

工程与应用

基于SVM和KNN的蛋白质耐热性分类

丁彦蕊^{1,2}, 蔡宇杰^{2,3}, 孙俊¹, 须文波¹

1.江南大学 信息工程学院, 江苏 无锡 214122

2.江南大学 工业生物技术教育部重点实验室, 江苏 无锡 214036

3.江南大学 生物工程学院, 江苏 无锡 214036

收稿日期 修回日期 网络版发布日期 2007-5-19 接受日期

摘要 以氨基酸含量为特征向量, 研究了SVM和KNN预测蛋白质耐热性的准确度。结果表明, 基于SVM的分类效果较好, 其局部预测率和全局预测率分别为82.4%和83.4%; 而基于KNN方法的局部预测率和全局预测率分别为77.6%和79.9%。两种方法的预测率均表明氨基酸含量是影响蛋白质耐热性的主要因素。

关键词 [氨基酸含量](#) [SVM](#) [KNN](#) [蛋白质耐热性](#)

分类号

Classification of protein thermostability using Support Vector Machines and K-Nearest Neighbors

DING Yan-rui^{1,2}, CAI Yu-jie^{2,3}, SUN Jun¹, XU Wen-bo¹

1.School of Information Technology, Southern Yangtz University, Wuxi, Jiangsu 214122, China

2.Key Laboratory of Industrial Biotechnology, Wuxi, Jiangsu 214036, China

3.School of biotechnology, Wuxi, Jiangsu 214036, China

Abstract

Regarding amino acid composition as eigenvector, protein thermostability is classified using Support Vector Machines and K-Nearest Neighbors. It is found that the result of using support vector machines is better than K-Nearest Neighbors. The local accuracy and global accuracy are 82.4% and 83.4% respectively. But the local accuracy and global accuracy are 77.6% and 79.9% respectively using K-Nearest Neighbors. The prediction accuracy of two kinds of methods can both prove that the amino acid composition is the main factor that influences the protein thermostability.

Key words [amino acid composition](#) [SVM](#) [KNN](#) [protein thermostability](#)

DOI:

通讯作者 丁彦蕊 [E-mail: yanrui_ding76@yahoo.com.cn](mailto:yanrui_ding76@yahoo.com.cn)

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(802KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [参考文献](#)

服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ [本刊中 包含“氨基酸含量” 的相关文章](#)

▶ [本文作者相关文章](#)

· [丁彦蕊](#)

· [蔡宇杰](#)

· [孙俊](#)

· [须文波](#)