Engineering Applications of Artificial Intelligence 34 (2014) 199–207

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal–ordinal support vector classifier

J. Sánchez-Monedero [a], S. Salcedo-Sanz [b,*], P.A. Gutiérrez [a], C. Casanova-Mateo [c], C. Hervás-Martínez [a]

[a] Department of Computer Science and Numerical Analysis, Universidad de Córdoba, Córdoba, Spain
[b] Department of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain
[c] Department of Applied Physics, Universidad de Valladolid, Valladolid, Spain

ARTICLE INFO

Article history:
Received 19 December 2013
Received in revised form
21 May 2014
Accepted 31 May 2014
Available online 2 July 2014

Keywords:
Rainfall occurrence
Rainfall amount
Nominal and ordinal classifiers
Ordinal regression

ABSTRACT

In this paper we propose a novel computational system for simultaneous modelling and prediction of rainfall occurrence and amount. The proposed system is based on a hierarchical system of nominal–ordinal support vector classifiers, the former focussed on the prediction of the rainfall occurrence, and the latter centered in the expected rainfall amount from a set of three different ordinal classes. In addition to the proposed model, we use a novel set of predictive meteorological variables, which improve the classifiers performance in this problem. We evaluate the proposed system in a real problem of rainfall forecast at Santiago de Compostela airport, Spain, showing that the system is able to obtain an accurate prediction of occurrence and rainfall amount, and we discuss the usefulness of the proposed system as part of the airport weather forecast and warning system, in order to improve airport operations.

1. Introduction

Rainfall modelling is a very important problem that arises in many applications such as agriculture (Wei et al., 2005), water resources management (Yu et al., 2006; Kisi and Cimen, 2011; Wu et al., 2009; Kannan and Ghosh, 2013; Salvi et al., 2013), hydrology (Chen and Chau, 2006; Taormina et al., 2012; Chau et al., 2005; Cheng et al., 2005) or facilities maintenance and control (Benedetto, 2002), among others (Tapiador et al., 2012; Muttil and Chau, 2006). Currently, numerical weather prediction models have improved their performance, but they are still unable to provide accurate models for expected precipitation amount at high spatial and time resolutions. Different previous works have applied soft-computing approaches to overcome this difficulty, mainly based on neural networks and related approaches. These approaches have several advantages over global numerical models: they are much more easy and fast to train, they can be applied to data from a specific point of measurement, and their performance is really competitive compared to global techniques.

Neural computation models for precipitation prediction started to be applied about 20 years ago (Kuligowski and Barros, 1997; Navone and Ceccatto, 1994; Hall et al., 1999; Luk and Ball, 2001). Some of these first works applied multi-layer perceptrons to a set of predictive variables, carefully chosen to be related to rainfall, and with data from precipitation gauges (pluviometers) to obtain rainfall quantity (Hall et al., 1999; Kuligowski and Barros, 1998; Hung et al., 2009). The majority of these approaches considered short-term precipitation prediction, from 6 h to 24 h time-horizons, obtaining good results in the prediction (Yuval and Hsieh, 2003). There are other approaches focused on long-term rainfall prediction and precipitation trends in a given zone, such as Philip and Joseph (2003), where the rainfall trend in the southern part of Indian Peninsula is analyzed by using an Adaptive Basis Function Neural Network with a back-propagation training algorithm. In Silverman and Dracup (2000) a multi-layer perceptron is applied to a problem of long-term precipitation prediction in California. More recently, in Nastos et al. (2013) an artificial neural network has been applied to model and forecast precipitation in Athens, Greece. In Shukla et al. (2011) a neural network was applied to forecast precipitation during the summer Monsoon station in India, using El Niño South Oscillation (ENSO) indices. In Manzato et al. (2007) a neural network approach is applied to the short-term forecasting of thunderstorms rainfall.

Alternative classification and regression techniques have also been applied to problem of rainfall prediction and modelling. In Ingsrisawang et al. (2008) a comparison of machine learning algorithms (decision trees (DT), neural networks (ANN) and Support

* Correspondence to: Department of Signal Processing and Communications, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain. 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.2014.05.016
0952-1976/

Vector Machines (SVMs)) has been carried out for a short-term precipitation prediction problem in Thailand. In Wu et al. (2010) a hybrid SVM for regression with particle swarm optimization was applied to a problem of rainfall prediction. In Lu and Wang (2011) a SVM approach with different kernel functions is presented to predict monthly rainfall in a region of China. In Kisi and Cimen (2012a) a novel wavelet-SVM approach was applied to precipitation forecasting from past data. SVMs have also been recently applied to precipitation related studies, such as precipitation downscaling (Tomassetti et al., 2009; Chen et al., 2010) or streamflow prediction (Kisi and Cimen, 2012b). One of the main advantages of machine learning classifiers is that they do not consider strong assumptions about the distribution of the training data, as opposite to other traditional statistical methods.

In spite of this huge work on rainfall prediction, there are not many papers focussed on the modelling and forecast of precipitation occurrence and amount together. There are two main articles dealing with this problem. In Harpham and Wilby (2005) several types of neural network models are applied to solve a problem of rainfall occurrence and amount modelling in northwest and southeast of England. The input data of this study are different measurement stations and also some large-scale climate predictors such as atmospheric circulation, thickness or moisture content at the surface, 850 and 500 hPa. More recently, in Hasan and Dunn (2010) a simple model for modelling rainfall occurrence and amount simultaneously has been proposed. It is based on a tweedy generalized linear modelling and the authors show that it performs well in modelling both occurrence and precipitation amount in Australia. Data from over 200 measurement stations spread all over Australia are used as inputs to the model. The use of joint models for simultaneous modelling of rainfall occurrence and amount is a hot topic in hydrology, since it provides information that can then be used in agriculture production systems and other applications.

In this paper we propose a novel system for simultaneous prediction of rainfall occurrence and amount, based on a hierarchical classifier, composed of a nominal and ordinal SVM classifier. First, a nominal SVM is used to set the rainfall occurrence model, that is, no-rain or rain prediction in the following 6 h. A second ordinal SVM is then hybridized with the previous nominal classifier, in order to obtain the expected rainfall amount (for those cases in which occurrence is predicted by the first model) from a set of three different ordinal classes. The proposal includes a cost sensitive SVM model to alleviate the imbalanced nature of the problem and we simultaneously optimize the hyper-parameters associated with both SVM models. In addition to the proposed model, we use a novel set of predictive variables, which improve the classifiers performance in this problem. First, we consider significant meteorological variables from atmospheric soundings. We also include as predictive variable the synoptic configuration of the atmosphere (synoptic situation using Hess–Brezowsky classification), that, to our knowledge, has not been either considered in precipitation prediction studies with machine learning techniques, in spite of its significance to establish precipitation regimes in mid-latitude regions (Trigo and DaCamara, 2000). We also evaluate the importance of other predictive variables such as humidity and Equivalent Potential Temperature (both measured in vertical soundings), and of groups of these variables in the proposed hierarchical SVM performance. Regarding the objective variables, real rainfall data from a measurement station at Santiago de Compostela (Airport), Spain, are considered to establish the performance of the proposed system.

The rest of this paper is structured as follows: next section presents a review of the main predictive variables and precipitation data used in the study. We also state the exact modelling carried out, which includes the estimation/forecasting of rainfall occurrence and amount in the next 6 h. Section 4 presents the proposed nominal and ordinal hierarchical SVM for rainfall modelling. Section 5 presents the experimental part of the paper. Finally, we give some concluding remarks for closing the paper in Section 6.

2. Predictive and objective variables used

Rainfall requires the existence of adequate clouds to produce precipitation. Therefore in order for precipitation to occur, three basic factors should be combined in an adequate way: condensation nuclei, enough water vapor (moist) and vertical movements (updrafts and downdrafts as well as the atmospheric stability). As a consequence, data selection should cover all these three elements so as to obtain a robust group of predictive meteorological variables related to the physical processes involved in the production of precipitation. Fortunately, an adequate number of condensation nuclei (such as smoke from industrial, particles of salt) on which water vapor undergoes condensation to form water droplets or deposition to form ice crystals are almost always present in the atmosphere. Then, it is only necessary to select meteorological variables related to the presence of enough water vapor and vertical movements.

As has been shown in some studies (Hall et al., 1999; Kuligowski and Barros, 1998), it is difficult to determine the criteria that should be followed to select the best set of meteorological variables to use in machine learning classifiers, based solely on an understanding of the physical mechanism of precipitation. Moreover, because precipitation is highly dependent on small-scale processes and local geography (Applequist et al., 2002) a standardized pool of meteorological variables to forecast precipitation would be difficult to set. Nevertheless, considering the satisfactory results obtained in Hall et al. (1999) and Kuligowski and Barros (1998) using neural network approaches with backpropagation training algorithm, it is possible to choose a reasonable group of meteorological variables following similar criteria.

In our study we combine observed variables, taken from an upper air sounding station, and meteorological variables derived from a numerical weather prediction model, plus the observed precipitation.

As mentioned before, observed precipitation (target variable) data was obtained from Santiago de Compostela Airport ground automatic station (latitude: 42.89; longitude: −8.41; altitude: 370 m). We chose this target area because Santiago de Compostela is located in one of the rainiest areas of the Iberian Peninsula, without a dry season and with an average annual precipitation of 1886 mm (AEMET, 2004). This station is part of the State Meteorological Agency of Spain (AEMET) surface observing network and reports all meteorological data every 10 min (it calculates the average value for each meteorological variable every 10 min). Although the data are available on a ten-minute basis, we consider the rainfall prediction in a time horizon of 6 h. Thus, the precipitation data's temporal resolution selected for this study is 6 h. The meteorological data and variables employed for this study span the dates from 1 September 2009 to 31 August 2010, i.e., this study covers the 2009–2010 hydrological year.

We have used different predictive variables in order to predict precipitation occurrence and amount. First of all, data from La Coruña (latitude: 43.36; longitude: −8.41; altitude: 67 m) radiosonde station is considered, which is the nearest upper air station to our study area. This station belongs to AEMET and its data are freely available on the Internet http://weather.uwyo.edu/upperair/sounding.html. The second set is formed by the data from the medium-range global prediction model GFS (Global Forecast System) maintained by the National Center for Environmental