Support Vector Machines for predicting protein structural class

ABSTRACT

The Support Vector Machine method and the covariant discrimination algorithm, which is an elegant component-coupled method, are expected to combine their computational capabilities in predicting protein structural classes.

INTRODUCTION

Many disciplines have developed graphical representations of complex systems to communicate, archive and analyze knowledge. Although electronic circuit diagrams and architectural plans are accessible to knowledgeable individuals, there is no universal'real' describing functional systems in biology. Knowledge in specific areas, such as metabolic pathways, gene networks, signaling networks and molecular interactions, is stored and archived in various forms in print or online, but these representations are not standardized to account for cross-disciplinary systems – for example, the interactions between genes and metabolism at the inter- and intra-molecular levels. Inadequate cartoon diagrams are frequently used by biologists to represent multidisciplinary biological problems at various levels, such as intramolecular and disease phenotype, due to their ambiguity and lack of clarity. We propose that a standardized visual biological description language would offer easier and more clear communication, and, with computational resources, provide underlying conditions for distributed searchable archives of functional knowledge, as well as CAD functionality for simulating and analyzing biological systems. Our team is currently developing a prototype biological description language, BioD, to test its conceptual foundation, explore its utility, and identify critical issues surrounding its implementation.

CONCLUSION

Our proposal involves the use of'shaving' methods to isolate intriguing gene clusters from DNA microarray experiments. These methods can either be unsupervised or supervised, depending on available information about the samples, such as a class label or survival time. The proposed shaving methods aim to identify gene clusters that exhibit significant variation across samples and maintain coherence across them. Simple clustering or individual gene thresholding based on sample variation cannot address these aspects. We have developed our model-based shaving method, which allows for the inclusion of other prognostic factors to assist in finding intriguing gene clusters. If a specific outcome is available for each sample, the method searches for matched genes in the group with corresponding column average genes, who may influence the outcome and potentially other contributing factors. The microarray data xij is the first one that we have examined, but it is only available at real-valued expression levels. Other arrays also generate different types of data, such as array methods that detect single-nucleotide polymorphisms (SNPs) and one of k 2 unordered values. The shaving methods described below can be easily modified to handle this type of information. Detailed: We make k data matrices X1, x2... Yk with m [n =], where j[h] is considered 1 (if phys. adj) and d[k] otherwise. Let ijj = 1, 2,.... kW] as the variance matrix for penalty, then we apply principal component shaving so that there are no determinants of each molecule in each expression but some kind of supervision term can be added.This principle allows for both parties to maintain quality