Development of Intelligent Systems with Predictive Capability for the Elaboration of Environmental Data

L. Trizio, P. Bruno, M. Caselli, G. de Gennaro

Department of Chemistry, University of Bari, Via E. Orabona 4, 70125 Bari, Italy

Abstract

The problem of air pollution is a frequently recurring situation and its management has social and economic considerable effects.

Air pollution is caused by industrial emissions and vehicular traffic. Moreover, the reduction of the produced pollution implies high costs due to depuration plants or to the temporary interruption of the productive activities.

Given the interaction of the numerous factors involved in the raising of the atmospheric pollution rates, it should be considered that the relation between the intensity of emission produced by the polluting source and the resulting pollution isn t immediate.

So it can be important to develop mathematics models able to perform forecasts about the pollutants concentration to inform the population about the pollution in particular sites or to take restrictive measures as, for example, the traffic block [1, 2].

The aim of this study is to realize a support decision system with the neural networks, that, correlating the air quality data with the meteorological information, is able to predict the critical pollution events to help the Public Administration about the interventions to perform on the territory.

Introduction

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information [3]. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process [4, 5].

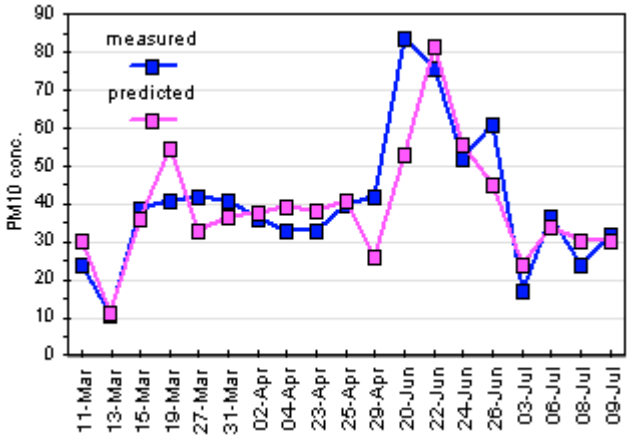

Fig. 1; Application of the neural network on a test set

Learning in biological systems involves adjustments to the synaptic connections that exist between the neurones. This is true of ANNs as well.

Materials & Methods

In this paper a feed forward back propagation neural network is used. The network used is constituted by 15 input neurones (that are the data of PM10 of the previous days (from 12 a.m. to 12 a.m.) respect to the foreseen day and data of whether forecast), by 7 neurones in the Hidden layer and by one output neuron (PM10 daily mean).

About the prediction of atmospheric particulate matter, a principal source of data is that of measures performed by monitoring stations displaced in critical sites of the city.

The weather forecast (temperature, wind speed, pressure, relative humidity) available on the network (www.wunderground.com) are used as input meteorological data.

Results

The learning of the net has been performed using the PM10 data of a network of Municipality of Bari in the year 2005.

The transfer function used is "logsig", while for the weights updating the function used is "trainrp" (Resilient Backpropagation) that limits the problems resulting from the size order of the gradient components (that tend to humble themselves when the neurons are next to saturation) using only the information about the individual derivative sign to determine the weight direction.

The performance of the network has been optimized changing various parameters (as the numbers of layers, the learning rate and others parameters typical of the network).

The training set is constituted by 200 data of daily PM10 from January 2005 to February 2006 in San Nicola monitoring network. The training error, calculated as the difference mean between the real and the predicted value is 0.00023. After the training, the net has been submitted to a test set of 19 data.

The MSE (mean squared error) calculated as the mean difference between the real and the foreseen value is 10 percent.

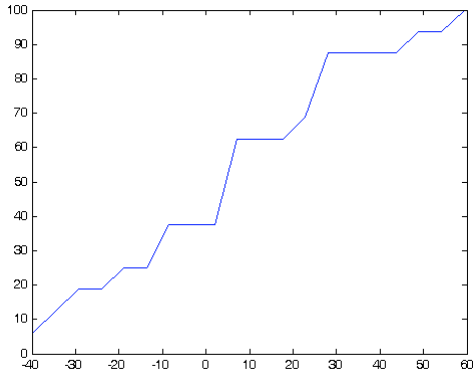

Fig. 2; Integral plot of percentage errors vs percent of data

In figure 1 the comparison between the PM10 real values and the PM10 forecasted values is shown. In the days 20 and 22 June, in which there were overcomings due to Saharan dust, the network responded in a correct way predicting high values of PM10.

As shown in figure 2, the 80% of data has an error ranging from - 30% to +30%.

Conclusions

The networks gives good results. Moreover daily the network is tested to predict the levels of concentration of PM10 in the Municipality of Bari.

The next phase will regard:

- the hourly prediction of the PM10 using the hourly meteorological data forecasted.
- The prediction of PM10 long - term

Infact, in this way the network could be a valid tool for the Public Administration to take measure about possible traffic blocks or others precautions.

Moreover, to demonstrate the advantages of the neural network on others models, we ll compare the results obtained with these obtained by others models as PCR or PLS.

References

1) M.W. Gardner, S.R. Dorling, Artificial Neural Networks (The Multilayer Perceptron) - A Review of Applications in the Atmospheric Sciences, Atmos. Environ. 32 (14/15), (1998) 2627-2636

2) G. Nunnari, A.F.M. Nucifora, C. Randieri, The application of neural techniques to the modelling of time-series of atmospheric pollution data, *Ecol. Model.* 111 (1998) 187–205

3) R. Hecht-Nielsen, Theory of the backpropagation neural network, Proc. IEEE-IJCNN89 at Washinghton DC, IEEE Press, New York, (1998) vol.1, pp. 543-611

4) P. PeHrez, A. Trier, J. Reyes, Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile, *Atmos. Environ.* 34 (2000) 1189-1196

5) M. Kolehmainen, H. Martikainen, J. Ruuskanen, Neural networks and periodic components used in air quality forecasting, *Atmos. Environ.* 35 (2001) 815-825