L073

A Considerable Improvement in Protein Fold Recognition Using a Novel Kernel Fusion Framework

Pooya Zakeri, KU Leuven /iMinds, Belgium
Ben Jeuris (KU Leuven , Dept. of Computer Science Belgium); Raf Randebril (KU Leuven, Dept. of Computer Science Belgium); Yves Moreau (KU Leuven /iMinds, Department of Electrical Engineering - ESAT, SCD-SISTA/ Future Health Department Belgium);

Abstract: Useful knowledge on function of proteins can be provided by information of their tertiary structure; hence determination of this structure is among the most essential objectives in biological science. From the other point of view, information about the protein fold can provide much better understanding about its three-dimensional structure. Various protein sequence feature-based techniques have been used in the classification of protein folds. Nevertheless, it is worthwhile to find a efficient and cost-effective method for integrating theses existing different views on data. On the other hand, The heterogeneous biological data sources can intelligibly be integrated as partial integration like kernel-based data fusion. Most of proposed techniques for integrating multiple kernels focus on learning convex linear combination of base kernels. In addition to limitation of linear combination, going with such approaches could cause losing potentially fruitful data information.

We present a new method which combines kernel matrices in taking geometric mean instead of convex linear combination. We employ twenty six different representative models of protein samples based on physicochemical properties and primary structural information of protein sequences, sequence evolutionary information and local sequence alignments. Furthermore, our computational model is developed by incorporating the functional domain composition of proteins through hybridization model. We evaluate our method for classification on the SCOP PDB-40D benchmark dataset for protein fold recognition. By using our proposed hybridization model, the protein fold recognition accuracy is improved to 87.7% that is superior to the results of the best existing approaches.