Title: Comparison of Deep Learning and Classical Regression Approaches for Multivariate and Multi Step Time Series Forecasting

Author:
Ayesha Shafique

Email: aieshashafique@gmail.com

Abstract
Ozone (O3), one of the most important air quality and climate change pollutants, has a negative impact on human health, the climate, and vegetation. Medical research shows that polluted air containing Cl2 gas damages the ozone layer in large amounts. It directly affects the increased number of diseases, especially skin cancer; therefore, predicting the concentration of surface ozone is very important for the protection of human health and the environment. The forecasting of air pollution data for O3 in the real world time series is challenging because it has multiple input variables. This paper presented three approaches for multivariate and multi-step time series forecasting, Vector Auto Regression (VAR), Multilayer Perceptron (MLP) and Long Short Term Memory (LSTM) to analyze air pollution data in multivariate time series. These forecast models not only take the current data as their input but also what they previously recognized in time to generate new O3 forecasts. We examined the performance of the proposed models and observed improvements of 10 % to 20 % in forecast evaluation matrices, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) using LSTM in comparison with the VAR and MLP models in the O3 layer prediction.

Keywords—Multivariate time series, Multi-step time series, Long short term memory (LSTM), Vector Auto Regression (VAR), and Multilayer Perceptron (MLP).

1. Introduction
Ozone (O3) is an incredible oxidant. This layer absorbs 93-99% of the sun's high frequency ultraviolet light, which is potentially damaging to life on earth [1]. The people most in jeopardy from breathing air containing ozone include people with asthma, children, older adults and alfresco workers. In integration, people with certain genetic characteristics and people with reduced intake of certain nutrients, such as vitamins C and E, are more liable to be exposed to ozone. Impacts on populations residing in areas where ozone levels are high for longer periods are more difficult to detect and are still questionable.

Statistical models still exist for the prediction of air quality based on meteorological data [2]. However, these models have some limitations and were mainly limited to the simple use of standard classification or regression models that neglected the nature of the problem itself or ignored the correlation between sub models in different time slots [3-6]. Various machine learning techniques have been used to predict the time series forecasting.

A number of classical time series forecasting methods have been used to predict the future forecast. Widely used methods include Autoregression (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving-Average (SARIMA), Seasonal Autoregressive Integrated Moving-Average with
Exogenous Regressors (SARIMAX), Simple Exponential Smoothing (SES), Holt Winter’s Exponential Smoothing (HWES).

Many solutions were proposed to predict the future value of the O3 layer in the coming years, but these classical models were not so progressive as to be able to work on multivariate data to determine the level of the O3 layer that depends on other multiple environmental air quality factors [7].

In this study, we presented three approaches for multivariate and multi-step time series forecasting, Vector Auto Regression (VAR), Multi-layer perceptron (MLP) and Long Short Term Memory (LSTM) and compared their performance to analyze air pollution data in multivariate time series [13-18]. These forecast models not only incorporate the current data, but also what they have previously recognized in time to generate new O3 forecasts.

2. Material and Methodology

2.1 Data Source

The presented methodology was tested on air quality data from the open source machine learning repository of the UCI [8]. The dataset contains 9358 instances of average hourly responses from a range of 5 metal oxide chemical sensors embedded in an air quality chemical multi-sensor. The device was in a highly polluted area on the ground, on the road, in an Italian city. Data from March 2004 to February 2005(one year) representing the longest available records of the response of chemical sensor devices deployed in the field.

2.2 Data Preparation

In a data set of 9358 instances, missing values were found in some records, which were replaced by the mean attribute value. The challenge in pre-processing was to turn the multivariate time domain problem into a supervised learning problem where dependent and independent variables can be targeted. We had to perform the multivariate, multi time step forecasting, so that future forecasting could not only be done on current input but also on what occurred at previous state; that is the concept of multi-step forecasting. Since our goal was to predict the Ozone O3 level, we have defined O3 explicitly as our target variable [9-12].

The data is divided into a training and test cohort with a ratio of 60:40 respectively to train the model using the training data set and validate it with the test data set. Unlike normal machine learning models, this split is carried out in time without the instances being shuffled.

2.3. Forecasting Machine Learning Models

2.3.1. Vector Auto Regression (VAR)

It is a multivariate linear time series model design to capture the join dynamic of multiple time series. For forecasting purpose reduced form VAR’s is sufficient. VAR model is a multi-equation system where all the variables are treated as dependent. The vector auto regression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model. The VAR model is useful for describing the dynamic behavior of multivariate time series and for forecasting. The superior forecasts to those from univariate
time series models and elaborate theory-based simultaneous equations models can be provided by using VAR models. Forecasting is quite flexible since they can be made conditional on the potential future paths of specified variables in the model [25].

2.3.2. Multi-Layer Perceptron (MLP)

A Multi-layer perceptron (MLP) is a deep artificial neural network. It is composed of more than one perceptron. They are consist of input layer to receive signal, an output layer that makes a decision to predict about the input and in between those two arbitrary number of hidden layer. If we have multiple hidden layer with nonlinear activation function then it gives better prediction. The perceptron consist of weight which is the summation processor and an activation function. The input values are presented to perceptron if the output which we predicted as same as the desire output then the performance is good and no changes to the weight are made. However if the output does not fulfilled the output which we desire. Then the weight need to be change to reduction of error.

2.3.3. Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specific recurrent neural network (RNN) architecture that was designed to model temporal sequences and their long-range dependencies more accurately than conventional RNNs. The LSTM contains special units called memory blocks in the recurrent hidden layer [23]. LSTM have chain like structure. From forget gate operation we take input from current time step and previous time step and concatenate them. We pass this value through a sigmoid function. Which give output between 0 and 1 through update gate operation we sum the value from current time step and also the previous time step. Then pass this value through a tanh function. We produce candidate value and by passing it through a sigmoid function. We choose values to be selected from candidate. From output gate operation we summed up the value from current time step and also from previous time step and pass it through a sigmoid function. To choose which value we use as output, we take the cell state and applying a tanh function which let only selected output.

3. Experiment and Discussion

The forecast models used in this study to deal with multivariate and multi-step time series are Vector Auto Regression (VAR), Multi-layer perceptron (MLP) and Long Short Term Memory (LSTM). Training of the forecasting model is carried out on the data set of training [19-21]. Once our models have been trained, we evaluated these models further with unseen test data to validate our proposed models.

The comparison graphs between prediction and actual data on training and testing data reveals that how the forecasting models predicted the time series vs the actual time series on train and test cohort respectively.
As shown above, the results of forecasting using VAR are not reliable and satisfactory for multivariate and multi-step time series forecasting to predict Ozone (O3) in our case.
Fig. 2. (a) Shows the comparison between actual and predicted forecasting on train cohort using MLP, (b) presents the comparison between actual and predicted forecasting on test cohort using MLP.
Fig. 3. (a) Shows the comparison between actual and predicted forecasting on train cohort using LSTM, (b) presents the comparison between actual and predicted forecasting on test cohort using LSTM.

The comparison graphs between actual and predicted forecasting seems much better using MLP (Fig.2) and LSTM (Fig.3) models in comparison with VAR (Fig.1) to predict Ozone (O3). It is clear from the comparison graph that deep neural networks performed better on multivariate and multi-step time series to better manage multiple parameters of the time series forecasting with large window of dependencies to predict the future value of targeted variable, in our case, Ozone (O3).
4. Results

The performance metrics used to evaluate the quality of forecast time series models are Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The results of the forecast metrics for predicting the O3 layer level are shown in Table 3 and Table 4. On the training dataset, it was seen that the forecasting model generated by Long Short Term Memory (LSTM) and Multi-layer perceptron (MLP) had the lowest errors rate, MAPE (< 14%) in contrast to Vector Auto regression (VAR) which had MAPE (>33%). However, when these trained multivariate and multi-stage forecast models were validated using test data to predict the O3, the LSTM [22-23] forecast model outperformed to predict the future value of O3 based on its previous input.

Table 1: Comparison of forecasting models on training data

<table>
<thead>
<tr>
<th>Evaluation Matrices</th>
<th>VAR</th>
<th>MLP</th>
<th>LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>271.622</td>
<td>116.355</td>
<td>107.717</td>
</tr>
<tr>
<td>MAPE</td>
<td>33.202</td>
<td>12.715</td>
<td>11.266</td>
</tr>
<tr>
<td>RMSE</td>
<td>342.448</td>
<td>158.612</td>
<td>150.343</td>
</tr>
</tbody>
</table>

Table 2: Comparison of forecasting models on testing data

<table>
<thead>
<tr>
<th>Evaluation Matrices</th>
<th>VAR</th>
<th>MLP</th>
<th>LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>346.21</td>
<td>116.16</td>
<td>116.01</td>
</tr>
<tr>
<td>MAPE</td>
<td>36.27</td>
<td>11.733</td>
<td>11.223</td>
</tr>
<tr>
<td>RMSE</td>
<td>448.03</td>
<td>159.79</td>
<td>159.39</td>
</tr>
</tbody>
</table>

In conclusion, the LSTM model has the lowest error rate (Fig.4, 5 and 6) in our case to predict the O3, because the LSTM model is more context-based and capable of learning long-term dependencies than using MLP, which is a key forecast.
Figure 4: Comparison of MAPE

Figure 5: Comparison of RMSE

Figure 6: Comparison of MAE
Conclusions

In this study, we investigated recent advances in the prediction of ozone levels (O3) using multivariate and multi-step time series forecast models. We compared the VAR, MLP and LSTM models and as a result, the performance matrices (MAE, MAPE, and RMSE) using LSTM are 10%-20% more accurate than VAR and MLP models for prediction of Ozone (O3) future values based on previous N sequential, measurement record. We observed that the LSTM model is better able to learn long-term dependencies than to use MLP for Ozone (O3). This proposed approach is therefore used not only to forecast the multivariate time series, but can also handle multiple step time series with less amount of error ingesting real time inputs and generating future prediction for O3 layer.

Conflict of Interest: The author declares that they have no conflict of interest.

References


23. Hasım Sak, Andrew Senior, Franoise Beaufays. Long Short Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. Google, USA
