基于支持向量机的人类5'非翻译区剪接位点识别

晏春*、杜耀华、高青斌、王正志
国防科技大学自动控制系

基因非编码区域剪接位点的识别是基因识别中一个非常具有挑战性的问题，尤其是5'非翻译区（5'UTRs）中剪接位点的识别。与一般剪接位点不同，5'UTRs剪接位点的两侧不存在由编码到非编码的状态转移，所以通常的剪接位点识别算法在UTRs区域的性能不太理想。本文采用了支持向量机（SVM）对5'UTRs中的剪接位点进行识别。为了提高识别精度，我们利用了基于矩阵相似性度量的核函数参数确定方法，它能够简单快速地确定合适的核函数参数，进而提高核函数的识别性能。通过实验验证，经过参数选择后的SVM能够较好的识别5'UTRs剪接位点

Identification of 5' UTR Splice Sites In Human Gene Based On Support Vector Machine

Identification of splice sites in non-coding regions of genes is one of the most challenging aspects of gene structure recognition, especially the identification of splice sites embedded in human 5' untranslated regions (UTRs). Different from the conventional splice sites identification, there is no transition from coding to non-coding in 5' UTRs, so conventional splice sites prediction methods perform poorly in UTRs. In this paper, we use Support Vector Machines to identify 5' UTR splice sites. To increase recognition accuracy, we select the kernel's parameters based on the measurement of matrix similarity. By doing this, we can get the apropos parameters quickly and simply, thereby improve the identification performance. Experiment results show that 5' UTR splice sites can be identified well based on SVM with the selection of parameters.

关键词