

# 基于支持向量机识别真核生物DNA中的翻译起始位点

詹泳\*、周艳红、卢正鼎  
华中科技大学计算机学院

翻译起始位点(TIS)的识别是真核生物基因预测的关键步骤之一,近年来一直得到研究人员的高度重视。基于TIS附近序列的统计特性,出现了一些辨识TIS的判别方法,但识别精度还有待进一步提高。针对传统支持向量机(SVM)方法中存在的不足,提出了基于数据优化法的SVM,它通过其它统计学模型优化训练数据集,进而提高分类器的辨识精度。实验结果表明基于数据优化法的SVM分类器在翻译起始位点的辨识上可获得比其他判别方法更好的效果。

## SVM BASED RECOGNITION OF TRANSLATION INITIATION SITES IN EUKARYOTIC DNA

As one of crucial steps of gene prediction, an intensive attention was paid to the recognition of translation initiation site (TIS) in recent years. On the basis of the statistical feature of sequences around TIS, some methods have been applied to identify TIS. However, their prediction accuracy remains to be improved. Considering the deficiency of traditional SVM, data optimization based SVM is presented. Its scheme is to improve the prediction accuracy of classifier by optimizing the training dataset with other statistical model. Experiment results show that data optimization based SVM is better than other methods.

### 关键词

支持向量机(Support vector machine); 翻译起始位点(Translation initiation site); 数据优化(Data optimization); 敏感度(Sensitivity)