



Usefull(ness) of NAO index for forecast of Monthly rainfall in Lisbon

António R. Tomé(1,2) and Pedro de Almeida(1)

(1)Universidade da Beira Interior, (2)Centro de Geofísica da Universidade de Lisboa,
artome@ubi.pt

In this work we apply an approach based on data mining techniques for predicting the next month rainfall in Lisbon. Our approach departs from a number of historical time series that are expected to contain useful information to solve the desired prediction task. Among those time series, we included NAO index and derived ENSO indices. These time series are then pre-processed. Using domain knowledge as a guide, a number of derivative variables based on different combinations of the original data are composed. Then the derivative variables are discretized and automatically filtered, using a number of variable evaluation measures, and are progressively developed in successive iterations of that composition/filtering cycle. After the final selection of the derivative variables, a pattern detection and prediction phase is performed. This phase uses two very simple machine learning algorithms. The first is a nearest neighbours "lazy" algorithm. The second is the classical Naïve Bayes probability based algorithm. This pattern detection/prediction phase is a data mining "classification" task, since its output is a prediction of precipitation in terms of 3 classes: for each month, we predict "unusually heavy rain conditions", "typical rain conditions" or "unusually dry weather", and also we perform prediction on 10 equally distributed classes.

Using this system, we are achieving prediction accuracies (in the described 3 classes layout) above 80% in winter months, and slightly less than that in the "drier" months. One were also able to conclude that in spite of the high correlation of NAO index with the monthly Lisbon rainfall it has a value almost null as a valid predictor, in spite of being used as such in several rainfall models of monthly Lisbon rainfall forecast.