贝叶斯聚类在基因表达谱知识挖掘中的应用

包雷、李泽、孙之荣
清华大学生物科学与技术系

在大规模基因表达谱的数据分析中引入了一种全新的基于贝叶斯模型的聚类算法。从生物学背景出发，研究了该算法应用在大规模基因表达谱中的理论基础和算法优越性，并应用该算法对两个公共的基因表达数据集进行了知识再挖掘。结果表明，与其他聚类算法相比，该算法在知识发现方面具有显著的优越性。挖掘出的生物学知识对该领域研究人员的实验设计也有一定的启发性。

APPLICATION OF BAYESIAN CLUSTERING TO THE ANALYSIS OF GENE EXPRESSION PROFILES

A novel clustering algorithm based on Bayesian model was introduced into the analysis of large-scale gene expression profiles. It was applied to analyze two public gene expression data sets and some new biological knowledge was discovered. Compared with other clustering algorithms used in the gene expression profile analysis, this algorithm worked better both in the algorithmic and the biological sense.

关键词

基因表达谱(Gene expression profiles)；聚类算法(Clustering algorithm)；数据挖掘(Knowledge discovery)；AUTOCLASS程序(AUTOCLASS)