Optimized Multilayer Perceptrons by Weighted Fisher Criteria for Protein Secondary Structure Prediction

Emerson Cordeiro Morais, Rubem Mondaini
Federal University of Rio de Janeiro, UFRJ, Center of Technology, COPPE,
Ilha do Fundão, 21.945-972, P. O. Box 68511, Rio de Janeiro, Brazil.
E-mail: emersonc, mondaini@cos.ufrj.br

ABSTRACT

Artificial Neural Networks usually provide better approaches to problems than they request: pattern classification; pattern recognition; pattern association; identification; resistance to noise; approximation of functions and machine learning. From the accuracy of results of previous works [1][3][2], it is believed that multilayer perceptrons can be applied perfectly in the prediction of secondary structures of proteins.

A common classifier of neural networks which is usually applied to the prediction of secondary structures of proteins is the Feed-Forward *MLP* (MultiLayer Perceptron), with backpropagation algorithm.

Input arrays and the corresponding output arrays are used for training the *MLP* network. This is an update process to weights and *biases*, which works until the *MLP* approximates the function for associating input arrays with specific output arrays. The generalization property could favor the *MLP* training with an input/output representative group and will obtain better results in the prediction of proteins which have not been classified yet.

The performance of *MLP* network in the process of non-linear and multidimensional mapping from protein collection lead us to think about it as an ideal classifier for prediction of secondary structures.

Besides, to solve the secondary structure prediction problem using neural networks, we hope to be able to project a *MLP* network with two layers based on the Fisher Linear Discriminant Analysis [4] in order to optimize a *MLP* network with two layers that optimizes the initialization parameters and the architecture of the network.

We also hope that the *optimized MLP* (*oMLP)* will be effective in the process of dimensionality reduction of the search space in extensive groups of protein data. In comparison with a conventional *MLP* network, which works with random initialization, we hope to obtain improvement in most of the performance measures by including the accuracy of the secondary structure prediction and convergence properties.

We understand that the *oMLP* possibly will have a better performance in relation to the *conventional MLP* (*cMLP*) networks, for all the different subsets of proteins that will be tested.

References

[1] H. J. Bohr; S. Bohr; R.M. Brunak *et al.*, Protein secondary structure and homology by neural networks. FEBS Letters, v.241, n.1, pp. 223-228 (1998).

[2] G. Pollastri; D. Przybylski; B. Rost *et al*, 2002, Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins: Structure, Function and Genetics, v.47, n.2, pp. 228-235 (2002).

[3] B. Rost & C. Sander, Third generation prediction of secondary structure. Protein Structure Prediction: Methods and Protocols, v.143, n.1, pp. 71-95 (2000).

[4] Z. Wang; Y. Wang; J. Xuan *et al.*, Optimized multilayer perceptrons for molecular classification and diagnosis using genomic data. Bioinformatics, v.22, n.6, pp. 755-761 (2006).