FEATURE SELECTION OF PROTEIN STRUCTURAL CLASSIFICATION USING SVM CLASSIFIER

Zbigniew Krajewski, Ewaryst Tkacz

Silesian University of Technology, Gliwice, Poland

Abstract

Recursive feature elimination method (RFE), cross validation coefficient (CV) and accuracy of classification of test data are applied as a criterion of feature selection in order to find relevant features and to analyze their influence on classifier accuracy. Feature selection method was compared to principal component analysis (PCA) to understand the effectiveness of feature reduction. Support vector machine classifier with radial basis function (R**B**F) kernel is applied to find the best set of features using grid model selection and to select and assess relevant features. The best selected feature set is then analyzed and interpreted as the source of knowledge about the protein structure and biochemical properties of amino acids included in the protein domain sequence.

Keywords: pseudo amino acid composition, support vector machine, principal component analysis, recursive feature elimination, feature selection, SCOP database