Agricultural Water Management 97 (2010) 208–214

Contents lists available at ScienceDirect

Agricultural Water Management

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

Application of the relevance vector machine to canal flow prediction in the Sevier River Basin

John Flake [a], Todd K. Moon [a,*], Mac McKee [b], Jacob H. Gunther [a]

[a] Department of Electrical and Computer Engineering, Utah State University, 4120 Old Main Hill, Logan, UT 84322, USA
[b] Utah Water Research Laboratory, Utah State University, Logan, UT 84321, USA

ARTICLE INFO

Article history:
Received 13 June 2008
Accepted 12 September 2009
Available online 22 October 2009

Keywords:
Water demand
Irrigation canals
Flow rate modeling
Forecasting
Supervised machine learning

ABSTRACT

This work addresses management of water for irrigation in arid regions where significant delays between the time of order and the time of delivery present major difficulties. Motivated by improvements to water management that will be facilitated by an ability to predict water demand, it employs a data-driven approach to developing canal flow prediction models using the relevance vector machine (RVM), a probabilistic kernel-based learning machine. A search is performed across model attributes including input set, kernel scale parameter and model update scheme for models providing superior prediction capability using the RVM. Models are developed for two canals in the Sevier River Basin of southern Utah for prediction horizons of up to 5 days.

1. Introduction and background

One of the biggest challenges in areas with limited water is getting the necessary amounts of water to the desired places at the appropriate times, with the ever-present objective of providing the water with minimal loss in transmission and minimal excess. Meeting this challenge is problematic when, as is often the case, the amounts of water needed, the locations and times of need, and the losses that will occur are not precisely known at the time when water management and diversion decisions are made. One important area of focus, then, is the development of models for predicting water demand. Such has become a focus for research on the Sevier River Basin.

The Sevier River Basin is a closed river basin in south central Utah covering approximately 12.5% of the state's area. Due to the arid climate of the region, irrigation is essential to crop growth. Various efforts have been used to improve water management in the Sevier River Basin. A system of reservoirs and canals has been developed to manage the water needs in the basin (Berger et al., 2003). More recently, in an attempt to improve water management practices in the basin, the canal system has been heavily instrumented for measurement and control purposes (Berger et al., 2001). The instrumentation system includes measurement devices as well as communication hardware and software which collect hourly data for many points in the basin and log this data in a publicly accessible Internet database. Measurements include water levels and flow rates as well as several weather indices collected at weather stations in the basin. This automated data collection has been ongoing since the year 2000, so that there are now several years of data for many measurement points within the basin (SRWUA, 2000). These data have been used mainly for monitoring purposes, until recently, when some work has been done with statistical learning machines to predict reservoir releases (Khalil et al., 2005). Meeting with some success, this work has prompted further interest in investigations of potential improvements to water management that may come as a result of an increased ability to predict water demands in the basin.

Many challenges confront the water users in the Sevier River Basin. Depending on their location in the basin, farmers must place water orders as many as 5 days in advance of the time they can expect to receive it. A large portion of available water is lost in transmission from reservoir to field. The mechanism for water delivery is relatively inflexible: delivery times are rigid and order cancellation is generally not an option.

The work of this paper is the investigation and development of canal flow prediction capability in the Sevier River Basin with relevance vector machine (RVM) models trained using the existing database of canal measurements. The methods and tools used for prediction in the Sevier River Basin are expected to have application to other regions where water is in high demand. This work differs from the previous work (Khalil et al., 2005), since this

* Corresponding author. Tel.: +1 435 7972970; fax: +1 435 7973054.
E-mail address: Todd.Moon@usu.edu (T.K. Moon).

0378-3774/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.agwat.2009.09.010

deals with prediction of water delivered to the farmers, while the previous work deals with prediction of reservoir releases.

At present, canal operators make flow decisions based on farmer orders, transmission loss rates, canal limitations and experience. To provide for the water needs of the farmers, operators are dependent on the receipt of water orders. When setting the flow to meet current orders the canal operator has little knowledge of future orders or the future flow. More knowledge of future orders would enable improvements to water management. However, predictions must rely on data that are consistently and readily available, particularly for a system that intends to provide automated predictions at regular time steps. The primary source for consistent data in the basin is the aforementioned database where data are limited to reservoir levels, canal flow rates, and weather data (temperature, humidity, solar radiation, wind speed, and precipitation). Since other data – such as crop type and acreage – are not readily available, a physical model is problematic. Instead, a data-driven model is considered here.

We seek a functional relationship between a set of inputs and an output, where the output is the item we desire to predict. While we have spoken of water orders as the quantity we would like to predict, unfortunately, order information is not included in the database, nor are they readily available otherwise. Instead we will choose the canal flow itself as the item of prediction. This choice fills the same role as water orders and is arguably a better choice. We justify this as follows: in setting canal flow, individual farmer orders are combined additively to form a total water order. Expected water loss is accounted for with a multiplicative factor. Some modifications are likely made by the canal operator based on his strategies for respecting canal limitations, maintenance needs, and other objectives. These result in a quantity that can be thought of as the intended canal flow. The actual canal flow differs from this intended flow only by limitations in the precision of the operator at meeting his intentions. Such control limitations are a matter of the tools at the disposal of the operator for setting canal flow; they can be modeled as noise. Finally, the measured canal flow – which is the data item available in the database – is the actual canal flow with noise introduced through measurement. The measured canal flow, then, is the inclusion of control and measurement noise on a flow that is intended to meet the water orders given by the farmers. For purposes of setting canal flow we can predict this intended flow directly, which is equivalent to predicting water orders and then determining the intended flow from the orders. The direct approach eliminates computations while suiting itself to the available data.

Intended flow, which we will hereafter call demand, is directly related to the water orders placed by farmers and is generated to match those orders by taking into account the losses associated with transmission while remaining within the bounds of operation for the canal. The type of inputs that would be used to predict farmers' orders are generally the same inputs that will be effective in predicting demand.

We choose the RVM as our tool for prediction. The RVM is a learning machine with a model function formed as a linear combination of data-centered basis functions. It yields equivalent and often superior results to other kernel-based learning machines both in terms of generalization ability and model sparseness. Having chosen the RVM and given the data items available in the database, forming a model for prediction is a matter of experimenting with the choice of inputs to find the set of inputs that produce a model with the lowest prediction error.

2. Predictive function estimation and the relevance vector machine

In this section we very briefly introduce the relevance vector machine to establish some notation and concepts for following discourse. A considerably more complete discussion and derivation of the results appears in Flake (2007); the relevance vector machine appears initially in Tipping (2001).

Prediction is the deduction or estimation of a system condition based on some functional or intuitive relationship between that condition and other conditions in the system. The task of machine learning is to determine or estimate this functional relationship between the desired condition and other conditions in the system from a set of paired examples or observations of the same. In other words, if we call the value of the desired condition a *target*, and denote it t_n, and call the vector value of the system conditions that yield the target an *input*, and label it $\mathbf{x}_n$, then the task of machine learning is to estimate the functional relationship that relates inputs $\mathbf{x}_n$ to their associated targets t_n using a finite set of examples of the same, $\{t_n, \mathbf{x}_n\}_{n=1}^{N}$, hereafter referred to as the training data.

For the problem at hand, our purpose is to find good values for the model weights that will generalize to unseen data so predictions can be performed. Each target is modeled as the function on the corresponding input with additive white Gaussian noise to accommodate measurement error on the target:

$$t_n = y(\mathbf{x}_n, \mathbf{w}) + \epsilon_n. \tag{1}$$

With this formulation – given that we know $y(\mathbf{x}_n)$ – each target is independently distributed as Gaussian with mean $y(\mathbf{x}_n)$, and variance σ^2 equal to that of the noise process:

$$p(t_n|y(\mathbf{x}_n), \sigma^2) \sim \mathcal{N}(t_n|y(\mathbf{x}_n), \sigma^2). \tag{2}$$

A widely used approach to machine learning is to employ models of the form:

$$y(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{N} w_i K(\mathbf{x}, \mathbf{x}_i) + w_0. \tag{3}$$

In this formulation, the vector $\mathbf{x}$ is an input vector (consisting of potentially many elements) and the value $y(\mathbf{x}, \mathbf{w})$ is the corresponding output value. The function $K(\mathbf{x}, \mathbf{x}_i)$ is referred to as a kernel function. The kernel used in our experiments is the Gaussian kernel of form $K(\mathbf{x}, \mathbf{x}_n) = \exp\{-\eta\|\mathbf{x} - \mathbf{x}_n\|^2\}$ with scale parameter η, where (unless otherwise indicated) we take $\eta = 1$.

Roughly speaking, the vector of weights $\mathbf{w}$ can be selected by training so that across a given set of training data, the likelihood function (2) is maximized. However, this direct approach tends to overfit to the training data, so that generalization to new observations is poor. To avoid this, the training also employs a prior distribution on the weights (Flake, 2007; Tipping, 2001), specifically designed to favor small, potentially zero-valued weights, since zero-valued weights effectively remove the corresponding basis functions from the model, leaving only those basis kernel functions that are "relevant" training vectors. This gives rise to the term "relevance vector machine." Models such as these having many zero-valued weights (termed "sparse") tend to have better generalizing performance.

In our experiments below, the model (3) is compared against the linear multiple-regression (MR) model:

$$y(\mathbf{x}, \mathbf{a}) = \sum_{i=1}^{P} a_i x_i = \mathbf{a}^{\mathrm{T}}\mathbf{x}.$$

Again we model training targets as the function on the training input with additive noise. Stacking for all training vectors and constraining the error and data to be orthogonal leads to the set of normal equations:

$$\mathbf{a} = \mathbf{R}^{-1}\mathbf{p},$$

where $\mathbf{R}$ is the Grammian matrix and $\mathbf{p}$ is the correlation vector (Moon and Stirling, 2000).