计算机应用 2013, 33(08) 2188-2193 DOI: ISSN: 1001-9081 CN: 51-1307/TP

本期目录 | 下期目录 | 过刊浏览 | 高级检索 [打印本页] [关闭]

数据库技术

相对行常量差异共表达双聚类挖掘算法

谢华博,尚学群,王淼

西北工业大学 计算机学院，西安 710129

摘要： 在生物信息学上，挖掘差异共表达双聚类有助于研究衰老、癌变类变化的生物过程。以往的差异共表达双聚类定义仅仅从一组基因的角度来衡量差异，导致包含了很多噪声。为了克服上述缺点提出新的差异共表达支持度MiSupport，可以将一组基因的差异细化到基因级别；并由此定义提出MiCluster算法，可以在两个真实的基因芯片数据中挖掘最大的差异共表达双聚类。MiCluster算法首先基于两个基因芯片数据构建差异共表达权值图，然后基于权值图，采用样本扩展和层次扩展，并利用精确的候选产生方法和高效的剪枝策略，挖掘出最大的差异共表达双聚类。实验结果证明,MiCluster算法比现有的算法快速高效，而且通过均方误差(MSE)测试和基因本体(GO)评价，挖掘出来结果具有更大的统计意义和生物学意义。

关键词： 基因芯片 基因共表达 双聚类 差异 行常量

Differential co-expression relative constant row bicluster mining algorithm

XIE Huabo,SHANG Xuequn,WANG Miao

School of Computer Science, Northwestern Polytechnical University, Xi'an Shaanxi 710129, China

Abstract: Bioinformaticly, it is useful to study the change process of biology, such as aging and canceration, by mining differential co-expression bicluster. The definition in the past only measured from the perspective of all set of genes, thus containing a lot of noise. Therefore, a new definition named MiSupport was put forward to measure the difference on gene level, and on the basis of MiSupport, an algorithm named MiCluster was proposed to mine the maximal differential co-expression bicluster in two real gene chips. Firstly, MiCluster constructed a differential weighted undirected sample-sample relational graph in two real-valued gene expression datasets. Secondly, the maximal differential biclusters was produced in the above differential weighted undirected sample-sample relational graph with efficiently pruning techniques and accurately generating candidates method by sample-growth and level-growth. The experimental results show that MiCluster is more efficient than the existing methods. Furthermore, the performance is evaluated by Mean Square Error (MSE) score and Gene Ontology (GO). The results show that this algorithm can find better statistical and biological significance.

Keywords: gene chip gene co-expression bicluster differential constant row

收稿日期 2013-03-05 修回日期 2013-04-15 网络版发布日期 2013-09-11

DOI:

基金项目:

国家973计划项目;国家自然科学基金资助项目

通讯作者: 谢华博

作者简介: 谢华博(1987-),男，江西于都人,硕士研究生,主要研究方向：生物数据挖掘、差异共表达;
尚学群(1973-),女，陕西西安人,教授,博士,主要研究方向：数据库、数据挖掘、生物信息学;
王淼(1981-),男, 河南义马人,博士,主要研究方向：数据挖掘、生物信息学。
作者Email: 282525634@qq.com

参考文献：

[1] MADEIRA S C, OLIVEIRA A L. Biclustering algorithms for biological data analysis: a survey ［C］// IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2004, 1(1):24-45.

[2] HARTIGAN J A. Direct clustering of a data matrix ［J］. Journal of the American Statistical Association, 1972, 67 (337):123-129.

[3] GETZ G, LEVINE E, DOMANY E. Coupled two-way clustering analysis of gene microarray data ［J］. Proceedings of the Natural Academy of Sciences of United States of America, 2000, 97(22): 12079-12084.

[4] CORMEN T H, ELEISERSON C E, RIVEST R L, et al. Introduction to algorithms ［M］. 2nd ed. Cambridge: MIT Press, 2001.

[5] LAZZERONI L, OWEN A. Plaid models for gene expression data ［J］. Statistica Sinica, 2002, 12: 61-86.

扩展功能
本文信息
Supporting info
PDF(1080KB)
[HTML全文]
参考文献[PDF]
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
引用本文
Email Alert
文章反馈
浏览反馈信息
本文关键词相关文章
基因芯片
基因共表达
双聚类
差异
行常量
本文作者相关文章
谢华博
尚学群
王淼
PubMed
Article by Xie,H.B
Article by Chang,H.Q
Article by Yu,m

［6］KOSTKA D, SPANG R. Finding disease specific alterations in the coexpression of genes ［J］. Bioinformatics, 2004, 20 (Suppl. 1): i194-i199.

［7］OKADA Y, INOUE Y. Identification of differentially expressed gene modules between two-class DNA microarray data ［J］. Bioinformation, 2009,4(4): 134-137.

［8］SERIN A, VINGRON M. DeBi: discovering differentially expressed biclusters using a frequent itemset approach ［J］. Algorithms for Molecular Biology, 2011,6(1):18.

［9］BURDICK D, CALIMLIM M, GEHRKE J. MAFIA: a maximal frequent itemset algorithm for transactional databases ［C］// Proceedings of the 17th International Conference on Data Engineering. Piscataway: IEEE, 2001: 443-452.

［10］ODIBAT O, REDDY C K, GIROUX C N. Differential biclustering for gene expression analysis ［C］// Proceedings of the ACM Conference on Bioinformatics and Computational Biology. New York: ACM, 2010: 275-284.

［11］FANG G, KUANG R, PANDEY G, et al. Subspace differential coexpression analysis: problem definition and a general approach ［C］// Proceedings of the 15th Pacific Symposium on Biocomputing. Singapore: World Scientific Publishing, 2010:145-156.

［12］WANG M, SHANG X Q, LI X Y, et al. Efficient mining differential co-expression constant row bicluster in real-valued gene expression datasets ［J］. Applied Mathematics & Information Sciences, 2013, 7(2):587-598.

［13］CHENG Y, CHURCH G M. Biclustering of expression data ［C］// Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology. ［S.l.］: AAAI, 2000: 93-103.

［14］ZAHN J M, POOSALA S, OWEN A B, et al. AGEMAP: a gene expression database for aging in mice ［J］. PLoS Genetics, 2007, 3(11): e201

［15］The Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource ［J］. Nucleic Acids Research, 2004, 32(1): D258-D261.

本刊中的类似文章

1. 刘军 谭德庆.供应链促销-定价决策与内生时机[J]. 计算机应用, 2013,33(04): 971-975

2. 陈骅 檀结庆.基于空间分布差异度的分块彩色图像检索方法[J]. 计算机应用, 2012,32(06): 1539-1543

3. 何骞 卓碧华.一种远程文件同步方法[J]. 计算机应用, 2012,32(02): 566-568

4. 谭义红 陈治平 李学勇 林亚平.基于k-完美差异图的超节点拓扑结构构造[J]. 计算机应用, 2011,31(08): 2021-2024

5. 林亚忠 郝刚 顾金库.利用邻域差异性信息的FCM改进算法[J]. 计算机应用, 2011,31(02): 375-378

6. 王筠 郭莹 杨萍 杨美红.领域需求差异分析方法与应用研究[J]. 计算机应用, 2010,30(8): 2177-2180

7. 冯慧军 陈斌 赵向辉 夏凡.基于能量最小的拉普拉斯流域分割算法[J]. 计算机应用, 2009,29(2): 462-464

8. 汤周文 叶东毅.基于层次聚类的差异化属性约简算法[J]. 计算机应用, 2009,29(2): 419-420

9. 王金林 赵辉.基于DE的ε-SVRM参数优化研究[J]. 计算机应用, 2008,28(8): 2074-2076

10. 黄晓春，晏蒲柳，夏德麟，陈健.基于差异—相似矩阵的文本降维方法[J]. 计算机应用, 2005,25(08): 1821-1823

11. 张光前，邓贵仕.基于事例推理中差异驱动的事例修改策略研究[J]. 计算机应用, 2005,25(07): 1658-1660