Environmental Modelling & Software 47 (2013) 207–217

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

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

Recursive partitioning techniques for modeling irrigation behavior

Sanyogita Andriyas [a,b,*], Mac McKee [a,b]

[a] Department of Civil & Environmental Engineering, Utah State University, Logan, UT, USA
[b] Utah Water Research Laboratory, Department of Civil & Environmental Engineering, Utah State University, Logan, UT, USA

A R T I C L E I N F O

Article history:
Received 29 June 2012
Received in revised form
14 May 2013
Accepted 21 May 2013
Available online 21 June 2013

Keywords:
Farmer
Irrigation behavior
Recursive partitioning
Trees
Decision

A B S T R A C T

Accurate forecasts of short-term irrigation demands can provide information useful for canal operators to manage water diversions and deliveries more efficiently. This can be accomplished by analyzing the actions of the farmers who make water use decisions. Readily available data on biophysical conditions in farmers' fields and the irrigation delivery system during the growing season can be utilized to anticipate irrigation water orders in the absence of any predictive socio-economic information that could be used to provide clues into future irrigation decisions. Decision classification and the common factors, form a basis for division of farmers into groups, which can be then used to make predictions of future decisions to irrigate. In this paper, we have implemented three tree algorithms, viz. classification and regression trees (CART), random forest (RF), and conditional inference trees (Ctree), to analyze farmers' irrigation decisions. These tools were then used to forecast future decisions. During the training process, the models inferred connections between input variables and the decision output. These variables were a time series of the biophysical conditions during the days prior to irrigation. Data from the Canal B region of the Lower Sevier River Basin, near the town of Delta, Utah were used. The main crops in the region are alfalfa, barley and corn. While irrigation practices for alfalfa are dependent on the timing of cuts, for barley and corn the critical crop growth stages are often used as indicators of farmer decisions to irrigate. Though all the models performed well in forecasting farmer decisions to irrigate, the best prediction accuracies by crop type were: 99.3% for alfalfa using all the three models; 98.7% for barley, using the CART model; and 97.6% for corn, with Ctree approximately. Crop water use, which is the amount of water lost through evapotranspiration, was the prime factor across all the crops to prompt irrigation, which complies with the irrigation principles. The analyses showed that the tree algorithms used here are able to handle large as well as small data sets, they can achieve high classification accuracy, and they offer potential tools to forecast future farmer actions. This can be subsequently useful in making short-term demand forecasts.

1. Introduction

In this data-rich world, there is a lack of pertinent information about certain phenomena that are either hard to model or lack a complete physically based cause–effect description of the problem. This presents challenges in the use of conventional approaches such as deterministic models to predict future conditions. Such a problem exists in understanding and predicting a farmer's decision to irrigate. Substantial scientific theory and large quantities of data are available to analyze the irrigation problem and forecast short-term irrigation demand, but the problem of accurately anticipating short-term water demand of an individual irrigator still remains. This is due to a limited understanding of the irrigation practices that are followed by different farmers and how farmer preferences influence decisions about the timing of irrigation.

The site selected for this study is equipped with technologies to monitor reservoir releases and canal diversions, and it has dependable forecasts of evapotranspiration (ET). Some of the irrigated fields have real-time soil moisture measurements to study agricultural water use in the area. In spite of these developments, day-to-day irrigation demands are difficult to forecast. Information about such demands can be vital to help irrigation system operators achieve greater efficiency in water deliveries. In an on-demand irrigation delivery system, farmers make the basic water use decisions. Hence it is essential to consider their decision-making mechanisms in forecasting short-term irrigation demand.

Irrigation behavior has rarely been a topic of research. Each farmer has personal goals to achieve in a season, ranging anywhere

* Corresponding author. Utah Water Research Laboratory, Department of Civil & Environmental Engineering, Utah State University, Logan, UT, USA. Tel.: +1 435 797 2932; fax: +1 435 797 1185.
E-mail addresses: sandriyas@gmail.com, s.andriyas@aggiemail.usu.edu (S. Andriyas), mac.mckee@usu.edu (M. McKee).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.05.011

from profit maximization, to crop quality, to being environmentally conscious about saving water. A farmer whose primary profession is agriculture will make different choices from the one who considers agriculture as a secondary occupation. These characteristics make it more difficult to forecast behavior. The few studies that have dealt with irrigation behavior have been inconclusive in understanding the factors that contribute to decisions regarding if and when to irrigate. To find out the scope of studies done on farmer's behavior previously, we are presenting some of the notable ones in the field.

Becu et al. (2006) used a multi-agent system for a study of water sharing between two villages located at the upstream and downstream ends of a watershed. The objective was to evaluate various options to allot water to the villages and provide feasible solutions to different water users for dealing with water scarcity. The solutions were found by analyzing the impact of different land use and water management options on water deficit. Since it involved water use decisions, farmer's behavior was considered in terms of what crops are planted, when they are harvested, and how they are irrigated during the season. Farmers were grouped into various classes on the basis of different cropping patterns identified in the region. This study simulated irrigation decisions taken by the farmers on the basis of the crops they were growing. The paper primarily evaluated the use of a multi-agent system to support collective decision-making in a participatory modeling framework. The farmers initially had misunderstood the model as being the representation of the real world but the model tested scenarios and suggested possible solutions. The study showed that the players involved from the upstream village were concerned about the impact of water scarcity on both villages, while the ones downstream were only locally concerned. Bontemps and Couture (2002) developed a sequential decision model to study water use behavior under conditions when the farmers paid a negligible amount to obtain water and there was no charge for supplying it. The model required precisely calibrated crop-growth simulation models, irrigation practices and information about land use and climate for the region of interest. Data was obtained by integrating an agronomic model, and a solution-searching optimization methodology connected to an economic model. Irrigation demand functions were estimated using non-parametric methods. Three different functions were estimated keeping three types of representative weather conditions in mind. The method was applied to estimate crop demands in southwest France. Results for all types of weather conditions were found to be same (demand function curves had the same shape: decreasing and non-linear), suggesting that irrigation demand was inelastic for the small amount of water available, but if the total quantity of water was increased, the demand became more elastic. The results showed that the threshold price at which alteration of price-response seems to take place, depends on weather conditions. Le Bars et al. (2005) developed a multi-agent systems paradigm to simulate farmer-agents and their decisions over a number of years, under conditions where water supply was limited. The water manager controlled, the amount of water given to a farmer by using allocation rules that were based on the amount of water requested by farmers at the beginning of the season. The farmer-agents each owned a farm with several plots and could decide their own cropping plan. Weather variables were random. This agent-based model helped the negotiations between water suppliers, farmers, public servants and environmentalists by presenting the impacts of water allocating rules based on different criteria. In other words, rules can be tested and resulting consequences can be seen. It was found that for global corn profits, based on the information of whether the previous agent knows about the allocated water to the agent before them, the differences between farmers could be decreased. This decrease would also show a drastic effect on water use efficiency.

From the limited literature on farmer irrigation decision behavior, it is clear that few studies have been conducted to analyze decisions already made or to forecast future irrigation decision under simulated conditions. Models that could provide such forecasts could be potentially useful for improving irrigation system operations. The study reported here is a first attempt at analyzing farmers' decisions using "decision" trees. We use data about the biophysical conditions during the growing season to isolate information available to the farmer about differences on the days leading up to the time of irrigation. We also look into the possibility of using those differences to forecast farmer decision-making.

2. Theory

A wide range of machine learning techniques is available today to address modeling problems, where missing information is an issue. These show promise for the analysis of problems involving the forecasting of decision behavior under conditions where it is not possible to quantify all of the process-specific factors that affect the decision. Fig. 1 shows a tree structure. The nodes are the variables related to the process in form of root, intermediate or terminal nodes (which do not have any child nodes). As we descend in the tree the importance of the variable to the process decreases. The variable at the root node is the most important. The effect of all the variables leading to the terminal node is collective.

Trees are used to understand systems that have little *a priori* information about how and which variables are related. Classification trees have been used by Kastellec (2010) to analyze judicial decisions and laws. Random forests have been used to model ecology applications (Cutler et al., 2007). Das et al. (2009) used conditional inference trees to assess crash severity in road accidents and found the factors involved. These are some applications of trees to real problems.

Decision trees have been a powerful tool for classification and forecasting. The features and capabilities of the trees are described ahead (Hill and Lewicki, 2007). They can give insights into non-parametric, non-linear relationships between a large number of continuous and/or categorical predictor inputs, and output variables, which may be continuous or categorical. When the output variable is continuous it is a regression analysis and when categorical then a classification problem. They divide a heterogeneous group of features into small homogeneous groups with respect to the output variable. A binary tree formed by two child nodes split from each parent node is one such structure. The split is best when it separates the data into groups with two different predominant classes. "Levels" in a tree are referred to as the depth of the tree, and "size" is the number of nodes in the tree. The measure often used to estimate the split is known as "purity." The best split is defined as

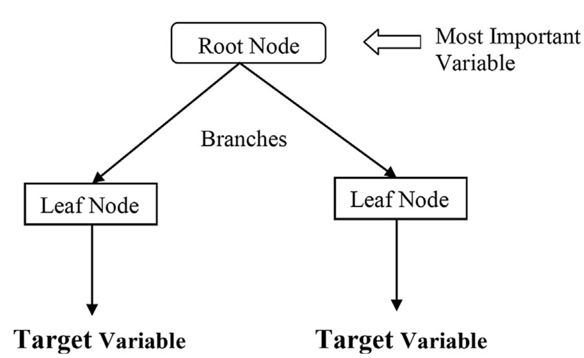

Fig. 1. A tree structure, showing root/parent nodes, branched into leaf nodes which can be intermediate nodes or terminal nodes and the variable of interest or the target variable.