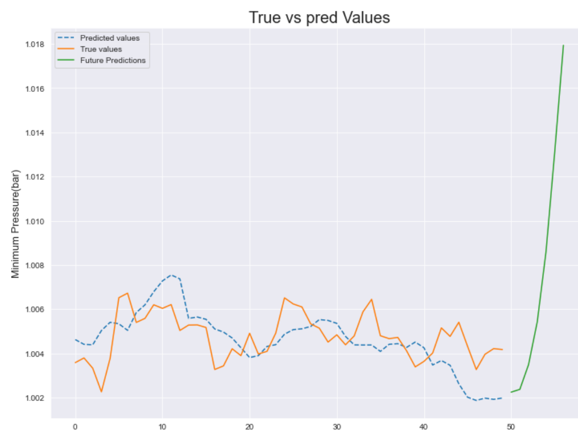


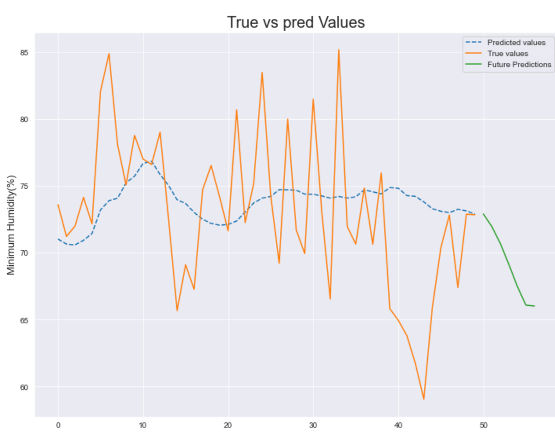






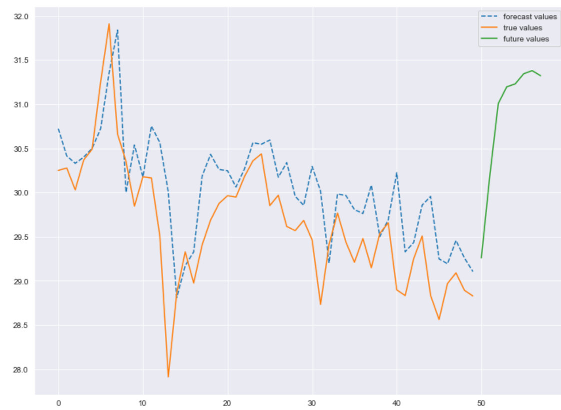
Training and Validation on Trivandrum Dataset





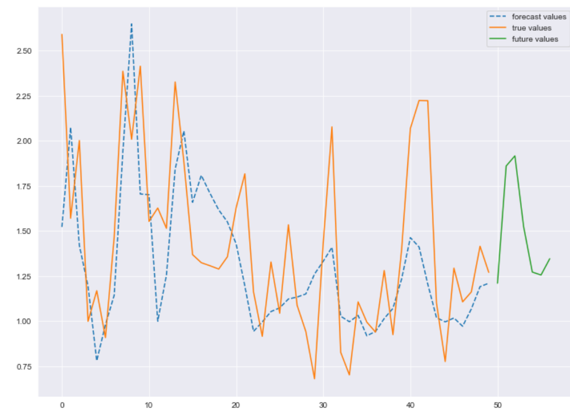
Validating data using Shenzhen

Average Temperature



No. of days

Wind Speed

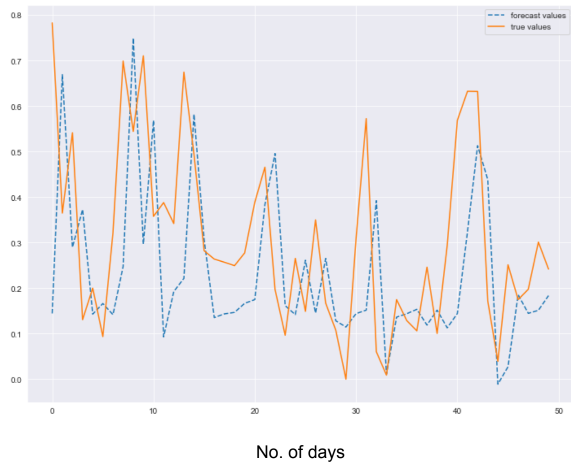


No. of days

Name of Feature	mse	mae
Average Temperature	0.0032	0.04145
Wind Speed	0.1722	3.2276

Validation using Palakkad Dataset

Wind Speed



On palakkad dataset the mae and mse are 1.4233 and 0.05109 respectively.

For shenzhen dataset

	Feature name	Mean Squared Error	Mean absolute error
1	Average Pressure	0.00150	0.0258
2	Minimum Humidity	0.0037234	0.05005
3	Minimum Pressure	0.00155619	0.0309285
4	Average Temperature	0.002058	0.03558

For trivandrum dataset

	Feature name	Mean Squared Error	Mean absolute error
1	Average Temperature	0.004145	0.003115
2	Minimum Pressure	6.575e-06	0.002112
3	Minimum Humidity	0.00853	0.06974

Conclusions

LSTM and EMD learning neural network fusion model for enhancing the accuracy of weather forecasts. The advantages of EMD in the decomposition of non-stationary data with seasonal trends are fully utilized by the method of filtering the correlation coefficients of the components of each variable decomposed by EMD and then recombining the data into an LSTM network, which also minimizes the impact of data noise and seasonal fluctuations.

If there is enough data, we can train the model immediately by using the best IMF and then training it with more characteristics to determine the future forecast for the following days. Otherwise, we can validate the model with data that will produce good results and train it on a similar dataset.

References

https://www.mdpi.com/2073-4433/13/8/1208

https://www.google.co.in/books/edition/Long_Short_term_Memory_Networks_with_Pyt/ONpdswEACAAJ?hl=en

https://www.researchgate.net/publication/221534245_Empirical_Mode_Decomposition_-_an_introduction

https://towardsdatascience.com/decomposing-signal-using-empirical-mode-decomposition-algorithm-explanation-for-dummy-93a93304c541