Engineering Applications of Artificial Intelligence 43 (2015) 44–53

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Significant wave height and energy flux range forecast with machine learning classifiers

J.C. Fernández [a], S. Salcedo-Sanz [b,*], P.A. Gutiérrez [a], E. Alexandre [b], C. Hervás-Martínez [a]

[a] Department of Computer Science and Numerical Analysis, University of Córdoba, Rabanales Campus, Albert Einstein Building, 3rd Floor, 14071 Córdoba, Spain
[b] Department of Signal Processing and Communications, University of Alcalá, 28871 Alcalá de Henares, Madrid, Spain

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form
12 March 2015
Accepted 20 March 2015
Available online 29 April 2015

Keywords:
Wave energy prediction
Ordinal classification
Multi-class classification
Significant wave height
Flux of energy
Wave energy converters

ABSTRACT

In this paper, the performance of different ordinal and nominal multi-class classifiers is evaluated, in a problem of wave energy range prediction using meteorological variables from numerical models. This prediction could be used in problems of wave energy conversion in renewable and sustainable systems for energy supply. Specifically, the work is focused on *ordinal classifiers*, that have provided excellent performance in previous applications. The proposed techniques are novel with respect to alternative classification and regression techniques used up to date, the former not considering the order relation between classes in a multi-class problem and the latter needing, in general, more complex models. Another important novelty of the paper is to consider meteorological variables from numerical models as inputs of the classifiers, which has not been done before, to our knowledge, in this context. For this, a data matching is carried out between meteorological data, obtained from NCEP/NCAR Reanalysis Project in four points around the two buoys subjected to study (a buoy in the Gulf of Alaska and another one in the Southeast of United States), and the wave height or wave period collected by sensors in each buoy. Using this matching, the problem is tackled as an ordinal multi-class classification problem and the objective is to predict the range of height of the wave produced in each buoy and the range of energy flux generated. The classifiers to be compared and the model proposed are fully evaluated in both buoys. The results obtained are promising, showing an acceptable reconstruction by ordinal methods with respect to nominal ones in terms of wave height and energy flux.

1. Introduction

Marine energy is currently a hot topic in renewable and sustainable systems for energy supply (López et al., 2013; Heras-Saizarbitoria et al., 2013). It refers to a set of different technologies that exhibit a clear potential for sustainable growth, do not generate greenhouse gases, and that are potentially able to convert part of the huge energy of oceans into electricity. Some of the marine energy technologies that are currently under exploitation are off-shore wind energy, ocean thermal and tidal and wave energy conversion. Among these technologies, off-shore wind energy is currently the most exploited, but wave energy conversion is pushing hard as a new and promising energy source many countries are betting for. In fact, a number of public institutions and companies in British islands (Lawrence et al., 2013; The Offshore Renewable Energy in Scotland Website; The Pelamis Wave Power Website) are aiming at exploiting the huge wave energy resource (Arinaga and Cheung, 2012; Esteban and Leary, 2012), which, according to The Pelamis Wave Power Website, is equivalent to three times the current UK electricity demand. A review of the major issues involved in the generation of electricity from ocean energy, including wave energy conversion and its related economical aspects, can be found in Bahaj (2011).

Ocean wave energy is the energy source that converts potential and kinetic energy of waves into electricity, using devices called Wave Energy Converters (WECs) (Falcão, 2010; Lindroth and Leijon, 2011; Alamian et al., 2014). WECs transform the energy of waves into electricity by means of either the vertical oscillation of waves or the linear motion of waves, and exhibit some important advantages when compared to the other sources of ocean energy sources: (1) WECs have a much lower impact on ecosystems than tidal devices; (2) there are "ideal" areas for marine wave energy in Europe (Norway, UK, Ireland, Portugal), North coast of US and Canada, southern coast of Australia, northern coast of New Zealand, and Japan (in the sense of having great wave power density located near populated regions demanding energy); (3) wave energy is the most concentrated form of

* Corresponding author. Tel.: +34 91 885 6731; fax: +34 91 885 6699.
E-mail address: sancho.salcedo@uah.es (S. Salcedo-Sanz).

http://dx.doi.org/10.1016/j.engappai.2015.03.012
0952-1976/

renewable energy (about 1000 times denser than wind) and is less variable on an hourly basis than wind energy; (4) the wave energy resource in a given area can be predicted within hours in advance, though this prediction is difficult because of their stochastic nature and the large amount of factors which influence wave height (meteorology, sea depth, closeness to the coast, etc.). A comprehensive review of these and other beneficial properties of WEC-based technologies can be found in Falcão (2010).

Even though wave energy exhibits a number of advantages when compared to tidal energy conversion, however, waves are more difficult to characterise and predict because of their stochastic nature. Even more, wave energy flux can exhibit nonlinear variability, with irregular extreme events (Reikard, 2013). As a consequence of this complexity, wave resource prediction become a crucial topic for the design, deployment, and control of WECs (Richter et al., 2013; Fusco and Ringwood, 2010a,b), that require a proper characterisation of waves. Alternative applications of wave height prediction are decision making in operational works at sea, risk evaluation of marine energy facilities, etc. Note that the wave height characterisation can be based on either physical models or data-driven ones. Data for characterising waves can be basically obtained from radars and buoys arrays, which generate time series. Using these time series, the corresponding wave spectrum, $S(f)$, can be computed. In turn, based on its spectral moments, a number of wave parameters, which are particularly useful to estimate the *wave power density* at a given location, can be defined. The most important wave parameters in this regard are the *significant wave height* (H_s) and the *wave energy period* (T_e). Note that the wave's *flux of energy* (F_e), which will be used by the WEC in order to generate electricity, can be obtained from H_s and T_e as $F_e = 0.49 \cdot H_s^2 \cdot T_e$. As mentioned, waves stochastic nature makes very difficult the prediction of wave energy resource, so the research work on this topic has been intense in the last few years. Focusing on machine learning approaches (Bishop, 2006) to wave energy prediction, most of them use artificial neural networks. One of the first approaches to predict H_s is due to Deo and Naidu (1998), who proposed the use of artificial neural networks in this problem. Improvements on this proposed system were presented in a posterior work (Agrawal and Deo, 2004). In Tsai et al. (2002), H_s and T_e are predicted from the observed wave records using time series neural networks. In Zanaganeh et al. (2009) (an improvement of Kazeminezhad et al., 2005), a hybrid genetic algorithm-adaptive network-based fuzzy inference system model was developed to forecast H_s and the peak spectral period in Lake Michigan from wind speed, fetch length and wind duration. In this methodology, both clustering and rule base parameters are simultaneously optimised using genetic algorithms and artificial neural networks. Recently, in Castro et al. (2015), a neural network is applied to estimate the wave energy resource in the northern coast of Spain. There have been other works that apply more efficient regression methods, based on machine learning and soft computing (Mahjoobi et al., 2008), such as support vector regression (Mahjoobi and Mosabbeb, 2009), genetic programming (Nitsure et al., 2012) or fuzzy logic (Özger, 2011). Alternative approaches using numerical models of atmosphere and ocean, usually hybridised with time series prediction, can also be found in the literature (Reikard et al., 2011; Akpinar and Kömürcü, 2013).

This paper deals with a problem of marine wave energy prediction (specifically H_s and F_e prediction), including different novelties in its resolution. Firstly, it is proposed the application of machine learning classifier algorithms for this problem. This approach is completely novel, since the previous works had tackled the problem, to the best of our knowledge, by applying regression techniques. Secondly, the problem is tackled as an ordinal multi-class classification problem, from the point of view that any real variable can be discretised in different categories, reducing the amount of information involved and simplifying this way the problem: a predefined number of ordered categories or ranges in terms of H_s or F_e is enough for obtaining practical information. The categories obtained through discretisation of a real value exhibit an order, and this order can be taken into account both in classifier construction and evaluation, choosing appropriate performance measures for that (Verwaeren et al., 2012; Cruz-Ramírez et al., 2014; Baccianella et al., 2009). Ordinal techniques studied take this consideration, achieving an error in the predictor that usually improves that of nominal classifiers. Modelling this prediction problem as a multi-class classification challenge allows the application of novel and powerful approaches existing in the literature, that have not been discussed before in a problem of marine wave energy prediction. This kind of multi-class classification problems are sometimes referred to as *ordinal regression* (Chu and Keerthi, 2007; Sun et al., 2010; Gutiérrez et al., 2012) and this work will compare the performance of standard multi-class classifiers and ordinal regression ones in this problem of H_s and F_e prediction. Another novelty of the problem is that the different classification techniques discussed in this paper are applied by using meteorological predictive variables from global numerical models. Specifically, meteorological data from The NCEP/NCAR Reanalysis Project are used in four different grid points around each buoy subjected to study in order to conform the independent variables for H_s and F_e prediction. For this purpose, a matching procedure is carried out every 6 h between the meteorological variables obtained from NCEP/NCAR Reanalysis Project and between H_s and F_e hourly collected by each buoy, which are obtained from the National Data Buoy Center (NDBC) (National Oceanic and Atmospheric Administration), belonging to the National Data Buoy Service of the United States. The performance of the different classifiers discussed in this problem has been evaluated in the prediction of H_s and F_e from real data of a buoy located in the Gulf of Alaska and another one located in the Southeast of United States.

The rest of this paper is structured as follows: Firstly, the next section briefly presents the formulation for an ordinal classification problem. Secondly, the more important ordinal classification methods in the literature for tackling this problem are described. In Section 4, the datasets used in this paper, including the buoys and reanalysis data are detailed. The specific problem modelling considered, which includes the estimation/prediction of H_s and F_e in a 6 h time-horizon, are also stated in this section. Section 5 presents the experimental part of the paper and the results. Finally, Section 6 gives some concluding remarks for closing the paper.

2. Ordinal classification

The classification of items into naturally ordered classes is commonly found in many supervised learning problems (Gutiérrez et al., 2013; Sánchez-Monedero et al., 2014), which are referred to as ordinal classification (also known ordinal regression) problems, where an ordinal scale (Lippmann, 1989) is used to label the patterns – for instance, a teacher who rates his/her students using labels (A, B, C, and D) that have a natural order among them ($A \succ B \succ C \succ D$). However, they are usually tackled by using methods intended for the classification of nominal classes, where the order relation is not considered.

Taking into account the nature of the problem studied in this paper, different class labels can be defined based on different ranges of H_s and F_e, in such way that an ordered relationship exists (the smallest to the largest, in increasing order) between the labels ($\mathcal{C}_1 \prec \mathcal{C}_2 \prec \cdots \prec \mathcal{C}_Q$, being Q the number of classes or labels). A class with lower order provides less energy than a class with higher order. This ordered rank implies an additional restriction for the classification problem. As a result, the nominal multi-class classification problem turns into an ordinal classification one.

When the ordinal nature of the target variable is not obvious or has been defined a posteriori, nominal classifiers can also be applied to ordinal problems (Bender and Grouven, 2006) and they can even yield