一种新聚类算法在基因表达数据分析中的应用

曹 晖,席 斌,米 红
CAO Hui,XI Bin,MI Hong

厦门大学 信息科学与技术学院 模式识别与智能系统研究所,福建 厦门 361005
College of Science and Technology,Xiamen University,Xiamen,Fujian 361005,China
E-mail:sunnyxx2005@gmail.com

CAO Hui,XI Bin,MI Hong.Application of new clustering algorithms in gene expression data.Computer Engineering and Applications,2007,43(18):234-238.

Abstract: Self-Organizing Maps (SOM) and the hierarchical clustering are two of the most classical clustering technologies for analyzing gene expression data,which exist own advantages and disadvantages on account of the complexity and the instability of gene expression data.Therefore,on base of comparing difference of the two clustering technologies,this article creatively proposes one new algorithm,that is first clustering gene expression data with SOM and second clustering the weight of nerve cells corresponding the clustering from the first step.In succession,the new algorithm is applied to the published data of yeast gene expression to prove that it conquers some bug of SOM and improves the efficiency of gene clustering through emulation mode.
Key words: Self-Organizing Maps(SOM) algorithm;hierarchical clustering;gene expression data

摘 要:自组织特征映射神经网络与层次聚类算法是两种较经典的分析基因表达数据的聚类算法,但由于基因表达数据的复杂性与不稳定性,这两种算法都存在着自身的优劣。因此,在比较两种算法差异性的基础上,创造性地提出了一种新算法,即通过SOM算法对基因表达数据进行聚类,再用层次聚类将每个类对应的神经元权值二次聚类,并将此算法应用在酵母菌基因表达数据中,用实验证明改进算法克服了自组织算法的一些缺陷,提高了基因聚类的效能。
关键词:SOM算法;层次聚类;基因表达数据
文章编号:1002-8331(2007)18-0234-05 **文献标识码**:A **中图分类号**:TP301

1 引言

随着20世纪80年代末期人类基因组计划(Human Genome Project,简称HGP)的全面启动,生物信息学(Bioinformatics)这一新兴的交叉学科应运而生,并在之后的20多年里得到了蓬勃的发展。人类可以从分子粒度上来研究生物,这一突破也使得生物科学技术飞速发展。同时,功能基因组和蛋白质组的大量数据开始涌现,基因芯片[1]的发明使得同时研究和比较大量基因的特性成为可能。然而面对随之产生的海量的基因表达数据,如何运用信息科学与计算机技术对这些数据进行分析处理,从中挖掘出对生物学实验有指导意义的信息或知识成为当前生物信息学研究的一大新课题。

数据挖掘(Data Mining)作为知识发现的重要手段,近年来已经在越来越多的领域得到广泛地应用。而在生物信息学领域,已有多种数据挖掘和信息处理技术应用于基因表达数据分析,主要包括:聚类分析、多元统计、模式识别及神经网络几大类。其中,聚类分析能够检测具有相似表达谱的基因群,并将功能相关的基因按表达谱的相似程度归纳成类,有助于对未知功能的基因进行研究,是目前基因表达分析研究的主要计算技术之一。基因芯片实验得到的大量数据通过聚类分析,可以得到很多有用的信息,其成功应用已广泛涉及到生物医学研究中的各个领域。

人工神经网络(Artificial Neural Network,简称ANN)是采用大量的仿人工神经细胞模型相互连接构成的一种分布式并行信息处理网络,具有与人脑相类似的学习记忆能力和输入信息特征提取能力。其中自组织特征映射网络是神经网络用于聚类的一个例子,具有非人控和自适应的特点。层次聚类是分析基因表达数据最广泛的聚类算法,是基因表达数据聚类方面事实上的标准。这两种算法已基本发展成熟,并在基因表达数据的分析中得到大量的应用。但基于算法原理的不同,这两种经典的算法存在各自的优缺点。

2 SOM算法与层次聚类算法评析

2.1 自组织特征映射神经网络(Self-Organizing Map)算法

自组织特征映射网络是由芬兰赫尔辛基大学神经网络专家Kohonen教授在1981年提出的,是人工神经网络用于聚类分析中的一个例子。这种网络模拟大脑神经系统自组织特征映射功能,它是一种竞争式学习网络,在学习中能够无监督地进

基金项目:厦门大学985研究项目。
作者简介:曹晖(1984-),男,硕士研究生,研究方向为数据挖掘技术与应用;席斌(1963-),男,副教授,研究方向为人工神经网络;米红(1962-),男,系统研究所所长,厦门大学信息科学与技术学院自动化系重点岗位教授,厦门大学人口资源环境与GIS研究中心主任。

行自组织学习。

在自组织映射里,神经元被放置在一维或二维的网格节点上,在竞争学习过程中,神经元变化依不同输入模式(刺激)或输入模式的类别而选择性的调整。调整后的神经元的位置彼此之间成为有序的,使得对于不同的输入特性,在网格上建起有意义的坐标系。因此,自组织映射由输入模式的拓扑映射结构所表征,其中网格神经元的空间位置表示输入模式包含的内在统计特征。

其本质是:输入模式的连续输入空间通过网格中神经元之间的竞争过程映射到神经元的离散输出空间,见图 1。

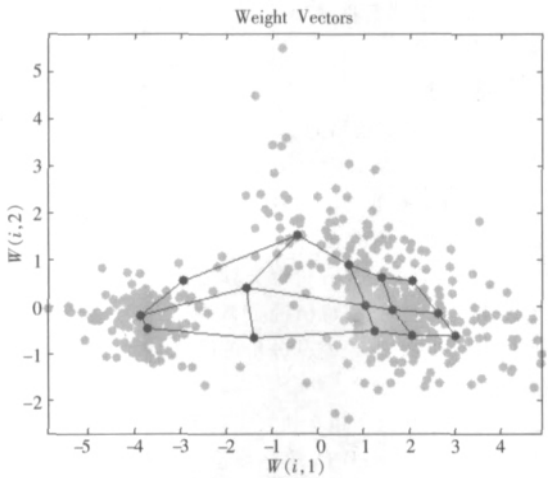

图 1　二维自组织特征映射示意图:数据点与神经元

图 2 为 SOM 的网络结构,它由输入层和竞争层组成。输入层神经元数为 n,竞争层由 $M=m^2$ 个神经元组成的二维平面阵列,输入层与竞争层各神经元之间实现全互连接。

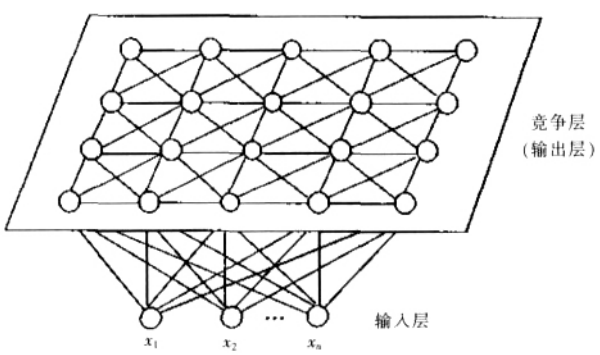

图 2　自组织特征映射网络图

在网络的竞争层,各神经元竞争对输入模式的响应机会,最后仅一个神经元成为胜利者,并对那些获胜神经元有关的各权重朝着更有利于它竞争的方向调整,即以获胜神经元为圆心,对近邻的神经元表现出兴奋性侧反馈。这说明在竞争层,近邻神经元相互激励,远邻神经元相互抑制,比远邻更远的神经元则表现弱激励,通常用‘墨西哥草帽函数’(见图 3)对神经元侧反馈进行计算[2]。

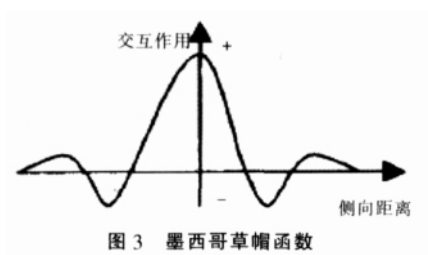

图 3　墨西哥草帽函数

应用侧反馈原理,在每个获胜神经元附近形成一个‘聚类区’。学习的结果总是使聚类区内各神经元的权重向量保持与输入向量逼近的趋势,从而使具有相近特性的输入向量聚集在一起。这种自组织聚类过程是系统自主、无导师指导的条件下完成的。

自从 1999 年 Tamayo 教授[3]将自组织特征映射神经网络应用在区分白血病 HL-60 样本以来,该算法在后来的聚类基因表达数据方面发挥了巨大的作用。其优点如下:

(1)由于神经网络的自适应性,SOM 算法对基因表达数据的聚类具有较高的稳定性和智能性,这一点在聚类有缺失数据的基因表达中有较好的体现;

(2)算法的结果实现的对高维数据的降维映射;

(3)能够产生更为结构化的初始中心,也就是权重矢量矩阵,同时把聚类结果在空间中体现;

(4)可以通过改变学习率大小来避免开始学习的慢或最后不稳定的学习状态,在输入量比较小的时候可以将数据多次输入,以得到较好的训练效果。

虽然 SOM 算法具有较高的稳定性,但有些情况下仍存在较大缺陷,甚至出现非常糟糕的聚类结果:

(1)SOM 算法最大的缺陷在于:需要预先指定参数,即选择输出的神经元数。而在实际应用中很难预测类数,需要通过多次实验,比较其结果,并从生物学角度对结果进行验证。当神经元数选择过少时,聚类的效果可能出现较严重的失误,而当神经元数目过多时,聚类效果又不具有分析性,这在分析指定的基因表达数据分类中往往产生不理想的结果。

(2)存在聚类‘边界’问题[4],例如一个 5×5 的神经元网络和一个 10×10 的神经元网络就具有不同的边界距离;

(3)对于过高维的基因表达数据,竞争胜利点会耗费大量的时间。

2.2　层次聚类算法

层次聚类又称系统聚类或等级体系聚类,具有较好的通用性,使用最为广泛。层次聚类分两种,一种是聚合式(Agglomerative)层次聚类,另外一种是分裂式(Divisive)层次聚类。它们是通过一系列合并或者一系列的分裂来得到聚类结果,两个过程互为逆过程。

这两种方法都产生一棵二叉树,树的顶点是一个包含所有对象的类,叶子节点是单独的对象,中间节点包含了其两个子节点中的所有对象。在层次聚类算法中计算两个类之间相似性度量主要有三种方式,分别为(1)简单连接;(2)完全连接;(3)平均连接。图 4 为层次聚类的一个例子,显然,采用汇聚方式更

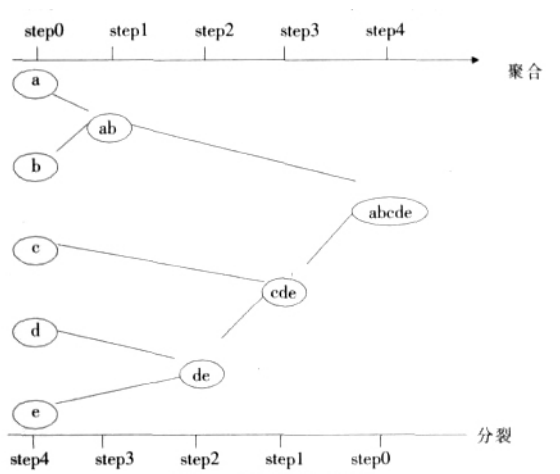

图 4　层次聚类示例图

方便和直观。

作为应用最广的基因表达数据聚类方法,层次聚类是基因表达数据聚类方面事实上的标准。其最大的优点在于所得的结果可以方便地进行可视化观察,更有意义的是它能够表达出基因之间相似度的大小关系。此外,层次聚类的优点还包括:

(1)实现起来相对容易;

(2)运行机制相对简单,易于理解,同时类间融合所要遵循的标准比较清晰。

层次聚类的不足之处在于:

(1)层次聚类产生一树状结构,树枝高度与类间距离成正比,但最后选取某一类间距离的类数作为最终结果,此一步主观性较大,对聚类结果的影响也较强;

(2)不论是聚合式还是分裂式层次聚类都具有较大的时间复杂度,与所分析的表达谱数目的平方成正比,对于较大的基因表达数据集而言是一个大问题;

(3)类的个数必须作为输入量或者随后被估计,一些层次聚类算法存在'链'式效应,导致聚类效果受到影响。

3 新算法的提出与实证应用

Yeung是最早进行聚类评判分析的学者之一,基于他提出的FOM(Figure Of Merit)评价聚类质量的算法,国内有些学者综合了Entropy(信息熵)评价法与FOM算法提出了一种新的称为Entropy.FOM的聚类结果评价方法[5]。图5为6种常见的聚类算法在Ferea的酵母数据集中聚类的评价结果。

从图5中可以清楚的看到,在以上6种聚类算法中,SOM算法具有相对高的聚类效能,而层次聚类的average聚类、single聚类、complete聚类3种算法则相差不大。

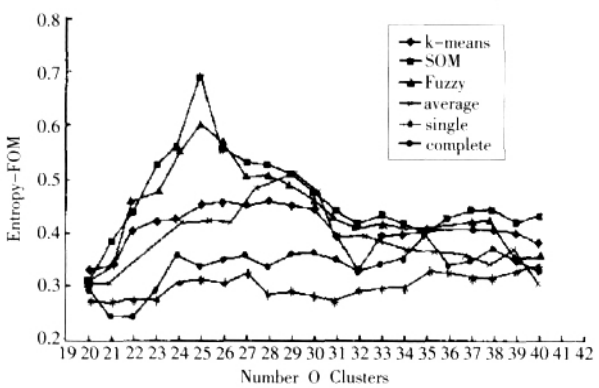

图5 Entropy.FOM聚类结果评价图

因此,基于上述两种算法优劣性的比较,本文提出一种新的算法,即采用SOM算法结合层次聚类的改进算法来对基因表达数据进行聚类。

3.1 新算法的提出

第一步:通过SOM算法对基因表达数据进行聚类,得出一定数量的类;

第二步:将每个类对应的神经元权值作为层次聚类的输入,二次聚类后得出最终结果。

本文将通过酵母菌基因表达数据的聚类例子来检验该算法。其中引用的基因数据得到了美国贝勒医学院蔡蔚文教授的大力帮助,特此感谢。

3.2 实验准备

3.2.1 数据来源与实验目的

实验采用斯坦福大学分子生物系网站上的6 178条酵母基因片段(Yeast Cell Cycle)在cln、clb、alpha、cdc15、cdc28以及elu6个独立实验下的表达[6]。由于该数据中已有已知功能的基因,因此它们的聚类对评价聚类算法在基因表达数据中的效果有直接意义。同时,酵母(Yeast)基因是人类已经全部测出基因序列的真核生物之一,对其分析可以对未来人类基因的分析有尝试性意义。

通过对这些基因在相互独立实验条件下的表达进行聚类,按照相同周期阶段内cDNA表达峰值模式相近或一致的标准,将这些调控基因分为不同的类。同类基因的表达模式相对类间基因更为接近,因而其发挥的调控功能应该是一致的,按照这一思路可知这6 000多条基因是分属于细胞分裂周期S-G2-M-G1等不同阶段的功能基因,通过与同类中的已知基因进行比较,聚类技术有助于发现未知功能的调控基因。

3.2.2 数据预处理

由于实验数据集比较大,需要从6 178条基因中取出已知功能的93条基因,每条基因具有77维属性,通过查阅生物学相关资料,这93条基因分别处于酵母细胞分裂周期的5个阶段:G1(44条)、G2/M边界期(19条)、M/G1边界期(14条)、S(11条)、S/G2边界期(5条)。基因分类情况如表1。

3.3 算法的实现

通过减去平均值后除以标准协方差的方法对实验所选数据进行标准化,同时运用matlab7.0中的isnan()、nanmedian()

表1 93条酵母基因在5个分裂阶段的分类

G1(44)			G2/M(19)		M/G1(14)	S(11)	S/G2(5)
YAR007C	YER095W	YMR199W	YAL040C	YMR001C	YBR067C	YBL002W	YDR150W
YBL035C	YGL163C	YMR307W	YAR018C	YNL145W	YBR083W	YBL003C	YKL096W
YBR088C	YGR044C	YNL082W	YBR054W	YOR058C	YCL027W	YBR009C	YKL096W-A
YDL003W	YGR109C	YNL102W	YBR202W	YPR119W	YCL055W	YBR010W	YLR210W
YDL055C	YJL115W	YNL262W	YDR033W		YDL179W	YDR224C	YMR198W
YDL102W	YJL173C	YNL289W	YDR077W		YER111C	YDR225W	
YDL127W	YJL187C	YNL312W	YDR146C		YJL194W	YGL225W	
YDL164C	YKL042W	YOL090W	YFL026W		YKL185W	YJL092W	
YDL197C	YKL045W	YOR074C	YGL116W		YLR079W	YNL030W	
YDL227C	YKL101W	YPL153C	YGR092W		YLR274W	YNL031C	
YDR097C	YKL113C	YPL256C	YGR108W		YLR452C	YPR159W	
YDR309C	YLR103C	YPR120C	YHR152W		YNL192W		
YDR356W	YLR286C	YPR141C	YIL106W		YNL327W		
YER001W	YLR342W	YPR175W	YJL157C		YNR044W		
YER070W	YML021C		YLR131C				

以及 repmat()函数实现对基因表达数据中空缺值的查找,并用同一实验下所有数据的平均值取代,最后得到的基因表达数据为的矩阵。

在 SOM 算法与层次聚类的结合上,运用 matlab7.0 开发环境中 Toolboxes 中的 Bioinformatics、Neural Network 以及 Statistics3 个工具来实现算法。

其中 SOM 算法网络结构如下:

输入层:93×7 维基因表达数据矩阵。

输出层(竞争层):5×5 神经元。

运用 Neural Network 工具箱中的 newsom()函数建立自组织特征映射网络,train()函数实现对网络的训练,训练次数为 1 000 至 5 000 次,初始学习速率为 0.9。

层次聚类算法的实现:

输入:自组织特征映射网络输出的 25 个神经元权值

输出:采用 3 种相似性度量中的平均连接,输出类数为 5 类。

运用 Statistics 工具箱中的 cluster()函数进行层次聚类。

当网络训练成功,且 25 个神经元已经被分成 5 大类后,将基因表达数据代入 SOM 网络进行仿真,可以观察到每个基因分别映射到竞争层的哪个神经元,从而将其分成 5 类。

3.4　实验结果与分析

3.4.1　自组织特征映射神经网络聚类结果

在进行改进算法实验前,先用 SOM 算法对基因表达数据进行聚类。本文使用的自组织算法包含在美国麻省理工学院开发的 GeneCluster2.0[7]程序中。值得一提的是 GeneCluster 程序已经被全世界成百上千个生物基因和医药研究中心用来分析基因表达数据,包括:Fred Hutchinson 癌症研究中心、斯坦福大学医学院、遗传学协会等,并在一些高水平的出版物中作为主要的分析工具。

将 93×7 维基因表达矩阵作为输入层,输出层 1×5 为的神经元,即直接将 93 个基因分为 5 类,从最后的实验结果中,删除几组较好与较坏的结果,从中等结果中取出 5 组,表 2 为这 5 次实验的结果。

表 2　基于 GeneCluster 的 SOM 聚类正确率分析

实验次数	聚类正确率/%					
	G1(c3)	G2/M(c0)	M/G1(c1)	S(c4)	S/G2(c2)	总正确率
1	100	94.12	76.47	72.73	12.00	67.70
2	100	88.89	76.47	72.73	6.67	60.22
3	100	100.00	76.47	72.73	15.38	68.82
4	100	83.33	68.42	64.71	0.00	64.52
5	100	83.33	72.22	71.43	4.17	62.37
平均值	100	89.93	74.01	70.87	8.00	64.73

图 6 与表 3 分别为其中最好的一次聚类结果输出图和 c2 类结果。

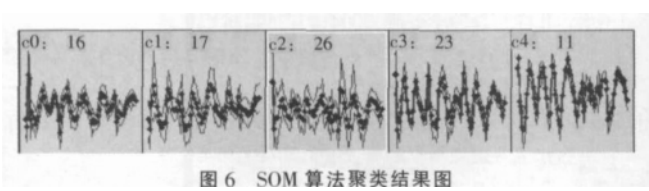

图 6　SOM 算法聚类结果图

从表 2 中可以清楚的看到 SOM 算法对 G1、G2/M、M/G1、S 四个类的聚类正确率比较高,平均达到 70%以上,而且在 5 次实验中聚类的结果变化不大,这也证实了神经网络的稳定性和高效性。另一方面,可以看到在对 S/G2 类基因的聚类结果中出现了较严重的失误,正确率很低。在训练次数较小的第 4 次实验中,该类基因甚至完全被忽略。通过对 5 次实验聚类结果的分析,以及从表 3 中可以发现,S/G2 类的 5 个基因在每次实验过程中几乎都被较大的类(如 G1 类、G2/M 类)"淹没",不能较好的聚成一类。对于 G1 类基因,虽然 5 次实验的正确率均达到 100%,但是每次聚在该类中的基因在 25 个左右,这就是说对这一大类(44 条)基因的缺失很厉害。

表 3　c2 类 26 个基因

基因名称	原本类别	基因名称	原本类别	基因名称	原本类别
YAL040C	G2/M	YGL225W	S	YLR342W	G1
YDL102W	G1	YJL092W	S	YML021C	G1
YDL164C	G1	YJL173C	G1	YMR198W	S/G2
YDL197C	G1	YJL187C	G1	YNL102W	G1
YDL227C	G1	YKL042W	G1	YNL289W	G1
YDR150W	S/G2	YKL096W-A	S/G2	YNL312W	G1
YDR356W	G1	YKL101W	G1	YPR141C	G1
YER111C	M/G1	YKL113C	G1	YPR159W	S
YGL163C	G1	YLR210W	S/G2		

究其原因有两点:(1)选择的神经元作为输出不适合该项基因数据的聚类,因为 G1 类基因数较大,而 S/G2 类基因数较少,对于 5 类的小分类数,小类数据可能被当作"噪声"处理;(2)自组织算法产生了"边界"效应,S/G2 类神经元可能正好处于 G1 类或 G2/M 类神经元的边界上。这两项结果也是造成最后聚类总正确率一般的原因。

3.4.2　改进算法聚类结果

在 matlab7.0 实验环境下,用 SOM 算法结合层次聚类的改进算法对实验数据进行聚类,此次将 SOM 神经网络的输出改为 5×5 的神经元,即最大可分为 25 类,具体类数由算法自行完成,在二次聚类中,用层次聚类把输出结果再聚成 5 类。同样取出实验结果中等的 5 组,得到的结果如表 4。

表 4　改进算法聚类正确率分析

实验次数	聚类正确率/%					
	G1	G2/M	M/G1	S	S/G2	总正确率
1	97.22	93.33	65.00	64.71	80.00	82.80
2	90.70	84.21	76.92	69.23	80.00	83.87
3	92.85	100.00	68.42	76.92	80.00	86.02
4	92.68	90.48	78.57	69.23	66.67	84.95
5	95.12	100.00	65.00	76.92	83.33	86.02
平均值	93.71	93.60	71.00	71.40	78.00	84.73

从表 4 数据中可以比较得出,改进算法具有较高的聚类效能,对 5 个类的聚类效果都比较突出,平均正确率达到 71%以上,总正确率平均高达 84.73%,这对生物学研究可以说是非常高的正确率。

图 7 与表 5 分别为其中最好一次聚类结果输出图和输出结果分析。

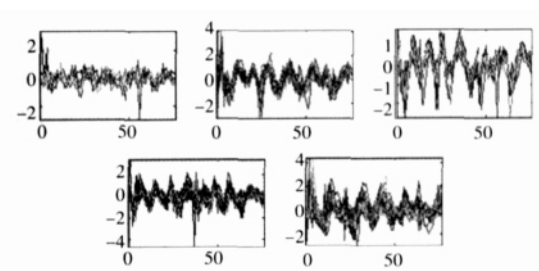

图 7　聚类结果输出图

表 5 输出结果分析表

基因分类	正确数/该类总数(总:80/93)	错误基因	缺失基因
c0(G2/M)	13/13	无	YAL040C、YDR077W、YFL026W、YHR152W、YJL157C、YNL145W
c1(M/G1)	13/20	YDR077W、YFL026W、YGR044C、YHR152W、YJL157C、YLR286C、YNL145W	YER111C
c2(S/G2)	5/6	YAL040C	无
c3(G1)	39/41	YER111C、YPR159W	YDL055C、YGR044C、YLR286C、YLR342W、YMR307W
c4(S)	10/13	YDL055C、YLR342W、YMR307W	YPR159W

同时,利用层次聚类的可视性方面的优越性,从图 8 中可以清楚的看到 25 个自组织的神经元中的距离大小关系,从而为生物学家检测真实的基因功能提供了先后顺序:通过观察类中的基因,可以知道离已知功能越近的未知基因有更大的可能与已知基因有相似的功能,做生物实验的时候可以优先检测这些基因。

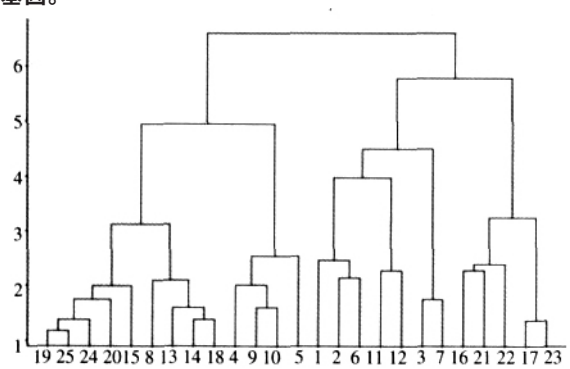

图 8 层次聚类树状图

3.4.3 实验结论

通过比较自组织特征映射神经网络与改进算法的聚类结果,可以发现两种算法在稳定性上相差不大,都具有较高的稳定性,这主要是由于神经网络的自适应特性。在改进算法的实验中,虽然对 G1 类基因的聚类正确率略微有所下降,但基因缺失率却比前者大大减少,同时也保持了较高的正确率。对于 S/G2 类基因,改进算法的效果更是显著,不仅该类基因被独立出来,而且保证了正确率。该类中的 5 个基因得到了完全的聚类。分析原因可能是因为放大了 SOM 输出的神经元网络,从而减小了小类基因被当作'噪声'点的可能,再利用层次聚类进行二次聚类,有效的克服了神经元之间的'边界'效应。这说明改进算法在正确划分小类基因以及减少大类基因缺失率方面比 SOM 算法单独聚类具有更高的效能。

4 总结与展望

随着分子生物学、信息科学的发展,生物数据量空前增长,把数据挖掘技术应用到这些数据的分析中,从而获得生物结构、功能方面的信息,是生物信息学研究的核心目标,也是后基因组计划取得突破的决定性步骤。聚类技术对于分析基因表达数据,推断未知基因的可能功能上具有举足轻重的作用。本文创造性地运用自组织特征映射神经网络结合层次聚类的算法,从神经网络"人工智能"的角度对基因表达数据进行聚类分析,收到了一定的成效。各聚类中的基因都是表达模式相近的,可以预测其功能类似。此外,由于不同的算法都有其一定的适应范围和限制条件,针对某一次基因表达数据的聚类成功与否并不代表了该算法就完全适合或不适合用于基因表达数据的研究。作为 GeneCluster 这样的通用性软件也不能保证对每种基因表达数据的聚类都得到较好的结果。值得一提的是,生物信息学本身只是为生物学的研究提供参考,这些信息能提高研究的效率或提供研究的思路,但问题的最终解决以及可靠性分析还要生物学家通过实验的方法验证。(收稿日期:2006 年 11 月)

参考文献:

[1] Sorin Draghici.Data analysis tools for DNA microarrays[M].[S.l.]:Chapman & Hall/CRC,2003.
[2] 闻新,周露,王丹力,等.Matlab 神经网络应用设计[M].北京:科学出版社,2001.
[3] Tamayo P,Slonim D,Mesirov J,et al.Interpreting gene expression with self-organizing maps:methods and application to hematopoeitic differentiation[C]//Proc Natl Acad Sci,USA,1999,96:2907-2912.
[4] Akinobu Sugiyama,Manabu Kotani.Analysis of gene expression data by using self-organizing maps and k-means clustering [J].IEEE,2002,0-7803-7278-6.
[5] 易东,杨梦苏,李辉智,等.基因表达数据聚类分析结果的评价方法研究[J].中国卫生统计,2002,12(6):332-335.
[6] Reich M,Ohm K,Tamayo P,et al.(2004)GeneCluster 2.0.An advanced toolset for bioarray analysis[J].Bioinformatics,2004.
[7] Naoki Yano,Manabu Kotani.Clutering gene expression data using self-organizing maps and k-means clustering [J].SICE, 2003,PR0001/03/0000-1289.
[8] 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,1999,286:531-537.
[9] 陈京民.数据仓库与数据挖掘技术[M].北京:电子工业出版社,2002.
[10] 黄德双.神经网络模式识别系统理论[M].北京:电子工业出版社,1996.
[11] 胡永钢,须文波.数据挖掘在生物信息学中的应用[J].生物信息学,2004(3):40-42.
[12] 潘金灯,郭腾冲,涂序彦.生物信息学中的智能模型[J].计算机工程与应用,2003,39(28):81-84.
[13] 王长本,刘兴晖.基因表达数据的聚类分析综述[J].国外医学临床生物化学与检验学分册,2004(4):359-362.
[14] 王富刚,陈先农.基因芯片数据的聚类分析[J].国外医学临床生物化学与检验学分册,2004(2):98-101.
[15] 杨春梅,万柏坤,高晓峰.基因表达聚类分析技术的现状与发展[J].生物化学与生物物理进展,2003,30(6):974-979.