A Crossover Between SVMs and HMMs for Protein Secondary Structure Prediction

Ioannis Tsochantaridis, Yasemin Altun, and Thomas Hofmann
Department of Computer Science
{it,th,altun}@cs.brown,edu
Brown University, Box 1910, Providence, RI 02912

November 15, 2002

Protein secondary-structure prediction constitutes an important step towards solving one of the most challenging problems in molecular biology: the *protein folding problem*. Essentially this prediction problem can be modeled as a mapping from a string of amino acids (primary structure) to a string of structure labels, for example, a string over the alphabet {helix, sheet, coil}. We have investigated the use of machine learning techniques for sequence label learning and segmentation in this context. Based on a novel learning framework of joint input/output space kernel functions, we have developed a powerful learning architecture that combines large margin principles (SVMs) with the possibility to introduce inter-label dependencies of the type that have been used in generative models like HMMs. We propose an on-line algorithm in the spirit of perceptron learning as well as a quadratic programming formulation for margin maximization. Both techniques use Viterbi decoding as part of the learning procedure. Preliminary results demonstrate the competitiveness of our approach.