SPLICE SITE RECOGNITION BY SUPPORT VECTOR MACHINES

Kunal Agrawal, Ameya Virkar, Guolinag Li, Hwey Ting, Han Hao, Suisheng Tang, Vladimir B Bajic

BioDiscovery Group, Laboratories for Information Technology, Singapore
bajicv@lit.org.sg

If there is an accurate system for recognition of splice-sites that separate exons from introns in complex eukaryotic genes, then a significant portion of the gene structure prediction problem would be resolved. Unfortunately no such system is available yet, and a lot of scope for improvements of current methods exits. We present a new system for recognition of donor and acceptor splice-sites in human DNA. The system is designed on the basis of a number of selected higher-order features derived from the splice-site neighborhood, and uses the Support Vector Machine (SVM) technology to build the intelligent system for recognition of the functional splice-sites.

We performed comparison analysis of our system with HMMgene, NetGene 2, NNSPLICE, etc., on several large data-sets. We found that our system performs considerably better then these systems.