

论文

一种新的基因表达数据聚类方法

王文俊;张军英

(西安电子科技大学 计算机学院, 陕西 西安 710071)

摘要:

提出了一种基于样本间关系的新聚类方法,从基因表达数据中通过pearson相关系数获得样本间的关系,并用网络的方法表示这种关系,通过该网络的空间结构特征来提取样本间的关系特征,并在这种关系特征空间中进行样本的聚类.该方法能更好地揭示不同类样本间的差异性,具有聚类空间维数低而无需降维的特点.分别采用本方法与现有的聚类方法对真实的基因表达数据进行了聚类分析,实验结果说明该方法能获得更高的聚类正确率,且对于分布混杂的数据的聚类效果也较好.

关键词: 聚类 样本关系网络 结构特征 关系特征

New method for clustering gene expression data

(School of Computer Science and Technology, Xidian Univ., Xi'an 710071, China)

(School of Computer Science and Technology, Xidian Univ., Xi'an 710071, China)

Abstract:

A new clustering method based on the relationship between patterns is proposed. The relationship between patterns is obtained from gene expression data through the pearson correlation coefficient, which is denoted by a network, the relation feature between patterns is extracted by discovering the structure feature of the network, and clustering is performed in the relation feature space. The proposed method uncovers the dissimilarity between patterns belonging to different classes more effectively, and the dimensionality of the clustering space is so low than there is no need to reduce dimensions. The comparison of the method with the conventional ones shows that the method can obtain a much higher clustering efficiency than other methods and it can lead to a better efficiency even for those data with promiscuous distribution.

Keywords: clustering pattern relation network structure feature relation feature

收稿日期 2008-02-28 修回日期 2008-04-02 网络版发布日期 2009-05-25

DOI:

基金项目:

国家自然科学基金资助(60371044)

通讯作者: 王文俊

作者简介:

参考文献:

- [1] Witten I H, Frank E. Data Mining: Pracitcal Machine Learning Tools and Techniques [M]. Sna Francisco: Morgan Kaufmann Publishers, 2005.
- [2] Golub T R, Slonim D K, Tamayo P, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring [J]. Science, 1999, 286(5439): 531-537.
- [3] Karypis G, Han E H, Kumar V. CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling [J]. Journal of Computer Using Dynamic Modeling, IEEE Computer, 1999, 32(8): 68-75.
- [4] Eisen M B, Spellman P T. Cluster Analysis and Display of Genome-wide Expression Ppatterns [J]. The National Academy of Science, 1998, 95(25): 14863-14868.
- [5] Herwig R, Poustka A J. Large-scale Clustering of cDNA-fingerprinting Data [J]. The National

扩展功能

本文信息

Supporting info

PDF(573KB)

[HTML全文](1KB)

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

▶ 聚类

▶ 样本关系网络

▶ 结构特征

▶ 关系特征

本文作者相关文章

▶ 王文俊

▶ 张军英

PubMed

Article by Yu,W.J

Article by Zhang,J.Y

Academy of Science, 1999, 9(11): 1093-1105.

[6] Kohonen T. The Self-Organizing Map [J]. Proc IEEE, 1990, 78(9): 1464-1480.

[7] 宫改云, 毛用才, 高新波, 等. 模糊c-均值聚类的微阵列基因表达数据分析 [J]. 西安电子科技大学学报, 2004, 31(2): 291-295.

Gong Gaiyun, Mao Yongcai, Gao Xinbo, et al. Fuzzy C-mean Clustering Method for Analyzing Microarray Gene Expression data [J]. Journal of Xidian University, 2004, 31(2): 291-295.

[8] Rougemont J, Hingamp P. DNA Microarray Data and Contextual Analysis of Correlation Graphs [J]. BMC Bioinformatics, 2003, 4(1): 15.

[9] Qu Yi, Xu Shizhong. Supervised Cluster Analysis for Microarray Data Based on Multivariate Gaussian Mixture [J]. Bioinformatics, 2004, 20(12): 1905-1913.

[10] Herrero J, Valencia A, Dopazo J. A Hierarchical Unsupervised Growing Neural Network for Clustering Gene Expression Patterns [J]. Bioinformatics, 2001, 17(2): 126-136.

本刊中的类似文章

1. 石恒华; 许鑫. 基于层次聚类法的网络延迟监测点放置

[J]. 西安电子科技大学学报, 2007, 34(7): 185-188

2. 马秀丽; 焦李成. 联合模型初始化独立谱聚类算法

[J]. 西安电子科技大学学报, 2007, 34(5): 768-772

3. 张宏怡^{1,2}; 张军英¹. 延迟基因调控网络重构问题研究

[J]. 西安电子科技大学学报, 2007, 34(5): 809-813

4. 马文萍; 尚荣华; 焦李成. 免疫克隆优化聚类技术

[J]. 西安电子科技大学学报, 2007, 34(6): 911-915

5. 杨勋¹; 谢维信^{1,2}; 黄建军². 量子球壳聚类

[J]. 西安电子科技大学学报, 2008, 35(1): 43-48

6. 赵峰^{1,2}; 张军英²; 刘敬². 核最优变换与聚类中心的算法

[J]. 西安电子科技大学学报, 2009, 36(1): 127-133

7. 伍忠东¹; 高新波¹; 谢维信². 基于核方法的模糊聚类算法 [J]. 西安电子科技大学学报, 2004, 31(4): 533-537

8. 徐学洲; 卢科学; 杨儒亮. 储层识别的神经网络方法实现的几个问题 [J]. 西安电子科技大学学报, 1998, 25(6): 0-0

9. 徐斌; 陈平; 杨刚. VLSI及PCB设计中的核度聚类布局算法 [J]. 西安电子科技大学学报, 1998, 25(6): 0-0

10. 张正阳; 段勇; 吴成柯. 一种快速高效的VQ初始码书设计方法 [J]. 西安电子科技大学学报, 1998, 25(5): 0-0

11. 暂时无作者信息. 数据驱动的二维点集自适应聚类算法 [J]. 西安电子科技大学学报, 1999, 26(3): 300-305

12. 暂时无作者信息. 模糊联想记忆系统规则提取的新方法 [J]. 西安电子科技大学学报, 1999, 26(4): 452-457

13. 暂时无作者信息. 基于遗传算法的曲线段连接与拟合 [J]. 西安电子科技大学学报, 2000, 27(1): 44-49

14. 暂时无作者信息. 一种基于特征加权的模糊c-均值聚类算法 [J]. 西安电子科技大学学报, 2000, 27(7): 80-84

15. 裴继红; 龚忻; 范九伦. 基于贴近度的聚类有效性函数 [J]. 西安电子科技大学学报, 2000, 27(7): 89-93

16. 徐松涛^{1,2}; 杨绍全¹. 被动单脉冲导引头干扰源角度分辨 [J]. 西安电子科技大学学报, 2004, 31(1): 102-105

17. 罗明; 杨绍全. 一种多进制频移键控信号的调制分类及解调方法 [J]. 西安电子科技大学学报, 2005, 32(1): 52-55

18. 高新波; 范九伦; 谢维信. 区间值数据模糊c-均值聚类新算法 [J]. 西安电子科技大学学报, 1999, 26(5): 604-610

19. 宫改云¹; 毛用才²; 高新波²; 刘三阳¹. 基于模糊c-均值聚类的微阵列基因表达数据分析 [J]. 西安电子科技大学学报, 2004, 31(2): 291-295

20. 暂时无作者信息. 医学图像分割的神经网络方法 [J]. 西安电子科技大学学报, 1998, 25(5): 0-0

21. 暂时无作者信息. 基于数据划分最大信息的聚类有效性函数 [J]. 西安电子科技大学学报, 2001, 28(6): 781-785

22. 王俊平^{1,2}; 郝跃¹; 荆明娥¹. IC真实缺陷的骨架提取方法 [J]. 西安电子科技大学学报, 2005, 32(2): 206-209

23. 范九伦; 裴继红; 谢维信. 模糊相关度与聚类有效性 [J]. 西安电子科技大学学报, 1998, 25(1): 0-0

24. 汪涛; 卢朝阳. 一种新的汽车牌照字符切分算法 [J]. 西安电子科技大学学报, 2005, 32(6): 931-934

25. 李良群; 姬红兵. 基于最大熵模糊聚类的快速数据关联算法 [J]. 西安电子科技大学学报, 2006, 33(2): 251-256

26. 赵恒; 杨万海; 张高煜. 模糊K-Harmonic Means聚类算法 [J]. 西安电子科技大学学报, 2005, 32(4): 603-606

27. 黄健斌^{1,2}; 姬红兵¹. 基于模糊概念格的Web搜索结果聚类算法 [J]. 西安电子科技大学学报, 2005, 32(6): 856-860

28. 暂时无作者信息. 聚类神经网络的通用设计方法 [J]. 西安电子科技大学学报, 1997, 24(1): 0-0

29. 暂时无作者信息. 聚类的密度函数方法 [J]. 西安电子科技大学学报, 1997, 24(4): 0-0

30. 田小林; 焦李成; 缙水平. 加权空间函数优化FCM的SAR图像分割

[J]. 西安电子科技大学学报, 2008, 35(5): 846-852

31. 田小林; 焦李成; 缙水平. 视觉特征方向流邻域加权PCM的SAR图像分割

[J]. 西安电子科技大学学报, 2008, 35(4): 624-631

32. 冯少荣^{1;2};肖文俊¹.一种提高DBSCAN聚类算法质量的新方法
[J]. 西安电子科技大学学报, 2008,35(3): 523-529

33. 黄文龙¹;焦李成¹;贾建^{1;2}.一种结构自适应免疫抗体竞争网络的非监督图像分割
[J]. 西安电子科技大学学报, 2008,35(3): 444-448

34. 许录平;谢维信;李周昱.人机界面中彩色的定量选择与调色板获取方法研究[J]. 西安电子科技大学学报, 1997,24(1): 0-0

文章评论

序号	时间	反馈人	邮箱	标题	内容
1	2009-10-21	caragon	caragon@googlemail.com		?? &????????????????????????f???ugg ukugg saleugg bootsUGG Bailey Buttonsupra shoesnike dunkMBT Shoes discountugg sale ugg shoes ugg