

杨枝青, 包含飞, 周强, 鲍颖洁: 计算机应用与软件 (信息技术应用论文集),
Vol. 3, (增刊), 102-104, 2006_

中医知识整合论的若干基础研究

杨枝青 包含飞 周强 鲍颖洁

上海中医药大学 中医信息化-标准化研究室

摘要 本文按照生物医学知识整合论的观点对中医的认知科学本质, 中医数据生成器及其数据类型, 中医知识的信息单元, 中医语义学及其背景空间, 中医证的症状泛权模型等信息学基本问题进行探讨。

关键词 中医 生物医学知识整合论 中医信息学 认知科学

Some Basic Researches of TCM Based on the Principles of BMKI

Zhiqing Yang Hanfei Bao Qiang Zhou Yingjie Bao

Lab of Informationization and Standardization of TCM

Shanghai Univ. of Traditional Chinese Medicine

Shanghai, China, bhf2002@online.sh.cn, bhflhl@yahoo.com.cn

Abstract This article studied the cognitive essence, the types of data and data generators, the information units, semantics and its background spaces, symptom-frequency pan-right of the syndromes of Traditional Chinese Medicine(TCM), by the views of The Theory of BioMedical Knowledge Integration(BMKI).

Keywords TCM; The Theory of BioMedical Knowledge Integration; Informatics of TCM; Cognition Science

1. 中医知识的认知科学本质

如前所述，认知（cognition）一词源于拉丁语 *cognitio*（the action or faculty of knowing or learning），是心理学研究的重要领域之一，指人获得知识或学习的心智机理或过程，也是任何科学产生发展的心智机理或过程，当然也是中医知识体系产生发展的心智机理或过程。在 20 世纪 60 年代后，随着计算机的产生，系统论、控制论和信息论的出现，认知一词被引进到了人工智能领域，并最终促成了现代认知心理学的诞生，认知一词的意义逐渐发生了变化。现代认知心理学家将信息加工、整合过程与人类的认知过程联系起来，创造了一种新的研究人类认知活动的方法——信息加工论。他们认为人的一切行为受其认知过程的制约。所以，他们主张通过研究认知活动本身的过程与结构，揭示人类心智的本质。从此，认知与信息学有了不可分割的联系。^[1-5]

中医学是整个医学大家庭的一个组成部分，但在认知对象和手段上与现代医学有所差别或互补。由于中国古代传统文化的影响，在认病和辨病方面，中医一般着重对人体自然征象的观察和识别，并未对人体进行解剖或作更具体更细化研究。在治疗手段方面，中医着重自然药物的整体（药味）及其复合（方剂）医学功效而不是其化学成分的医学功效。所以无论中医的观控对象和观控手段，都带有“综合性”特征。从现代医学来看，它们的行为是大量的更为细化的元素或机制作用的总和或最终结果。这种“**综合性-to-综合性**”是中医理论或中医认知科学的一个重要特点，也是中医作为复杂性科学的一个重要特点^[6-9]。

现代医学试图用微观实验发现的机制来解释综合性临床现象的目的远没有实现，可以预见前面的路途还十分漫长。而相比之下，中医已经形成了一套“**综合性事物-to-综合性事物**”的理论体系。那么形成中医理论的认知科学机制是什么？中医认知科学是如何处理不明确性和明确性这一重要关系的？从中医认知科学的观点出发，中医的理论是否需要进一步完善和发展，如何进一步完善和发展？

由包含飞教授开创的生物医学知识整合论（The Theory of Biomedical Knowledge Integration, BMKI）^[10, 11]是当前信息时代生物医学知识爆炸或危机

的大背景下的产物，它的建立依赖于坚实的、系统的知识或认知科学的基础。

BMKI 指出，由于认知动机 (cognitive motive) 的不同，人类认知目标 (cognitive goal) 可分为两大类：一类是本能求知性的认知目标，指“在人类求知本能和探索欲望驱使下的无穷目标”^[10]，生物学和医学的基础研究多属此类；另一类是在具体的物理性或意识性功利追求下的认知目标，指在具体的物理性或意识性功利追求驱使下的获取知识的努力所“希望达到的具体的物理的或意识的目的（事物和关系）的集合”^[10]，因此这种研究活动的目标是相对确切的，相对限定的。中医临床诊疗活动有明显的防病治病保健益寿的功利目的，主要应归属于功利性认知目标指导下的认知活动。

2. 中医知识的数据采集

2.1 广义数据生成器理论

与其他科学一样，中医知识的获取也是从数据采集开始。中医的数据采集同样遵循一般的认知规律，数据生成力求客观准确。根据包含飞教授广义数据生成器理论^[10]，数据生成器由广义观测器和广义判别器组成（见图），笔者参考了英国著名心理学家 Broadbent 的过滤器模型 (Filter Model)，认为广义的判别器的含义应包括诸如专业人士的知识背景空间，认知目标决定的观察角度、关注程度、社会关系或经验等广义的参与因素，还应考虑广义判别器对广义观测器的指导作用。因此充实和细化了包教授的广义数据生成器理论（见图 1）

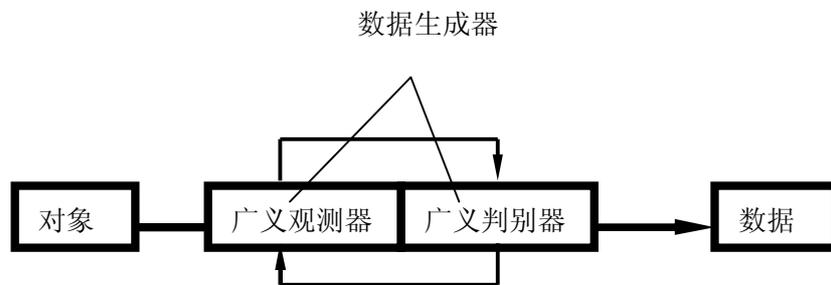


图 1 从认知对象通过数据生成器获得数据的认知过程。

2. 2 数据生成器类型

BMKI有关论文指出“不同类型的数据生成器所生成的数据的认知学属性如语义(覆盖性)、确定性、精确性方面存在差异”,被称为数据的生成属性(generation attribute)^[10]。下文举例说明各类数据生成器及其生成的各类具有中医学特色的临床信息:

1. I类数据生成器:医生基本感觉层次的测量与判断。如“发热,咳嗽,咳痰,两颧发红,双下肢浮肿,颜面浮肿,面色少华,形体消瘦”等。

2. II类数据生成器:医生心理层次的测量与判断。如“神清,精神萎软,言语欠流利”等。

3. III类数据生成器:医生专业知识层次的测量与判断。如“舌质红,苔薄白,脉弦细”等。

4. IV类数据生成器:病人或家属基本感觉层次的测量与判断。如“发热,口干,恶寒,咳嗽,咳痰,恶心,腰痛,腰酸,胸闷,心慌,心悸,夜寐不安,夜尿频多,纳差”等。

5. V类数据生成器:病人或家属心理层次的测量与判断。如“五心发热,烦躁,心中惴惴不安,终日惶恐,咽中如有赘肉,自觉有气从脐下往上冲”等。

3. 中医知识的信息单元

当前人类在人工智能方面最富想象力、最雄心勃勃也最具潜力的当数语义Web(Semantic Web)或本体Web(Ontology Web),1999年由Tim Berners-Lee等人提出。其目的是为了发表在网上信息资源不仅操作者能理解,而且计算机也能“理解”、“推导”和“运算”,这将可能根本上提高网络资源的利用性和有效性。

为此Tim Berners-Lee于在2000年提出了语义Web的体系结构,其最底层是统

一资源标志符(Uniform Resource Identifier, URI)和统一字符编码(Unicode),这是Web上资源标识和字符表示的标准。

容易理解,语义Web的信息或知识资源必须用统一格式来表示。1997年10月,W3C正式发布了RDF(Resource Description Framework,即资源描述框架)草案。后来又用RDF的词汇描述语言RDF Schema(RDFS)进一步丰富了RDF的描述功能和互联功能。

RDF采用三元组形式:资源,属性或特征,值表示资源,三个成分也被分别称为主体(subject),谓词(predicate),客体(object)。

为了进行生物医学知识整合探索,BMKI也对信息单元这一重要的基础概念作过独立的探讨,事实上BMKI对信息单元的研究较之其他同类研究更为深入,更为丰富。包含飞教授在BMKI有关论文中指出:“从根本上讲,任何信息单元需要回答三个问题:(1)我们的观察对象是什么?(2)我们关注的是对象的那一方面?(3)观察的结果如何?”^[11]。因此关于信息单元最基本结构,BMKI与其他相关研究得到了一致的结果。它将信息单元分为三个基本成分:信息载体C(观测对象,Carrier),变量F(观测方面,Feature)和变量值V(观测值,Value),被称为信息单元三联体。

BMKI指出:“说到底,任何信息单元的任何变量都是一种关系,即一种二元或多元关系,包括多元序关系。我们按照信息变量的惯常性与特殊性、恒定性与变化性、单纯性和复合性可把变量分为静态变量和动态变量。前者包括特性(或状态)变量及关系(明的、潜的)变量,后者可看作静态变量的连续有序的复合系列”^[11]。(图2)

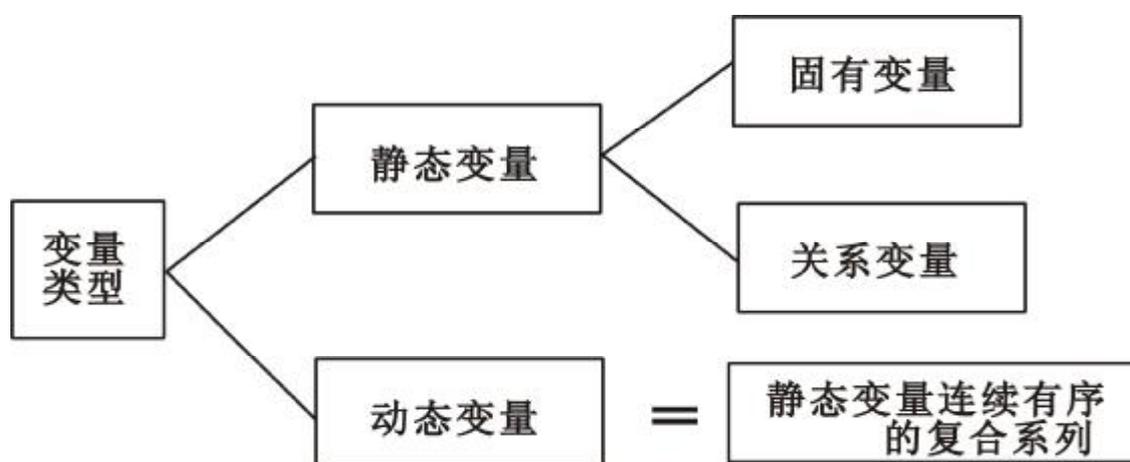


图 2 信息单元的变量类型

中医资源是整个世界医学知识资源的一部分, 中医自己也有潜力和能力为人类提供语义或本体 Web 资源。其基本信息单元的结构也与一般认知科学无异。表 1 是关于中医信息的静态变量的实例。

BMKI 认为: 所谓关系变量是一种“势态”或“潜质”, 很多情况下, 只有在相关的角色因子或“配偶”存在或加入, 关系才会出现。可表示为 $V = f(y, x)$, 其中 x, y 为广义的关系角色, f 为关系, V 为广义的变量值。表 2 为中医学中一些关系变量的实例。

表 1 中医病证信息单元实例

载体 (C)	变量 (F)	变量值 (V)
面部	颜色	黄色
口唇	颜色	红色
右手指甲	颜色	青色
左耳朵	外形	肿大
左上肢部分皮肤	颜色	玫瑰红
左上肢部分皮肤	外形	高起 (形小如粟粒)
舌体	颜色	淡白色
舌苔	颜色	黄色
舌体	外形	胖大
脉象 (桡动脉的搏动)	速率	快, 一息五至以上 (数)
脉象 (桡动脉的搏动)	强度	无力 (虚、微、弱)
脉象 (桡动脉的搏动)	节律	不齐 (促、结、代)

表 2 中医关系变量实例

关系角色 1 (x)	关系角色 2 (y)	关系 (f)	变量值 (V)
患儿食指络脉	医生右手拇指	用力作用于 (推)	络脉显, 至气关
患儿食指络脉	医生右手拇指	用力作用于 (推)	络脉显, 呈深紫色
足背肿块	医生手指	用力作用于 (触摸)	硬
肋下	医生手指	用力作用于 (触摸)	未有异常感觉
大椎穴	毫针	由...治疗 (刺入)	高热下降
直肠	蜂蜜	由...治疗 (注入)	便秘缓解

4. 中医知识的语义及其背景空间

语义关系多为二元关系, 语义网络 (Semantic Network) 是图象化的二元关系。语义网络是支持人类大脑联想型思维的有力工具, 但如果没有足够详尽的二元关系权值 (广义) 的描述或给定, 语义网络难以进行推导。

在医学信息学方面, 工程最为浩大的是由美国国家医学图书馆主持的一体化医学语言系统 (UMLS)。UMLS 的超级叙词表包含约 900, 000 个概念, 它的语义网络有 135 种语义关系。

中医之论病, 其实涉及多个空间, 它们可以以单独的形式论病, 但一般以复合的形式或相互关系论病。就中医之论肺癆而言, 涉及的认知空间有**物理空间**: (1) 环境空间, (2) 机体群体空间; (3) 机体空间 (素体-体质), (4) 脏腑空间, (5) 病理状态空间, (6) 功能状态空间; 和**准形式空间**: (7) 八纲空间, (8) 五行空间, (9) 时间或其他描述空间等。

肺癆是具有傳染性(流行病特征：**機體群體空間**)、慢性(過程時性特征：**時間或其他描述空間**)、虛損性(虛實性特征：**八綱空間**)的疾患，以咳嗽、咳血、潮熱、盜汗(四大主症)(生理-病理功能特征：**功能狀態空間**)及身體逐漸消瘦(虛實性特征：**八綱空間**)為主要特征。肺癆的病因主要為感染“瘵蟲”(病原特征：**環境空間-機體空間的關係**)和正氣虛弱(病因特征：**機體空間(素體-體質)**)；病理機制主要有三點：(一)主要病位在肺(病位特征：**臟腑空間**)，(二)病理性質主要在於陰虛(**八綱空間**)，(三)晚期病情嚴重(過程程度特征：**時間或其他描述空間**)，可發生傳變，主要損及腎陰、脾氣(臟腑傳變特征：**臟腑空間入五行空間**)；病理性質可由陰虛導致火旺、氣陰兩虛或陰陽兩虛(病機因果特征：**病理狀態空間內的關係**)。^[12, 13]

按中醫傳統的描述，肺癆病證演變關係(語義關係和空間轉換)可見圖4。

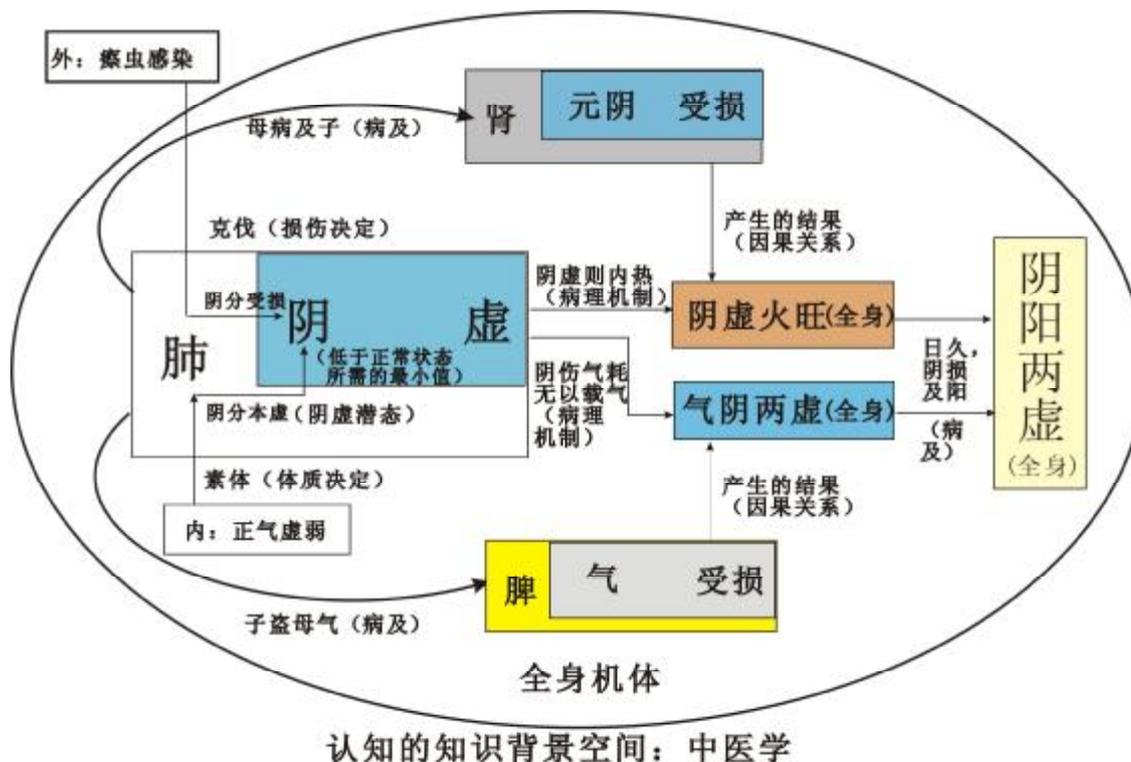


圖4 肺癆病證演變關係整合圖：不同認知空間的關係和轉化

如果我們模仿 UMLS 以統一的實體和關係體系的語義網絡來描述肺癆病的相

关证的演变，则形成包括 13 个实体(脾疾病，脾气虚，肺气虚，素体阴虚，素体肺阴虚，肺阴虚，肾疾病，阴虚火旺（全身），阴虚内热（全身），肾元阴虚损，肾元阳虚损，阴阳两虚（全身）(全身))和 8 种关系（包含，子盗母气，阴伤气耗，被克伐，母病及子，阴虚则热，阴损急阳，两种关系的复合）的语义网络（图 5）。

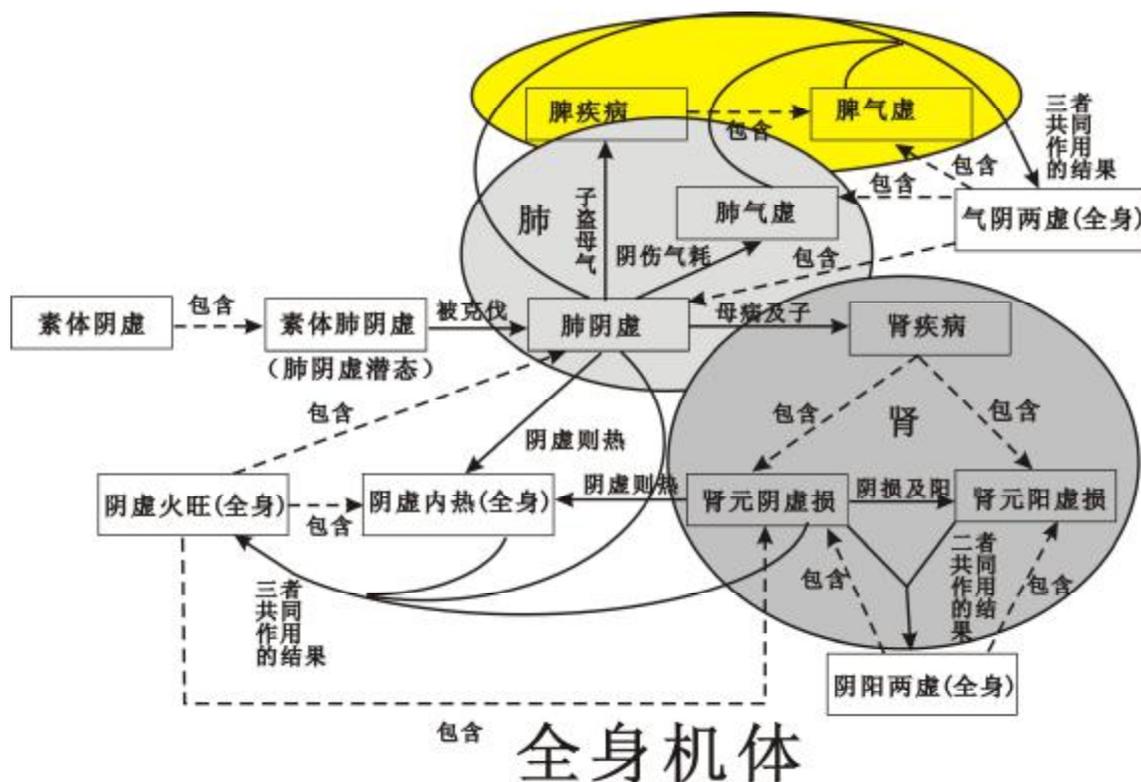


图 5 肺癆病的证语义网络图

如果我们如果我们按六阈法^[14]对图 5 的实体和关系作更为细化或更为明确的描述，则可得肺癆病证演变关系整个过程的的准形式化描述（见图 6），其中 A 为阴态，B 为火热态，E 为气态，C 为阳态。

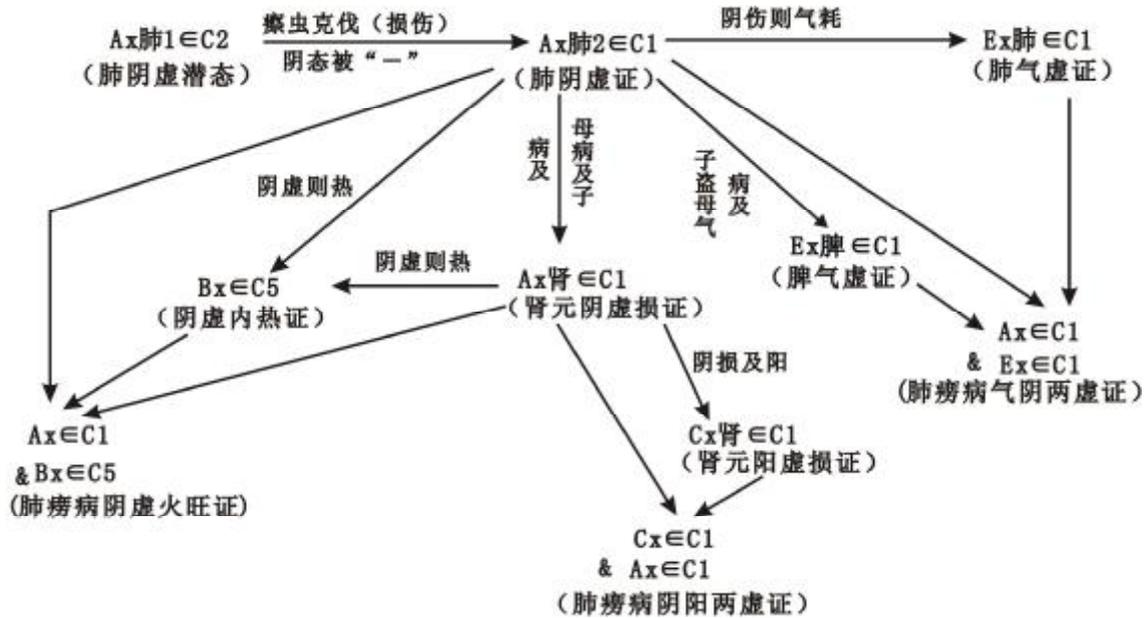


图 6 肺痿病的证演变关系准形式化描述

5. 中医证的症状频率泛权研究

在中医临床各科，几乎所有的证，均以如下肺阴（亏）虚证类似的形式进行描述：“肺阴亏虚，虚热内扰，以干咳少痰，或痰粘不易咯出，或痰中带血，口燥咽干，或音哑，潮热颧红，或有盗汗，舌红少津，脉细数等为常见症”^[12, 13]。

因此传统的证或体质的描述是理想的一系列症状的“捆绑”，是“不带泛权”的证或体质。设Z为某一证，B为症状集合或表象集合，则Z的传统描述为 $Z = B = \{症_1, 症_2, \dots\}$ 。

然而，实际临床是所见的证所包含的症状或表象不一定是Z的全部，某些应有症可以缺如，不应有的症可以出现，并且症有轻重之别。因而，临床的证所包含的症应带有频率泛权，其更为精确的形式为 $Z = \{症_1 \times W_1, 症_2 \times W_2, \dots\}$ ，其中 W_i 为症 i 的频率泛权。

临床医生对临床症状的记录往往与其认知特点有关。涉及如下类型：

一、**有或无型**：这些症状具有排斥性，例如如果没有“咳嗽”记录，则表示“无咳嗽”，无第3种情况存在。此类症状包括咳嗽、发热、咳痰等；

二、**难于取值型**：这类症状的值往往连续变化，临床医生易受各种主客观因

素影响，临床取值相对地难以确定，医生之间易于产生不一致性。此类症状有舌质红、脉细等；

三、贡献不大型：这类症状往往较少特异性，即与某一证的特定关联度不高。如饮食、二便、夜寐、面色、形体等全身性基本情况等。这类症状对全面了解机体状态是有价值的，但对于某一特定证的诊断意义并不大。若此类症状未作记载，一般多视为“正常”或“未见异常”。

本文通过岳阳医院2002年至2005年住院各科的电子病历收集到气阴两虚证（100例）、肺阴亏虚证（75例）、肝肾阴虚证（50例）、脾肾阳虚证（50例）、肺脾气虚证（40例）。

气阴两虚证带泛权的模式（见图7）为：

有或无型： 0.4口干（39例）+0.4头晕（36例）+0.4咳嗽（35例）+0.3发热（33例）+0.3胸闷（25例）+0.2腹胀（24例）+0.2咳痰（19例）+0.2气急（15例）+0.1盗汗（13例）+0.1恶心（12例）+0.1心悸（10例）+0.1腹痛（10例）+0.1呕吐（9例）+0.1气喘（9例）+0.1恶寒（7例）+0.1心慌（6例）+0.1自汗（6例）+0.1双下肢浮肿（6例）；

难以取值型： 0.5脉细（46例）+0.3苔薄（30例）+0.3脉弦（28例）+0.3舌质红（25例）+0.3苔腻（25例）+0.2舌质淡红（16例）+0.2苔白（15例）+0.1苔黄（14例）+0.1苔少（13例）+0.1脉数（9例）+0.1舌质暗红（8例）；

贡献不大型： 0.8乏力（79例）+0.7神清（71例）+0.6二便尚可（57例）+0.4夜寐不安（44例）+0.4纳差（44例）+0.4纳可（37例）+0.4夜寐安（36例）+0.4面色少华（36例）+0.3形体消瘦（29例）+0.1大便干结（12例）+0.1精神萎软（12例）+0.1小便量少（6例）。

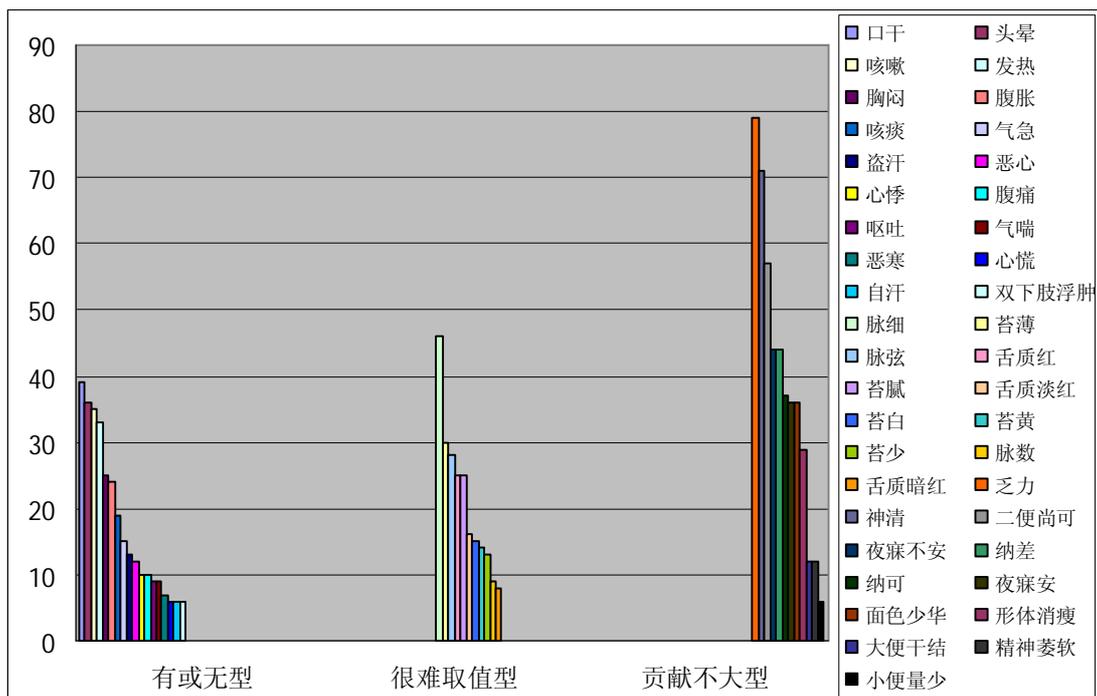


图7 气阴两虚证带泛权的常见症状统计图

参考文献

- [1]王甦，汪安圣著. 认知心理学[M]. 第一版. 北京：北京大学出版社，1992，31、47、80、138-139、203.
- [2]杨光华主编. 中医临床思维研究[M]. 第一版. 南昌：江西科学技术出版社，1992，28-43，52-77.
- [3]赵艳芳著. 认知语言学概论[M]. 第一版. 上海：上海外语教育出版社，2001，1-2、55、80.
- [4]【美】司马贺（Herbert A. Simon）著，荆其诚、张厚粲译. 人类的认知——思维的信息加工理论[M]. 第一版. 北京：科学出版社，1986，117-118.
- [5]【加拿大】G. 隆多著，刘钢，刘健译. 术语学概论[M]. 第一版. 北京：科学出版社，1985，22-23.

- [6] (2) 包含飞, 严世芸: 中医学概念的信息学属性及中西医学的互通性初探——中医学标准化预备性研究之一, 上海中医药大学学报, 15(4):8-11, 2001
- [7] (3) 包含飞: 初议中医学是复杂性科学——中医标准化预备研究之二, 上海中医药大学学报, vol. 17, No. 2, 3-6, 2003
- [8] (4) 包含飞: 续议中医学是复杂性科学——中医标准化预备研究之三, 上海中医药大学学报, 17(1): 3-6, 2003
- [9] (5) 包含飞, 严世芸, 周强, 鲍颖洁, 杨枝清: 中医证候术语学与语义码初探——中医标准化预备研究之四, 中华医学论坛, 2005, No(9-10), 71-73
- [10] 包含飞. 生物医学知识整合论(I-X). 医学信息, 2003-2005
- [11] 包含飞. 概念性本体、实验性本体和罗盘—灯塔式本体——BMKI 的概念, 原理和方法学(I). 计算机应用与软件, 2005, 22(增刊): 113-115.
- [12] 邓铁涛主编. 中医诊断学[M]. 第一版. 上海: 上海科学技术出版社, 1984, 34、175.
- [13] 张伯臾主编. 中医内科学[M]. 第一版. 上海: 上海科学技术出版社, 1985, 76-82.
- [14] 包含飞, 杨枝青, 周强, 鲍颖洁: 中医概念的准形式化描述探索(待发表)