2010 Third International Joint Conference on Computational Science and Optimization

A Novel Bayesian Additive Regression Trees Ensemble Model Based on Linear Regression and Nonlinear Regression for Torrential Rain Forecasting

Jiansheng Wu and Liangyong Huang
Department of Mathematical and Computer Science
Liuzhou Teachers College
Liuzhou, Guangxi, China
Email: wjsh2002168@163.com, huangliangyong@126.com

Xiaoming Pan
Department of Physics and Information Science
Liuzhou Teachers College
Liuzhou, Guangxi, China
Email: 1882001@163.com

***Abstract*—In order to improve the accuracy of precipitation forecasting with the linear regression of traditional statistical model and the nonlinear regression of Neural Network (NN) model, especially in torrential rain, a novel Bayesian Additive Regression Trees (BART) ensemble model is proposed in this paper. Firstly, three different linear regression model are used to extract the linear characteristic of rainfall system with the Partial Squares Least Regression, the Quantile Regression and the M-regression. Secondly, three different NNs model are used to extract the nonlinear characteristics of rainfall system with the General Regression Neural Network (GR–NN), the Radial Basis Function Neural Network (RBF–NN) and the Levenberg-Marquardt Algorithm Neural Network (LMA–NN). Finally, the BART is used for ensemble model based on linear and nonlinear regression. For illustration, a summer daily rainfall example is utilized to show the feasibility of the BART ensemble model in improving the accuracy of torrential rainfall with linear regression and nonlinear regression model. Empirical results obtained reveal that the torrential rainfall prediction by using the BART ensemble model is generally better than those obtained using other models presented in this paper in terms of the same evaluation measurements. Our findings reveal that the BRAT ensemble model proposed here can be used as an alternative forecasting tool for a Severe Weather application in achieving greater forecasting accuracy and improving prediction quality further.**

***Keywords*-Bayesian Additive Regression Trees; Neural Network; Rainfall Forecasting; Ensemble**

I. Introduction

Without doubt accurate forecasting of torrential rainfall has been one of the most important issues in hydrological research, because early warnings of severe weather, made possible by timely and accurate forecasting can help to prevent casualties and damages caused by natural disasters. Developing a rainfall forecasting and flood warning system for typical catchments is not considered a simple task [1]. In general, rainfall forecasting involves a rather complex nonlinear data pattern, for example pressure, temperature, wind speed, meteorological characteristics of the catchments and so on. Although the climate dynamics method for rainfall forecasting has been making great progress, given the short time scale, the small catchments area, and the massive costs associated with collecting the required meteorological data, it is not a feasible alternative in most cases because it involves many variables which are interconnected in a very complicated way. Even if the weather dynamics model can be the effective establishment, it is difficult to solve because it involves a complex numerical model of computing technology [2], [3].

At present business applications in meteorology forecasting, statistical prediction is the main method based on the history observed rainfall data with linear regression, time series analysis, etc. For traditional statistical methods, it is extremely difficult to capture the nonlinear characteristics of rainfall system, especially for outliers data which contain important predictive information on severe weather. With the development of science and technology, as well as the variety of intelligent computing technology have been developing in the past few decades, many emerging techniques, such as Neural Networks (NNs),were widely used in the rainfall forecasting and obtained good results [4], [5].

NNs are based on a model of the human neurological system which consists of a series of basic computing elements (called neurons) interconnected together to allow recognition of incidents that have had a similar pattern to the current input [6]. It has proven to be very successful in dealing with complicated problems without understanding the physical laws and any assumptions of traditional statistical approaches required, such as function approximation and pattern recognition. Due to their powerful capability and functionality, NNs provide an alternative approach for many engineering problems that are difficult to solve by conventional approaches (especially for nonlinear and dynamic evolutions) [7].

Empirical results indicated that the NNs model with lower lag outperformed in terms of forecasting accurate index. Moreover, the results of many experiments have shown that the generalization of single neural network is not unique in the practical application. That is, the results of neural networks are not stable [8]. As neural network approaches want of a rigorous theoretical support, effects of applications strongly depend upon operators experience. If carelessly used, it can easily learn irrelevant information (noises) in

978-0-7695-4030-6/10 $26.00 © 2010 IEEE
DOI 10.1109/CSO.2010.15

the system (over-fitting). Such a model might be doing well in predicting past incidents, but unable to predict future events. Therefore, applied effects of neural networks vary with operators. That is to say, even the same method is applied to solve the same problem, different operators can probably work out different results, and limit applications of neural networks in the practical application [9], [10].

Different from the previous work, in this paper, a novel method is presented for summer torrential rainfall forecasting model in terms of Bayesian Additive Regression Trees (BART) ensemble based on linear regression and nonlinear regression, which three linear regression models extract linear features and three NNs extract nonlinear features. The rainfall data in Guangxi is predicted as a case study for the development of daily rainfall forecasting model. The rest of this study is organized as follows. Section 2 describes the model building process. For further illustration, this work employs the method to set up a prediction model for daily rainfall forecasting in Section 3. Finally, some concluding remarks are drawn in Section 4.

II. The Building Process of the Bayesian Additive Regression Trees Ensemble Model

The weather process is a complex system that contains many uncertain factors, it is hardly exactly speaking that it is merely a linear or nonlinear system. Therefore, the modeling of precipitation forecasting should contain some linear and nonlinear characteristics.

A. Extract the Linear Features of Rainfall System

Standard regression has been one of the most important statistical methods for applied research for many decades. Assume that $\{y_i, x_i, i = 1, 2, \cdots, n\}$ are independent bivariate observations from the pair of response–explanatory variables $\{X, Y\}$, To describe the relationship between Y and X, a typical regression model is described by

$$y_i = m(x_i, \alpha) + \varepsilon_i \tag{1}$$

where α is a vector of unknown parameters, and $\varepsilon_i, i = 1, 2, \cdots, n$ are i.i.d. zero-mean errors. The least squares estimator of α minimizes the objective function

$$\sum_{i=1}^{n}(y_i - m(x_i, \alpha))^2 \tag{2}$$

Partial least squares (PLS) regression analysis was developed in the late seventies by Herman [11]. It is a statistical tool that has been specifically designed to deal with multiple regression problems where the number of observations is limited, missing data are numerous and the correlations between the predictor variables are high. PLS regression is a recent technique that generalizes and combines features from principal component analysis and multiple regressions. It is particularly useful when we need to predict a set of dependent variables from a large set of independent variables (i.e., predictors). Readers interested in a more detailed introduction about these PLS regression are referred to the related literature [12].

Quantile regression has emerged as an important statistical methodology [13]. By estimating various conditional quantile functions, quantile regression complements the focus of classical least squares regression on the conditional mean, and offers a systematic strategy for examining how covariates influence the entire response distribution. It offers a more complete statistical model than mean regression and now has widespread applications [14]. Estimation proceeds by minimizing

$$\sum_{i=1}^{n}\rho_p(y_i - x_i^T\alpha) \tag{3}$$

where ρ_p is the “check” function given by

$$\rho_p = puI_{[0,\infty)}(u) - (1-p)uI_{(-\infty,0)}(u) \tag{4}$$

where p indexes the conditional quantitle of current interest. In this paper, the local linear fit is to approximate the unknown pth quantile $u(x)$ by a linear function

$$q_p(z) = q_p(x) + q_p'(x)(z-x) \equiv a + b(z-x) \tag{5}$$

Quantile regression is rewrited as

$$\sum_{i=1}^{n}\rho_p\{Y_i - a - b(X_i - x)\}K(\frac{x - X_i}{h}) \tag{6}$$

where h and K are the bandwidth and kernel function. Quantile regression see to the literature [15]

M–regression is nonparametric regression method based on M–estimates by using either kernel functions. It has been popularly used in curve fitting, signal denosing, and image processing. In such applications, the underlying functions (or signals) may vary irregularly, and it is very common that data are contaminated with outliers [16]. M–regression minimizes

$$\sum_{i=1}^{n}\rho_{\sigma H}(y_i - m(x_i, \alpha))^2 \tag{7}$$

where $\rho_{\sigma H}$ is the Huber function [17]. The Huber function replaces the square function with a less rapidly increasing function–the absolute value function when absolute standardized residuals $|\frac{y_i - m(x_i,\alpha}{\sigma}|$ exceed the tuning constant H. The robustness is obtained due to less contribution to Equation (5) from outliers with large absolute standardized residuals. More details about M–regression are introduced by the literature [18].

B. Extract the Nonlinear Features of Rainfall System

NNs are one of the technologies soft computing. They provide an interesting technique that theoretically can approximate any nonlinear continuous function on a compact domain to any designed of accuracy. The network learns by adjusting the interconnections (called weights) among layers. In this research, there are three NN methods to capture nonlinear pattern in rainfall system, such as General Regression Neural Network (GR–NN), the radial basis function network network (RBF–NN) and multi–layer perceptrons based on the Levenberg-Marquard algorithms (MLP–NN).

The GRNN was firstly proposed and developed by Specht [19]. It is a variation of the radial basis neural networks, which is based on kernel regression networks and and other non-parametric functional approximations. Although the GRNN does not require an interactively training procedure as MLP networks, It is able to approximate any arbitrary function between input and output vectors by drawing the function estimate directly from the training data. This type of network model has unique advantages, such as learning swiftly, simple and straightforward training algorithm, discriminative against infrequent outliers and so on [20].

Radial basis function was introduced into the neural network literature by Broomhead and Lowe [21]. The RBF–NN model is a network with local neurons which was motivated by the presence of many local response neurons in human brain. Other motivation came from numerical mathematics. On the contrary to the other type of NN used for nonlinear regression like back propagation feed forward networks, the RBF–NN learns quickly and has a more compact topology. The Gaussian RB–NN is found suitable not only in generalizing a global mapping but also in refining local features [22].

In this paper, the Levenberg-Marquardt Algorithm [23] is used to train for MLP–NN. The training process has some advantages, such as converges quickly, acquisition more subtle features of a complicated mapping, etc. It is a kind of quick algorithm in which the network's matrix of weights is updated based on the Jacobian matrix, *J*, which collects the partial derivatives of the network error $F(\omega, \theta)$ with respect to the weights. In other words, the matrix $\Delta\omega$ collecting the corrections of the weights in matrix ω is computed according to

$$\Delta\omega = [J^T J + \mu I)]^{-1} J^T e \tag{8}$$

where μ is the learning rate which is to be updated by using the β depending on the outcome. In particular, μ is multiplied by decay rate $\beta(0 < \beta < 1)$ whenever error function decreases, whereas μ is divided by β whenever function decreases or increases in a new step.

C. Bayesian Additive Regression Trees Ensemble

BART is a nonparametric Bayesian regression approach which uses dimensionally adaptive random basis elements. Motivated by ensemble methods in general, and boosting algorithms in particular, BART is defined by a statistical model: a prior and a likelihood. This approach enables full posterior inference including point and interval estimates of the unknown regression function as well as the marginal effects of potential predictors. By keeping track of predictor inclusion frequencies, BART can also be used for model free variable selection [24].

Consider the regression model

$$Y = f(x) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \tag{9}$$

where x is the data matrix of size $n \times p$ containing the predictor variables in its columns, and Y is the data vector of size $n \times 1$ containing the dependent variable. To do this, BART consider at least approximating $f(x) = E(Y|x)$, the mean of Y given x, by a sum of m regression trees

$$f(x) \approx h(x) \equiv \sum_{j=1}^{m} g(x; T_j, M_j) \quad \varepsilon \sim N(0, \sigma^2) \tag{10}$$

where for each binary regression tree T_j and its associated terminal node parameters M_j, $g(x, T_j, M_j)$ is the function which assigns $u_{ij} \in M_j$ to x. When the number of trees $m > 1$, each u_{ij} here is merely a part of $E(Y|x)$, represent a main effect when $g(x, T_j, M_j)$ depends on only one component of x (i.e., a single variable), also represent an interaction effect when $g(x, T_j, M_j)$ depends on more than one component of x (i.e., more than one variable). Interested readers can be referred to [24] for more details.

In this paper, the traditional regression model can extract linear characteristics, and the neural network model has highly mapping feature to nonlinear characteristics. Therefore, firstly, we use Partial Least Square Regression model, Quantile Regression and M–regression model to extract linear characteristic of the stock market system, and use different neural network algorithms to extract nonlinear characteristic of the stock market system. Finally, the BART is used to generate the predictive output.

III. Selection of Data and Method of Model Building

A. Empirical Data

Real–time ground rainfall data has been obtained in June from 2005 to 2008 in Guangxi by observing 89 stations, which 120 samples are modelling from June 2005 to June 2008, other 30 samples are tested modelling in June of 2009. Method of modelling is one–step ahead prediction, that is, the forecast is only one sample each time and the training samples is an additional one each time on the base of the previous training.

In this paper, the candidate forecasting factors are selected from the numerical forecast products based on 48h forecast field, which includes:

(1) the 17 conventional meteorological elements and physical elements from the T213 numerical products of China Meteorological Administration, the data cover the latitude from 15°N to 30°N, and longitude from 100°E to 120°E, with $1° \times 1°$ resolution, altogether there are 336 grid points.

(2) the fine-mesh precipitation data from the East Asia of Japanese Meteorological Agency, the data cover the latitude from 15°N to 30°N, and longitude from 100°E to 120°E, with $1.25° \times 1.25°$ resolution, altogether there are 221 grid points.

The main forecasting factors are exploited by the correlation coefficients test based on historical precipitation data and the candidate forecasting factors. We can get 15 variables as the main forecasting factors.

Due to the complex terrain of Guangxi and inhomogeneous rainfall, the region has been divided into three regional precipitation based on historical precipitation data by the cluster analysis method to reduce the difficulty of forecasting. Statistics for each district in the average daily precipitation is used as the forecasting object. Figure 2 shows three region maps.

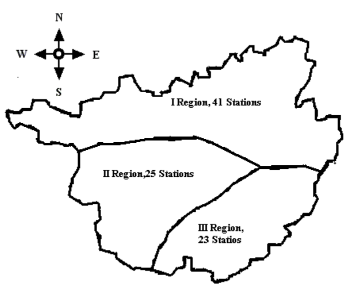

Figure 1. The Group Average Region Map of Guangxi Rainfall.

In this paper, the first region as an example explain the process of establishing the model. In order to measure effectiveness of the BART ensemble method, four types of errors, such as, the Root Mean Squares Error (RMSE), the Pearson Relative Coefficient (PRC), the Error Value more than 25*mm* (F_1) and the Error Value less than 5*mm* (F_2). Interested readers can be referred to [12] for more details. For the purpose of comparison, we have also built two other ensemble forecasting models: (1) simple average all the available forecasting output based on linear and nonlinear model(SA); (2) output of T213 numerical prediction product.

B. Analysis of the Results

Fig. 3 gives graphical representations of the fitting results for the rainfall in the first regions with different models, which are used to fit the rainfall of 120 training samples for comparison. Tab. 2 illustrates the fitting accuracy and efficiency of the model in terms of various evaluation indices for 120 training samples. From the graphs and tables, we can generally see that learning ability of BART model outperforms the other two models under the same input and parameters. The more important factor to measure performance of a method is to check its forecasting ability of testing samples for actual rainfall application.

From the graphs and table, we can generally see that the forecasting results are very promising in the rainfall

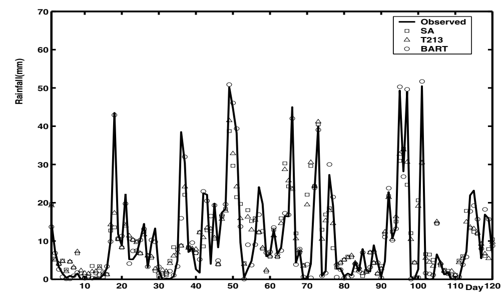

Figure 2. Fitting of Training 120 Samples in The First Region.

Table I
A COMPARISON OF RESULT OF DIFFERENT MODELS FOR TRAINING AND TESTING SAMPLES.

Results	Training Samples(120)			Testing Samples(30)		
Errors	SA	T213	BART	SA	T213	BART
RMSE	9.00	8.28	3.14	14.93	15.97	4.41
PRC	0.73	0.76	0.97	0.58	0.36	0.97
F_1	7	5	1	5	8	0
$F_1 \cdot 100$	5.83%	4.17%	0.83%	16.67%	26.67%	0%
F_2	70	71	113	13	12	25
$F_2 \cdot 100$	58.33%	59.17%	94.17%	43.33%	40.0%	83.33%

forecasting of the first regions under the research where either the measurement of fitting performance is goodness fitting, such as RMSE and PRC, where the forecasting performance performance (refer table 1).

If the error value less than 5mm is reference information, the reference information of BART model is 94.17% in fitting samples, which is 83.33% in testing samples. If the error value more than 25mm is unreliable information, the BART model has no unreliable information, which unreliable information of SA is 5.83% in fitting samples, is 16.67% in forecasting samples, unreliable information of SA T213 is 4.17% in fitting samples, is 26.67% in forecasting samples. Those shows that the forecasting of BART ensemble model for a Meteorological application can prove more reference information and can avoid no unreliable information.

Furthermore, we use the same method to train precipitation data and predict precipitation in June in the other two regions. The experimental results also show that BART ensemble method has better generalization ability than SA and T213 method.

IV. CONCLUSION

The rainfall system is one of the most active dynamic weather systems including linear and nonlinear ingredient, and its interaction is also the most complex. And because it is influenced by many changeable factors, it is very difficult to predict the results. In this paper, three kinds of different linear regression models are used to extract the

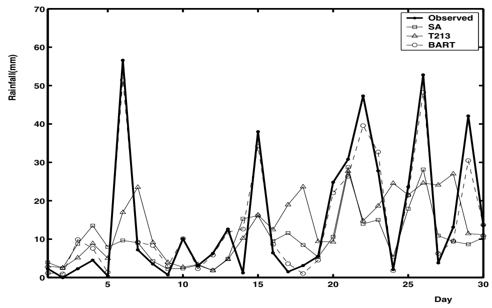

Figure 3. Forecasting of Testing 30 Samples in The First Region.

linear features of the rainfall system, three kinds of different neural network algorithms are used to extract the nonlinear. Two groups of prediction individuals are resembled with BART to generate the final conclusions, examples of calculation shows that the method can significantly improve the system's predictive ability, prediction accuracy, and with a high prediction accuracy. Empirical results obtained reveal that the proposed nonlinear combination technique is a very promising approach to rainfall forecasting.

ACKNOWLEDGMENT

This work was supported by the Department of Guangxi Education under Grant No. 2007MS098.

REFERENCES

[1] K. C. Luk, J. E. Ball and A. Sharma, "An application of artificial neural networks for rainfall forecasting," Mathematical and Computter Modelling, vol. 33, pp: 683–693, 2001.

[2] Jiansheng Wu, A novel nonparametric regression ensemble for rainfall forecasting using particle swarm optimization technique coupled with artificial neural network, Lecture Note in Computer Science, Vol. 5553, No. 3, pp: 49–58, Springer-Verlag Berlin Heidelberg, 2009.

[3] Jiansheng Wu, A novel artificial neural network ensemble model based on K–nn nonparametric estimation of regression function and its application for rainfall forecasting, Proeedings of the 2nd Internatioal Joint Conference on Computational Sciences and Optimization, eds. Lean Yu, K. K. Lai and S. K. Mishra, IEEE Computer Society Press, vol. 2, pp:44–48, 2009.

[4] G. M. Tseng, C. D. Jan, J. S. Wang and C. M. Wang, "Application of an artificial neural network to typhoon surge forecasting," Ocean Engineering, Vol. 34, pp: 1757–1768, 2007.

[5] Jiansheng Wu and Long Jin, "Study on the meteorological prediction model using the learning algorithm of neural network based on pso algorithms," Journal of Tropical Meteorology, Vol. 16, No. 1, pp: 83–88, 2009.

[6] G. F. Lin and L. H. Chen, "A non–linear rainfall–runoff model using radial basis function network," Journal of Hydrology, Vol. 289, No. 4, pp: 1–8, 2004.

[7] W. C. Hong, "Rainfall forecasting by technological machine learning models," Appliaion Matchemaics and Computation, Vol. 200, pp: 41–57, 2008.

[8] M. H. Fredric and K. Ivica, Principles of Neurocomputing for Science & Engineering. McGraw-Hill, New York 2001.

[9] A. Jain and A. M. Kumar, "Hybird neural network models for hydrologic time series foreasting," Applied Soft Computing, Vol. 7, pp: 585–592, 2007.

[10] Jiansheng Wu and Long Jin, "Neural Network with Partial Least Square Prediction Model Based on SSA-MGF," Journal of Catastrophology, Vol. 21, no. 2, pp: 17–23, 2006.

[11] P. H. Garthwaite, "An Interpretation of Partial Least Squares. Journal of the American Statistical Association," Vol. 89, No. 425, pp: 122–27, 1994.

[12] H. Abdi, PLS-Regression, Multivariate analysis. In M. Lewis-Beck, A. Bryman and T. Futing (Eds): Encyclopedia for research methods for the social sciences. Thousand Oaks: Sage, 2003.

[13] R. Koenker and G. J. Bassett, "Regression quantiles. Econometrica," Vol. 46, pp: 33-50, 1978.

[14] Keming Yu, Zudi Lu and J. Stander, "Quantile regression: applications and current research areas," The Statistician, Vol. 52, No. 3, pp: 331–350, 2003.

[15] Keming Yu and M. C. Jones, "Local liear quantile regression," Journal of the Amercian Statistical Association, Vol. 93, No. 441, pp: 228–237, 1998.

[16] P. Hall and M. Jones, "Adaptive M–estimation in nonparametric regression," Annals of Statistics, Vol. 18, pp: 1712–1728, 1990.

[17] C. Chen, "Bayesian adaptive nonparametric M–regression," Statistics and Its Interface, Vol. 2, 71–81, 2009.

[18] C. J. Stone, "Nonparametric M-regression with free knots," Journal of Statistical Planning and Inference, Vol. 130, pp: 183–206, 2005.

[19] D. Specht, "A general regression neural network," IEEE Transactions on Neural Networks, Vol. 2, pp: 568–576, 1991.

[20] M. T. Leung, A. S. Chen, H. Daouk, Forecasting exchange rates using general regression neural networks. Computers and Operations Research, 27 (2000) 1093-1110.

[21] D. S. Broomhead, D. Lowe, Multivariable functional interpolation and adaptive networks. Complex Systems, 2(1988) 321-355.

[22] P. K. Dash, A. K. Pradhan, G. Panda, Application of minimal radial basis function neural network to distance protection. IEEE Transactions on Power Delivery, 16(1)(2001),68-74.

[23] M. T. Hagan, M. B. Mehnaj, Training Feedforward Networks with the Marquardt Algorithm, IEEE Transaction on Neural Networks, 5(6)(1994) 989-993.

[24] H. A. Chipman, E. I. George, R. E. McCulloch, "BART: Bayesian Additive Regression Trees,"