

# 基于大数据深度学习的中医“证候”到“方剂”的新型算法研究

谢天宇,曹继忠,赵姝婷,温川飙\*

(成都中医药大学,四川 成都 610075)

**摘要:**目的:探索中医师类似于近代控制论“黑箱原理”的辨证论治、选法定方的思维过程。方法:采用一种学习深层非线性网络结构,利用挖掘出的信息建立的数据仓库、训练模型,确定新型算法中更为高效的网络权重。结果:设计出了“证候”到治法“方剂”的具有复杂线性关系与大数据等特点的新型算法,实现“证候”到治法“方剂”的相对准确对应。结论:新型算法的研究有望在现有条件下取得93%以上的选法定方准确率,达到临床和科研对计算机外脑论治的要求,来实现“证候”到“方剂”的准确对应,为中医客观化探索一种新的数据分析和研究方法,以规范中医论治过程。

**关键词:**中医药大数据;中医证候;中药方剂;新型算法

中图分类号:R22

文献标识码:A

文章编号:1673-2197(2018)01-0051-02

DOI:10.11954/ytctyy.201801017

证候是疾病发生和演变过程中某阶段以及患者个体当时所处特定内外环境本质的反映,是一个非线性的“内实外虚”“动态时空”和“多维界面”的复杂巨系统,包括“证”与“候”两个方面。证候一旦确定,治法及方剂也将相对应,但是我国中药资源种类繁多,中药间的相互配伍组成方剂,使“证型”与“方剂”的相互对应变得极其复杂,数据量也相当庞大。人工神经网络具有非线性映射特性<sup>[1]</sup>,利用网络自动学习能力,充分辨识信息,将信息分布于网络联接中,模拟证候模型挖掘其内在规律,采用人工神经网络技术,通过网络训练和学习,从样本中对证候要素的自动提取,建立一种非线性的映射关系,模拟证候的诊断模型,并将其运用于对新病例的诊断。但是鉴于证候的研究多采用BP算法,缺乏学习标准,只能用证候的原始数据学习训练网络,依靠网络的自学习能力,来判断未知的证候类型,学习速度慢、诊断准确率不高,限制了BP网络的应用。样本含量少,信息也不全面,在一定程度上影响准确率。因此,我们采用一种学习深层非线性网络结构,利用挖掘出的信息建立的数据仓库、训练模型,确定新型算法中更为高效的网络权重,设计出“证候”到治法“方剂”的具有复杂线性关系与大数据等特点的新型算法,实现“证候”到治法“方剂”的相对准确对应。本文总结了人工神经网络在证候研究中的应用现状,新型算法的研究方法、创新之处以及应用前景。

## 1 人工神经网络在证候研究中的应用现状

证候研究中的人工神经网络技术大多采用的是反向传播神经网络、径向基(RBF)神经网络、贝叶斯网络。骆文斌等<sup>[2]</sup>通过优化神经网络参数,以病位、病性证素诊断为核心,探讨其用于肺系证诊断研究的可行性。孙贵香等<sup>[3]</sup>采用矩阵实验室(MATLAB)神经网络对496例冠心病患者信息构建冠心病证候模型,准确率可达90.5%,对于新采集的132例冠心病患者诊断,准确率可达到91.4%,说明它可以获取证候的内在规则及规律,然后用于证候诊断。温宗良等<sup>[4]</sup>运用人工神经网络共轭梯度算法从60例患者样本中获取规则,建立高血压证候诊断模型,并运用检测样本对模型进行评价,识别率达89.5%,对19例新患者进行证候诊断。岳彩青等<sup>[5]</sup>利用聚类分析和RBF网络相结合的方法,建立证候诊断模型,用测试样本检验其证候诊断的有效性,拓宽了径向基神经网络的应用。李建生等<sup>[6]</sup>用聚类分析的方法以及(RBF)神经网络,对2型糖尿病病人的资料进行了分析,把样本作为学习训练样本,构建了关于证的模型,一部分样本为检验样本的证候模型进行检验,诊断的准确率达到94.4%,验证了其用于中医证候诊断的有效性,优化了证候诊断模型,扩大了RBF神经网络的应用范围,为证候的研究提出了新思维和新方法。但是,人工神经网络

收稿日期:2017-04-06

基金项目:国家中医药管理局中医药健康服务信息化项目(SATCM-2016-JKFW-XXH05);四川省教育厅自然科学一般项目(17ZB0147)

作者简介:谢天宇(1993—),男,成都中医药大学硕士研究生,研究方向为针灸推拿、中药数字化。

通讯作者:温川飙(1970—),男,成都中医药大学副教授,硕士生导师,研究方向为中医药数字化。E-mail:wcb@cdutcm.edu.cn

技术的不足之处是:算法易于停止在局部最优解,算法的培训时间太长时,过度拟合状况出现,限制了人工神经网络的应用。

## 2 中医“证候”到“方剂”的新型算法研究方案及关键技术

### 2.1 新型算法研究方案

建立中医优势病“证”“法”“方”数据仓库 通过文献挖掘和采集临床真实数据,以疾病为单位整理出中医优势病种的“证型”与相应的“治法”“方剂”,进一步进行数据预处理和数据的挖掘,再建立深度学习数据仓库。

运用受限玻尔兹曼机(restricted Boltzmann Machines, RBM)神经网络设计中医“论治”深度学习模型。以受限玻尔兹曼机神经网络作为单层网络的基本构架,分层训练来实现深层网络的训练。深度学习主要分为三个环节:用无监督方式进行训练系统、调准以及测试。

开展深度学习算法训练,并运用 BP 神经网络进一步验证算法,调整优化网络权重。用大量文献挖掘数据和临床的真实数据对深度学习算法进行训练,每次用无监督训练一层,将训练的结果作为更高一层的输入,逐层的提取学习特征。分析并学习各层之间的相关性,初步拟定各层网络的权重。运用 BP 神经网络,采用反向传播学习算法,从输出层经过各个中间层逐层的修正各连接权值,再回到输入层,此过程反复交替的进行,直到网络全局误差趋于给定的极小值,完成其学习过程。

### 2.2 新型算法研究的关键技术

2.2.1 关键技术路线 新型算法研究的关键技术路线见图 1。

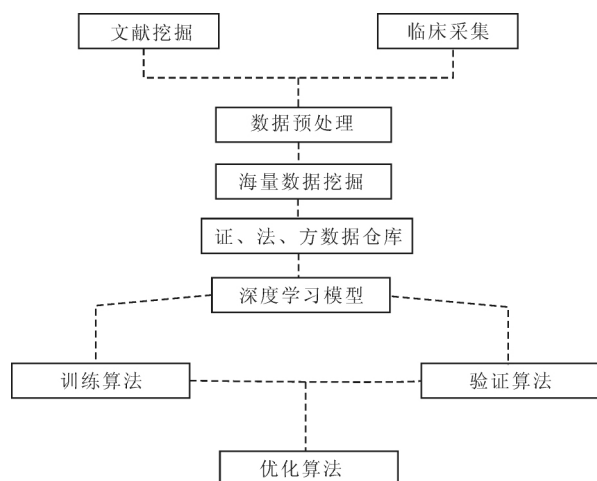


图 1 新型算法研究的关键技术路线

2.2.2 关键技术 (1) 中医优势病种“证、法、方”人工神经网络模型的设计。① 隐含层的确定。根据中医“因、位、性、势”理论,把“因位性势”“治法”作为隐含层,这种双隐含层可以提高“证”到“方”的匹配准确度。② 隐含层的节点数确定,神经网络通常需要大量的彼此连接的神经元,研究表明,神经网络的隐含层节点数目越多,网络对函数的逼近能力就越好,但是从神经网络的学习速度来看,隐含层节点数不宜过多。本项目拟采用基于灰关联—灵敏度的隐含层结

构优化算法,以使整体网络的实际输出值与期望输出值之间的误差最小。③ 深度学习算法的确定:本项目拟采用改进的 BP 算法,一共轭梯度 BP 算法,这种深度学习算法可以极大地提高神经网络的训练速度和收敛性能。(2) 进一步调整优化深度学习的网络权重。共轭梯度 BP 神经网络有较强的非线性映射能力,实现了从输入到输出的映射,神经网络能够以任意的精度逼近任意的非线性连续函数。使其特别适合求解内部机制复杂的问题。这种网络有高度的自学习、自适应能力:在训练时,能通过学习自动提取输入、输出数据间的规则,通过自适应的方式将学习到的内容记忆在网络的权值中。

### 3 新型算法的特色与创新

将中医“因、位、性、势”理论与深度学习技术相融合。算法在深度学习技术逐层提取和学习数据特征的基础上,在“证”与“法”之间根据中医辨证论治理论设置“因、位、性、势”的隐含层,使各层之间的关联更为紧密,更有针对性的提取、学习相应的关键特征。

使用大量临床真实数据训练算法,并运用共轭梯度 BP 神经网络验证算法修正权值。前期近百万条临床真实数据进行算法训练,提供了数据支撑,保证训练得出的参数真实有效,结果可靠。在得出初步的权值后,运用共轭梯度 BP 神经网络,采用反向传播学习算法,从输入层到各中间层逐层的修正各连接之间的权值,直到输出层,以验证算法,得出最优化网络权重。

改进神经网络深度学习模型预测的结果准确性提高。以改进的神经网络深度学习模型为新算法的基本架构,利用挖掘出的信息建立的数据仓库、训练模型,确定新型算法中更为高效准确的网络权重,提高“证”到“方”的匹配准确度。根据预测结果核查中医治疗过程是否规范合理,检出误治、失治病例,从而提高临床疗效,使中医治疗过程实现规范化与标准化。

### 4 研究成果应用

以四川省中医药管理局的“中医数字化诊疗平台”门诊电子病历数据为基础,目前该平台的电子病历数据来源于德阳、绵阳、成都等市以及其下辖县的中医医院,上传数据量已达 100 多万条。然后采用大数据的海量数据挖掘技术,选择合适的数据挖掘方法,如关联规则、分类和聚类等,针对中医的优势病种如胃炎,以电子病历的海量真实数据进行数据挖掘,从而挖掘出中医证候、治法以及处方等信息,建立数据仓库,解决了样本采集的质量和容量问题。再以神经网络作为新算法的基本架构,采用一种学习深层(深层的好处是可以用较少的参数表示复杂的函数)非线性网络结构,实现复杂函数逼近的深度学习技术与中医的“因、位、性、势”相结合,利用挖掘出的信息建立的数据仓库、训练模型,确定新型算法中更为高效的网络权重,提高“证”到“方”的匹配准确度。

对照同类研究以及前期研究经验,新型算法研究有望在现有条件下取得 93% 以上的选法定方准确率,达到临床和科研对计算机外脑论治的要求,来实现“证候”到“方剂”

# 何首乌及其制剂的安全性研究进展及风险因素分析

罗春颖<sup>1</sup>, 黄晓婧<sup>2</sup>, 史翌川<sup>1</sup>, 蒲旭峰<sup>2\*</sup>

(1. 成都中医药大学 药学院, 四川 成都 611137; 2. 成都市食品药品检验研究院, 四川 成都 610045)

**摘要:**近年来关于何首乌及其制剂致肝损伤的报道逐渐增多,其存在的安全性问题引起社会高度关注。通过整理何首乌及其制剂的安全性研究进展并分析其风险因素,为后续建立系统安全性评价提供研究思路。药材质量方面,从何首乌的基源、种植采收、炮制三个角度分析提出最可能引起安全性问题的因素是基源混用及炮制不当;临床应用方面,分析复方配伍、剂量对何首乌毒性的影响,长期超剂量使用可致肝损伤而配伍是否有影响有待研究;制备工艺方面,不同工艺条件可能使其活性成分产生变化造成不良反应。个体因素也是其导致安全性问题不可忽视的因素。后续系统安全性评价需要着重研究炮制及制备工艺对何首乌及其制剂安全性的影响。

**关键词:**何首乌;何首乌制剂;安全性;风险因素;肝损伤

中图分类号:R282;R283

文献标识码:A

文章编号:1673-2197(2018)01-0053-04

DOI:10.11954/ytctyy.201801018

何首乌为蓼科植物何首乌(多花蓼)*Polygonatum multiflorum* Thunb. 的干燥块根,性微温,味苦、甘、涩。在临床上何首乌有生品与炮制品之分。生何首乌能解毒、消痈、截疟、润肠通便;制何首乌能补肝肾、益精血、强筋骨、乌须发、

化浊降脂<sup>[1]</sup>。

何首乌作为我国传统中医常用的补益类中药,《开宝本草》称之“久服长筋骨,益精髓,延年不老”。在历代古籍中均鲜有记载其毒性,民间也流传着许多关于何首乌神奇功

的准确对应,为中医客观化探索一种新的数据分析和研究方法,以规范中医论治过程。

## 参考文献:

- [1] 张书河,陈群.数据挖掘及人工神经网络技术在中医证候研究中的应用概述[J].医学信息,2008(12):2145-2150.
- [2] 骆文斌,吴承玉,徐征,等.肺系证诊断的人工神经网络方法探讨[J].中国临床研究,2010(10):926-927.
- [3] 孙贵香,袁肇凯.人工神经网络在中医证候研究中的应用[J].中华中医药学刊,2007(7):1450-1452.
- [4] 温宗良,岳桂华,杨靖,等.基于共轭梯度算法的BP神经网络在高血压证候诊断中的应用[J].山东中医药大学学报,2012(3):183-184.
- [5] 岳彩青,常青美,庞学民,等.基于聚类分析的RBF网络建模方法及应用的研究[J].计算机仿真,2006(1):120-123.
- [6] 李建生,胡金亮,余学庆,等.基于聚类分析的径向基神经网络用于证候诊断的研究[J].中国中医基础医学杂志,2005(9):685-687.

- [7] 任廷革,李宇航,陈永义,等.中医辨证论治信息研究评述[J].中国中医药信息杂志,2006(3):94-96.
- [8] 倪荣,彭明德,沈玉强,等.模式识别与中医智能化辨证[J].中华中医药学刊,2015(5):1061-1064.
- [9] 朱敬,朱翰学.论中医“证”及“辨证论治”[J].中华中医药杂志,2017(1):21-24.
- [10] 何光明,李晓鹤.浅谈辨证、辨病、辨证论治三结合诊疗体系[J].江苏中医药,2015,47(5):13-14.
- [11] 孙占全,刘艳骄,西广成,等.证候与方剂相关性的研究[J].中国中医基础医学杂志,2006,12(2):150-152.
- [12] 洪芳,何建成,曹雪滨.人工神经网络在中医证候研究中的应用现状与趋势[J].辽宁中医杂志,2013(1):13-15.
- [13] 赵铁牛,于春泉,王惠君,等.人工神经网络在中医证候学中的应用初探[J].中华中医药杂志,2014(3):831-833.
- [14] 王永炎,盖国忠,陈仁波.中医辨证论治思维的研究方法与发展方向[J].环球中医药,2014(1):1-5.

(编辑:宋勇刚)

收稿日期:2017-06-05

基金项目:2015年国家药品标准提高课题(2015-21)

作者简介:罗春颖(1990—),女,成都中医药大学硕士研究生,研究方向为中药新制剂、新剂型、新技术。

通讯作者:蒲旭峰(1963—),男,成都市食品药品检验研究院研究员,主任中药师,研究方向为中药分析及中药新药研发。E-mail:pxf68@263.net