International Journal of Recent Trends in Engineering & Research
ISSN(ONLINE) : 2455-1457
IMPACT FACTOR : 3.344

Weather Prediction by Mass Neural Networks through an Integrated Model approach

Dr. B S Pradeep[1], Naveen L[2]
[1]*International R&D dept., Linyi Top Network Co. Ltd*
[2]*Dept. of ISE, Rajarajeswari College of Engineering*

Abstract- The active and disordered nature of weather makes weather forecasting a challenging and provocative task. Various numerical models have been developed and applied for this purpose, however usually it's difficult to conclude accurate predictions. Although artificial neural networks have been considerably applied for weather forecasting, even then marking the precise result was still a question! Consequently some researchers proposed to use Cluster models of neural networks for the prediction task. When considering multiple neural networks, the redundancy caused by having multiple models and also combining the results of different networks are still the main challenges. In this project, a new Composite model is proposed for weather forecasting based on the Cluster of neural networks. Here, the redundancy issue has been addressed by introducing an integrated model in which a feature selection module is first applied to the data. Introduction of a Mutual information approach to tackle the challenge of combining the results of different networks and reducing the redundancy in the Composite model. The evaluation of result presents an outperformance of the proposed method compared to previous work.

Keywords- Classification, Cluster Neural Network; Modular Neural Network; Mutual Information; Weather prediction

I. INTRODUCTION

Weather Forecasting is the use of science and technology simultaneously in order to predict weather for a specific location. Because of the atmosphere conditions are dynamic and disorder in nature, forecasting results are not usually precise. Hence Neural networks (NN) are one of the most powerful artificial intelligence methods that are used for nonlinear modeling and particularly for weather forecasting. Many researchers have applied different types of neural networks in the field of weather forecasting, and back propagation algorithm [6] is utilized for training. However, this algorithm is a gradient-based method that often gets stuck in local minima. The output of such practice is the weighted sum of each network, then concluding that Radial Basis Function (RBF) network could provide the best result [11], and Hopfield network had the poor outcome such as Multi-Layered Perceptron (MLP) [3] or (RBF) for predicting the weather.

At macro level, weather forecasting is usually done using the data gathered by remote sensing satellites. Weather parameters like maximum temperature, minimum temperature, level of rainfall, cloud conditions, wind streams and their directions, are projected using images taken by these meteorological satellites to assess future trends. The satellite-based systems are inherently expensive and require complete support system. The variables defining weather conditions like temperature, relative humidity, rainfall etc., vary constantly with time, forming time series of each parameter and can be used to develop a forecasting model either statistically or using some other means that uses this time series data. The true power and advantages of neural networks lies in the ability to represent both linear and nonlinear relationships directly from the data being modeled.

Traditional linear models are simply inadequate when it comes for true modeling data that contains nonlinear characteristics. A neural network model is a structure that can be adjusted to produce a mapping from a given set of data to features of or relationships among the data. The model is adjusted, or trained, using a collection of data from a given source as input. After successful training, the neural network will be capable to perform classification, estimation, prediction, or simulation on new data from the same or similar sources. An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the new structure of the information processing system. It is composed of a huge number of highly interconnected processing elements working in accord to solve specific problems.

II. PRPOSED MODEL

The proposed model for temperature prediction is composed of three different stages. At the first stage data are normalized and useful features are selected based on stepwise regression. At the second stage, it's proposed to have a Cluster of four different neural networks each of which is trained based on the data provided by the first stage. Then at the third stage, a mutual information approach in order to select the best networks and with that, the results of the selected networks are combined.

Data Preprocessing- The representation and quality of data is done first before consecutively an analysis in data preprocessing. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection etc. The product of data pre-processing is the final training set.

Normalization- Changes in expression are independent of abundance. Rare transcripts are as likely to change in response to a given stress as common ones. The probe-level data must be cleaned and processed to obtain biologically meaningful measurement. Background correction: eliminate signals due to non-specific binding. Some of the assumptions can be made as follows:

- Most transcripts are not differentially expressed in response to a given stress.
- Expression ratio of typical spot: tumor/control= 1.
- Range of abundance begins at 0. "Negative transcripts" = error measurement. Outliers are biologically relevant.
- Lowess is a technique for fitting a smoothing curve to a dataset. An intensity dependent normalization. Predicted lowess value is subtracted from the data to decrease the standard deviation and place the mean log ratio at 0.
- All samples in the dataset are corrected independently.
- Lowess normalization can be applied to complete or incomplete datasets.

Feature Selection Methods- The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. In statistics, the most popular form of feature selection is stepwise regression, which is a wrapper technique. It is a greedy algorithm that adds the best feature (or deletes the worst feature) at each round. The main control issue is deciding when to stop the algorithm. In machine learning, this is typically done by cross-validation. In statistics, some criteria are optimized. This leads to the inherent problem of nesting. More robust methods have been explored, such as branch and bound and piecewise linear network.

Feature selection is often an essential data preprocessing step prior to applying a classification algorithm such as Multi-Layer Perception (MLP). The term feature selection is taken to refer to algorithms that output a subset of the input feature set. One factor that plagues classification algorithms is the quality of the data. If information is irrelevant or redundant or the data is noisy and unreliable then knowledge discovery using training is more difficult. Regardless of whether a learner attempts to select features itself or ignores the noise, feature selection prior to learning can be beneficial. Reducing the dimensionality of the data reduces the size of the hypothesis space and allows algorithm to operate faster and more effectively. Figure 1 shows the composite model of the cluster neural network.

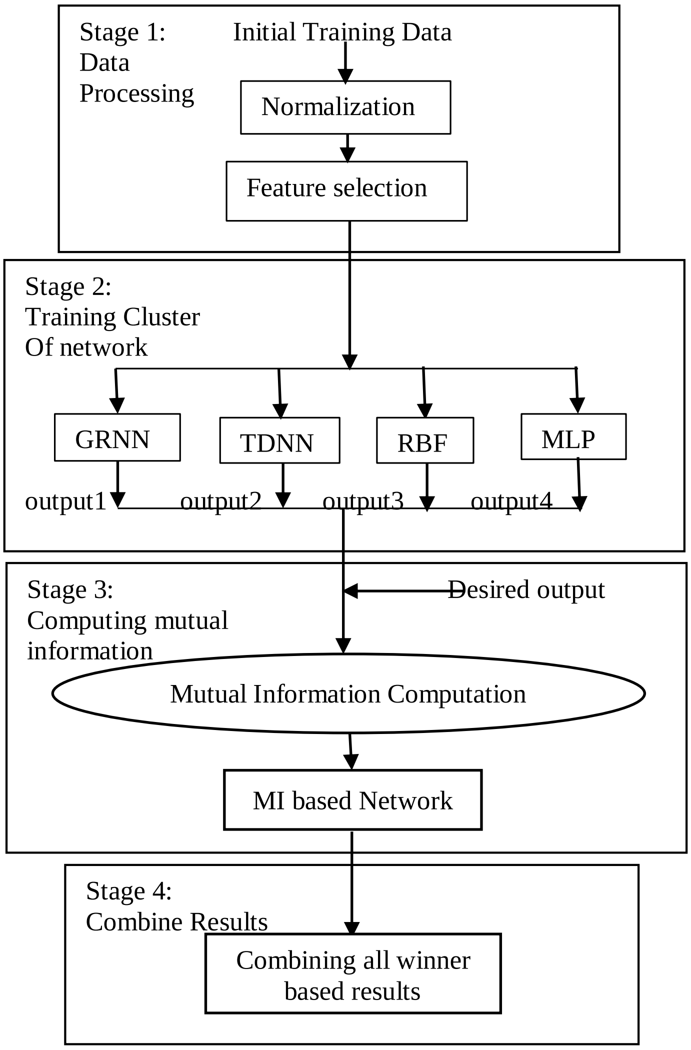

Figure 1. Composite Model using Cluster of Neural Networks

III. METHODOLOGY

An Artificial Neural Network (ANN) is a system that can be used to perform classifications on datasets. The ANN can access a solution space that is unavailable to computational systems based on logical inferences. They have the ability to relate n-inputs to m-output classes and operate on classification problems which are non-linear. Given a series of continuously variable input values, ANNs are able to place each case into a particular output class. If the input variables are represented by n, the output classes by m, then n > m where n and m are members of the set of real numbers. The Methodology for proposed Composite Model involves Classification, Cluster of Neural Networks, and Mutual Information

3.1. Classification

Classification involves four types: General Regression Neural Network (GRNN), Time Delay neural Network (TDNN), Radial Basis Function (RBF), and Multi-layer Perceptron (MLP).

Generalized Regression Neural Network- GRNN is a variation of the radial basis neural networks, which is based on kernel regression networks. A GRNN does not require an iterative training procedure as back propagation networks. It approximates any arbitrary function between input and output vectors, drawing the function estimate directly from the training data. In addition, it is consistent that as the training set size becomes large, the estimation error approaches zero. A GRNN consists of input layer, pattern layer, summation layer and output layer. The number of input units in input layer depends on the total number of the observation parameters. The first layer is connected to the pattern layer in which each neuron presents a training pattern and its output. The pattern layer is connected to the summation layer. The summation and output layer together perform a normalization of output set. In training of network, radial basis and linear activation functions are used in hidden and output layers. Each pattern layer unit is connected to the two neurons in the summation layer, S and D summation neurons. S summation neuron computes the sum of weighted responses of the pattern layer. On the other hand, D summation neuron is used to calculate un-weighted outputs of pattern neurons. The Generalized Regression Network consists of three layers of nodes with entirely different roles: The input layer, where the inputs are applied, the hidden layer, where a nonlinear transformation is applied on the data from the input space to the hidden space, the linear output layer, where the outputs are produced. GRNN a very useful tool to perform predictions and comparisons of system performance in practice.

Time Delay Neural Network- TDNN is an artificial neural network architecture whose primary purpose is to work on sequential data. The TDNN units recognize features independent of time-shift and usually form part of a larger pattern recognition system, converting continuous audio into a stream of classified phoneme labels for speech recognition. An input signal is augmented with delayed copies as other inputs, the neural network is time-shift invariant since it has no internal state. The basic model presented a perceptron network whose connection weights were trained with the back-propagation algorithm, this may be done in batch or online. The TDNN architecture addresses both problems by imposing certain restrictions on the network topology and by the way in which weights are updated. Hidden units are connected to a limited number of input units that represent a consecutive pattern in the input window. These hidden units have a receptive field that are only sensitive to a part of the input window. The positions and a receptive field covers a subsequence of three positions, there must be eight hidden units with the same receptive field. Since the corresponding weights in all copies of a receptive field are forced to have the same values, these hidden units are said to have linked receptive fields.

Radial basis function neural network: In the field of mathematical modeling, a radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction classification, and system control. Radial basis functions are powerful techniques for interpolation in multidimensional space. RBFs have been applied in the area of neural networks where they may be used as a replacement for the sigmoid hidden layer transfer characteristic in multi-layer perceptron. RBF networks have two layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. The RBF chosen is usually a Gaussian. Like Gaussian Processes, and unlike SVMs, RBF networks are typically trained in a Maximum Likelihood framework by maximizing the probability of the data under the model. SVMs take a different approach to avoiding over fitting by maximizing instead a margin. RBF networks are outperformed in most classification applications by SVMs. In regression applications they can be competitive when the dimensionality of the input space is relatively small.

Multilayer perceptron: Typically, one unit model is placed at each node of the network & interconnections run between units. The network model is operated by a training algorithm, which itself has a number of associated variables. The network used in this development employs the back propagation technique for training. This network is a feed forward multi-layer network. It should contain, 1 Input Layer of Units, At least 1 Hidden Layer of Units, and 1 Output Layer of Units. It is important to note that weights are assigned to the interconnections between each layer in the network. The back propagation technique is employed as a means of reducing error in the network's classification, by initially calculating this error and the propagating it back through the network for reduction. Inputs are propagated to the first layer of hidden units, whose output is calculated and propagated to the next hidden later. This process is repeated until the output layer is reached. Each output layer unit calculates the activation, from the sum of weighted inputs from previous layers. The error on the initial output is computed and propagated back to the first hidden layer, where the weight matrix is updated. A multilayer perceptron is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron with a nonlinear activation function. MLP utilizes back propagation for training the network. MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable. The multilayer perceptron consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes and is thus considered a deep neural network. Each node in one layer connects with a certain weight Wij to every node in the following layer. Some people do not include the input layer when counting the number of layers and there is disagreement about whether Wij should be interpreted as the weight from i to j or the other way around. We represent the error in output j node in the nth data point (training example) by $e_j(n) = d_j(n) - y_j(n)$, where d is the target value and y is the value produced by the perceptron.

3.2. Cluster of neural networks

Two main categories exist in combining neural networks: modular networks and Cluster of networks. In modular network the problem is divided into different parts and each part is solved by one module, while in Cluster of networks, a collection of independent networks is considered and each of them learns and solves the whole problem independently. The proposed model contains Cluster of four neural networks: MLP, RBF, general regression neural network and time delay neural network. MLP, which is a feed forward neural network, uses the back-propagation algorithm for training. RBF is also a feed forward network and a combination of supervised and unsupervised methods is used in the training task. GRNN is a four-layered neural network and it usually works better that MLP while at the same time, it can be trained faster only within a single iteration.

3.3. Mutual information

Mutual information (MI) is a linear value that is used for measuring correlation between two linear or nonlinear random variables. In fact, mutual information between two random variables is the knowledge provided by one of them that reduces uncertainty about the other. MI, I(X;Y), showing the amount of information being shared by X and Y, is defined as

$$\mathbf{I(X:Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}} \qquad \textbf{(1)}$$

Based on this definition, if X and Y are closely related, then I(X; Y) will be large, otherwise when X and Y are independent, I(X, Y) will be zero [9]. In fact, mutual information between two random variables is the knowledge provided by one of them that reduces uncertainty about the other. We apply mutual information to calculate the correlation between the output of each network and the desired output. Then, the networks that have produced an output closer to the desired one are selected. The outputs of the chosen networks are then combined together at the last stage of the model based on the Winner-Takes-All strategy. We apply mutual information to calculate the correlation between the output of each network and the desired output. Then, the networks that have produced an output closer to the desired one are selected. The outputs of the chosen networks are then combined together at the last stage of the model based on the Winner-Takes-All strategy.

IV. RESULTS AND DISCUSSION

The numerical data used in this research consists of six different features of the atmosphere and they are Dew Point, Humidity, Sea Level Pressure (hPa), Visibility (Km), Wind Speed (Km/h) and Precipitation (mm). Data are divided into three datasets: one for training, one for validation, and one as test data to report the error of the proposed model. By performing the feature selection at the first stage of the model, four features out of six features are selected, which are: Dew Point, Humidity, Sea Level Pressure and Visibility. The validation set consists of 366 data samples. At the second stage of the model, each of the four networks in the model is trained based on the training data. Then at the third stage, the mutual information between the actual output (real temperature) and the output of each network is calculated, as shown in Table I.

Table 1. MI between the output of each network & the actual output.

MI(MLP)	MI(RBF)	MI(TDN)	MI(GRNN)
0.12	2	0.1	1

Figure 2 shows comparision between proposed model and the original temperature. In order to validate the proposed work with other types of models in the literature, five different methods are chosen to compare with: GRNN, RBF, and three methods based on the Clusters of networks as presented in [2, 3]. We can note from the below figure that proposed temperature is almost equivalent to that of actual one.

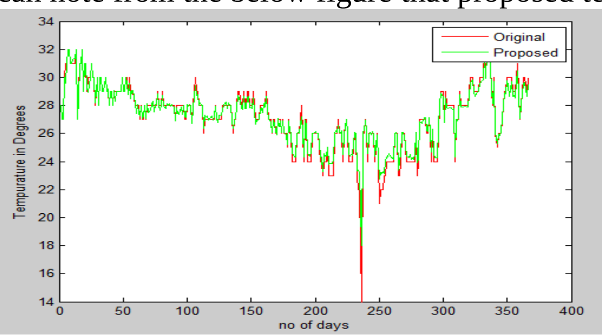

Figure 2. Estimated temperature based on proposed Method

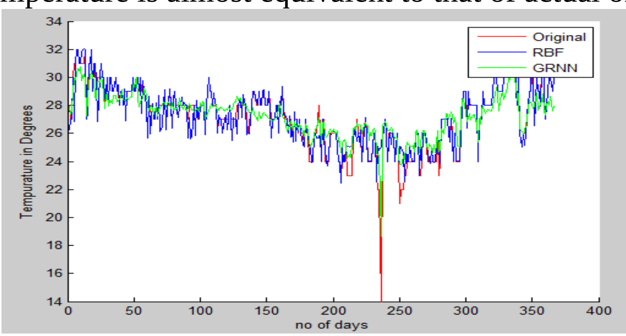

Figure 3. Original temperature based on RBF, GRNN

In Figure 3 the output of GRNN, RBF and the proposed model are compared with the actual real temperature, based on the validation dataset.

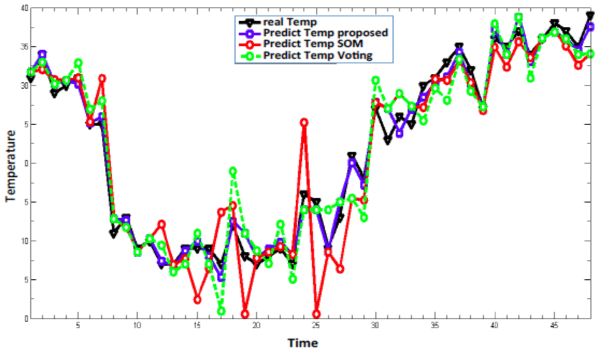

Figure 4. Estimated temperature based on the proposed method as compared to SOM and voting scheme in [2]

The combined network result as in [2] for two different methods are presented in figure 4 in terms of validation set as compared: the first one is based on a competitive network like SOM, and the second one is based on a voting scheme. Table 2 shows mean-squared error (MSE) between the actual temperature and the estimated one for different methods, based on the testing data. As we can see from the results in this table, the proposed model can provide better results comparing with other types of methods in terms of MSE.

Table 2. Comparing the performance of the proposed method with other methods in terms of Mean Square Error(MSE)

Prediction Method	MSE
Proposed Method	0.17
RBF	0.99
GRNN	1.114
SOM	1.23

V. CONCLUSION

The comparison results in terms of the validation set are presented with the outcome of the research work. The proposed model is also an integrated model since a feature selection module is considered at the first stage of the model. In addition, to reduce the redundancy caused by considering Cluster of networks, and also decrease the complexity, it's proposed to calculate the correlations between the output of each network and the actual result. A comparison between the results of the proposed model and some similar models in the literature shows an outperformance of the proposed model.

REFERENCES

1. AJWM tenBerg and Lambert Spaanenburg, "Modular and hierarchical specialization in neural networks,"IU Journal of Electrical & Electronics Engineering, vol. 3, no. 1, pp.719–726, 2003.
2. Gasser Auda and Mohamed Kamel, "Modular neural networks: a survey", International Journal of NeuralSystems,vol. 9, no. 02, pp. 129–151, 1999.

3. Imran Maqsood, Muhammad Riaz Khan, and Ajith Abraham, "An Cluster of neural networks for weatherforecasting," Neural Computing & Applications, vol. 13, no. 2, pp.112–122, 2004.
4. James W Taylor and Roberto Buizza, "Neural network load forecasting with weather Cluster predictions,"PowerSystems, IEEE Transactions on, vol. 17, no. 3, pp. 626–632, 2002.
5. Nguyen Q Hung, Mukand S Babel, S Weesakul, and NK Tripathi, "An artificial Neural network model forrainfall forecasting in bangkok, thailand.," Hydrology & Earth System Sciences, vol. 13, no. 8, 2009.
6. Pooja Malik, Prof. Saranjeet Singh, Binni Arora, "An Effective Weather Forecasting Using Neural Network",International Journal of Emerging Engineering Research & Technology Volume 2, Issue 2, May 2014, PP 209-212.
7. Roberto Battiti, "Using mutual information for selecting features in supervised neural net learning," NeuralNetworks, IEEE Transactions on, vol. 5, no. 4, pp. 537– 550,1994.
8. S Santhosh Baboo and I Kadar Shereef, "An efficient weather forecasting system using artificial neural network," International journal of environmental science and development, vol. 1, no. 4, pp. 2010–0264, 2010.
9. Sanjay Mathur Paras, Avinash Kumar, and Mahesh Chandra, "A feature based neural network model forweather forecasting," World Academy of Science, Engineering and Technology, vol. 34, pp. 66–73, 2007.
10. Shahrokh Asadi, Esmaeil Hadavandi, Farhad Mehmanpazir, and Mohammad Masoud Nakhostin, "Compositeizationof evolutionary levenberg–marquardt neural networks and data pre-processing for stockmarketprediction,"Knowledge-Based Systems, vol. 35, pp. 245–258, 2012.
11. Tiruvenkadam Santhanam and AC Subhajini, "An efficient weather forecasting System using radial basisfunctionneural network.," Journal of Computer Science, vol. 7, no. 7, 2011.
12. Tony Hall, Harold E Brooks, and Charles A Doswell III, "Precipitation forecasting Using a neural network.,"Weather& Forecasting, vol. 14, no. 3, 1999.
13. Zhi-Hua Zhou, JianxinWu, andWei Tang, "Ensembling neural networks: many could be better than all,"Artificial intelligence, vol. 137, no. 1, pp. 239–263, 2002.