基于 BP 神经网络的中医辨证模型构建方法

郭荣传 曾青霞 胡鑫才[1]

(江西中医药大学 岐黄国医书院，江西 南昌 330025)

【摘 要】：[目的]以太阴风湿表证辨证模型为例，探讨中医辨证模型构建方法。[方法]以江西中医药大学中医门诊规范化培训改革试点基地的“岐黄中医门诊规培系统”中 2600 例中医电子病历为样本数据，在中医理论指导下创建医案症状关键字词典，训练词向量模型，将其作为 BP 神经网络的输入，将 2080 例医案作为训练数据，剩余 520 份病例作为测试数据。[结果]该辨证模型的准确率为 88.29%。[结论]本文利用 BP 神经网络技术，构建了太阴风湿表证的系统中医辨证模型，准确率较高，为名老中医智能辨证提供了一条新的途径，值得推广。

【关键词】：太阴风湿表证 BP 神经网络 辨证模型

【中图分类号】:TB**【文献标识码】**:A

0 引言

中医辨证论治，是中医诊治过程中最为核心的部分。它是指中医师面对疾病时，根据中医理论，通过收集四诊(望闻问切)信息，然后根据中医的辨证理论确定具体的证型。根据证型，确定治则，然后决定相应的治疗方法。根据治法，确定具体的方剂，最后观察临床疗效，判断辨证用药的准确性。然而，在辨证论治的过程中，很多症状体征无法具体描述，给诊断的系统化和客观化研究带来了很大困难。

BP(backpropagation)神经网络是一种基于误差逆向传播算法训练的多层前馈神经网络，应用非常广泛。它是在 1986 年由 RineHart 和 McClelland 为首的科学家团队提出的一个的概念。BP 神经网络主要根据生物大脑神经元之间的联系，建立输入到输出的非线性映射关系，从而模拟实现人类大脑学习的过程，并且能够通过自身调节神经元的权重，让输出结果接近于我们预设的期望值，相对传统统计方法，有着更好地的系统适应性、容错性及自组织性等优点。中医的辩证论治是通过中医师的望、闻、问、切来收集患者的各种症状体征信息，通常都是一些非线性数据。因此，通过大量的中医电子病历数据对神经网络模型进行训练，同时不断调整模型中不同参数的权重，使模型的输出结果更接近于真实。本文以太阳风湿表证为例，利用中医临床门诊真实的电子病历数据，通过 BP 神经网络技术，构建中医辨证模型，为名老中医辨证经验的挖掘提供一种新思路。

1 资料与方法

1.1BP 神经网络原理与算法

1943 年，科学家根据大脑神经元模型提出了人工神经元模型，用于模拟大脑神经元之间的信息传输过程，为了增强神经网络的表达能力，科学家又引入了激活函数。BP 神经网络是一种多层前馈型神经网络，通常包含三层或更多的神经元网络层，每

[1]**作者简介**：郭荣传(1981-),男，硕士，江西中医药大学岐黄国医书院讲师，研究方向：中医数据挖掘与数据库、计算机教学;曾青霞(1995-),女，硕士，助教，研究方向：中医药数据挖掘、文本挖掘;胡鑫才(1983-),男，医学博士，讲师，研究方向：中医临床与教学。

基金项目：江西省中医药管理局科技计划一般项目(2019A092,2020A0318);江西省卫生健康委科技计划项目(202131059)

个神经元被激活就会产生一个输出信号，通常整个网络包括输入层、隐含层和输出层，每一层都可以含有多个神经元。其中，输入数据的多少决定了输入层的神经元个数，隐含层和输出层的神经元个数可以进行动态调整。而隐含层数量也不止一层，每层之间是全连接的，即各个层的神经元之间是相互连接的，如图 1 所示。

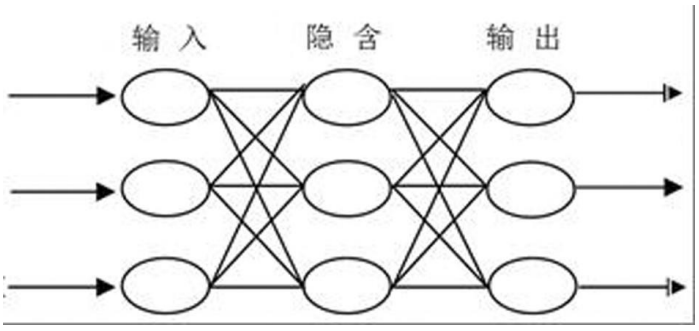

图 1 基本的神经网络结构

BP 神经网络算法是一种有导师的学习算法，整个网络学习过程包括正向的信号传播和误差的反向信号传播，然后为了是输出值达到研究需要的期望值，需要不断地调整神经元的权重。正向信号传播时，首先从输入层输入样本数据，经过每个隐含层的相互传递，最后到达输出层。如果输出层的输出值没有达到我们的期望值，则计算它们之间的误差，然后转送误差的反向传播；如果输出层的实际输出值到达了研究需要钱期望值，则结束学习算法。误差的反向传播是将正向传播时输出的信号误差从隐含层开始原路反向传送、计算，直至开始的输入层，在这个反向传送的过程中，需要将误差按照给定的算法分给各层的每个神经元节点，获得它的误差信号，并将其作为调整各节点权值的依据。在上述信号的误差反向传播过程中，需要对传递函数，进行求导计算。因为要进行微分运算，所以一般采用的激活函数为 Sigmoid 函数的对数、正切函数或线性函数，而计算的过程一般使用梯度下降算法，通过调整各层节点的权值，最大限度地降低误差信号值。信号正向传播与误差反向传播过程中，反复调整节点的权重，直到网络的输出误差达到预先设定的学习训练次数，这个过程就是神经网络的学习过程和训练过程。

1.2 数据来源

为了提高系统辨证模型的准确率，实验数据来自于中医门诊的临床真实病例。电子病历数据来源于江西中医药大学中医门诊规范化培训(简称规培)基地(岐黄国医书院)的门诊临床电子病历数据，从基地的“中医门诊规培”平台中选取 2012 年 5 月至 2019 年 5 月的中医门诊临床电子病历数据共 2600 份。将电子病历数据导出为 Excel 文件，字段主要包含患者的编号、就诊时间、主诉、现病史、中医诊断、西医诊断、治法和方剂、证型等。

1.3 数据预处理

由于电子病历数据都是中医医师在门诊临床时录入的，数据可能存在缺失、错误、不规范等问题，因此，在仿真构建系统模型之前需要进行数据预处理。对特异值、缺失值进行处理，数据中的症状、中医诊断、西医诊断等不是按照国家中医术语标准录入，因此，在输入数据建模之前，必须根据中国国家中医药管理局发布的“GB/T20348-2006 中医基础理论术语”和“GB/T16751.2-1997 中医临床诊疗术语-证候部分”规范化这些数据。

然后，对标准化后的数据进行标签处理，即将各种症状进行数值化处理，初始化语料库，然后采用 Word2vec 模型预训练词向量，即嵌入矩阵，建立字典，即语料的词汇表，且字典中的每个词都能通过嵌入矩阵表示成一个固定长度的一维向量。对于给定的句子，构建向量矩阵，作为模型的输入参数，即在预训练好的嵌入矩阵中查找每个词对应的词向量，将给定的句子映射并进行纵向拼接。

1.4 模型参数设置

BP 神经网络模型各层的参数设置，主要分为三层，如下：(1)输入层参数设计：输入层神经元的个数与词向量维度一致，本研究中选取自诉、主诉、脉诊、舌诊、望诊、查体等字段的关键词作为神经元，共 64 个。(2)隐含层的设计：按照以往的经验，如果隐含层神经元个数太少，不能充分训练网络模型，会出现很多未学习过的样本数据无法识别；但是如果设置的隐含层神经元个数太多，会充分地训练网络模型，不足之处便是模型的适应性不高，例如当输入参数与训练样本变化时，除了会导致样本不能识别外，有可能会增加模型的训练时间，出现过度拟合。因此，我们根据以往的经验，隐含层单元个数设置为 16 个。(3)输出层参数设计：为了使输出层的神经元个数与期望输出的神经元个数一致，本模型研究的输出神经元个数为 1 个，即辩证结果是太阴风湿表证，则输出 1;如果辩证结果不是太阴风湿表证，则输出为 0。其模式示意图，如图 2 所示。

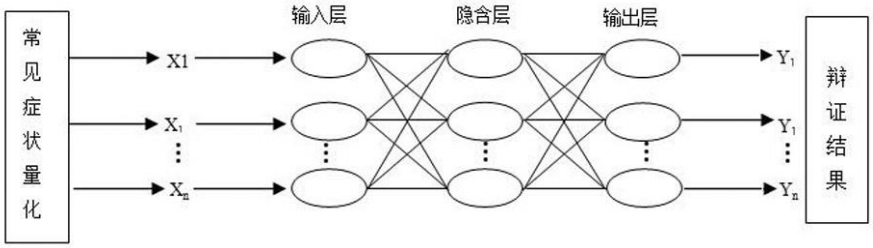

图 2 基于 BP 神经网络的太阴风湿表证辨证模型示意图

2 实验结果与分析

本研究采用 python 软件作为仿真实验平台，在 python 软件中编写 BP 神经网络程序，实现中医辨证模型。将 2600 份的病历数据以 8∶2 的比例划分为训练集和测试集，其训练模型结果如表 1。

表 1 模型结果

	precision	recall	F1-score	Test-accuracy
0	0.87	0.91	0.89	
1	0.89	0.85	0.87	0.8829
avg	0.88	0.88	0.88	

通过 2080 份训练数据集模拟训练神经网络模型，用于模型训练的数据集包含输入值与输出值，通过不断的输入数据来动态调整隐含层神经元权值，使输出值尽可能达到我们的预期值。当实验模型经过 2080 份电子病历数据训练之后，使之达到最佳结果，最后通过结果以数据的形式获取太阴风湿表证的辩证结果。

如表 1 模型结果所示，本文利用运用 BP 神经网络技术建立了太阴风湿表证的中医辨证分类模型，其辨证分类的准确率(precision)达到了 0.89,特异度为 0.87,预测的一致率(Test-accuracy)达到了 0.8829。实验结果说明本研究中的 BP 神经网络中医辩证模型在中医辩证分类中研究的可行性。为了提高模型预测的准确率，可以增加训练集病历数据的样本数量以及对各层模型参数的优化选择，该模型的建立，为探索中医辨证论治提供了一种全新思路，同时具有一定的适应性。

然而，BP 神经网络自身也存在着不足，比如固定的学习速率导致算法的收敛速度较慢，训练时间很长；神经网络中隐含层个数的选择，一般是根据实际经验，通过实验进行确定，没有一个统一通用的参数，导致网络模型还会存在一定的冗余性，增加了模型的训练时间。

3 结束语

本文主要从太阴风湿表证的真实电子病历数据出发，利用神经网络技术构建了一个中医辨证的系统模型，实验结果说明了 BP 神经网络在中医智能化辩证研究方面具有可行性。同时，增加模型输入层训练样本集的数量，模型的参数设置还可以进一步优化。此外，该方法的建立，为构建中医智能化辩证研究提供了一种新思路和方法。

参考文献：

[1]刘龙，许玲，孙大志，等.一种胃癌模糊辨证模型的建立[J].中西医结合医学杂志，2008,6(11):1117-1121.

[2]闻新，张兴旺，朱亚萍，等.智能故障诊断技术：MATLAB 应用：北京航空航天大学出版社，2015,09.

[3]蒋亮.BP 神经网络的优化研究及应用[D].南昌：南昌大学，2014.

[4]王俊杰，陈景武.BP 神经网络在疾病预测中的应用[J].数理医药学杂志，2008,21(3):259-262.

[5]司建波，杨芳，郭蔚莹，等.基于 BP 神经网络的两阶段疾病预测模型[J].吉林大学学报：工学版，2013,(S1):481-484.

[6]石凤贵.基于自然语言处理的 Word2Vec 词向量应用[J].黑河学院学报，2020,11(07):173-177.