ISSN 2278-3091
Volume 8, No.4, July – August 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse58842019.pdf
https://doi.org/10.30534/ijatcse/2019/58842019

Linear Program Boosting Classification with Remote Sensed Big Data for Weather Forecasting

Pooja S.B[1,*] R.V Siva Balan[2]
[1]Department of Computer Science, Noorul Islam Centre for Higher Education, Kumaracoil, 629180 Tamil Nadu, India. poojasnair100@gmail.com
[2]Department of MCA, Noorul Islam Centre for Higher Education,Kumaracoil 629180,Tamil Nadu, India. rvsivan@gmail.com

ABSTRACT

Big data is the term includes a large and complex data used for predictive analytics and other data analytics to extract the valuable information. Classification is a fundamental data mining technique for analyzing big data for weather forecasting applications. Many classification techniques have been introduced for improving the prediction performance of weather data. But the existing technique failed to get efficient results in reasonable time complexity. To improve classification Linear Program Boosting Classification (LPBC) technique is introduced. Classification is carried out using the Linear Program Boosting technique for predicting future outcomes by constructing weak learners. The boosting classifier considers the feed-forward artificial neural classifier as weak learners to categorize the input data into different classes by performing the polytomous regression analysis. Linear Program Boosting technique combines the results of all weak learners to form strong classifier for improving the prediction accuracy and minimizing false positive rate. The result observed that the proposed LPBC technique obtains high prediction in terms of classification accuracy, false positive rate, and time complexity. Based on the observations, LPBC technique is more efficient in predictive analytics than the other methods.

Key words: Big data, weather forecasting, Point-Biserial Correlation Coefficient, Linear Program Boosting, artificial neural classifier, polytomous regression analysis.

1. INTRODUCTION

In general, big data comprises structured data and unstructured data. The Structured data are relatively simple to analyze since the data exist in databases in the form of rows and columns. Whereas, the unstructured datais not a pre-defined data and it is not stored in any of the databases. Analyzing big data helps the weather predictor [1].In our big data analytics, the classification is an essential data mining technique that classifies unstructured data into the structured class. There are several applications of big data like a business, telecommunication, healthcare, weather forecasting and so on. In this work, the weather forecasting is considered with the remote sensed big data collected through remote sensing device. Weather forecasting is used to predict the atmospheric circumstances for a particular location and time. Major severe weather actions cause a considerable loss of life and property [2]. Therefore analyzing of weather data is a significant one in big data. The aim of classification is to categorize the different data with the features in the dataset. The input big data comprises many attributes or features. The classification with several features causes more time consuming and the dimensionality of the data also very high. In order to improve the weather prediction accuracy, the classification is performed with the relevant features.

A hybrid neural model was developed in [3] to obtain high weather prediction accuracy with the most important input features. The model minimizes training time and complexity but the error detection and minimizationwerenot performed to obtain high accuracy. introduced in [4] for evaluating the snow depth with the passive microwave (PM) remote data. But the classification accuracy was not improved since it failed to select the relevant features.

A conceptual weather environmental forecasting system (CWEFS) was presented in [5] for predicting the weather condition using supports vector regression and decision tree classifier. The system does not enhance the accuracy in weather condition prediction. A Probabilistic Rain Diagnostic Model was developed in [6] for identifying the majority of “affecting” cyclones. The model does not use any machine learning technique to improve cyclone prediction accuracy. A novel machine learning approach depends on fuzzy information retrieval and genetic programming was introduced in [7] for forecasting the joint wind speed and direction. The approach failed to select the features for minimizing the complexity in the prediction [8]. In-Mapper combiner based Map-Reduce algorithm was designed in [9] for processing the big climate data. The algorithm has a high complexity while accessing big climate data.

A data with meteorological forecasting problems were addressed in [10] by introducing a set of informed default encoding combined the meteorological principle with visualization practice. But, it failed to improve forecasting performance with large weather data. Artificial Neural Networks was introduced in [11] for weather forecasting using Data mining and curve fitting techniques. This technique does not minimize the error rate in the weather prediction. A multi-stage probabilistic machine learning approach was designed in [12] and estimated for predicting the monthly rainfall. The approach failed to select the relevant feature for improving the prediction accuracy. The most significant issuesidentified from the above-said methods such as lesser prediction accuracy, high error rate, lack of feature selection, high complexity and so on. In order to overcomesuch kind of issues, an efficient novel technique called Linear Program Boosting Classification (LPBC) is introduced.

The proposed LPBC technique performs classification for accurate weather prediction. Linear Program Boosting technique is applied for categorizing the weather data into different classes. This boosting machine learning technique constructs weak learners as feed-forward artificial neural network. The weak learner considers the number of weather data in the input layer. In a hidden layer, polytomousregression function is used to analyze the input with the selected features and classified the weather data into different classes and results are displayed at the output layer. The outputs of all weak learners are combined and the similar weight is assigned. Then the error is calculated between the actual and predicted output of the weak learner. The Linear Program Boosting technique classifies all the data with minimum error rate. This helps to improve the classification accuracy and minimizes the false positive rate.

The paper is organized into five different sections. In section 2, the review of related works is presented along with their capabilities and limitations are discussed. Section 3 provides an overview of the proposed LPBC technique workflowwith neat diagram. In section 4, the experimental evolution is performed with the big weather dataset. The experimental results and discussions are presented in Section 5. Finally, the paper is concluded in Section 6.

2. RELATED WORKS

Weather prediction a challenging problem around the world. The different data mining techniques were developed in [13] for classifying the weather parameter. But this technique does not work.

A Dynamic Modeling approach was developed in [14] for weather forecasting with the attributes. But the approach failed to measure the interrelationship among the weather attributes for accurate prediction [15].

A Support Vector Machine (SVM) integrate with the distance correlation was presented in [16] for predicting the geomagnetic storms. The classification technique failed to minimize the incorrect prediction. A neural network model combined with the smart grid was introduced in [17] with historical weather data. The model does not consider the variable weather conditions [18].

An optimized artificial intelligence algorithm was designed in [19] for forecasting the multistep wind speed. The algorithm does not improve the forecasting accuracy with minimum time complexity.

An online support vector algorithm (LaSVM)-based urban air pollution forecasting was presented in [20]. The algorithm does not significantly enhance the performance and reliability of the prediction. In [21], a Random Forests (RF) classification algorithm was introduced for predicting the rainfall with the textural and temporal features of clouds. Though the algorithm minimizes the error in the classification, the complexity was not minimized.

Machine learning methods namely regression and classification technique were developed in [22] for improving the drought forecast with minimum error. The classification performance was not improved for efficient weather condition prediction. The author has proposed [23]. The author has Implemented a Web based Weather Monitoring Station and Data Storage System.[24]

The major concerns in the above-said reviews are overcome by introducing a technique called Linear Program Boosting Classification (LPBC). The detailed explanation of the LPBC technique is presented in the following section.

2.1. Linear Program Boosting Classification For Big Weather Data Forecasting

Notations	definition
f_i	Features in the particular set
m_f	Mean values of the features in the particular subset
X_i	Number of weather data
Y_i	Prediction output results
d_i	Number of data
$R(d_n)$	Polytomous regression function
$P(h_n(d_i) = k)$	Predicted probability of a Polytomous regression function
$h_n(dp_i)$	Observed output of the base learner
α_l	Weight coefficient
d_i	Weather data
b_n	Bias
T_f	Transfer function
s	Learned values of data at the output layer
Y	Strong learner output
ω_i	Similar weights to all the weak learners
e	Error of each weak learner
ω'	Updated weight of the weak learner
ξ_n	Non-negative vector of the slack variable
δ_l	Margin of the classes

Weather forecastingis a complex task in the meteorological sector due to the changing behavior of the climates. In this case, accurate prediction is a major problem. Therefore, an essential technique is required for predicting the future events of the weather condition. In weather forecasting, remotely sensed data are analyzed over the different circumstances from the big datasets. The big dataset comprisesalarge volume of sensed data. The major problem of big data in the weather forecasting is the 'curse of dimensionality', and it directs to computational complexity.Classification is used for accurate prediction with minimum error. Based on this motivation, Linear Program Boosting Classification (LPBC) technique is introduced in this paper for accurate weather forecasting with less time complexity.

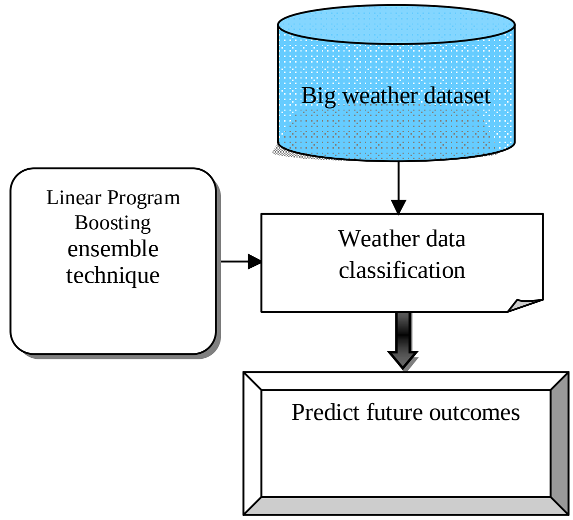

Figure 1: Flow process of LPBC technique

After taking the data from the big weather Linear Program Boosting (LPB) technique is applied for predicting future outcomes. Linear Program Boosting is a machine learning ensemble algorithm for improving the classification performances. The ensemble technique combines the several weak classifiers into one predictive model for increasing accuracy and minimizes the error when compared to the standard classifier. The flow process of LPBC technique is illustrated in figure 1.

As shown in figure 1, the flow processes of LPBC technique are illustrated with the big weather data. The big dataset contains a number of weather data and features. The features are extracted from the weather dataset. The LPBC technique weather data classification is done with the selected features using a Linear Program Boosting ensemble technique. The ensemble classifier effectively analyzes the input data with several weak learners and combined into one classifier to provide strong classification results. Here thefeed-forwardartificial neural network acts as a weak learner of the ensemble classifier. These two processes are explained in the following subsections.

2.2 Linear Program Boosting classifier for predictive analytics

The process in the LPBC technique is to perform the weather data classification with the selected features for predicting future outcomes. Classification is the process of categorizes the data into different classes for more accurate predictions and analysis.LPBC technique uses linearprogram boosting classifier from the boosting family. Linear Program Boosting Classifier is an ensemble classifier includes a set of weak learners and creates a strong classifier. A weak learner is a classifier categorizes the data with less accuracy. In contrast, linear programming is also known as a linear optimization to achieve better results by maximizing or minimizing an objective function. The linear program boosting act as a strong learner categorizes the data with two objective functions such as maximize the accuracy and minimize the error rate. Therefore, the proposed LPBC technique uses a boostingclassifier for improving the weather prediction. The boosting classifier uses a weak learner as an artificial neural network. The artificial neural network is the subdivision of Artificial Intelligence (AI). The proposed artificial neural network uses feed-forward approach and it is made up of units or nodes called artificial neurons which are interconnected together. Hence the name is called as Artificial Feed Forward Neural Network (AFFNN). The flow process of Linear Program Boosting Classifier is illustrated in figure 2.

Figure 2 illustrates a flow process of Linear Program Boosting Classifier for obtaining high classification accuracy with minimum time. Let us consider training sets (X_i, Y_i) where X_i represents a number of weather data (d_i) and Y_i represents prediction output results.

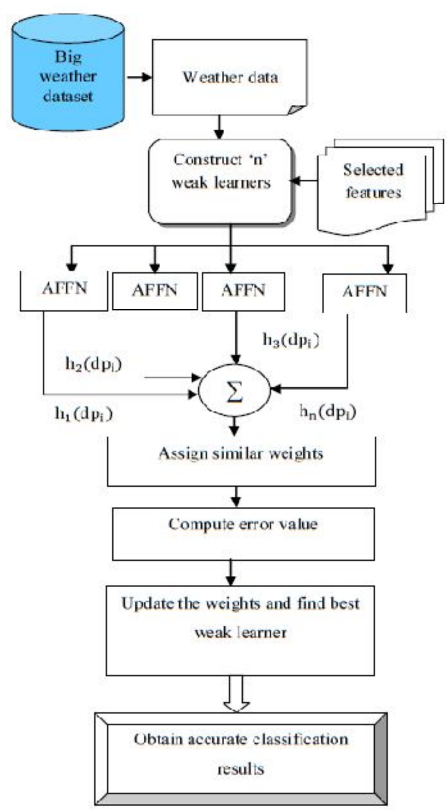

Figure 2: Flow process of the Linear Program Boosting Classifier

The classification is done with the set of weaklearners to achieve the different class labels. The base learners categorize the weather data into different classes with the selected features.

Let us taken the number of weathers data from the big dataset D^w

$$d_i = d_1, d_2, d_{3,\ldots.} d_n \in D^w \quad (1)$$

From (4),d_i denotes a number of data$d_1, d_2, d_{3,\ldots.} d_n$, D^w denotes a big weather dataset. Then,LPBC technique constructs the number of weak learners (i.e. AFFNN) to classify the given data with the selected relevant features. The process of AFFNN is described as follows.

The AFFNN is made up of the neurons (i.e. nodes) that are arranged into three different layers. The neurons in the one layer are connected withanother layer to form a network. The connection between the two nodes is differentiated by the weight coefficient. The weight coefficient reveals the degree of significance of the specified connection in the neural network.The first layer is an input layer where the input is fed into the network. The second layer is the hidden layer where the input data are learned with the selected features. The final layer is called anoutput layer that provides the classification results.

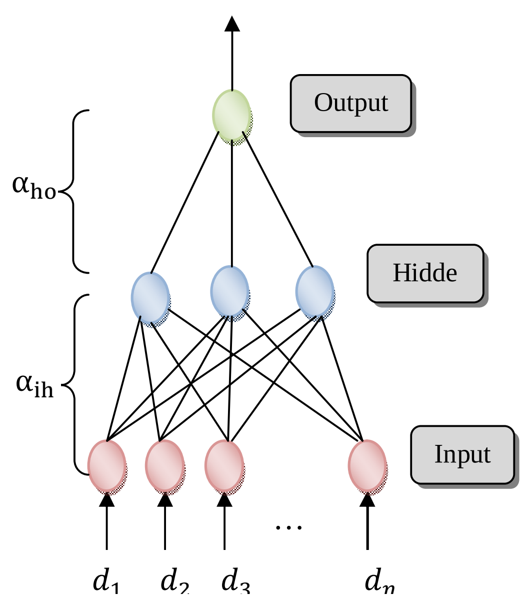

Figure 3: Structure of the AFFNN

Figure 3 shows the structure of the artificial feedforward neural network where three layers and the nodes are connected with the weight coefficient.

In figure 3, α_{ih} denotes a weight coefficient which connects the nodes in the input layer and the hidden layer. α_{ho}denotes a weight coefficient which connects the nodes in the hidden layer and output layer. In a feed-forward neural network, the information (i.e. weather data) moves one direction and it has no "feedback" from the outputs towards the inputs.

As shown in figure 3, the nodeis an artificial neuron and connection from the output of one node to the input of another nodeis formed by an arrow symbol. In figure 3, each layer in the network isfully connected to another layer in a forward manner.

The numbers of weather data are given to the input of the AFFNN classifier. The input layer does not perform any computation but it only comprises the input weather data. After receiving the input, it fed into the hidden layer. In this layer, the data and the features are analyzed by using Polytomous Regression functions. Regression analysis is broadly used for prediction and forecasting process, where it's combined with the field of machine learning. Regression analysis is performed to with the selected features.

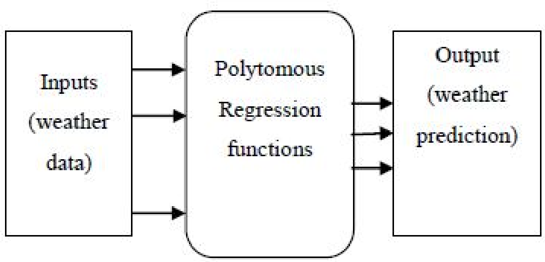

Figure 4: Structure of Polytomous regression function

Figure 4 shows the Polytomous regression function which includes a number of independent data $d_1, d_2, d_{3,....}d_n$as input and selected features provides the prediction outcomes (i.e. output). The regression analysis is performed using the following mathematical formula,

$$R(d_n) = \frac{\exp(c_j)}{\sum_{i=1}^{n} \exp(d_i)} \quad \text{Where } j = 1,2,3 \ldots k \ (2)$$

From (2),$R(d_n)$ represents a Polytomous regression function of weather data. Then, the output of Polytomous regression functions is used as a probability distribution over k different possible outcomes and the predicted probability of jth class specified as a sample input data d_i.Then the predicted probability of the Polytomous regression functions is expressed as follows,

$$P(h_n(d_i) = k) = \frac{\exp(c_j)}{\sum_{i=1}^{n} \exp(d_i)} j = 1,2,3 \ldots k.. \quad (3)$$

From (3), $P(h_n(d_i) = k)$ denotes predicted probability of a Polytomous regression function classifies the weather data into k number of classes based on feature values. The regression function used for relating the input with the outputand provides the results. The output of the hidden layer is fed into the final output layer of the AFFNN. The output of the artificial feed forward neural network is computed as follows,

$$h_n(dp_i) = T_f(\sum_{i=1}^{n} \alpha_{ho} * d_i + b_n) \quad (4)$$

From (4),$h_n(dp_i)$ denotes anobserved output of the base learner at the output layer, α_l denotes a weight coefficient, d_i denotes a weather data, b_n is the bias, T_f denotes a transfer function which is also called as activation function.The activation function performs the non-linear transformation and it able to learn the more data as well as perform more complex tasks. The proposed AFFNN uses softsign activation function at the output layer which is expressed as follows,

$$T_f = \frac{s}{1+|s|} \quad (5)$$

From (5,$'s''$indicates learned values of data at the output layer.The soft sign activation function outputs value ranged from '-1' to '+1'. The positive results '+1' show that the weather data are correctly categorized into a different class. The weak learner output has some training loss resulting in minimizes prediction accuracy. Such kinds of problems are addressed by combining all weak learners into a strong one. The strong classification output of the boosting classifier is expressed as follows,

$$Y = \sum_{i=1}^{n} h_n(d_i) \quad (6)$$

From (6),Y denotes a strong learner output after combining the results of all weak learners$h_n(d_i)$. Then the boosting classifier assigns the weight to each weak learner.

$$Y = \sum_{i=1}^{n} \omega * h_n(d_i) \quad (7)$$

From (7),ω_i denotes similar weights to all the weak learners. After assigning the similar weights, the training error of each weak learner is computed as follows,

$$e = (y_i - h_n(d_i))^2 \quad (8)$$

From (8),e represents the error of each weak learner, y_i denotes an actual output, $h_n(d_i)$denotesanobserved output of the weak learner. After computing the It means that the weight is increased then the weak learner is classified incorrectly. The weight is reduced then the weak learner correctly categorizes the data into the different classes. The updated weight of the weak learner is represented asω'.

Then the linear program boosting classifier finds the classifier with a minimum error rate for attaining the higher accuracy. The output of the final strong classifier is expressed as follows,

$$Y = \sum_{i=1}^{n} \omega' * h_n(d_i) \quad (9)$$

From (9),Y represents strong classification resultsto improve the prediction results. In order to improve the classification of all the data into any of the relevant classes, the linear program boosting classifier maximizes the margin of the different classes. The strong classification results subject into the margin,

$$\sum_{i=1}^{n} \omega' * h_n(d_i) + \xi_n \geq \delta_l \ldots.(10)$$

Where $\xi_n \geq 0$

From (10),ξ_n denotes a non-negative vector of the slack variable, δ_l represents the margin of the classes. The resulted outcomes are greater than the margin that provides all the data points categorized into particular class resulting maximizes the classification accuracy.By this way, the ensemble classifier obtains the true class labels and minimizes the incorrect class labels i.e. false positive rate. The algorithmic process of proposed LPBC technique is described as follows.

Algorithm 1 describes The classification results of the weak learners attained at the output layer. Then the boosting classifier sums the entire weak learners and distributes the same weight value. After that, the training error for each weak learner is calculated. The initial weight is adjusted based on the error value. The boosting classifier finds the weak learner with minimum error. Then it also maximizes the margin between the classes for obtaining high classification accuracy and minimizing the false positive rate.

Input: Weather big datasetD^w, number of features $f_1, f_2, f_3, \ldots, f_n$, Number of data $d_1, d_2, d_{3,\ldots.} d_n$, $c_j$denotes a different classes.

Output: Improve feature selection and classification accuracy

Begin

\\ **weather data classification**

1. **For weather data** d_i
2. Construct 'n' weak learners $h_n(d_i)$
3. Perform regression analysis $'R(d_n)'$ at hidden layer with weight α_{ih}
4. Obtain the classification output $h_n(d_i)$
5. **Classifies** d_i into different classes c_j
6. Combines all weak learners $h_n(d_i)$
7. Assign similar weights ω to $h_n(d_i)$
8. Calculate error $'e'$ for $h_n(d_i)$
9. Update the weight ω' of $h_n(d_i)$
10. Find classifier $h_n(d_i)$with $\arg\min e$
11. **if** $(\sum_{i=1}^{n} \omega_i' * h_n(d_i) + \xi_n \geq \delta_l)$**then**
12. Classify all data d_i into c_j
13. **end if**
14. Obtain strong classification results
15. **End for**

End

Algorithm 1: Linear Program Boosting Classification

3. EXPERIMENTAL SETTINGS

Experimental evaluations of proposed LPBC technique and existing methods namely hybrid neural model[3] and SVR [4] are implemented using Java language. The Atlantic hurricane database is used for predicting the cyclone with the relevant features.The database is taken from https://www.kaggle.com/noaa/hurricane-database.
The database comprises the 22 attributes.Among the 22 attributes, relevant attributes are selected for classification. The National Hurricane Center

contains data on tropical cyclones that occurred within the Atlantic Ocean and Eastern Pacific Ocean. The database comprises the 49,106 instances for the Atlantic Ocean and 26,138 instances for the eastern Pacific Ocean. Among these, tropical cyclones occurredwithin the Atlantic Oceanis considered for performing the experimental evaluation.

For experimental consideration, 1000-10000 instances are considered. Performance analysis of LPBC technique and existing methods namely hybrid neural models [1] and SVR [2] are carried out with certain parameters such as feature selection accuracy, classification accuracy, false positive rate and time complexity.While studying the impact of the cyclone prediction from weather dataset, the response variable is prediction of cyclone (Y) and the independent variables classification accuracy (x_2), false postive rate (x_3) and time complexity (x_4). $Y = f(x_1, x_2, x_3, x_4)$. For example, while considering the number features as 4, the proposed LPBC technique correctly selects the features for cyclone prediction as This result provides thatOut of 1000 data, 896 data is correctly classified and provides the classification accuracy (x_2) as 90%. Based on this, the output of response variable is varied. The results obtained from the experimental evaluation are described in the next section.

4. RESULTS AND DISCUSSION

The results and discussion of LPBC technique and hybrid neural model [3] and SVR [4] are described in this section with different parameters as feature selection accuracy, classification accuracy, false positive rate,and time complexity. With the help of these parameters, the results of proposed and existing methods are compared using graphical representations.

4.1 Impact of classification accuracy

Classification accuracy is defined as the numbers of weather data are classified correctly to the total number of data in the given dataset. The mathematical formula for classification accuracy is expressed as follows,

$$Classification\ accuracy = \frac{Number\ of\ data\ classified\ correctly}{n} * 100 \quad (15)$$

From (15) $'n'$represents the total number of data. The classification accuracy is measured in terms of percentage (%).

Sample calculation for classification accuracy:

Proposed LPBC technique: Number of data correctly classified is 896 and the total number of data is 1000. The classification accuracy is computed as follows,

$$Classification\ Accuracy = \frac{896}{1000} * 100 = 90\ \%$$

Existing hybrid neural model: Number of data correctly classified is 823 and the total number of data is 1000. The classification accuracy is computed as follows,

$$Classification\ Accuracy = \frac{823}{1000} * 100 = 82\ \%$$

Existing SVR: Number of data correctly classified is 769 and the total number of data is 1000. The classification accuracy is computed as follows,

$$Classification\ Accuracy = \frac{769}{1000} * 100 = 77\ \%$$

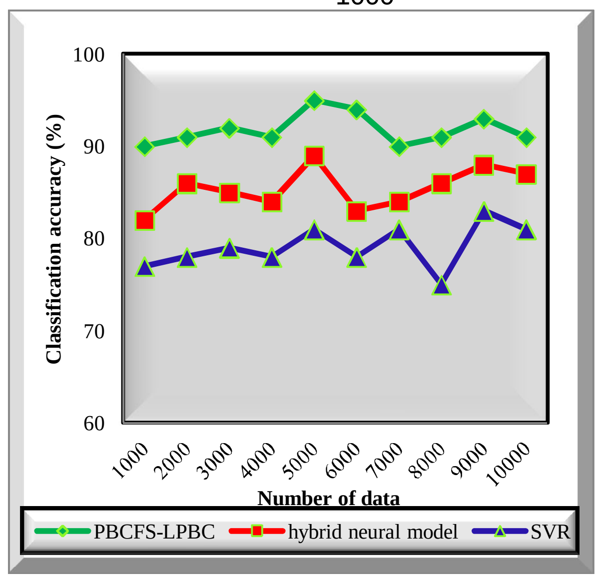

Figure 5 : Performance results of classification accuracy

Figure 5 shows the experimental results of classification accuracy based on the number of weather data. Totally ten runs are performed with a different number of weather data. For the experimental consideration, the numbers of weather data are varied from 1000 to 10000. In figure 6, the number of data taken in of LPBC technique are significantly improved when compared to the other two conventional methods. This is because the big weather data is collected from the Atlantic hurricane database. After that, cyclone data classification is performed with the selected features.

The LPBC technique uses the ensemble classification technique to predict the various types of cyclones by classifying the data. The existing hybrid neural model [1] and SVR [2] methods do not use any boosting concept to obtain high classification accuracy.

Totally 10 various runs are performed with a number of big data (i.e. weather data). For each run, the different outcomes are obtained for three different methods. The performance results of LPBC technique compared with the two conventional methods. The comparison resultsshow that the LPBC technique considerably improves the classification accuracy by 8% and 16% when compared to two state-of-the-art methods namely hybrid neural model [3] and SVR [4] respectively.

4.2 Impact of false positive rate

The False Positive Rate is computed as the number of weather data are incorrectly classified to the total number of data. The false positive rate is mathematically computed as follows,

$$false\ positive\ rate = \frac{Number\ of\ data\ incorrecly\ classified}{n} * 100 \quad (16)$$

From (16), n represents the number of data. The false positive rate is measured interms of percentage (%).

Sample calculation for false positive rate

Proposed LPBC technique: Number of data incorrectly classified is 104 and the total number of data is 1000. Then false positive rate is calculated as follows,

$$false\ positive\ rate = \frac{104}{1000} * 100 = 10.4\,\% \approx 10\,\%$$

Existing hybrid neural model: Number of data incorrectly classified is 177 and the total number of data is 1000. Then false positive rate is calculated as follows,

$$false\ positive\ rate = \frac{177}{1000} * 100 = 17.7\% \approx 18\,\%$$

ExistingSVR: Number of data incorrectly classified is 231 and the total number of data is 1000. Then false positive rate is calculated as follows,

$$false\ positive\ rate = \frac{231}{1000} * 100 = 23.1\,\% \approx 23\,\%$$

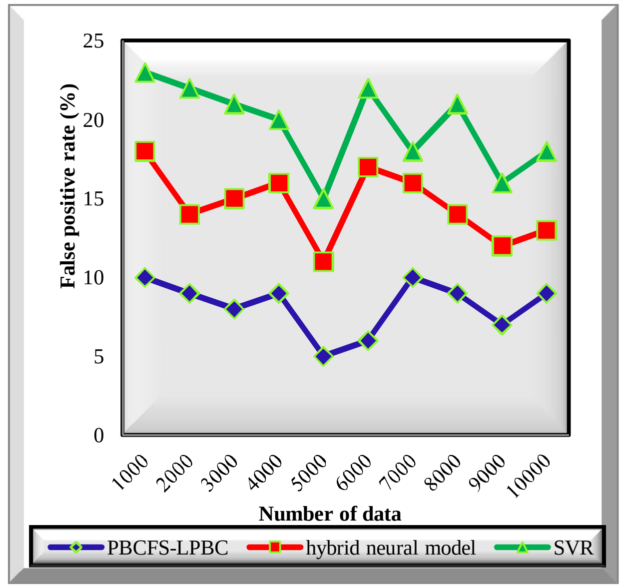

Figure 6: Performance results of false positive rate

As shown in figure 6, the experimental results of the false positive rate versus a number of data are illustrated. The above two dimensional graphical results show that the false positive rate of LPBC technique is comparatively minimized than the two conventional methods namely hybrid neural model [3] and SVR [4]. The incorrect classification of the LPBC technique is minimized by applying the linear program boosting classification technique.

Let us consider the 1000 data for computing the false positive rate. The proposed LPBC technique obtains 10% of the false positive rate whereas the false positive rate of the hybrid neural model [1] and SVR [2]is 18% and 23% respectively. Similarly, the nine remaining runs are performed to show the performance of the LPBC technique. The observed results prove that the LPBC techniqueminimizes the false positive rate by 44% when compared to the hybrid neural model [3] and 58% compared to the SVR [4].

4.3 Impact of time complexity

Time complexity is defined as the amount of time required to classify the data into different classes. The time complexity is computed using the followingmathematical formula,

$$TC = Number\ of\ data * time\ (classifying\ one\ data) \quad (17)$$

From (17), TC represents the time complexity. The time complexity is measured in terms of milliseconds (ms)

Sample calculation for time complexity

Proposed LPBC technique: Total number of data is 1000 and the time taken for classifying one weather data is 0.023. Then time complexity is calculated as follows,

$$TC = 1000 * 0.023 = 23ms$$

Existing hybrid neural model: Total number of data is 1000 and the time taken for classifying one weather data is 0.032. Then time complexity is calculated as follows,

$$TC = 1000 * 0.032 = 32\ ms$$

Existing SVR:Total number of data is 1000 and the time taken for classifying one weather data is 0.041. Then time complexity is calculated as follows,

$$TC = 1000 * 0.041 = 41\ ms$$

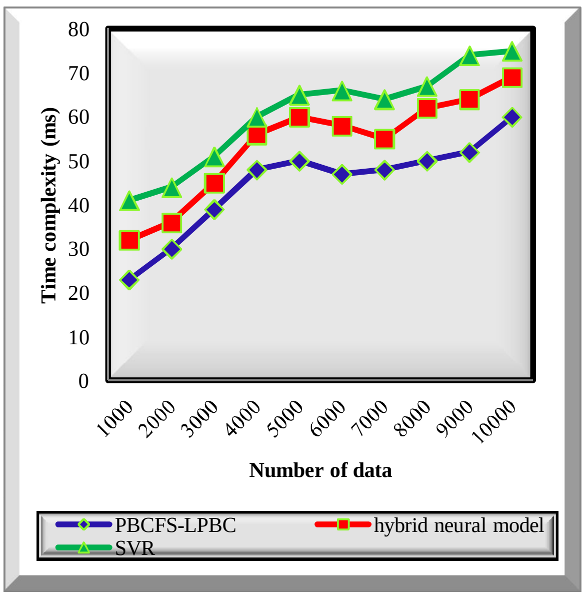

Figure 7: Performance results of time complexity

Figure 7 depicts the experimental results of time complexity versusa numberof weather data. The above figure clearly evident that the performance results of time complexity arecomparatively minimized using LPBC technique than the other two existing methods.This is because of LPBC technique accurately selects the relevant features from the Atlantic hurricane database. This database includes several features and number of cyclones data.Among the features, the relevant features are selected for minimizing the time complexity in the classification process.In addition, the boosting classification technique classifies the data into different classes by constructing the number of weak learners. The weak learner results are combined to provide a strong classification for improving the types of cyclones occurred in the Atlantic Ocean with minimum time complexity. Moreover, the proposed LPBC technique access the thousands of data in the classification process and eachassessment typically consumes a significant amount of time. This helps to minimize the complexity of the cyclone prediction.

The boosting classifier considers1000 data from the Atlantic hurricane database. The time complexity in the data classification using LPBC technique is 23ms, and time complexity of hybrid neural model [3] and SVR [4] are 32ms and 41ms. Similarly, all the runs are carried out toshow the performance ofclassificationresults. The final comparison results observed that thetime complexity of LPBC technique is minimized by 17% and 27% when compared to the existing hybrid neural model [3] and SVR [4]respectively.

From the above results and discussion, the LPBC technique considerably improves the cyclone prediction in the Atlantic Ocean with relevant features and minimizes the time complexity as well as false positive rate.

5.CONCLUSION

An efficient technique calledLinear Program Boosting Classification (LPBC) is developed for predicting future outcomes with higher classification accuracy and lesser time complexity. The LPBC technique considers big weather dataset. Initially, the LPBC technique is applied for categorizing the weather data for predicting future outcomes by constructing the weak learners. The weak learnerperforms polytomous logistic regression analysis for classifying the data into various classes. Then the weak learners are combined to construct the strong one and increase the margin between the different classes. This helps for improving the classification performance with the minimumfalse positive rate. Experimental evaluation is carried out using an Atlantic hurricane database with the parameters such as feature selection accuracy, classification accuracy, false positive rate,and time complexity. The experimental result of LPBC technique improves the classification and feature selection accuracy with a minimum false positive rate as well as time complexity than the state-of-art methods.

REFERENCES

[1] Richard D. De Veaux, Roger W. Hoer, and Ronald D. Snee, “Big Data and the Missing Links”, Statistical analysis and data mining, Wiley, Volume 9, Issue 6, 2016, Pages 411-416 https://doi.org/10.1002/sam.11303

[2] Amy McGovern, David John Gagne, Nathanie Troutman, Rodger A. Brown, Jeffrey Basaraand John K. Williams, “Using Spatiotemporal Relational

Random Forests to Improve Our Understanding of Severe Weather Processes", Statistical analysis and data mining, Wiley, Volume 4, Issue 4, 2011, Pages 407-429
https://doi.org/10.1002/sam.10128

[3] Tanzila Sabar, Amjad Rehman, Jarallah S. AlGhamdi, "Weather forecasting based on the hybrid neural model", Applied Water Science, Springer, Volume 7, Issue 7, 2017, Pages 3869–3874
https://doi.org/10.1007/s13201-017-0538-0

[4] Xiongxin Xiao, Tingjun Zhang, Xinyue Zhong, Wanwan Shao, Xiaodong Li, "Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data", Remote Sensing of Environment, Elsevier, Volume 210, 2018, Pages 48–64
https://doi.org/10.1016/j.rse.2018.03.008

[5] Chih-Chiang Wei, "Conceptual weather environmental forecasting system for identifying potential failure of under-construction structures during typhoons", Journal of Wind Engineering & Industrial Aerodynamics, Elsevier, Volume 168, 2017, Pages 48–59
https://doi.org/10.1016/j.jweia.2017.05.010

[6] V. Iordanidou, A. G. Koutroulis, and I. K. Tsanis, "A Probabilistic Rain Diagnostic Model Based on Cyclone Statistical Analysis", Advances in Meteorology, Hindawi Publishing Corporation, Volume 2014, June 2014, Pages 1-11
https://doi.org/10.1155/2014/498020

[7] Pavel Krömer, Jan Platoš, "Simultaneous Prediction of Wind Speed and Direction by Evolutionary Fuzzy Rule Forest", Procedia Computer Science, Elsevier, Volume 108, 2017, Pages 295-304
https://doi.org/10.1016/j.procs.2017.05.195

[8] S. B. Pooja and R. V. S. Balan, "An Investigation Study on Clustering and Classification Techniques for Weather Forecasting", Journal of Computational and Theoretical Nanoscience vol. 16, no. 2, pp. 417–421, 2019.
https://doi.org/10.1166/jctn.2019.7742

[9] Gunasekaran Manogaran, Daphne Lopez, Naveen Chilamkurti, "In-Mapper combiner based MapReduce algorithm for processing of big climate data", Future Generation Computer Systems, Elsevier, Volume 86, 2018, Pages 433-445
https://doi.org/10.1016/j.future.2018.02.048

[10] P. Samuel Quinan and Miriah Meyer, "Visually Comparing Weather Features in Forecasts", IEEE Transactions on Visualization and Computer Graphics, Volume 22, Issue 1, January 2016, Pages 389-398
https://doi.org/10.1109/TVCG.2015.2467754

[11] Prasanta Rao Jillella S.S, P Bhanu Sai Kiran, P. Nithin Chowdary, B. Rohit Kumar Reddy, Vishnu Murthy, "Weather Forecasting Using Artificial Neural Networks and Data Mining Techniques", International Journal Of Innovative Technology And Research, Volume 3, Issue.6, 2015, Pages 2534 – 2539

[12] Mumtaz Ali , Ravinesh C.Deo, Nathan J.Downs, Tek Maraseni, "Multi-stage hybridized online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm for rainfall forecasting", Atmospheric Research, Elsevier, Volume 213, 2018, Pages 450-464
https://doi.org/10.1016/j.atmosres.2018.07.005

[13] M Ramzan Talib, Toseef Ullah, M Umer Sarwar, M Kashif Hanif and Nafees Ayub, "Application of Data Mining Techniques in Weather Data Analysis", International Journal of Computer Science and Network Security, Volume 17, Issue 6, 2017, Pages 22-28

[14] Jyotismita Goswami and Alok Choudhury, "Dynamic Modeling Technique for Weather Prediction", International Journal of Computer Science & Engineering Technology (IJCSET), Volume 5, Issue 05, 2014, Pages 524-531

[15] S. B. Pooja and R. V. S. Balan, "Iterative Gradient Ascent Expected Maximization Clustering for Weather Forecasting", International Journal of Recent Technology and Engineering (IJRTE) no. 6, pp. 412–418, 2019.

[16] J.Y.Lu, Y.X.Peng, M.Wang, S.J.Gu, M.X.Zhao, "Support Vector Machine combined with Distance Correlation learning for Dst forecasting during intense geomagnetic storms", Planetary and Space Science, Elsevier, Volume 120, 2016, Pages 48-55
https://doi.org/10.1016/j.pss.2015.11.004

[17] Arif I. Sarwat, Mohammadhadi Amini, Alexander Domijan, Jr.Aleksandar DamnjanoviC, Faisal Kaleem, "Weather-based interruption prediction in the smart grid utilizing chronological data", Journal of Modern Power Systems and Clean Energy, Springer, Volume 4, Issue 2, 2016, Pages 308–3
https://doi.org/10.1007/s40565-015-0120-4

[18]S. B. Pooja and R. V. S. Balan, "Weather Data and Its Future Selection Using Principal Component Regression Technique", Journal of Advance Research in Dynamical & Control Systems, Vol. 11, 04-Special Issue, 2019.

[19] Zhongshan Yang, Jian Wang, "A combination forecasting approach applied in multistep wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm", Applied Energy, Elsevier, Volume 230, 2018, Pages 1108-1125
https://doi.org/10.1016/j.apenergy.2018.09.037

[20] Z. Ghaemi , A. Alimohammadi , M. Farnaghi, "LaSVM-based big data learning system for dynamic prediction of air pollution in Tehran", Environmental Monitoring and Assessment, Springer, Volume 190, Issue 300, 2018, Pages 1-17
https://doi.org/10.1007/s10661-018-6659-6

[21] Fethi Ouallouche, Mourad Lazri, Soltane Ameur, "Improvement of rainfall estimation from MSG data using Random Forests classification and regression", Atmospheric Research, Elsevier, Volume 211, 2018, Pages 62-72
https://doi.org/10.1016/j.atmosres.2018.05.001
[22] Jinyoung Rhee and Jungho Im, "Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data", Agricultural and Forest Meteorology, Volumes 237, 2017, Pages 105-122.
https://doi.org/10.1016/j.agrformet.2017.02.011
[23] Lai, Wai Yan. (2019).A Study on Sequential K-Nearest Neighbor (SKNN) Imputation for Treating Missing Rainfall Data, International Journal of Advanced Trends in Computer Science and Engineering, 8(3), 363-368.
https://doi.org/10.30534/ijatcse/2019/05832019
[24] Cesar A. Llorente. (2019), Implementation of a Web based Weather Monitoring Station and Data Storage System, International Journal of Advanced Trends in Computer Science and Engineering, 8(3), 527-530.
https://doi.org/10.30534/ijatcse/2019/29832019