An analysis of Data mining TC-CMELPBC Technique in Weather Prediction

Namrata Kumari[1]*, Dr. S. M. Asif Ali[2]

[1] Research Scholar, Dept. of Computer Science & IT, Magadh University, Bodh Gaya, Bihar, India

Email- namratakrishnasinha@gmail.com

[2] Assit. Prof. & Head, Dept. of Physics Mirza Ghalib College, Gaya, Bihar India

Email- asif_ali_gaya@yahoo.com

Abstract: The study was analysed "An analysis of Data mining TC-CMELPBC technique in weather prediction." A random sample 1000-10,000 prediction reports on whether Atlantic hurricane Database has been served as subjects. The hypothesis was that whether the data mining TC-CMELPBC technique in prediction time using Atlantic hurricane Database has more efficient on weather forecasting. Results shows that experiment results of prediction time for existing the TC-CMELPBC technique. For the simulation setup, the number of data is considered in the range of 1000 to 10000. All three methods reduce the prediction time for detecting weather data. In big data, weather forecasting gives significant information about future weather. The changes in weather conditions are effectively identified by clustering and classification of similar data. Some of the existing research works developed for performing weather prediction while considering the large volume of data.

Keywords: Data mining techniques, Climate change, Weather prediction and Weather forecasting.

- X -

INTRODUCTION

Weather acts as a significant factor in human life. Many daily works and business is based on weather circumstances. Besides, an unpredicted weather condition causes a vast life and property. If people are able to predict the weather conditions for the future, then these losses can be avoided or restricted. The weather of earth keeps on changing every time resulting in various seasons such as summer, winter, spring, autumn, monsoon, etc. This weather transform is a normal and normal phenomenon of earth. The earth's climate is influenced and altered through natural causes. The data mining procedure is extracting predictive information from enormous databases. It is an influential technology for assisting associations to focus on significant information in data warehouses. Data mining tools forecast future trends and actions to produce proactive and knowledge-driven decisions. In weather forecasting, data mining techniques are employed for producing weather reports by plotting the acquired data sets of preceding weather data sets to forecast future weather circumstances. This prediction is applied for all purposes and business applications depended on the weather.

By gathering the data about the present state of the atmosphere, weather forecasts are achieved. It also employs scientific recognition of atmospheric processes to plan how the atmosphere changes.

Weather prediction is based on chronological time series data. The essential data mining functions and numerical methods are applied to obtain a significant pattern from a huge volume of data set. Various testing and training circumstances are carried out to obtain the exact prediction result. Feature selection is performed for identifying the minimal set of significant features so that the classification error is reduced. An optimal subset of features is to be selected based on certain conditions which helps in measuring the goodness of feature subsets. In several forecasting works, clustering and classification are employed as efficient methods. The weather report for a specific month is collected and based on prior warning of the occurrence of rainfall and changes in temperature, significant steps are taken to reduce the losses and damages caused because of bad weather.

To improve the performance of weather prediction, many conventional forecasting methods are developed. These methods helped in acquiring exact prediction results of upcoming weather data with the exploitation of feature selection, clustering, and classification techniques. Also the existing methods such as conjunct space Cluster-Based Adaptive Neuro-Fuzzy Inference System (CF-ANFIS),[1] designed k-means algorithm developed by Corporal-hybrid neural model[2] a non-linear regression technique depended on Support Vector Regression (SVR) introduced[3]. Dynamic Self-Organized

Multilayer Neural Network With Immune Algorithm (DSMIA)[4] novel clustering technique[5] sliding window algorithm,[6] temporal growth of uncertainty in ensembles of weather forecasts[7] were introduced to attain exact prediction results of upcoming weather condition at minimum time consumption. We discuss various prediction techniques using artificial neural network. In order to forecast the weather condition accurately, these existing methods considered several clustering and classification methods.[8] But, it failed in performance is due to a lack of efficient feature selection technique for selecting relevant features for attaining exact prediction results. In addition, some of the existing methods provide poor prediction accuracy and consumes more time in forecasting the weather conditions. The false-positive rate of some conventional routing methods also remains higher that is to be refined for achieving efficient forecasting results.

MOTIVATION

Weather refers to the state of air on earth at a specified place and time. The application of science and technology is significantly used for forecasting the condition of the atmosphere in future time for a specified location due to its effectiveness in human life. Weather forecasting is significant for different purposes such as aviation, shipping, fisheries, and many other unique uses like forecasts for the common public. The prediction of upcoming disasters helps in taking the required precautions and security measures. Weather warnings are considered as significant forecasts for securing life and property. Based on temperature and precipitation forecasts is crucial in agriculture and traders in commodity markets. Temperature forecasts are applied by utility companies for approximating demand over the coming days. Forecasts exploited to plan the events and live on them. The prediction of weather has to be consistent and better enough to promote the society. The consistency of a particular prediction is based on the level of recognition of the process, amount, and quality of data. When the fundamental physical process of weather is reasonably recognized and high quality of data is being gathered, it is possible for achieving exact predictions with respect to future weather changes. In order to predict the atmospheric conditions accurately, different methods have been developed. A significant scheme is required for wind speed forecasts during typhoons as under-construction structures repeatedly fail in typhoon winds. The typhoon wind speeds are essential for the construction industry. Typhoons easily cause substantial casualties and economic losses for the construction industry. As a result, an accurate warning system was required for forecasting surface wind speeds and organizing construction hazards by typhoons. A Conceptual Weather Environmental Forecasting System (CWEFS) designed a construction industry. To detect wind velocity, data-driven models were employed as forecasting approaches.[9] By structure reference load analysis, reference load on under-construction experimental task is acquired. In addition, shorter prediction windows are used to attain higher prediction accuracy. However, the efficient prediction of weather conditions is not made possible.

Big data Analysis

A big data processing framework is implemented to accumulate and process the big climate data.[10] It is capable of observing the association among climate parameters and dengue the incidence in an incessant manner. The implemented framework is designed in a Hadoop cluster environment. The number of cluster nodes in the Hadoop environment differs depended on the evaluation and optimization approach. The efficiency in accumulating big climate data and forecast association among the number of dengue cases and climate parameters were enhanced. However, the implemented framework causes high complexity while handling big climate data.

Some other discussed a representation of both issues and data linked with meteorological forecasting. A system for informed defaults is considered to enable meteorologists for producing a broad variety of effective visualizations depended on meteorological and visualization principles.[11] Then, the users are also allowed to associate multiple isocontour features. In addition, Weaver is used as an open-source tool to interactively represent weather forecasts. But, the accuracy in weather condition prediction is not improved efficiently. To sum up, the conventional feature selection based clustering and classification methods developed to forecast weather conditions issues due to poor prediction accuracy, time complexity, and false-positive rate. Therefore, it is understood that there is a need for an efficient approach to improve the process of weather forecasting which addresses the above-mentioned hitches in achieving accurate prediction results.

Weather is the condition of daily air in a specific place and covers a narrow area. Daily weather information is essential for performing various activities. Weather forecasting detecting the weather condition in specified future time. It helps in acquiring critical information about the upcoming weather conditions. The process of attaining the exact prediction of weather conditions remains a complex task because of the dynamic nature of the atmosphere. Weather condition at any instance is denoted by a number of variables. Out of those variables, only one is found to be more significant. This variable has to be chosen for performing efficient weather prediction. Therefore, efficient feature selection techniques are to be considered for attaining exact prediction results. Weather forecasting is essential for performing activities associated with weather conditions. But, weather forecasting fails due to inaccurate forecasting results of possible errors. The meteorologist faces complexities due to poor accuracy of weather analysis and its prediction. Various data mining techniques are utilized to predict meteorological parameters. In addition, the increase in the availability of climate data also inspired the

Namrata Kumari[1]*, Dr. S. M. Asif Ali[2]

significance of improving the accuracy rate to examine various patterns from massive data. Therefore, data mining is to be concerned to identify hidden patterns present in huge data. The mined information helps to recognize climate change and prediction. The conventional data mining techniques developed for weather prediction lacks inefficient feature selection and also increase the error rate and time complexity. Hence, the aim of the proposed research work is to develop Weather forecasting system which performs efficient prediction of weather condition using to attain accurate results.

OBJECTIVES

This research attempts to overcome the restrictions involved in weather foresting for predicting future atmosphere. The research focus is to predict the weather condition with higher accuracy at a minimum time by performing clustering and classification of big weather data with efficient feature selection techniques.

This work proposes techniques to predict the weather data with higher feature selection accuracy at minimum time complexity which performs more efficient than other conventional feature selection, clustering, and classification methods used in weather forecasting.

HYPOTHESIS

Whether the data mining TC-CMELPBC technique in prediction time using Atlantic hurricane Database has more efficient on weather forecasting.

PURPOSE OF THE STUDY

The proposed work involves three processes namely, feature selection, clustering, and classification. The developments in data collection and storage technologies have caused the invention of huge, high-dimensional, complicated, and heterogeneous datasets. This in turn also made the clustering and classification task more demanding. As the data dimensionality increase, the complexity of data and the availability of noninformative features also increases. Only some features have the significance of performing the task.

RESULTS

As explained in the above algorithm 5.2, the proposed TC-CMELPBC technique is employed to perform both clustering and classification to increase the weather prediction. At first clustering process is performed to increase the weather prediction accuracy. In clustering process, the number of clusters and centroids are initialized. Then the expected probability for each data and cluster centroid is determined. Subsequently, the determined expected probability is increased by applying MAP function for grouping similar data into that certain cluster. This in turns, the similar type of weather data is grouped with less time. With the

results of the clustering process, the classification process is carried out with the help of a linear program boosting technique.

Algorithm- Combinatorial of MAP Expected Clustering and Linear Program Boosting Classification

Table -1 shows the experiment results of prediction time for existing the TC-CMELPBC technique. For the simulation setup, the number of data is considered in the range of 1000 to 10000. All three methods reduce the prediction time for detecting weather data. Comparatively, the proposed TC-CMELPBC technique minimizes the time to predict the weather data than the other methods. According to the value in above the graph is plotted between numbers of data versus prediction time.

Table-1 Cell frequency of prediction time using Atlantic hurricane Database

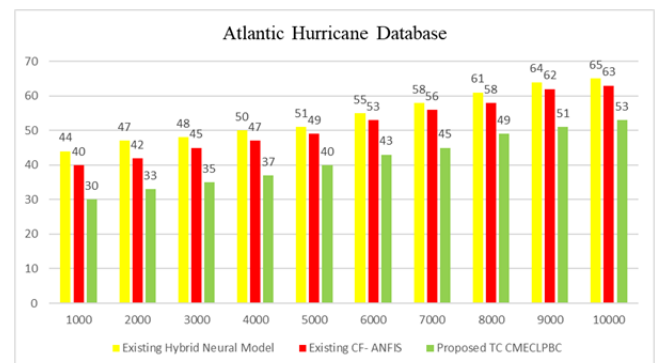| Number of Data	Prediction Time (ms)		
	Existing Hybrid Neural Model	Existing CF- ANFIS	Proposed TC CMECLPBC
1000	44	40	30
2000	47	42	33
3000	48	45	35
4000	50	47	37
5000	51	49	40
6000	55	53	43
7000	58	56	45
8000	61	58	49
9000	64	62	51
10000	65	63	53

Table-1 describes the experimental result analysis of prediction time with respect to a number of data. The performance of the proposed TC-CMELPBC technique is compared with the existing hybrid neural model. As observed in the graph, the

Namrata Kumari[1]*, Dr. S. M. Asif Ali[2]

prediction time using the proposed TC-CMELPBC technique is reduced when compared to existing methods. This is because of choosing relevant features from the weather dataset instead of processing huge features in the TC-CMELPBC technique. Feature selection is performed with the help of measuring the Tanimoto correlation coefficient. From the correlation measure, the higher similarity features are chosen to execute the weather prediction. Besides, the low similarity features are discarded. After that, the combinatorial of clustering and classification is carried out to classify the weather data with minimal time consumption. This, in turn, the time utilized to perform weather prediction is effectively reduced in the proposed TC-CMELPBC technique. Therefore, the proposed TC-CMELPBC technique minimizes the prediction time using the Atlantic hurricane database up to 20% and 24% when compared to existing hybrid neural models respectively. Similarly, the time to predict the weather data is reduced in the proposed TC-CMELPBC technique by considering the pacific hurricane database up to 20% and 25% than the state-of-the-art methods. So, the weather forecasting is the technology employed to detect the climate changes in for a particular region and time period. Due to the number of weather features and missing values of a dataset, the weather forecasting is a complex task.

CONCLUSION

In big data, weather forecasting gives significant information about future weather. The changes in weather conditions are effectively identified by clustering and classification of similar data. Some of the existing research works developed for performing weather prediction while considering the large volume of data. But the accuracy and time consumed for analysing the weather remain a major concern. Therefore, the research work designs three proposed techniques with the objective of solving above said existing issues in weather prediction by grouping and categorizing a similar type of data in an accurate manner.

In our proposed research work, the space complexity of data remained unaddressed during the clustering and classification process. This leads to increase the complexity of weather prediction as well as reduce the accuracy of the classification results. Thus, the future development of proposed work is to overcome the above-mentioned issue for increasing the performance of weather prediction.

REFERENCE

1. Duong et al., (2016), ENSO-based tropical cyclone forecasting using CF-ANFIS, *Vietnam Journal of Computer Science*, 3(2), 81–91.

2. Corporal-Lodangco et al., (2014), Cluster analysis of North Atlantic tropical cyclones, *Procedia Computer Science,* 36, 293–300.

3. Saba et al., (2017), Weather forecasting based on hybrid neural model, *Applied Water Science*, 7(7), 3869–3874.

4. Xiao et al., (2018), Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data, *Remote sensing of environment*, 210, 48–64.

5. Kapoor P. and Bedi S. S., (2013), Weather forecasting using sliding window algorithm, *ISRN Signal Processing,* 2013.

6. Kumpf et al., (2017), Visualizing confidence in cluster-based ensemble weather forecast analyses, *IEEE* transactions on visualization and computer graphics, 24(1), 109–119.

7. Ferstl et al., (2016), Time-hierarchical clustering and visualization of weather forecast ensembles, *IEEE* transactions on visualization and computer graphics, 23(1), 831–840.

8. Nayak et al., (2013), A survey on rainfall prediction using artificial neural network, International *Journal of Computer Applications*, 72(16).

9. Wei (2017), Conceptual weather environmental forecasting system for identifying potential failure of under-construction structures during typhoons, *Journal of Wind Engineering and Industrial Aerodynamics,* 168, 48–59.

Corresponding Author

Namrata Kumari*

Research Scholar, Dept. of Computer Science & IT, Magadh University, Bodh Gaya, Bihar, India

Email- namratakrishnasinha@gmail.com