Abstract

The three dimensional structure of proteins is useful to carry out the biophysical and biochemical functions in a cell. Protein contact maps are 2D representations of contacts among the amino acid residues in the folded protein structure. Proteins are biochemical compounds consisting of one or more polypeptides, facilitating a biological function. Many researchers make note of the way secondary structures are clearly visible in the contact maps where helices are seen as thick bands and the sheets as orthogonal to the diagonal. In this paper, we explore several machine learning algorithms to data driven construction of classifiers for assigning protein off diagonal contact maps. A simple and computationally inexpensive algorithm based on triangle subdivision method is implemented to extract twenty features from off diagonal contact maps. This method successfully characterizes the off-diagonal interactions in the contact map for predicting specific folds. NaiveBayes, J48 and REPTree classification results with Recall 76.38%, 91.66% and 80.32% are obtained respectively.

References

Acids Res. ;28:257-259

- Dynameomics http://www. dynameomics. org/
- Weka http://www. cs. waikato. ac. nz/ml/weka/
- Protein Data Bank http://www. rcsb. org/pdb/home/home. do

Index Terms
Computer Science
Pattern Recognition

Keywords
Protein Contact Maps  Classification  Protein Data Bank  SCOP  J48  REPTree
Naive Bayes