An Application of Data Mining and Machine Learning for Weather Forecasting

Conference Paper *in* Advances in Intelligent Systems and Computing · July 2018

DOI: 10.1007/978-3-319-60663-7_16

3 authors:

Risul Islam Rasel
Chittagong Independent University
29 PUBLICATIONS **253** CITATIONS
SEE PROFILE

Nasrin Sultana
University of Chittagong
14 PUBLICATIONS **81** CITATIONS
SEE PROFILE

Phayung Meesad
King Mongkut's University of Technology North Bangkok
169 PUBLICATIONS **1,619** CITATIONS
SEE PROFILE

An Application of Data Mining and Machine Learning for Weather Forecasting

Risul Islam Rasel[1](✉), Nasrin Sultana[2], and Phayung Meesad[3]

[1] Department of Computer Science and Engineering,
International Islamic University Chittagong, Chittagong, Bangladesh
risul-islam@cse.iiuc.ac.bd

[2] Department of Computer Science and Engineering, University of Chittagong,
Chittagong, Bangladesh
nasrin_cse@cu.ac.bd

[3] Department of Information Technology Management,
KMUTNB, Bangkok, Thailand
phayung.m@it.kmutnb.ac.th

Abstract. Weather forecasting for an area where the weather and climate changes occurs spontaneously is a challenging task. Weather is non-linear systems because of various components having a grate impact on climate change such as humidity, wind speed, sea level and density of air. A strong forecasting system can play a vital role in different sectors like business, agricultural, tourism, transportation and construction. This paper exhibits the performance of data mining and machine learning techniques using Support Vector Regression (SVR) and Artificial Neural Networks (ANN) for a robust weather prediction purpose. To undertake the experiments 6-years historical weather dataset of rainfall and temperature of Chittagong metropolitan area were collected from Bangladesh Meteorological Department (BMD). The finding from this study is SVR can outperform the ANN in rainfall prediction and ANN can produce the better results than the SVR.

Keywords: Data mining · Machine learning · SVM · ANN · Weather forecasting · Temperature · Rainfall

1 Introduction

The climate is the condition of the environment, whether it is hot or cool, wet or dry, quiet or stormy, clear or shady [1–3]. Most climate marvels happen in the troposphere, just beneath the stratosphere. Weather prediction is one of the most challenging tasks to accomplish because many natural and man-made components are involved in weather change such as change of seasons, greenhouse effect, deforestation etc. Those collectively make weather prediction more challenging [4, 5].

Weather prediction plays a significant role in many components in decision making related to many fields such as agriculture, business, tourism, energy management, human and animal health etc. [1]. Climate determining includes anticipating how the present circumstance with the air will change in which display atmosphere conditions

P. Meesad et al. (eds.), *Recent Advances in Information and Communication Technology 2017*, Advances in Intelligent Systems and Computing 566,
DOI 10.1007/978-3-319-60663-7_16

are taken by ground recognition, for example, from boats, plane, Radio sound, Doppler radar, and satellites. The gathered information is then sent to meteorological centers in which the data are assembled, examined, and made into a combination of frameworks, maps, and diagrams. Calculations trade countless onto the surface and upper air maps, and draw the lines on the maps with help from meteorologists. Calculations draw the maps and foresee how the maps will take a gander eventually later on. The determination of atmosphere condition utilizing calculations is plot as numerical or computational weather prediction. Generally the climate and atmosphere expectation issues have been seen as various disciplines [1, 4]. Numerical Weather Prediction (NWP) is urgently subject to characterizing an exact starting state and running at the most astounding conceivable resolutions, while atmosphere prediction has tried to incorporate the full multifaceted nature of the Earth framework keeping in mind the end goal to precisely catch long time-scale varieties and inputs deciding the present atmosphere and potential atmosphere change. The idea of a unified or seamless structure for climate and atmosphere expectation has pulled in a lot of consideration in the most recent couple of years the field of Data Mining and Machine learning has progressed rapidly over the last few decades [5–7]. Predictive analysis has gone to a very new level with the use of machine learning techniques. Weather data used in this study data are dependent on their nature and thus, their estimation is not effectively made with numerous quantitative methodologies. However, they can be portrayed, estimate and arranged quantitatively by utilizing probability theory.

The goal of this paper is to find the challenging pattern of weather of Chittagong, Bangladesh and to predict the weather. To tackle these challenges, we use a jointly predicts rainfall and temperature across space and time. The study combines a bottom-up predictor for each individual variable with a Support Vector Regression (SVR) [11] and an Artificial Neural Network (ANN) model [12] to determine an effective and efficient model. So, a comparative study between these algorithms is done in this study. The climate of Chittagong is described by tropical storm atmosphere. The dry and cool season is from November to March; the pre-storm season is from April to May which is exceptionally hot. The sunny and the rainstorm seasons are from June to October, which is warm, overcast and wet.

2 Methodology

2.1 Support Vector Regression (SVR)

SVM regression [11] performs linear regression in the high dimension feature space using ε – insensitivity loss and, at the same time tries to reduce model complexity by minimizing $\|\omega\|^2$. This can be described by introducing slack variables ξ_i and ξ_i^* where $i = 1, \ldots, n$ to measure the deviation of training sample outside ε - sensitive zone [8, 9].

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{1}$$

$$\min\begin{cases} y_i - f(x_i, \omega) \leq \varepsilon + \xi_i^* \\ f(x_i, \omega) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, \quad i = 1\ldots n \end{cases} \tag{2}$$

This optimization problem can transform into the dual problem and solution is given by

$$f(x) = \sum_{i=1}^{n_{sv}} (\alpha_i - \alpha_i^*) K(x_i, x) \tag{3}$$

$$\text{Subject to,} \quad 0 \leq \alpha_i^* \leq C, 0 \leq \alpha_i \leq C$$

Where n_{sv} is the number of support vector (SVs) and the kernel function

$$K(x, x_i) = \sum_{j=1}^{m} g_j(x) g_j(x_i) \tag{4}$$

It is well-known that SVM generalization performance depends on a good setting of meta-parameters C, ε and kernel parameters [9].

2.2 Artificial Neural Network (ANN)

ANN [10, 12] is based on a large collection of artificial neurons mathematically simulating the biological brain in solves problems. ANN can perform as a linear or non-linear function mapping from input data to output target. Multilayer perceptron (MLP) is a one of the most well-known neural networks able to learn any nonlinear function if there are enough training data and given a suitable number of neurons.

The activation function of the artificial neurons in ANNs implementing the back-propagation algorithm is a weighted sum (the sum of the inputs x_i multiplied by their respective weights w_{ji}) [12]:

$$A_j(\bar{x}, \bar{w}) = \sum_{i=0}^{n} x_i w_{ji} \tag{5}$$

The activation depends only on the inputs and the weights. If the output function would be the identity, then the neuron would be called linear. The most common output function is the sigmoidal function [12]:

$$O_j(\bar{x}, \bar{w}) = \frac{1}{1 + e^{A_i(\bar{x}, \bar{w})}} \tag{6}$$

The goal of the training process is to attain a desired output when certain inputs are given. Since the error is the difference between the actual and the desired output, the error depends on the weights, and we need to adjust the weights in order to minimize the error. We can define the error function for the output of each neuron [12]:

$$E_j(\bar{x}, \bar{w}, d) = (O_j(\bar{x}, \bar{w}) - d_j)^2 \tag{7}$$

The error of the network will simply be the sum of the errors of all the neurons in the output layer:

$$E(\bar{x}, \bar{w}, \bar{d}) = \sum\nolimits_j (O_j(\bar{x}, \bar{w}) - d_j)^2 \tag{8}$$

The backpropagation algorithm now calculates how the error depends on the output, inputs, and weights. After that, it adjusts the weights using the gradient descendent method [12]:

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} \tag{9}$$

The goal of the backpropagation algorithm is to find the derivative of E in respect to w_{ji}. First, we need to calculate how much the error depends on the output, which is the derivative of E in respect to O_j (from (7)).

$$\frac{\partial E}{\partial O_j} = 2(O_j - d_j) \tag{10}$$

And then, how much the output depends on the activation, which in turn depends on the weights. From (5) and (6):

$$\frac{\partial O_j}{\partial w_{ji}} = \frac{\partial O_j}{\partial A_j} \frac{\partial A_j}{\partial w_{ji}} = O_j(1 - O_j)x_i \tag{11}$$

And from (10) and (11) it can be seen that:

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial O_j} \frac{\partial O_j}{\partial w_{ji}} = 2(O_j - d_j)O(1 - O_j)x_i \tag{12}$$

The adjustment to each weight will come from (9) and (12):

$$\Delta w_{ji} = -2\eta(O_j - d_j)O_j(1 - O_j)x_i \tag{13}$$

Now, we can use the Eq. (13) for training an ANN with two layers [12].

3 Evaluation Process

3.1 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) [9, 10] or Root Mean Square Deviation (RMSD) is a commonly used measure of the contrasts between qualities anticipated by a model or an estimator and the qualities really observed. The RMSE serves to total the extents of the errors in forecasts for different times into a solitary measure of prescient force. RMSE is a decent measure of precision, however just to analyze estimating blunders of various models for a specific variable and not between variables, as it is scale-dependent.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_t - \widehat{y}_t)^2}{n}} \tag{14}$$

Here, y_t is the original value of a point for a given time t; n is the total number of fitted points, and $\widehat{y}_t$ is the fitted forecast value for the time t [9].

3.2 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) [9, 10] is a typical measure of figure mistake in time series data where the expressions "Mean Absolute Deviation" is occasionally utilized as a part of perplexity with the more standard meaning of mean absolute deviation.

$$MAE = \frac{SAE}{N} = \frac{\sum_{i=1}^{n} |x_i - \widehat{x}_i|}{N} \tag{15}$$

Here, x_i is the actual observations time series, $\widehat{x}_i$ is the estimated or forecasted time series. SAE is the Sum of the Absolute Error. N is the number of non-missing data points [9].

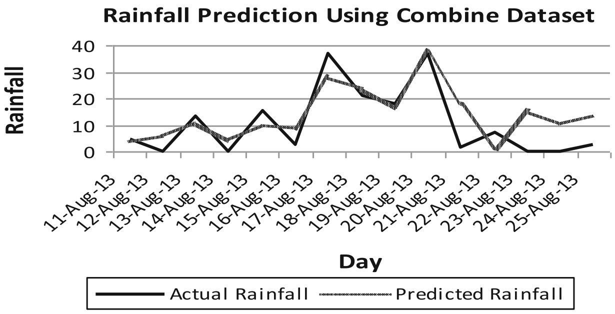

Fig. 1. Rainfall prediction using SVR with combine dataset

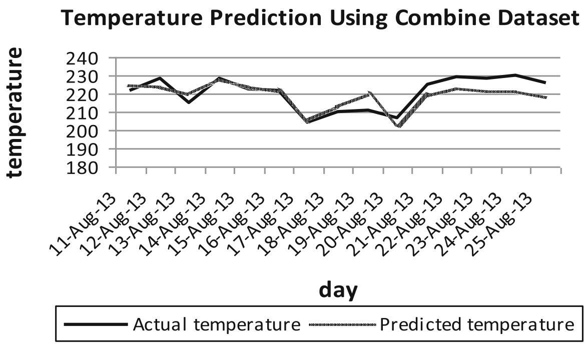

Fig. 2. Temperature prediction using SVR with combine dataset

4 Experiment Design

4.1 Windowing Operator Analysis

Windowing operator transforms a given example set containing series data into a new example set containing single valued examples. For this, windows with a specified window and step size are moved across the series and the attribute value lying horizon values after the window end is used as label which should be predicted [8].

Table 1 illustrates the parameter set up for windowing input into support vector regression model. The values of parameters here included only the best-optimized combination for forecasting rainfall and temperature. Horizon means how many days a-head to predict where training and testing window width are the major part for training the model in order to predict future value based on that learning. Step size is the sliding a-head value of the window that feed the input set into machine learning process. From Table 1, it is cleanly seen that all the parameters setup are same for both domain. These values are obtained by doing repetitive simulation process for 1 day a-head, 7 days a-head, and 10 days a-head future value prediction.

Table 1. Sliding window parameter for SVR

Model	Horizon	Training window width	Step size	Testing window width	Cumulative training
Rainfall	1	5	1	5	No
	7	5	1	5	
	10	5	1	5	
Temperature	1	5	1	5	
	7	5	1	5	
	10	5	1	5	

Table 2 describes the parameter setting for Artificial Neural Network (ANN) model. From Tables 1 and 2, there is a clear indication about training and testing window setup

Table 2. Sliding window parameter for ANN

Model	Horizon	Training window width	Step size	Testing window width	Cumulative training
Rainfall	1	2	1	2	No
	7	2	1	2	
	10	2	1	2	
Temperature	1	2	1	2	
	7	2	1	2	
	10	2	1	2	

that SVR needs more input than the ANN for producing good prediction results and ANN needs less input for training the model.

Kernel parameter analysis is one of the most important parts of the SVR simulation. Because appropriate kernel selection and optimized kernel parameter findings play vital rules for producing less erroneous results. In this work RBF, Gaussian, Polynomial, ANOVA and Neural kernels are analyzed with different parameter but only ANOVA produced better results among those kernels with priory mentioned parameters setup in Table 3.

Table 3. Kernel analysis for SVR

Model	Horizon	Kernel type	C
Rainfall	1	ANOVA	100
	7		120
	10		200
Temperature	1		100
	7		150
	10		300

Table 4 shows the optimized parameter setting for ANN models. Here training cycle, learning rate and values of *M* for every model setup is almost same. Every model uses 2 hidden layers for producing weighted input values for machine learning process.

Table 4. ANN parameter settings

Model	Horizon	Training cycle	Learning rate	M	Hidden layer
Rainfall	1	120	0.3	0.2	2
	7	120	0.3	0.2	
	10	110	0.3	0.2	
Temperature	1	120	0.3	0.2	
	7	100	0.3	0.2	
	10	110	0.3	0.2	

5 Experiment Results

Table 5 shows the result analysis for rainfall prediction using SVR and ANN model. Two types of simulations were undertaken, one is using only historical rainfall dataset and other is using a combined rainfall and temperature dataset for predicting only rainfall. Two evaluation processes RMSE and MAE were applied for understanding the error. From Table 5, it can be said that rainfall has a clear impact on temperature because when combined dataset were used the error rate were minimal than the only rainfall dataset produced. In addition, SVR outperformed ANN in predicting rainfall as it produced 0.95% and 0.17% error rate in both single and combined dataset.

Table 5. Rainfall prediction result

Model	Horizon	Rainfall using only rainfall dataset (Aug'13–Dec'14)		Rainfall using rainfall and temperature combine dataset (Aug'13–Dec'14)	
		(RMSE)	(MAE)	(RMSE)	(MAE)
SVR	1	20.33	**0.95**	19.88	1.93
	7	27.68	1.71	27.6	**0.17**
	10	28.57	2.25	30.96	4.51
ANN	1	21.41	3.54	18.43	2.42
	7	31.97	**10.87**	27.53	**11.33**
	10	27.34	1.02	25.84	3.31

Bold symbolizes the maximum and minimum error rate among the others values.

Table 6 shows the outcomes for temperature prediction using both SVR and ANN. From Table 6 it shows that ANN outperforms SVR for both single and combined dataset in temperature prediction. For ANN, the activation function which has the most significance in modeling needs non negative or non positive values as input rather than

Table 6. Temperature prediction result

Model	Horizon	Temperature using only rainfall dataset (Aug'13–Dec'14)		Temperature using rainfall and temperature combine dataset (Aug'13–Dec'14)	
		(RMSE)	(MAE)	(RMSE)	(MAE)
SVR	1	4.27	5.3	5.03	6.25
	7	9.82	4.18	11.7	**5.71**
	10	9.98	**2.56**	12.32	6.34
ANN	1	3.31	2.29	4.14	4.86
	7	7.89	**0.72**	8.4	**1.72**
	10	7.96	5.46	8.03	1.98

Bold symbolizes the maximum and minimum error rate among the others values.

the zero values for proper execution of the model. So in this experiment we used Leaky rectified linear unit (*Leaky ReLU*) which allows a small, nonzero gradient when the unit is not active [13].

6 Conclusion and Future Works

The purpose of this study was to observe weather forecasting performance of different Machine Learning and Data Mining techniques to propose a weather forecasting model to forecast weather with high accuracy. Two well-known data mining techniques: Support Vector Regression (SVR) and conventional Artificial Neural Network (ANN) were used to conduct the study. The data were fed to the algorithms using conventional windowing technique to train and test the model. A sliding window validation process was done to find convenient amount of training and testing input set to feed into machine learning process. Experiments were done using the same size of window for both SVR and ANN. However, Tables 1 and 2 show only the best fit of the training and testing window parameters settings, which have produced good results in forecasting rainfall and temperature. RMSE and MAE error calculation approaches were applied to calculate the error margin between actual and predicted values. The finding from this study is; SVR can outperform the ANN in rainfall prediction with marginal error rate using both types of dataset and ANN can produce the better results than the SVR with acceptable deviation of error rate.

In this study, dataset from a single station of a country have been used, other datasets were not applied into proposed techniques in order to compare the results. Only 6-year dataset was considered to build the models. For ANN models, maximum 3 hidden layer networks were used in this study. In future work, different dataset from different areas of the world will be applied and different settings of hidden layers in ANN and other different types of kernel for Support Vector Regression will be experimented.

References

1. Xiong, L., O'Connor, K.M.: An empirical method to improve the prediction limits of the glue methodology in rainfall–runoff modeling. J. Hydrol. **349**(1), 115–124 (2008)
2. Wu, J., Huang, L., Pan, X.: A novel Bayesian additive regression trees ensemble model based on linear regression and nonlinear regression for torrential rain forecasting. In: Third International Joint Conference on Computational Science and Optimization (CSO), vol. 2, pp. 466–470 (2010)
3. Wu, J., Chen, E.: A novel nonparametric regression ensemble for rainfall forecasting using particle swarm optimization technique coupled with artificial neural network. In: 6th International Symposium on Neural Networks, pp. 49–58 (2009)
4. Lin, G.F., Chen, L.H.: Application of an artificial neural network to typhoon rainfall forecasting. Hydrol. Process. **19**(9), 1825–1837 (2005)
5. Hong, W.C.: Rainfall forecasting by technological machine learning models. Appl. Math. Comput. **200**(1), 41–57 (2008)

6. Lu, K., Wang, L.: A novel nonlinear combination model based on support vector machine for rainfall prediction. In: Fourth International Joint Conference on Computational Sciences and Optimization (CSO), pp. 1343–1346 (2011)
7. Mellit, A., Pavan, A.M., Benghanem, M.: Least squares support vector machine for short-term prediction of meteorological time series. Theor. Appl. Climatol. **111**(1–2), 297–307 (2013)
8. Rasel, R.I., Sultana, N., Meesad, P.: An efficient modeling approach for forecasting financial time series data using support vector regression and windowing operators. Int. J. Comput. Intell. Stud. **4**(2), 134–150 (2015)
9. Hasan, N., Nath, N.C., Rasel, R.I.: A support vector regression model for forecasting rainfall. In: 2nd International Conference on Electrical Information and Communication Technology (EICT), pp. 1–6 (2015)
10. Rasel, R.I., Sultana, N., Hasan, N.: Financial instability analysis using ANN and feature selection technique: application to stock market price prediction. In: International Conference on Innovations in Science, Engineering and Technology (ICISET-2016), pp. 1–4 (2016)
11. Gunn, S.R.: Support vector machines for classification and regression. Technical Reports, University of Southampton (1998)
12. Gershenson, C.: Artificial neural networks for beginners. Technical Reports, University of Sussex
13. Rectifier (neural networks). https://en.wikipedia.org/wiki/Rectifier_(neural_networks)