Second International Conference on Bioinformatics (InCoB'03). September 8-10, 2003. Penang, Malaysia.

Predicting Protein Secondary Structure with Neural Network and Dempster-Shafer Theory

Satya N. V. Arjunan[1] Safaai Deris[2] Rosli Md Illias[3] Mohd Saberi Mohamad[2]
Azmee Awang

[1]Laboratory of Bioinformatics, Institute for Advanced Biosciences
Tsuruoka Town Campus of Keio University, 14-1, Baba-cho, Tsuruoka, Yamagata, 997-0035 Japan
Tel: +81-235-29-0800, Fax: +81-235-29-0809, Email: satya@ttck.keio.ac.jp

[2]Department of Software Engineering, Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia, 81310 Skudai, Johore, Malaysia
Tel: +60-7-557-6160 x 3860, Fax: +60-7-556-5044, Email: safaai@fsksm.utm.my

[3]Department of Bioprocess Engineering, Faculty of Chemical and Natural Resources Engineering
Universiti Teknologi Malaysia, 81310 Skudai, Johore, Malaysia
Tel: +60-7-550-5313, Fax: +60-7-558-1463, Email: r-rosli@utm.my

Abstract

Protein secondary structure prediction has benefited largely from the improvements in machine learning techniques. A novel neural network classifier based on the Dempster-Shafer theory has been introduced by Denoeux. It demonstrated favourable performance in some artificial and real world problems when compared against methods such as Support Vector Machines and Kernel Fisher Discriminants. However, it has not been used for protein secondary structure prediction. In this work, we examined the applicability of this technique in secondary structure prediction. We have identified two main drawbacks of this network in predicting secondary structure; the first being its high computational demand when training on large data sets. We alleviated the computational requirements by implementing the training procedure using the data parallel approach in a distributed and shared memory parallel computing environment. Secondly, the Euclidean distance measure used in the neural network is impaired as a result of the high dimensionality of the data. By replacing the distance measure with Multilayer Perceptrons, we are able to improve the classification accuracy. Moreover, at the level of sequence-to-structure prediction, its performance was comparable to the PHD (Profile network from Heidelberg) method, which is one of the best secondary structure prediction schemes.

Keywords: protein secondary structure prediction, neural network, Dempster-Shafer theory, parallel computing