

Research Article

Weather Forecasting Using Sliding Window Algorithm

Piyush Kapoor¹ and Sarabjeet Singh Bedi²

¹ *Kvantum Inc., Gurgaon 122001, India*

² *MJP Rohilkhand University, Bareilly 243006, India*

Correspondence should be addressed to Piyush Kapoor; piyushkapoor7@yahoo.com

Received 7 June 2013; Accepted 19 August 2013

Academic Editors: W.-L. Hwang and G. A. Tsihrintzis

Copyright © 2013 P. Kapoor and S. S. Bedi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To predict the future's weather condition, the variation in the conditions in past years must be utilized. The probability that the weather condition of the day in consideration will match the same day in previous year is very less. But the probability that it will match within the span of adjacent fortnight of previous year is very high. So, for the fortnight considered for previous year a sliding window is selected of size equivalent to a week. Every week of sliding window is then matched with that of current year's week in consideration. The window best matched is made to participate in the process of predicting weather conditions. The prediction is made based on sliding window algorithm. The monthwise results are being computed for three years to check the accuracy. The results of the approach suggested that the method used for weather condition prediction is quite efficient with an average accuracy of 92.2%.

1. Introduction

Weather forecasting is mainly concerned with the prediction of weather condition in the given future time. Weather forecasts provide critical information about future weather. There are various approaches available in weather forecasting, from relatively simple observation of the sky to highly complex computerized mathematical models. The prediction of weather condition is essential for various applications. Some of them are climate monitoring, drought detection, severe weather prediction, agriculture and production, planning in energy industry, aviation industry, communication, pollution dispersal, and so forth, [1]. In military operations, there is a considerable historical record of instances when weather conditions have altered the course of battles. Accurate prediction of weather conditions is a difficult task due to the dynamic nature of atmosphere. The weather condition at any instance may be represented by some variables. Out of those variables, one found that the most significant are being selected to be involved in the process of prediction. The selection of variables is dependent on the location for which the prediction is to be made. The variables and their range always vary from place to place. The weather condition of any day has

some relationship with the weather condition existed in the same tenure of precious year and previous week.

A statistical model is designed [2] that could predict the rainfall and temperature with the help of past data by making use of time-delayed feed forward neural network. Artificial neural network was combined with the genetic algorithm to get the more optimized prediction [3]. An improved technique that uses artificial neural network with photovoltaic system was proposed by Isa et al. [4] that utilizes perceptron model with Levenberg Marquardt algorithm. Apart from neural network Fuzzy logic has also been being used in weather prediction models. The rainfall was classified into three fuzzy sets which can be predicted by making use of simple fuzzy rules [5]. Also a fuzzy self-regression model was proposed by Lu Feng and Xu xiao Guang [6] which makes use of the form of self-related sequence number according to observed number. The self-related coefficients were computed by making use of Fuzzy Logic [7]. A combined approach of neural network with Fuzzy Logic is being proposed for the weather prediction system. The work has applied principle component analysis technique to the fuzzy data by making use of Autoassociative neural networks.

But the major shortcoming in the techniques proposed above that they utilized the previous weather conditions to predict the ones in future, but the underlying relationship that exists between previous data had not been being mathematically described and analyzed. The techniques using artificial neural networks (ANN) were only concerned with the adjustment of weights in order to get correct output from the given input. But no relationship among the data was mathematically defined. Also the ANN techniques suffered from anomalies like local minima, overfitting, and so forth. Another problem is that it is hard to decide how much training data is sufficient to adjust weights so that optimal accuracy of the predicted weather conditions can be achieved. The number of other techniques for weather forecasting that used regression with machine learning algorithms was proposed in [8, 9]. But a mathematical model that could represent the relationship among previous data that could be used for prediction is still desired. A new sliding window approach for the same is being proposed in this text for weather prediction.

2. Proposed Work

2.1. Methodology. There is always a slightly variation in weather conditions which may depend upon the last seven days or so variation. Here variation refers to difference between previous day parameter and present day's parameter. Also there exists a dependency between the weather conditions persisting in current week in consideration and those of previous years. In this work a methodology is being proposed that could mathematically model these two types of dependency and utilize them to predict the future's weather conditions. To predict the day's weather conditions this work will take into account the conditions prevailing in previous week, that is, in last seven days which are assumed to be known. Also the weather condition of seven previous days and seven upcoming days for previous year is taken into consideration. For instance if the weather condition of 16 November 2012 is to be predicted then we will take into consideration the conditions from 09 November 2012 to 15 November 2012 and conditions from 09 November to 22 November 2011 for previous years. Now in order to model the aforesaid dependencies the current year's variation throughout the week is being matched with those of previous years by making use of sliding window. The best-matched window is selected to make the prediction. The selected window and the current year's weekly variations are together used to predict the weather condition. The reason for applying sliding window matching is that the weather conditions prevailing in a year may not lie or fall on exactly the same date as they might have existed in previous years. That is why seven previous days and seven ongoing days are being considered. Hence a total period of fortnight is checked in previous condition to find the similar one. Sliding window is quite good technique to capture the variation that could match the current year's variation.

2.2. Sliding Window Algorithm. The work proposes to predict a day's weather conditions. For this the previous seven days weather is taken into consideration along with fortnight weather conditions of past years. Suppose we need to predict

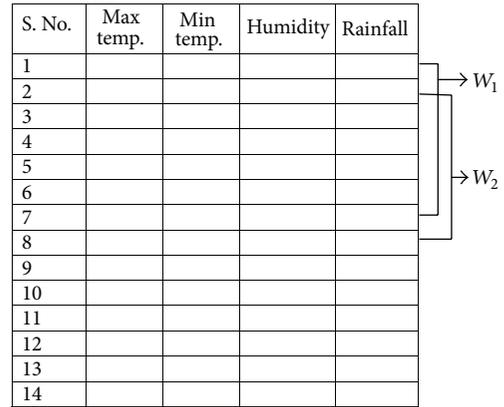


FIGURE 1: Sliding window concept where W_1 represents Window number 1 and W_2 represent window number 2.

weather of 23rd August 2013 then we will take into consideration the weather conditions of 16th August 2013 to 22nd August 2013 along with the weather conditions prevailing in the span of 16th August to 29th August in past years. Then the day by day variation in current year is computed. The variation is also being computed from the fortnight data of previous year. In this work the four major weather parameters will be taken into consideration, that is, maximum temperature, minimum temperature, Humidity and Rainfall. Hence the size of the variation of the current year will be represented by matrix of size 7×4 . And similarly for past year the matrix size would be 14×4 . Now, the first step is to divide the matrix of size 14×4 into the sliding windows. Hence, 8 sliding windows can be made of size 7×4 each. The concept of sliding window is shown in Figure 1.

Now the next step is to compare every window with the current year's variation. The best-matched window is selected for making the prediction. The Euclidean distance approach is used for the purpose of matching. The reason for taking Euclidean distance is its power to represent similarity in spite of its simplicity. Following are the parameters used for the weather condition prediction:

- (1) mean: mean of day's weather conditions, that is, maximum temperature, minimum temperature, humidity, and rainfall. After adding each separately, and divide by total day's number

$$\text{Mean} = \frac{\text{Sum of parameter}}{\text{number of days}}, \quad (1)$$

- (2) variation: calculate day by day variation after taking difference of each parameter. This tells how the next day's Weather is related to previous day's weather;
- (3) euclidean distance: it compares data variation of current year and previous year.

By this we are able to mathematically model the aforesaid defined dependencies. That the relationship between previous year and previous week data is being defined mathematically can be used to predict the future conditions.

Step 1. Take matrix “CD” of last seven days for current year’s data of size 7×4 .
 Step 2. Take matrix “PD” of fourteen days for previous year’s data of size 14×4 .
 Step 3. Make 8 sliding windows of size 7×4 each from the matrix “PD” as $W_1, W_2, W_3, \dots, W_8$
 Step 4. Compute the Euclidean distance of each sliding window with the matrix “CD” as $ED_1, ED_2, ED_3, \dots, ED_8$
 Step 5. Select matrix W_i as
 $W_i = \text{Corresponding_Matrix}(\text{Min.}(ED_i))$
 $\forall i \in [1, 8]$
 Step 6. For $k = 1$ to n
 (i) For WC_k compute the variation vector for the matrix “CD” of size 6×1 as “VC”.
 (ii) For WC_k compute the variation vector for the matrix “PD” of size 6×1 as “VP”.
 (iii) $\text{Mean}_1 = \text{Mean}(\text{VC})$
 (iv) $\text{Mean}_2 = \text{Mean}(\text{VP})$
 (v) Predicted Variation “V” = $(\text{Mean}_1 + \text{Mean}_2)/2$
 (vi) Add “V” to the previous day’s weather condition in consideration to get the predicted condition.
 Step 7. End

ALGORITHM 1

The sliding window used for predicting the “ n ” number of weather conditions ($WC_1, WC_2, WC_3, \dots, WC_n$) is shown in Algorithm 1.

The main logic behind using sliding window approach is that the weather conditions prevailing at some span of day in the year might not have existed in the same span of days in previous year. For instance the weather condition in first week of February 2010 might not have existed in the first week of February in 2009. The similar weather conditions might have prevailed in previous year but not necessarily in same week but in some days. The probability of finding the similar weather conditions are maximum at the considered fortnight spam.

3. Results and Discussion

The previous algorithm is being tested against weather data for the years 2006 to 2010 of the Champawat city, Uttaranchal. The data has been taken from Pantnagar Weather Forecasting Centre. The algorithm has been executed and tested in Matlab 2010a version. Thus, in the algorithm in consideration the previous year’s data is being utilized for predicting the weather conditions. Hence, the algorithm is tested to predict weather condition for three years, that is, 2008–2010, which is being tested against the available data. Also it can be concluded that learning approach used in the algorithm is supervised. In the test four weather conditions are taken into consideration, that is, minimum temperature, maximum temperature, humidity and rainfall. Temperature, in general, can be measured to a higher degree of accuracy relative to any of the other weather variables. The data of these four factors are taken daywise for the previously mentioned four years. The algorithm is also being tested daywise.

Figures 2, 3, 4, and 5 show the variation of actual and predicted four weather conditions for the year 2010 day wise.

These graphs clearly shows least variation among the actual and predicted weather conditions. The monthwise accuracy of predicted weather conditions is being given in Table 1.

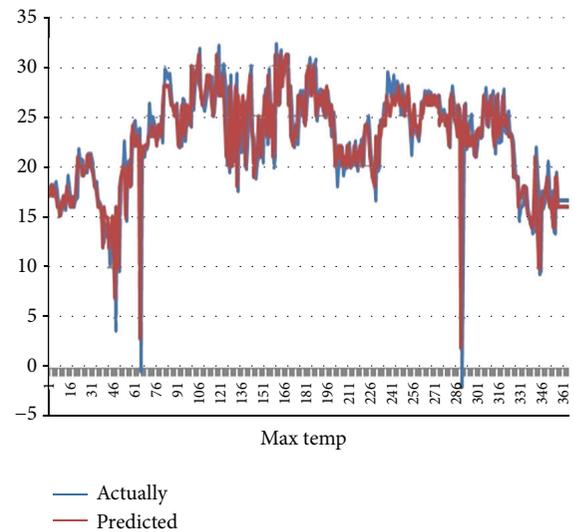


FIGURE 2: A graph representing predicted versus actual maximum temperature for year 2010.

TABLE 1

Month	Accuracy of predicted data
Jan 10	97.24%
Feb 10	88.94%
Mar 10	92.25%
Apr 10	98.17%
May 10	93.49%
Jun 10	99.02%
Jul 10	87.19%
Aug 10	78.83%
Sep 10	79.23%
Oct 10	94.35%
Nov 10	98.78%
Dec 10	95.73%

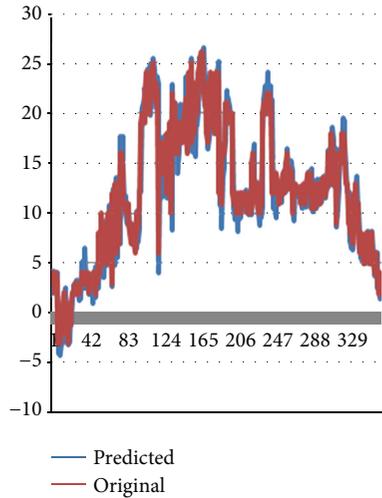


FIGURE 3: A graph representing predicted versus actual minimum temperature for the year 2010.

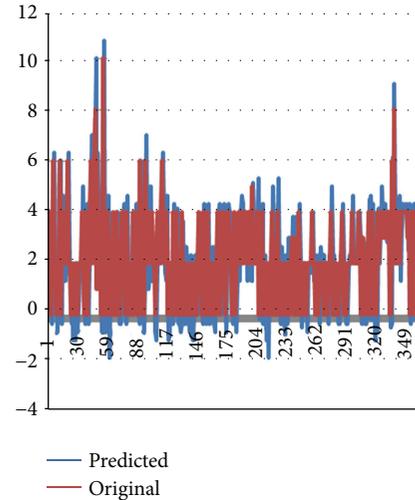


FIGURE 5

FIGURE 5: A graph representing predicted versus actual rain for the year 2010.

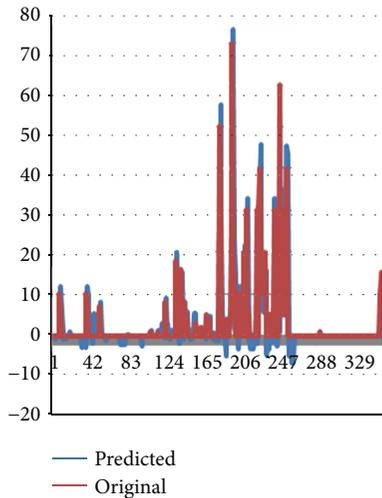


FIGURE 4: A graph representing predicted versus actual humidity for the year 2010.

The above result of weather conditions have been from an Indian city. India has a typically tropical type of weather, that is, the weather which has all varieties. The Champawat city lies in the state of Uttaranchal which lies in the plains of Ganges. The monthwise accuracy in Table 1 can be understood by the following facts. The months of April, May, and June are considered to be of summers which correspond to high temperature. The months of November, December, and January are winters having low or cold temperature conditions. Thus, the factors like temperature are quite fixed in these months and hence the accuracy for these is also high. In contrast the months like February, March, August, and September are considered to be the months when weather changes, that is, a phase of transition from one season to another. In the months of February and March, the winter season is shifted to summer. And in the month of August and September, the summer is getting over and, winter starts

coming. And hence the weather condition becomes highly unpredictable in these months. Also it is observed that the weather conditions vary greatly in these months from year to year. This is also being reflected in the results.

4. Conclusion and Future Work

The comparison of weather condition variation using sliding window approach has been found to be highly accurate except for the months of seasonal change where conditions are highly unpredictable. The results can be altered by changing the size of the window. Accuracy of the unpredictable months can be increased by increasing the window size to one month. Since ANN techniques are very good in mapping Inputs and outputs, the sliding window algorithm if incorporated with ANN could improve the results drastically even for the months of seasonal change.

References

- [1] Y. Radhika and M. Shashi, "Atmospheric temperature prediction using support vector machines," *International Journal of Computer Theory and Engineering*, vol. 1, no. 1, pp. 1793–8201, 2009.
- [2] L. L. Lai, H. Braun, Q. P. Zhang et al., "Intelligent weather forecast," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 4216–4221, Shanghai, China, August 2004.
- [3] J. Gill, B. Singh, and S. Singh, "Training back propagation neural networks with genetic algorithm for weather forecasting," in *Proceedings of the 8th IEEE International Symposium on Intelligent Systems and Informatics (SIISY '10)*, pp. 465–469, September 2010.
- [4] I. S. Isa, S. Omar, Z. Saad, N. M. Noor, and M. K. Osman, "Weather forecasting using photovoltaic system and Neural Network," in *Proceedings of the 2nd International Conference*

on Computational Intelligence, Communication Systems and Networks, pp. 96–100, July 2010.

- [5] L. Zuoyong, C. Zhenpei, and L. Jitao, “A model of weather forecast by fuzzy grade statistics,” *Fuzzy Sets and Systems*, vol. 26, no. 3, pp. 275–281, 1988.
- [6] L. F. Lu Feng and X. X. G. Xu xiao Guang, “A forecasting model of fuzzy self-regression,” *Fuzzy Sets and Systems*, vol. 58, no. 2, pp. 239–242, 1993.
- [7] T. Denoeux and M.-H. Masson, “Principal component analysis of fuzzy data using autoassociative neural networks,” *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 3, pp. 336–349, 2004.
- [8] R. Collobert and S. Bengio, “SVM Torch: support vector machines for large-scale regression problems,” *Journal of Machine Learning Research*, vol. 1, no. 2, pp. 143–160, 2001.
- [9] P.-S. Yu, S.-T. Chen, and I.-F. Chang, “Support vector regression for real-time flood stage forecasting,” *Journal of Hydrology*, vol. 328, no. 3-4, pp. 704–716, 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

