Kernel-Based Approaches to Classification and Design of Protein Sequences and Chemical Compounds

Tatsuya Akutsu

Kyoto University

Support vector machines and kernel methods have been applied to various classification problems in bioinformatics. In order to apply kernel methods to bioinformatics problems, it is usually required to develop a kernel function that provides a kind of measure of similarity between two objects. In this talk, we briefly review kernel functions developed for classification of protein sequences and chemical compounds. Then, we present our results on these problems: the local alignment kernel for protein sequences and extensions of the marginalized graph kernel for chemical compounds. On the other hand, a new approach was recently proposed for designing and/or optimizing objects using kernel methods. In this approach, a desired object is computed as a point in the feature space using suitable objective function and optimization technique and then the point is mapped back to the input space. We present some theoretical results on this problem. Though the results are not yet practical, it may suggest future directions of the approach, This talk is based on joint works with Daiji Fukagawa, Pierre Mahe, Jean-Luc Perret, Hiroto Saigo, Nobuhisa Ueda and Jean-Philippe Vert.