A New Trend on Weather Forecasting Prediction Using Big Data Sits Application

[1]Adhithyan R, [2]Abhiram Babu, [3]Joseph George

[1,2,3]Student
[1]Santhigiri College of Computer Science, Vazhithala, Thodupuzha, India

Abstract: Nowadays weather forecasting influences the daily lives of people extensively, by evaluating the present conditions of the atmosphere. Big data analytics is the procedure of analysing the massive data to uncover, trends, patterns and correlations of data in large amount to produce a more valuable outcome. A huge amount of meteorological data is being gathered from various locations on each hour to evaluate the present status of the atmosphere. The analysis of the data using conventional methods has become exceedingly resource-intensive. Weather analysis and forecasting by evaluating such huge data is a tough task. So big data technology is the perfect choice to do this complicated mission.

Keywords: Weather Forecasting,Bigdata,Weather Predition

I. INTRODUCTION

The first weather predictions were made in the nineteenth century. Weather forecasting is the study of atmospheric information such as temperature, radiation, air pressure, wind speed, wind direction, humidity, and rainfall. There must be a large amount of data generated or collected in order to anticipate the weather. These data are also not well arranged. Originally a labor-intensive human effort, weather forecasting has evolved into a computational process, necessitating high-tech machinery. The delicacy of protrusions can be impacted by a number of effects. Some of these important variables are season, location, input data accuracy, weather classifications, lead time, and validity period.When it comes to weather forecasting, big data can facilitate better decision-making. We may employ big data analytics in weather forecasting to meet this objective because it is crucial to make exact predictions.Historically, forecasting relied heavily on human forecasters; however, in the information era, forecasting is now conducted using technology and data. Rainfall, humidity, air pressure, radiation, sun intensity, data gathering, etc. are all included in atmospheric data sets. Additionally, we want a significant quantity of datasets obtained from other sources (big data).This volume of data must be processed using sophisticated technology and software [1-2].

II. WEATHER FORECASTING

Weather is the name for the everyday variations in the atmosphere.The current state of the weather can be seen in the weath er data, such as air pressure, temperature, wind speed and direction, humidity, etc., collected in different meteorological stat ions from sea observation, land observation, radar observation, and other types of observation.The process of finding trends in these data using various applications and algorithms is known as weather forecasting. Weather forecasting was arised in the nineteenth century, but operations have advanced significantly during the 1970s and 1980s by use of Numerical Weather Prediction (NWP). The mentioned models are essentially composed of mathematical equations. These equations are solved by the computer software in order to track changes in meteorological attribute values. It can be finished either day by day or a few days ahead of time. The ability to predict the weather is a crucial process that affects many people's daily lives and may have an impact on industries like agriculture, irrigation, and marine trade. It can also prevent many fatal accidents. Weather forecasting has multiple uses, and it has an impact on many aspects of daily life, including business, agriculture, disaster relief, and energy management[3-4].

A. INDUSTRY

An industry is a group of businesses that create certain commodities for consumers to buy and offer desired services. The primary source of revenue for a business or organisation operating within that industry determines how that industry is classified.

B . AGRICULTURE/ FOOD

Undoubtedly, there is a greater need for food production as the population grows. Weather forecasting as well as big data analytics in this area should be used to improve the quality of agricultural productions. Farmers can be mindful of soil erosion, overwatering, and drought with the aid of weather forecasts. Farmers can more precisely plan their crops and determine the cost of food by predicting rainfall[5].

III. BIG DATA

The amount of data is increasing swiftly in comparison to computer technology, and there are many different definitions of big data. According to one of these theories, using new tools, analytics, and technical architectures is necessary for big data in order to create high-value sources for organisations and to unlock large amounts of hidden information through analytics. Big data is characterised by five key characteristics: variety, variability, volume, velocity, and value;

- Variety includes different data types and forms. Data from a variety of sources, including sensor data, web logs, web pages, e-mail, documents, and social networking sites, is completely diversified[6].
- Volume demonstrates the size and scope of the data. The current data is measured in petabytes, and it is anticipated that this will change to zettabytes in the future.
- Velocity is related to the rate at which data is created, changed, or transferred from one location to another. This feature has to do with both the data stream's and the data entry's speed.

- Variability considers the data stream's inconsistency. Maintaining the data that people install onto their devices is quite difficult, particularly with the growth of social media.
- Data is useless and invalid by itself; value is created when it is turned into priceless knowledge[7].

IV. WEATHER BIG DATA ANALAYTIC FACTORS

Weather conditions are always subject to modest variations, some of which may be related to variations during the previous seven days or so.Variation here refers to the difference between the parameter from yesterday and the parameter today. Additionally, there is a correlation between the current week's weather and weather conditions from earlier years. In this paper, a method is put forth for modelling these two types of dependencies mathematically and using them to forecast weather in the future. This study will consider the conditions that prevailed in the previous week, or in the preceding seven days, which are supposed to be known, in order to predict the weather conditions for the day.Additionally, the weather from the previous year's seven prior days and seven prospective days is taken into account. For instance, if it is necessary to forecast the weather for January 16, 2023, we will evaluate the conditions from January 9, 2023, through January 15, 2023, as well as the conditions from January 9, 2022, through January 22, 2022. To make the prediction, the window with the best match is chosen. The current year's weekly variations are combined with the chosen window to anticipate the weather. Since the weather in a year may not be exactly the same as it may have been in prior years, sliding window matching is used to account for this possibility.

A.SLIDING WINDOW ALGORITHM

To read the rainfall for a full day is the purpose of the design. This is determined by taking into account both the once-everytwo-week rainfall circumstances as well as the rainfall over the previous seven days. Suppose we have to read.

S. No.	Max temp.	Min temp.	Humidity	Rainfall
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

$\rightarrow W_1$ $\rightarrow W_2$

Fig1.In this figure where $W1$ stands for Window number 1 and $W2$ stands for window number 2 .

We will analyse the weather from January 9 through January 15 in 2023 as well as the weather that prevailed from January 9 through January 22years before determining the weather on January 16. The current year's daily variation is computed after that. In this study, the maximum temperature, minimum temperature, humidity, and rainfall will all be taken into account as important meteorological characteristics. As a result, a matrix of dimension 7*4 will be used to show how large the fluctuation of the current year is. The matrix size would be 14*4 for the previous year, too. The 14*4 matrix must now be divided into sliding windows as the first phase. In light of this, 8 sliding windows.Thus, 8 sliding windows with a 7*4 size each can be constructed. Figure 1 illustrates the idea of a sliding window.

The following step is to compare each window to the variation for the current year.

The prediction is made using the window with the best match. For the purpose of matching, Euclidean distance method is applied.

1.Mean: The average of the day's weather parameters, including the highest and lowest temperatures, the relative humidity, and the amount of precipitation. Divide the total number of days after adding each separately.

$$\text{Mean} = \frac{\text{Sum of parameter}}{\text{number of days}},$$

2. Variation: Determine daily variation after comparing each parameter's values. This information explains how the weather of one day will compare to that of another.

This allows us to formally model the dependencies that have been previously defined. The mathematical definition of the relationship between the previous year's data and the prior week's data can be used to forecast the future. we use some algorithm for this.

Alogrithm

- Step 1: Create a matrix "CD" with the most recent seven days' worth of data for the current year, measuring 7*4 in size.
- Step 2: Create a matrix "PD" with a size of 14 by 4 days for data from the preceding year.
- Step 3: Create 8 sliding windows, each measuring 7*4 inches, using the matrix "PD" as W1, W2, W3,...,W8.
- Step 4: Using the matrix "CD" as ED1, ED2, ED3,..., ED8, calculate the Euclidean distance of each sliding window.
- Step 5: Determine the matrix Wi by choosing Wi = Correponding Matrix (Min.(EDi)) I [1, 8]. \
- Step 6 :for k = 1 to n.
- Calculate the variation vector (VC) for the matrix "CD" of dimension 61 for WCk.
- Calculate the variation vector for the matrix "PD" of dimension 61 as "VP" for WCk.
- Mean1 = Mean (VC), Mean2 = Mean (VP), etc. (Mean1 + Mean2)/2 is the predicted variation .
- To get the anticipated condition, multiply the weather from the previous day by "V."

- Step 7:Finish.

Algorithm 1 displays the sliding window used to forecast the "n" number of weather situations (WC1, WC2, WC3,..., WCn).
The primary justification for employing the sliding window strategy is because it's possible that the weather at a particular time of day during the year wasn't the same time of day the year before. For instance, the weather during the first week of February in 2010 may not have been present during the same period in 2009. Although not necessarily during the same week, identical weather conditions may have existed in years past on some days. The fortnight considered has the highest chance of having identical weather conditions [8].

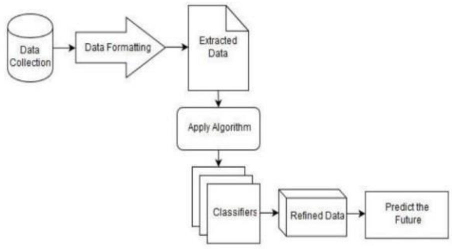

Fig2.Data Flow

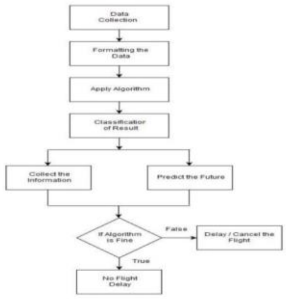

Fig3.Data Flow Diagram [9]

V. OPEN ISSUES OF BIG DATA ANALYTICS IN WEATHER FORECASTING

- Multi-objective optimization
- Optimization algorithm.
- Implementation challenge.
- Trust and data cleaning.
- Data mining algorithm for MapReduce solution.
- Scalability.
- Integration and accuracy.

VI. WEATHER BIG DATA ANALYTIC FACTORS

To compare existing techniques and solutions withthose that are suggested, each weather big data analysis methodology has specificevaluation criteria thatinclude weather and QoS (Quality of Service) variables. Following are definitions of the crucialQoSandweather variables.

- Precision, accuracy, andmemory, We require QoSfactors in order to assess a big data approach, compare it to earlier methods to find a more efficient algorithm, and also to find out what it's good and bad for. Some of these factors cited in different papers include precision, accuracy, and recall. The keyQoSconsiderations for big data analytics in weather forecasting are illustrated in the following.

Recall, accuracy, and precision are defined. The percentage of correct predictions is called accuracy.
Precision is the percentage of correctly predicted favourable outcomes. Recall is the proportion of events that are projected to be positive.
Accuracy =(TP + TN) /(TP + TN + FP +FN).
Precision is equal to TP /(TP + FP)
Recall equals TP/(TP + FN)
Additional parameters used in articles for evaluation and comparison include execution time (second), response time (second), scalability, reliability (a number between 0 and 1 that indicates the overall consistency of a measure), temperature (Celsius), wind (metres per second), precipitation (percent), humidity (percent), rainfall (millimetres), and pressure (Pascal) [10-11].

VII. APPLICATIONS

- Weather-related businesses can employ big data analytics to apply to weather data to enhance their business intelligence.
- Accurate and nearly real-time analytics may be possible with big data and predictive modelling.
- Stakeholders can quickly process large volumes of data using big data and predictive analytics technology, and they can do so with high degree of accuracy.
- Ahead of time, the public is informed of the day's weather conditions.
- It gives the company important data that it can utilise to decide on its long-term business plans.
- Uses ways most businesses prefer to deliver forecasts in a visual format.

• Aids in the purchase and sale of cattle by agricultural groups; • Aids in the establishment of pastures, crops, and water sources for the farming business.

CONCLUSION

Big data analytics for weather forecasting were provided in this paper along with a thorough literature analysis. First, the idea of big data was discussed, after which the idea of weather forecasting and the significance of big data in this day and age were examined. Big data is much more than simply a trendy term and isn't just hype. Today's weather departments are learning that in order to compete successfully in the market, they not only need to manage escalating massive data volumes in their real-time systems, but also evaluate that data. On the basis of a vast amount of data outliers, big data analytics on meteorological data assists in drawing the appropriate conclusion. Consequently, this will assist in developing an appropriate business model that is employed for As a result, it will be easier to develop a sound business model, which is utilised to make intelligent business decisions. It also shows them where they can improve, which will lead to improved prediction techniques. With the aid of big data technologies, we improved the effectiveness of processing a lot of weather data. The main use of these systems is the forecasting of the weather, though. There are a lot more forecasts for weather features to investigate. in the interest of our upcoming work.

REFERENCES

1. Progress in climate prediction and weather forecast operations in China, according to Xiao Z, Liu B, Liu H, and Zhang D (2012).
2. Adv. Atmos. Sci. 1980's The Weather Forecast by L. BengtssonGeophys Pure Appl 119(3):515-537.
3. L. Kan, L. Yu-Shu (2005) a fuzzy neural network approach for predicting the weather based on rough sets. Pages. 1888– 1892 in Volume 3 of the 2005 International Conference on Machine Learning and Cybernetics.
4. L. Kan, L. Yu-Shu (2002) Forecasting the weather using fuzzy case-based reasoning. pp. 107–110 in International Conference on Machine Learning and Cybernetics Proceedings, Volume 1.
5. The role of climate forecasts in smallholder agriculture: insights from participatory research in two Senegalese communities, Roudier Petal, 2014. Clim Risk Mana 2:42.
6. Big data: concerns, difficulties, tools, and best practises. Katal A, Wazid M, Goudar RH (2013). Sixth international conference on modern computing, 2013, (IC3).pp 404–409.
7. Pathak AR, Rautaray S, Pandey M (2018) building the huge data using taxonomies, analytics, and methods. 237–259 in Iran J ComputSci 1(4). 99
8. https://www.hindawi.com/journals/isrn.
9. Images through Google.
10. MapReduce and its uses, difficulties, and architecture: a thorough analysis and recommendations for further study, Khezr SN, Navimipour NJ 2017. 15(3):295-321 J Grid Comput. Hong S, Lee J-H, and Lee J (2014) a genetic algorithm-based prediction of heavy rain based on huge meteorological data. The 8th international conference on ubiquitous information management and communication proceedings, Siem Reap, Cambodia.