

Protein subcellular localization prediction based on compartment-specific features and structure conservation

Emily Chia-Yu Su^{1,2}, Hua-Sheng Chiu³, Allan Lo^{1,4}, Jenn-Kang Hwang², Ting-Yi Sung^{3,*}, and Wen-Lian Hsu^{3,*}

¹Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan, ²Institute of Bioinformatics, National Chiao Tung University, Hsinchu, Taiwan, ³Bioinformatics Lab., Institute of Information Science, Academia Sinica, Taipei, Taiwan, and ⁴Department of Life Sciences, National Tsing Hua University, Hsinchu, Taiwan

ABSTRACT

Motivation: Protein subcellular localization is crucial for genome annotation, protein function prediction, and drug discovery. However, since determining subcellular localization by experimental approaches is time-consuming, predicting localization via computational approaches is desirable.

Results: We propose a prediction method for Gram-negative bacteria that adopts a one-versus-one support vector machines model, in which compartment-specific features are incorporated. The method achieves marked improvements of 1.6%~2.9% in overall accuracy using ten-fold cross-validation on the benchmark data sets. Our results demonstrate that biological features derived from Gram-negative bacteria translocation pathways yield a significant improvement. Combining a structure homology approach based on secondary structure alignment improves our method by a further 1.0%~1.6%, which suggests that structure conservation could be a useful indicator for inferring localization. In addition, non-redundant data sets are used to show that general performance for prediction should be approximately 80%.

Contact: hsu@iis.sinica.edu.tw; tsung@iis.sinica.edu.tw

Supplementary Information: Supplementary material is available online at <http://www.iis.sinica.edu.tw/~cysu/PSL101/Supplement.pdf>
