An Efficient Approach of Support Vector Machine for Runoff Forecasting

Satanand Mishra Vinayak Choubey, , S.K.Pandey, Dr. J.P.Shukla

ABSTRACT- This research presents the survey of the one of the most recent and broadly applicable approach of Data mining that is support vector machine. In this research, a hybrid cum combined model approach for runoff forecasting with the help of the technique of chaotic identification, least square support vector machine and wavelet analysis have been considered for rainfall-runoff prediction and accuracy of this new approach evaluated through the statistics of root mean square error (RMSE), mean absolute error (MAE), and correlation (R).

Index Terms-Support Vector machine, Least square support vector machine, chaotic identification, Wavelet analysis, Regression analysis, combined model.

1 Introduction

Various forecasting model and methods for prediction were developed in hydrological Science and research community. Data mining refers to extracting or mining knowledge from large amounts of data. These large and vast amounts of data may be of linear, nonlinear, time series or continuity based. Time series is an important type data objects as well as it can be easily obtained from scientific and financial applications. Time series data mining is dedicated to the development and application of novel computational techniques and patterns for the analysis of large temporal databases. Many interesting techniques of time series data mining are proposed and shown to be useful in many applications. . Support vector machine in one of the technique of machine learning along with data mining. Support vector machine is a simple and novel machine learning algorithm which is based onstatistical learning theory and later it advanced by V.N.Vapnik using theory of the VapnikChervonenkis(VC) dimension and Structural Risk Minimization. Data driven model based onstructural risk minimization principalwhich minimizes a bound on a generalized error for nonlinear data as opposed to the empirical risk minimization principal exploited by conventional Regression technique [1].Researchers are engage in finding valid outcome in the form of pattern, trend, forecast on the basis of advanced support vector machine technique.

Satanand Mishra is a research scholar of a National Institute of Technology , Durgapur and also working as a Scientist in CSIR-AMPRI, Bhopal. India. Email: snmishra07@gmail.com

Vinayak Choubey is a M Tech Student of Computer science and Engineering in a VNS Group, RGPV Bhopal. Email: vinayak.nri@gmail.com

S K. Pandey is a Asstt. Professor of a Computer Science and Engineering in a VNS Group, RGPV Bhopal. Email: shashikant3017@gmail.com

Dr. J. P. Shukla has completed his PhD from BU, Bhopal and also working as a Principal scientist in a CSIR – AMPRI, Bhopal. Email:jpshukla@yahoo.com

The Suykens and Vanderwalls [41] proposed the least square support vector machine (LSSVM) model to simplify the SVM. LSSVM and SVM have almost similar advantages but the LSSVM has an additional advantage, e.g. it needs to solve only a linear system of equation which is much easier to solve and predict results [40]. Along with LSSVM, wavelet theory has been introduced in the field of hydrology. It has been identified as a useful tool for describing both rainfall and runoff time series.

1.1 Motivation

The main purpose of this survey paper is to propose a new performance metric in the field of hydrology using above mentioned data mining technique. This survey builds on existing works significantly expanding the discussion in several directions. This survey paper can become the starting point for anyone trying to understand, evaluate, deploy or create prediction analysis.

1.2 Organization of the Paper

The organization of the paper is as follows: Section 2 presents a variety of prediction approaches. Section 3 describes the proposed approach and section 4 describes evaluation and performance analysis for application of prediction analysis using Support Vector machine. In section 5, we discuss research issues and challenges. Opportunities for future research and concluding remarks are presented in Section 6.

2 Approaches of Support Vector Machines in Hydrological Prediction

2.1 Support Vector Machines Based Approaches

Support Vector Machine (SVM) was first heard in 1992, introduced by Boser, Guyon, and Vapnik in COLT-92. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression [1]. They belong to a family of generalized linear classifiers. In another terms, Support Vector Machine (SVM) is a classification and regression

ISSN 2229-5518

prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data. Support Vector machines can be defined as systems which use hypothesis space of a linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory.

There is vast mining tools are available for data analysis, clustering, regression and pattern discovery. One of the tool is Support vector machine so a guide for users for the better understand ability of Support Vector Machines as a widely used classifier and describe the effect of the SVM parameters on the resulting classifier, how to select good values for those parameters, data normalization, factors that affect training time, and software for training SVMs. Author provides the user with a basic understanding of the theory behind SVMs and focus on their use in practice.[1]

The exploration using SVM models, which were initially developed in the Machine Learning community, has recently been shifted in hydrological prediction, with the focus on the identification of a suitable model structure and its relevant parameters for rainfall runoff modelling. This paper found that exhaustive search of an optimum model structure and its parameter space is prohibitive due to their sheer size and unknown characteristics. The paper further explored the relationships among various model structures, kernel functions, scaling factor, model parameters and composition of input vectors. [2]

To familiarize the concept of Support Vector Machines for data mining this paper or tutorial introduce support vector machine and highlight the advantages thereof over existing data analysis techniques.[5]

In its several proposed approach there is an approach to select the structure of the RBF networks based on the support vectors (SVs) of the support vector machines. In this paper, the modeling of the relationship between rainfall and river discharges of the Fuji River using the SVRBFN is presented. The main advantage of this approach is that the structure of the network can be obtained objectively, as the SVs of the SVM are obtained from the constrained optimization for a given error bound. [6]

Support vector regression is a new regression procedure in water resources for predicting suspended sediments load in rivers. The method was applied to the stream flow and suspended sediments data of two rivers in the USA. The estimated suspended sediments values were found to be in good agreement with the observed ones. The result indicates that this approach may give better performance than those described in the literature using different methodologies. This study provides the best suspended sediments estimates according to the error criteria. [7]

A novel regression approach, termed as the Rough Margin Support Vector Regression (RMSVR) network, is the approach to adopt the concept of rough sets to construct the model obtained by SVR and fine tune it with a robust learning algorithm. Simulation results of the proposed approach have shown the effectiveness of the approximated function in discriminating against outliers. [10]

The purpose of the study of the SVM, is to develop a parsimonious model used little operation gage that accurately simulates semi-arid regions by using SVM models.[11]

Recommended learning in hydrological prediction environment are, firstly, to examine the limitations in applying conventional method for evaluating the data driven forecasting model performance, and, secondly, to present the new performance evaluation methods that can be used to evaluate hydrological data-driven forecasting model with consistency and objectively. The Relative Correlation Coefficient (RCC) is used to estimate the forecasting efficiency relative to the naïve model in data driven forecasting. [12]

Support vector machine is a useful tool for knowledgeable data discovery. A novel clustering based Short Term Load Forecasting (STLF) using support vector machine (SVM). The forecasting is performed for the 48 half hourly loads of the next day. The data considered for forecasting contains 2 years of half hourly daily load and daily average temperature. The proposed architecture is implemented in Matlab. The result obtained from clustering the input patterns and without clustering is presented and the results show that the clustering based approach is more accurate. [15]

Time series data analysis is a new approach along with weighted support vector machine. Typically nonlinear features are introduced through a nonlinear kernel function. While these studies propose a new feature induction algorithm for SVMs. Specifically, new features are induced iteratively. At each step, it weights training examples differently according to the outputs of the current SVM model. The weighted training examples are then used to train a classification model that becomes a new feature for the SVM model. [16]

The accuracy of support vector machines (SVM), which are regression procedures, in modeling reference to evapotranspiration (ET0). Obtained result comparison reveals that the support vector machines could be employed successfully in modeling the ET0 process. [17]

In this paper four simple dynamic methods and two supervised learning techniques including a linear regression model, a quadratic regression model, an original grey prediction model, a back-propagation neural network model, and an epsilon-SVM regression model were investigated for the forecasting of flood stage one hour ahead for early warning of flooding hazards.[18]

Li H.P.(2010) et al introduced a modified SVM based prediction framework which improves the predictability of the inflow using climate data from the prior period. [19]

Lingras and Butz (2004) presented unique classifier techniques with support vector machine and rough set theory. This paper shows how the classification obtained from a support vector machine can be represented using interval or rough sets. Such a formulation is especially useful for soft margin classifiers. This proposed scheme is useful in practical situations when the feature space transformed by an SVM is not linear separable. If the dimensionality is increased, it may be possible to obtain a linear separable space. The paper describes how the hyper plane obtained by the soft margin classifier can be used to create a rough set based classification scheme. [22]

LiongYui-Shie and S.Chandrasekaran (2002) study reveals a novel machine learning and regression technique and application of Support Vector Machine as a robust forecasting tool has been shown with its implementation in the Bangladesh flood stage data. SVMs inherent properties give it an edge in overcoming some of the major lacunae in the application of ANN. [20]

In this paper Lin Y.J. et al (2006) presented SVM as a promising method for hydrological prediction. The SVM prediction model is tested using the long term observation of discharges of monthly river flow discharges in the Manwan Hydropower Scheme. Obtained result demonstrates SVM as a very potential candidate for the long term discharges. [23]

Mamat M and Samad A.S. compare the performance of Radial Basis Function and Support Vector Regression in time series forecasting. Both methods were trained to produce one step ahead forecasting on two chaotic time series data: Mackey Glass and Set A data. The criterions for comparison are based on the coefficient of determination (R2) and Root Mean Square Error (RMSE) between actual and forecasted output. Results show that SVR outperformed RBF significantly on the both data sets. [27]

Misra et al (2009) demonstrated to use Support Vector Machines (SVM) to simulate runoff and sediment yield from watersheds. They have simulated daily, weekly, and monthly runoff and sediment yield from an Indian watershed, with monsoon period data, using SVM, a relatively new pattern-recognition algorithm. [32]

Mountrakis et al (2010) presented a review over remote sensing implementations of support vector machines. This review is timely due to the exponentially increasing number of works published in recent years. Most of the findings show that there is empirical evidence to support the theoretical formulation and motivation behind SVMs. The most important characteristics is SVM's ability to generalize well from a limited amount and quality of training data. Compared to alternative methods such as BPN, SVM's can yield comparable accuracy using a much smaller training sample size. [33]

Solomatine P.D. (2002) described the application of the artificial neural networks, fuzzy rule-based systems, M5 model trees, support vector machine, chaos theory to the problems of water management and control in this research. [40]

In this paper Shahbazi and Pilpayeh (2012) have implemented the concept support vector machine for forecasting of river flow downstream through examining the data collected from upstream. Results have shown that SVM acts well and efficient in preserving the monthly and yearly features. [37]

Torabi M. and Hashemi S. (2012) presented an approach to forecast three-day ahead hourly electric energy consumption. To exact energy consumption patterns, Neural Network and Support Vector Machine are adapted in a novel manner. [42]

Wu et al (2004) apply support vector regression (SVR) for travel-time prediction and compare its result to other baseline travel time prediction methods using real highway traffic data. Results indicated that compared to other baseline predictors, the SVR predictor can significantly reduce both relative mean errors and root-mean-squared errors of predicted travel time. [46]

Yang et al (2009) in this issue used the concept of non-stationary and volatile data with localized support vector regression model. This model is demonstrated to provide a systematic and automatic scheme to adapt the margin locally and flexibly, while the margin in the standard SVR is fixed globally. Therefore the LSVR can tolerate noise adaptively and generate better predictive value for financial time series indices. [48]

Yi D. and Wei C. experiment shows that the fitting value which obtains using the return to first would be moreprecise than directly. Using this method to food

ISSN 2229-5518

processing forecast, its accuracy is superior to other production forecasting method. [49]

Xian and Zeng presented a two-phase, efficient, and fair evaluation method for DRMs (digital right management system) basing on SVM. Influence of three difference methods and test set number on evaluation result id discussed. [47]

In this paper Yoon H. et al developed two nonlinear time series models for predicting ground water level fluctuations using artificial neural networks (ANNs) and support vector machines (SVMs). The results of the model performance show that root mean square value of ANN models is lower than those of SVM in model training and testing stages. [49]

2.2 Least Square Support Vector Machine Approach

Classification problems have arisen in many applications, attracting many researches to develop advanced classifier techniques. A method called Least Square Support Vector Machines (LS-SVM) for pattern recognition and function estimation has been introduced in the framework of statistical learning theory. Since then there is a growing interest on this kernel method for its interesting features. In this section, a least squares version (LS-SVM) is explained. LS-SVM expresses the training in terms of solving a set of linear equations instead of quadratic programming as for the standard SVM case. Iterative training algorithm for LS-SVM based on a conjugate gradient method is then applied.

Bhagwat P. and Maity R (2013) have been explored potential of least square support vector regression model in the context of streamflow prediction using Hydroclimatic inputs. Four different Hydroclimatic variables namely rainfall, maximum temperature, minimum temperature and streamflow values of previous days are used as input variables. Parameters of LS-SVM regularization parameters and RBF kernel parameter are estimated based on the model performance during calibration period. [2]

Espinoza M. presented a primal dual formulation of Least Square Support Vector Machine. In this study it is done by using a sparse approximation of the nonlinear mapping induced by the kernel matrix, with the active selection of support vectors based on quadratic Renyi entropy criteria. The results show that the nonlinear regression in primal space improves their accuracy with larger values of dataset. [8]

Hwang H.S. (2012) et al presents a Least Square Support vector machine (LS-SVM) approach for forecasting nonlinear hydrological time series. This paper is to examine the feasibility using LS-SVM in the forecasting of nonlinear hydrological time series by comparing it with a statistical method such as Multiple Linear Regression (MLR) and a heuristic method such as a Neural Network using Back-Propagation (NNBP). In the experimental results, LS-SVM showed superior forecasting accuracies and performances to those of MLR and NNBP. [13]

Samsudin et al (2011) proposes a novel hybrid forecasting model known as GLSSVM, which combines the group method of data handling (GMDH) and the least squares support vector machine (LSSVM). The GMDH is used to determine the useful input variables which work as the time series forecasting for the LSSVM model. The performance of this model was compared with the conventional ANN models, ARIMA, GMDH and LSSVM models using the long term observations of monthly river flow discharge. The result of the comparison indicates that the new hybrid model is a useful tool and a promising new method for river flow forecasting. [35]

Shabri A. and Suhartono (2012) investigate the ability of least square support vector machine (LSSVM) model to improve the accuracy of streamflow forecasting. Cross-validation and grid-search methods are used to automatically determine the LSSVM parameters in the forecasting processes. The performance of the LSSVM model is compared with the conventional statistical models using various statistical measures. The results of the comparison indicate that the LSSVM model is a useful tool and a promising new method for streamflow forecasting. [36]

Shah Shiloh R.(2005) described the concept and characteristics of least squares support vector machines and then produced detailed result using Least Square Support Vector Regression which shows significant improvement over Support Vector Machine (SVM). [38]

Suykens et al (2001) introduced a unique method which can overcome the drawbacks of sparseness and the estimation of the support vectors values. They also introduce a sparse approximation procedure for weighted and un-weighted version of LS-SVM. The result showed how to obtain robust estimates within the LS-SVM framework in case of outliers and heavy tailed non-Gaussian error distributions. [41]

2.3 Hybrid Model Approach

A new comprehensive and more refine approach where machine learning technique and data mining technique are combined together are termed as Hybrid model approach for hydrological prediction. Classification and regression analysis have been performed in an efficient way, leads to a accurate result simulation. In this section, hybrid approach is studied through a various research papers.

In this study Botsis D. et al (2011) presented a performance comparison between Support vector regression and multilayer feed-forward neural network models with respect to their forecasting capabilities. The two models have been designed to estimate the relationship between rainfall and runoff, which describes the most complex phenomenon of hydrological science. This relationship is nonlinear and it is suggested that nonlinear models like SVR and MFNN may have notable advantages in estimating rainfall-runoff mapping. Predicted results demonstrate that SVR models have a better performance than the ANN models in simulation. However both of them have limitations to forecasting outside values of runoff and period with low or zero values. [3]

Flint and Flint (2012) proposed an evaluation on the environmental impacts of climate change on water resources and biological components of the landscape is an integral part of hydrologic and ecological investigations. The methodology, which includes a sequence of rigorous analyses and calculations, is intended to reduce the addition of uncertainty to the climate data as a result of the downscaling while providing the fine scale climate information necessary for ecological analyses. [9]

Ismail et al (2012) proposed a hybrid model based on a combination of two methods: Self Organizing Map (SOM) and Least Square Support Vector Machine (LSSVM) model, referred to as a SOM-LSSVM model for river flow forecasting. The hybrid model uses the SOM algorithm to cluster the entire dataset into several disjointed cluster, where the monthly river flows data with the similar input pattern are grouped together from a high dimensional input space onto a low dimensional output layer. Then an individual LSSVM is applied to forecast the river flow. [14]

Lin F. et al (2013) propose a hybrid model which combines locally linear embedding (LLE) algorithm and support vector machine to predict the failure of firms based on past financial performance data. By making use of the LLE algorithm to perform dimension reduction for feature extraction, is then utilized as a pre-processor to improve business failure prediction capability by SVM. [21]

M.Abbasi (2013) Study proposes a novel method of forecasting municipal solid waste (MSW) generation. Here support vector machine as an intelligent tool combined with partial least square (PLS) as a feature selection tool was used to weekly prediction of MSW generated. Model performance evaluated and compared by statistical indices of Relative Mean Errors, Root Mean Squared Errors, Mean Absolute Relative Error and coefficient of determination. The result analysis indicated that the PLS-SVM model had more robustness than SVM and had a lower sensitivity to change of input variables. [24]

Ma Xixia et al (2011) presented a combined hybrid model of chaos theory; wavelet and support vector machine was built to overcome the limitations including challenges in determination of order of nonlinear models and low prediction accuracy which the simulated accuracy is high in runoff forecasting. Firstly, runoff series were decomposed into different frequency runoff components in application of wavelet. Secondly, phase space was reconstructed in chaotic analysis. Thirdly, support vector machine was used to predict each component. Finally, all components were combined into a model to predict runoff. The results indicated that the simulated accuracy and predicted accuracy were for more accurate than parallel conventional parallel running data mining technique. [25]

Maity R. and Khasid S.S. (2009) presented an approach for monthly streamflow prediction. In this paper using the concept of hydroclimatological association rainfall-runoff relationship over a catchment is explained. Keeping various factors which influence the streamflow,large scale hydroclimatological data having taken into account along with genetic programming approach which populates the better offspring as a result and shows significant improvement in streamflow prediction. [26]

Mishra et al (2013) presented the analysis based on data mining technique in hydrological daily discharge time series of the panchratna station in the river Brahmaputra under Brahmaputra and Barak Basin Organization in India. K-means, Dynamic Time Wraping (DTW), and agglomerative hierarchical clustring are used to cluster and discover the discharge pattern in terms of the modelling. [28,29,30]

Through this paper Mohammadi K.et al (2006) tried to achieve a goal to minimize the error of specific season of the year as well as for the complete series. Goal programming was used to estimate the ARMA model parameters. [31]

The study of Pandhiani and Shabri (2013) explores the least square support vector and wavelet technique (WLSSVM) in the monthly stream flow forecasting. This is the new hybrid technique. The Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Correlation(R) statistics are used for evaluating the accuracy of the WLSSVM and WR models. [34]

Sivakumar et al (2000) by employing the correlation dimension method, provided preliminary evidence of the existence of Chaos in the monthly rainfall-runoff process at the Gota basin in Sweden. The study analyses the monthly rainfall, runoff and runoff coefficient series using the nonlinear prediction method, and the presence of chaos is investigated through an inverse approach. [39]

Wang et al (2010) presents a SVM model with chaotic genetic algorithm (CGA) as a promising method for hydrological prediction. The experimental results reveal that the SVM model with chaotic genetic algorithms results in satisfying predictions. [43]

Wang W.C. et al (2013) describes an adaptive data analysis methodology, ensembles empirical mode decomposition (EEMD), for decomposing annual rainfall

ISSN 2229-5518

series in a rainfall-runoff model based on a support vector machine (SVM). In addition, the particle swarm optimization (PSO) is used to determine free parameters of SVM. [44]

Wei L.H. and Billings A.S. (2006) proposed the long term prediction of non-linear dynamical time series, based on multiresolution wavelet models, from historically observed data sets is investigated and directed prediction approach is introduced. Prediction results based on the new direct scheme are compared with those from iterative methods and it is shown that improved predictions can be obtained using the new approach. [45]

The study by Yu X. (2004) demonstrates a combined application of chaos theory and support vector machine (SVM) in the analysis of chaotic time series with a very large sample data record the various parameters inherent in chaos technique to yield the minimum prediction error. [51]

3 PROPOSED APPROACH

Proposed approach is basically inspired from the research work of Pandhiani and Shabri [34] where they implemented the concept of Wavelet and least square support vector machine i.e. WLSSVM. In our methodology we enhanced the available concept using the method of identification of chaotic time series which ultimately helps in more accurate prediction of hydrological rainfall-runoff. By the help of below mentioned flow diagram we can get a brief view of our methodology.

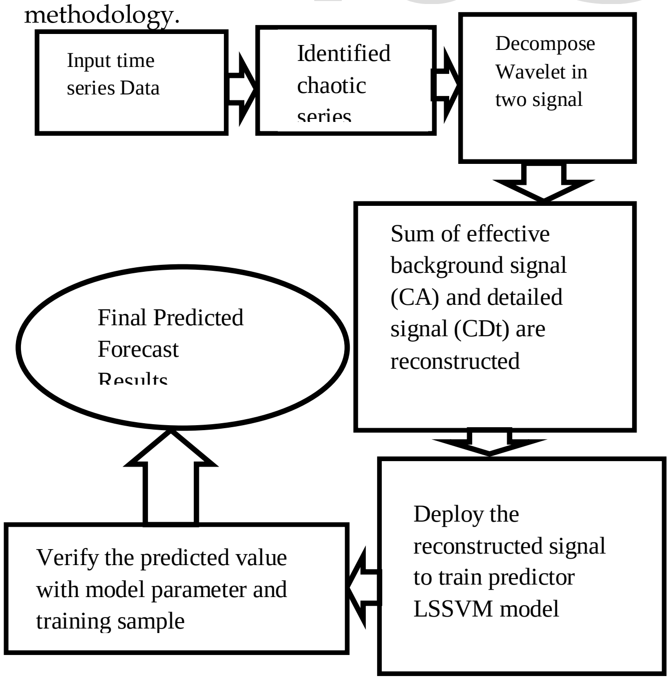

Fig.1. Flow Diagram of Proposed Approach

4 EVALUATION AND PERFORMANCE ANALYSIS

4.1 Performance Analysis

4.1.1 SVM Based Approach

A large number of techniques have been deployed on or before implementation of SVM technique of Data Mining in field of hydrological predictions. However it is unambiguously found that the performance of prediction using SVM technique is simulating more accurate prediction as compared to more general approaches of Data Mining such as ANN, Regression Analysis, Fuzzy Network, and Genetic Algorithm based approaches.

4.1.2 LS-SVM Based Approach

Most of the predictive approaches of hydrological science analyse datasets information on the basic parameters as Root Mean Square Error (RMSE), Absolute Mean Error (AME) and Correlation coefficient (R). A more general and accurate approach which comparative performed enhanced and better simulation results than SVM that is LS-SVM. In LS-SVMs a least squares cost function is used and the solution of the binary LS-SVM follows from solving a linear Karush-Kuhn-Tucker system instead of a QP problem results in more generalized prediction accuracy.

4.1.3 Hybrid Model Based Approach

A hybrid approach is the most efficient approach used for hydrological prediction and the main difference from other is that it combines two or more techniques together. In the above discussed section we have studied so many papers and generated a detailed view that prediction accuracy is increased if we use hybrid techniques. This hybrid formulation is related to a simple regression approach to Classification using complex targets. In this sense, it is also related to regularization networks and Gaussian Processes regression. The use of the bias term in the formulation and the primal-dual formulation also allow relating to the LS-SVM or SVM classifier to other techniques.

4.2 EVALUATION ANALYSIS

Evaluating a real time series based data prediction is a difficult task due to several reasons. First, it is difficult to get high quality datasets for performing evaluation due to privacy and competitive issues. Second, even if real life datasets were available, calculation requires an enormous amount of time for human expertise. Third, the constant change of real time datasets can not only introduce new type of data instances or variables but can also change aspects of normal behaviour, thus making useful predictions even more difficult. In spite of all these issues, below we talk about commonly used evaluation and criteria for predictions.

The evaluation of each approaches for both testing and training data are evaluated by using the mean-square error (MSE), mean absolute error (MAE), and correlation coefficient (R) which is widely implemented for evaluating effective results of time series forecasting. MSE, MAE and R are defined as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t^0 - y_t^f)^2$$

ISSN 2229-5518

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}(y_t^o - y_t^f)$$

$$R = \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t^o - \bar{y}_t^o)(y_t^f - \bar{y}_t^f)}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t^o - \bar{y}_t^o)^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t^f - \bar{y}_t^f\right)^2}}$$

Where y_t^o and y_t^f are the observed and forecasted values at time t , respectively and n is the number of data points. The criteria to judge for the best model are relatively small of MAE and MSE in the training and testing. Correlation coefficient measures how well the flows predicted correlate with the flows observed. R value close to the unity indicates a satisfactory result, while a low value or close to zero implies an inadequate result.

5 RESEARCH ISSUES& CHALLENGES

There are two important qualities that need to be concentrated while hydrological predictions: Correctness and performance or timeliness. Prediction approaches are very limited in the degree to which they can quantify correctness and performance or timeliness. Some important research issues in this area listed next:

Due to the voluminous size of datasets and the constant changing of data patterns as well as the presence of noise in data, it is a challenging task to build normal prediction behaviour. Further research towards finding appropriate machine learning or soft computing methods in this regard is necessary.

Based on our analysis of existing methods, Hybrid model based approach has been found to be more effective. These hybrid model based methods are highly sensitive to input parameters and their simulation prediction are often found to be scenario dependent. Therefore, development of a generic hybrid method across different scenarios is a challenging issue.

Due to lack of availability of labelled datasets for training and validation of the models, most approaches in many false alarms that require attention. Thus, minimization of false alarm is challenging issues.

Several approaches of data mining have been implemented for prediction analysis. So it is important issue that everyone must be aware about the most accurate method through which accurate predictive simulation result achieved.

6 CONCLUSIONS

In this paper, we have examined the state of modern predictive approaches. The discussion follows well known criteria for prediction based approaches: RMSE, AME, R. Experiments demonstrate that for different type’s datasets different approaches may be more successful. Research prototypes combining data mining and machine learning analysis for prediction have shown great promise. Such prediction approaches tend to have lower false alarming rates.

7 REFERENCES

[1] Ben-Hur A. and Weston J., A User’s Guide to Support Vector Machines.

[2] Bhagwat P. and Maity R., 2013, Hydroclimatic stream flow prediction using Least Square Support Vector Regression, ISH Journal of Hydraulic Engineering, vol. 19, No. 3, pp 320-328.

[3] Botsis D., Latinopulos P. and Diamantaras K., 2011, Rainfall-Runoff Modeling Using Support Vector Regression and Artificial Neural Networks, CEST2011- Rhodes, Ref No. XXX, Greece.

[4] Bray M. and Han D., 2004, Identification of Support Vector Machines for runoff modeling, journal of hydroinformatics, 06.4, IWA Publishing.

[5] Burbridge Robert and Buxton Bernard, An Introduction to Support Vector Machines for Data Mining, UCL, Gower Street, WC1E 6BT, UK.

[6] Choy Y.K. and Chan W.C., 2010, Modeling of river discharges and rainfall using radial basis function networks based on support vector regression, International Journal of Systems Science, vol.34, numbers14-15, pp763-773.

[7] Cimen M., 2008, Estimation of daily suspended sediments using support vector machines, Journal of Hydrological Sciences, vol.53, number 3.

[8] Espinoza M. et al, Load Forecasting using Fixed-Size Least Squares Support Vector Machines.

[9] Flint E.L. and Flint L.A., 2012, Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis., Ecological Processes, A Springer Open Journal. http://www.ecologicalprocesses.com/content/1/1/2.

[10] Hsiao C.C., Su F.S. and Chuang C.C. 2011, A Rough-based Robust Support Vector Regression Network for Function Approximation, IEEE International Conference on Fuzzy Systems 978-1-4244-7317-5/11.

[11] Hu H.C., Wu Z., Wang J. and Liu L. 2011, Application of the Support Vector Machine on Precipitation-Runoff Modelling in Fenhe River, IEEE, 978-1-61284-340-7/11.

[12] Hwang H S, Ham H. D. and Kim H.J. 2012, A new measure for assessing the efficiency of hydrological data-driven forecasting models, Hydrological Sciences Journal, 57(7), 1257-1274.

[13] Hwang H.S., Ham H.D. and Kim H.J., 2012, Forecasting Performance of LS-SVM for Nonlinear Hydrological Time Series, KSCE Journal of Civil Engineering (2012) 16(5):870-882.

[14] Ismail S., Shabri A. and Samsudin R., 2012, A hybrid model of self organizing maps and least square support vector machine for river flow forecasting, Hydrology and Earth System Sciences, 16, 4417-4433.

[15] Jain A. and Satish B., 2009, Clustering based Short Term Load Forecasting using Support Vector Machines, proceedings of 2009 IEEE Bucharest PowerTech, Romania. Report No: IIIT/TR/2009/212.

[16] Jin R. and Liu H., 2004, Robust feature Induction for Support Vector Machines, appearing in proceedings of the 21st International Conference on Machine Learning, Bnaff, Canada.

[17] Kisi O. and Cimen M., 2009, Evapotranspiration modeling using support vector machine, Hydrological Sciences Journal, 54(5), 918-928.

[18] Lai H.C. and Tseng H.M., 2010, Comparison of regression models, grey models, and supervised learning models for forecasting flood stage caused by typhoon events, Journal of the Chinese Institute of Engineers, 33:4, 629-634.

[19] Li H.P., Kwon H.H., Sun L., Lall Upmanu and Kao J. J., 2010 A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan, International Journal Of Climatology 30: 1256-1268.

[20] LiongYui-Shie and Chandrasekaran S., 2002, Flood Stage Forecasting With Support Vector Machines, Journal Of The American Water Resources Association, Vol.38, No.1.

[21] Lin F., Yeh C.C. and Lee Y.M., 2013, A Hybrid Business Failure Prediction Model Using Locally Linear Embedding and Support Vector Machines, Romanian Journal of Economic Forecasting.

[22] Lingras P. and Butz C., 2004, Interval set Classifiers using Support Vector Machines, IEEE, 0-7803-8376-1/04.

[23] Lin Y.J., Cheng T.C. and Chau W.K., 2006, Using support vector machines for long term discharge prediction, Hydrological Science Journal, 51(4), pp599-612.

[24] M.Abbasi, Abduli M.A., Omidvar B. and Baghvand A., 2013, Forecasting Municipal Solid waste Generation by Hybrid Support Vector Machine and Partial Least Square Model, Int. J. Environ. Res., 7(1): 27-38, ISSN: 1735-6865.

[25] Ma Xixia, Ping J., Yang L., Yan M. and Mu H., 2011, Combined model of chaos theory, wavelet and support vector machine for forecasting runoff series and its application, IEEE 978-1-61284-340-7/11.

[26] Maity R. and Khasid S.S., 2009. Hydroclimatological approach for monthly streamflow prediction using genetic programming, ISH Journal of Hydraulic Engineering, vol. 15, No. 2.

[27] Mamat M. and Samad A.S., Performance of Radial Basis Function and Support Vector Machine in Time Series Forecasting.

[28] Mohammadi K., Eslami R.H. and Kahawita R., 2006, Parameter estimation of an ARMA model for river flow forecasting using Goal programming, ELSEVIER Sciences, Journal of Hydrology 331, 293-299.

[29] Mishra S., Majumder S., and Dwivedi V.K. **,** pattern discovery in hydrological time series data mining in a Sustainable Water resources Management And Climate Change Adaptation, Vol.-II, pp.107-115, February 17-19, 2011, NIT Durgapur.

[30] Mishra S., Dwivedi V.K. , Sarvanan C. and Pathak K. K., Pattern Discovery in Hydrological Time Series Data Mining during the Monsoon Period of the High Flood Years in Brahmaputra River Basin, IJCA(0975-8887), doi-10.5120/11397-6698, Vol. 67, No.6, pp.7-14,April, 2013.

[31] Mishra S., Sarvanan C., Dwivedi V.K. , and Pathak K. K., Discovering Flood Rising Pattern in Hydrological Time Series Data Mining during the Pre Monsoon Period, Indian. J.of Zeo-Marine Science, Accepted on 12/01/2014.

[32] Misra D., Oommen T., Agarwal A., Mishra S.K. and Thompson M.A., 2009, Application and analysis of support vector machine based simulation for runoff and sediment yield, ELSEVIER Science: 103(2009), 527-535.

[33] Mountrakis G., Im J. and Ogole C., 2010, Support vector machines in remote sensing: A review, ISPRS Journal of Photogrammetry and Remote Sensing, doi:10.1016/j.isprsjprs.2010.11.001.

[34] Pandhiani M.S. and Shabri B.A., 2013. Time Series Forecasting Using Wavelet-Least Squares Support Vector Machines and Wavelet Regression models for Monthly Stream Flow Data, Open Journal of Statistics, 2013, 3, 183-194.

[35] Samsudin R., Saad P. and Shabri A., 2011, River flow time series using least squares support vector machines, Hydrology and Earth System Sciences, 15, 1835-1852, 2011.

[36] Shabri A. and Suhartono, 2012, Streamflow forecasting using least-square support vector machines, Hydrological Sciences Journal, 57 (7), 1275-1293.

[37] Shahbazi N.A. and Pilpayeh AR., 2012, River flow forecasting using support vector machines, 14th ISCCBE, Moscow, Russia.

[38] Shah Shiloh R., 2005, Least Squares Support Vector Machines.

[39] Sivakumar B., Berndtsson R., Olsson J., Jinno K. and kawamura A., 2000, Dynamics of monthly rainfall-runoff process at the Gota basin: A search of chaos, Hydrology and Earth System Sciences, 4(3), pp407-417.

[40] Solomatine P.D., 2002, Computational intelligence techniques in modeling water systems: some applications, IEEE, 0-7803-7278-6/02.

[41] Suykens K.A.J, Brabanter D.J., Lukas L. and Vandewalle J., 2001, Weighted least support vector machines: robustness

ISSN 2229-5518

and sparse approximation, ELSEVIER Science Neurocomputing: 48, 0925-2312/02, pp85-105.

[42] Torabi M. and Hashemi S., 2012, A Data Mining paradigm to Forecast Weather Sensitive Short-Term Energy Consumption, IEEE 978-1-4673-1479-4/12.

[43] Wang W., Xu D. and Qiu L., 2010, Support Vector Machine with Chaotic Genetic Algorithms for Annual Runoff Forecasting, ICNC IEEE, 978-1-4244-5961-2/10.

[44] Wang W.C., Xu M.D., Chau W.K. and Chen S., 2013, Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD, IWA publishing, Journal of Hydroinformatics, vol. 15, No. 4, pp 1377-1390.

[45] Wei L.H. and Billings A.S., 2006, Long term prediction of non-linear time series using multiresolution wavelet models, International Journal of Control, vol. 79, No. 6, pp 569-580.

[46] Wu H.C., Ho M.J. and Lee T.D. 2004, Travel-Time Prediction with Support Vector Regression, IEEE Transactions on Intelligent Transportation Systems, vol. 5, No.4, 1524-9050/04.

[47] Xian M.G. and Zeng Q.B., 2008, A novel evaluation method basing on support vector machines, IEEE 978-0-7695-3134-2/08.

[48] Yang H., Huang K., King I. and Lyu R.M., 2009, Localized support vector regression for time series prediction, ELSEVIER Sciences Neurocomputing 72(2009) 2659-2669.

[49] Yi D., Wei C. and Shengfeng L., 2011, A new Regression Method Based on SVM classification, Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 978-1-61284-181-6/11.

[50] Yoon H., Jun C.S., Hyun Y., Bae O.G. and Lee K.K., 2011, A comparative study of artificial neural network and support vector achiness for predicting groundwater levels in a costal aquifer, ELSEVIER Sciences, Journal of Hydrology No. 396, pp 128-138.

[51] Yu X., Liong Y.S. and Babovic V., 2004, EC-SVM approach for real time hydrological forecasting, Journal of Hydroinformatics, 06.3.