

有监督的 Kohonen 神经网络聚类算法在癌症诊断中的应用

马 卫¹, 马全富²

(1 中南民族大学 计算机科学学院, 湖北 武汉 430074;

2 华中科技大学 同济医学院附属同济医院肿瘤生物医学中心, 湖北 武汉 430030)

摘 要: 为了提高癌症诊断识别的正确率, 提出了一种基于有监督的 Kohonen 神经网络的癌症诊断模型. 在原有 Kohonen 神经网络的基础上添加了一个输出层, 使原本无监督的神经网络变成有监督的神经网络 S_Kohonen, 并且网络输出个数与分类个数相同. 实验结果表明, 采用有监督的 Kohonen 神经网络模型能对基因表达水平样本进行有效聚类, 从而提高癌症诊断识别的正确率和科学性, 为癌症的治疗研究提供一定的参考价值.

关键词: 神经网络; 癌症; 诊断

中图分类号: TP3

文献标识码: A

文章编号: 1000-7180(2014)12-0108-03

Application of Supervised Kohonen Neural Network in Cancer Diagnosis

MA Wei¹, MA Quan-fu²

(1 Department of Computer Science, South Central University for Nationalities, Wuhan 430074, China;

2 Molecular Cancer Research Center, Tongji Hospital, Tongji Medical College,

Huazhong University of Science and Technology, Wuhan 430030, China)

Abstract: In order to improve the accuracy of cancer diagnosis, a model is proposed for cancer diagnosis, which based on the supervised Kohonen neural network. Adding an output layer on the basis of the Kohonen neural network, make the unsupervised neural network into supervised neural network- S_Kohonen which has the same network output number as the classification. The experimental results show that the supervised Kohonen neural network model is effective in clustering to the samples of the level of gene expression, so as to improve the accuracy and scientificity of cancer diagnosis, provide the certain reference value for the research on the treatment of cancer.

Key words: Kohonen neural network; cancer; diagnosis

1 引言

对于癌症患者, 早发现早治疗是最行之有效的办法, 但传统的诊断技术和方法不能对癌症进行快速准确的诊断, 漫长的诊断过程往往让患者错过了癌症治疗的最佳时机. 因此, 如何尽早快速而准确地发现癌症, 对癌症患者的治疗有着重要的意义. 而神经网络作为计算机辅助诊断系统的核心技术, 在近几年得到了快速发展, 文献[1-4]中分别讨论了某些神经网络在不同癌症诊断中的应用, 但没有涉及

Kohonen 神经网络. 本文从 Kohonen 神经网络的角度出发, 探讨了一种改进的 Kohonen 神经网络聚类算法在癌症诊断中的应用.

2 Kohonen 神经网络模型

Kohonen 神经网络是芬兰赫尔辛基大学教授 Teuvo Kohonen 提出的一种自组织竞争型网络, 该网络为无监督学习网络, 通过自组织特征映射调整网络权值, 使神经网络收敛于一种表示形态. 在这一形态中, 一个神经元只对某种输入模式特别敏感.

收稿日期: 2014-02-18; **修回日期:** 2014-04-10

基金项目: 国家自然科学基金青年项目(81101571)

Kohonen 神经网络结构为包含输入层和竞争层两层前馈神经网络,神经网络第 1 层为输入层,输入层神经元的个数同输入样本的维数一致,取输入层神经元数为 m . 第 2 层为竞争层,竞争层节点呈二维阵列分布,取竞争层节点数为 n . 输入节点和输出节点之间以可变权值全连接,连接权值为 ω_{ij} ($i=1, 2, \dots, m; j=1, 2, \dots, n$). 如图 1 所示.

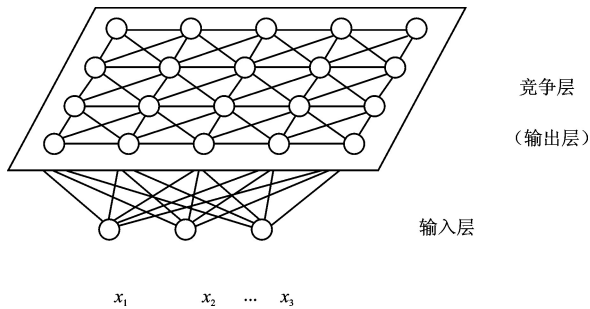


图 1 Kohonen 神经网络拓扑结构示意图

Kohonen 网络的工作原理是:在网络学习过程中,首先输入学习样本,然后通过其自组织竞争的特点,计算输入样本和竞争层神经元权值之间的欧几里得距离,距离最小的神经元为获胜神经元. 接下来进一步调整获胜神经元和其邻域内神经元的权值,使获胜神经元及周边权值更加靠近相应的输入样本. 最终,通过反复训练,各神经元的连接权值的分布结果使具有数据相似性的神经元组织在一起,从而实现数据的有效聚类. 上述过程是系统自主的,在无监督学习条件下完成的.

但由于 Kohonen 神经网络只是起到聚类^[5]的作用,网络收敛速度慢,并且对于快速而准确的分类,效果不佳,因此在进行癌症诊断分类处理时,需要对 Kohonen 神经网络进行改进,实现有监督的学习^[6-7].

3 改进的 Kohonen 神经网络聚类算法

3.1 算法的思想

改进的 Kohonen 神经网络在原有 Kohonen 神经网络的基础上添加了一个输出层,使原本无监督的神经网络变成有监督的神经网络 S_Kohonen,并且网络输出个数与分类个数相同,每个输出节点代表一个数据类别. 输出层节点采用全连接的方式与竞争层的节点相连. 在训练网络时, S_Kohonen 神经网络不仅要调整输入层同竞争层获胜节点的权值,还需要调整竞争层获胜节点同输出层节点之间的权值. 训练完毕的网络可以对未知输入样本实现

有效的聚类,聚类的实现通过计算与未知样本距离最近的竞争层节点作为获胜节点,而与获胜节点连接权值最大的输出层节点所代表类别即为未知输入样本的类别.

3.2 模型的建立

基于 S_Kohonen 神经网络的癌症诊断聚类算法流程如图 2 所示.

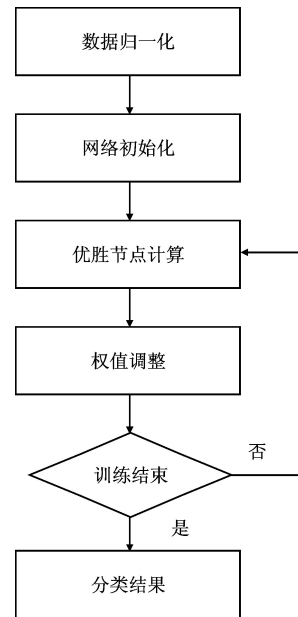


图 2 算法流程

首先对癌症病人的基因表达水平样本进行归一化处理;然后根据数据的特点,确定输入层节点数、竞争层节点数以及输出层节点数,建立训练网络;接下来进行网络训练,计算与输入样本距离最近的节点作为优胜节点,并调整相应的权值,同时也要调整优胜节点与输出节点间的权值;最后,当网络训练结束时,权向量已经趋于稳定,网络能够区别具有不同特征的输入样本,从而实现数据的聚类功能.

3.3 算法的实现

S_Kohonen 神经网络的网络学习训练算法步骤如下:

(1)数据归一化.

(2)网络初始化. 初始化输入层与竞争层之间的网络权值 ω_{ij} 、竞争层与输出层之间的网络权值 ω_{jk} 为 $[0,1]$ 区间的任意随机数,学习速率 $\eta(t)$ 的初始值 $\eta(0)$ 在 $(0,1)$ 区间取值. 确定领域学习半径 r 和总学习次数.

(3)输入训练样本 $X = (x_1, x_2, \dots, x_n)$, 并计算输入样本与竞争层神经元权值之间的欧几里得距离 $d_j, d_j = \sqrt{\sum_{i=1}^m (x_i - \omega_{ij})^2} \mid j = 1, 2, \dots, n$.

(4) 寻找获胜神经元. 通过计算所得的欧式距离, 选取距离最小的竞争层神经元 c 作为最佳匹配的获胜神经元.

(5) 权值调整. 调整输入层同竞争层获胜神经元领域内节点权值, 调整公式为式(1):

$$\omega_{ij} = \omega_{ij} + \eta_i (X_i - \omega_{ij}) \quad (1)$$

调整竞争层获胜神经元领域内节点同输出层节点权值, 调整公式为式(2):

$$\omega_{jk} = \omega_{jk} + \eta_p (Y_k - \omega_{jk}) \quad (2)$$

式中, η_i 、 η_p 为学习速率, 一般随学习次数的增加而线性下降, Y_k 为样本所属类别.

(6) 判断算法是否结束, 若没有结束, 则返回步骤(2).

4 实验结果分析

为验证本算法的性能, 实验在 matlab R2010b 环境中利用神经网络工具箱创建网络模型. 实验数据采取包含 84 个个体基因表达水平的样本. 其中有 30 个是正常人的基因表达水平样本, 有 30 个是癌症病人的基因表达水平样本, 剩下的 24 个是待检测的未知的基因表达水平样本, 每个样本中测量了 110 个基因特征.

实验 1: 将 30 个正常人的基因数据分别作为 kohonen 网络和 s_kohonen 网络的输入数据, 对网络进行训练, 当训练结束时, 两种网络所得到的结果分类和实际分类的对比如图 3 所示. 测试样本的分类结果对应着输出神经元的编号, 1 表示癌症类别, 0 表示正常类别. 从实验结果可以看出, 对于正常人的基因数据分类, s_kohonen 神经网络的精确度更高, 达到了 93.33%.

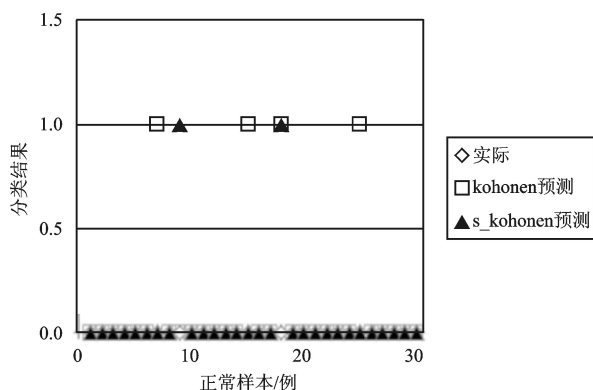


图 3 正常样本分类结果比较

实验 2: 将 30 个癌症病人的基因数据分别作为 kohonen 网络和 s_kohonen 网络的输入数据, 对网

络进行训练, 当训练结束时, 两种网络所得到的结果分类和实际分类的对比如图 4 所示. 从图中可以看出, 对于癌症基因的有效分类, s_kohonen 网络的诊断精确度也略高于 kohonen 网络.

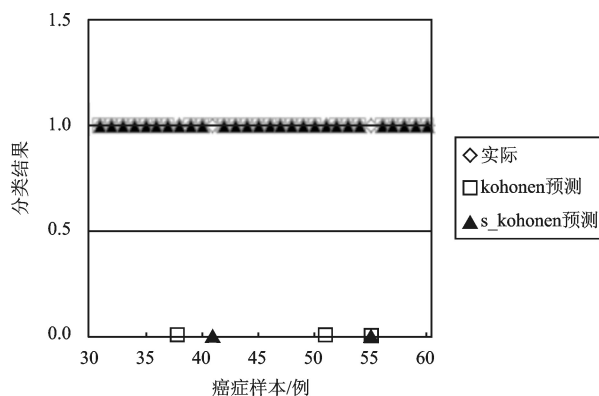


图 4 癌症样本分类结果比较

实验 3: 对 24 个未知的待检测基因样本在已经训练好的 s_kohonen 网络中进行预测, 分类结果如表 1 所示.

表 1 待检测样本数据的分类结果

编号	1	2	3	4	5	6	7	8
类别	1	0	1	0	1	1	1	1
编号	9	10	11	12	13	14	15	16
类别	1	0	1	1	1	1	1	1
编号	17	18	19	20	21	22	23	24
类别	0	1	1	1	1	0	1	1

5 结束语

在进行癌症诊断时, 一般的诊断方法主要是医生根据自身的专业知识和临床经验, 对患者的检查结果进行主观判断, 存在着较多的主观性, 易出现误诊现象, 诊断结果不完全可信. 本文将改进后的 kohonen 神经网络模型用于癌症诊断, 从实验结果来看, 改进后的 s_kohonen 神经网络模型比 kohonen 神经网络诊断精确度更好, 具有一定的医学应用价值.

参考文献:

- [1] 吕冬姣, 张珏, 王霄英, 等. 人工神经网络在前列腺癌诊断中的应用[J]. 北京大学学报: 医学版, 2009, 41(4): 469-472.
- [2] 王波, 杜晓昕, 金梅. LVQ 神经网络在乳腺肿瘤诊断中的应用[J]. 计算机仿真, 2012, 29(8): 171-174.
- [3] 高阳, 周根宝. 基于 T 检验分析的 SOM 神经网络算法在癌症诊断中的应用[J]. 内蒙古农业大学学报, 2013, 34(2): 142-146.

(下转第 113 页)

间也有所减少。

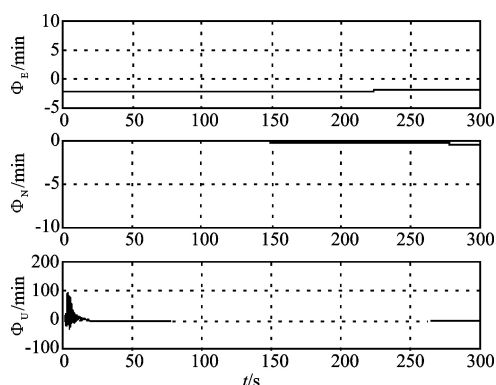


图4 小波滤波后初始对准姿态误差

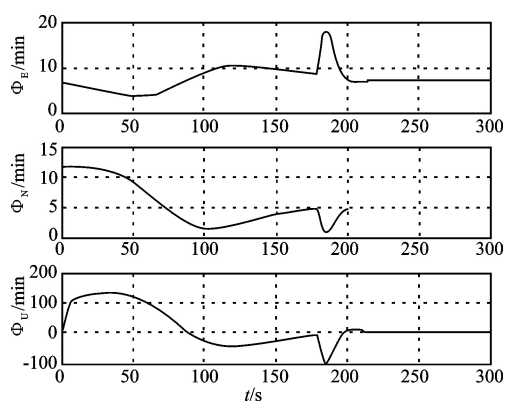


图5 未经滤波处理的初始对准姿态误差

6 结束语

在对小波分析的理论进行深入的研究的基础上,将多种小波滤波方法应用于光纤陀螺随机漂移数据的处理中,通过滤波效果与小波阈值滤波方法相结合的方法,最终选定出在工程应用中最佳的小波滤波方法,能够最大程度上抑制陀螺的随机漂移。应用 Allan 方差分析方法对该小波进行性能评估,验证了此方法的正确性,并将此最佳小波滤波方法应用于初始对准试验中,并对试验结果进行分析,得出有效结论,再次验证了最佳小波滤波方法的先进

性,具有较高的工程意义。

致谢 感谢西北工业大学张科教授给予的指导和帮助。

参考文献:

- [1] 王继成. 小波分析在信号去噪中的应用[J]. 数学的实践与认识, 2008, 38(23): 134-136.
- [2] 张涛, 徐晓苏. 基于小波和人工智能的舰载捷联系统动基座对准[J]. 吉林大学学报, 2010, 40(2): 549-554.
- [3] 董绪荣, 张守信, 华仲春. GPS/SINS 组合导航定位及其应用[M]. 长沙: 国防科技大学出版社, 1998.
- [4] 王雅静, 申晋, 郑刚. 基于多尺度变换的动态光散射粒径反演范围的自适应调整[J]. 强激光与粒子束, 2010, 22(12): 2979-2983.
- [5] 张立保, 敖鹏亮. 基于小波变换的图像混合噪声自适应滤除算法[J]. 强激光与粒子束, 2010, 22(11): 2540-2544.
- [6] 邓义君, 严高师. 一种基于小波变换的红外小目标去噪算法[J]. 强激光与粒子束, 2007, 22(3): 399-402.
- [7] 于旭东, 张鹏飞, 谢元平. 单轴旋转惯导系统轴向陀螺漂移预测方法[J]. 强激光与粒子束, 2013, 22(4): 399-402.
- [8] Donohodl. De-noising by soft-thresholding[J]. IEEE Transactions on Information Theory, 2005, 41(3): 613-626.
- [9] Daubechies. Orthogonal bases of compactly supported wavelets[J]. Comm Pure Appl Math, 1998(41): 909-996.
- [10] 徐正光, 沈晓冬, 张利欣. 基于小波变换和神经网络集成的人脸表情识别[J]. 微电子学与计算机, 2006(7): 143-146.

作者简介:

白俊卿 女, (1983-), 博士生. 研究方向为多传感器组合导航系统信息融合技术, 车载定位定向技术等方面的研究. E-mail: baihua-2001@163.com.

(上接第 110 页)

- [4] 袁前飞. 基于支持向量机的癌症诊断研究[D]. 重庆: 重庆大学, 2007.
- [5] 葛彦强, 汪向征, 于江德. 一种利用自组织映射和径向基函数神经网络的网络拥塞预测方法[J]. 微电子学与计算机, 2012, 29(12): 176-179.
- [6] 吴柯, 方强, 张俊玲, 等. 基于改进 Kohonen 神经网络的遥感影像分类[J]. 测绘信息与工程, 2007, 32(2): 47-49.

- [7] 雷璐宁, 石为人, 范敏. 基于改进的 SOM 神经网络在水质评价分析中的应用[J]. 仪器仪表学报, 2009, 30(11): 2380-2383.

作者简介:

马卫 女, (1980-), 研究生, 讲师. 研究方向为数据挖掘与网络安全. E-mail: mw0626@163.com.

马全富 男, (1978-), 博士, 研究方向为妇科肿瘤防治研究.