

## · 证的研究 ·

# 中风病证候与神经功能缺损程度的关系

徐俊峰<sup>1</sup> 高 颖<sup>2</sup> 王振华<sup>3</sup> 侯忠生<sup>3</sup> 刘 超<sup>3</sup> 张严心<sup>3</sup>

(1. 中国康复研究中心北京博爱医院中西医结合康复科, 北京市丰台区角门北路 10 号, 100068;

2. 北京中医药大学东直门医院神经内科; 3. 北京交通大学)

**[摘 要]** **目的** 采用数据挖掘方法研究中风病证候与神经功能缺损程度的关系。**方法** 提取建立在中风病临床采集系统基础上的缺血性中风的临床数据, 通过应用 SPSS 12.0 和遗传算法验证中风病证候诊断量表和美国国立卫生研究院卒中量表(NIHSS)之间是否存在线性关系, 若不存在线性关系, 则通过使用多层前馈(BP)神经网络、学习向量量化(LVQ)神经网络、模糊自适应映射(Fuzzy ARTMAP)神经网络、支持向量机(SVM), 验证两者之间的非线性关系。**结果** 通过应用 SPSS 12.0 和遗传算法, 验证了中风病证候诊断量表和 NIHSS 之间不存在线性关系, 通过使用 BP 神经网络、LVQ 神经网络、Fuzzy ARTMAP 神经网络、SVM 验证两者之间存在复杂的非线性关系。**结论** 中风病证候诊断量表和 NIHSS 之间存在复杂的非线性关系。

**[关键词]** 中风病; 证候; 神经功能缺损; 数据挖掘

## Study on the Relationship between the TCM Syndromes and Neurological Impairment of Stroke

XU Junfeng<sup>1</sup>, GAO Ying<sup>2</sup>, WANG Zhenhua<sup>3</sup>, HOU Zhongsheng<sup>3</sup>, LIU Chao<sup>3</sup>, ZHANG Yanxin<sup>3</sup>

(1. Beijing Boai Hospital, China Rehabilitation Research Center, Beijing 100068; 2. Dongzhimen Hospital, Beijing University of Chinese Medicine; 3. Beijing Jiaotong University)

**ABSTRACT** **Objective** To study on the relationship between the TCM Syndromes and neurological impairment of stroke patients. **Methods** The clinical data of ischemic stroke which based on the clinical information collecting platform was extracted to test whether there was a linear relationship between the Criterion for Diagnosis and Differentiation of Syndromes of Stroke (CDDSS) and National Institutes of Health Stroke Scale (NIHSS) by means of SPSS 12.0 and genetic algorithm. If there was not a linear relationship between CDDSS and NIHSS, the nonlinear relationship was to be tested by means of BP nerve net, LVQ nerve net, Fuzzy ARTMAP nerve net and Support Vector Machine (SVM). **Results** There was no linear relationship existed between CDDSS and NIHSS with the application of SPSS 12.0 and genetic algorithm. There existed a complicated nonlinear relationship between CDDSS and NIHSS by using BP nerve net, LVQ nerve net, Fuzzy ARTMAP nerve net and Support Vector Machine (SVM). **Conclusion** Between CDDSS and NIHSS, there isn't a linear relationship, but a complicated nonlinear relationship.

**Key Words** Stroke; Syndrome; Neurological impairment; Data mining

证候和疾病关系的研究是目前研究热点问题之一。中风病作为危害人体健康的疾病, 也是目前研究的重点。中风病通过《中风病辨证诊断标准》<sup>[1]</sup> 的评分得到中风病中医证候特征, 而现代医学通过神经功能缺损程度的评价来判断病情程度, 目前主要应用的量表为美国国立卫生研究院卒中量表(NIHSS)<sup>[2]</sup>。无论是中医证候的评判还是西医神经功能缺损的评判, 针对的都是同一患者。因此, 中、西医评分量表之间一定存在着某种关系, 而这种关系可以在一定层面上反映疾病与证候的关系。本项研究采用数据挖掘的方法对中西医量表之间的关系进行了研究。

## 1 临床资料

### 1.1 诊断标准

西医诊断标准: 参照 1995 年中华医学会第四届全国脑血管病学术会议通过的《各类脑血管疾病诊断要点》<sup>[3]</sup>。中医诊断标准: 参照国家中医药管理局全国中医脑病急症科研协作组制定的《中风病诊断与疗效评定标准》<sup>[4]</sup>。

### 1.2 纳入与排除标准

纳入符合中医中风病诊断; 符合西医脑梗死、脑出血、蛛网膜下腔出血的诊断者。排除既往曾有中风病史, 而此次因出现智能障碍而入院的患者及精神病者。

### 1.3 病例来源

本研究的数据来源于北京市科委重大课题“中风病个体诊疗规律与综合治疗方案的研究”的临床信息

采集系统。所有病例均为 2003 年 11 月~2006 年 12 月北京中医药大学东直门医院、北京中医药大学东方医院、首都医科大学附属朝阳医院、中国中医科学院西苑医院、中国中医科学院广安门医院、北京市平谷区中医医院和北京市中西医结合医院的住院患者。建立在该信息采集系统上的所有数据中包含脑梗死、脑出血等数据,其中大部分为脑梗死(缺血性中风)的数据,经过数据抽提,选取缺血性中风患者的数据进行统计分析。将所有缺血性中风患者按发病 3 天内入院、发病 3~14 天内入院、发病 14 天~3 个月入院和发病 3 个月以后入院 4 个组别进行数据抽取。每个组别患者分别按入院 0~1 天、入院 2~3 天、入院 6~8 天和入院 12~16 天予以抽取。预处理后病例数据分布情况见表 1。

表 1 预处理后数据分布表(例次)

| 发病及入院时间        | 原始数据                | 入院 0~1 天 | 入院 2~3 天 | 入院 6~8 天 | 入院 12~16 天 |
|----------------|---------------------|----------|----------|----------|------------|
| 发病 3 天内入院      | 919                 | 919      | 904      | 605      | 704        |
| 发病 3~14 天入院    | 380                 | 380      | 377      | 206      | 288        |
| 发病 14 天~3 个月入院 | 521                 | 521      | 514      | 225      | 431        |
| 发病 3 个月以后入院    | 217                 | 217      | 215      | 98       | 181        |
| 总例次            | 所有组别所有时段共计 6 785 例次 |          |          |          |            |

2 方法

2.1 临床信息采集方法

对符合纳入标准的患者,根据临床治疗情况,如实记录在中风病临床采集系统中,临床信息采集内容包括:①患者的一般信息;②中医四诊信息;③神经系统评估<sup>[2]</sup>[Glasgow 昏迷评分、NIHSS、Barthel 指数、简易精神状态量表(MMSE)];④实验室检查(血常规、心电图、血脂、血凝等);⑤中风病证候(《中风病辨证诊断标准》<sup>[1]</sup>);⑥治则治法;⑦中医方药;⑧针灸;⑨中西成药使用情况。临床信息采集时段为:①所有患者入院后连续采集 3 天,每天采集 1 次;②入院时病程 1 个月以内的患者每周采集 2 次;③入院时病程超过 1 个月的患者每周采集 1 次;④急、危、重患者病情出现变化时随时采集。

2.2 数据质量评估方法

NIHSS 是临床上使用比较广泛的量表,用于评价患者的神经功能缺损程度,具有较好的信度和效度。依据文献[5],NIHSS 轻、中、重划分的边界分别为轻(0~5 分)、中(6~13 分)、重(14 分以上)<sup>[5]</sup>。

本项研究中采集了 NIHSS 评价患者的神经功能缺损情况。为了对采集的数据进行初步的临床质量评估,选择了缺血性中风发病 3 天内入院患者的信息,将入院 0~1 天、入院 2~3 天、入院 6~8 天及入院 12~16 天 4 个时点的信息汇总,总计 3 132 例次,使用 k-means

算法进行三类聚类,将 NIHSS 分为轻、中、重三类,得出相应中心点,以相邻两个聚类中心点之间的中点作为两个类别的划分边界,结果与文献[5]显示的 NIHSS 三类划分边界相同。

表 2 发病 3 天内入院患者的聚类结果

| 程度 | 基于 k-means 算法的聚类中心点 | 聚类后划分的边界 | 文献[5]显示的边界 |
|----|---------------------|----------|------------|
| 轻  | 2.30                | 0~5      | 0~5        |
| 中  | 8.46                | 6~13     | 6~13       |
| 重  | 17.99               | 14       | 14         |

表 2 示,此次聚类的结果与文献[5]标准一致。说明此次研究所采用的数据真实性、可靠性、科学性较好,基于此数据的相关数据挖掘结果可以为临床、科研提供可靠的规律和证据。

2.3 量表之间关系验证方法

假定 NIHSS 量表与中医证候量表之间存在线性关系,通过 SPSS 12.0 软件做线性回归,并用遗传算法进一步验证它们之间是否存在线性关系,如果不存在线性关系则假定它们之间存在着一一定的非线性关系,然后使用多层前馈(BP)神经网络、学习向量化(LVQ)神经网络、模糊自适应映射(Fuzzy ARTMAP)神经网络、支持向量机(SVM)对这种假定的非线性关系做出验证。

2.3.1 线性关系验证

基于 SPSS 12.0 回归分析的线性关系验证:假设原数据存在线性关系,即以西医量表作为因变量 y,中医的 6 项证候量表作为自变量 x,假设存在如表达式 1 的关系:

$$y=b_0+\sum_{i=1}^6b_i\cdot x_i$$

表达式 1

使用 SPSS 12.0 软件分析各自变量与因变量的相关性。根据相关系数判断 NIHSS 和中风病证候量表之间是否存在线性关系。

基于遗传算法的线性关系验证:假设 NIHSS 与中医 6 证候之间的线性关系如表达式 1 所示,则可以对其中的参数进行编码,适应度函数定义如表达式 1 所示:

$$f_{\text{适应度}}=f_{d(\text{西轻,中轻})}+f_{d(\text{西中,中中})}+f_{d(\text{西重,中重})}$$

表达式 2

其中 f<sub>d</sub> 为偏差度函数,定义为以西医量表为依据评判的病情轻、中、重集合与对应的以中医量表加权和后的 k-means 聚类结果为中医轻、中、重集合的偏差数。例如,假设全集有 10 个病例 U<sub>all</sub>={1,2,3,4,5,6,7,8,9,10},U<sub>西轻</sub>={1,2,3,8,10},U<sub>中轻</sub>={2,3,6,7};则使用西医量表属于轻类别的患者与使用中医量表属于轻类别的病例的偏差度 f<sub>d</sub>(西轻,中轻)=(U<sub>西轻</sub>∩U<sub>中轻</sub>)∪(U<sub>西轻</sub>∩U<sub>中轻</sub>) 中元素的个数=集合 {1,8,10,

6, 7 中元素的个数=5。

通过不断调整初始群体、遗传代数、变异率、代沟等参数,进行验证。

### 2.3.2 非线性关系验证

如果中风病中、西医量表之间不存在线性关系,则假定它们之间存在某种非线性关系,其非线性关系可以用函数表达式<sup>3</sup>给予近似表达。因此,需要使用一些研究非线性关系的方法加以分析。神经网络及支持向量机在研究非线性问题方面应用十分广泛。

$$y=f(x_1,x_2,x_3,x_4,x_5,x_6) \quad \text{表达式 3}$$

其中,人工神经网络(ANN)研究非线性化问题的思路是,根据输入与输出训练得到神经元组成的结构及相互关系,以此来表示中西医两表之间的复杂非线性关系,而支持向量机则通过升维将非线性问题线性化并结合函数的特性用最优理论加以求解。通过使用BP、LVQ、Fuzzy ARTMAP 3种神经网络以及SVM对这种非线性关系做出分析和验证。

输入端为中医证候的分值,输出端为NIHSS的轻、中、重三类划分,分类准确率均采用10-折交叉验证的平均分类准确率。

## 3 结果

### 3.1 证候量表与NIHSS之间的线性关系验证

3.1.1 基于SPSS 12.0回归分析的线性关系验证  
使用SPSS 12.0软件分析了发病3天内入院的病例的第1时段(0~1天)的919例数据的线性回归特性,各自变量与因变量的相关系数如表3。

表3 中、西医量表之间线性分析的相关系数表(Pearson 相关系数)

| 变量    | NIHSS  | 风     | 火      | 痰      | 瘀      | 气虚     | 阴虚阳亢   |
|-------|--------|-------|--------|--------|--------|--------|--------|
| NIHSS | —      | 0.254 | 0.304  | 0.289  | 0.000  | 0.238  | -0.021 |
| 风     | 0.254  | —     | 0.254  | 0.247  | 0.074  | 0.100  | 0.029  |
| 火     | 0.304  | 0.254 | —      | 0.164  | 0.013  | -0.094 | 0.129  |
| 痰     | 0.289  | 0.247 | 0.164  | —      | 0.222  | 0.289  | -0.283 |
| 瘀     | 0.000  | 0.074 | 0.013  | 0.222  | —      | 0.188  | -0.060 |
| 气虚    | 0.238  | 0.100 | -0.094 | 0.289  | 0.188  | —      | -0.035 |
| 阴虚阳亢  | -0.021 | 0.029 | 0.129  | -0.283 | -0.060 | -0.035 | —      |

注:相关系数表示了变量之间线性关系的密切程度,取值范围在0~1之间,越接近1表示线性关系越强。

表3示,自变量与因变量的互相关系数都很低,说明NIHSS值与中医证候之间并没有简单的线性关系存在。

3.1.2 基于遗传算法的线性关系验证  
通过不断调整初始群体、遗传代数、变异率、代沟等参数,对发病3天内入院的病例0~1天的评分量表总计919例进行分析,遗传算法的结果显示,证候量表与NIHSS轻、中、重类别之间并不存在线性关系。因此,综合以上分析可

得出结论,中、西医量表之间并不存在简单的线性关系。

### 3.2 证候量表与NIHSS之间的非线性关系验证

针对证候量表与NIHSS之间非线性关系,在3种人工神经网络中,BP神经网络(表4)的效果最差,LVQ神经网络(表5)与Fuzzy ARTMAP神经网络(表6)的效果相当,但LVQ网络(表5)的结果更稳定,Fuzzy ARTMAP(表6)对测试集选取比较敏感。SVM(表7)不仅结果优于这3种神经网络方法的结果,而且稳定性也更好。其分类准确率保持在60%~80%之间。

以上结果可以看出,疾病与证候之间并非简单的一一对应关系,疾病层面的神经功能缺损程度(NIHSS)与中风病证候之间存在复杂的非线性关系。在不同的病程阶段均存在复杂的非线性关系。

表4 基于BP神经网络不同组别患者在不同时段的分类准确率(%)

| 发病及入院时间     | 不同打分时间点 |       |       |        |
|-------------|---------|-------|-------|--------|
|             | 0~1天    | 2~3天  | 6~8天  | 12~16天 |
| 发病3天内入院     | 58.21   | 53.75 | 62.13 | 62.32  |
| 发病3~14天内入院  | 55.74   | 59.63 | 64.43 | 68.46  |
| 发病14~90天内入院 | 42.33   | 44.57 | 48.38 | 53.47  |
| 发病90天后入院    | 52.26   | 47.05 | 54.94 | 43.55  |

表5 基于LVQ神经网络不同组别患者在不同时段的分类准确率(%)

| 发病及入院时间     | 不同打分时间点 |       |       |        |
|-------------|---------|-------|-------|--------|
|             | 0~1天    | 2~3天  | 6~8天  | 12~16天 |
| 发病3天内入院     | 67.14   | 63.27 | 72.24 | 72.86  |
| 发病3~14天内入院  | 65.63   | 69.37 | 76.88 | 81.25  |
| 发病14~90天内入院 | 52.31   | 54.22 | 58.14 | 63.26  |
| 发病90天后入院    | 52.00   | 57.00 | 64.25 | 50.75  |

表6 基于Fuzzy ARTMAP神经网络不同组别患者在不同时段的分类准确率(%)

| 发病及入院时间     | 不同打分时间点 |       |       |        |
|-------------|---------|-------|-------|--------|
|             | 0~1天    | 2~3天  | 6~8天  | 12~16天 |
| 发病3天内入院     | 63.88   | 61.73 | 71.22 | 69.59  |
| 发病3~14天内入院  | 68.25   | 76.88 | 79.37 | 81.25  |
| 发病14~90天内入院 | 48.61   | 54.11 | 58.23 | 55.89  |
| 发病90天后入院    | 57.68   | 62.50 | 60.71 | 52.37  |

表7 基于SVM算法不同组别患者不同时间点的分类准确率(%)

| 发病及入院时间     | 不同打分时间点 |       |       |        |
|-------------|---------|-------|-------|--------|
|             | 0~1天    | 2~3天  | 6~8天  | 12~16天 |
| 发病3天内入院     | 68.98   | 66.12 | 71.43 | 73.47  |
| 发病3~14天内入院  | 73.12   | 74.38 | 78.13 | 81.25  |
| 发病14~90天内入院 | 52.56   | 54.78 | 58.67 | 69.89  |
| 发病90天后入院    | 65.75   | 63.25 | 71.75 | 70.00  |

对比分析可以看出,各类方法中前两组别(发病3天内、发病3~14天)患者的分类准确率要明显高于后

两类组别(发病 14~90 天、发病 90 天后)。在病程的不同时段,分类准确率不同。随着病程的增加,分类准确率下降。说明在临床治疗过程中除了注重对疾病的治疗外,越早对证候进行干预效果越好,这体现了早期治疗的重要性。

4 讨论

4.1 中风病证候诊断量表和 NIHSS 之间存在复杂的非线性关系,在临床上具有重要的意义

证候是疾病发生和演化过程中某阶段本质的反映,它以一组相关的症状不同程度地揭示病因、病位、病性、病势等,为论治提供了依据<sup>[6]</sup>。中医多采用中风病证候诊断量表进行证候的量化评分,并在此基础上进行相关的统计分析和研究。NIHSS 量表是目前国际公认的评测神经功能缺损程度的量表。两者都能在不同的侧面反映中风病患者疾病发生、发展过程中的变化。但由于两个量表所建立的理论基础有所不同,中风病证候诊断量表建立在中医传统理论的基础上,重视患者症状、舌象、脉象的变化,注重患者的自我感觉变化,而 NIHSS 建立在现代医学的基础上,在中风病疾病层面反映神经功能缺损程度的变化。本项研究结果显示两者存在复杂的非线性关系,该研究在临床具有重要意义。随着中风病病程的推进,患者神经功能缺损程度出现变化,中风病证候也随之改变,但两者的变化呈现不平行和不同步性,神经功能缺损程度变化相对较慢,而患者的某些自我感觉症状、舌象、脉象变化相对较快,因此某些证候的变化相对较快,如中风病急性期,风证与其他证候的组合较为突出,随着时间的推移,风证的值迅速下降<sup>[7]</sup>,神经功能缺损程度和中风病证候变化的不平行性决定了两者存在复杂的非线性关系。在临床治疗时要多方位、多角度评测患者的病情变化,NIHSS 量表与中风病证候诊断量表的评测不能互相取代,可以同步评测,互为补充。

4.2 数据挖掘的方法为证候与疾病关系的研究提供了方法学基础

数据挖掘是从大量数据中抽取有效的、新颖的、潜在有用的以及最终可被理解的模式的“非平凡过程”<sup>[8]</sup>。涉及数据库、人工智能、统计学、模式识别、可视化技术、并行计算等众多领域知识。

医学数据挖掘是一门涉及面广、技术难度大的新兴交叉学科,需要从事智能信息处理、计算机、应用数学领域的科研人员与医务工作者通力合作。将数据挖掘技术应用到中医学信息数据库中,可以发现其中的医学诊断规则和模式,为临床医疗提供数据支持。

本次研究为开放型研究,开放性中医临床研究的数据特点,决定了一般的数据统计方法难以满足中医临床研究的需要。采用数据挖掘技术对临床实时采集的中风病结构化数据信息进行分析,是对中医临床数据的处理方法进行了有益的探索。根据中医临床的需求,寻找合适的数据挖掘方法,总结一定的规律,经过专家论证,最终形成可供应用的知识,这是实现由中医临床经验向知识转变的有效途径。

4.3 交叉学科的介入为病证结合的研究提供了新的思路

本研究将交叉学科的知识引入到中医的研究领域,通过研究中风病证候与神经功能缺损程度之间的关系来实现病证结合的研究,为病证结合的研究提供了新思路和新方法。

4.4 非线性关系的研究结果为疾病预测的研究提供了基础

此次研究的结果显示,中风病证候与疾病之间存在复杂的非线性关系,这种非线性关系可以通过神经网络和 SVM 得到较好的表达。在今后的研究中通过研究重症疾病证候特征,可以为疾病的预测提供方法学支持。这为建立数学模型,通过证候判断疾病严重的程度奠定了基础,为建立基于数据挖掘的预测模型提供了基础。

参考文献

[1]国家中医药管理局脑病急症科研组. 中风病辨证诊断标准(试行)[J]. 北京中医药大学学报, 1994, 17(3): 64-66.

[2]北京神经病学学术沙龙·BNC 脑血管病临床指南[M]. 北京:人民卫生出版社, 2002: 245-263.

[3]中华神经科学会, 中华神经外科学会. 各类脑血管疾病诊断要点[J]. 中华神经科杂志, 1996, 29(6): 379.

[4]国家中医药管理局脑病急症协作组. 中风病诊断与疗效评定标准(试行)[J]. 北京中医药大学学报, 1996, 19(1): 55-56.

[5]冯丽洁, 沈洪. 第 34 例——急性脑卒中溶栓治疗和脑保护(Internet 网上专题讨论)[J]. 中国危重病急救医学, 2001, 13(8): 506-508.

[6]季绍良, 李鹏图. 全国中医诊断专题学术会议述要[J]. 中国医药学报, 1992, 7(3): 57-59.

[7]“中风病证候学与临床诊断的研究”科研协作组. 《中风病证候诊断标准》的临床验证研究[J]. 北京中医药大学学报, 1994, 17(6): 41-44.

[8]史忠植. 知识发现[M]. 北京:清华大学出版社, 2002: 8-9.

(收稿日期: 2009-06-25; 修回日期: 2009-09-30)

[编辑: 叶晨]