

基于人工神经网络的《伤寒论》 方证知识库的构建*

杨 涛, 吴承玉**

(南京中医药大学基础医学院 南京 210023)

摘 要:目的 运用人工神经网络算法探讨“方”与“证”之间的非线性映射关系。方法 整理《伤寒论》中有症有方的条文,利用人工神经网络对 245 条数据进行训练,建立方证知识库,先选择 100 条作为测试样本,然后分别以桂枝汤证、麻黄汤证、小柴胡汤证、白虎汤证的症状表现输入模型,进行“症状”到“药物”的预测分析。结果 模型预测正确率 79%,对桂枝汤证及麻黄汤证各药物的预测误差均在 0.1 以内,对小柴胡汤、白虎汤证中主药的预测误差在 0.1 以内,对其它药物的预测误差在 0.3 以内。结论 利用人工神经网络能较好的模拟方证之间的非线性映射,该思路可进一步应用于方证辨证论治体系中,为方证规范化及中医信息化服务。

关键词: 伤寒论 方证 人工神经网络 知识库

doi: 10.11842/wst.2013.09.030 中图分类号: R2-03 文献标识码: A

《伤寒论》法寓于方,方由证立,其中“桂枝证”、“柴胡证”等皆以方名证,开创方证辨证论治之先河^[1]。“方证对应”理论产生于对方证对应变化规律的研究与实践,自张仲景开创以来,在诸多伤寒家的实践下,逐步发展成为一门相对独立的学说,此处的“方”不仅是指药物的特定组合,而且指具有明确应用指征的药物^[2]。“证”是对四诊信息表达的机体病理及生理变化整体反应状态的概括,具有内实外虚、动态时空及多维界面的表现特征^[3]。方证之间的对应关系纷繁复杂,“一方多证”、“随证加减”及“但见一证便是”等影响了人们对方证的认知和把握。

方证体系是一个非线性的复杂系统,不能使用线性回归方法研究,而人工神经网络^[4,5](Artificial Neural Network, ANN)是由大量处理单元互连而成的网络,类似于大脑神经突触连接结构进行信息处理的数学模型,通过对输入及输出样本的自动学

习,能够以任意精度逼近任意复杂的非线性映射,输入及输出数据之间的映射规则由神经网络在学习阶段自动抽取并分布存储在网络的所有连接中,利用 ANN 这一特性,可以构建相应的数学模型来模拟方证之间的映射关系。

1 资料及方法

1.1 数据来源及预处理

以人民卫生出版社 2005 年第 1 版中医临床必读丛书系列中的《伤寒论》^[6]为蓝本,整理其中有症有方的条文(包括症状原文及方剂组成),并根据《中医临床术语证候部分》对症状描述进行了合并、修改及删除等规范化处理,如《伤寒论》第 12 条:“太阳中风...啬啬恶寒,淅淅恶风,翕翕发热,鼻鸣干呕者,桂枝汤主之”,症状归纳为“恶寒、恶风、发热、鼻鸣、干呕、脉浮缓”,最终得到 245 条记录,涉及 354 个症状、113 方(包括蜜煎方)及 98 种药物。将“症状”作为属性,在某一记录中,如果出现该症

收稿日期:2013-03-16

修回日期:2013-04-21

* 江苏省自然科学基金面上项目(BK2009429)心系基础证的研究,负责人:吴承玉;江苏省教育厅研究生培养创新工程项目(CXZZ13_0607):基于软计算方法的心系基础证诊断知识库的构建及应用,负责人:杨涛。

** 通讯作者:吴承玉,教授,博士生导师,主要研究方向:病证规范化研究。

状则属性对应“1”,否则为“0”,如表1所示。

同理,根据《中华本草》对药名进行规范,如“肥
梔子”统一为“梔子”;“栝楼根”统一为“天花粉”,
“黄檗”统一为“黄柏”;“桂心”统一为“肉桂”等。以
“药物”作为属性,将数据格式化,如表2所示。

利用 MATLAB 7.11 构建人工神经网络模型,见图1,其中单个节点结构如图2所示。输入层354个
节点,对应354个症状;隐含层18个节点(通过比较不同节点下的网络误差,发现当隐含层节点数为18时,网络误差最小);输出层98个节点,对应98
种药物。传递函数选择对数函数“logsig”,训练函数
选择共轭梯度算法“trainsecg”,最大训练次数10 000
次,最小均方误差0.01。

当进行2 927次运算时,误差均方根为0.099,
小于设定的最小均方误差0.01,则判定模型达到了
最优解,此时模型自动产生权值向量及阈值向量,
其中包括输入层(354个神经元)与隐含层(18个神
经元)之间的二维权值向量w(354×18)及一维阈值
向量θ(18);隐含层(18个神经元)与输出层(98个
神经元)之间的二维权值向量w1(18×98)及一维阈
值向量θ1(98)。此时,神经网络通过自主学习,建立

了症状向量X与药物向量Y之间的关联,即建立了
症状→药物对应(即方证)的数学模型如下:

$$Y_j = \varphi(\sum_{i=0}^n W_{ij} X_i - \theta_j)$$

其中,Y_j为输出层第j个神经元的预测值;φ为
神经网络的激活函数;W_{ij}为第i个神经元与下一层
第j个神经元之间的连接权值;X_i为第i个神经元的
输入值;θ_j为第j个神经元的阈值。神经网络训练
存在误差,预测值Y_j浮动在0~1之间,由于对应药
物的取值只有0或1两类模式,因此,在程序设计
时,通过设定判别阈值λ区分两类模式,即Y_j>λ,则
Y_j=1,否则,Y_j=0。

2 模型预测分析

张仲景制方用药精专,法度严谨,君、臣、佐、使
层次分明。通过比较模型对经典方证中不同药物的
实际值与原始预测值(未经阈值判定的数据)的差
异,来分析模型的学习效果。设定神经网络模型的
判别阈值Lanmude=0.5,进行“症状”到“药物”的预
测分析。随机选择训练样本中的100条数据作为测
试样本,以100×354的矩阵形式输入模型,对比预

表1 条文-症状对应表

编号	恶寒	恶风	发热	呕吐	潮热	烦躁
1	1	1	1	1	0	0
2	1	0	1	0	0	0
.....
244	0	0	0	0	0	0
245	0	0	0	1	0	0

表2 方-药对应表

编号	方剂	桂枝	芍药	甘草	水蛭	虻虫	升麻
1	桂枝汤	1	1	1	0	0	0
2	桂枝加葛根汤	1	1	1	0	0	0
.....
244	麻黄升麻汤	0	0	0	0	0	0
245	干姜黄芩黄连人参汤	0	0	0	0	0	0

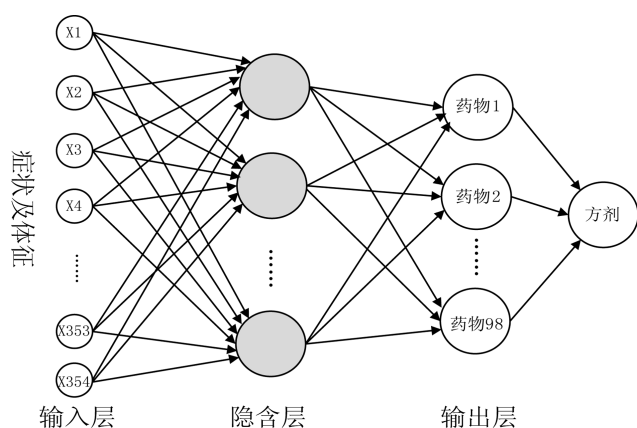


图1 方证神经网络模型

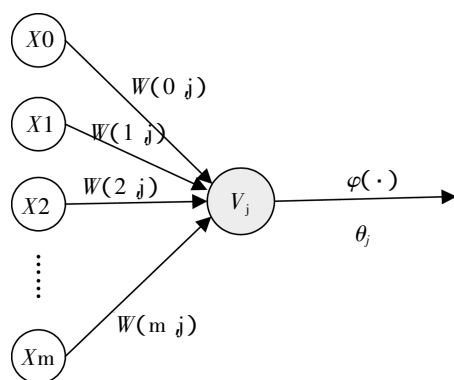


图2 单个节点结构

测值与实际值差异,结果显示,模型预测正确79条,正确率79%。分别以桂枝汤证、麻黄汤证、小柴胡汤证及白虎汤证的症状表现作为输入值,代入模型运算后,抽取模型对主要药物原始预测值,利用MATLAB绘制预测值与实际值的对比图,见图3,可知模型对桂枝汤证及麻黄汤证各药物的预测误差(实际值与原始预测值的差值)均在0.1以内;对小柴胡汤的柴胡、黄芩预测误差均在0.1以内,对其它药物的预测误差在0.3以内;对白虎汤证中石膏、知母的预测误差在0.1以内,对其它药物的预测误差在0.3以内。由此可见,模型较好地模拟了“症状”到“药物”的关系,且对主药(君、臣药)的预测准确率较高。

3 分析及讨论

通过神经网络算法构建的《伤寒论》方证知识库,可以模拟方证之间非线性映射,实现“症状”到“药物”预测分析。该研究思路及方法可以为后世张仲景学术思想的研究,为方证辨证论治体系的研究及临床决策支持服务。当然,研究过程中也存在诸多不足之处,具体如下:

3.1 症状归类欠标准

《伤寒论》对症状的描述可谓“以词达意,以意

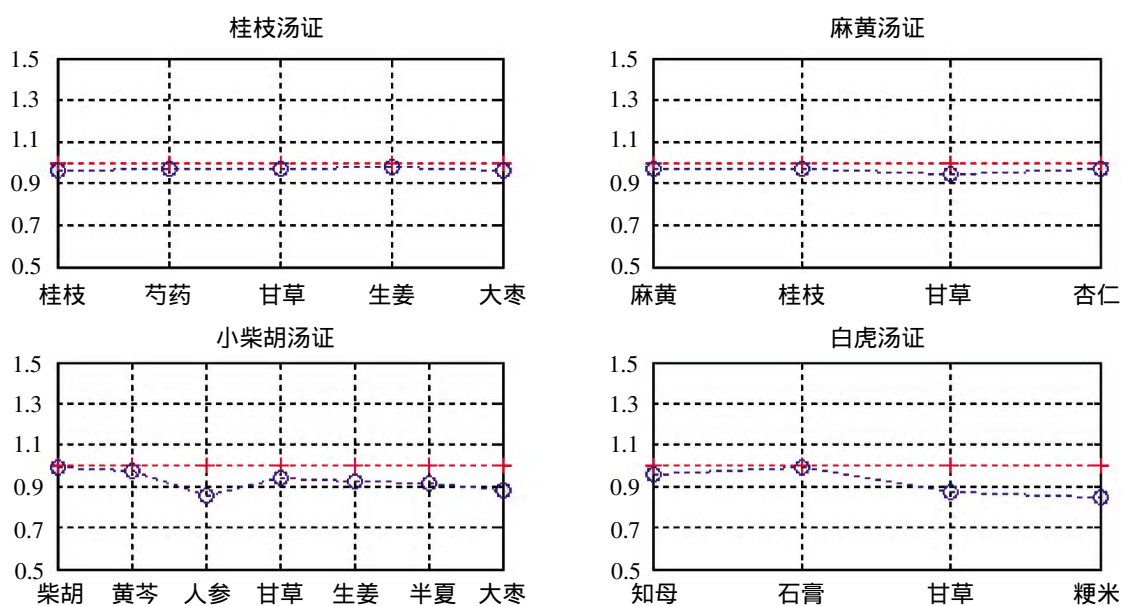


图3 方证原始预测值与真实值对比
注: x为真实值, O为预测值。

传神”,按照《中医临床术语证候部分》规范后,抹杀了其灵动及神韵。如《伤寒论》第12条:“啬啬恶寒,淅淅恶风,翕翕发热...”规范为“恶寒、恶风、发热”;又如第64条:“发汗过多,其人叉手自冒心...”规范为“心悸”。如何规范症状且不失其神韵值得思考。

3.2 条文整理欠完整

《伤寒论》文语简练,条文整理过程中需要仔细推敲。如论中第31条:“太阳病,项背强几几,无汗恶风...”其症状应当包括太阳病提纲证的内容。又如第101条:“伤寒中风,有柴胡证,但见一证便是,不必悉具...”应将伤寒、中风的症状与邪犯少阳主症排列组合后录入样本库。

3.3 方药描述欠精确

《伤寒论》对药物剂量的描述没有统度量单位,如“枚、铢、斤、两、升、如鸡子大”等,故本研究只是定性的分析主症与药物之间的关系。然而,某些方的变化就体现在剂量上,如“桂枝麻黄各半汤”与“桂枝二麻黄一汤”;又如“桂枝汤”与“桂枝加芍药汤”等。这造成模型拟合时将“多方多证”误认为“一方多证”。

3.4 模型拟合欠优化

隐含层数目及激活函数会影响模型的收敛速

度及准确度,当前模型具有训练时间长、时空复杂度高及正确率不高等缺陷,设定合适的隐含层数目以及设计符合方证对应特点的激活函数仍需进一步探索。

本文利用神经网络方法探讨《伤寒论》方证对应关系,初步构建了方证对应的数学模型,经测试及案例验证,证明方证辨识过程符合中医理论,但准确率有待进一步提高。在后续的研究中,应将方证的模糊定性与方药的精确定量引入知识库中,将后世医家优质方证“知识”丰富到知识库中,为张仲景学术思想的传承及方证研究进行深入探索。

参考文献

- 1 顾武军.《伤寒论》方证辨证探析.南京中医药大学学报,1995,11(2):20~21.
- 2 黄煌.论方证相关及其意义.中国中医基础医学杂志,1998,4(6):11~12.
- 3 郭蕾,王永炎,张志斌.证候概念的诠释.北京中医药大学学报,2003,26(2):5~8.
- 4 傅荟璇,赵红.MATLAB神经网络应用设计.北京:机械工业出版社,2010.
- 5 王士同.人工神经网络教程.北京:电子工业出版社,2006.
- 6 钱超尘,郝万山.中医临床必读丛书—伤寒论.北京:人民卫生出版社,2005.

Construction of Knowledge Base Model of Syndromes and Prescriptions in *Shanghan Lun* Based on Artificial Neural Network

Yang Tao, Wu Chengyu

(Preclinical Medicine College, Nanjing University of Chinese Medicine, Nanjing 210046, China)

Abstract: This study was aimed to explore the nonlinear relationship between syndromes and prescriptions by using the artificial neural network (ANN) algorithm. The clauses composed of syndromes and prescriptions in *Shanghan Lun* (*Treatise on Cold Damage and Miscellaneous Diseases*) were sorted out. A total of 245 clauses were trained with ANN to establish the knowledge base model of syndromes and prescriptions. After that, 100 clauses were tested. And then, the symptoms of *Gui-Zhi-Tang* (GZT) syndrome, *Ma-Huang-Tang* (MHT) syndrome, *Xiao-Chai-Hu-Tang* (XCHT) syndrome and *Bai-Hu-Tang* (BHT) syndrome were input into the model to predict the Chinese herbs. The results showed that the test accuracy of the model was 79%. For the herbs to GZT and MHT syndrome, the test error was less than 0.1. For the main herbs to XCHT and BHT syndrome, the test error was less than 0.1. While for other herbs, the test error was less than 0.3. It was concluded that the ANN algorithm can simulate the nonlinear relationship between syndromes and prescriptions, which can be applied to study the syndrome differentiation, in order to contribute to the syndromes and prescriptions standardization and the traditional Chinese medicine (TCM) information.

Keywords: *Shanghan Lun*, syndromes and prescriptions, artificial neural networks, knowledge base model

(责任编辑 叶丽萍 张志华,责任译审 王 晶)