Chapter 5
Complex-Valued Neural Networks: A New Learning Strategy Using Particle Swarm Optimization

Mohammed E. El-Telbany
Electronics Research Institute, Egypt

Samah Refat
Ain Shams University, Egypt

Engy I. Nasr
Ain Shams University, Egypt

ABSTRACT

In this chapter, the authors will try to go through the problem of learning the complex-valued neural networks (CVNNs) using particle swarm optimization (PSO); which is one of the open topics in the machine learning society. Quantitative structure-activity relationship (QSAR) modelling is one of the well developed areas in drug development through computational chemistry. This relationship between molecular structure and change in biological activity is center of focus for QSAR modelling. Machine learning algorithms are important tools for QSAR analysis, as a result, they are integrated into the drug production process. Predicting the real-valued drug activity problem is modelled by the CVNN and is learned by a new strategy based on PSO. The trained CVNNs are tested on two drug sets as a real world bench-mark problem. The results show that the prediction and generalization abilities of CVNNs is superior in comparison to the conventional real-valued neural networks (RVNNs). Moreover, convergence of CVNNs is much faster than that of RVNNs in most of the cases.

DOI: 10.4018/978-1-7998-0414-7.ch005

INTRODUCTION

The problem of drug design is to find drug candidates from a large collection of compounds that will bind to a target molecule of interest. The development of a new drug is still a challenging, *time-consuming* and *cost-intensive* process and due to the enormous expense of failures of candidate drugs late in their development. Designing '*drug-like*' molecules using computational methods can be used to assist and speed up the drug design process (Lipinski, 2004; Leeson, *et al*., 2004; Li, 2005). The major bottlenecks in drug discovery ware addressed with computer-assisted methods, such as QSAR models (Hansch, 1969), where the molecular activities are critical for drug design. The QSAR models used to predict the drug activity within a large number of chemical compounds using their descriptors that are often generated with high- noise in high-dimensional space. Nowadays, machine learning algorithms have been used in the modelling of QSAR problems (Duch, *et al*., 2007; Chin, & Chun, 2012; Gertrudesa, *et al.*, 2012). They extract information from experimental data by computational and statistical methods and generate a set of rules, functions or procedures that allow them to predict the properties of novel objects that are not included in the learning set. Formally, a learning algorithm is tasked with selecting a hypothesis that best supports the data. Considering the hypothesis to be a function f mapping from the data space X to the response space Y; i.e., f: $X \rightarrow Y$. The learner selects the best hypothesis f^* from a space of all possible hypotheses F by minimize errors when predicting value for new data, or if our model includes a cost function over errors, to minimize the total cost of errors.

As shown in Figure 1, the QSAR modelling is heavily dependent on the selection of molecular descriptors; if the association of the descriptors selected to biological property is strong the QSAR model can identify valid relations between molecular features and biological property/activity. Thus, uninformative or redundant molecular descriptors should be removed using some feature selection methods during (filters) or before (wrappers) the learning process. Subsequently, for tuning and validation of the predictively of learned QSAR model, one of the validation strategy can be applied likes cross-validation, leave-one-out or the full data set is divided into a training set and a testing set prior to learning (See (El-Telbany, 2014) for a survey).

Figure 1. General steps of developing QSAR models