An Efficient Clustering Method for Atmospheric Conditions Prediction using ART Algorithm

[1]Ankita Singh, [2]Dr. Bhupesh Gour, [1]Anshul khandelwal, [1]Harsha Lackhwani
[1] Students of Computer Sc. & Engg. Jai Narain College of Technology, Bhopal
[2] Head of Dept. of Computer Sc. & Engg. Jai Narain College of Technology, Bhopal

***Abstract*-Ambient air temperatures prediction is of a concern in environment, industry and agriculture. The increase of average temperature results in global warming. The aim of this research is to develop artificial neural network based clustering method for ambient atmospheric conditions prediction in Indian city. In this paper, we presented a clustering method that classifies cities based on atmospheric conditions like Temperature, Pressure and Humidity. Data representing month-wise atmospheric conditions are presented to Adaptive Resonance Theory Neural Network to form clusters which represents association in between two or more cities. Such associations predict atmospheric conditions of one city on the bases of another. ART based clustering method shows that the months of two cities which fall in the same cluster, represent similar atmospheric conditions in them.**

***Keywords*- Atmospheric conditions, artificial neural network, Adaptive Resonance Theory, clustering**

I. INTRODUCTION

Climate prediction is of a concern in environment, industry and agriculture. The climate change phenomenon is as the first environmental problem in the world threatening the human beings. The industrial activities are so effective in this problem and cause the global warming which the world has been faced with. The weather is a continuous, data-intensive, multidimensional, dynamic and chaotic process, and these properties make weather forecasting a formidable challenge. Knowing the variability of ambient temperature is important in agriculture because extreme changes in air temperature may cause damage to plants and animals. Air temperature forecasting is useful in knowing the probability of tornado, and flood occurrence in an area .Due to chaotic nature of the atmosphere, the massive computational power is required to solve the equations that describe the atmosphere, error involved in measuring the initial conditions, and an incomplete understanding of atmospheric processes. The use of ensembles and model helps narrow the error and pick the most likely outcome. Several steps to predict the temperature are

- Data collection(atmospheric pressure, temperature, wind speed and direction, humidity),
- Data assimilation and analysis,
- Numerical weather prediction,
- Model output post processing.

Advent of digital computers and development of data driven artificial intelligence approaches like Artificial Neural Networks (ANN) have helped in numerical prediction of atmospheric conditions.

ANNs provide a methodology for solving many types of non-linear problems that are difficult to solve by traditional techniques. Most meteorological processes often exhibit temporal and spatial variability, and are further plagued by issues of non-linearity of physical processes, conflicting spatial and temporal scale and uncertainty in parameter estimates. With ANNs, there exists the capability to extract the relationship between the inputs and outputs of a process, without the physics being explicitly provided. Thus, these properties of ANNs are well suited to the problem of weather forecasting under consideration.

One of the best artificial neural network approaches is ART (Adaptive Resonance Theory).ART structure is a neural network for cluster formation in an unsupervised learning domain. The adaptive resonance theory (ART) has been developed to avoid the stability-plasticity dilemma in competitive networks learning. The stability-plasticity dilemma addresses how a learning system can preserve its previously learned knowledge while keeping its ability to learn new patterns. ART architecture models can self-organize in real time producing stable recognition while getting input patterns beyond those originally stored.

This algorithm tries to fit each new input pattern in an existing class. If no matching class can be found, i.e., the distance between the new pattern and all existing classes exceeds some threshold, a new class is created containing the new pattern. The novelty in this approach is that the network is able to adapt to new incoming patterns, while the previous memory is not corrupted. In most neural networks, such as the back propagation network, all patterns must be taught sequentially; the teaching of a new pattern might corrupt the weights for all previously learned patterns. By changing the structure of the network rather than the weights, ART1 overcomes this problem.

II. RELATED WORK

Many works were done related to the temperature prediction system. They are summarized below.

Y.Radhika and M.Shashi presents an application of Support Vector Machines (SVMs) for weather prediction. Time series data of daily maximum temperature at location is analysed to predict the maximum temperature of the next day at that location based on the daily maximum temperatures for a span of previous n days referred to as order of the input. Performance of the system is observed over various spans of 2 to 10 days by using optimal values of the kernel.

Mohsen Hayati studied about Artificial Neural Network based on MLP was trained and tested using ten years (1996-2006) meteorological data. The results show that MLP network has the minimum forecasting error and can be considered as a good method to model the short-term temperature forecasting [STTF] systems.

Brian A. Smith focused on developing ANN models with reduced average prediction error by increasing the number of distinct observations used in training, adding additional input terms that describe the date of an observation, increasing the duration of prior weather data included in each observation, and re-examining the number of hidden nodes used in the network. Models were created to predict air temperature at hourly intervals from one to 12 hours ahead. Each ANN model, consisting of a network architecture and set of associated parameters, was evaluated by instantiating and training 30 networks and calculating the mean absolute error (MAE) of the resulting networks for some set of input patterns.

Arvind Sharma briefly explains how the different connectionist paradigms could be formulated using different learning methods and then investigates whether they can provide the required level of performance, which are sufficiently good and robust so as to provide a reliable forecast model for stock market indices. Experiment results reveal that all the connectionist paradigms considered could represent the stock indices behaviour very accurately.

Mike O'Neill focus on two major practical considerations: the relationship between the amounts of training data and error rate (corresponding to the effort to collect training data to build a model with given maximum error rate) and the transferability of models‟ expertise between different datasets (corresponding to the usefulness for general handwritten digit recognition).

Henry A. Rowley eliminates the difficult task of manually selecting non face training examples, which must be chosen to span the entire space of non-face images. Simple heuristics, such as using the fact that faces rarely overlap in images, can further improve the accuracy. Comparisons with several other state-of-the-art face detection systems are presented; showing that our system has comparable performance in terms of detection and false-positive rates.

Dr. S.Santhosh Baboo and I.Kadar Shereef proposed a model using BPN neural network that has potential to capture the complex relationships between many factors that contribute to certain temperature. The results were compared with actual working of mutual meteorological department and these results confirm that the model had the potential for successful application to temperature forecasting.

III. ARTIFICIAL NEURAL NETWORK APPROACH

Adaptive Resonance Theory

The best suitable method to accomplish this aim is neural network approach. One of the best neural network approaches is ART (Adaptive Resonance Theory).ART structure is a neural network for cluster formation in an unsupervised learning domain. In this architecture, the number of output nodes cannot be accurately determined in advance.

The adaptive resonance theory (ART) has been developed to avoid the stability-plasticity dilemma in competitive networks learning. The stability-plasticity dilemma addresses how a learning system can preserve its previously learned knowledge while keeping its ability to learn new patterns. ART architecture models can self-organize in real time producing stable recognition while getting input patterns beyond those originally stored. An ART system consists of two subsystems, an attentional subsystem and an orienting subsystem. The stabilization of learning and activation occurs in the attentional subsystem by matching bottom-up input activation and top-down expectation. The orienting subsystem controls the attentional subsystem when a mismatch occurs in the attentional subsystem. In other words, the orienting subsystem works like a novelty detector.

The basic ART system is an unsupervised learning model. It typically consists of a comparison field and a recognition field composed of neurons, a vigilance parameter, and a reset module. The vigilance parameter has considerable influence on the system: higher vigilance produces highly detailed memories (many, fine-grained categories), while lower vigilance results in more general memories (fewer, more-general categories). The comparison field takes an input vector (a one-dimensional array of values) and transfers it to its best match in the recognition field. Its best match is the single neuron whose set of weights (weight vector) most closely matches the input vector. Each recognition field neuron outputs a negative signal (proportional to that neuron's quality of match to the input vector) to each of the other recognition field neurons and inhibits their output accordingly. In this way the recognition field exhibits lateral inhibition, allowing each neuron in it to represent a category to which input vectors are classified. After the input vector is classified, the reset module compares the strength of the recognition match to the vigilance parameter. If the vigilance threshold is met, training commences. Otherwise, if the match level does not meet the vigilance parameter, the firing recognition neuron is inhibited until a new input vector is applied; training commences only upon completion of a search procedure. In the search procedure, recognition neurons are disabled one by one by the reset function until the vigilance parameter is satisfied by a recognition match. If no committed recognition neuron's match meets the vigilance threshold, then an uncommitted neuron is committed and adjusted towards matching the input vector.

ART1-Algorithm

The architecture of ART1 NN based clustering is given in Fig. 2. Each input vector activates a winner node in the layer F2 that has highest value among the product of input vector and the bottom-up weight vector. The F2 layer then reads out the top-down expectation of the winning node to F1, where the expectation is normalized over the input pattern vector and compared with the vigilance parameter ρ. If the winner and input vector match within the tolerance allowed by the ρ, the ART1 algorithm sets the control gain G2 to 0 and updates the top-down weights corresponding to the winner. If a mismatch occurs, the gain controls G1 & G2 are set to 1 to disable the current node and process the input on another uncommitted node. Once the network is stabilized, the top-down weights corresponding to each node in F2 layer represent the prototype vector for that node.

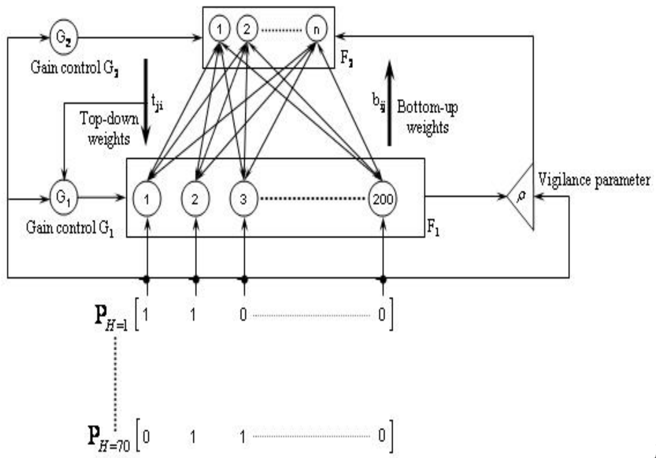

Fig 2- Architecture of our ART1 neural network based clustered. The pattern Vector PH, which represents the access patterns of the host H is the input to the Comparison layer F1. The vigilance parameter determines the degree of mismatch that is to be tolerated. The nodes at the Recognition layer F2 represent the clusters formed. Once the network stabilizes, the top-down weights corresponding to each node in F2 represent the prototype vector for that node.

Phases of ART1: Processing in ART1 can be divided into four phases, (1) recognition, (2) comparison, (3) search, and (4) learning.

(1) Recognition : Initially, in the recognition or bottom-up activation, no input vector I is applied disabling all recognition in F2 and making the two control gains, G1 and G2, equal to zero. This causes all F2 elements to be set to zero, giving them an equal chance to win the subsequent recognition competition. When an input vector is applied one or more of its components must be set to one thereby making both G1 and G2 equal to one. Thus, the control gain G1 depends on both the input vector I and the output X2 from F2,

$G_1 = \{\ 1 \quad \text{if } \ I \neq 0 \ \& \ X_2 \neq 0$

$G_1 = \{\ 0 \quad \text{otherwise}$

In other words, if there is an input vector I and F2 is not actively producing output, then G1 = 1. Any other combination of activity on I and F2 would inhibit the gain control from exciting units on F1. On the other hand, the output G2 of the gain control module depends only on the input vector I,

$G_2 = \{\ 1 \quad \text{if } \ I \neq 0$

$G_2 = \{\ 0 \quad \text{otherwise}$

In other words, if there exists an input vector then G2 = 1 and recognition in F2 is allowed. Each node in F1 receiving a nonzero input value generates an STM pattern activity greater than zero and the node's output is an exact duplicate of input vector. Since both X1i and Ii are binary, their values would be either 1 or 0,

X1 = I, if G1 = 1

Each node in F1 whose activity is beyond the threshold sends excitatory outputs to the F2 nodes. The F1 output pattern X1 is multiplied by the LTM traces W12 connecting from F1 to F2. Each node in F2 sums up all its LTM gated signals

$V_{2j} = \sum_i X_{1i} \ W_{12ji}$

These connections represent the input pattern classification categories, where each weight stores one category. The output X2j is defined so that the element that receives the largest input should be clearly enhanced. As such, the competitive network F2 works as a winner-take-all network described by.

$V_{2j} = \{\ 1 \quad \text{if } \ G_2 = 1 \cap V_{2j} = \max_k \{V_{2k}\} \ \forall \ k$

$V_{2j} = \{\ 0 \quad \text{otherwise}$

The F2 unit receiving the largest F1 output is the one that best matches the input vector category, thus winning the competition. The F2 winner node fires, having its value set to one, inhibiting all other nodes in the layer resulting in all other nodes being set to zero.

(2)Comparison: In the comparison or top-down template matching, the STM activation pattern X2 on F2generates a top-down template on F1. This pattern is multiplied by the LTM traces W12 connecting from F2 to F1. Each node in F1 sums up all its LTM gated signals

$$V_{1i} = \sum_j X_{2j} W_{21ij}$$

The most active recognition unit from F2 passes a one back to the comparison layer F1. Since the recognition layer is now active, G1 is inhibited and its output is set to zero. In accordance with the "2/3" rule, stating that from three different input sources at least two are required to be active in order to generate an excitatory output, the only comparison units that will fire are those that receive simultaneous ones from the input vector and the recognition layer. Units not receiving a top down signal from F2 must be inactive even if they receive input from below. This is summarized as follows

$$X_{1i} = \{ 1 \qquad I_i \cap V_{1i} = 1$$

$$X_{1i} = \{ 0 \qquad \text{otherwise}$$

If there is a good match between the top-down template and the input vector, the system becomes stable and learning may occur. If there is a mismatch between the input vector and the activity coming from the recognition layer, this indicates that the pattern being returned is not the one desired and the recognition layer should be inhibited.

*(3)Search :*The reset layer in the orienting subsystem measures the similarity between the input vector and the recognition layer output pattern. If a mismatch between them, the reset layer inhibits the F2 layer activity. The orienting systems compares the input vector to the F1 layer output and causes a reset signal if their degree of similarity is less than the vigilance level, where ρ is the vigilance parameter set as $0 < \rho \leq 1$.
The input pattern mismatch occurs if the following inequality is true, $\rho < X_1 / I$
If the two patterns differ by more than the vigilance parameter, a reset signal is sent to disable the firing unit in the recognition layer F2. The effect of the reset is to force the output of the recognition layer back to zero, disabling it for the duration of the current classification in order to search for a better match.
The parameter ρ determines how large a mismatch is tolerated. A large vigilance parameter makes the system to search for new categories in response to small difference between I and X2 learning to classify input patterns into a large number of finer categories. Having a small vigilance parameter allows for larger differences and more input patterns are classified into the same category.
When a mismatch occurs, the total inhibitory signal from F1 to the orienting subsystem is increased. If the inhibition is sufficient, the orienting subsystem fires and sends a reset signal. The activated signal affects the F2 nodes in a state-dependent fashion. If an F2 node is active, the signal through a mechanism known as gated dipole field causes a long-lasting inhibition. When the active F2 node is suppressed, the top-down output pattern X2 and the topdown template V1 are removed and the former F1 activation pattern X1 is generated again.
The newly generated pattern X1 causes the orienting subsystem to cancel the reset signal and bottom-up activation starts again. Since F2 nodes having fired receive the longlasting inhibition, a different F2 unit will win in the recognition layer and a different stored pattern is fed back to the comparison layer. If the pattern once again does not match the input, the whole process gets repeated. .
If no reset signal is generated this time, the match is adequate and the classification is finished. The above three stages, that is, recognition, comparison, and search, are repeated until the input pattern matches a top-down template X1. Otherwise a F2 node that has not learned any patterns yet is activated. In the latter case, the chosen F2 node becomes a learned new input pattern recognition category.

(4)Learning: The above three stages take place very quickly relative to the time constants of the learning equations of the LTM traces between F1 and F2. Thus, we can assume that the learning occurs only when the STM reset and search process end and all STM patterns on F1 and F2 are stable. The LTM traces from F1 to F2 follow the equation

$$T_1\, dW_{12ij} / dt = \{ (1-W_{12ij})L-W_{12ij}(X-1) \quad \text{if } V_{1i} \,\&V_{1j} \text{ are active}$$

$$T_1\, dW_{12ij} / dt = \{ 0 \qquad \text{if only } V_{ij} \text{ is inactive}$$

$$T_1\, dW_{12ij}/ dt =\{ -X_1 W_{12ij} \qquad \text{if only } V_{ij} \text{ is active}$$

where τ1 is the time constant and L is a parameter with a value greater than one. Because time constant τ is sufficiently larger than the STM activation and smaller than the input pattern presentation, the above is a slow learning equation that converges in the fast learning equation

$$W_{12ij} =\{ L /(L-1+X_1) \qquad \text{if } V_{1i} \,\&\, V_{1j} \text{ are active}$$

$$W_{12ij} = \{ 0 \qquad \text{if only } V_{ij} \text{ is active}$$

$$W_{12ij}= \{ \text{no change} \qquad \text{if only } V_{ij} \text{ is inactive}$$

The initial values for W12ij must be randomly chosen while satisfying the inequality $0 < W12ij < L / (L-1 + M)$, where M is the input pattern dimension equal to the number of nodes in F1.

The LTM traces from F2 to F1 follows the equation, $T_2 \; dW_{21ji} / dt = X_2 (-W_{21ji} + X_{1i})$ where τ2 is the time constant and the equation is defined to converge during a presentation of an input pattern. Thus, the fast learning equation of the for W21ji is

$$W_{21ji} = \{ 0 \qquad \text{if only } V_i \text{ is inactive}$$

$W_{21ji} = \{ 1 \qquad \text{if } V_{1i} \text{ and } V_{1j} \text{ are active}$

The initial value for W21ji must be randomly chosen to satisfy the inequality 1 ≥ W21ji(0) > C (where C is decided by the slow learning equation parameters. However, all W21ji(0) may be set 1 in the fast learning case.

IV. EXPERIMENTATION AND RESULTS

Collection of data represents the month-wise atmospheric conditions under two parameters namely, temperature and pressure for ten different cities of India namely - Delhi, Kolkata, Bhopal, Mumbai, Jaipur, Amritsar, Cochin, Lucknow, Bhubaneswar, Guwahati, which are geographically well separated from each other.

NEW DELHI	TEMP	PRESSURE
JAN	21	1014
FEB	25	1013
MAR	29	1011
APR	36	1007
MAY	34	1002
JUN	35	1000
JUL	36	999
AUG	35	1002
SEP	33	1005
OCT	32	1011
NOV	27	1014
DEC	23	1014

Fig 1. Data collected for city New Delhi –Temperature and Pressure.

After data collection of various cities, data have been converted into normalised form i.e. binary values of temperature and pressure of various cities have been recognized.

Now on this observed binary form of data, ART1 clustering algorithm is applied so as to generate class codes for various inputs, so as to form clusters to show correlation in between the months of one city to another.

	Delhi	Kolkata	Bhopal	Mumbai	Jaipur	Amritsar	Cochin	Lucknow	Bhuvane	Gwahati
Jan	2	2	1	2	2	2	1	2	1	1
Feb	2	2	1	2	2	2	1	2	1	2
March	2	2	1	2	2	2	1	1	2	1
April	2	2	2	2	1	2	1	1	2	2
May	2	1	2	2	1	2	1	1	2	2
June	2	1	2	2	2	2	1	1	2	1
July	2	1	2	1	2	2	2	1	1	1
Aug	1	1	1	1	1	2	2	1	1	1
Sept	1	2	1	1	1	2	2	1	1	1
Oct	1	1	2	1	1	1	1	1	2	2
Nov	1	2	2	2	1	1	1	2	2	2
Dec	2	1	1	2	1	1	1	2	2	2

Fig 3 Cities categorized into clusters

NEW DELHI	TEMP	PRESSURE
JAN	0	1111
FEB	100	1110
MAR	1000	1100
APR	1111	1000
MAY	1101	11
JUN	1110	1
JUL	1111	0
AUG	1110	11
SEP	1100	110
OCT	1011	1100
NOV	110	1111
DEC	10	1111

Fig 2 Normalized form of input data

As a final result the two major clusters formed are encircled and are represented by cross and check in the figure 4. These encircled figures show the correlation between two cities. These clusters show that there is a close association in between atmospheric conditions of the cities. This correlation helps to predict atmospheric conditions of one city with the help of other cities' atmospheric conditions of other cities in the same cluster.

It also shows that the months of any particular cities which in the same category are either having same atmospheric conditions or the conditions are going to become similar. Such associations are meaningful in prediction of atmospheric conditions which are helpful in protection of loss of human, cattle life and crops.

V. CONCLUSION

Temperature warnings are important forecasts because they are used to protect life and property. Temperature forecasting is the application of science and technology to predict the state of the temperature for a future time and a given location. These are made by collecting quantitative data about the current state of the atmosphere. The Neural Networks package supports different types of training or learning algorithms. One such algorithm is Adaptive Resonance Theory (ART) based on artificial neural network (ANN) technique. The proposed ART1 based clustering method is shown very efficient in correlating the atmospheric conditions in between two or more cities and hence helps in prediction of atmospheric conditions in one particular city based on atmospheric conditions of another city of the same cluster.

The main advantage of ART algorithm is that it can fairly form clusters of similar properties so that prediction of atmospheric conditions of various cities can be differentiated. The simple meaning of this term is that our model has potential to capture the complex relationships between many factors that contribute to certain atmospheric conditions

V. REFERENCES

[1] Dr. S. Santhosh Baboo and I.Kadar Shereef: "An Efficient Temperature Prediction System using BPN Neural Network", IACSIT,2011.

[2] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.

[3] Andreas Nürnberger and Marcin Detyniecki, "Content Based Analysis of Email Databases Using Self-Organizing Maps", Proceedings of the European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems EUNITE'2001, Tenerife, Spain, pp. 134-142, December, 2001.

[4] Arvind Sharma, Prof. Manish Manoria," A Weather Forecasting System using concept of Soft Computing,".

[5] Bhupesh Gour, et al., "ART Neural Network Based Clustering Method Produces Best Quality Clusters of Fingerprints in Comparison to Self Organizing Map and K-Means Clustering Algorithms", IEEE Communications Society Explore, pp. 282-286, Dec. 2008.

[6] Bhupesh Gour, et al., "Fingerprint Clustering and Its Application to Generate Class Code Using ART Neural Network", IEEE Computer Society Explore, pp 686-690, July 2008.

[7] Smith Brian.A , McClendon Ronald.W , and Hoogenboom Gerrit," Improving Air Temperature Prediction with Artificial Neural Networks" International Journal of Computational Intelligence 3;3 2007.

[8] D. Jianga. Y. Zhanga. X. Hua. Y. Zenga. J. Tanb. D. Shao. Progress in developing an ANN model for air pollution index forecast. Atmospheric Environment. 2004, 38: pp.7055-7064.

[9]S. Palani. P. Tkalich. R. Balasubramanian. J. Palanichamy. ANN application for prediction of atmospheric nitrogen deposition to aquatic ecosystems. Marine Pollution Bulletin. 2011, in press.

[10] Gour Bhupesh, Khan Asif Ullah (2012). Atmospheric Condition based Clustering using ART Neural in International Journal of Information and Communication Technology Research.Bhopal, 2012.

[11] Abbas Osama Abu, "Comparision between Data Clustering Algorithm" In International Arab Journal of Information Technology,2008.