Analyzing the Impact of Meteorological Parameters on Rainfall Prediction

Muhammad Salman Pathan*, Jiantao Wu*, Yee Hui Lee†, Jianzhuo Yan‡, and Soumyabrata Dev*

* ADAPT SFI Research Centre, School of Computer Science, University College Dublin, Ireland
† School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore
‡ Ministry of Education Faculty of Information Technology, Beijing University of Technology, China

Paper #2026
Presentator: Dr. Muhammad Salman Pathan

This research has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 801522, by Science Foundation Ireland and co-funded by the European Regional Development Fund through the ADAPT Centre for Digital Content Technology grant number 13/RC/2106_P2
Contents

- Rainfall Prediction and its Importance
- Dataset
- Correlation Analysis
- Feature Importance
Rainfall Prediction and its Importance

- Rainfall is a climatic event which has an impact on wide range of activities like power generation, construction etc.
- Rainfall prediction is one of the most important tasks as it has a significant importance in different sectors mainly agriculture.
- An accurate and timely prediction of rainfall could save a lot of efforts and resources during critical events.
- The field of artificial intelligence has established a potential space in the field of meteorology.
- Despite the availability of advanced technology and enough weather data, the prediction of rainfall event is extremely complex.
- In order to develop efficient rainfall prediction systems, firstly a thorough study of various meteorological parameters should be made in the context of rainfall occurrence.
Dataset

• The rainfall data is collected from National Oceanic and Atmospheric Administration Climate Data Online service (CDO\(^1\)).

• The downloaded data is daily rainfall data from Jan 2015-Dec 2020, recorded from the weather station located in Alpena Regional Airport, Michigan, U.S.

• The meteorological parameters utilized are precipitation (PRCP), average wind speed (AWND), direction of fastest 2-minute wind (WDF2), direction of fastest 5-second wind (WDF5), fastest 2-minute wind speed (WSF2), fastest 5-second wind speed (WSF5), average temperature (TAVG), minimum temperature (TMIN), maximum temperature (TMAX).

• Feature PRCP is rainfall measured in inches (in inches) using a rain gauge, and we consider this feature as the rainfall indicator.

1: https://www.ncdc.noaa.gov/cdo-web/
## Dataset

### Daily Summaries Station Details

| STATION DETAILS              |                  |
|-----------------------------|-----------------|
| Name                        | ALPENA CO REGIONAL AIRPORT, MI US |
| Network:ID                  | GHCND:USW00094849 |
| Latitude/Longitude          | 45.0716°, -83.5644° |
| Elevation                   | 208.5 m          |

### PERIOD OF RECORD

|                          |                  |
|--------------------------|-----------------|
| Start Date\(^1\)         | 1916-10-21      |
| End Date\(^1\)           | 2021-10-27      |
| Data Coverage\(^2\)      | 99%             |

\(^1\) Data availability starts from 1916-10-21, ends on 2021-10-27.

\(^2\) Data coverage is 99%. 

![Map of Alpena area]
Correlation Analysis

• The correlation values among the features were computed by calculating the correlation coefficients of feature matrix $X$, having dimension $m \times n$, denoted as: $X = [v_1, v_2, ..., v_n]$, where $v_1, v_2, ..., v_n$ are the vectors of $n$ number of meteorological features. Each vector is of length $m$ indicating a weather recording at a particular time.

• Looking at figure 1, there is a significant correlation between PRCP, and the features related to data type wind (i.e., WSF5, WSF2, WDF5, WDF2, AWND). The reason of the positive correlation is because winds carry an amount of moisture in it which can highly affect the amount of precipitation.

• Furthermore, a negative correlation exists between rainfall and temperature (i.e., as temperature increases, rainfall drops). The amount of precipitation gets lower with higher temperatures and vice versa. Hence, a strong negative correlation occurs on land, as temperature favor more dry conditions.

Fig. 1: The correlation values of the different meteorological parameters and precipitation.
Feature Importance

We have used least absolute shrinkage and selection operator (LASSO) technique which is an effective technique for feature selection. Lasso regression typically leads to a sparse solution in the feature space, meaning that the regression coefficients of redundant or the irrelevant features are shrunk to zero.

\[
L = \sum_j(y_j - \sum_k \beta_k v_{jk}) + \lambda \sum_k \| \beta_k \|_1
\]

Eq. 1

here \( v_{jk} \) denotes the \( kth \) meteorological feature in the \( jth \) datum, \( y_j \) is the response feature value in this datum and \( \beta_k \) represents the regression coefficient of the \( kth \) feature. Because of \( \ell_1 \) function \( \sum_k \| \beta_k \|_1 \), the LASSO regression typically produces estimates in which some of the feature coefficients are set exactly to zero thereby performing feature selection.
Feature Importance

Figure 2 shows the feature importance values of each meteorological variable in terms of precipitation prediction. We can observe that features $TMIN$, $WSF2$, $WSF5$ have the highest values. Whereas $TMAX$ and $AWND$ posses the lowest scores meaning that they have a low influence in detecting the rainfall. Looking at the importance values, we can observe a similarity with the correlation results listed above i.e features related with wind has a significant influence in the prediction of the $PRCP$ whereas the features related to temperature doesn’t have an impact on rainfall prediction task.
Thank you