SELF-ORGANIZING NEURAL NETWORKS AS FILTERS OF CLEAN START CODONS

Ameya Virkar, Guolinag Li, Kunal Agrawal, Hwey Ting, Han Hao, Suisheng Tang, Vladimir B Bajic
BioDiscovery Group, Laboratories for Information Technology, Singapore
bajicv@lit.org.sg

We present a new method for recognition of functional translation initiation sites (TIS) in humans. The method is based on the use of self-organizing map (SOM) artificial neural networks (ANNs). It uses two SOM ANNs, each in a specific filter of the functional TIS context. The primary use of the method is as the add-on system for preliminary filtering functional TISs during massive downloads of sequences from public databases, such as with the FIE system (http://sdmc.krdl.org.sg/FIE/). The method is designed to make virtually no false negatives, while it maintains a reasonable level of true negative predictions. The method is compared with several other systems for prediction of TISs on several datasets, and although its design was aimed at a very specific type of applications, very favorable results are achieved. On a large test set the method achieves sensitivity of 100% and specificity of 44%, which reflects its primary domain of application. However, the sensitivity and specificity of this method is balanced at the level of 90% on anonymous DNA, and thus implies its possible use as an independent TIS-finder, or as a part of more complex systems for gene finding and gene structure prediction.