

An efficient approach for Weather forecasting using Support Vector Machines

Tarun Rao ¹, N Rajasekhar ² Dr T V Rajinikanth³

¹ Lead, Education and Research, Infosys Limited

² Assistant Professor, VNRVJIET, JNTU University

³ Professor & Head of the Department, GRIET, JNTU University

Abstract. Weather forecasting vital role in the area of investigating the weather parameters like time series data of daily maxima and minima temperatures, relative humidity, highest rain fall, sudden weather destructions, abnormal weather changes have been forecasting or predicting using many data mining techniques, as well as from the weather parameters itself. Support Vector Machine (SVM) represents a novel approach neural network technique, which is used for classification, regression analysis and forecasting. In this paper, we shall be showing a novel approach for weather forecasting using SVM. The experiment study and outcome is compared with Multi Layer Perceptron Learning algorithm.

Keywords: Back propagation, Weather forecasting, Multi Layer Perceptron, Support Vector Machine

1. Introduction

Recently, weather forecasting/prediction is very complex process and one of the very key tasks for researchers and academicians [1] [2] [3]. Now the weather forecasting is continuously increasing based on the traditional users, such as agriculture, air traffic services and other sectors like energy, environment, that require reliable information on the present and future weather. In addition to this, the forecasters have to cope with increasing data volume, notably from Numerical Weather Prediction models; meteorological satellites, radars and other observation systems such as AWS, wind profilers and radiometers etc. Forecasters focus on weather forecasting capability instead of losing time accessing the information, Enhance and upgrade their forecasting proficiency with many types of data presentation. Share their proficiency with their colleagues and transfer their know-how to junior forecasters, Browse, visualize and animate all the existing data intuitively with a limited number of actions, Take advantage of multi-screens with a GUI (Graphical User Interface) based on multi-windows, Use collaborative tools for graphical production of expertise data. Accurate forecast of weather parameters is a very difficult task due to dynamic nature of the atmosphere. Various techniques like linear regression, auto regression, Multi Layer Perception (MLP), Radial Basis Function networks are applied to calculate approximately atmospheric parameters like temperature, wind speed, rainfall, meteorological pollution etc. [4][5] [6] [7] [8] [9].

Weather information that also includes temperature, wind speed, sky cover, humidity etc., have also been used in many of the short - term load forecasting works [10]. But there is one complication: for mid-term load forecasting, temperature information for numerous weeks in advance is needed. If we wish to use temperature information in our model, we will also want to foretell the temperature. The usage of temperature, therefore, is not an only option when forecasting is performed for period longer than one week. Yet, the temperature forecasting is a much complex problem than load forecasting.

A neural network model is a structure which can be attuned to create a mapping from a given set of data to features of or relationships between the data. The model is attuned, or trained, using a compilation of data from a known source as input, usually referred to as the training set. An Artificial Neural Network (ANN)

[11] is an information processing model which is stirred by the method biological nervous systems, such as the brain, processes information. A back propagation network [12] consists of at least three layers (multi layer perception): an input layer, at least one intermediary hidden layer, and an output layer.

2. BACKGROUND

In recent years, weather prediction as drawn much attention in many research communities because it helps in safeguarding human life and their wealth. A part from that it is useful for enhancing natural calamities, agriculture yield growth, air traffic control, marine navigation, forest growths and defense purposes.

2.1. Support Vector Machines(SVM):

The Support Vector Machine (SVM), well known as kernel machine, was developed at AT & T Bell laboratories by Vapnik and his team [13], and it is a tool that works based on the statistical learning theory. The principal thought behind the Support Vector Machines is that it tries to map the primal data X into a feature space termed as F with a high dimensionality through a non-linear mapping function and thus builds the best possible hyper plane in a novel space. An application using Support Vector Machines (SVMs) for weather prediction was presented in [1]. Time series data of maximum temperature at various locations on a day to day basis is deliberated so as to forecast the maximum temperature of the subsequent day at that location depending on the daily maximum temperatures for a period of preceding n days referred to by the order of the input. Performance of the method is analyzed for a combination of spans of 2 to 10 days by making use of the best possible values of the kernel.

Another supervised learning method that can be used for estimation tasks is Support Vector Regression (SVR). The SVR algorithm is an addition to the accepted classification tool Support Vector Machines (SVM). SVM is a machine learning tool which has its ancestry in statistical learning theory [14].

In this paper we devise the linear support vector regression to predict the maximum weather at a location. In order to solve regression problems, we are given training data $(x_i, y_i), \forall i = 1, 2, 3 \dots l$, where 'x' is a d-dimensional input with $x \in R^d$ and the output is $y \in R$. The linear regression model can be written as shown below [15]:

$$f(x) = \langle \omega, x \rangle + b, \forall \omega, x \in \mathfrak{R}^d, b \in \mathfrak{R} \dots \dots \dots (1)$$

Where $f(x)$ is a target function and $\langle \cdot, \cdot \rangle$ denotes the dot product in \mathfrak{R}^d

So, we measure the empirical risk [15], we should state a loss function. Several other alternatives are available. The most common loss function is the ϵ -insensitive loss function. The ϵ -insensitive loss function proposed by Vapnik is defined the following function:

$$L_\epsilon(y) = \begin{cases} 0 & |f(x) - y| \leq \epsilon \\ |f(x) - y| - \epsilon & \text{otherwise} \end{cases} \dots \dots \dots (2)$$

the best possible parameters and b in Eq.(1) are found by solving the primal optimization problem(L.P.Wang,[15]):

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i^- + \xi_i^+) \dots \dots \dots (3)$$

with constraints: $y_i - \langle \omega, x_i \rangle - b \leq \epsilon + \xi_i^+, \langle \omega, x_i \rangle + b - y_i \leq \epsilon + \xi_i^-, \xi_i^-, \xi_i^+ \geq 0, i = 1, 2, \dots, l$

where C is a pre-defined value which determines the trade-off in between the flatness of f(x) and the amount up to which deviations better than the precision are tolerated. The slack variables ξ_i^+ and ξ_i^- represent the deviations from the constraints of the ϵ -tube. This primal optimization problem can as well be reformulated as a dual one defined as follows:

$$\max_{x, x^*} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i^* - a_i)(a_j^* - a_j) \langle x_i, x_j \rangle + \sum_{i=1}^l y_i (a_i^* - a_i) - \epsilon \sum_{i=1}^l (a_i^* + a_i) \dots \dots \dots (4)$$

with constraints: $0 \leq a_i, a_i^* \leq C, i = 1, 2, \dots, l$ and $\sum_{i=1}^l (a_i - a_i^*) = 0$. Solving the optimization problem defined by Eq.(4) and these constraints gives the best possible Lagrange multipliers α and α^* , while $\bar{\omega}$ and \bar{b} are given by $\bar{\omega} = \sum_{i=1}^l (a_i^* - a_i)x_i$, $\bar{b} = -\frac{1}{2} \langle \bar{\omega}, (x_r + x_s) \rangle$, where x_r, x_s are the support vectors. According to the computed value of $\bar{\omega}$, the $f(x)$ in Eq.(1) can be written as:

$$f(x) = \sum_{i=1}^N (a_i - a_i^*) \langle x_i, x \rangle + b \dots \dots \dots (6)$$

Hence is the precise formulation of the cost function and the use of the Lagrange theory. This solution has several interesting properties. It can be proven that the solution thus found is always global because the problem is convex.

3. MLPs are trained with back propagation algorithm

Artificial Neural Networks (ANNs) are a parallel as well as dynamic system of vastly interconnected interacting parts based on the neurobiological models. There is closed analogy between the structure of a biological neuron and a processing element of an ANN called artificial neuron. Each neuron accepts input signal in the form of mathematical amount then generates the output signal which in turn activates other neurons that directly connected to it. As an outcome, the input signal of a neuron is considered the summary signal of all the outputs of neurons standing before it in the network. The activation mechanism of each neuron depends on its inner method is called activation function and the signals communicates between neurons are represented in weights.

A multi-layer network (alongwith one or more than one hidden layers) can learn any uninterrupted mapping to an arbitrary precision [16]. One hidden layer is considered adequate for data learning model coping with complicated data factors.

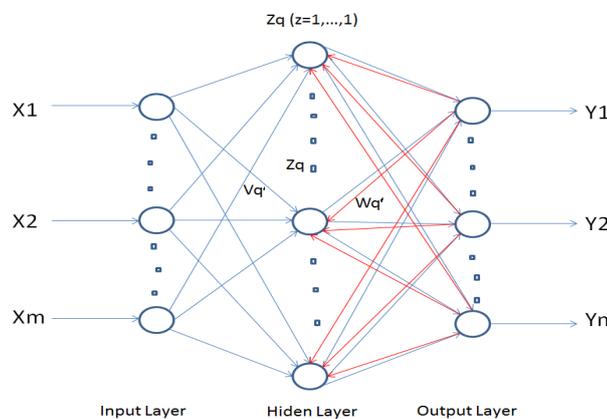


Fig. 1: Three-layered feed forward neural network.

Algorithm:

- Step 1: Initialize the weights that are present in the network (often randomly)
 - Step 2: Do
 - Step 3: For every example e in the training set
 - Step 4: O:= neural-net-output (network, e) ; then forward pass
 - Step 5: T:= teacher output for e
 - Step 6: Calculate the error (T - O) at the said output units
 - Step 7: Compute delta_wh for all the said weights from hidden layer to the output layer; then backward Pass
 - Step 8: Compute delta_wi for all the weights from input layer to the hidden layer; backward pass
- Continued

Step 9: Update all the weights in the network

Step 10: Until all the examples classified properly or stopping criterion are satisfied

Step 11: Return the said network

In the fig.1 the output of neuron in a layer form which moves the all the neuron. Therefore each neuron has its very own input weights, the weights are fixed to one for each input layer i.e. weights are not changed. The output is obtained by the applying the input values to the input layer, while passing the output of every neuron to the following layer as input.

4. Experimental Analysis

The real world databases are highly dirty to noisy and missing values. Therefore the data can be normalized and preprocessed to improve the quality, prediction results of the data. In this work we collected data sets from University of Cambridge for a span of 5 years (2003-07) is used so as to build the models. In this experiment maximum temperature of a day is estimated based on the maximum temperature of previous n days where n represents the best possible length of the span. The value of n is established by the method of experimentation. The available data is again sub-divided into training, validation and test sets. Training set is used so as to build the model, validation set is used so as to perform parameter optimization and finally test set is used so as to assess the model. Separate models are developed using SVM and MLP making use of back propagation algorithm.

The performance of Multi Layer Perceptron(MLP) trained along with the back propagation algorithm and SVM for diverse orders in terms of Mean Square Error (MSE) is established in Fig 2. From the derived outcomes it has been noticed that the window size will not have a key effect on the performance of MLP trained along with the back propagation algorithm and SVM. Nonetheless it can also be noticed that irrespective of the order, SVM performs superior to MLP. The Mean Square Error in the case of MLP varies from the values 8.10 to 11.1 depending on the order, whilst it is in the range of 7.06 to 7.68 in the case of SVM's. It can be observed that in the case of SVM there is no apparent variation in the performance of the system ahead of the order 5 and MLP performed finest for order 5. The error seems to lessen to some extent for higher orders in the case of SVM's but the training time also increases proportionately along with the increase in order. Thus order 5 is selected as best possible window size.

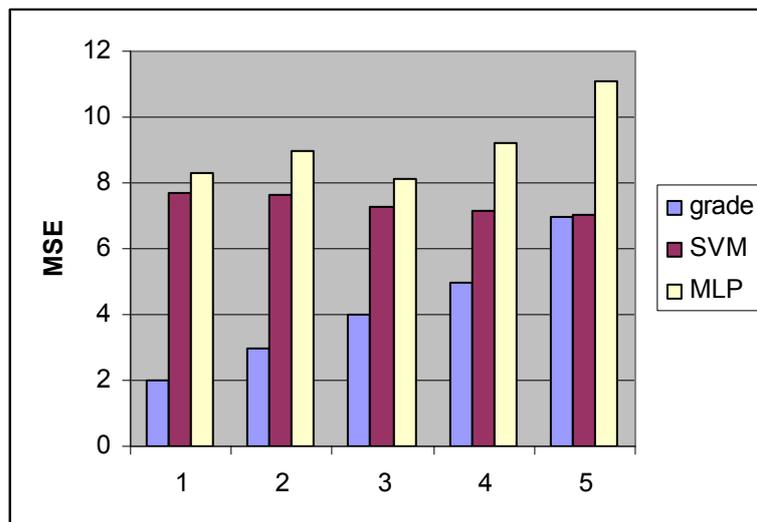


Fig. 2 Comparison in between MLP and SVM for diverse grades

5. Conclusion

In this paper, we have effectively proposed the weather estimate based on support vector regression model and we compared both the SVM with MLP for different degrees. Thus the outcome also shows that SVM performs better than MLP trained with back propagation algorithm for all degrees. It was observed that

parameter choice in SVM's case has a noteworthy effect on the by and large performance of the model. As a result we can conclude that through proper selection of the different parameters, Support Vector Machines can replace few of the neural network based models for weather prediction applications.

6. Acknowledgements

The author would like to thank all colleagues who contributed to this study. I am grateful to Infosys Limited for permitting me to come out with this paper.

7. 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, April 2009,
- [2] Denis Riordan, and Bjarne K Hansen, "A fuzzy case-based system for weather prediction." *Engineering Intelligent Systems*, Vol .10, No.3. 2002
- [3] Guhathakurtha, P., "Long-Range monsoon rainfall prediction of 2005 for the districts and Sub-division kerala with artificial neural network." *Current Science*, Vol.90, No.6, 25. 2006.
- [4] Jae H.Min., Young-chan Lee. "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters." *Expert Systems with Applications*, 28, pp.603-614. 2005.
- [5] Mohandes, M.A., Halawani, T.O., Rehman, S and Ahmed Hussain, A. "Support vector machines for wind speed prediction." *Renewable Energy*, 29, pp.939-947. 2004
- [6] Pal, N.R., Srimanta Pal, Jyotirmoy Das, and Kausik Majumdar, "SOFM-MLP: A Hybrid Neural Network for Atmospheric Temperature Prediction." *IEEE Transactions on Geoscience and Remote Sensing*, Vol.41, No, 12, pp.2783-2791. 2003
- [7] Pao-Shan Yu., Shein-sung Chen., I-Fan Chang. "Support vector regression for real- time flood stage forecasting." *Journal of Hydrology*, 328, pp. 704-716. 2006.
- [8] Stanislaw Osowski and Konrad Garanty, "Forecasting of daily meteorological pollution using wavelets and support vector machine." *Engineering Applications of Artificial Intelligence*, 20, pp.745-755. 2007.
- [9] Wei-Zhen Lu. Wen-Jian Wang. "Potential assessment of the support vector machine method in forecasting ambient air pollutant trends." *Chemosphere*, 59, pp.693-701. 2005.
- [10] A. Jain and B. Satish, "Clustering based short term load forecasting using support vector machines," in *PowerTech Conference*, Bucharest, Romania, July 2009.
- [11] Mohsen Hayati, and Zahra Mohebi," *Application of Artificial Neural Networks for Temperature Forecasting*," *World Academy of Science, Engineering and Technology* 28 2007.
- [12] Surajit Chattopadhyay," *Multilayered feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India* ," 2006.
- [13] Haykin, S., "Neural Networks- A Comprehensive Foundation." Prentice Hall. 1999.
- [14] Cortes, C. and V. Vapnik (1995). *Support vector networks*. *Machine Learning* 20 (3), 273–297.
- [15] L.P.Wang. (2005). *Support Vector Machines: Theory and Application*, Springer, Berlin, 2005.
- [16] Fundamentals of Neural Networks architectures, algorithms, and applications. Laurence Fausett, *Florida Institute of Technology*. Prentice-Hall, Inc. 1994.