Applied Soft Computing 19 (2014) 372–386

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Review Article

Support vector machine applications in the field of hydrology: A review

Sujay Raghavendra. N*, Paresh Chandra Deka

Department of Applied Mechanics and Hydraulics, National Institute of Technology Karnataka, Surathkal, India

ARTICLE INFO

Article history:
Received 23 September 2013
Received in revised form 9 December 2013
Accepted 2 February 2014
Available online 4 March 2014

Keywords:
Support vector machines
Hydrological models
Statistical learning
Optimization theory

ABSTRACT

In the recent few decades there has been very significant developments in the theoretical understanding of Support vector machines (SVMs) as well as algorithmic strategies for implementing them, and applications of the approach to practical problems. SVMs introduced by Vapnik and others in the early 1990s are machine learning systems that utilize a hypothesis space of linear functions in a high dimensional feature space, trained with optimization algorithms that implements a learning bias derived from statistical learning theory. This paper reviews the state-of-the-art and focuses over a wide range of applications of SVMs in the field of hydrology. To use SVM aided hydrological models, which have increasingly extended during the last years; comprehensive knowledge about their theory and modelling approaches seems to be necessary. Furthermore, this review provides a brief synopsis of the techniques of SVMs and other emerging ones (hybrid models), which have proven useful in the analysis of the various hydrological parameters. Moreover, various examples of successful applications of SVMs for modelling different hydrological processes are also provided.

Contents

* Corresponding author. Tel.: +91 9035219180.
E-mail address: sujayraghavendran@ymail.com (S. Raghavendra. N).

1568-4946/$ – see front matter
http://dx.doi.org/10.1016/j.asoc.2014.02.002

1. Introduction

Hydrological models provide us a wide range of significant applications in the multi-disciplinary water resources planning and management activities. Hydrological models can be formulated using deterministic, probabilistic and stochastic approaches for the characterization of surface and ground water systems in conjunction with coupled systems modelling such as hydro-ecology, hydro-geology and climate. However, due to resource constraints and the confined scope of available measurement techniques, there are limitations to the availability of spatial-temporal data [1]; hence a need exists to generalize data procured from the available measurements in space and time using special techniques like support vector machines and its hybrid models.

Hydrological model applications have a wide variety of objectives, depending on the problem that needs to be investigated [1]. To analyze hydrologic data statistically, the user must know basic definitions and understand the purpose and limitations of SVM analysis. The application of SVM methods of hydrological analyses requires measurement of physical phenomena. The modeller has to evaluate the accuracy of the data collected and should have a brief knowledge on how the data are accumulated and processed before they are used in modelling activities. Some of the commonly used data in hydrologic studies include rainfall, snowmelt, stage, streamflow, temperature, evaporation, and watershed characteristics [2].

SVMs have been recently introduced relatively new statistical learning technique. Due to its strong theoretical statistical framework, SVM has proved to be much more robust in several fields, especially for noise mixed data, than the local model which utilizes traditional chaotic techniques [3]. The SVM has brought forth heavy expectations in recent few years as they have been successful when applied in classification problems, regression and forecasting; as they include aspects and techniques from machine learning, statistics, mathematical analysis and convex optimization. Apart from possessing a strong adaptability, global optimization, and a good generalization performance, the SVMs are suitable for classification of small samples of data also. Globally, the application of these techniques in the field of hydrology has come a long way since the first articles began appearing in conferences in early 2000s [4].

This paper aims at discussing the basic theory behind SVMs and existing SVM models, reviewing the recent research developments, and presenting the challenges for the future studies of the hydrological impacts of climate change.

2. Theory of support vector machines

The support vector machines (SVMs) are developed based on statistical learning theory and are derived from the structural risk minimization hypothesis to minimize both empirical risk and the confidence interval of the learning machine in order to achieve a good generalization capability. SVMs have been proven to be an extremely robust and efficient algorithm for classification [5] and regression [6,7]. Cortes and Vapnik [8] proposed the current basic SVM algorithm. The beauty of this approach is twofold: it is simple enough that scholars with sufficient knowledge can readily understand, yet it is powerful that the predictive accuracy of this approach overwhelms many other methods, such as nearest neighbours, neural networks and decision tree. The basic idea behind SVMs is to map the original data sets from the input space to a high dimensional, or even infinite-dimensional feature space so that classification problem becomes simpler in the feature space. SVMs have the potential to procreate the unknown relationship present between a set of input variables and the output of the system. The main advantage of SVM is that, it uses kernel trick to build expert knowledge about a problem so that both model complexity and prediction error are simultaneously minimized. SVM algorithms involve the application of the three following mathematical principles:

- Principle of Fermat (1638)
- Principle of Lagrange (1788)
- Principle of Kuhn–Tucker (1951)

2.1. Support vector classification

The preliminary objective of SVM classification is to establish decision boundaries in the feature space which separate data points belonging to different classes. SVM differs from the other classification methods significantly. Its intent is to create an optimal separating hyperplane between two classes to minimize the generalization error and thereby maximize the margin. If any two classes are separable from among the infinite number of linear classifiers, SVM determines that hyperplane which minimizes the generalization error (i.e., error for the unseen test patterns) and conversely if the two classes are non-separable, SVM tries to search that hyperplane which maximizes the margin and at the same time, minimizes a quantity proportional to the number of misclassification errors. Thus, the selected hyperplane will have the maximum margin between the two classes, where margin is defined as a summation of the distance between the separating hyperplane and the nearest points on either side of two classes [5].

SVM classification and thereby its predictive capability can be understood by dealing with four basic concepts: (1) the separation hyperplane, (2) the hard-margin SVM (3) the soft-margin SVM and (4) kernel function [9].

SVM models were originally developed for the classification of linearly separable classes of objects. Referring to Fig. 1. Consider

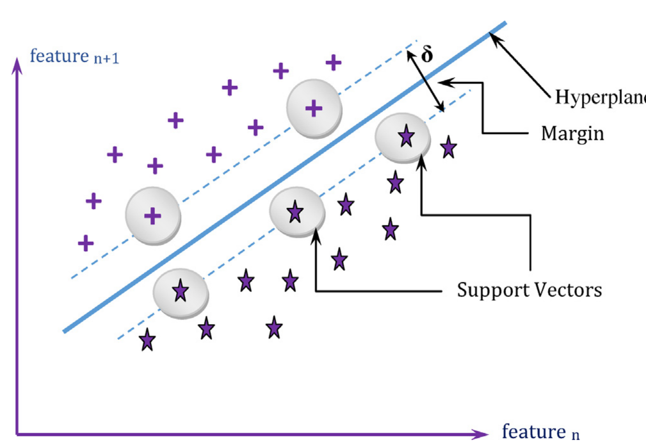

Fig. 1. Maximum separation hyperplane.
Adapted from [61].