

PROTEIN SUBCELLULAR LOCALIZATION PREDICTION BASED ON COMPARTMENT-SPECIFIC BIOLOGICAL FEATURES

Chia-Yu Su^{1,2}, Allan Lo^{1,3}, Hua-Sheng Chiu⁴, Ting-Yi Sung⁴, Wen-Lian Hsu⁴

¹Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica,
Taipei, Taiwan

²Institute of Bioinformatics, National Chiao Tung University, Hsinchu, Taiwan

³Department of Life Sciences, National Tsing Hua University, Hsinchu, Taiwan

⁴Bioinformatics Lab., Institute of Information Science, Academia Sinica, Taipei, Taiwan

Prediction of subcellular localization of proteins is important for genome annotation, protein function prediction, and drug discovery. We present a prediction method for Gram-negative bacteria that uses ten one-versus-one support vector machine (SVM) classifiers, where compartment-specific biological features are selected as input to each SVM classifier. The final prediction of localization sites is determined by integrating the results from ten binary classifiers using a combination of majority votes and a probabilistic method. The overall accuracy reaches 91.4%, which is 1.6% better than the state-of-the-art system, in a ten-fold cross-validation evaluation on a benchmark data set. We demonstrate that feature selection guided by biological knowledge and insights in one-versus-one SVM classifiers can lead to a significant improvement in the prediction performance. Our model is also used to produce highly accurate prediction of 92.8% overall accuracy for proteins of dual localizations.