

人工神经网络在染色体自动分析系统中的应用☆

闫文忠¹, 封孝辉²

Application of artificial neural network in chromosome automatic analysis system

Yan Wen-zhong¹, Feng Xiao-hui²

Abstract

BACKGROUND: The classification and identification of human chromosome is a basic mission of medical genetics. Analyzing and identifying human chromosome automatically with computer technology is an important research subject in human chromosome image analysis.

OBJECTIVE: To introduce basic principle, technique advantages and operating methods of artificial neural network and explore the applications of artificial neural network in chromosome automatic analysis.

METHODS: A computer-based online search of EBSCO database (<http://search.ebscohost.com>) and Wanfang database (<http://www.wanfangdata.com.cn>) was performed for articles about application of artificial neural network in chromosome automatic analysis system with the key words "chromosome, artificial neural network" in English and Chinese.

RESULTS AND CONCLUSION: The aim of studying chromosome automatic analysis is to reduce technician labor intensity and allow them free from repeated laboring. Finally, these systems are used in clinic for molecular cytogenetics identification and aristogenesis etc. Despite the significant research effort and progress of applications of artificial neural network in chromosome automatic analysis over these years, there are still some localizations. The classified chromosome database is needed, which is not easy to get for general researchers. In addition, the accuracy of the classification result is not good enough, which even cannot reach the level of well-trained cytology researchers. Moreover, the neural network has its inherent shortcomings such as huge training data and huge training time. Therefore, further research and development is required to improve and consummate the network through optimizing network structure, selecting effective characteristic, and reducing otiose operation.

Yan WZ, Feng XH. Application of artificial neural network in chromosome automatic analysis system. Zhongguo Zuzhi Gongcheng Yanjiu yu Linchuang Kangfu. 2010;14(39):7371-7373. [<http://www.crter.cn> <http://en.zglckf.com>]

摘要

背景: 人体染色体的分类与识别是医学遗传学中的一项目基本任务,应用计算机技术实现人体染色体自动分析与识别是人体染色体图像分析技术的重要研究课题。

目的: 介绍人工神经网络的基本原理、技术优势、应用方法,研究人工神经网络在染色体自动分析系统中应用。

方法: 由第一作者检索 1994/2009 EBSCO 数据 (<http://search.ebscohost.com>) 及万方数据库 (<http://www.wanfangdata.com.cn>) 有关人工神经网络在染色体自动分析系统中应用方面的文献,英文检索词为 "chromosome, artificial neural network", 中文检索词为 "染色体, 人工神经网络"。

结果与结论: 研究染色体自动分析系统的目的,就是要减轻技术人员的劳动强度,使他们从繁琐的重复劳动中解放出来,并最终将这些系统应用于临床,进行肿瘤患者的细胞遗传学鉴定、优生优育的检查等工作。尽管神经网络在染色体自动分析系统中的应用已经经过了多年的发展与完善,但仍存在一定的局限性:①它首先需要一套准确无误的已分类染色体数据库,这对一般研究人员来说是不易得到的。②分类的结果不够准确,甚至达不到训练有素的细胞学研究者的水平。③神经网络有它固有的缺点——训练数据量庞大、训练时间长。因此在今后的研究中,应该从优化神经网络的结构、提取有效特征、减少不必要的运算着手,继续改进和完善这个网络。

关键词: 人工神经网络; 染色体自动分析; 分类; 数字化神经医学; 综述文献

doi:10.3969/j.issn.1673-8225.2010.39.037

闫文忠, 封孝辉. 人工神经网络在染色体自动分析系统中的应用[J]. 中国组织工程研究与临床康复, 2010, 14(39):7371-7373. [<http://www.crter.org> <http://cn.zglckf.com>]

0 引言

染色体是遗传物质的载体,它上面带有遗传因子,决定人体的形态特征和生理功能。对于各种生物,其细胞核内染色体不但有一定数目还有一定形状,如果由于某种原因(如突然变异、放射线、化学等污染影响),使染色体的数目和形状发生异常,就会产生多种遗传性疾病。这些疾病都是与染色体变异有关的,因此通过

对染色体的分析来诊断疾病得到了越来越多的重视^[1-2]。染色体自动分析系统借助于计算机,应用模式识别和图像处理的理论与技术,对光学显微镜输入的染色体图像作定量计算与分析,然后对染色体进行分类,并配对产生核型图或者发现其中变异的染色体^[3-4]。

目前,随着神经计算学的发展,人工神经网络所具有的模式识别功能及其在医学图像分析中的应用受到了越来越多的关注^[5]。1993年, Jannings 等^[6]用多层感知器神经网络(MLP)后向

¹Department of Computer,
²Department of Mechanical and Electrical Engineering, North China Institute of Science and Technology, Beijing 101601, China

Yan Wen-zhong☆, Doctor, Lecturer, Department of Computer, North China Institute of Science and Technology, Beijing 101601, China yanwenzhong@ncist.edu.cn

Received: 2010-03-20
Accepted: 2010-04-21

华北科技学院,
¹计算机系,²机电工程系,北京市 101601

闫文忠☆,男,1976年生,河南省焦作市人,汉族,2008年北京邮电大学毕业,博士,讲师,主要从事数字图像处理研究。 yanwenzhong@ncist.edu.cn

中图分类号:R318
文献标识码:A
文章编号:1673-8225 (2010)39-07371-03

收稿日期:2010-03-20
修回日期:2010-04-21
(20100319010/GW·A)

传播算法(BP)和 Kohonen 自组织特征映射训练网络, 其中 MLP/BP 网络工作较好, 然而, 由于 MLP/BP 网络的训练需要很长时间而一直没被重视。1994 年, Sweeney 等^[7]把概率神经网络用于染色体自动解析取得了令人满意的结果。在此基础上, Lerner 等^[8-9]1992/1999 一直致力于神经网络应用于染色体自动分析系统的研究, 使这项工作日趋完善。可见, 神经网络在染色体自动分析系统中的应用得到了越来越多的重视。本文对此作了综述, 介绍了神经网络在染色体自动分析系统中的应用, 并着重介绍了神经网络在染色体分类中的应用。

1 资料和方法

1.1 资料来源 由第一作者检索 1994/2009 EBSCO 数据 (<http://search.ebscohost.com>) 及万方数据库 (<http://www.wanfangdata.com.cn>) 有关神经网络在染色体自动分析系统中应用方面的文献, 英文检索词为 “chromosome, artificial neural network”, 中文检索词为 “染色体, 人工神经网络”。

1.2 纳入及排除标准

纳入标准: ① 文章所述内容需与神经网络及染色体自动分析系统等方面的研究密切相关。② 同一领域选择近期发表或在权威杂志上发表的文章。

排除标准: 重复性研究。

1.3 数据的提取 计算机初检得到 41 篇文献, 阅读标题和摘要进行初筛, 排除因研究目的与本文无关及内容重复的研究 21 篇, 共保留其中的 20 篇归纳总结。

1.4 质量评估 符合纳入标准的 20 篇文献中, 文献 [1-14] 探讨了神经网络在染色体自动分析系统中的基本应用, 文献 [15-20] 探讨了神经网络在染色体分类中的应用。

2 结果

2.1 神经网络概述 神经网络(简称神经网络)是由大量简单的信息处理单元(神经元)广泛连接而成的复杂网络, 用来模拟人脑神经系统的结构和功能, 是一种简化的人脑数学模型。随着计算机、人工智能尤其是专家系统的发展, 神经网络逐渐成熟并开始用于解决一系列的实际问题。

神经网络主要由处理单元、网络拓扑结构、训练规则组成。处理单元是神经网络的基本操作单元, 用以模拟人脑神经元的功能。一个处理单元有多个输入、输出路径, 输入端模拟脑神经的树突功能, 起信息传递作用; 输出端模拟脑神经的轴突功能, 将处理后的信息传给下一个处理单元(如图 1 所示)。

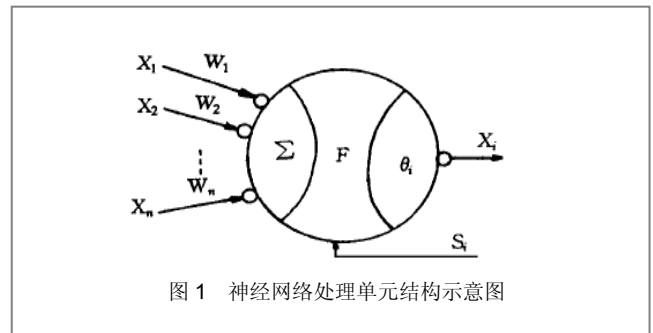


图 1 神经网络处理单元结构示意图

具有相同功能的处理单元构成处理层。常用的多层感知机由输入层、输出层和若干隐含层组成(如图 2 所示), 神经网络的拓扑结构决定了各处理单元、各层之间信息的传递方式与途径。训练规则利用转换函数 $f(x)$ 进行数据处理的加权及求和, 训练网络系统进行模式识别, 并将其转换成潜在的输出值^[10]。

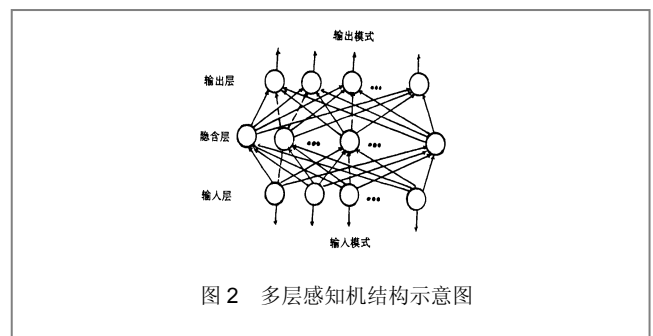


图 2 多层感知机结构示意图

2.2 神经网络在染色体自动分析系统的应用 神经网络具有自适应性、并行处理的能力和非线性处理的特性, 允许通过网络训练应用专家知识和经验, 使基于神经网络的染色体自动分析成为可能。1995 年, Lerner 等^[8]研究了基于中轴转换的人类染色体分类神经网络, 该网络通过提取染色体中轴的特征运用多层感知器对染色体进行分类, 取得较好的效果。

1997 年, Lerner 等^[11]又提出了用 Sammon 非线性映射算法运用非线性变换, 把原始特征映射成较少的特征, 这个过程试图在模式从高维空间到低维空间映射时保持固有数据结构。而这种固有结构的保持则通过保持映射下模式间的距离实现。在此基础上, Lerner^[12]1998 年又优化了该系统, 使之成为了“完全自动的基于神经网络的人类染色体分析系统”, 主要对染色体图像预处理, 解决相接触染色体的分离问题。此外, Musavi 等^[13-14]还把神经网络的方法应用于老鼠染色体分类, 由于老鼠染色体是端着丝粒型染色体, 不能像人类染色体一样把着丝粒位置作为分类的重要指标, 所以识别老鼠染色体主要依赖于带型。这里使用的是快速正交算法训练的径向基本函数神经网络分类器 (RBF/FOS), 实验证明该网络用于老鼠染色体分类时, 取得了良好的效果。

2.3 神经网络在染色体分类中的应用 为了提高染色体

自动分类的准确性, 人工智能和机器学习方法广泛应用于染色体的分类中。其中, 神经网络由于具有可以在对问题了解较少情况下模拟人脑作决定的能力, 以及简单的结构和训练过程比较容易等优点, 而成为最受欢迎的工具。许多基于特征和象素值分类的神经网络应用到了染色体的分类中, 包括有监督型的神经网络和无监督型的神经网络^[15]。BP 网络广泛应用于神经网络的训练当中。在多层前向网络中, 输出神经元的数目通常是固定的(从 1 到 24), 但是输入神经元的数目、隐含神经元的数目、学习速度、循环学习的次数等都是可编程控制的。这些参数的确定对神经网络的实现和健壮性具有非常重要的影响。研究表明, 在图像质量很高的情况下, 经过 BP 网络的训练, 一个简单的神经网络就可以对染色体进行分类。但是, 早期的研究也表明神经网络得到的结果比用简单的统计方法(如最大似然法)得到的结果要差一些。同时, 在神经网络充分优化的过程中, 如果训练过度会降低神经网络的健壮性^[16]。

在一项研究中, 使用基于多层感知器的神经网络对训练数据库得到的错误率为 0, 对测试数据库得到的错误率为 24.2%^[14]。为了提高神经网络的准确率, 另一个研究得到了如下结论: 通过减小神经网络的大小可以将测试的准确率从 75.8% 提高到 88.3%^[17]。

为了提高传统多层神经网络的性能, 许多其他神经网络被提出并应用于这个领域。一种模糊 Hopfield 神经网络由神经元模型和模糊算法联合组成, 它和传统多层神经网络的主要区别在于它具有模糊聚类和学习机制, 可以从有噪声的训练样本中得到目标(染色体)的相关信息。该网络用来识别人体染色体并将其中的每条染色体进行归类。通过对 100 条染色体的检验, 模糊 Hopfield 神经网络取得的识别率高达 96.67%^[18]。

郭宏宇等^[19]将一类模糊神经网络引入染色体自动分析系统中, 采用模糊前向网络, 而算法仍然采用反向传播(BP)算法。实验证明该方法对许多特征相似的染色体具有很好的识别效果。

另外, 一些无监督非线性的学习方法也用于测试和评估神经网络的最优化。但是, 研究发现无监督非线性方法的结果要比有监督非线性方法的结果差。

阮晓钢^[5]提出了一种基于自组织径向基人工神经网络的染色体模式分类的方法, 该方法将染色体长度、着丝粒位置、带纹分布等特征作为自组织径向基网络的输入, 利用自组织径向基网络的混合学习算法对网络进行训练, 使人工神经网络对染色体具有了自动分类的能力。虽然在模式识别和染色体的分类中, 神经网络是一种非常强大的机器学习工具, 但它还是具有自身的一些弱点。首先, 当复杂医学图像所描述的信息不够准确时, 神经网络的健壮性比较弱。第二, 由于在最优化时使用的方法是“黑盒子”方法, 使用者无法得知

其运行机制, 所以这是神经网络的一个主要缺点^[20]。

3 讨论

研究染色体自动分析系统的目的, 就是要减轻技术人员劳动强度, 使他们从繁琐的重复劳动中解放出来, 并最终将这些系统应用于临床, 进行肿瘤患者的细胞遗传学鉴定、优生优育的检查等工作。尽管神经网络在染色体自动分析系统中的应用已经经过了多年的发展与完善, 但仍存在一定的局限性: ①它首先需要一套准确无误的已分类染色体数据库, 这对一般研究人员来说是不易得到的。②分类的结果不够准确, 甚至达不到训练有素的细胞学研究者的水平。③神经网络有它固有的缺点—训练数据量庞大、训练时间长。因此在今后的研究中, 应该从优化神经网络的结构、提取有效特征、减少不必要的运算着手, 继续改进和完善这个网络^[3]。

4 参考文献

- [1] 叶志生, 胡晓飞, 高慧敏, 等. 人类染色体自动解析系统[J]. 中国医疗器械杂志, 2001, 25(1): 16-25.
- [2] 闫文忠. 染色体自动分析系统的研究现状及未来目标[J]. 中国组织工程研究与临床康复, 2009, 13(13): 2544-2546.
- [3] 张斌, 赵卫东, 马世雄, 等. 人类显带染色体的图象分析与识别系统[J]. 中国图像图形学报, 1996, 1(1): 428-432.
- [4] 闫文忠, 沈树群. 染色体自动分析系统关键技术研究[J]. 仪器仪表学报, 2007, 28(4): 531-533.
- [5] 阮晓钢. 基于自组织径向基网络的染色体分类[J]. 北京工业大学学报, 1999, 25(2): 81-86.
- [6] Jennings M, Graham J. A neural network approach to automatic chromosome classification. Medical Biology. 1993; 38: 959-970.
- [7] Sweeney WP, Musavi MT, Guidi JN. Classification of chromosomes using a probabilistic neural network. Cytometry. 1994; 16: 16-24.
- [8] Lerner B, Guterman H, Dinstein I, et al. Medial axis transform-based features and a neural network for human chromosome classification. Pattern Recognition. 1995; 28(11): 1673-1683.
- [9] Lerner B, Guterman H, Dinstein I. A classification-driven partially occluded object segmentation(CPOOS) method with application to chromosome analysis. IEEE Trans Signal Processing. 1998; 46(10): 2841-2847.
- [10] 江东, 王建华. 神经网络在遥感中的应用与发展[J]. 国土资源遥感, 1999, 11(2): 12-18.
- [11]] Lerner B, Rosenberg B, Levinstein M. Media axis transform based features and a neural network for human chromosome classification. World Congress on Neural Networks, W CNN' 94, San Diego, 1994: 173-178.
- [12] Lerner B. Toward a completely automatic neural-network-based human chromosome analysis. IEEE Trans Syst Man Cybern-B: Cybern. 1998; 28: 544-552.
- [13] 叶志生, 胡晓飞, 高慧敏, 等. 神经网络在染色体自动解析系统中的应用[J]. 国外医学: 生物医学工程分册, 2000, 23(4): 229-233.
- [14] Musavi MT, Bryant RJ, Qiao M. Mouse chromosome classification by radial basis function network with fast orthogonal search. Neural Networks. 1998; 11: 769-777.
- [15] Mitchell TM. 1997 Machine Learning (Boston, MA: WCB/McGraw-Hill). Boston, MA: WCB-McGraw-Hill. 1997.
- [16] Cho J. Chromosome classification using backpropagation neural networks. IEEE Eng Med Biol Mag. 2000; 19: 1928-1933.
- [17] Delshadpour S. Reduced size multi-layer perceptron neural network for human chromosome classification Proc. 25th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, 2003: 2249-2252.
- [18] Ruan X. A classifier with the fuzzy Hopfield network for human chromosomes, intelligent control and automation Proc. 3rd World Congress on Intelligent Control and Automation, 2000: 1159-1164.
- [19] 郭宏宇, 鲍旭东, 蒋春涛. 基于模糊人工神经网络的染色体识别[J]. 中国生物医学工程学报, 2004, 23(2): 116-120.
- [20] Wang XW, Zheng B, Wood M, et al. Development and Evaluation of Automated Systems for Detection and Classification of Banded Chromosomes: Current Status and Future Perspectives. J Phys D: applied Phys. 2005; 38(15): 2536-2542.