A Stochastic-based Path Prediction of Wind Storms

[1]Tharun Sai Vaddi, [2]Kalyan D, [3]Chaitanya U, [4]Dr. D. Vijaya Lakshmi

[1,2] UG Student, [3] Assistant Professor, [4] Professor
Dept. of Information Technology
MGIT, Hyderabad, India.

Abstract- **The objective of this paper is to analyze and apply the best clustering algorithm to the path prediction system of wind storms. Wind storms cause much damage to both human life and also human properties and hence predicting before it comes will be a great advantage to evacuate the place as early as possible and go to a safer place. Hence to build this we need a path prediction system. To build this system, as there will be a huge number of previous datasets of recorded windstorms, we need a clustering algorithm which needs to give accurate output. So, for this purpose, the analysis is done on both the agglomerative clustering algorithm and the K-means clustering algorithm, and the best one will later be applied to the Markov chain (MC) rule and finally will predict the accurate path. Here the path of the wind storm is predicted given the genesis point, which is the latitude and longitude of the starting point of the windstorm. Hence, applying the best clustering algorithm gives an accurate path and is critical in reducing economic loss and saving human lives.**

Keywords- **Agglomerative clustering, data pre-processing, k-means clustering, markov chain model, path prediction, stochastic model, wind storm.**

1. INTRODUCTION

The world has witnessed an escalation of devastating wind storms over the last three decades. Every year, the time between June 1 and November 30 signifies the North Atlantic Storm season. During this period, the warm waters in the Atlantic give birth to tropical cyclones, and a few of these tropical cyclones end up making landfall causing major casualties and loss of property. Thousands of people suffer every year due to the wind storms which occur mainly due to the Atlantic Ocean. Here, The prediction of the windstorm path is done using both the agglomerative clustering algorithm and the Kmeans clustering algorithm, and then after applying to the Markov chain rule the results are compared with the actual path traversed by the storm and after the comparison, the algorithm which gives the accurate path is applied.

By using the Atlantic Windstorm Database, The dataset provides location details Latitude & Longitude of every storm at every 3-hrs interval from the point of genesis to decay. Short-term forecasting of the windstorm path is relatively straightforward. A long-term prediction model is built that can predict the path of the windstorm a few days or weeks in advance. With the help of historical data, the model should be able to route the entire path of the storm from the starting genesis point to the decay point of the storm. This model should also locate the places that will be most affected by the storm in order to take all the precautionary measures needed. Accurate frequency, intensity, and landfall location forecasting are crucial for reducing the risk of these expensive disasters. This will be helpful for every individual to know whether that particular wind storm might cause damage to them or not. If yes then they can take precautionary measures before itself and can protect themselves and reduce their damage loss.

2 RELATED WORK

Taylor S. Cox, et al. proposed an Accurate Model for Hurricane Trajectory Prediction. The Weighted Asset, Realistic-Distance Storm Trajectory Predictor (WARD-HTP) is a new model that aims to increase the precision of current data mining methods for hurricane trajectory prediction. In this case, the haversine distance formula, weighted training data, and training data from 1950 to 2000 were all used. WARD-HTP, which is the result of the three contributions, has a hurricane trajectory forecast correctness ratio of 75.0 percent, which is 10% higher than the existing benchmark.

Automatically Locate Tropical Cyclone Centers Using Top Cloud Motion Data Derived FromGeostationary Satellite Images model is proposed by Gang Zheng, et al. Here, The new data triggered our development of a new method to automatically and accurately determine TC centers based on cloud concentric motion, The methodology used here are Top Cloud Motion Data And Motion Field Decomposition.

A new data mining model for hurricane intensity prediction is proposed by Y. Su, S. Chelluboina, M. Hahsler, & M. H. Dunham where the weights of the features are learned by a genetic algorithm (GA) using historical hurricane data. This describes a weight feature learning EMM model for hurricane intensity prediction. This model is divided into two faces. In first phase a genetic algorithm is used and in second phase an EMM constructed and used to predict the future intensities of windstorms.

The novel method for hurricane trajectory prediction based on data mining by integrating association analysis technology and using the real American Atlantic hurricane data is proposed by X. Dong, et al. First, utilising association analysis technology, all frequent trajectories in the historical storm trajectory database are mined, and the related association rules are constructed as motion patterns. The motion patterns are then linked with the present cyclone tracks to make predictions. This experiments show that the prediction accuracy is ideal and satisfactory

Suraj Singh Chouhan et al. proposed disaster management and risk management techniques to trace the losses incurred in a controlled way. The existing research involves 3 prime contributions for prediction based technique development which being, the pre-processing technique,k-means clustering and hidden Markov model. Thereafter come the prediction and the performance evaluation of these techniques.

3 METHODOLOGY

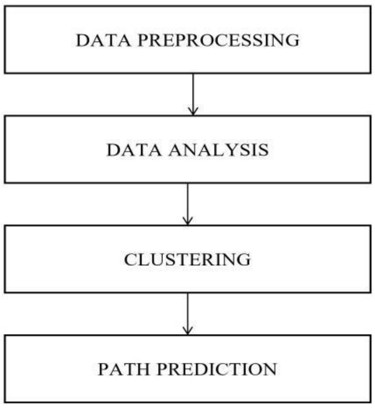

Figure 3.1. Path Prediction Architecture

3.1 DATA PREPROCESSING

A data mining approach called data preprocessing entails putting raw data into a comprehensible format. Real-world data is frequently inaccurate and lacking in specific behaviours or trends. It is also frequently inconsistent and incomplete. Preprocessing data is a tried-and-true technique for handling these problems. Real-world data are typically incomplete because they lack attribute values, lack specific attributes that are of interest, or simply contain aggregate data. Noisy: containing errors or outliers. Inconsistent: containing discrepancies in codes or names. Hence to solve these issues in the dataset, the data preprocessing is to be done in order to get an accurate output.

3.2 DATA ANALYSIS

The process of analyzing information by utilising logical and analytical reasoning to look at each part of the information presented. One of the various procedures that must be carried out when performing a research experiment is this kind of analysis. To arrive at a finding or conclusion, information from many sources is collected, examined, and then analyzed. There are a variety of specific data analysis methods, some of which include data mining, text analytics, business intelligence, and data visualizations.

In order to comprehend the issues that an organisation faces and to meaningfully analyse data, data analysis is crucial in business. Data is just a collection of facts and figures. Data analysis puts the data in context by organizing, interpreting,structuring, and presenting the data. From the dataset given the analysis can be made on the frequency of storms that have occurred in the past years and with the help of this one can try to recognize if there is any pattern in these series of storms occurring. And also identify the years in which most intensive windstorms have occurred and try to predict the damage that could be caused by that kind of storm if it occurred now or in the future.

3.3 CLUSTERING

A cluster is a collection of items from the same class. To put it another way, things that are dissimilar are grouped in a different cluster from those that are similar. The climatology (genesis points, sea surface temperature, energy, etc) determines the trajectory of hurricanes and the probability of making landfall. We will divide the hurricanes into four groups in order to take the climatology into account. The clustering analysis is made to gain some valuable insights from the data by seeing what groups the data points fall into when the clustering algorithm is applied.
So, here we have the two best clustering algorithms which are K-means clustering and Agglomerative clustering. Now, our aim is to analyze these two algorithms and apply the best to the Markov chain model.

3.3.1 Analyzing K-means clustering and Agglomerating clustering algorithms

Firstly, we see the differences between these two clusterings.
i) Agglomerative clustering is one where each data point starts in its own cluster and thereafter merges a similar pair of clusters successively resulting in a hierarchy whereas K-means uses a straightforward process to divide a given data set into a number of clusters, each of which is designated by the predetermined letter "k." The clusters are then placed as points, and all observations or data points are connected to the closest cluster, calculated, and adjusted. The procedure is then repeated using the new adjustments, and so on, until the desired outcome is obtained.
ii) The second difference is that in agglomerative it gives you visibility on the proximity of all the data points through a dendrogram. Whereas K-means uses the elbow method to find the optimal number of clusters. So, in this case, agglomerative clustering is more useful in analyzing similarities and differences between individual data points or clusters.

iii) Agglomerative clustering follows a bottom-up approach whereas K-means clustering doesn't follow any method but only the centroid is considered to form clusters.
iv) If we compare according to the size of the datasets, the execution time required for the large datasets in the K-means clustering technique is less compared to the Agglomerative clustering.
v) Finally, the agglomerative clustering gives a better and more accurate output when compared to the K-means for the path prediction of wind storms.

3.4 MARKOV CHAIN MODEL

After getting the clusters by using agglomerative clustering, now we need to apply the Markov chain model. This is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. The changes in the state of the system are called transitions. The probabilities associated with various state changes are called transition probabilities. The process is characterized by a state space, a transition matrix describing the probabilities of particular transitions, and an initial state or initial distribution across the state space. The Markov Chain model is the best model that can be used in order to predict the entire route of the storm from the start point to its decay point of it. So from the genesis point, it starts predicting the next point in the path one after the other according to the probabilities that are present in the transition matrix. The highest probability point is selected as the next point and it continues till it reaches the end. These points will be only the points that are present in the clusters, which are generated using the agglomerative clustering algorithm.

Finally, we get the required predicted path of wind storms after applying Data preprocessing, Data analysis, Agglomerative clustering and Markov Chain model and the results are shown in the next session.

4 RESULTS

Firstly the data preprocessing is done and the below figure 4.1 is the datset before preprocessing.

Out[2]:

	ID	Name	Date	Time	Status	Latitude	Longitude	Maximum Wind	Minimum Pressure
44827	AL072006	FLORENCE	20060913	1200	EX	45.5N	55.6W	70	967
22949	AL041952	BAKER	19520907	1800	HU	40.4N	60.1W	95	969
18184	AL031936	UNNAMED	19360626	1800	TS	26.2N	95.5W	40	-999
23583	AL031954	ALICE	19540625	1400	HU	25.0N	97.6W	95	-999
32598	AL101977	DOROTHY	19770928	1200	HU	35.5N	59.7W	65	988

Figure 4.1. Dataset before Preprocessing

Now after Data Preprocessing the foolowing figure 4.2 is result where two individual columns named time and date are merged into one column which reduces some space.

:[6]:

	ID	Name	Status	Latitude	Longitude	Maximum Wind	Minimum Pressure	TIME
49100	AL122015	KATE	EX	41.3N	50.4W	55	981.0	2015-11-12 12:00:00
49101	AL122015	KATE	EX	41.9N	49.9W	55	983.0	2015-11-12 18:00:00
49102	AL122015	KATE	EX	41.5N	49.2W	50	985.0	2015-11-13 00:00:00
49103	AL122015	KATE	EX	40.8N	47.5W	45	985.0	2015-11-13 06:00:00
49104	AL122015	KATE	EX	40.7N	45.4W	45	987.0	2015-11-13 12:00:00

Figure 4.2. Dataset after Preprocessing

After preprocessing the dataset, now the data is analyzed and the below figure 4.3 is the frequency of storm occurrence in a year after the data analysis.

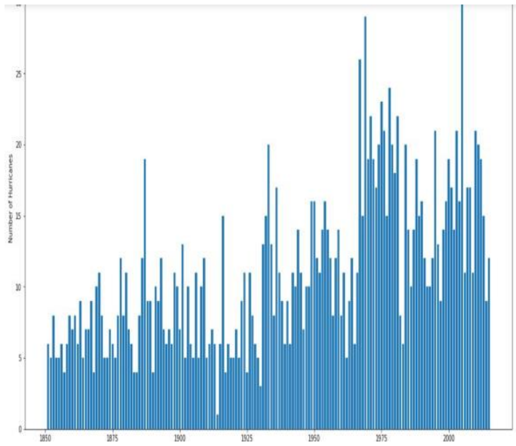

Figure 4.3. Frequency of storm occurence in a year

Same as the above, the frequency of storm occurence in a month is shown infigure 4.4.

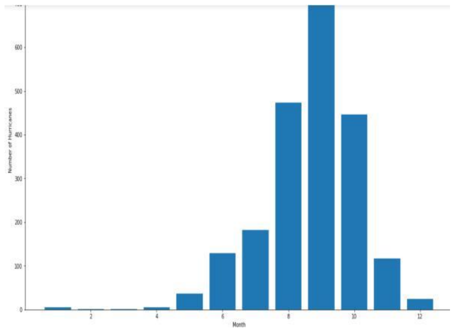

Figure 4.4. Frequency of storm occurrence in a month

And also the frequency of various kinds of storms are also shown in below figure 4.5

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x18cb4bd6128>
```

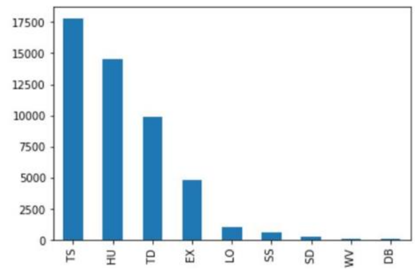

Figure 4.5. Frequency of various kinds of storms

Input to the Path Prediction System is shown in below figure 4.6 where we give the latitude and longitude of the genesis point of wind storm.

```
In [43]: predict_path(([-40], [15]))
```

Figure 4.6. Input to the path prediction system

Output of the Path Prediction System using Agglomerative clustering is shown in below figure 4.7 where the path is indicated with the blue line on the map.

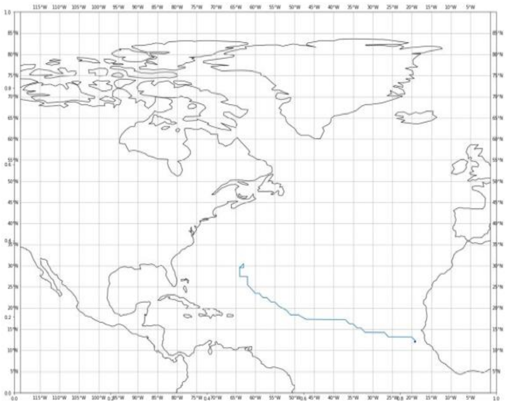

Figure 4.7. Predicted Path using Agglomerative clustering

Output of the Path Prediction System using K-means clustering is shown in below figure 4.8 where the path is indicated with the blue line on the map.

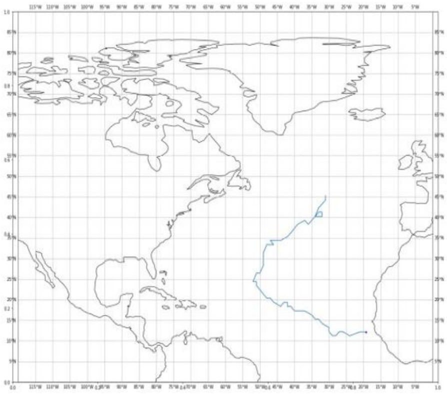

Figure 4.8. Predicted Path using K-means clustering

Trajectory of real wind storm is shown in the figure 4.9 to compare the two clustering results with the real path and consider the accurate one.

Figure 4.9. Path of real wind storm.

5 CONCLUSION AND FUTURE SCOPE

Prediction of trajectory path of wind storm is a model which helps in predicting the path or route of a wind storm. Firstly, Data preprocessing takes place where it is a data mining technique that involves transforming raw data into an understandable format. After that the Data Analysis process takes place. In this step Data from various sources is gathered, reviewed, and then analyzed to form some sort of finding or conclusion. Now as the Agglomerative Clustering gives accurate output than the K-means clustering we can apply this clustering algorithm to the system. Finally, we use the Markov model for tracking a real storm and determine the threat that it poses to a particular geographical location. As we successfully analyzed the two clustering algorithms and applied the best one, so we can now deploy this model to a web page and by developing the web page with a good user interface, it can be accessible to the world where people themselves can see the predicted path of the wind storms and hence this deployment will be a great advancement.

REFERENCES:

[1] Cox, Taylor S., et al. "An accurate model for hurricane trajectory prediction." 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). Vol. 2. IEEE, 2018.
[2] Zheng, Gang, et al. "Automatically locate tropical cyclone centers using top cloud motion data derived from geostationary satellite images." IEEE Transactions on Geoscience and Remote Sensing 57.12 (2019): 10175-10190.
[3] Su, Yu, et al. "A new data mining model for hurricane intensity prediction." 2010 IEEE International Conference on Data Mining Workshops. IEEE, 2010.
[4] Dong, X., and D. C. Pi. "Novel method for hurricane trajectory prediction based on data mining." Natural Hazards and Earth System Sciences 13.12 (2013): 3211-3220.
[5] Chouhan, Suraj Singh, and Ravi Khatri. "Data mining based technique for natural event prediction and disaster management." *International Journal of Computer Applications* 139.14 (2016).
[6] Jaiswal, Shweta, Atish Mishra, and Praveen Bhanodia. "Grid host load prediction using gridsim simulation and hidden markov model." *International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001: 2008 Certified Journal, Volume 4, Issue 7* (2014).
[7] For gathering information about wind storms https://www.almanac.com/content/hurricaneforecast-facts-and-common-questions
[8] For understanding the concept of synoptic times and to get a clearer idea of why is thatan important measure in measuring the time of occurrence of a particular wind storm https://nsidc.org/cryosphere/glossary/term/synoptic-hour
[9]For using Agglomerative Clustering for the model for prediction of the route - https://scikitlearn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html
[10] Using Keras, Preprocessing data, Model Creation – Machine learning Mastery - machinelearningmastery.com