

## 基于信息熵的关联度的血瘀证量化研究<sup>\*</sup>

□李海霞 王阶<sup>\*\*</sup> 胡元会 何庆勇 刘如秀

(中国中医科学院广安门医院 北京 100053)

孙占全 西广成 (中国科学院自动化研究所 北京 100080)

**摘要:**应用基于信息熵的相关度对血瘀证临床诊断数据进行了分析,从数据中发现血瘀证与各症状间的相关性,在诊断血瘀证时获得了令人满意的结果。表明信息熵的相关度适合于解决中医定量诊断问题。它可以揭示众多症状间以及症状与证候间的复杂关系,从中发现证候的主要症状和次要症状并定量确定其诊断价值,有助于帮助确定证候诊断的标准和规范。

**关键词:**血瘀证 信息熵 规范化

### 一、引言

众所周知,中西医具有不同的理论体系和临床治疗的价值取向,这就造成作为国际上主流医学的西医学难以理解和接受中医,也就严重阻碍了中医药走向国际化。建立公认的中医证候诊断标准成为促进中医走向国际化的方法之一。目前,中医中证候与症状体征的相关性研究大都处于定性的研究,定量的相关性研究也都是基于统计学的线性相关研究,但这些方法只适用于解决变量之间的线性相关,而证候与症状体征之间不是简单的线性关系,因此必须用非线性方法

对证候与症状体征的相关性进行定量研究,通过研究证候与症状体征相关度的大小确定某种证候的特异性指标,为中医辨证提供客观的标准,从而为中医的辨证的规范化起到重要的作用。运用熵方法对证候相关科学规律的研究,可以用世界都能理解的“熵语言”揭示证候与证候之间的关联度、证候与症状体征之间的关联度,有了这些关联度的数据,中医就容易被理解和接受。现在以血瘀证为例,阐述基于熵的关联度在中医证候中的应用。总之,通过熵方法对证候相关科学规律的阐释,对于突破制约中医发展的瓶颈,推动中医现代化和国际化的进程,都有重要意义。

收稿日期:2006-12-06

修回日期:2007-05-19

<sup>\*</sup> 国家重点基金研究发展规划项目(2003CB517103):证候规范及其与疾病、方剂相关的基础研究—冠心病心绞痛病症结合诊断标准及疗效评价体系研究,负责人:王阶;国家中医药管理局课题(02-03JP47):血瘀证病位分类与诊断研究,负责人:王阶;国家自然科学基金重大项目面上项目(90209011):基于计算智能的血瘀证量化诊断方法学研究,负责人:王阶。

<sup>\*\*</sup> 联系人:王阶,本刊学术副主编,主任医师,博士生导师,主要研究方向:中西医结合心血管病防治病证结合方证对应研究, Tel:010-88001238, Email:wangjie0103@yahoo.com.cn。

## 二、关于信息熵

### 1. 信息熵理论概述

熵最早是 Clausius 于 19 世纪 60 年代作为一个热力学概念而引入的。早在 1925 年和 1928 年, R. A. Fisher 和 R. V. L. Hartley 就试图定量地描述信息量的大小。但直到 1948 年, Shannon<sup>[1]</sup> 才真正为信息论奠定了理论基础。20 世纪 50 年代初, S. Kullback 在概率理论与数理统计的研究中, 成功地将 Shannon 信息熵与 Fisher 信息函数联系起来, 建立了 Kullback - Leibler 信息函数<sup>[2]</sup>, 成为当今数理统计学通用的判别信息函数。

### 2. 熵理论在生命科学中的研究现状

将熵理论应用于生命科学和中医学的研究已受到了学术界的重视, 相关研究人员进行了一些有益的具有启发性的工作。张渡淮采用普利高津的耗散结构理论建立了人体熵流的数学模型<sup>[3]</sup>。陈阜新用信息论方法探讨了中药四性与微量元素的关系<sup>[4]</sup>。熵理论在脑科学, 特别是脑的高级功能产生机理的研究中, 获得了重要的应用。诺贝尔奖获得者 Edlman 等人在 1998 年用互信息理论研究了一类发生在不同脑区分布式的神经过程, 在学术界产生了重要影响<sup>[5]</sup>。众多科学家对熵理论的重要性和掌握熵理论的必要性做了肯定的评述, 这都表明熵在科学发展中的重要地位和

自身的生命力。我们认为, 将熵理论应用到中医科学中, 是一个非常具有前途的发展方向, 并有望取得好的成果。

### 3. 基于熵的关联度理论

利用信息论中的熵理论来研究非线性相关度的方法在模式识别领域已经得到了广泛的应用, 该关联度可以度量变量之间的任意统计相关性, 对变量的分布类型没有任何特殊要求, 它不仅能描述变量间的线性相关关系, 也能描述变量间的非线性相关关系。

对于一个复杂系统, 可以表示为矢量:

$$x = (X_1, X_2, \dots, X_i, \dots, X_p)^T \quad (1)$$

其中  $X_i = (X_{ia}) (i=1, 2, \dots, p, a=1, 2, \dots, q)$  是描述系统特征的变量。令  $C_i (i=1, 2, \dots, p)$  为  $X_i$  分类的集合,  $C_i$  的第  $a$  个元素  $C_{ia} = a$ , 则有  $C_i = \{1, 2, \dots, a, \dots, k\}, k \leq q$ , 并令  $n_a (\sum_{a=1}^k n_a = q)$  为事件  $X_i$  属于  $C_i$  第  $a$  类的数量, 则变量  $X_i$  的熵定义为:

$$H(X_i) = - \sum_{a=1}^k \frac{n_a}{q} \log \frac{n_a}{q} \quad (2)$$

$X_i$  和  $X_j$  的联合熵定义为:

$$H(X_i, X_j) = - \sum_a \sum_b \frac{n_{ab}}{q} \log \frac{n_{ab}}{q} \quad (3)$$

其中  $n_{ab}$  表示事件  $X_i$  属于  $C_i$  的第  $a$  类, 同时  $X_j$  属于  $C_j$  的第  $b$  类的数量。式(2)、(3)可分别表示成:

表 1 重要变量排序列表

| 序号 | 变量   | 项目     | 关联度      | 序号 | 变量   | 项目      | 关联度       |
|----|------|--------|----------|----|------|---------|-----------|
| 1  | X66  | 舌质紫暗   | 0.587841 | 16 | X16  | 眩晕      | 0.017055  |
| 2  | X49  | 舌体瘀斑   | 0.177551 | 17 | X57  | 疼痛固定    | 0.018575  |
| 3  | X23  | 舌脉曲张   | 0.178307 | 18 | X27  | 面色黑     | 0.010747  |
| 4  | X48  | 脉弦     | 0.154827 | 19 | X61  | 月经色黑    | 0.3365    |
| 5  | X22  | 腭粘膜征   | 0.15236  | 20 | X118 | 红细胞压积升高 | 0.2345    |
| 6  | X21  | 齿龈色黑   | 0.148051 | 21 | X109 | 全血黏度升高  | 0.2219    |
| 7  | X67  | 口唇色黑   | 0.203133 | 22 | X112 | 血浆黏度升高  | 0.21679   |
| 8  | X69  | 肌肤甲错   | 0.196976 | 23 | X63  | 月经血块    | 0.175338  |
| 9  | X70  | 皮肤瘀斑   | 0.15662  | 24 | X107 | 总胆固醇升高  | 0.167845  |
| 10 | X68  | 脉结代    | 0.120517 | 25 | X20  | 五心烦热    | 0.15239   |
| 11 | X24  | 五心烦热   | 0.067021 | 26 | X108 | LDL 升高  | 0.14567   |
| 12 | X106 | 甘油三酯升高 | 0.157051 | 27 | X105 | HDL 降低  | 0.125436  |
| 13 | X56  | 头昏     | 0.049896 | 28 | X62  | 痛经      | 0.1243445 |
| 14 | X59  | 畏寒肢冷   | 0.021537 | 29 | X119 | 红细胞变形性  | 0.11241   |
| 15 | X26  | 关节疼痛   | 0.01923  | 30 | X121 | 一氧化氮降低  | 0.12142   |

$$H(X_i) = \log q - \frac{1}{q} \sum_{a=1}^k n_a \log n_a$$

$$H(X_i, X_j) = \log q - \frac{1}{q} \sum_a \sum_b n_{ab} \log n_{ab}$$

有了上述熵的定义,下面给出关联度的定义。

定义 1:假设  $X_i \cap X_j = \phi$ , 则称熵

$$\mu(X_i, X_j) = H(X_i) + H(X_j) - H(X_i, X_j) \quad (4)$$

为  $X_i$  和  $X_j$  之间的关联度。

定义 2:假设对任意  $i, j (j \neq i)$ ,  $X_i \cap X_j = \phi$ ,  $p$  为任意正整数, 则称

$$\mu(X_1, X_2, \dots, X_p) = \sum_{i=1}^p H(X_i) - H\left(\sum_{i=1}^p X_i\right) \quad (5)$$

为  $X_1, X_2, \dots, X_p$  之间的关联度。

### 三、信息熵的关联度在中医血瘀证中的应用

#### 1. 资料的选择

本研究资料来自国家中医药管理局课题“血瘀证病位分类与诊断研究”中的西苑医院及云南省昆明医学院附属医院的病例 1002 例, 包括脑梗塞、冠心病心绞痛、肝硬化、月经不调、类风湿性关节炎等 5 种病, 共计 116 项指标, 其中血瘀证 598 例, 非血瘀证 404 例。症状体征是本次研究的目标变量, 通过这 5 种病中一些可测量和易观察的指标提取总体血瘀证的症状体征, 以观察他们与中医传统经验的异同, 从而为血瘀证诊断标准的规范化和客观化提供合适的诊断方法。数据分析前进行了预处理, 如异常值的删除, 缺失值的填充(用该类中具有最大出现概率的类的类标号进行填充)。

#### 2. 关于血瘀证的症状体征

表 1 和图 1 分别给出了重要变量排序和症状体征对血瘀证的贡献度。从中可以看出, 利用信息熵的关联度得到的这 30 个血瘀证的相关症状体征和理化检查指标, 与中医临床经验相符。关联

度越大, 代表对血瘀证的贡献度越大, 而甘油三酯、总胆固醇、高密度脂肪蛋白(HDL)、低密度脂肪蛋白(LDL)、全血黏度、血浆黏度、红细胞变形性、红细胞压积、一氧化氮的变化, 与现代血瘀证实验室检查的结果相同, 属于血瘀证实质研究的一部分。因此认为基于信息熵的相关度对血瘀证症状体征的提取具有临床意义。

### 四、讨论

利用信息论中的熵理论来研究非线性相关的方法在模式识别领域已经得到了广泛的应用。文献[6]提出了基于信息熵的关联度概念, 该关联度可以度量变量之间的任意统计相关性, 对变量的分布类型没有任何特殊要求。它不仅能描述变量间的线性相关关系, 也能描述变量间的非线性相关关系。该方法在中医中已经有所应用, 并取得了较好的效果<sup>[7]</sup>。本文应用基于信息熵的相关度对血瘀证临床诊断数据进行了分析, 从数据中发现了各症状间的相关性, 在诊断血瘀证时获得了令人满意的结果。这表明信息熵的相关度适合于解决中医定量诊断问题。它可以揭示众多症状间以及症状与证候间的复杂关系, 从中发现证候的主要症状和次要症状并定量确定其诊断价值, 有助于确定证候诊断的标准和规范。作为一种知识发现方法, 它从数据出发的特质不仅能提高中医专家诊断的客观性和准确性, 还有可能发现新的医学知识以丰富中医理论。可以预见, 应用信息熵等数据挖掘技术, 摸索出中医定量诊断的新方法, 将有可能带来中医定量诊断研究的突破。

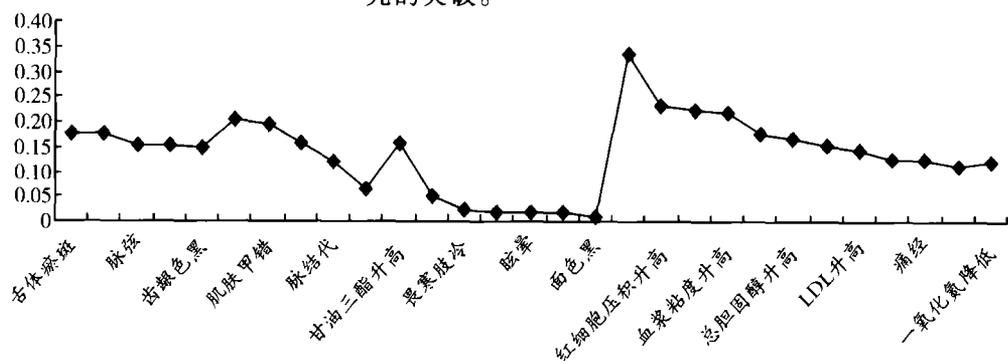


图 1 症状体征对血瘀证的贡献度

## 参考文献

- 1 Shannon C, E. A Mathematical Theory of Communication. *Bell Syst. Tech J*, 1948, 27: 379 ~ 423 & 623 ~ 656.
- 2 Kullback S. *Information Theory and Statistics*. New York: Willey, 1959.
- 3 张渡淮. 人体熵流的数学模型. *生物医学工程杂志*, 1995, 12(1): 64 ~ 66.
- 4 陈阜新. 信息论法讨论中药四性与微量元素含量关系. *Journal of mathematical medicine*, 2001, 14(2): 108 ~ 111.
- 5 Giulio Touoni, Gerald M Edelman. *Consciousness and Complexity*. Science, 1998, 282: 1846 ~ 1851.
- 6 西广成. 复杂系统分划的熵方法. *自动化学报*, 1987, 13(3): 216 ~ 220.
- 7 赵亚丽, 西广成, 刘艳骄. 用相关系数法探讨齿痕舌出现的特点. *中国中医基础医学杂志*, 2003, 9(9): 77 ~ 79.

## A Quantization Study of Blood Stasis Syndrome Using Entropy Correlation

Li Haixia, Wang Jie, Hu Yuanhui, He Qingyong, Liu Ruxiu

(Guang'anmen Hospital, China Academy of Chinese Medicine Science, 10053 Beijing)

Sun Zhanquan Xi Guangcheng

(Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, China Academy of Sciences, 100080, Beijing, China)

We analyzed the clinical data of blood stasis syndrome, in an attempt to verify the BSS diagnosis, and to abstract the symptoms of blood stasis syndrome, using entropy correlation. The result implies that the entropy correlation degree is a useful indicator in addressing the quantitative aspect of TCM based diagnosis. It improves both objectiveness and accuracy of TCM diagnosis through identifying data trait. It also helps us to discover new medical information leading to enriching the theory of traditional Chinese medicine.

Keywords: blood stasis syndrome; entropy; standardization

(责任编辑:张述庆, 责任译审:邹春申)

## 临床中药研发促进学科发展

广东省佛山市中医院围绕临床开展中药制剂的研究, 目前已开发出获得批准文号的中药制剂 208 种, 不仅方便了病人, 提高了疗效, 也促进了医院中医学术的发展。为确保中药制剂质量, 该院制剂中心已拥有一支素质优良的技术队伍。医院近年来投入 3000 余万元, 建立了制剂室, 装备了快速冷却灭菌器、真空冷冻干燥机、多功能提取罐等现代化仪器, 并已能生产颗粒剂、片剂、丸剂、软膏剂等 26 个剂型, 其中中药巴布剂新剂型解决了传统贴膏剂载药量少、刺激性大、不能反复使用的缺点, 使传统制剂重新焕发出了“青春”。(刘燕玲) 广东

省佛山市中医院围绕临床开展中药制剂的研究, 目前已开发出获得批准文号的中药制剂 208 种, 不仅方便了病人, 提高了疗效, 也促进了医院中医学术的发展。为确保中药制剂质量, 该院制剂中心已拥有一支素质优良的技术队伍。医院近年来投入 3000 余万元, 建立了制剂室, 装备了快速冷却灭菌器、真空冷冻干燥机、多功能提取罐等现代化仪器, 并已能生产颗粒剂、片剂、丸剂、软膏剂等 26 个剂型, 其中中药巴布剂新剂型解决了传统贴膏剂载药量少、刺激性大、不能反复使用的缺点, 使传统制剂重新焕发出了“青春”。

(信息来源: 中国药材 GAP 网)