Decision Trees Classification Algorithm for the Prediction of Wave Parameters

J.Mahjoobi [1], E.Mahjoobi [2]
1-Ministry of Energy, Water Research Institute
jmahjoobi@gmail.com
2- College of Civil Engineering, Sharif University of Technology
mahjoobi@civil.sharif.ir

Abstract:
Wind waves play a significant role in ocean and coastal activities. In this study, the performance of decision trees classification for prediction of wave parameters was investigated. The data set used in this study comprises of wind and wave data gathered in Lake Ontario from October to November, 2004 and further from November to December, 2005. The data set was gathered by National Data Buoy Center (NDBC) in station 45012 at 43˚ 37′ 09″N and 77˚ 24′ 18″W. The data set was divided into two groups. The first one that comprises of 26 days (611 data points of year 2005) wind and wave measurements was used to train the models. The second one that comprises of 14 days (326 data points of year 2004) wind and wave measurements was used to verify the models. Training and testing data include wind speed, wind direction, fetch length and wind duration as input variables and significant wave height (H_s) and peak spectral period (T_p) as output variables. For building classification trees, C5 algorithm was invoked. Wave heights and wave periods for whole data set were grouped into wave height bins of 0.25 m and wave period bins of 1.0 s. Then a class was assigned to each bin. For evaluation of the developed models, the index of each predicted class was compared with that of the observed data. Results indicate that as a novel method, the decision tree model using C5 algorithm is an efficient approach with an acceptable range of error for wave parameters prediction.

Keywords: Wave Prediction; Data Mining; Classification; Decision Trees; *C5* Algorithm

1. INTRODUCTION

Wave parameters are required for coastal and offshore structures design, sediment transport estimation and other coastal engineering works. In the literature, several approaches have been proposed to wave predictions classified as empirical based, soft-computing based and numerical based approaches. Different empirical methods have been developed for wave prediction such as SMB [Bretschneider, 1970], Wilson [Wilson, 1965], JONSWAP [Hasselmann et al., 1970], Donelan [Donelan, 1980, 1985], Shore Protection manual [U.S. Army, 1984] and Coastal Engineering Manual [U.S. Army, 2003]. With development of high speed processors, sophisticated numerical models such as MIKE 21 SW [DHI, 2004] and SWAN [Booij, 1999] have been developed for wave prediction. However, due to their complexity of implementation, high amount of processor time required, and the need for accurate local bathymetric surveys implementation of them is not an easy task [Browne et al. 2007]. Recently, soft computing techniques such as Artificial Neural Networks (ANNs), Fuzzy Inference System (FIS) and Adaptive-Network-based Fuzzy Inference System (ANFIS) have been used to develop wave prediction models (e. g. Deo et al. 2001, Tsai et al. 2002, Agrawal and Deo 2004, Makarynskyy et al. 2005, Kazeminezhad et al. 2005, Ozger et al. 2007). In this work, the performance of decision trees classification for wave parameters prediction was investigated. For building classification tree we selected C5.0 due to its speed, small memory requirement, boosting and cross-validation features, which greatly improve predictive accuracy. Since C5.0 can generate rules that have a straightforward interpretation, it is also quite robust in problems such as missing data and large numbers of fields.