Chapter 20
Applications of Self-Organizing Maps to Address Environmental Studies

M.P. Gómez-Carracedo
University of A Coruña, Spain

D. Ballabio
University of Milano-Bicocca, Italy

J.M. Andrade
University of A Coruña, Spain

R. Fernández-Varela
University of A Coruña, Spain

V. Consonni
University of Milano-Bicocca, Italy

ABSTRACT

An overview on the basic principles of Kohonen Self-Organizing Maps (SOMs), Counter-Propagation Artificial Neural Networks (i.e., a SOM with a supervised layer, called the Grossberg layer) and the so-called MOLecular Map of Atom-level Properties (MOLMAP) and their usefulness in analytical chemistry, particularly in the environmental field, was presented. Two case studies dealing with environmental studies of soil pollution by road traffic and sea pollution by spilled hydrocarbons were discussed in detail to exemplify different benefits derived from the use of these techniques.

INTRODUCTION

Artificial Neural Networks (ANNs) are increasing in uses related to several working areas within analytical chemistry and can be considered already as one of the most important tools in multivariate analysis. For this reason, several different ANN architectures and learning strategies have been proposed in literature so far (Zupan, 1994; Zupan and Gasteiger, 1993).

Kohonen Maps (or Self-Organizing Maps, SOMs, Kohonen, 1988) are one of the most

DOI: 10.4018/978-1-61520-893-7.ch020

popular Artificial Neural Networks nowadays. They mimic the action of a biological network of neurons, where each neuron accepts different signals from neighbouring neurons and processes them. Kohonen Maps are self-organising systems which are capable of solving unsupervised problems, such as clustering, exploratory analysis of data structure, and selection of relevant samples. Counter-Propagation Artificial Neural Networks (CP-ANNs) are essentially based on the Kohonen method, but combine characteristics from both supervised and unsupervised learning (Hecht-Nielsen, 1987; Zupan et al., 1995; Zupan et al., 1997). CP-ANNs are a development of Kohonen Maps, which can handle both qualitative and quantitative responses and, consequently, they allow to derive both classification and regression models. Both Kohonen Maps and CP-ANNs have been successfully applied in different research fields (Arakawa et al., 2006; Ballabio et al., 2007a; Cosio et al., 2006; Fermo et al., 2004; Marengo et al., 2006; Marini et al., 2004). Application of SOMs in the field of analytical chemistry is exploding and several examples can be cited just as a matter of example (a complete review is out of the scope of this chapter).

Lee and Scholz (2006) used SOMs to elucidate heavy metal removal mechanisms and predict heavy metal concentrations in experimental constructed wetlands treating urban runoff. Marini et al. (2007a) employed SOMs to authenticate extra virgin olive oil varieties. Samecka-Cymerman et al. (2009) applied SOMs to unravel distinct groups of soils and *Robinia pseudoacacia* leaves and bark, depending on traffic intensity. A map was developed and, then, used to recognize types of pollution in the same geographical area. Torrecilla et al. (2009) used SOMs and learning vector quantification network (LVQ) models to explore the identification of edible and vegetable oils and detect adulteration of extra virgin olive oil using the most common chemical substances in these oils, *viz*. saturated fatty (mainly palmitic and stearic) acids, oleic and linoleic acids. Fonseca et al., (2006) showed that Kohonen SOMs classified samples of crude oils on the basis of gas chromatography-mass spectrometry (GC-MS) descriptors, in terms of geography origin, with a high degree of accuracy.

A reason that can explain the success of SOMs in analytical chemistry is their ability to solve both supervised and unsupervised problems. Besides, and very important from a pragmatic point of view, they are based on relatively easy-to-understand algorithms. Moreover, when dealing with supervised classification issues (a task that must be performed often by analytical chemists) CP-ANNs demonstrated that they can account for non-linear dependence between input and output vectors and, generally, can model classes with non-linear boundaries. On the other hand, since neural networks are non-parametric statistical methods based on adaptable parameters, most of the learning schemes require the use of a test set to optimise the structure of the model. Indeed, one of the major disadvantages of SOMs, as well as other ANN methods, is probably related to model optimisation, because this procedure suffers from some arbitrariness and can be sometimes computationally time-demanding.

Some refinements, modifications and specific implementations of the original SOM algorithm have been proposed in the literature in order to deal with specific issues (Marini et al., 2006 and 2007b; Melssen2007b et al., 2006 and 2007; Schmuker2007 et al., 2007). One of these is the so-called MOLMAP (MOLecular Map of Atom-level Properties) approach used for molecule description in the molecular modelling field (Zhang and Aires-de-Sousa, 2005 and 20072007). In this approach, the objects used for training the neural network were chemical bonds and the input variables were selected properties of bonds. By using a trained SOM, bonds in a molecule were mapped into the SOM and the pattern of activated neurons was interpreted as a fingerprint of the bonds of the molecule.