

# 基于形式概念分析的中医辨证可视化方法

刘旭龙, 洪文学\*, 张 涛, 樊凤杰, 宋佳霖

(燕山大学 电气工程学院, 河北 秦皇岛 066004)

**摘要:** 计算机辅助中医辨证研究是一个热点, 但是目前尚缺乏一种标准的数学语言用于描述辩证的体系架构。本文提出了一种基于形式概念分析的可视化方法, 这种方法将证和证素定义为对象与属性并将它们的概念和层次关系数学化, 最后利用概念格将其可视化。最后通过一个实例初步说明了将形式概念分析应用于计算机辅助中医辨证体系研究是可行的。

**关键词:** 形式概念分析; 证; 证素; 可视化;

**中图分类号:** TP399 **文献标识码:** A

## 0 引言

辨证论治是中医理论的核心, 是中医诊断学的特点与精髓<sup>[1]</sup>。目前中医诊断学所研究的核心问题是如何科学地辨证, 涉及证候规范化研究、证与证素、证素之间、方证相应关系等方面<sup>[2]</sup>。作为中医辨证论治体系重要组成部分的“证”, 是一个开放的非线性复杂网络系统<sup>[3]</sup>。在研究证与证素这一复杂网络的关系中引入数据挖掘技术是目前中医辨证研究的一个热点。在国内, 传统的数据挖掘技术在中医的证候研究中已有大量应用, 主要方法有: 关联规则、神经网络、聚类分析、贝叶斯网络、决策树、粗糙集等<sup>[4]</sup>。但是, 这些传统的数据挖掘技术所共同面临的一个问题就是, 如何将证与证素的概念和概念层次以及它们所构成的网络关系以一种纯数学化的形式语言加以描述、演绎和推理。

形式概念分析是以数学化的概念和概念层次为基础的应用格理论的一个分支<sup>[5]</sup>, 其基本思想是基于对象与属性之间的二元关系建立一种概念层次结构, 其相应的 Hasse 图可以生动简洁地展现概念之间的复杂网络关系。因此形式概念分析被认为是进行数据挖掘和可视化的有力工具。本文应用形式概念分析的基本方法对证与证素的关系进行研究和可视化表示, 首先将证和证素定义为对象和属

性, 组成形式背景, 然后进行形式背景净化, 提取概念, 最后利用 Hasse 图进行可视化和分析。本文在第3节通过一个实例对上述方法加以说明, 最后的结论表明将形式概念分析应用于中医辨证体系的可视化数据挖掘是可行的和具有前景的。

## 1 形式背景分析

**定义 1** 一个形式背景是一个三元组  $(G, M, I)$ ,  $G$  和  $M$  是集合,  $I \subseteq G \times M$ ,  $G$  的元素成为对象,  $M$  的元素成为属性,  $(g, m) \in I$  或  $gIm$  表示对象  $g$  具有属性  $m$ <sup>[5]</sup>。

**定义 2** 设  $A$  是对象集合  $G$  的一个子集, 定义

$$f(A) = \{m \in M \mid \forall g \in A, gIm\}$$

( $A$  中对象共有属性的集合),

相应地  $B$  是对象集合  $M$  的一个子集, 定义

$$g(A) = \{g \in G \mid \forall m \in B, gIm\}$$

(具有  $B$  中所有属性的对象的集合)<sup>[5]</sup>。

**定义 3** 背景  $(G, M, I)$  上的一个形式概念是二

收稿日期: 2009-12-20 基金项目: 国家自然科学基金资助项目 (60904100)

作者简介: 刘旭龙 (1984-), 男, 河北保定人, 博士研究生, 主要研究方向为中医工程、可视化模式识别; \*通信作者: 洪文学 (1953-), 男, 黑龙江依安人, 教授, 博士生导师, 主要研究方向为信息整合、可视化模式识别和中医工程学, Email: hongwx@ysu.edu.cn.

元组 $(A, B)$ , 其中 $A \subseteq G, B \subseteq M$ , 而且满足 $f(A)=B, g(B)=A$ 。称 $A$ 是概念 $(G, M, I)$ 的外延,  $B$ 是概念 $(G, M, I)$ 的内涵, 用 $\mathcal{B}(G, M, I)$ 表示背景 $(G, M, I)$ 上的所有概念的集合<sup>[5]</sup>。

如果 $(G, M, I)$ 是一个背景,  $A, A_1, A_2 \subseteq G$ 是对象的集合,  $B, B_1, B_2 \subseteq M$ 是属性的集合, 则有以下一些性质:

- 1)  $A_1 \subseteq A_2 \Rightarrow f(A_2) \subseteq f(A_1),$   
 $B_1 \subseteq B_2 \Rightarrow g(B_2) \subseteq g(B_1);$
- 2)  $A \subseteq g(f(A)), B \subseteq f(g(B));$
- 3)  $f(A) \subseteq f(g(f(A))),$   
 $g(B) \subseteq g(f(g(B)));$
- 4)  $A \subseteq g(B) \Leftrightarrow B \subseteq f(A)。$

定义4 若 $(A_1, B_1), (A_2, B_2)$ 是某个背景上的两个概念, 而且 $A_1 \subseteq A_2 \Rightarrow f(A_2) \subseteq f(A_1)$ , 则称 $(A_1, B_1)$ 是 $(A_2, B_2)$ 的子概念,  $(A_2, B_2)$ 是 $(A_1, B_1)$ 的超概念, 并记作 $(A_1, B_1) \leq (A_2, B_2)$ , 关系 $\leq$ 称为是概念的“层次序”(简称“序”)。 $(G, M, I)$ 的所有概念用这种序组成的集合用 $\mathcal{B}(G, M, I)$ 表示, 称它为背景 $(G, M, I)$ 上的概念格<sup>[5]</sup>。

定义5 一个背景 $(G, M, I)$ 称为净化的, 如果任意两个满足 $f(g) \subseteq f(h)$ 的元素 $g, h \in G$ 都有 $g=h$ , 而且对偶地任意两个满足 $g(m) \subseteq g(n)$ 的元素 $m, n \in M$ , 都有 $m=n$ <sup>[5]</sup>。

## 2 可视化算法

概念格是一个特定形式背景下所有概念及其层次序所组成的集合, 利用概念格可以清晰地表示证与证素、证与证之间的层次关系。概念格的构造过程实际上是概念聚类的过程。因此, 在概念格中, 建格算法具有很重要的地位。对于同一批数据, 所生成的格是唯一的, 即不受数据或属性排列次序的影响, 这也是概念格的优点之一。概念格的建造算法主要要完成两项任务: 一是要生成所有的格节点(即形式背景下的全部集合); 二是要建立这些格节点之间前驱后继的关系。因此按照这两项任务完成次序的不同, 可以采用如下途径来完成: 首先生成全部的概念集合, 然后再找出这些概念之间的前驱后继关系。

定义6 若 $A, B \subseteq G$ , 而且 $A$ 与 $B$ 从左开始第1个不等元素中 $A$ 的较大,  $B$ 的较小, 则称 $A$ 字典地小于 $B$ , 形式上

$$A < B: \Leftrightarrow (\exists i \in B - A) \text{ 使得}$$

$$A \cap \{1, 2, \dots, i-1\} = B \cap \{1, 2, \dots, i-1\} \text{ [5]。}$$

产生给定背景全部外延的算法: 按字典序列的最小对象是 $g(f(\emptyset))$ 。对于一个集合 $A \subseteq G$ , 检查 $G-A$ 中的所有元素, 从最大的一个开始沿着下降序列继续检查直到第一次出现 $A < iA \oplus i$ 。  $A \oplus i$ 就是下一个对象<sup>[5]</sup>。

## 3 实例分析

本文将在下面的实例中演示如何运用形式概念分析的基本方法对证与证素的概念和概念层次进行可视化表示和推导。在形式概念分析中, 对象和属性的笛卡尔积构成一个形式背景, 本文将证的集合 $G$ 定义为对象, 证素的集合 $M$ 定义为属性, 表1为“证与证素”的背景, 图中的资料来源于朱文锋教授的《证素辨证学》一书<sup>[6]</sup> 中心脾综合证的相关内容, 为了能够对基本方法进行清晰简洁地演示, 下图中证的表达只采用了必有证, 未考虑兼有证。

形式概念就是具有相同属性集的对象所构成的二元组。可以利用定义2分4步来挖掘一个形式概念:

1) 选择一个对象 $A$ , 如表1所示, 选择4号对象心脾气血两虚证 $A_4$ ;

2) 利用定义2求 $A_4$ 的属性集

$$f(A_4) = \{m \in M \mid \forall g \in A_4, gIm\} = \{a, b, c, d, g\};$$

3)  $g(f(A_4)) = \{4', 8'\}$ 即{心脾气血两虚证, 脾不统血心血亏虚证};

4)  $(g(f(A_4)), f(A_4))$ 即为一个形式概念。

依据定义5,  $g(a)=g(b)$ ,  $a, b$ 为满列属性, 那么 $a, b$ 一定是可约属性。同理,  $A_4, A_8$ 为可约对象, 因此“证与证素”形式背景是可以净化的, 将对象 $A_4, A_8$ , 属性 $a, b$ 合并即得到“证与证素”的净化

背景,依据定义 4 和 5,可以画出“证与证素”的概念格,如图 1 所示。图中每个圆形节点为一个概念,上半圆部分为概念内涵,下半圆部分为概念外延,箭头表示概念之间的层次关系,一个概念所包

含的内涵(证素)由当前节点按与箭头相反的方向(向上)检索所有相关节点(用线相互连接的节点)的全部内涵组成。例如“心脾气虚证”的内涵即为“心脾两虚证”的内涵{心、脾、虚、气虚}加上“湿”。

表 1 “证与证素”的背景

Tab. 1 Formal context for syndrome with syndrome elements

	心(a)	脾(b)	虚(c)	气虚(d)	水停(e)	阳虚(f)	血虚(g)	湿(h)	血瘀(i)
心脾两虚证 $A_1$	×	×	×	×					
心脾气虚证 $A_2$	×	×	×	×				×	
心脾气虚水停证 $A_3$	×	×	×	×	×				
心脾气血两虚证 $A_4$	×	×	×	×			×		
心脾阳虚证 $A_5$	×	×				×			
心脾阳虚血瘀水停证 $A_6$	×	×	×		×	×			×
心脾阳虚水饮内停证 $A_7$	×	×			×	×		×	
脾不统血心血亏虚证 $A_8$	×	×	×	×			×		

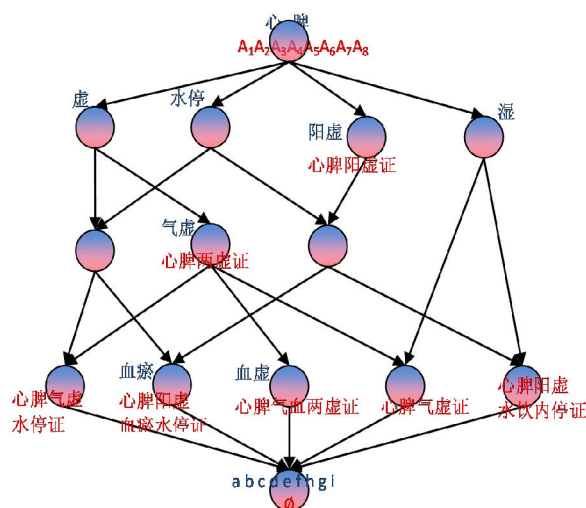


图 1 “证与证素”的概念格

Fig. 1 Concept lattice of syndrome with syndrome elements

利用形式概念分析的基本原理分析表 1,图 1,可以得到 3 点结论:

1) 在心脾综合征的形式背景下,与心脾发生关联的证素分为 3 个层次,第 1 层为{虚,水停,阳虚,湿},第 2 层为{气虚},第 3 层为{血瘀,血虚}。层次之间从上到下为包含与被包含的关系,层次越高越能体现出证素与心、脾联系的普遍性,反之层次越低越能体现出特异性。

2) 虚为图 1 中节点的最大频繁项,即为心脾综合征中最普遍的病性因子。

3) 心脾阳虚证与心脾两虚证为心脾综合征的基础证,因为从图 1 可以看出其余证型节点都为此

二节点所派生。

#### 4 结束语

本文初步探讨了利用形式概念分析的方法将证与证素的概念和层次关系可视化,为今后证候的自动归类和自动化证名生成提供了数学基础,最后通过一个实例进一步展示了如何应用概念格对证与证素的关系进行可视化和分析,初步说明了将形式概念分析应用于计算机辅助中医辨证研究是具有前景的。

#### 参考文献

- [1] 何军锋,朱文锋,黄碧群,等. 证素辨证体系与计量诊断算法实践[J]. 中医研究, 2008,21(1): 1-3.
- [2] 薄敏敏. 中医证素研究[J]. 时珍国医国药, 2008,19(6): 1490-1491.
- [3] 晏峻峰,朱文锋. 数据挖掘技术在中医诊断学中的应用研究[J]. 中国数字医学, 2007,2(10): 24-27.
- [4] 魏华凤,季光,郑培永,等. 证候诊断规范化研究的现状分析[J]. 中西医结合学报, 2007,5(2): 115-121.
- [5] Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations [M]. New York: Springer-Verlag, 1999.
- [6] 朱文锋. 证素辨证学 [M]. 北京: 人民卫生出版社, 2008.

(下转第 168 页)

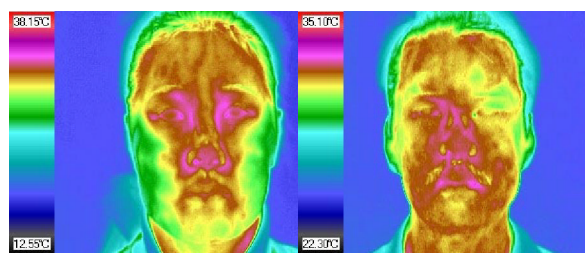


图3 面瘫热像图与健康者的对比

Fig. 3 Comparison of thermography in paralysis and healthy people

#### 参考文献

[1] Zhang Dong. A method of selecting acupoints for acupuncture

treatment of peripheral facial paralysis by thermography [J]. Am J Chin Med, 2007,35 (6): 967-975.

[2] 陈钱, 柏连发, 张保民. 红外图像直方图双向均衡技术研究 [J]. 红外与毫米波学报, 2003,22 (6): 428-430.

[3] 陈国群, 付冬梅, 常晓辉. 基于聚类算法的红外图像伪彩色增强 [J]. 激光与红外, 2007,37 (4): 384-386.

[4] 王超, 叶中付. 红外图像的变分增强算法 [J]. 红外与毫米波学报, 2006,25 (4): 306-310.

[5] 唐先武, 丁海曙, 腾秩超. 用于显示乳房病灶组织红外热图像的伪彩色方法 [J]. 光谱学与光谱分析, 2009,29 (3): 611-615.

[6] 毛玉星, 孙才新, 徐泽宇. 一种红外图像细节增强算法 [J]. 计算机工程与应用, 2009,45 (31): 183-185.

[7] 腾升华, 沈怡平. 颜色空间在图像彩色化处理中的应用分析 [J]. 计算机科学, 2009,36 (10): 277-279.

## A visual approach used to assess severity of facial paralysis

LIU Xu-long<sup>1</sup>, HONG Wen-xue<sup>1</sup>, LIU Ping<sup>1</sup>, WU Zhen-ying<sup>2</sup>

(1. College of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China; 2. Department of Acupuncture, Qinhuangdao Chinese Medicine Hospital, Qinhuangdao, Hebei 066000, China)

**Abstract:** Based on asymmetry of temperature distribution in the early onset of facial paralysis, a complete infrared temperature data visualization methods used to assess the severity of facial paralysis is presented. First the extreme points are found and replaced, then histogram equalization based on the neighborhood measure operator is used in the image, finally color sequence in which color differences are equal is used and the brightness is corrected to generate the pseudo-color image in which the temperature distribution be visualized. The final experiment shows that the image obtained by this method is good, better able to help physicians assess the severity of facial paralysis.

**Key words:** facial paralysis; visualization; temperature distribution; assessment

(上接第 164 页)

## A visualization method for differential diagnosis in TCM based on formal concept analysis

LIU Xu-long, HONG Wen-xue, ZHANG Tao, FAN Feng-jie, SONG Jia-lin

(College of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

**Abstract:** Computer-aided syndrome differentiation of TCM is a hot spot, but currently, there is no standard mathematical language used to describe the dialectical architecture. In this paper, a visualization methods based on formal concept analysis is presented. The syndromes and syndrome elements is defined as the objects and attributes, then their concepts and the hierarchical relationships are described in mathematical language, finally their hierarchy is visualized based on concept lattice. The experimental results demonstrate the feasibility of using the formal concept analysis to computer-aided system of syndrome differentiation of TCM.

**Key words:** formal concept analysis; syndrome; syndrome elements; visualization; traditional Chinese medicine