Applied Mathematical Sciences, Vol. 9, 2015, no. 13, 613 - 616
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2015.411953

Annual Mean Temperature Prediction of India Using K-Nearest Neighbour Technique

M. Mallika, S. Meenakshi Sundaram and M. Nirmala

Department of Mathematics
Sathyabama University, Chennai – 119, India

Abstract

Weather can be defined as the condition of air on earth for a given time and at a given place. Weather prediction has been of great interest and a challenging task for researchers for so many years. To predict weather so many factors have to be considered like temperature, atmospheric pressure, humidity, wind pressure etc. In particular temperature prediction plays a vital role for planning in so many fields like agriculture, industry, climatic conditions and so on. Various statistical and computational methods are applied for temperature prediction so far and literature survey shows that machine learning methods outperformed the standard traditional methods. In this paper, one of the machine learning algorithm K-Nearest Neighbour algorithm (KNN) is applied for predicting annual mean temperature of India based on the annual maximum and annual minimum temperatures of India. The proposed model predicts annual mean temperature and the findings are compared using the error measure MAPE.

Keywords: Temperature prediction, K-Nearest Neighbour, MAPE

1 Introduction

Weather prediction which is highly complex in nature and also nonlinear has created a great interest among researchers in recent years. Many scientists are predicting the weather parameters throughout the world using different techniques. Y. Radhika and M. Shashi applied Support Vector Machines for temperature prediction. The results are compared with MLP trained with back-propagation al-

gorithm and have shown that the performance of SVM is found to be consistently better [2]. In another paper Dr. Santhosh Baboo and I. Kadar Shereef implemented Back Propagation Neural Network (BPN) technique for temperature prediction .The results are compared with actual working of meteorological department and shown that the model can be successfully applied to temperature forecasting [4]. Shaminder Singh, Pankaj Bhambri and Jasmeen Gill have proposed a time series based temperature prediction model using integrated Back Propagation with Genetic Algorithm technique. The effect of under training and over training the system is also discussed [5]. Khushbu Khamar compared various algorithms such as Support Vector Machines, Naïve Bayes,K-Nearest Neighbor,etc for short text classification and concluded that kNN gives better accuracy [1]. Sagar S. Badhiye, Nilesh U.Sambhe and P. N. Chatur used Knn to predict temperature and humidity for a specific region and the results were established [3]. The objective of this study is to develop KNN models for temperature prediction.

2 Methodology

KNN is a method in which classification is done based on the closest training points in the data set. This technique assumes that all instances correspond to points in n-dimensional space. Thus an arbitrary instance say x is given by the vector ($a_1(x)$, $a_2(x)$,…$a_n(x)$) where $a_r(x)$ denotes the value of the r^{th} attribute of instance x. The closeness of these vectors can be measured using a distance metric. Most commonly used metric is Euclidean distance which is given by

$D(x_i, x_j) = \sqrt{\sum_{r=1}^{n}\left(a_r(x_i) - a_r(x_j)\right)2}$ where x_i and x_j are two instances.

KNN is basically a classifier which can also be used for prediction. For making predictions select the value of 'k' and the 'k' nearest neighbours of the point in question are found from the entire data set using distance metric. In this research work, the distance metric is taken as Euclidean squared and the unknown value is calculated from the 'k' nearest neighbours using three models one using equal weights(average) and the remaining two using unequal weights. One of these two approaches (unequal weights) is by giving weights exponentially in the form $W[x, x_i] = \frac{\exp[-D(x, x_i)]}{\sum_{i=1}^{k}\exp[-D(x, x_i)]}$ where $D(x, x_i)$ denotes the distance between the query point and one of the 'k' nearest neighbor [6]. The second of this kind (unequal weights) is by Inverse Distance Weighting (IDW) in which each of the 'k' nearest neighbor is given a weight equal to the inverse of its distance from the query point. i.e, $W[x, x_i] = \frac{1}{D(x, x_i)}$.

3 Study Materials and Discussion

For the prediction of Annual mean temperature of India a total of 62 years from 1951 to 2012 temperature data is taken as the database where the first 50 years are considered as initialization set and the remaining 12 years as the test set. The data is based on the surface air temperature data from more than 350 stations spread over. The time series data is taken from Open Government Data (OGD) Platform India. In this research work, Annual maximum temperature and Annual Minimum temperature of India is taken as the input (2-dimensional vector) and the Annual Mean temperature of India is taken as the output.

4 Results and Conclusion

For three values of 'k' all three models were tested for the test set and the analysis was made using Mean Absolute Percentage Error which shows that KNN IDW gives a minimum MAPE value when compared to the other models KNNAV and KNNEXP. Figure1, 2 and 3 shows the graph of actual and forecasts of various models when k =3,5and 7 respectively. Table 1 gives the MAPE values for the three models for the three 'k'values.

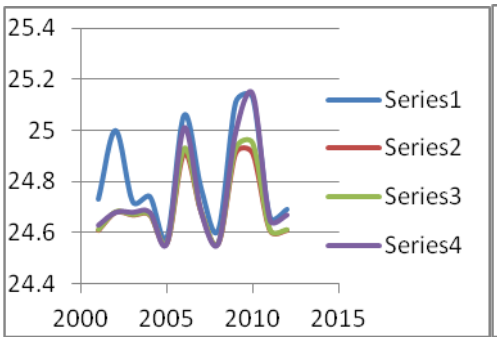

Fig .1: Graph of Actuals and Forecasts of various models when k =3

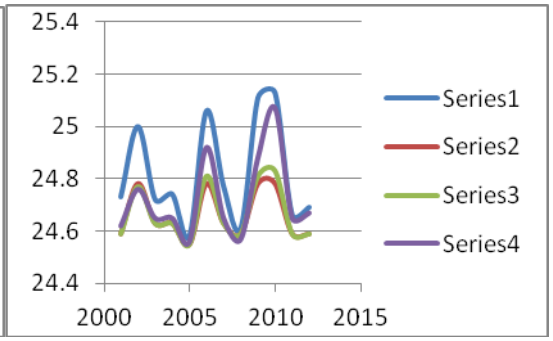

Fig.2: Graph of Actuals and Forecasts of various models when k =5

Table. 1 : MAPE values for different models

'k' value	MAPE for average	MAPE for exponential	MAPE for IDW
3	0.475	0.44	0.30
5	0.63	0.59	0.38
7	0.745	0.705	0.48

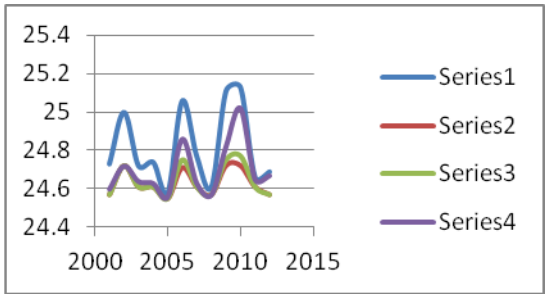

Fig.3: Graph of Actuals and Forecasts of various models when k =7

In the above figures 1, 2 and 3, Series 1, 2, 3 and 4 denote the values of Actual, Forecast by average, Forecast by exponential and Forecast by IDW respectively.

The results have shown a lower mean absolute percentage error for this model. Thus in this article a model for predicting annual mean temperature of India was made.

References

[1] Khusbhu Khamar, Short Text Classification Using kNN Based on Distance Function, *International Journal of Advanced Research in Computer and Communication Engineering*, **2** (2013), 1916-1919.

[2] Y. Radhika and M. Shashi, Atmospheric Temperature Prediction using Support Vector Machines, *International Journal of Computer Theory and Engineering*, **1**(2009), 55-58. http://dx.doi.org/10.7763/ijcte.2009.v1.9

[3] Sagar S. Badhiye, Nilesh U. Sambhe and P. N. Chatur, KNN Technique Analysis and Prediction of Temperature and Humidity Data, *International Journal of Computer Applications*, **61** (2013), 7-13. http://dx.doi.org/10.5120/9994-4847

[4] Dr. S. Santhosh Baboo and I. Kadar Shereef, An Efficient Temperature Prediction System using BPN Neural Network, *International Journal of Environmental Science and Development*, **2** (2011), 49-54. http://dx.doi.org/10.7763/ijesd.2011.v2.96

[5] Shaminder Singh, Pankaj Bhambri and Jasmeen Gill, Time Series based Temperature Prediction using Back Propagation with Genetic Algorithm Technique, *International Journal of Computer Science Issues*, **8** (2011), 28-32.

[6] www.statsoft.com/textbook/k-nearest-neighbour.

Received: December 1, 2014; Published: January 14, 2015