Classification of co-expressed genes from DNA regulatory regions

ID - 201

Pavesi Giulio[1], Valentini Giorgio[2]

[1]Dept. of Biomolecular Science and Biotechnology, University of Milano, Milano
[2]Dept. of Computer Science, University of Milano, Milano

Motivation

The analysis of non-coding DNA regulatory regions and the identification of functional elements in them is one of the most challenging open problems in molecular as well as computational biology. In this work, we focused on elements regulating transcription, and investigated whether functional information about genes can be predicted by using information extracted from their promoter sequences coupled with expression data.

Methods

We formalized this problem as a classification problem, and we applied Support Vector Machines (SVMs) with non linear kernels to predict classes of co-expressed genes obtained from clustering procedures. SVMs were trained both by using information about selected motifs extracted from DNA regulatory regions through combinatorial and statistical methods, and by identifying putative regulatory elements with descriptors of the binding specificity of known transcription factors.

Results

Our experiments have indeed shown that functional classes of genes can be predicted from biological sequence data in yeast, and in some cases also on higher eukaryotes like human, achieving results competitive with those recently presented in literature. In general, our methodolgy can be applied with a twofold purpose: either for predicting whether a given gene can be associated with a known cluster of co-regulated or co-expressed genes from sequence data only or, vice versa, for identifying which sequence features are essential to discriminate a given cluster of genes from other clusters or random sets of genes, providing additional strategies for the identification of conserved motifs.

Email: pavesi@dico.unimi.it