DS 670 – LAB 10

Aakash Parwani
April 06, 2017

Comparison between competitor's algorithm and my algorithm

Competitor	My Algorithm
• This paper emphasizes importance of the average daily air temperature's long-range forecast. Further, daily values may be used as very good basis for week, month, and season forecasts.	• My algorithm is using multi linear regression model to better understand the relationship between the independent and the dependent variables.
• In this paper, long-range forecasting average daily air temperature using inductive method was proposed.	• Important thing in the proposed weather analysis is first considering error free prediction of wind speed and then performs forecasting of daily air temperature. This operation took less than 3 seconds using zeppelin.
• Data from 1 January 1973 to 20 April 2013 represent air temperature adequately because 35-year period is within this interval.	• **Loading and cleaning** the data. This operation took less than 1 minute. This operation was run on Zeppelin.
• The principle of high-impact weather events substantiates the different places interaction by atmosphere, hydrosphere, landmass, biosphere, etc.	• Used pandas, numpy, scipy to ensure optimized **data transformation and analysis** on big weather dataset to get a clear picture of relationship between variables. This operation took less than 1 minute in zeppelin.
• The forecasting method used has **significant errors** at long-range period (more than 20% in the mean).	• My algorithm has used short range forecasting method. That has **error** less than 10% in the mean.
• With the algorithm average **waiting time** is approximately 15 minutes.	• With Spark interpreter on Zeppelin IDE. **Process time** is approximately 4 minutes.
• The algorithm considers precipitation has close relation to air temperature and vice versa.	• Performed **pearson correlation** using zeppelin and found **p-value** for temperature and wind speed is than less than **0.05**. That is a strong evidence for good relation.
• **Pearson product-moment** correlation coefficient is greater than 0.8 and less than 0.85 in absolute value.	• **Pearson product-moment** correlation coefficient calculated using Zeppelin IDE is greater than 0.85 (chances for improvement).
• The model designed is combination of air temperature time series analysis	• Here, multi linear regression analysis is performed. That considers relations

for different places. So the **accuracy** depends on the number of places into consideration.	between variables like: Temperature, Wind speed, Wind direction, precipitation etc. That improves the **accuracy** of model to 85 percent.
• Only air temperature and precipitation is used in the long range forecasting that reduces the **reliability level**.	• All the variables (air temperature, precipitation, windspeed, pressure, dew point etc.) are considered for short range forecasting that set's **reliability level** of this model to 85 percent.
• Recent research based on the inductive methods showed possibility of the long-range (half-year lead-time) forecast with **mean absolute error** (MAE) up to 8^0 F.	• **Mean absolute error (MAE)** of my algorithm is 5.7^0 F. And there are chances of improvement.

References

Dmytro, Z. (2012). Improving Long-Term Weather Global Using Cloud Computing. *Developing Cloud Computing's Novel Computational Methods* , 17.

Pedersen, L. (2007). NETWORK ARCHITECTURE FOR SMALL X-BAND WEATHER RADARS. *TEST BED FOR AUTOMATIC INTER-CALIBRATION AND NOWCASTING* , 7.

Zubov, D. (2013). Average Daily Air Temperature's Long-Range Forecast. *Average Daily Air Temperature's Using Sattelite Data* , 7.