ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Two Phase Neural Network Model for Weather Forecast Along-with Logistic and Linear Regression

Anil Kumar [1], Praveen Sharma[2]

M.Tech Student, Department of Computer Science & Engineering, N.G.F College of Engineering and Technology at Palwal, Haryana, India[1]

Assistant Professor, Department of Computer Science & Engineering, N.G.F College of Engineering and Technology at Palwal, Haryana, India[2]

ABSTRACT: Day by day Weather determining is utilized for different reasons in various zones like horticulture, vitality supply, transportations, and so forth. Exactness of climate conditions appeared in conjecture reports is exceptionally important. In this task, the scheme is directed to research a superior approach for estimating which looks at numerous systems, for example, Two Phase Neural Network, Logistic and Linear Regression for the mean of fuzzy bunching, and so on which are utilized for various kinds of measuring technique. Among which neural system with the back track calculation performs expectation with negligible mistake. Neural system is a mind boggling system which is self-versatile in nature. It learns independent from anyone else utilizing the preparation information and produces some insightful examples which are helpful for anticipating the climate. The system utilizes 7 input parameters to conjecture the day by day climate as far as temperature, precipitation, mugginess, cloud condition, and climate of the day thus predicting the temperature for next 4 to 5 days will be forecasted.

KEYWORDS: Two Phase Neural Network, Logistic Regression, Linear Regression, Artificial Neural Network.

I. INTRODUCTION

Climate expectation is measure of future climate condition. Weather state of air, humidity, pressure and wind direction at given time as far as climate factors like temperature, weight, wind bearing and so on. The precision of the expectation generally relies upon information of winning climate condition over a wide region. Climate is non-straight and dynamic process it shifts everyday even moment to-minute. As the climatic dataset is profoundly nonlinear so Artificial Neural Network (ANN) can be utilized for climate forecast and arrangement. Data mining using machine learning is one of them which we have utilized as a part of this paper to oversee climate related information and predict the forecast and certain condition of future weather. Under this scheme we suggest that how to utilized the data mining and in order retrieval of data using machine learning in the expectation of climate and forecasting of the weather. Presently and now a days, we the people of India experiencing changing bad weather, pollution and their reactions. Typically in horticulture field, ranchers are confronting numerous issues because of surprising climate conditions. Climate anticipating is straightforwardly rely on the regular particles display noticeable all around like dew, moisture, cloud density, heat index, air pressure, wind direction and so on. In this paper we have concentrated on particular area i.e. Delhi. To decrease these reactions up to some degree there are numerous strategies and calculations through which we can foresee the climate on the premise of given information using 2 Phase NN whereas twice the regression i.e. Logistic and Linear is modelled with 2Phase NN classification model to derive the results.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Number	Name	Value
1	Classification	Clear
2	Maximum	Temperature (F) 57
3	Minimum	Temperature (F) 33
4	Mean Humidity	Humidity 43
5	Mean Pressure	Atmospheric Pressure in 30.13

Table 1: Sample data from January 1, 2017, with the number, name, and value of each of the five features.

II. RELATED WORK

Related works included a wide range of and fascinating systems to attempt to perform climate figures. While a lot of current determining innovation includes reenactments in light of material science and differential conditions, numerous new methodologies from computerized reasoning utilized essentially machine learning strategies, generally neural systems while some drew on probabilistic models, for example, Bayesian systems. Out of the three papers on machine learning for climate expectation we inspected, two of them utilized neural systems while one utilized help vector machines."Neural systems appear to be the prominent machine learning model decision for climate determining on account of the Neural systems appear to be the prominent machine learning model decision for climate determining on account of the capacity to catch the non-direct conditions of past climate patterns and future climate conditions, dissimilar to the straight relapse and practical relapse models that we utilized. This gives the upside of not accepting basic direct conditions of all highlights over our models. Of the two neural system approaches, one [3] utilized a mixture demonstrate that utilized neural systems to show the material science behind climate estimating while the other [4] connected adapting all the more specifically to anticipating climate conditions. Likewise, the approach utilizing bolster vector machines [6] additionally connected the classifier straightforwardly for climate forecast yet was more restricted in scope than the neural system approaches. Different methodologies for climate guaging included utilizing Bayesian systems. One intriguing model [2] utilized Bayesian systems to model and make climate expectations however utilized a machine learning calculation to locate the most ideal Bayesian systems and parameters which was computationally costly due to the substantial measure of various conditions yet performed extremely well. Another approach [1] concentrated on a more particular instance of anticipating extreme climate for a particular topographical area which restricted the requirement for calibrating Bayesian system conditions however was constrained in scope.

1. Brian Dolan display the outline rationality, methods and experience giving MAD examination to one of the world's biggest promoting systems at Fox Audience Network, utilizing the Green plum parallel database framework. We depict database plan approaches that help the light-footed working style of examiners in these settings.

2. R. P. Singh clarify why a cloud-based arrangement is required, depict our model usage, and investigate some case applications we have executed that show individual information proprietorship, control, and examination. He address these issues by outlining and executing a cloud-based engineering that furnishes buyers with quick access and fine-grained control over their utilization information, and also the capacity To break down this information with calculations of their picking, including outsider applications that investigate that information in a protection saving style.

3. Jeffrey Dean depicts the essential programming model and gives a few cases. Numerous genuine errands are expressible in these models. Usage of Map Reduce keeps running on an extensive bunch of ware machines and is exceptionally adaptable: a regular Map Reduce calculation forms numerous terabytes of information on a huge number of machines. Software engineers and the framework simple to utilize: several Map Reduce programs have been actualized and upwards of one thousand Map Reduce employments are executed on Google's bunches each day.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

III. PROPOSED ALGORITHM

The most outrageous temperature, slightest temperature, mean clamminess, mean barometrical weight, and atmosphere gathering for consistently in the years 2011-2015 for Delhi, India were gained from Weather Underground. [7] Originally, there were nine atmosphere orders: clear, scattered fogs, to some degree shady, generally shady, dimness, overcast, rain, tempest, and snow. Since an extensive parcel of these requests are practically identical and some are meagrely populated, these were diminished to four atmosphere groupings by joining scattered fogs and not entirely shady into sensibly shady; generally shady, foggy, and shady into extraordinarily shady; and rain, tempest, and snow into precipitation. The data from the underlying four years were used to set up the counts, and the data from the latest year was used as a test set and the alluded data for January using the table 1 depicted parameters.

Input data is then pre-processed and cleaned. That means it is checked with any outlier and that is removed, missing values are entered, and data is checked if it is in the given range for the given parameter. Later ANN is designed with number of input and output nodes, hidden layers, activation function, and maximum number of epochs, weights, bias, goal and learning function. Neural network is trained with seventy percentages of the input data. Where the model is trained using this observed data to forecast the weather, followed by testing done using remaining thirty percentages of input data. Then the mean squared error and accuracy is calculated for the model by comparing the output of testing with target output. However, the confusion matrix will be produced from the Linear regression resulting the mean and variance and further the Logistic regression to produce the Slope and Maximum Likely Hood Estimation. The below diagram depicts the proposed framework which is implemented under the scheme to produce the results.

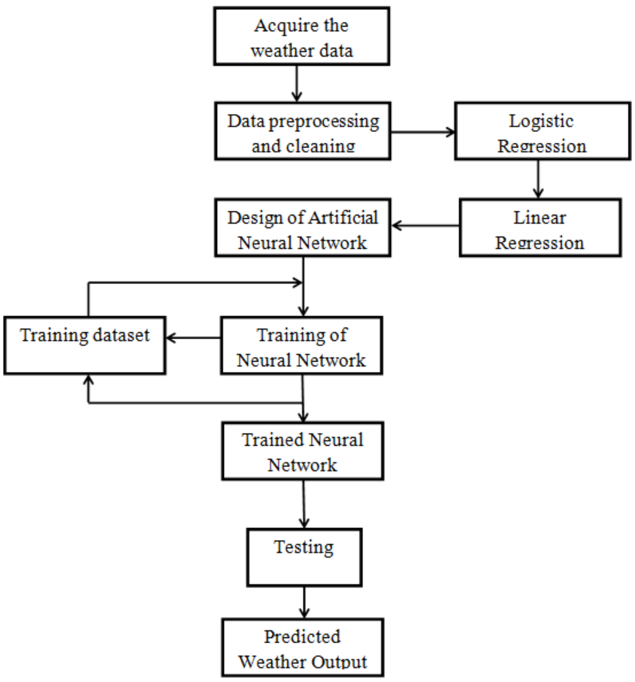

Figure1: Proposed Scheme vide Two Phase NN along-with Logistic and Neural Network

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

IV. PSEUDO CODE

To gain the desired goals and results in proposed scheme the probalastic scenarios i.e. linear regression ,logistic regression and 2Phase NN have be used. The below steps depicts the workflow and implementation of proposed scheme.

Logistic Regression : Because logistic regression predicts probabilities, rather than just classes, we can fit it using likelihood. For each training data-point, we have a vector of features, xi , and an observed class, yi . The probability of that class was either p, if $y_i = 1$, or 1 − p, if $y_i = 0$. The likelihood is then:-

$$L(\beta_0, \beta) = \prod_{i=1}^{n} p(x_i)^{y_i}(1 - p(x_i)^{1-y_i}$$

(I could substitute in the actual equation for p, but things will be clearer in a moment if I don't.) The log-likelihood turns products into sums:

$$\begin{aligned}
\ell(\beta_0, \beta) &= \sum_{i=1}^{n} y_i \log p(x_i) + (1 - y_i) \log 1 - p(x_i) \\
&= \sum_{i=1}^{n} \log 1 - p(x_i) + \sum_{i=1}^{n} y_i \log \frac{p(x_i)}{1 - p(x_i)} \\
&= \sum_{i=1}^{n} \log 1 - p(x_i) + \sum_{i=1}^{n} y_i (\beta_0 + x_i \cdot \beta) \\
&= \sum_{i=1}^{n} - \log 1 + e^{\beta_0 + x_i \cdot \beta} + \sum_{i=1}^{n} y_i (\beta_0 + x_i \cdot \beta)
\end{aligned}$$

where in the next-to-last step we finally use equation

Typically, to find the maximum likelihood estimates we'd differentiate the log likelihood with respect to the parameters, set the derivatives equal to zero, and solve. To start that, take the derivative with respect to one component of β, say β_j.

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta_j} &= -\sum_{i=1}^{n} \frac{1}{1 + e^{\beta_0 + x_i \cdot \beta}} e^{\beta_0 + x_i \cdot \beta} x_{ij} + \sum_{i=1}^{n} y_i x_{ij} \\
&= \sum_{i=1}^{n} (y_i - p(x_i; \beta_0, \beta))\, x_{ij}
\end{aligned}$$

Linear Regression: In totality, dataset was obtained from 4000 to 5000 (records from meteorological department is obtained for regression) for at slightest seven attributes are regressed by which waning combinations were calibrated under this scheme. In erstwhile expression, of the 5000 rows forming the data cluster is formerly selected for use in this study below depicts the linear regression model. The righteousness of fit character for the model calibrations are obtainable in below equation, and the calibrated coefficients are shown in table 1. However presents standard error (Se) calculated as:-

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

$$S_e = \sqrt{\frac{1}{n-m}\sum(y-\hat{y})^2}$$

where n is the number of observations,
m is the number of coefficients or exponents being calibrated,
y is the observed discharge (from the PeakFQ output), and
$\hat{y}$ is the predicted output calibrated by the regression tool.

Standard deviation (S_y) is calculated as

$$, = \sqrt{\frac{1}{n-1}\sum(y-\bar{y})^2},$$

where $\bar{y}$ is the mean of the discharges for the return period (T).

Explained variance (R^2) is calculated as

$$R^2 = \frac{1}{n^2 \cdot S_e^2 \cdot S_x^2}\left[\sum(y-\hat{y})\cdot(y-\bar{y})\right]^2$$

where

$$S_x = \sqrt{\frac{1}{n-1}\sum(\hat{y}-\bar{\hat{y}})^2},$$

in which $\bar{\hat{y}}$ is the mean of the predicted discharges for the return period

2Phase NN: To express the regression required by this scheme. Given four control points (consisting of neural network weight configurations) β = {α0 α1 α2 α3}, the cubic Catmull-Rom spline function Θ which produces network weights for arbitrary phase p can be defined as follows to predict the temperature on the dataset produced by Linear and Logistic Models :

$$\begin{aligned}\Theta(p;\beta) = \quad & \alpha_{k_1} \\ & + w\left(\tfrac{1}{2}\alpha_{k_2} - \tfrac{1}{2}\alpha_{k_0}\right) \\ & + w^2\left(\alpha_{k_0} - \tfrac{5}{2}\alpha_{k_1} + 2\alpha_{k_2} - \tfrac{1}{2}\alpha_{k_3}\right) \\ & + w^3\left(\tfrac{3}{2}\alpha_{k_1} - \tfrac{3}{2}\alpha_{k_2} + \tfrac{1}{2}\alpha_{k_3} - \tfrac{1}{2}\alpha_{k_0}\right) \\ w = & \frac{4p}{2\pi} \pmod 1 \\ k_n = & \left\lfloor \frac{4p}{2\pi} \right\rfloor + n - 1 \pmod 4.\end{aligned}$$

V. SIMULATION RESULTS

The simulation studies involve the trained model whereas the probabilistic approach is formed with the amalgamation of three machine learning models i.e. Neural Network ,Linear and Logistic Regression the following figures are the depicted for the proposed scenario and results:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

http://localhost:8080/w4casting/showResults.jsp

Logistic Regression : Maximum Likely Hood Estimation

SNo	Heat Index | Humidity | Pressure | Temperature based on Year-&-Month	MLE
64	[25309.024049944983, 60619.507799908955, 71547.31384993567, 209889.43779987886, 3852576.6427514288]	12181.0113
129	[50618.048049687655, 121239.01479968101, 143094.62734944423, 419778.8707999441, 7705153.234243807]	12216.6458
193	[75537.70244931147, 180925.9139991873, 213541.21294860778, 626439.2355994895, 1.1498459416668816E7]	12122.4287
257	[100457.3568488443, 240612.8131986414, 283987.79854771093, 833099.6003983855, 1.529176559910111E7]	12172.0106
321	[125377.01124837712, 300299.71239816863, 354434.38414668635, 1039759.9651972816, 1.908507178157276E7]	12222.8824
384	[149907.29604867488, 359054.0037977468, 423780.2418456778, 1243191.261797675, 2.2819107554968268E7]	12169.3268
447	[174437.58084920145, 417808.29519732494, 493126.0995446692, 1446622.558398121, 2.6553143328363776E7]	12102.0879
510	[198967.86564972802, 476562.5865969031, 562471.9572443797, 1650053.8549985671, 3.0287179101759285E7]	12214.4453
574	[223887.52005026295, 536249.4857965859, 632918.542844659, 1856714.2197990203,	12150.1368

Figure 2: Slope and MLE from Logistic Regression

localhost:8080/w4casting/showConfusionMatrix

Confusion Matrix

Year	Month	Humidity	Pressure	Temprature	Cloud Density
2010	1	92.6182	1022.5836	0.9091	0.0
2010	2	36.979	1020.3153	13.6449	0.0
2010	3	-21.8556	1007.5516	38.1329	4.0046
2010	4	-7.0902	978.7819	38.8298	3.121
2010	5	-11.5261	957.6311	41.1142	3.164
2010	6	-25.2393	986.3522	37.8471	0.0
2010	7	2.8592	1124.4366	25.1761	0.0
2010	8	52.5165	1714.9195	16.0671	0.0
2010	9	-34.7116	1057.1947	22.537	0.0
2010	10	19.1388	1018.0955	17.2067	0.0
2010	11	30.4591	1012.4807	13.1174	0.0

Figure 3: Confusion Matrix from Linear Regression

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Forecasting For Next 4 days

Weather Condition (Forecasted)	Temprature (Approx May Vary From -2 to +2 Degrees)
Blowing Sand	31
Haze	30
Sunny and Widespread Dust	36
Widespread Dust	36

Figure 4: Result using 2Phase NN based on MLE from Logistic Regression and Confusion Matrix Based on Linear Regression.

VI. Conclusion and Future Work

In the proposed scheme, we initially make a measure assessment of the attributes defined in 2 phase NN regression which will estimate and discover ground truth field estimations using trust relationship and mitigating the relation. Indeed, even in the inclination remedied case, the residuals attributes will act in the models which will initiate to have huge predisposition aim or means, demonstrating that there are indicators and with incorporation enhance the outcomes i.e. weather forecast. These outcomes inspired the advancement of information driven methodologies, for example, a 2 Phase NN. In building up information driven NN as demonstrated above, we found a general change of execution when utilizing earlier inputs together with the distinction amongst present and following day meteorological parameter estimates. Thus 2 Phase NN approach performed fundamentally superior to the event that we utilized just the future estimate factors. We additionally make a precision and recall, which delineated that the 2 Phase NN approach has prevalent forecasting aptitudes amid the subsequent days or the forthcoming weeks also. The 2 Phase NN aims to give a better and more accurate result as compared to the Neural Network approach since it also trains its model using the outputs generated in the first phase.

The weather Forecasting has a big challenge of predicting the accurate results which are used in many real time systems like electricity departments, airports, tourism centers, etc. The difficulty of this forecasting is the complex nature of parameters. Each parameter has a different set of ranges of values. This issue is addressed by ANN. It accepts all complex parameters as input and generates the intelligent patterns while training and it uses the same patterns to generate the forecasts.

The Artificial Neural Network model proposed in this project indicates all the parameters for input and output, training and testing data set, number of hidden layers and neurons in each hidden layer, weight, bias, learning rate and activation function. The Mean Squared Error between predicted output and the actual output is used to check accuracy.

For future work the same can be incorporated with Big-Data services based in Map-per and Reducer using YARN (Yet Another Resource Negotiator) or streaming model based on Apache Spark to handle the gigantic dataset and faster results.

References

1. Prasanta Rao Jillella S.S., P Bhanu Sai Kiran, P. Nithin Chowdary, B. Rohit Kumar Reddy, Vishnu Murthy, Weather Forecasting Using Artificial Neural Networks and Data Mining Techniques in (IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH, Volume No.3, Issue No.6, October - November 2015, 2534 – 2539.
2. Amartya Raj Gurung, Forecasting Weather System Using Artificial Neural Network (ANN): A Survey Paper in International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137, Volume – 02, Issue – 10, October – 2017, PP – 42-50
3. Abhishek Saxena, Neeta Verma, Dr. K C Tripathi, A Review Study of Weather Forecasting using Artifical Neural Network Approach in International Journal of Engineering and Research & Technology (IJERT), Vol. 2, Issue 11, November 2013
4. Meera Narvekar, Priyanca Fargose, Daily Weather Forecasting using Artificial Neural Network in International Journal of Computer Applications (0975 – 8887), Volume 121 – No.22, July 2015

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

5. Rishabh Gupta and Niharika Singh, Neural Network Based Application for Air Quality Visualization, 2nd International Conference on Next Generation Computing Technologies,2016, pp 169-172
6. Michaela, Sanda Florentina, and Marian, A Comparative Study of Computational Intelligence Techniques Applied to PM2.5 Air Pollution Forecasting, 6th International Conference on Computers Communications and Control (ICCCC), 2016, pp 103-108
7. Niharika, Venkatadri M, and Padma S.Rao, A survey on Air Quality forecasting Techniques, vol. 5(1). International Journal of Computer Science and Information Technologies, 2014, pp.103-107.
8. Santosh Kumar Nanda and S.S. Mahapatra, Application of Legendre Neural Network for Air-Quality Prediction, in International Conference on Engineering and Technology, 2011.
9. S.V. Barai, A.K. Dikshit and S. Sharma, Neural Network Models for Air Quality Prediction: A Comparative Study, in Springer-Verlag, Berlin, Vol. 20, 2007, pp. 290-3015.