Variable Selection and Prediction of Rainfall from WSR-88D Radar Using Support Vector Regression

BUDI SANTOSA
School of Industrial Engineering
University of Oklahoma
MICHAEL B. RICHMAN
School of Meteorology
University of Oklahoma
1100 East Boyd, Suite 1310, Norman, OK 73019, USA
THEODORE B. TRAFALIS
School of Industrial Engineering
University of Oklahoma
202 West Boyd, Suite 124, Norman OK, 73019, USA

Abstract:- This research utilizes linear programming support vector regression to perform variable selection and rainfall estimation. Variables selected from applying linear programming support vector regression are used to perform rainfall prediction tasks using standard support vector regression and a Bayesian neural network. Ground truth rainfall data are necessary to apply intelligent systems techniques. A unique source of such data is the Oklahoma Mesonet. With the advent of a national network of advanced radars (i.e., WSR-88D), massive archived data sets are available for data mining. The reflectivity measurements from the radar are used as inputs for the learning techniques tested. The application of linear programming support vector regression for variable selection is new for the estimation of rainfall by radar. Results show that by selecting subsets of pertinent variables, standard support vector regression is more accurate in terms of generalization error than application of either traditional regression or a rain rate formula used in meteorology. Moreover, support vector regression shows a better prediction than standard linear programming support vector regression, traditional linear regression and Bayesian neural network for rainfall estimation.

Key-Words:- Bayesian neural network, estimation, generalization error, mean squared error, regression analysis, linear programming support vector regression, neural networks, support vector regression

1. Introduction

Heavy rainfall, leading to flash floods, kills more people than lightning, tornadoes or hurricanes. Despite the impact of flash floods, our ability to estimate high rainfall rates from current state of the science technology is frequently inaccurate and, hence, there is ample impetus for improvement. Much of this inaccuracy arises from poor precipitation estimates from the Weather Surveillance Radar 1988 Doppler (WSR-88D) algorithms, which use only empirical techniques [5, 10, 11].

Regression models have been applied as an approach for modeling of linear and nonlinear systems. In regular regression, the errors are assumed to be normally, identically and independently distributed. Generally, those assumptions are not satisfied in real applications. In fact, the error distributions are likely to be contaminated by occasional bad values giving rise to outliers.

Alternative approaches to regression are Bayesian Neural Networks (BNNs) [6, 7], support vector machines (SVMs) and linear programming support vector regression (LP-SVR). These approaches are attractive as they relax distributional assumptions of the data..

SVMs, introduced by Vapnik [19], have been applied successfully to solve numerous problems in classification and regression. Müller et al. [9] used support vector regression (SVR) to predict time series data and compared the results with radial basis function networks, demonstrating that SVR had excellent performance. Trafalis and Santosa [16] used SVR along with a feedforward neural network and radial basis function (RBF) networks to predict

monthly flour prices in three cities. Again, results showed that SVR outperformed the two other methods. Linear programming support vector machine (LP-SVM) has been used successfully in classification and relevant feature identification in molecular profiling data [2] and for feature selection in tornado detection [17].

Since these techniques show promise in classification and prediction problems, the present work examines their efficacy for a rainfall estimation problem. The main objective of the paper is to develop an approach, new to radar meteorology, that uses Linear Programming Support Vector Regression (LP-SVR), SVR, and BNN to improve WSR-88D rainfall estimation over more traditional techniques.

Section 2 motivates and defines the problem. In sections 3 and 4 the data, preliminary analyses and the methodology are provided. Section 5 discusses the experimental setting and the analysis of the computational results. Section 6 concludes the paper.

2. Problem Statement

For over forty years, rainfall estimation from radar has been an active area of research. For the most part, the issue has been addressed through radar reflectivity-rainfall relations (known as Z-R). The Z-R relation was pioneered by Marshall and Palmer [8]. A method, based on matching the probabilities of the two variables, of deriving Z-R relation is presented by Rosenfeld et al. [13]. Currently, radar rainfall estimates are computed from a parametric Z-R relation that can be demonstrated in various ways. A common form of this relationship for rainfall rate (RR) can be written as

$$RR = 10\ C^{\ (0.0625)\ (Z)}$$

where C = 0.036 mm/hr and Z is the reflectivity. The RR equation is one of many developed for radar Z-R relations. Other Z-R relations have tuning parameters a, b in a regression format of $Z = a\ R^b$. The current WSR-88D default values for a and b are *300* and *1.4*, respectively [5]. However, the values of a and b vary from place to place, season-to-season, and time to time [20]. It is obvious that an estimation technique based on this relation will be suboptimal since no single value for a or for b will give a good estimate of the rainfall over a broad range of conditions. What does occur is a large uncertainty in estimating rainfall from reflectivity. The bias is not a constant and

cannot be corrected with existing algorithms. Therefore, it is essential to develop a new technique that reduce this variability of the Z-R relation.

The WSR-88D digital base data contains three variables: velocity (V), reflectivity (Z), and spectrum width (W). Current rainfall detection algorithms use only Z data [5]. This work proposes to

- determine the suitability of SVR, LP-SVR and BNN for precipitation prediction,
- compare the forecasts of precipitation given by SVR, LP-SVR and BNN to previous research, and
- determine the relevant variables for rainfall prediction.

By application of the SVR, LP-SVR and BNN, the aim of this research is to improve the precipitation processing subsystem (PPS) by creating new algorithms.

3. Data and Analysis

The rainfall data for calendar year 1998 are collected from the Chandler, Oklahoma Mesonet station. The Oklahoma Mesonet is a statewide network of 119 automated weather observing stations, designed with dense spatial coverage across the state [1]. Nine main parameters are measured at a time resolution of 5 minutes at all stations including rainfall. The Mesonet data are heavily quality controlled and archived for the period 1994 to present with 99.8 percent availability.

The radar (KTLX) used is located near Norman, OK and has complete coverage over a large number of Mesonet sites. The KTLX radar is situated approximately 54 km west-southwest of the Chandler raingauge. The WSR-88D radar has an effective range of approximately 230 km. The radar performs approximately 10 elevation scans that comprise a single volume scan. Each elevation scan is at a particular elevation angle between 0.5 and 15 degrees above the horizontal. For each elevation scan, the radar revolves a full 360 degrees about the vertical and makes 366 azimuthal scans within the revolution. This 54 km distance between the radar and raingauge is thought to be near optimal for sensing the atmosphere as the distance allows for measurement of the vertical extent of the convective clouds responsible for thunderstorms. For the five

azimuthals facing Chandler , OK, on the five lowest elevation scans, Z is retained. Z is measured once in every kilometer or at 230 points for every azimuthal scan and 5 one km range boxes are used for each of the 5 azimuth. Each volume scan measurement takes approximately 5 minutes to complete.

Given the 5-minute time resolution of both observing and measurement platforms, it is clear that there will be a large amount of data. For each elevation angle used (5), there are 25 radar spatial variables selected over and around the Chandler area. Therefore, the radar volume scan contains 125 input variables.   There were 576 five minute-period samples used for both the radar and raingauge information. Owing to the highly correlated structure of the data, it is convenient to preprocess these data and apply data reduction techniques. In the present study, we were able to reduce 125 input variables to 8 orthogonal dimensions accounting for 99 percent of the variance of the original data using principal components.

4. Methodology
4.1 Linear programming Support Vector Machines (SVM)

Suppose we have been given a set of training data $\{x_k, y_k\}_{k=1}^{\ell}$, with inputs $x_k \in R^n$ and ouputs $y_k \in R$. By support vector regression, one wants to find a function $f(x)$ that has at most ε deviation from the actual target y_k for all training data (Fig. 1).
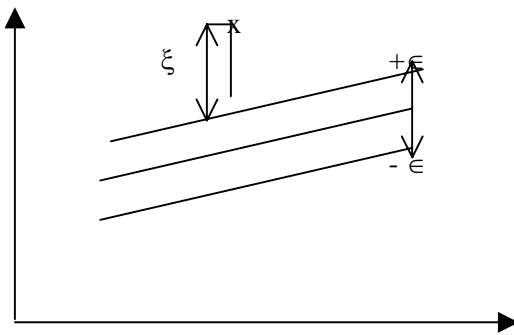

Figure 1. Linear SVR with slacks ξ. Points outside 2ε tube are penalized.

For the linear case, suppose we have the following function as a regressor:

$$f(x) = w^T x + b, \qquad (1)$$

where <.> denotes the dot product. The problems of regression and relevant variable identification can be solved concurrently by considering a sparse hyperplane, one for which the weight vector w has few non-zero elements. The hyperplane we are seeking is:

$$f(x) = w^T x + b = \sum_{p=1}^{P} w_p x_p + b = \sum_{\{p:w_p \neq 0\}} w_p x_p + b = 0$$

(2)

If a weight vector element is zero, $w_p = 0$, then feature p in the example vector does not decide the class of x and is thus "irrelevant". Only a variable for which the element is non-zero, $w_p \neq 0$, contributes to define hyperplane $f(x)=0$ and, thus is useful for regression. The problem of defining a small number of relevant variables can be thought of as synonymous with identifying a sparse hyperplane. Learning a sparse hyperplane can be formulated as an optimization problem. Minimizing the L_0 norm of the weight vector, $||w||_0$, minimizes the number of non-zero elements. The L_0 norm is $||w||_0 =$ number of $\{p: w_p \neq 0\}$. Unfortunately, minimizing an L_0 norm is NP-hard. However, a tractable, convex approximation is to replace the L_0 norm with the L_1 norm [4]. Minimizing the L_1 norm of the weight vector, $||w||_1$, minimizes the sum of the absolute magnitudes of the elements and sets most of the elements to zero. The L_1 norm is $||w||_1 = \sum_{p=1}^{P} |w_p|$.

We assume that there is a function f that approximates all pairs (x_i, y_i) with precision ε. In this case we assume that the problem is feasible. Therefore, we have to solve the following optimization problem in the primal weight space:

$$\min_{w,b} ||w||_1 \qquad (3)$$

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$$

$$||w||_1 = \sum_{p=1}^{P} |w_p|, \text{ where } |w_p| = \text{sign}(w_p)w_p.$$

Problem (3) can viewed as a special case of minimizing a weighted L_1 norm, $\min_{w} \sum_{p=1}^{P} a_p |w_p|$, in which the vector of weighting coefficients a is a unit vector, $a_p = 1, \forall p \in \{1,.., P\}$. In other words, all

variables are presumed to be equally good relevant variable candidates. Prior knowledge about the (un)importance of variable p can be encoded by specifying the value of a_p. In the case of infeasibility where some points might deviate from $f \pm \varepsilon$, one can introduce slack variables ξ, ξ^* to cope with infeasible constraints of the optimization problem. Then the above problem can be formalized as

$$\min_{w,b} \| w \|_1 + C \sum_{i=1}^{\ell} \xi_i \tag{4}$$

$$st \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi^*_i \geq 0 \end{cases}$$

The constant $C > 0$ determines the trade off between the flatness of function f and the amount up to which deviations larger than ε are tolerated. Any deviation more than ε will be penalized with C (Fig. 1). The value of C can be chosen more systematically via cross validation. Problem (4) can be recast as a linear programming problem by introducing extra variables u_p and v_p where $w_p = u_p - v_p$ and $|w_p| = u_p + v_p$. These variables are the pth elements of $u, v \in R^P$. The L_1 norm becomes $\| w \|_1 = \sum_{p=1}^{P} (u_p + v_p) = u + v$ and the problem can be rewritten in a standard form as follows:

$$\min_{w,b} u + v + C \sum_{i=1}^{\ell} \xi_i$$
$$st \quad y_i - \langle (u-v), x_i \rangle - b \leq \varepsilon + \xi_i \tag{5}$$
$$\langle (u-v), x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*, \ \xi_i, \xi_i^* \geq 0, \ i = 1,...\ell$$
$$u_p, v_p \geq 0, \ \forall p \in \{1,..,P\}$$

Problem (5) minimizes a linear function subject to linear constraints.

4.2 Standard Support Vector Regression (SVR)

The same problem described in section 4.1 can be formulated as a quadratic programming problem. For a detailed explanation see Schölkopf and Smola [14].

4.3 Bayesian Neural Network

The Bayesian evidence framework has been successfully applied to the design of multilayer perceptrons (MLPs) in the work of MacKay [6, 7]. Bayesian neural network is the extension of artificial neural networks (ANN). Bayesian methods have been proposed for neural networks to solve regression and classification problems. These methods claim to overcome some difficulties encountered in the standard approach, such as overfitting.

In conventional approaches, the training for ANN is based on the minimization of an error function, and is often motivated by some underlying principle such as maximum likelihood. The disadvantage of such approaches is that the designed networks can suffer from a number of deficiencies, including the problem of determining the appropriate level of model complexity. More complex models (e.g. ones with more hidden units, with more layers or with smaller values of regularization parameters) give better fits to the training data, but if the model is too complex it may give poor generalization (overfitting).

The Bayesian viewpoint provides a general and consistent framework for statistical pattern recognition and data analysis. In the context of neural networks, a Bayesian approach offers several important features including the following [3]:

- The conventional approach to network training, based on the minimization of an error function, can be seen as a specific approximation to a full Bayesian treatment.
- Similarly, the technique of regularization arises in a natural way in the Bayesian framework. The corresponding regularization parameters can be treated consistently within the Bayesian setting, without the need for techniques such as cross-validation.
- For classification problems, the tendency of conventional approaches to make overconfident predictions in regions of sparse data can be avoided.
- Bayesian methods provide an objective and principled framework for dealing with the issue of model complexity (for example, how to select the number of hidden units in a feed-forward network), and avoid many of the problems of overfitting which arise when using maximum likelihood.

5. Experiments
The SVR, LP-SVR and BNN [15] experiments are

performed in the MATLAB environment. First, LP-SVR is applied to identify the relevant variables with respect to the minimum mean squared error value of testing phase. MINOS solver [12], which is embedded in TOMLAB software package, is used to solve the linear programming problem resulting from LP-SVR formulation. The experiments are performed by applying the cross-validation technique to find the best constant cost C. After relevant variables are identified, then the data set with selected features is subject to SVR and BNN. For SVR, an RBF kernel function was used with some values of parameters C and spread σ (see section 4.1). The selection of parameters C and σ is based on cross-validation. It is found that the best parameters values are $C=10000$ and $\sigma = 0.7$ for this study. There are 10 samples for training and testing.

6. Results

Using LP-SVR reduces the number of variables from 125 to 30. The summary of the results is presented in Tables 1 and 2. The average testing MSE and R^2 are obtained by taking the average of over 10 testing samples. The best performance is given by SVR after variable selection (Table 2) which has average testing MSE of 0.236 and average testing R^2 of 0.627. Compared to using all variables (Table 1), there are 52% reduction in the average MSE and 43% increasing in average R^2. This decrease in error and increase in prediction is superior to previous results [18] using least square support vector machine (LS-SVR), with 8 principal components as inputs. That experiment [not shown] gave average MSE of 0.271 and average R^2 of 0.339.

Table 1. Average MSE and R^2 of testing with all variables (mm^2)

Method	Average Testing MSE	Average R^2
LP-SVR	0.5820	0.3751
SVR	0.3592	0.4373
BNN	0.7149	0.2822
LR	0.7125	0.2263

Table 2. MSE/R^2 of testing for selected variables (mm^2)

Method	Average Testing MSE	Average Testing R^2
LP-SVR	0.5820	0.3751
SVR	0.2360	0.6273
BNN	0.4899	0.4817
LR	0.4464	0.3845

7. Summary and Conclusions

Analysis of three-dimensional reflectivity patterns from the WSR-88D radar in Norman, OK for predicting rainfall rates are undertaken with LP-SVR, BNN and SVR. A 5 by 5 grid of 1 km boxes for the radar data is constructed, centered on the Chandler, OK Mesonet rain gauge. The goals of the research are variable selection and predictive data mining. From 125 variables, a reduction to 30 variables is possible using LP-SVR. This is a highly significant reduction resulting in a set of variables that enables. SVR to perform better compared to using all variables. Additionally, it is better than LP-SVR, BNN or the more traditional LR. Accordingly, variable selection is an important step in achieving more accurate reflectivity-rainfall parameterization. The technique should be tested operationally to determine the ability to predict flash floods more accurately.

References

[1] Basara, J. B., Brotzge, J. A., and Crawford, K. C. Investigating Land-Atmosphere Interactions Using the Oklahoma Mesonet. *Third Symposium on Integrated Observing Systems,* January 10-15, Dallas, TX, Amer. Meteor. Soc., 1999, pp. 120-123.

[2] Bhattacharyya, C. , Grate, L.R.,. Rizki, A., Radisky, D., Molina, F.J., Jordan, M.I., Bissell, M.J., Mian, I.S., Simultaneous Classification and Relevant Feature Identification in High-Dimensional Spaces: Application To Molecular Profiling Data, *Signal Processing*, Vol. 83 , Issue 4, 2003, pp 729 – 743.

[3] Bishop, C.M., *Neural Networks for Pattern Recognition*, University Press, Oxford, 1995.

[4] Donoho, D. and Huo, X. Uncertainty principles and ideal atomic decomposition, *Technical Report*, Statistics Department, Stanford University, 1999, http://www-stat.stanford.edu/~donoho/reports.html.

[5] Fulton, R. A., Breidenbach, J. P., Seo, D.-J., and Miller. D.A, The WSR-88D rainfall algorithm, *Weather and Forecasting,* Vol. 13*,* 1998, pp. 377-395.

[6] MacKay, D., A Practical Bayesian Framework for Backpropagation Networks, *Neural Computation*, Vol. 4, 1992, pp. 448-472.

[7] MacKay, D., The Evidence Framework Applied to Classification Networks, *Neural Computation*, Vol. 4, 1992, pp. 720-736.

[8] Marshall, J. S., and Palmer,W.M., The Distribution of Raindrops with Size. *Journal of Meteorology,* Vol. 5*,* 1948, pp. 165-166.

[9] Müller, K.R., Smola, A.J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., and Vapnik, V.N., Predicting Time Series with Support Vector Machines. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Articial Neural Networks* ICANN'97, Berlin. Springer Lecture Notes in Computer Science, Vol. 1327, 1997, pp. 999-1004.

[10] OFCM, Doppler Radar Meteorological Observations, Part B, Doppler Radar Theory and Methodology. *Federal Meteorological Handbook 11*, FCM-H11B-1990, Office of the Federal Coordinator for Meteorological Services and Supporting Research, Washington, D. C, 1990.

[11] OFCM, Doppler Radar Meteorological Observations, Part C, WSR-88D Products and Algorithm*s*. *Federal Meteorological Handbook 11*, FCM-H11C-1991, Office of the Federal Coordinator for Meteorological Services and Supporting Research, Washington, D. C., 1991.

[12] Murtagh, B. A. and Saunders, M.A., MINOS 5.5 USER'S GUIDE. *Technical Report SOL 83-20R*, Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, 1998.

[13] Rosenfeld, D., D. B. Wolff, and D. Atlas, General Probability-matched Relations Between Radar Reflectivity and Rain Rate. *Journal of Applied Meteorology, 32,* 1993, pp. 50-72.

[14] Schölkopf, B. and Smola, A., *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.

[15] Sigurdsson, S., Neural Regressor with Quadratic Cost Function, version1.0, http://mole.imm.dtu.dk/toolbox/ann/, 2002.

[16] Trafalis, T.B., Santosa, B. and Richman, M.B., Bayesian Neural Networks for Tornado Detection, *WSEAS Transaction on Systems*, Issue 10, Vol. 3, 2004, pp 3211-3216.

[17] Trafalis T.B., Santosa, B., Predicting Monthly Flour Prices through Neural Networks, RBFs and SVR, *Intelligent Engineering Systems Through Artificial Neural Networks*, (C.H. Dagli, A.L. Buczak, J. Ghosh, M Embrechts, O.Ersoy and S. Kercel, eds.), ASME Press, 11, 2001, pp.745-750.

[18] Trafalis, T., Santosa, B., and Richman, M., Prediction of Rainfall from WSR-88D Radar Using Kernel-Based Methods, *International Journal of Smart Engineering System Design*, 2004.

[19] Vapnik, V.N., *The Nature of Statistical Learning Theory*. John Wiley & Sons, New York, 1995.

[20] Wilson, J. W., and Brandes, E. A. Radar Measurement of Rainfall - A Summary. *Bulletin American Meteorological Society,*Vol. 60*,* 1979, pp.1048-1058.