Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine

Pijush Samui[1], Venkata Ravibabu Mandla[2], Arun Krishna[2] and Tarun Teja[2]
[1]Centre for Disaster Mitigation and Management
[2]School of Mechanical and Building Sciences, Center for Energy and Environment
VIT University, Vellore-632 014, TN, India
Email:pijush.phd@gmail.com

Abstract

This article adopts Support Vector Machine (SVM) and Relevance Vector Machine (RVM) for prediction of rainfall in Vellore (India). SVM is firmly based on the theory of statistical learning theory. RVM is a probabilistic basis model. SVM and RVM use air temperature (T), sunshine, humidity and wind speed (V_a) as input variables. This article uses SVM and RVM as a regression technique. Equations have been also developed for prediction of rainfall. The developed RVM gives variance of the predicted rainfall. This study shows the RVM is more robust model than the SVM.

Key Words: Rain-fall; SVM; RVM; Vellore district

Introduction

Rain is one of nature's greatest gifts and in third world countries like India; the entire agriculture depends upon rain. It is generally accepted that rainfall is unpredictable. However, there are certain periodicities in its long-term behavior (Nicholson and Entekhabi, 1986). Most of the deviations in the periodicity are related to solar cycles; variations in ocean currents and wind directions; sea surface temperature anomalies, etc. (Bhalme and Mooley, 1981; Ramage, 1971; Ananthakrishnan and Parthasarathy, 1984). Water resources assessment is defined by UNESCO and WMO (1988) as the process of assessing the source, scope, reliability, quantity and quality of water resources for the purposes of water resources utilization and management. Water resources assessment, which is important for water resources planning, is usually conducted using information that has been collected over a long period of time (over 20 years) and by normalizing historic series of observe driver discharges using water use data and water balance equations (Miloradov and Marjanovic,1998). Real-time water resources assessment can be defined as a rapid assessment of the water resources generated in a rainfall event or in a past period from a particular day of the year to the current rainfall event.

Applications of synthetic rainfall data may then be made in such diverse fields as flood modelling and urban drainage (Moretti and Montanari, 2004; Brath *et al.,* 2006; Dawson *et al.,* 2006; Hall *et al.,* 2006), pesticide fate modelling (Nolan *et al.,* in press), landslide modelling (Bathurst *et al.,* 2005), desertification vulnerability (Bathurst and Bovolo, 2004), water resource assessment (Fowler *et al.,* 2005) and flood risk assessment

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

(Kilsby *et al.,* 2000). Rainfall prediction has been the subject of extensive research in the recent past (Annamalai, 1995; Krishnakumar, 1994; Parthasarathy and Sontakke, 1988a; Parthasarathy *et al.,* 1988b; Prasad and Singh, 1992; Navone and Ceccatto, 1994). Application of neural networks for rainfall analysis and prediction is not new (Amin Talei *et al.,* 2010, Nayak *et al.,* 2006), Navone and Ceccatto, 1994; Sunyounget *et al.,* 1998). Navone and Ceccatto (1994) gave a brief introduction to neural networks and its application to predicting Indian monsoon rainfall in particular. They compared their results with conventional methods and showed that neural network based methods are able to produce reasonably accurate predictions compared to linear models popularly used for this purpose.

In this paper, we present support vector machine (SVM) and relevance vector machine (RVM) for studying and predicting the long-term variations in rainfall phenomena based on past observations. To illustrate the method, we have used the rainfall of Vellore town in Tamil Nadu State, India as an example. An hourly rain-fall flow data is been collected by Automatic Weather Station (AWS) at VIT-University campus (ISRO119 located at latitude: 12^0 91'N and longitude: 79^0 14'E measured at different ground levels) Vellore (Fig. 1).

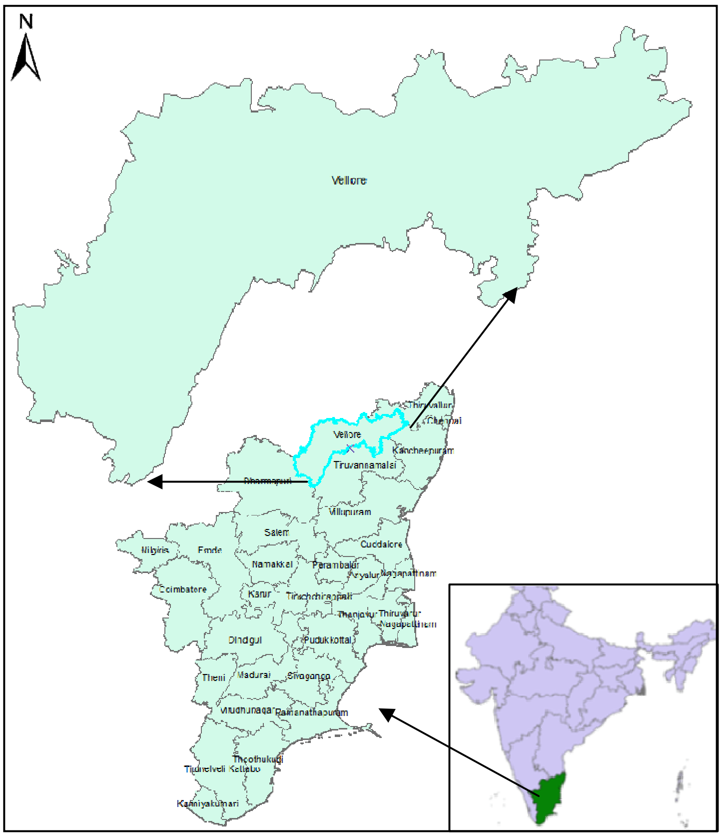

Fig. 1: Location map of study area.

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

Details of SVM

This study adopts SVM as a regression technique for prediction of rainfall. ε-insensitive loss function has used in this analysis. The expression of ε-insensitive loss function is given below:

$$L_{\varepsilon}(y) = 0 \text{ for } |f(x) - y| < \varepsilon \text{ otherwise } L_{\varepsilon}(y) = |f(x) - y| - \varepsilon \qquad (1)$$

Where x is input and y is output.

This defines an ε tube (Fig. 2) so that if the predicted value is within the tube the loss is zero, while if the predicted point is outside the tube, the loss is the magnitude of the difference between the predicted value and the radius, ε, of the tube.
Suppose f(x) takes the following form:

$$f(x) = (w.x) + b \quad w \in R^{n}, \qquad b \in r \qquad (2)$$

Where, w = is an adjustable weight vector, b = scalar threshold, R^n=n dimensional vector space and r=one dimensional vector space.

Then, we have to solve the following optimization problem:

$$\text{Minimize: } \frac{1}{2}\|w\|^2$$

$$\text{Subjected to: } y_i - (\langle w.x_i \rangle + b) \le \varepsilon, \; i = 1, 2, ..., l$$

$$(\langle w.x_i \rangle + b) - y_i \le \varepsilon, \; i = 1, 2, ..., l \qquad (3)$$

In the case where the constraints are infeasible, we introduce slack variables ξ_i and ξ_i^* (see Fig. 2) in optimization problem (3). The above optimization problem (3) can be written in the following way:

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)$$

$$\text{Subjected to: } y_i - (\langle w.x_i \rangle + b) \le \varepsilon + \xi_i, \; i = 1, 2, ..., l$$

$$(\langle w.x_i \rangle + b) - y_i \le \varepsilon + \xi_i^*, \; i = 1, 2, ..., l$$

$$\xi_i \ge 0 \text{ and } \xi_i^* \ge 0, \; i = 1, 2, ..., l \qquad (4)$$

The constant called capacity factor $0<C<\infty$ determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated (Smola and Scholkopf, 2004). This optimization problem (4) is solved by Lagrangian Multipliers (Vapnik, 1998), and its solution is given by:

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

$$f(x) = \sum_{i=1}^{nsv} \left(\alpha_i - \alpha_i^*\right)\left(x_i . x\right) + b \tag{5}$$

Where $b = -\left(\frac{1}{2}\right) w.[x_r + x_s]$, α_i, α_i^* are the Lagrangian Multipliers and nsv is the number of support vectors. An important aspect is that some Lagrange multipliers (α_i, α_i^*) will be zero, implying that these training objects are considered to be irrelevant for the final solution (sparseness). The training objects with nonzero Lagrange multipliers are called support vectors.

When linear regression is not appropriate, then input data has to be mapped into a high dimensional feature space through some nonlinear mapping (Boser et al. 1992). The two steps that are involved are first to make a fixed nonlinear mapping of the data onto the feature space and then carry out a linear regression in the high dimensional space. The input data is mapped onto the feature space by a map Φ. The dot product given by $\Phi(x_i).\Phi(x_j)$ is computed as a linear combination of the training points. The concept of kernel function [$K\left(x_i, x_j\right) = \Phi\left(x_i\right).\Phi\left(x_j\right)$] has been introduced to reduce the computational demand (Cristianini and Shwae-Taylor 2000; Cortes and Vapnik 1995). So, equitation (5) becomes written as:

$$f(x) = \sum_{i=1}^{nsv} \left(\alpha_i - \alpha_i^*\right) K\left(x_i . x_j\right) + b \tag{6}$$

Some common kernels have been used such as polynomial (homogeneous), polynomial (nonhomogeneous), radial basis function, Gaussian function, sigmoid etc for non-linear cases.

This study employs the above SVM model for prediction of rainfall in Vellore Town. We have collected 128 datasets. The data are normalized between 0 and 1. For developing SVM, the data are divided into the following two groups:

Training dataset: This is required to develop SVM model. This study uses 89 data for training.
Testing Dataset: This is required to evaluate the performance of the developed model. The remaining 39 data are used for testing dataset.
Radial basis function ($K(x_k, x) = \exp\left\{-\frac{(x_k - x)(x_k - x)^T}{2\sigma^2}\right\}$, Where σ is the width of radial basis function) has been used as a kernel function for the SVM model. The design

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

value of C and s will be determined by trail and error approach during training of SVM. The program of SVM has been constructed by using MATLAB.

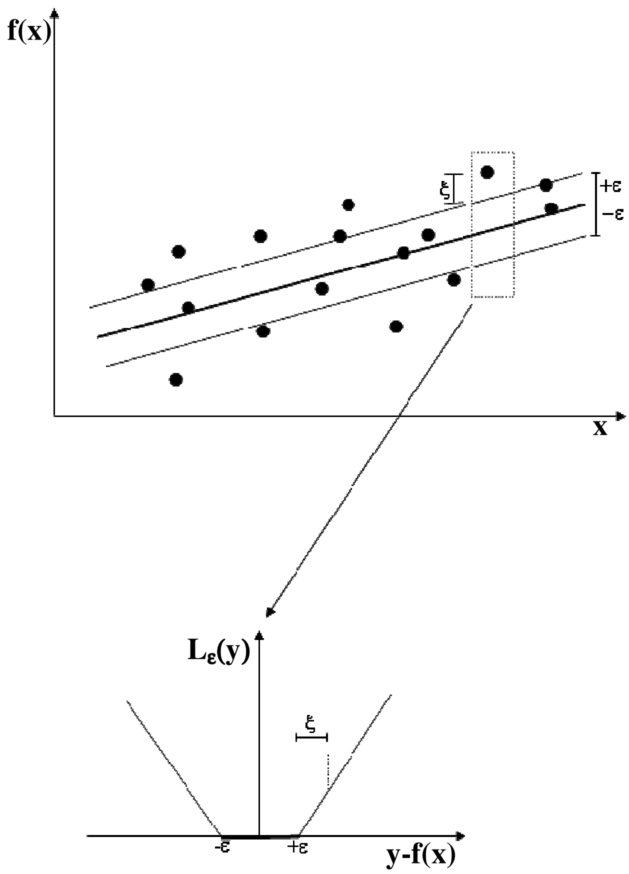

Fig 2: Prespecified Accuracy ε and Slack Variable ξ in support vector regression; Scholkopf (1997).

Details of RVM

The RVM was introduced by Bishop and Tipping (2000). The RVM uses the following generative model:

$$y(x;w) = \sum_{i=1}^{N} w_i k(x, x_i) + \varepsilon_n \qquad (7)$$

Where N is number of data points, $w = [w_1, ..., w_N]$ is weight, $k(x,x_i)$ is kernel function and $\varepsilon_n = N(0, \sigma^2)$ is error term with zero mean Gaussian process. The likelihood of the complete data set can be written as:

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

$$p\left(y/w,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\left\|y-\Phi w\right\|^2\right\} \quad (8)$$

Where $\Phi\left(x_i\right)=\left[1,K\left(x_i,x_1\right),K\left(x_i,x_2\right),\ldots,K\left(x_i,x_N\right)\right]^T$.

To prevent overfitting, automatic relevance detection (ARD) prior is set over the weights. An explicit zero-mean Gaussian prior probability distribution over the weights, w with diagonal covariance of α is proposed as follows:

$$p(w/\alpha)=\prod_{i=0}^{N}N\left(w_i/0,\alpha_i^{-1}\right) \quad (9)$$

With α is a vector of N+1 hyperparameters. Consequently, using Baye's rule, the posterior over all unknowns could be computed given the defined noninformative prior distribution:

$$p\left(w,\alpha,\sigma^2/t\right)=\frac{p\left(y/w,\alpha,\sigma^2\right)p(w,\alpha,\sigma)}{\int p\left(t/w,\alpha,\sigma^2\right)p\left(w,\alpha,\sigma^2\right)dw\,d\alpha\,d\sigma^2} \quad (10)$$

Full analytical solution of this integral (6) is obdurate. Thus decomposition of the posterior according to $p\left(w,\alpha,\sigma^2/t\right)=p\left(w/t,\alpha,\sigma^2\right)p\left(\alpha,\sigma^2/t\right)$ is used to facilitate the solution (Tipping, 2001). The posterior distribution over the weights is thus given by:

$$p\left(w/t,\alpha,\sigma^2\right)=\frac{p\left(t/w,\sigma^2\right)p(w/\alpha)}{p\left(t/\alpha,\sigma^2\right)} \quad (11)$$

The resulting posterior distribution over the weights is the multi-variate Gaussian distribution:

$$p\left(w/t,\alpha,\sigma^2\right)=\mathbf{N}(\mu,\Sigma) \quad (12)$$

Where the mean and the covariance are respectively given by:

$$\Sigma=\left(\sigma^{-2}\Phi^T\Phi+A\right)^{-1} \quad (13)$$

$$\mu=\sigma^{-2}\Sigma\Phi^T t \quad (14)$$

With diagonal $A=\mathrm{diag}(\alpha_0,\ldots,\alpha_N)$.

For uniform hyperpriors over α and σ^2, one needs only maximize the term $p\left(t/\alpha,\sigma^2\right)$:

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

$$p(t/\alpha,\sigma^2) = \int p(t/w,\sigma^2)p(w/\alpha)dw$$
$$= \left((2\pi)^{\frac{-N}{2}} \Big/ \sqrt{\left|\sigma^2 + \Phi A^{-1}\Phi^T\right|} \right) \times \exp\left\{ -\frac{1}{2} y^T \left(\sigma^2 + \Phi A^{-1}\Phi^T\right)^{-1} y \right\} \quad (15)$$

Maximization of this quantity is known as the type II maximum likelihood method (Berger, 1985; Wahba, 1985) or the "evidence for hyper parameter" (Mackay, 1992). Hyper parameter estimation is carried out in iterative formulae, *e.g.* gradient descent on the objective function (Tipping, 2001). The outcome of this optimization is that many elements of α go to infinity such that w will have only a few nonzero weights that will be considered as relevant vectors.

This study uses the above RVM model for prediction of rainfall in Vellore Town. In RVM model, the same training dataset, testing dataset, kernel function and normalization technique have been used as used in SVM model. The design value of σ has been determined by trail and error approach. The program of RVM has been constructed by using MATLAB.

Results and Discussion

This study adopts Coefficient of Correlation(R) to asses the performance of the developed SVM and RVM models. The value of R has been determined from the following equation:

$$R = \frac{\sum_{i=1}^{n}\left(r_{ai} - \bar{r}_a\right)\left(r_{pi} - \bar{r}_p\right)}{\sqrt{\sum_{i=1}^{n}\left(r_{ai} - \bar{r}_a\right)}\sqrt{\sum_{i=1}^{n}\left(r_{pi} - \bar{r}_p\right)}} \quad (16)$$

Where r_{ai} and r_{pi} are the actual and predicted r values, respectively, $\bar{r}_a$ and $\bar{r}_p$ are mean of actual and predicted E values corresponding to n patterns. For good model, the value of R should be close to one.

For SVM, the design value of C, ε and σ is 100, 0.02 and 1.1 respectively. For Best SVM model, the number of support vector is 84. Fig. 3 depicts the performance of training dataset. The performance of testing dataset has been illustrated in Fig. 4. The performance of training as well as testing dataset is almost same. So, the developed SVM does not show any overtraining. Therefore, it has good generalization capability. The developed SVM gives the following equation (by putting

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

$K(x_i, x) = \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2\sigma^2}\right\}$, b=0, σ=1.1 and N=89 in equation 11) for prediction of r:

$$r = \sum_{i=1}^{89} (\alpha_i - \alpha_i^*) \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2.42}\right\} \quad (17)$$

The value of $(\alpha_i - \alpha_i^*)$ has been given Fig. 5.

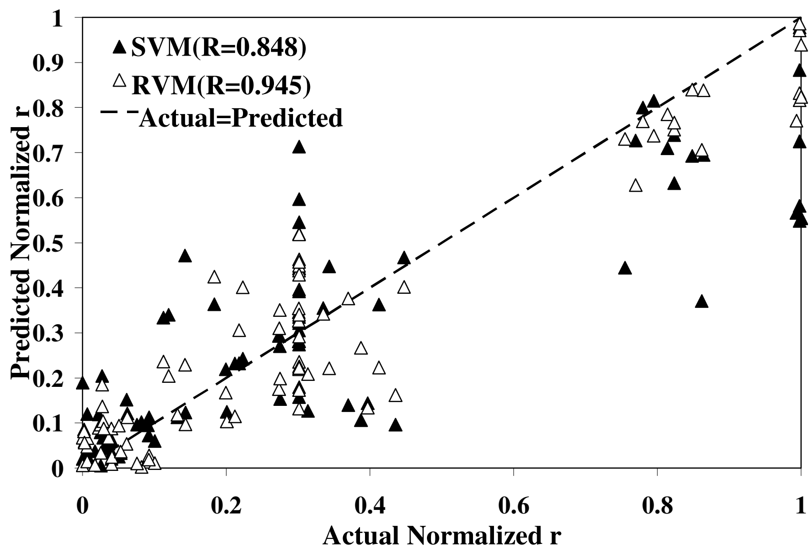

Fig. 3: Performance of training dataset.

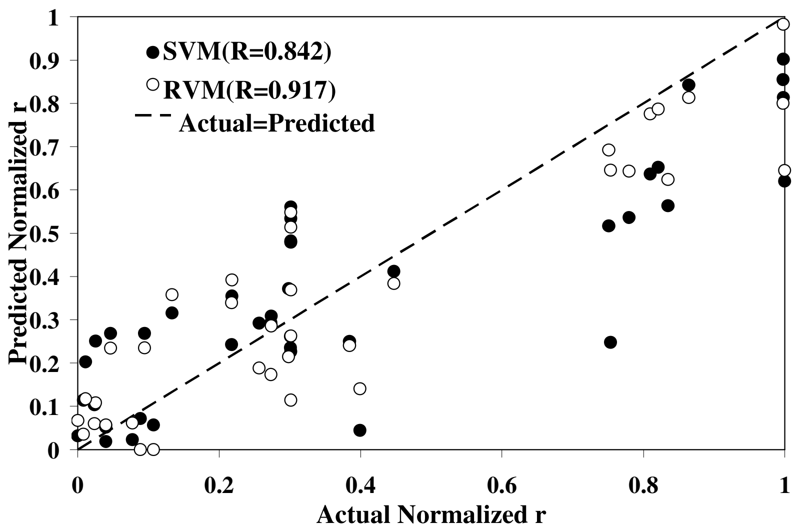

Fig. 4: Performance of testing dataset.

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

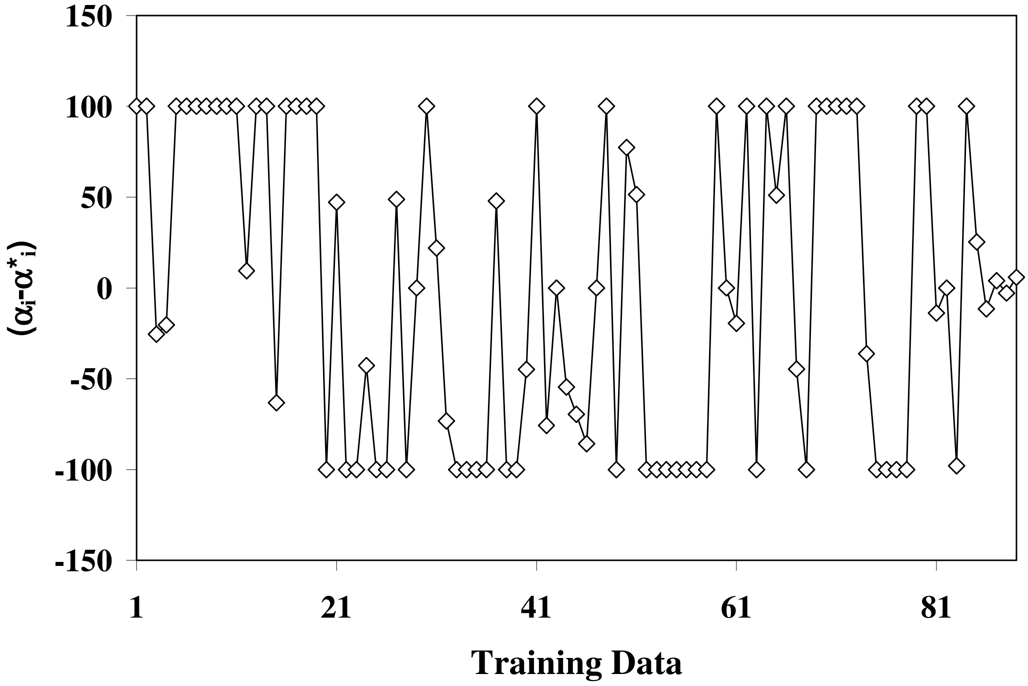

Fig. 5: Values of $(\alpha_i - \alpha_i^*)$ for SVM model.

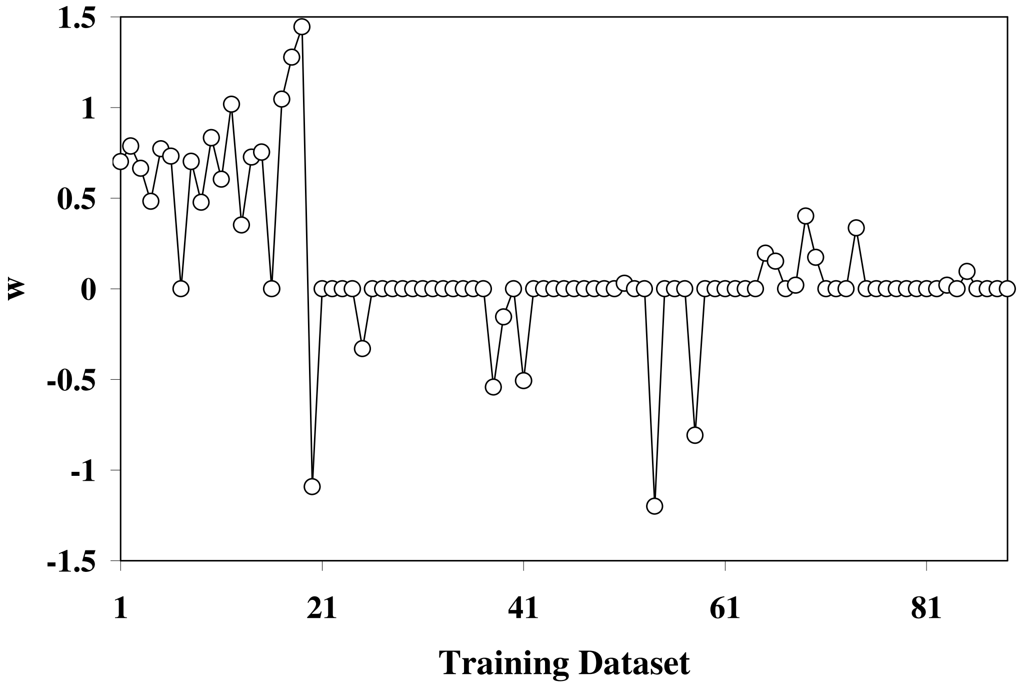

Fig. 6: Values of w for RVM model.

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

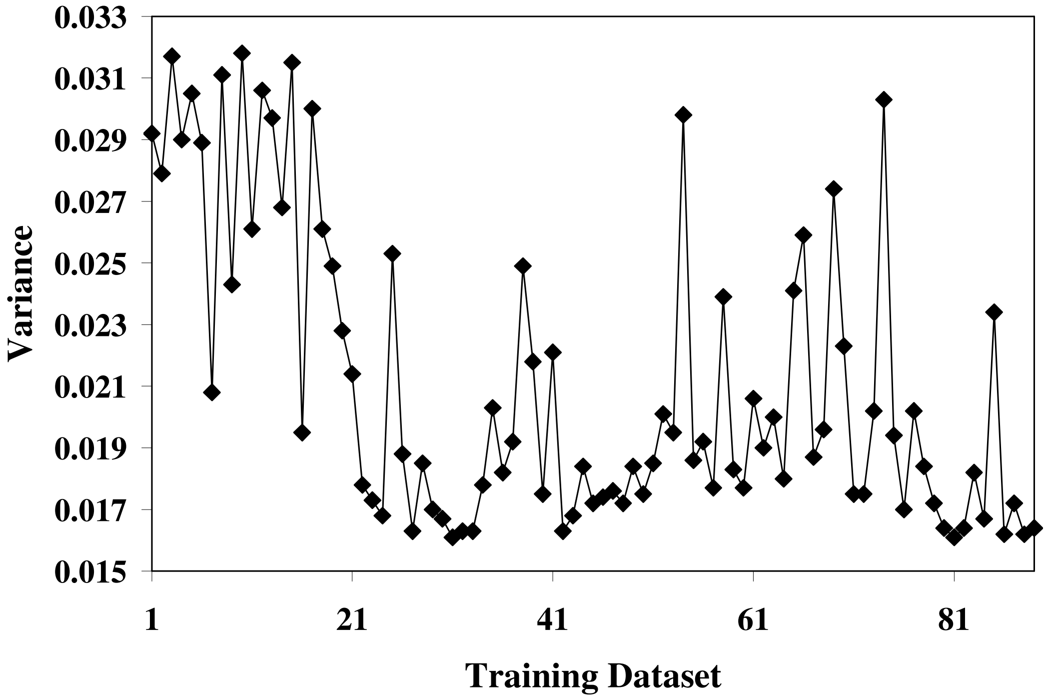

Fig. 7: Variance of training dataset.

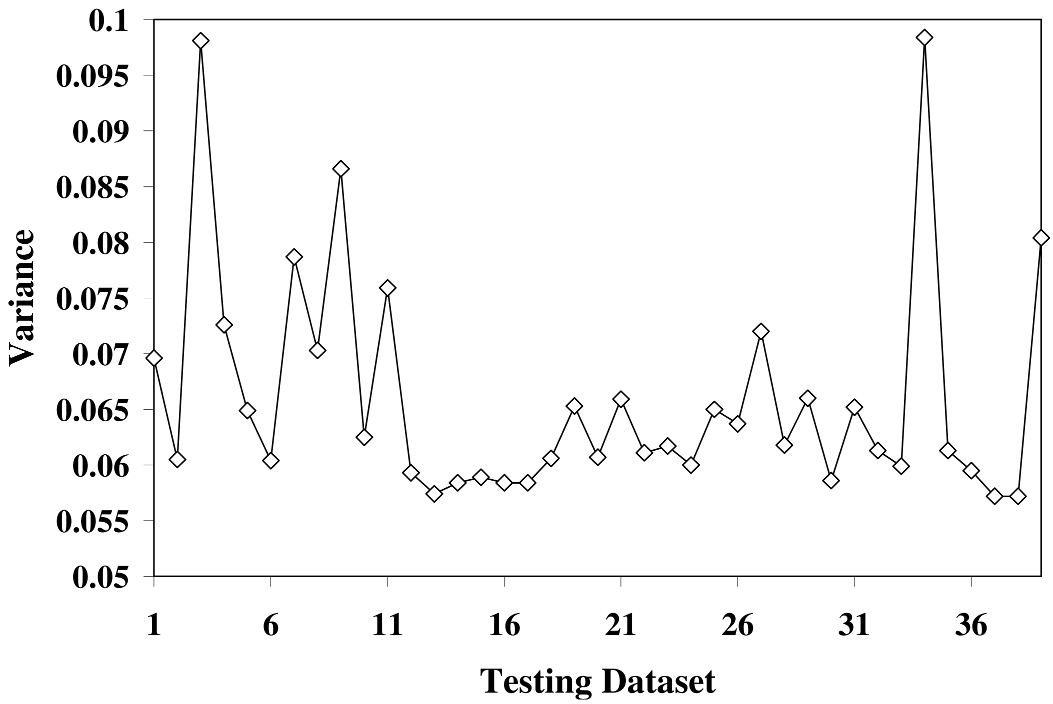

Fig. 8: Variance of testing dataset.

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

For RVM model, the design value of σ is 0.2. The developed RVM model produces 33 relevance vectors for best performance. Fig. 3 illustrates the performance of training dataset. It is observed from Fig. 3 that the value of R is close to one. Therefore, the RVM has captured relation between input and output for training dataset very well. Now, the trained RVM model has been used to determine the performance of testing dataset.

The performance of testing dataset has been depicted in Fig 4. Fig. 4 also shows that the value of R is close to one. So, the developed RVM can predict r reasonable well. The following equation has been developed for prediction of r from the RVM.

$$r = \sum_{i=1}^{89} w_i \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{0.08}\right\} \tag{18}$$

Fig. 6 shows the value of w.

The developed RVM has been also used to determine the variance of predicted r. Figs. 7 and 8 illustrate the variance of training and testing dataset respectively. The predicted variance can be used to determine prediction uncertainty.

Table- 1 shows the comparison between the SVM and RVM models. The value of R from RVM model is greater than the value of R from SVM. The performance of the RVM is better than the SVM. The developed RVM uses 33 training data as relevance vector. Whereas, the SVM uses 84 training data as support vector. So, the RVM produces more sparse solution than the SVM. The SVM uses three variables(C, ε and σ) as tuning parameter. However, RVM uses only one tuning parameter (σ).

Table-1: Comparison between the SVM and RVM models.

Models	Training Performance(R)	Testing Performance (R)	Number of training data used for final prediction	Number of tuning parameters
SVM	0.848	0.842	84	3(C,ε,σ)
RVM	0.945	0.917	33	1(σ)

Conclusion

This paper has successfully applied SVM and RVM for prediction of rainfall at Vellore. The both models show good generalization capability. However, the performance of RVM is better than the SVM. User can use the developed equation for prediction of rainfall at Vellore. The developed RVM gives variance of the predicted rainfall. The chance of overfitting in RVM is less than that of SVM due to the use of smaller number of training data. This study gives a robust model based on RVM for prediction of rainfall.

Acknowledgment: Authors would like to acknowledge VIT University management to do this project and thanks to Coordinator, ISRO-AWS, VIT University Campus.

Reference

Talei, A., Chua, L. H. C., and Quek, C. (2010) A novel application of a neuro-fuzzy computational technique in event-based rainfall–runoff modeling. Expert Systems with Applications, v. 37(12), pp. 7456–7468.

Ananthakrishnan, R. and Parthasarathy, B. (1984) Indian rainfall in relation to the sunspot cycle: 1871–1978. Inter. J. Climatology, v. 4, pp. 149–169.

Annamalai, H. (1995) Intrinsic problem in the seasonal prediction of the Indian summer rainfall. Meteorology and Atmospheric Physics, v. 55, pp.61–76.

Bathurst, J.C. and Bovolo, C.I. (2004) Development of Guidelines for Sustainable Land Management in the Agri and Cobres Target Basins, Deliverable 28 of the EU funded MEDACTION project, 37 pp. Available from: http://www.ncl.ac.uk/medaction.

Bathurst, J.C., Moretti, G., El-Hames, A., Moaven-Hashemi, A. and Burton, A. (2005) Scenario modelling of basin-scale, shallow landslide sediment yield, Valsassina, Italian Southern Alps. Natural Hazards and Earth System Sciences, v. 5, pp. 189–202.

Berger, J.O. (1985) Statistical Decision Theory and Bayesian Analysis, 2nd ed. Springer, New York.

Bhalme, H.N. and Mooley, D.A. (1981) Cyclic fluctuations in the flood area and relationship with the double (Hale) sunspots cycle. J. Applied Meteorology, v. 20, pp. 1041–1048.

Bishop, C. M. and Tipping, M. E. (2000) Variational relevance vector machine. Proceedings of the 16^{th} Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers, pp. 46-53.

Brath, A., Montanari, A. and Moretti, G. (2006) Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty). J. Hydrology, v. 324 (1–4), pp. 141–153.

Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992) A training algorithm for optimal margin classifiers. In: *D. Haussler (ed.)* 5th Annual ACM Workshop on COLT, Pittsburgh, PA, ACM Press, pp. 144-152.

Cortes, C. and Vapnik, V.N. (1995) Support vector networks. Machine Learning, v. 20, pp.273-297.

Cristianini, N. and Shawe-Taylor, J. (2000) An introduction to Support vector machine. London, Cambridge University press.

Smola, A.J. and Scholkopf, B. (2004) A tutorial on support vector regression. Statistics and Computing 14, pp. 199-222, Springer.

Dawson, R., Hall, J., Speight, L., Djordjevic, S., Savic, D. and Leandro, J. (2006) Flood risk analysis to support integrated urban drainage. Proceedings of the Fourth CIWEM Annual Conference on Emerging Environmental Issues and Future Challenges. 12–14 September 2006, Newcastle upon Tyne. Aqua Enviro.

Fowler, H.J., Kilsby, C.G., O'Connell, P.E. and Burton, A. (2005) A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change. J. Hydrology v. 308 (1–4), pp. 50–66.

Kilsby, C.G., Burton, A., Birkinshaw, S.J., Hashemi, A.M. and O'Connell, P.E. (2000) Extreme rainfall and flood frequency distribution modelling for present and future climates. Proceedings of the British Hydrological Society Seventh National Hydrology Symposium, pp. 3.51–3.56.

Krishnakumar, K. (1994) Forecasting of Indian summer monsoon rainfall on smaller regional scales using canonical correlation analysis technique. Project Report, International Research Institute for Climate Prediction. Lamont–Doherty Earth Observatory, New York, 28p.

MacKay, D.J. (1992), Bayesian methods for adaptive models, Ph.D. thesis, Dep. of Comput. and Neural System., Calif Inst. of Technol., Pasadena. Calif.

Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(IV), October, 2011, pp. 188 - 200
http://www.earthscienceindia.info/

Miloradov, M. and Marjanovic, P. (1998) Guidelines for conducting water resources assessment: a contribution to IHP-IV Project M-1-1(a).UNESCO Publishing, Paris, 120p.

Moretti, G. and Montanari, A. (2004) Estimation of the peak river flow for an ungauged mountain creek using a distributed rainfall–runoff model. In: *A. Brath, A. Montanari, E. Toth (eds.),* Hydrological Risk: Recent Advances in Peak River Flow Modelling, Prediction and Real-Time Forecasting – Assessment of the Impacts of Land-Use and Climate Changes. BIOS, Cosenza, Italy, pp. 113–128.

Navone, H.D. and Ceccatto, H.A. (1994) Predicting Indian monsoon rainfall: a neural network approach. Climate Dynamics, v. 10, pp. 305–312.

Nayak, P. C., Satyaji Rao, Y. R. and Sudheer, K. P. (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. Water Resources Management, v. 20(1), pp. 77–90.

Nicholson, S. E. and Entekhabi, D. (1986) The quasi-periodic behavior of rainfall variability in Africa and its relationship to the Southern Oscillation, J. Clim. Appl. Meteorol., v.34, pp. 331–348.

Parthasarathy, B. and Sontakke, N.A. (1988a) El-Nino/SST of Puerto Chicama and Indian summer monsoon rainfall statistical relationships. Geofisica Internacionale, v. 27, pp. 37–59.

Parthasarathy, B., Diaz, H.F. and Eischeid, J.K. (1988b) Prediction of all India summer monsoon rainfall with regional and large-scale parameters. J. Geophysical Research, v. 93 (5), pp. 5341–5350.

Philip, N.S. and Joseph, K.B. (2002) Adaptive basis function for artificial neural networks. Neurocomputing J., v. 47, pp. 21– 34 (http://www.iucaa.ernet.in/~nspp/index.html access on 24-Dec-2010).

Prasad, K.D. and Singh, S.V. (1992) Possibility of Indian monsoon rainfall prediction on reduced spatial and temporal scale. Journal of Climate, v. 5, pp. 1357–1361.

Sunyoung, L., Sungzoon, C. and Patrick, M.W. (1998) Rainfall prediction using artificial neural networks. J. Geographic Information and Decision Analysis, v. 2 (2), pp. 253–264.

Tipping, M.E. (2000) The relevance vector machine. Advances in Neural Information Processing Systems, v. 12, pp. 652-658.

Tipping, M.E. (2001) Sparse Bayesian learning and the relevance vector machine. J. Machine Learning Research, v. 1, pp. 211-244.

UNESCO, WMO (1988) Water Resources Assessment Activities: Hand book for National Evaluation. UNESCO Publishing, Paris,153 p.

Wahba, G. (1985) A comparison of GCV and GML for choosing the smoothing parameters in the generalized spline-smoothing problem. Ann. Stat., v. 4, pp. 1378-1402.