International Journal of Pure and Applied Mathematics
Volume 117 No. 20 2017, 579-590
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue
ijpam.eu

An Enhanced Weather Forecasting Approach Using Cascading Correlation Network

[1]K. Lavanya and [2]J. Saira Banu

1School of Computer Science and Engineering (SCOPE),
VIT University,
Vellore, Tamil Nadu, India.
lavanya.k@vit.ac.in

2School of Computer Science and Engineering (SCOPE),
VIT University,
Vellore, Tamil Nadu, India.
jsairabanu@vit.ac.in

Abstract

Voluminous geographic information has been gathered with current data procurement methods such as Global Positioning Systems (GPS), high-sensing wireless detection, location-specific amenities and web based volunteered geographic data. There is a pressing requirement for compelling and proficient techniques to concentrate obscure and unpredictable data from spatial data sets of remarkably vast size, high dimensionality, and intricacy. To address these difficulties, spatial data-set mining along with geographic knowledge revelation has developed as a dynamic research field, concentrating on the improvement of hypothesis, approach, and practice for the extraction of valuable data and learning from huge and complex spatial databases.

Key Words: Neural network, cascade correlation network, weather analysis.

1. Introduction

In the twentieth century, various strategies and investigations were performed by meteorologists to foresee the climate with more prominent viability after some time. Acute climate occasions have brought on more than $1 trillion in destruction since 1980 in the US alone. It's an amazing figure, however not almost as stunning as the loss of life related with these occasions-around 10,000 lives. The forecast of low-probability, high-impact occasions, for example, droughts, tornados, and storms and so forth, has demonstrated to have significant monetary and social effects with regards to restricting or avoiding mass property harms and sparing human lives. Be that as it may, paying little respect to the point, foreseeing climate has dependably been a dubious business. Some of the most common methods of predicting weather are persistence ie., using a barometer, regression based statistical calculations. But they regrets border or unexpected events. Most of these methods are deeply flawed with respect to the fact that they produce results with high index for error. This causes the outcomes to vary from the actual observed weather conditions. Some of these methods can only predict weather for a short period and is not beneficial to use in today's world. Also, with the development of IoT, data integration plays a crucial role in developing an efficient weather prediction system. Because of considerable advances in innovation, it is presently conceivable to figure the climate days and even months ahead of time which was not really conceivable before the mid-twentieth century. The utilization of PC models got to be distinctly broad principally all through the 1960s, as the main climate satellites were pioneered. The variants of PC models that are utilized as a part of estimates depend generally on the kind of atmosphere and climate conditions. Considering the aforesaid aspects, this paper overcomes the disadvantages of existing weather forecast system, giving results with reduced error index. Moreover, the accuracy of the obtained results can be increased by aggregating the amount of data.

2. Literature Survey

Culclasure states that neural network can be used to precisely predict weather even with noisy dataset. The paper concludes that RPROP and PSO can predict minimum temperature very accurately up to 24 hours regardless of dataset used. For classifying freezing event RBF performed very well whereas RPROP and PSO did not perform well. While classifying gusty events RBF neural network again outperformed RPROP and PSO with HSS (Heidke Skill Score score) above 0.6. It infers cross validation is a strong method for verification but it is not as good as real world validation [1].According to Mailk et al., back propagation neural network can approximate a lot of functions and neural network can be more accurate than numerical differentiation method used for weather forecasting using models. The paper compares two optimization algorithms gradient descent and Levenberg–Marquardt algorithm and they found out that gradient descent is slow and increasing the learning rate to speed gradient descent will result in destabilizing it. But LM algorithm is faster than gradient descent perhaps the fastest among other forecasting methods. Along with it Levenberg–Marquardt has better learning rate as compared to other algorithm [2]. Narvekar and Fargose states that compared to ensemble j ANN with back propagation algorithm performs the best with minimal error. They recommend ANN with back propagation for weather forecasting as it iteratively compares output of neural network and desired out and correct itself during training. As ANN also accepts complex parameters and approximates the output and it uses the same pattern for forecast [3]. Liang et al., states that the basic data mining operations and numerical methods are employed to get a useful pattern from a huge volume of data set.

They collected data sets of a particular region weather report from 1901 to 2001 with 11 attributes. The collected datasets undergo pre-processing. Then clustering operation, Curve fitting and extrapolation methods are applied, proceeding with back propagation. The back propagation and Extrapolation results are compared. They conclude that the method used is reliable as they tested it on the data from 2001 to 2005 and got satisfactory results. Their paper also shows that increasing hidden layers can increase the accuracy but more than a limit will reduce the accuracy. They got a 95% accuracy for the collected dataset and states that neural network are good at mapping nonlinear function [4]. Fazel et al., talks about building long-range systems for weather forecast conditions in India. There are mentions of various comparative studies which compare ANN to other trivial methods. The former was trained to be capable of detecting even minor climatic changes .The methods RBFN and BPN were the most commonly used methods by the researchers to predict the usefulness of artificial neural networks. Studies are focused on the systematic development of improved physics representations, which are suitable for climate prediction .It also describes the CWRF methodology in detail. This involves how CWRF analyzes climate changes and generates predictions at regional-local scales [6]. Kim et al., paper talks specifically about flood forecasts where there is heavy data contamination due to noise and other factors .In flood forecasting, the two most commonly used techniques are Root Mean Square (RMSE) and bayesian regularization .It also describes how to use the aforesaid methods with a mathematical model .In conclusion, it is observed that techniques using machine learning are robust enough to apply to many problems in the given domain [7]. Kiskac and Yardimci analyzes sensitivity of real-time storm surge forecasts .The input parameters are as follows: the wind speed, sea surface level, its depression rate, wind direction, sea level pressure and the typhoon position. Several datasets have been collected and have been analyzed and the results suggest that those obtained with neural networks with the input parameters are very close to the observed ones .This shows that NN are very effective in predicting rapid climate behaviors [8]. Domanska and Wojtylak have used Fuzzy logic and Artificial Neural networks to predict the weather. Hybrid weather forecast system in which data mining and data fusion has been used. They have used automated inputs and not human manual ones. This will give not a 100% precise result but the result will be more than satisfying [9]. Salim et al., have used Fuzzy logic to calculate the weather forecast. They have taken certain assumptions and have calculated to a certain extent of accuracy. The data was taken from COSMO LM model and the results have turned out to be smaller as compared to the original situation. In our point of view, they have done a very clean project with minute differences [10]. Li et al., have predicted the weather using a 'NOW CAST' system. The scope is really narrow for their paper. The research done does not even match the level of the Meteorological department of U.K .It delivers the summaries of the forecast done in a local context. Their paper gives us the accurate forecast of a particular locality and not of a whole region which is an advantage to the MET Department of U.K. On the whole, they offer a very particular weather forecast [11]. Hung et al., proposed the ANN advances in information technology to create a bridge between workflow systems and dynamic responses required for weather events. The paper proposes a model and also an algorithm to bridge the aforementioned gap. The algorithm uses myriad of observational data to stream mine data by classifying into storm patterns in response to unusual weather behavior. This has been modeled in the LEAD-CI project [12]. Fahlman and Lebiere showcases that the weather reports rarely publish the uncertainty in the forecast. This is partly because weather forecast uncertainty is probabilistic coupled with the research evidence that people understand and relate to frequencies better than probability. Susan et al conduct tests to check whether showcasing uncertainty as a frequency garnered it any edge over showcasing it as a probability. Results show that frequency doesn't get any edge albeit it does get an insignificant edge when proposed with a reference class [13]. Balázs presents the possibilities of ensemble forecasting for precipitation over China and comparing them to the operational suite of NCARs TIGGE for THORPEX program. The super ensemble strategy has resulted in accurate predictions the medium range. There is also a possibility of the strategy to have very high skill in long ranges [14].

3. Background Study

An Artificial Neural Network (ANN) is a data preparing model that is enlivened by the working of human sensory systems. For example, the cerebrum and it's generated stimuli responses. The key component of this model is the novel structure of the data processing framework. It is made out of countless interconnected handling components (neurons) working as one to tackle particular issues. An ANN is arranged for a particular application, such as data classification or pattern recognition, through a learning procedure. Learning in natural frameworks includes acclimations to the synaptic associations that exist between the neurons. ANN's are generally unrefined electronic models inspired by the neural structure of the mind. The cerebrum fundamentally gains from experience. This artificial cerebrum model additionally guarantees a less specialized approach to create machine models. This new way deals with giving a more elegant degradation amid framework overload than its more conventional counterparts. These biologically roused techniques for processing are thought to be the following real progression in the computing industry.

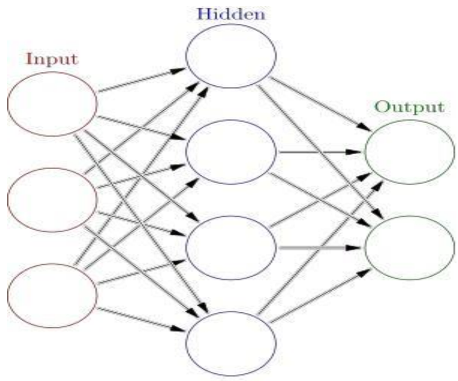

Figure 1: Basic Artificial Neural Network

Simple Predictive Algorithm

Despite the fact that there are helpful systems which contain just a single layer, or even one component, most applications require systems that contain at any rate the three ordinary sorts of layers-input, hidden, and output. The input layer of neurons gets the information either from input records or directly from electronic sensors in dynamic applications. The output layer sends data specifically to the outside world, to an optional PC handle, or to different gadgets, for example, a mechanical control framework. Between these two layers can be many shrouded layers. These inner layers contain a hefty portion of the neurons in different interconnected structures. The sources of outputs and inputs of each of these shrouded neurons just go to different neurons. In many systems every neuron in the hidden layer gets the signals from the majority of the neurons in a layer above it, normally an input layer. There are two sorts of these associations: One causes the summing component of the next neuron to include while alternate causes it to subtract. In more human terms one energizes while alternate hinders.

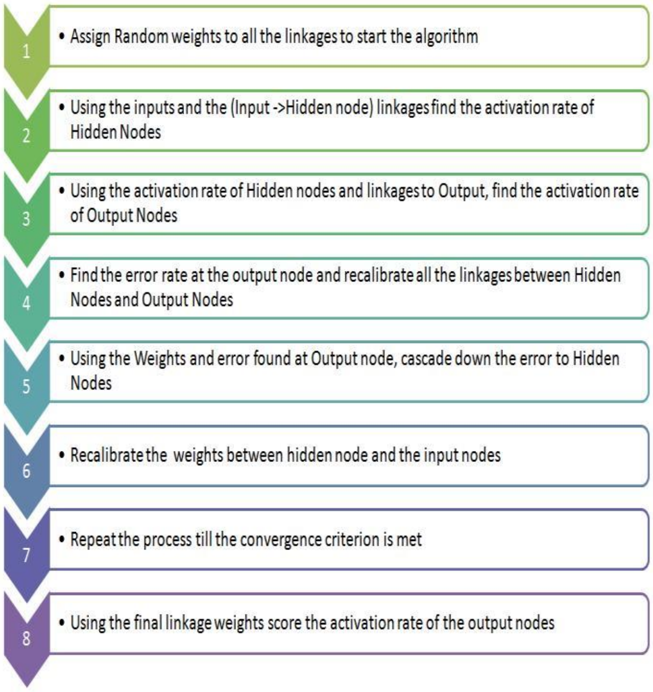

Figure 2: Neural Network Process Steps

Training a Neural Network

Once a network has been organized for a specific application, it is prepared to be trained using training data sets. To begin this handle the underlying weights are picked in a randomized manner. At that point, the learning, or training, starts. There are two ways to deal with training - supervised and unsupervised.

A. Supervised Learning

Supervised Training includes a component of furnishing the system with the wanted output either by manually reviewing the system's execution or by furnishing the coveted outputs with the inputs. In supervised training, both the sources of output and input are given. Then the network processes the data sources furthermore, thinks about its subsequent outputs against the coveted outputs. Erroneous data are then proliferated back through the framework, bringing about the network to alter the weights which control the network of nodes. This procedure happens again and again as the weights are constantly changed. The arrangement of information which empowers the training is known as the training set.

B. Unsupervised Learning

Unsupervised learning is the state the network needs to understand the contributions without outside assistance. The immeasurable mass of systems use supervised learning. Unsupervised learning is utilized to play out some underlying portrayal on information sources. Be that as it may, in the out and out feeling of being really self-learning, it is still only a sparkling guarantee that is not completely comprehended, does not totally work, and consequently is consigned to the lab.

Back Propagation Algorithm

Back propagation, also known as back propagation of errors, is a typical strategy for training artificial neural networks and utilized as a part of conjunction with an optimization technique, for example, gradient descent. The algorithm is designed to repeat a two phase cycle, weight update and propagation. At the point when an information vector is displayed to the network, it is propagated through the network, forward, layer by layer, until it achieves the output layer, as shown in Figure 3. The network output is then compared and contrasted with the coveted output, by using a loss function, and the measure of error computed for each of the neurons in the output layer. The error qualities are then propagated in reverse, beginning from the output layer, until every neuron has a related error value which generally reflects its contribution to its commitment to the first output.

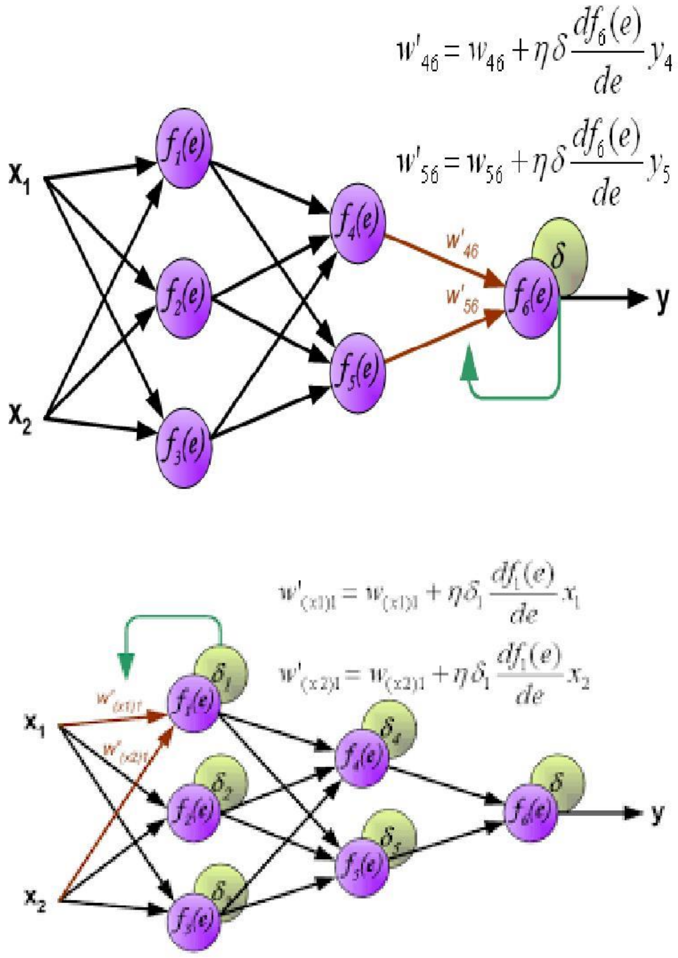

Figure 3: Back Propagation in Neural Network

Loss Function

Softmax function is used in neural networks to find the loss between the predicted output and the expected output. Softmax function calculates the probabilistic distributions among K different outcomes. This is used to calculate the probability of the output of neural network for each category during classification. The softmax function is used in various classification methods such as Naïve Bayes classifiers, artificial neural network and linear discriminant analysis.

$$\sigma(z)_j = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}} \qquad (1)$$

For j = 1,.. K ; Z=XW

Root mean square is the basic loss function widely used in neural networks. It is simple square root of arithmetic mean of squares of the given value. In neural network we find the root mean square of difference between expected output and the output received from the neural network. In the case of a set of *n* values {x1,x2…xn}, the RMS

$$\sqrt{\frac{\sum_{i=1}^{n} xi^{2}}{n}} \qquad (2)$$

Heidke Skill Score

The Heidke Skill Score is used to measure the skill in forecasting. In this paper, the neural network is evaluated on the basis of The Heidke Skill Score. It is defined as,

$$\frac{(NC - E)}{(T - E)} \qquad (3)$$

Where NC is the total number of correct predictions, T is the total number of test cases and E is the number of forecasts expectations and probabilistic occurrence. It can be calculated using Heidke Skill Score table.

Table 1: Heidke Skill Score Table

Observed Category		**Forecast Category**				
	1	2	…	m		Total
1	X11	X12	…	X1m		X1p
2	X21	X22	…	X1m		X2p
…	…	…	…	...		…
m	Xm1	Xm2	…	X1m		Xmp
Total	Xp1	Xp2	…	X1m		Xpp

Where ‘m’ is the total number of classes or categories to be predicted. NC, T and E are calculated as follows.

$$NC = \sum X_{ii} \qquad (4)$$

For i from 1 to m

$$T = X_{pp} \qquad (5)$$

$$E = \frac{\sum (X_{ip} X_{pi})}{T} \qquad (6)$$

A negative HSS indicates it is worse that a randomly generated result.

HSS = 0 means that one third of total predictions are true.

HSS = 0.5 means two third of all the predictions are true.

The perfect HSS result is equal to 1, which means all the predictions are correct.

Cascade Correlation Neural Networks

The biggest drawback of back propagation neural network is not knowing the correct architecture for neural network. Cascade correlation overcomes this problem by dynamically adding hidden node between input and output layers. Cascade correlation adds hidden nodes only when it helps in convergence. So, the architecture contains optimal number of nodes in the neural network such that addition of another node will not affect the output of the neural network. This reduces the overhead in the calculation during training.

4. Methodology

Dataset used for training and testing neural network was taken from http://tawn.tnau.ac.in for everyday during the time frame of 1997 to 2015. It contains data about temperature, dew point, humidity, sea level visibility, wind speed, precipitation and events. Event field contains the weather summary of the particular day and neural network is trained to predict it. The dataset was preprocessed to fill missing values and remove missing records but some of the noisy data was kept in the dataset to check networks performance, then feed into the neural network.

At the beginning the neural network only contains input and output layers. This neural network is trained using simple gradient descent algorithm applied for each output node individually. After this network achieves convergence, a new node is added to the network such that it is connected to all the input nodes. Now we freeze all the weights except the neurons connected to the new node that was added. This new network is trained till it achieves convergence. Then the output of this new network is compared to the old configuration, if the accuracy is increased then the newly (specified) added node is kept in the network and we repeat this process till accuracy remains same or desired accuracy is obtained [13].

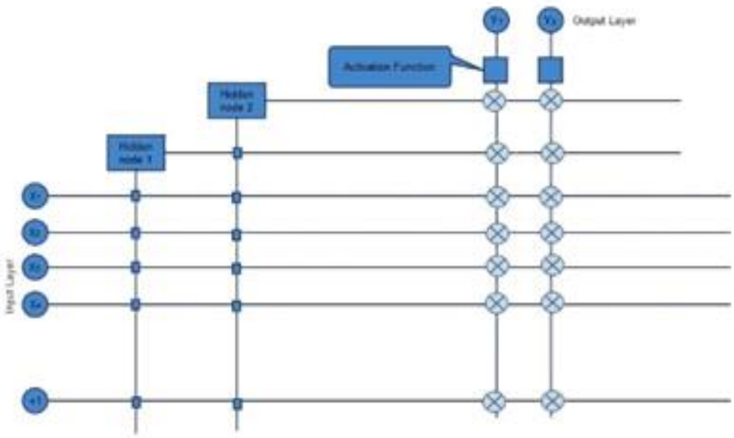

Figure 4: Cascade Correlation Neural Networks

Adding Node

Addition of new node can divided into two steps. First a new node is connected to the input layer that is it is connected to all the input nodes and previously added hidden nodes, the new node is not connected with output layer yet. Rest of the neural network's weight is frozen i.e. weights are not changed during back propagation.

Secondly the weights of neurons that are connected with the new node is trained. The goal of training is to maximize the co-variance(C) between output of network and newly(modified) added node with respect to the weights of the newly (modified) added node.

$$c = \sum_{o\in O}\left|\sum_{s\in S}(y_s - \bar{y})(e_{o,s} - \bar{e}_o)\right| \tag{7}$$

Here, S is the set of training samples, ys is the candidate's output for the sample s, eo,s is the error at output neuron o for the sample s and y, eo are the averaged values of ys and eo, s respectively over all the samples s ∈ S. This maximization can be done by the gradient ascent algorithm [14].

We can also use other measurements instead of co-variance like correlation or simple root mean square error. First, dataset was cleaned by removing missing records and filling missing values. The cascading neural network was feed with seventy percent of dataset for training. After the network achieved desired accuracy training will be stopped. Then the network was tested with thirty percent of dataset which was not used for training.

5. Results

Around 200 set of input was feed to the neural network over a 100 epochs with learning rate of 0.2.

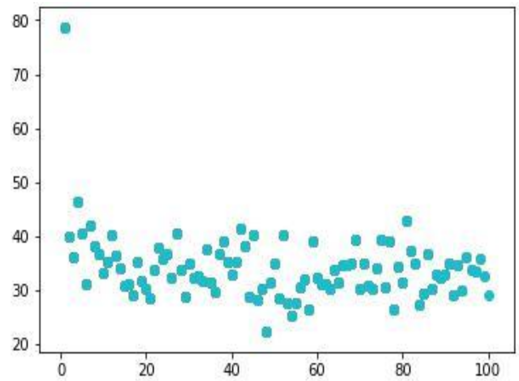

Figure 5: Loss During Training

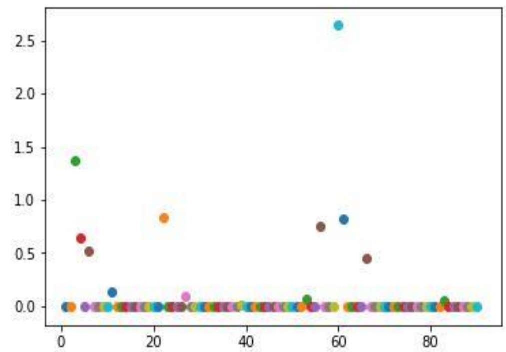

Figure 6: Loss During Testing

The graphs (Figure 5,6) above shows the loss of the neural network over 100 epochs with dropout. The network was then given a testing sample of 90 set of input which were not used in training of the network. Table 2 gives the recorded loss during testing. The network got 75 correct outputs and 15 outputs as wrongly** classified with an accuracy of 83.33%. **=can be changed into more sophisticated word

Table 2: Heidke Skill Score Table of the Network

			Forecast Category				
Observed	1	2	3	4		5	**Total**
Category							
1	16	0	2	0		0	18
2	1	14	6	0		0	21
3	0	0	14	0		0	14
4	1	0	0	14		1	16
5	0	1	0	3		17	21
Total	18	15	22	12		18	90

The network's performance is shown in Heidke Skill Score table. The network achieved a HSS score of 0.79523 with, E = 16.465, N = 75,T = 90 and HSS = 0.79523

6. Conclusion

The cascading neural network is very useful to mine results from a large data set. This also applies on the accuracy of the results we obtained by mining the forecast data set i.e. 83.33% irrespective of data size. The time taken to predict weather given the parameters such as temperature, dew point, humidity, sea level, wind speed, gust, visibility, cloud coverage, wind degrees and precipitation is quite less comparable to other neural network. The working of model is not a multiplex and the fact of estimation is acknowledged by straightforward arithmetic operations. The results asserts that ANN models may go about as an association between isolated exploratory information and prompted collaboration between the self-standing studies.

References

[1] Culclasure A., Using neural networks to provide local weather forecasts, (2013).

[2] Malik P., Singh S., Arora B., An effective weather forecasting using neural network, Int J Emerg Eng Res Technol., 2 (2) (2014), 209-212.

[3] Narvekar M., Fargose P., Daily weather forecasting using artificial neural network, International Journal of Computer Applications 121(22) (2015).

[4] Shrivastava G., Karmakar S., Kowar M.K., Guhathakurta P., Application of artificial neural networks in weather forecasting: a comprehensive literature review, International Journal of Computer Applications 51(18) (2012).

[5] Liang X.Z., Xu M., Yuan X., Ling T., Choi H.I., Zhang F., Chen L., Liu S., Su S., Qiao F., He Y., Regional climate–weather research and forecasting model, Bulletin of the American Meteorological Society 93 (9) (2012), 1363-1387.

[6] Fazel S.A.A., Mirfenderesk H., Tomlinson R., Blumenstein M., Towards robust flood forecasts using neural networks, IEEE International Joint Conference on Neural Networks (IJCNN) (2015), 1-6.

[7] Kim S.Y., Shiozaki S., Matsumi Y., Ota, T., A study of a real-time storm surge forecast system using a neural network at the Sanin Coast, Japan, In Oceans (2012), 1-7.

[8] Kiskac B., Yardimci H., Weather prediction expert system approaches, Ceng-568 Survey, (2004).

[9] Domanska D., Wojtylak M., Fuzzy weather forecast in forecasting pollution concentrations, In Proc. of International Conference on Chaotic Modeling and Simulation (2010), 1-8.

[10] Salim M., Best P.J., Carpenter G.F., September. Software engineering in the design of a weather-now cast system, IET Second International Conference on Software Engineering for Real Time Systems (1989), 146-150.

[11] Li X., Plale B., Vijayakumar N., Ramachandran R., Graves S., Conover H., Real-time storm detection and weather forecast activation through data mining and events processing, Earth Science Informatics 1(2) (2008), 49-57.

[12] Hung N.Q., Babel M.S., Weesakul S., Tripathi N.K., An artificial neural network model for rainfall forecasting in Bangkok, Thailand, Hydrology and Earth System Sciences 13 (8) (2009), 1413-1425.

[13] Fahlman S.E., Lebiere C., The cascade-correlation learning architecture, In Advances in neural information processing systems (1990), 524-532.

[14] Balázs G., Cascade-Correlation Neural Networks: A Survey, Department of Computing Science, University of Alberta, Edmonton, Canada (2009), 1-6.

[15] Solanki N., Panchal G., A Novel Machine Learning Based Approach for Rainfall Prediction. International Conference on Information and Communication Technology for Intelligent Systems. Springer, Cham, (2017), 314-319.

[16] Abbot J., Marohasy J., Skilful rainfall forecasts from artificial neural networks with long duration series and single-month optimization, Atmospheric Research (2017), 289-299.

[17] Ghasemieh H., Bazrafshan O., Bakhshayesh Manesh K., Artificial Neural Network for Monthly Rainfall Forecasting Using Tele connection Patterns (Case Study: Central Plateau Basin of Iran) Journal of the Earth and Space Physics 43(2) (2017), 405-418,

[18] Lavanya K., Durai M.A.S., Iyengar N.C.S.N., Site Specific Soil Fertility Ranking and Seasonal Paddy Variety Selection: An Intuitionistic Fuzzy Rough Set and Fuzzy Bayesian Based Decision Model, International Journal of Multimedia and Ubiquitous Engineering 10 (6) (2015), 311-328.