ACon: predicting protein contact-map using additive neural networks and subgroups balancing method

Sh.Ramesht [a], M.Mirzarezaee [a,c*], M.Sadeghi [b,c]

[a] Department of Computer Engineering, Science and Research branch, Islamic Azad University, Tehran, Iran
[b] National Institute of Genetics Engineering and Biotechnology ,Tehran, Iran
[c] School of Biological Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
*mirzarezaee@srbiau.ac.ir

Abstract: Proteins' contact-map is a binary matrix which shows contact points between residues in a protein sequence that is very essential in defining tertiary structure of proteins. Although contact-maps have been studied vastly over the last decade, predicting contact-maps from protein's sequence is a difficult task to do. Many algorithms are developed to predict contact-maps from protein sequences by using evolutionary information and physical constraints but the accuracy of almost all of them are not promising[1]. In this paper, we present a neural network based method called ACon. This method uses a multi-level additive neural networks to predict proteins residue-residue contact-maps. Inputs of each level of the network are provided by the output of previous levels. For training the unbalanced data of contact-map, we used subgroups method which shows better results over prevalent sampling methods. In this method, we keep all samples instead of randomly selecting a portion of them. We have devided our samples to seven groups. Each of these groups have trained separately. This method, divides these groups to L subgroups. Where L is the ratio of the number of non-contacts to the number of contacts in our contact map matrix. Each of these subgroups have trained separately and the results of each L subgroups have summed up by voting. ACon reaches the accuracy of 77 percent with TPR and SPC equals 74 and 78 percent respectively.

Keywords: protein structure prediction; contact-map; additive neural networks; unbalanced datasets

References

 [1] Zh. Wang, J. Xu, "Predicting protein contact map using evolutionary and physical constraints by integer programming" Bioinformatics. Vol. 29 (2013) 266-373.