ISSN: 2229 - 6646

Temperature Forecasting for Dehradun City Using Back Propagation Method of Data Mining Technique

shikha garg (shikha.garg36@gmail.com)
Uttarakhand technical university, dehradun

< >

Abstract-- Meteorological data mining is a form of data mining concerned with finding meaningful data patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge of weather data or climate data in a region is essential for business, society, agriculture and energy application. The main focus of this paper is to create models for decision making that predict future behavior based on analysis of past activity for meteorology and weather forecasting based on several parameters as minimum and maximum temperature, and to make tool due to which any prediction or graphic sample could be generated for future analysis.

Keywords-Artificial Neural Networks (ANN), Back Propagation (BPN),Meteorological data, Data Mining.

1-INTRODUCTION

Data mining is about processing data and identifying patterns and trends in that information so that we could decide or judge. Data mining principles have been around for many years, but, with the advent of *big data*, it is even more prevalent. Big data caused an explosion in the use of more extensive data mining techniques, partially because the size of the information is much larger and because the information tends to be more varied and extensive in its very nature and content. With large data sets, it is no longer enough to get relatively simple and straightforward statistics out of the system [1][2].

The process of data analysis, discovery, and model-building is often iterative as we target and identify the different information that you can extract. We must also understand how to relate, map, associate, and cluster it with other data to produce the result[4]. Identifying the source data and formats, and then mapping that information to our given result can change after we discover different elements and aspects of the data[7].

1-RELATED WORK

1. Dr.Sunil R. Gupta and Gaurav J.Sawale (2013) combine back propagation and Hopfield network model effective[5] to determine the non linear relationship that exist between the historical data (temperature,wind speed, humidity) and on this basis, make a prediction of what the weather would be in future.
2. Ch.Jyosthna Devi , B.Syam Prasad Reddy , K.Vagdhan Kumar ,B.Musala Reddy ,N.Raja Nayak [6]introduced the concept of neural network including back propagation. A 3 layer neural network is designed and trained with the existing dataset and obtained a relationship between the existing non linear parameters of weather.

2-BACK PROPAGATION METHOD

Artificial Neural Network [6][7] is a powerful data modelling tool that is able to represent complex input /output relationships. The motivation for the development of neural network technology stemmed from the desire to implement an artificial system that could perform intelligent tasks similar to those performed by the human brain. A **Back Propagation Network[6]** consists of at least three layers (multi layer perception): an input layer, at least one intermediate hidden layer, and an output layer. Typically, input units are connected in a feed-forward fashion with input units fully connected to units in the hidden layer and hidden units fully connected to units in the output layer. An input pattern is propagated forward to the output units all the way through the intervening input-to-hidden and hidden to output weights when a Back Propagation network is cycled. As the algorithm's name gives a meaning, the errors (and therefore the learning) propagate backwards from the output nodes to the inner nodes.

ALGORITHM:

1. First apply the inputs to the network and work out the output – remember this initial output could be anything, as the initial weights were random numbers.
2. While terminating condition is not satisfied, for each training tuple in data set.Propagate the input forward.
3. For each input unit j $O_j=I_j$, output of any input unit is its actual input value.
4. For each hidden and output layer j,calculate net input of unit x with respect to the previous layer.$I_x = \sum_{i=1}^{3} w_{ix} O_i$.
5. $O_x=\frac{1}{1+e^{-Ix}}$, calculate the outputof each unit until output node O_n .
6. Now back propagate errors for each unit in the output layer $Err_n= O_n (1 - O_n) (T_n - O_n)$,
7. Error will be calculated from the last to first hidden layer.$Err_x=O_x(1-O_x)(Err_{n*w}xn)$
8. Now weight updation will proceed in network $w_{xn} = w_{xn} + (Err_n * O_x)$.
9. We update all the weights in the output layer to input layer in this way.

1-METHODOLOGY

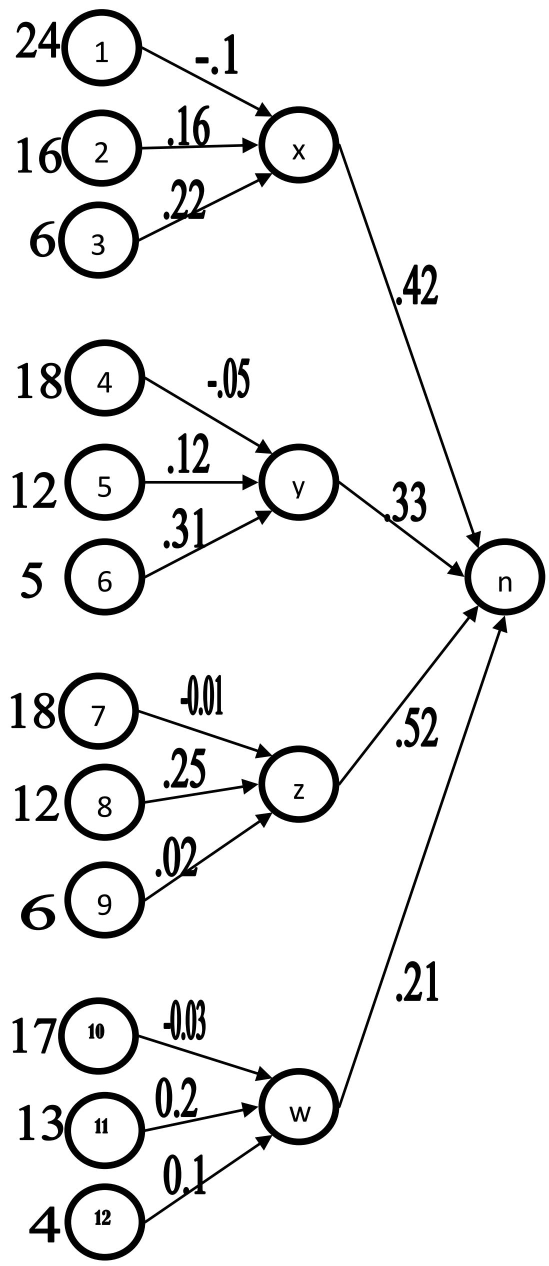

Fig.January temperature readings.

According to the work using back propagation method and by considering the following diagram,shows the temperature reading[3] for four years for the month January. It shows 12 input nodes ,4 hidden nodes ,1 output node . Node number 1,2,3 connected to the node x, node number 4,5,6 connected to the node y, node 7,8,9 connected to the node z , node 10,11,12 connected to the node w and node x, y, z, w connected to the node n. Each and every connection have a specified weight with a numberic value associated with it. Node 1 to 12 signifies the temperatures (max, avg, min) and node x, y, z, w signifies the year 2010,2011,2012,2013 and the final node n signifies the output layer through which we will back propagate the network and point out the error.

Working of Network -

Node 1,4,7,10 include the maximum temperature , node 2,5,8,11 includes the average temperature and node 3,6,9,12 includes the maximum temperature.

Now calculate I_x,I_y,I_z and I_w

Where,

$$I_x = \sum_{i=1}^{3} w_{ix} O_i$$

Where,

i = (1, 2, 3)

For an input unit I, its output ,O_i is equal to its input value I_i and W_{ix} is the input weight to the connected edges, Now

$$I_x=(w_{1x}*O_1) + (w_{2x}*O_2) + (w_{3x}*O_3)$$

Similarly I_y,I_z and I_w are calculated.Now according to the algorithm O_x is calculated. Formula for O_x is as follows:

$$O_x=\frac{1}{1+e^{-Ix}}$$

Similarly calculate O_y,O_z and O_w .

For the output node O_n is calculated as

$$O_n= (w_{xn}*O_x) + (w_{yn}*O_y) + (w_{zn}*O_z)+ (w_{wn}*O_w)$$

After calculating the following values now starts back propagate the network by calculating error at each node.Starting with the output node error at node n will be:

$$Err_n= O_n (1 - O_n) (T_n - O_n)$$

Where T_n is the target value, T_n is taken as approximately equal to the O_n here we take T_n as 1 during the whole calculations.After calculations error at O_n now calculate error at O_x,O_y,O_z and O_w .

Now err_x,err_y,err_z and err_w are calculated:

$$Err_x=O_x(1-O_x) (Err_{n*w}xn)$$

Similarly Err_y,Err_z and Err_w is calculated.

According to the algorithm now it is the time to update the weights of the edges and find out the differences between new weight and old weights.

Formula to update the weights.

$$w_{xn} = w_{xn} + (Err_n * O_x)$$

Similarly,

$$w_{yn} = w_{yn} + (Err_n * O_y)$$

$$w_{zn} = w_{zn} + (Err_n * O_z)$$

$$w_{wn} = w_{wn} + (Err_n * O_w)$$

Now we will back propagate it till private nodes until each and every weight is not updated.

$$w_{1x} = w_{1x} + (Err_n * O_1)$$

After calculating w_{2x} ,w_{3x},w_{4y},w_{5y},w_{xy},w_{7z},w_{8z},w_{9z},w_{10w},w_{11w} and w_{12w} compare this updated weight with the old weights , we notice a slight change between weights.

According to my calculation I take result of year 2010,2011,2012 and assume the temperature prediction for 2013. Now apply the back propagationon the network.

According to table I there are two attempts now we compare the error and weight changes into both the attemps.

After that again change the prediction of 2013 again but all weights & old readings were same. Then apply back propagation . So I notice a change in error.

Now finalize the reading with which the error would be less.

I. RESULT AND ANALYSIS

With the help of following tables, I can better define the whole process.

TABLE I

Shows January temperature for 4 years .(max,avg,min)

Year	Node	Node name	Temperature	
			I attempt	II attempt
year 2010	1	I1	24	24
	2	I2	16	16
	3	I3	6	6
year 2011	4	I4	18	18
	5	I5	12	12
	6	I6	5	5
Year 2012	7	I7	18	18
	8	I8	12	12
	9	I9	6	6
Year 2013	10	I10	17	19
	11	I11	13	11
	12	I12	4	5

 ISSN: 2229 - 6646

TABLE II

Weight taken

Weight Name	w_{1x}	w_{2x}	w_{3x}	w_{4y}	w_{5y}	w_{6y}	w_{7z}	w_{8z}	w_{9z}	w_{10w}	w_{11w}	w_{12w}	w_{xn}	W_{yn}	W_{zn}	W_{wn}
Weight	-.1	.16	.22	-.05	.12	.31	-.01	.25	.02	-.03	.2	.1	.42	.33	.52	.21

Weight will be same in both attempts

Now consider the table below.This table will show the comparative readings of both the attempts.

TABLE III

Error on the specified nodes in both attempt.

Node	Node Error	Error in 1st attempt	Error in 2nd attempt
X	Err_x	.0076	.0064
Y	Err_y	.0038	.0032
Z	Err_z	.0028	.0024
W	Err_z	.0018	.0013
N	Err_n	0.125	0.108

If we check this table then we notice that the reading in second attempt are less than 1st attempt.So after changing the temperature it will show a change.so second attempt readings of temperature will better.

TABLE IV

the change in weights in both the cases.

Weight Name	**Initial Weight**	**Weight in 1st attempt**	**Weight in 2nd attempt**
w_{1x}	**-0.1**	**.08**	**.05**
w_{2x}	**.16**	**.28**	**.26**
w_{3x}	**.22**	**.36**	**.22**
w_{4y}	**-.05**	**.018**	**.007**
w_{5y}	**.12**	**.165**	**.15**
w_{6y}	**.31**	**.329**	**.32**
w_{7z}	**-.01**	**.04**	**.03**
w_{8z}	**.25**	**.28**	**.27**
w_{9z}	**.02**	**.03**	**.03**
w_{10w}	**-.03**	**.0006**	**-.005**
w_{11w}	**.2**	**.22**	**.21**
w_{12w}	**.1**	**.1**	**.1**
w_{xa}	**.42**	**.52**	**.50**
Way	**.33**	**.44**	**.42**
w_{za}	**.52**	**.63**	**.62**
w_{wa}	**.21**	**.32**	**.31**

If we compare the weights in both attempts we conclude that 2nd attempt weights were less in comparison to 1st attempt reading and approximately equal to the initial weights.

So we find that 2nd attempt readings of temperature would be better.

CONCLUSION

In this paper,how back propagation is useful to check the future behaviour of temperature. The research shows that the following technique could give better predictions of temperature in future by back propagating the errors with non linear parameters.

REFRENCES

[1] Badhiye S. S., Wakode B. V., Chatur P. N. "Analysis of Temperature and Humidity Data for Future value prediction", IJCSIT Vol. 3 (1), 2012, 3012 – 3014

[2] Sarah N. Kohail, Alaa M. El-Halees, "Implementation of Data Mining Techniques for Meteorological Data Analysis", IJICT Journal Volume 1 No. 3, July 2011

[3]http://www.wunderground.com/history/wmo/42111/2011/7/28/MonthlyHistory.html.

[4] Meghali A.kalyankar, Prof.S.J.Alaspurkar , "Data Mining Techniques to Analyse the Metrological Data",IJARCSSE Vol.3 (2),February 2013

[5] Dr.Sunil R. Gupta and Gaurav J. Sawale, "Use of Artificial Neural Network in Data Mining For Weather Forcasting", IJCSA Volume 6 No. 2, April 2013

[6] Ch.Jyosthna Devi , B.Syam Prasad Reddy , K.Vagdhan Kumar ,B.Musala Reddy ,N.Raja Nayak, " ANN Approach for Weather Prediction using Back Propagation", IJETT Volume 3 Issue 1-2012

[7] J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000.