生物物理学报(2003) 19(2), 171-175

PDF文档

基于支持向量机和贝叶斯方法的蛋白质四级结构分类研究

张绍武、潘泉、张洪才、张云龙、王海瑜
西北工业大学自动控制系

用支持向量机和贝叶斯两种方法对蛋白质四级结构进行分类研究。结果表明，基于支持向量机的分类结果最好，其10CV检验的总分类精度、正样本正确预测率、Matthes相关系数和假阳性率分别为74.2%、84.6%、0.474、38.9%；基于贝叶斯的分类结果没有支持向量机的分类结果好，但其10CV检验的假阳性率最低(15.9%)。这些结果说明同源寡聚蛋白质一级序列包含四级结构信息，同时特征向量的确表示了埋藏在缔合亚基作用部位接触表面的基本信息。

CLASSIFICATION OF QUATERNARY STRUCTURE USING SUPPORT VECTOR MACHINES AND BAYES METHODS

The quaternary structure was classified using support vector machine method and Bayes method. It was found that the result of using support vector machine is the best, using 10-fold cross-validation test, the overall accuracy, true positive rate, Mattew's correlation coefficient and false negative rate are 74.2%, 84.6%, 0.474, 38.9% respectively; the result of Bayes method is not so good as that of the support vector machine method, the false negative rate of using 10-fold cross-validation test is the smallest. Those results show that the primary sequences of homo-oligomeric proteins contain quaternary information. The feature vectors appear to capture essential information about the composition and hydrophobicity of the residues in the surface patches that are buried in the interfaces of associated subunits. And they also show that the support vector machines is a specially effective method.

关键词

支持向量机(Support vector machines); 贝叶斯(Bayes); 蛋白质四级结构(Protein quaternary structure); 亚基(Subunits)