

Protein Sequences Classification Using Modular RBF Neural Networks

Dianhui Wang¹, N.K. Lee¹, T.S. Dillon¹ and N.J. Hoogenraad²

¹Department of Computer Science and Computer Engineering
La Trobe University, Melbourne, VIC 3083, Australia
Ph: +61-3-9479 3034 Fax: +61-3-9479 3060
dhwang@cs.latrobe.edu.au

²Department of Biochemistry
La Trobe University, Melbourne, VIC 3083, Australia

Abstract. A protein super-family consists of proteins which share amino acid sequence homology and which may therefore be functionally and structurally related. One of the benefits from this category grouping is that some hint of function may be deduced for individual members from information on other members of the family. Traditionally, two protein sequences are classified into the same class if they have high homology in terms of feature patterns extracted through sequence alignment algorithms. These algorithms compare an unseen protein sequence with all the identified protein sequences and returned the higher scored protein sequences. As the sizes of the protein sequence databases are very large, it is a very time consuming job to perform exhaustive comparison of existing protein sequence. Therefore, there is a need to build an improved classification system for effectively identifying protein sequences. This paper presents a modular neural classifier for protein sequences with improved classification criteria. The intelligent classification techniques described in this paper aims to enhance the performance of single neural classifiers based on a centralized information structure in terms of recognition rate, generalization and reliability. The architecture of the proposed model is a modular RBF neural network with a compensational combination at the transition output layer. The connection weights between the final output layer and the transition output layer are optimized by delta rule, which serve as an integrator of the local neural classifiers. To enhance the classification reliability, we present two heuristic rules to apply to decision-making. Two sets of protein sequences with ten classes of super-families downloaded from a public domain database, Protein Information Resources (PIR), are used in our simulation study. Experimental results with performance comparisons are carried out between single neural classifiers and the proposed modular neural classifier.

1 Introduction

The aim of classification is to predict target classes for given input patterns. There are many approaches available for classification tasks, such as statistical techniques, decision trees [9] and the neural networks [1]. Neural networks have been chosen as