Keynote Speech at the Sixth Asia-Pacific Bioinformatics Conference (APBC2008), Kyoto, Japan, Jan 14-17, 2008

Putting More Biology in the Learning Machine

Xuegong Zhang
MOE Key Laboratory of Bioinformatics and Bioinformatics Division, TNLIST /
Department of Automation, Tsinghua University
Beijing 100084, China
Jan 15, 2008

Machine learning methods such as support vector machines (SVMs) and artificial neural networks have become a major category of tools in mining today's high-throughput data in molecular and systems biology. New biological data and questions are emerging very rapidly, which brings challenges for machine learning methods. Enormous efforts have been put for improving performances of existing algorithms and developing new algorithms. These efforts have helped in better solving many biological problems and also advanced the study of the machine learning methods. However, for many current biological questions, the bottleneck may not be the algorithm itself, but the way the algorithm is tuned for the specific question. In this speech, I will present several examples of our recent study to illustrate how SVMs can be better tuned for different types of biological questions. The examples include the discrimination of alternative and constitutive splicing sites, the prediction of CpG island methylation and the recognition of transcription factor binding sites. Not only the performance of the methods were improved in terms of the prediction accuracy, but also from these computational analysis new insights were achieved on the regulation mechanism of alternative splicing, patterns of DNA methylation protection and transcription factor binding. The improvement and the new biological observations were not achieved by developing more advanced algorithms, but were achieved by putting more biological sense in the learning machine.