Name: Spoorthi Kondagadapu

Course: Big Data and Business Analytics[Capstone]

Assignment: 5

Date: Sunday, Feb 19

Apache Zeppelin:

Apache Zeppelin is a multipurpose web based notebook which brings the features like

Data exploration and Data visualization to Hadoop and Spark.

The components of the workflow are:

1.Data discovery

2.Data exploration

3. Data reporting

4. Data visualization

Algorithm:

First we download the Zeppelin from Apache.org

Then we install the pip package, seaborn package using the following commands in the terminal:

Sudo easy-install pip

Sudo pip install seaborn

Then we install the wordcloud package using the command

Sudo pip install wordcloud

We use the following command in the terminal to start Zeppelin:

bin/zeppelin-daemon.sh start

The next step is downloading the data from the appropriate link.

Then we create a new notebook

Then we import the data to Zeppelin using the following code:

// Imports

import org.apache.spark.sql.functions.\_

import org.joda.time.format.DateTimeFormat

// Load data - adjust the path to the location of your data

val inputPath =  "/Users/vsubramaniam/Downloads/pollution/\*"

val data1 = sqlContext.read

        .format("com.databricks.spark.csv")

        .option("header", "true") // Use first line of all files as header

        .option("delimiter", ",")

        .option("inferSchema", "true") // Automatically infer data types

        .load(inputPath)

data1.toDF().registerTempTable("data1")

Then we execute sql queries like:

%sql  
select DayOfYear, count(\*) as NrOfFlights, avg(DepDelay) as AvgDepDelay, avg(ArrDelay) as AvgArrDelay from air\_traffic group by DayOfYear having NrOfFlights > 100000

My idea is to implement the various algorithms to build a model to predict the weather more

accurately.

Theory:

Earlier the dynamic prediction was used, but now dynamic and statistical methods are used for

Weather forecasting.

The predicted variables and the preditors are the observed quantities of the Model.

Step wise regression is the technique used in selecting the predictors, used in the meteorology.

Continous as well as binary variables are used in estimating the probabilities in the equations.

We need a separate equation , if the minimum Root , mean square error of the wind is desired.

In the forecasting , the surface weather variables, the model output statistics technique is used.

The atmospheric boundary layer plays an important role in weather forecasting.

TheAir mass transformation is studied in the Weather forecasting.

The output of the air mass transformation is useful for air pollution and agricultural problems.

The air mass transformation model is used as a forecast tool. There are two types of forecast good forecast and bad forecast. Good forecast is one in which the conditions are similar to observed conditions and bad forecast is one in which the beneficial outcomes are not produced.

There are three types of goodness , which are consistency, quality , value.

Forecasters derive their conclusions from the knowledge base which contain the weather

Conditions.

The current statements about the weather forecast are called Judgements and about the

future are called forecasts.

The inherent element of uncertainity changes from event to event. The requisite forecast is one in

In which the users find the complete data to make good decisions.

The requisite forecast is synonymously called as forecast.

The forecasters base their judgement on the information contained in the knowledge base, however the complete information may not be required.

The forecasters judgement always corresponds to the requisite forecast.

The ultimate goal of the forecasting is to provide the best judgement to satisfy the users.

A forecast can contain less spatial specificity than the judgement. The forecasters judgement

Is supposed to predict whether there is a likelihood of the precipitation, if the value is 0.2, according to the knowledge base and the real value obtained is also 0.2, then we say that the consistency exists.

The forecasters control the type 1 goodness which is consistency. If the forecasts correspond to judgements, type 1 goodness is satisfied. This makes it different from the other two types of

goodness.

It is the totality of the information , which contributes to the quality of the forecast.

The other aspects of quality are referred to as discrimination.

The sharpness of the forecasts correspond to the marginal distribution.

The relative forecasting is known by comparing the two levels of accuracy. The forecasts are

assessed by the information provided to the user.

By achieving the three types of goodness, we serve the users interest as well as forecasters interest.

Ignoring the uncertainity element affects the quality and value. In the distributed-oriented approach,

It is possible to reduce the dimensionality and complexity.

No weather forecasting system can acquire hundred percent accuracy. Accuracy and skills are the two factors on which the weather forecasting depends upon.

The conditional mean dynamics reveal the time series modeling. In the weather derivates context,

The weather forecasting methods are useful.

If the prespecified weather events occur, then the weather derivates are a new type of security.

If the prespecified weather events occur, then the weather derivatesare a new type of security.

Precipitation[Rainf all , Snow], maximum temperature and humidity are the different kinds of weather

related variables.

Quantity hedging is determined by the weather forecasting, but price hedging is not.

The weather related changes in terms of quantities are protected by the weather derivative products.

Weather is a location specific commodity unlike crude oil. In the weather derivatives market, demand and supply sides are affected by weather forecasting.

Energy companies, utilities and insurance companies are affected by the weather driven poor performance.

The very large weather changes can pose very little risk, if they are predicted well.

Energy companies, utilities and insurance companies are affected by the weather driven poor performance.

The very large weather changes can pose very little risk ,if they are predicted well.

The physical models of atmospheric conditions produce the short horizon forecasts.

National climate Data Center collects data from all the measurement stations.

The unconditional temperature densities are shown by the seasonality in daily average temperature.

The two benefits of the fourier series are smooth seasonal pattern and parsimony.

Models like time series trend, Seasonal and cyclical components are greatly studied by the meteorologists to achieve good forecasts.

The trend in the daily average temperature is seen in most of the cities. The large increase in the

Trend is due to the air pollution.

Auto regressive forecasts are the type of forecasts which involve seasonal , trend, cycle model.

The no change forecast is the first bench mark forecast which is known as persistence forecast.

The two component seasonal and trend model is the second benchmark model which uses

dummy variables to study the daily seasonal effects.

The third benchmark forecast is produced by the Earthsat in real time which is highly sophisticated

one. The third type of forecast is achieved by the numeric weather prediction model.

The weather derivatives market depends on weather modeling and forecasting for both the demand and supply sides. To eliminate the weather derivatives we have to see how much weather noise exists.

To forecast daily average temperature in four US cities, the nonstructural time series approach works out well.

In the future weather forecasting ,multivariate analysis and cross hedging will be used. Weather

Surprises translate into earning and then translate to stock return movements.

The artificial neural networks has advantages over the linear combination. The data about the different

attribute are used to train and test different models. Accurate forecasts were produced by the artificial

neural networks.

Weather forecasting is a challenging task since weather is a data intensive and continuous process.

Artificial Neural networks help in solving many nonlinear problems of weather forecasting.

The process of Weather forecasting is done well by the artificial Neural Networks.

In The process of Neural networks , the different neural networks are trained and their predictions are collected. The weighted average of each network is the output of an ensemble.

The final decisions are based on the certainity of each classifier. The different results are produced

By the different networks. From the view point of bias, the neural networks bring about greater accuracy.

The final decision is obtained by the certainity of each classifier. Artificial Neural networks does the training and testing of data in its models.

To minimize the Mean square error, the initial weights are trained.

The lower error and a reasonable learning time are the characteristics of the optimal network.

The different data patterns result in the different values of temperature, humidity and wind speed.

The Neural networks are employed for hourly weather forecasting.

There is a challenge today to study the climate change and minimize the energy savings.

The room temperature is kept within the range to build climate control heating. The different

actuators keep the room temperature in the comfortable range. The Actuators settings are adjusted to minimize the energy costs.

For simulation and control, a bilinear model is used. The actuators are in each system. They are employ

ed for the following subsystems such as heating, cooling, ventilation.

The non linear problems are satisfied by the sequential linear programming.

Performance bound is used for perfect weather and its predictions.

There is an urgent need for reliable warning systems as the severity of floods increases.

The part of the uncertainities are predicted by the EPS forecasts.

In an urban atmosphere ,the real time forecasting is caused by the neural network.

The success index forecasting is 78% with the neural classifier. In the real world a model called Neurozone is being used. The training data is updated every year. The air quality agencies perform tasks like monitor pollutants and inform authorities and public. The first air quality model

Is a combination of chemistry, transport and dispersion.

The second model connects the level of pollutants directly to the meteorological conditions.

For the real time forecasting , the second approach is preferred. The daily ozone concentrations are affected by the daily emissions. The variability of the Ozone levels is best explained by the daily weather variations. The photochemical smog is produced in the summer where high temperature

And low humidity prevail. The air quality agencies use the available variables in the statistical regression function to predict the weather.

The regression models and Ozone modeling are studied in depth.

The autocorrelation of the Ozone variability , the seasonality and trend are modeled by the time series

Analysis.

The cart method is associated with the regression analysis. The ozone prediction is tested by the

non linear techniques. The neural network with a neural classifier and the Ozone modeling are

used to improve the prediction.

In the summer the ozone reaches peak values. The weather forecasts deliver the meteorological data.

The neural model is based on the multilayer perceptron network.

Computers are used to find the best parameters to avoid the time consuming process.

The pruning technique is used to avoid overtraining, which has poor performance on fresh data.

Overtraining occurs when the noisy details of the training data are learnt. To compute the

Confidence interval of the predicted values in the training data, a theory was proposed.

The particular observation on the fitted regression is a measure of leverage.

To optimize the parameters of the regression , the training data set is used. The performance of the regression is tested by the testing data set. The difference between the two sets is not there

At the 95% confidence level from the Anova test.

By using the neural model, there is significant improvement of the scores. The data by the National weather service is used in the meteorological forecasts. By using the neural model, there is significant improvement of the scores. The data by the National weather service was used in the meteorological

Forecasts. At the end of the year, validation data is used as training data and retrained to obtain new regression parameters.

In order to simulate , better non-linear problems in numerical weather predictions,

The non-linear functions are used. The windspeed and air temperature are the two essential parameters by which the numerical weather predictions are done.

The small flow characteristics can be characterized by increasing the model resolution.

The higher resolution increases the forecast skill.

The seasonal changes and the local effects are obtained from the Model output statistics.

The parameters air temperature and the wind speed are used in the study of the different order polynomials.

**References:**

1. Oldewurtel, F., Parisio, A., Jones, C. N., Morari, M., Gyalistras, D., Gwerder, M., ... & Wirth, K. (2010, June). Energy efficient building climate control using stochastic model predictive control and weather predictions. In *American control conference (ACC), 2010* (pp. 5100-5105). IEEE.
2. Bureau, U. W. (1959). Evaporation maps for the United States. *Tech. Paper*, *37*, 13.
3. Uno, I., Carmichael, G. R., Streets, D. G., Tang, Y., Yienger, J. J., Satake, S., ... & Matsumoto, K. (2003). Regional chemical weather forecasting system CFORS: Model descriptions and analysis of surface observations at Japanese island stations during the ACE‐Asia experiment. *Journal of Geophysical Research: Atmospheres*, *108*(D23).
4. George, J. J. (2014). *Weather forecasting for aeronautics*. Academic press.
5. Niziol, T. A., Snyder, W. R., & Waldstreicher, J. S. (1995). Winter weather forecasting throughout the eastern United States. Part IV: Lake effect snow. *Weather and Forecasting*, *10*(1), 61-77.
6. Kallos, G., Nickovic, S., Papadopoulos, A., Jovic, D., Kakaliagou, O., Misirlis, N., ... & Anadranistakis, E. (1997, October). The regional weather forecasting system SKIRON: An overview. In *Proceedings of the symposium on regional weather prediction on parallel computer environments* (Vol. 15, p. 17).
7. Godske, C. L., Bergeron, T., Bjerknes, J., & Bundgaard, R. C. (1957). *Dynamic meteorology and weather forecasting* (Vol. 605). Boston, Mass: American Meteorological Society.
8. Campbell, S. D., & Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, *100*(469), 6-16.
9. Gneiting, T., & Raftery, A. E. (2005). Weather forecasting with ensemble methods. *Science*, *310*(5746), 248-249.
10. Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and forecasting*, *8*(2), 281-293.
11. Holtslag, A. A. M., De Bruijn, E. I. F., & Pan, H. L. (1990). A high resolution air mass transformation model for short-range weather forecasting. *Monthly Weather Review*, *118*(8), 1561-1575.
12. Maqsood, I., Khan, M. R., & Abraham, A. (2004). An ensemble of neural networks for weather forecasting. *Neural Computing & Applications*, *13*(2), 112-122.
13. Shi, J., Lee, W. J., Liu, Y., Yang, Y., & Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, *48*(3), 1064-1069.
14. Shi, J., Lee, W. J., Liu, Y., Yang, Y., & Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, *48*(3), 1064-1069.
15. Monin, A. S. (1972). Weather forecasting as a problem in physics.
16. Hawking, S. W. (2014). Information preservation and weather forecasting for black holes. *arXiv preprint arXiv:1401.5761*.
17. Ratcliffe, R. A. S., & Murray, R. (1970). New lag associations between North Atlantic sea temperature and European pressure applied to long‐range weather forecasting. *Quarterly Journal of the Royal Meteorological Society*, *96*(408), 226-246.
18. Klein, W. H., & Glahn, H. R. (1974). Forecasting local weather by means of model output statistics. *Bulletin of the American Meteorological Society*, *55*(10), 1217-1227.
19. Bader, M. J., Forbes, G. S., Grant, J. R., Lilley, R. B. E., & Waters, A. J. (1995). *Images of weather forecasting: A practical guide for interpreting satellite and radar imagery*. Cambridge University Press.
20. Menzel, W. P., Holt, F. C., Schmit, T. J., Aune, R. M., Schreiner, A. J., Wade, G. S., & Gray, D. G. (1998). Application of GOES-8/9 soundings to weather forecasting and nowcasting. *Bulletin of the American Meteorological Society*, *79*(10), 2059-2077.