

基于模糊聚类方法分析干祖望教授 诊治耳鸣耳聋疾病群方特色

李文林¹, 陈涤平^{1*}, 曾莉¹, 陆建峰²

(1. 南京中医药大学中医学一级学科, 南京 210046; 2. 南京理工大学计算机系, 南京 210094)

[摘要] 目的: 对干祖望教授诊治耳鸣耳聋疾病的处方特点进行分析。方法: 从公开资料中选取干祖望教授治疗耳鸣耳聋的病案处方共 101 例, 运用模糊聚类及频繁模式增长算法, 分析挖掘其处方核心药群, 并通过药物构成、药群功效与临床公认的常用基本方的相似度匹配, 对不同药群中含有的基本方机构进行识别。结果: 101 个处方共分为 9 个药群, 识别出 11 个基本方。其中补中益气汤、四君子汤及八珍汤的组方结构在全部处方中所占比例最大。结论: 干祖望教授辨证治疗耳病多从调理脾胃、补益气血入手。数据挖掘结果能从一个侧面提示从脾胃论治耳病是干祖望教授病案处方的基本特色。

[关键词] 模糊聚类; 频繁模式; 干祖望; 耳鸣耳聋疾病方药分析

[中图分类号] R287 **[文献标识码]** A **[文章编号]** 1005-9903(2012)23-0004-05

[网络出版地址] <http://www.cnki.net/kcms/detail/11.3495.R.20121012.0934.024.html>

[网络出版时间] 2012-10-12 9:34

Study on Specialized Treatment Experience for Ear Disease of Professor GAN Zu-wang by Method of Fuzzy Clustering Analysis and Frequent Pattern Tree

LI Wen-lin¹, CHEN Di-ping^{1*}, ZENG Li¹, LU Jian-feng²

(1. Traditional Chinese Medicine Department of Nanjing University of Chinese Medicine, Nanjing 210046, China;
2. Department of Computer Science of Nanjing University of Technology, Nanjing 210094, China)

[收稿日期] 20120729(005)

[基金项目] 江苏高校优势学科建设工程项目、南京中医药大学中医学一级学科开放课题(YS2012ZYX308)

[第一作者] 李文林, 博士, 副教授, 从事基于文献的知识发现及名医经验分析研究, Tel:025-85811191, E-mail: njliwenlin@126.com

[通讯作者] * 陈涤平, 硕士, 教授, 博导, 从事基于文献的知识发现及名医经验分析研究, Tel:025-85811266, E-mail: chendp@126.com

[参考文献]

[1] 卢殿荣, 卢殿香, 冯利. 中药治疗恶性肿瘤骨转移疼痛应用概述[J]. 中国实验方剂学杂志, 2012, 18(2): 251.

[2] 沈欣, 李德凤, 宗桂珍, 等. 疼痛的中医症候实验研究[J]. 中国实验方剂学杂志, 2010, 16(6): 174.

[3] 程海波, 吴勉华. 癌性疼痛的中医理论探讨[J]. 中华中医药杂志, 2008, 23(1): 50.

[4] 刘稼. 聚类分析在中医药研究中的应用及意义[J]. 中医药学刊, 2004, 22(5): 927.

[5] 刘布鸣, 刘德翔, 林霄, 等. 十大功劳 HPLC 指纹图谱及聚类分析[J]. 中国实验方剂学杂志, 2012, 18(4): 95.

[6] 朱艳梅. 中医药治疗癌性疼痛的信息分析[J]. 湖南中医学院学报, 1998, 18(2): 70.

[7] 何清波, 苏丙华, 钱元. 医学统计学及其软件包[M]. 上海: 上海科学技术文献出版社, 2003: 285.

[8] 丁维, 蒋永光, 宋姚屏, 等. 基于中药药性和功效对清热解毒类药的聚类分析[J]. 广州中医药大学学报, 2007, 24(1): 3.

[9] 何前锋, 周雪忠, 周中眉, 等. 基于中药功效的聚类分析[J]. 中国中医药信息杂志, 2004, 11(6): 561.

[10] 苏晓宇. 数据挖掘聚类分析方法在中医临床中的应用[J]. 实用中西医结合临床, 2010, 10(6): 90.

[11] 卢殿荣, 卢殿香, 冯利. 癌性疼痛的中药外治临床应用概述[J]. 中医杂志, 2011, 52(24): 169.

[责任编辑 邹晓翠]

[Abstract] Objective: To analyse the effective combination drugs of GAN Zu-wang formulae in preventing and treating the ear disease. **Method:** One hundred and one formulas were selected from the public documents about professor GAN Zu-wang, the fuzzy clustering and frequent pattern tree algorithm were used to analyse the core medicine groups and the basic formulas by similarity matching in medicine efficacy and composition. **Result:** One hundred and one prescriptions were divided into 9 medicine groups, 11 basic formulas were recognized, and the mainly formulas were those for regulating spleen-stomach and benefiting qi-xie such as Bu-zhong-yi-qi-tang, Si-jun-zi-tang and Ba-zheng-tang. **Conclusion:** Regulating spleen and stomach was professor GAN Zu-wang's specialized treatment experience for ear disease, the method of data mining could provide interesting rules for understanding the academic thoughts of the famous and respected doctor of traditional Chinese medicine.

[Key words] fuzzy clustering analysis; frequent pattern tree; professor GAN Zu-wang; specialized treatment experience for ear disease

中医临床方剂数量浩繁,药物相互间的组配关系复杂,如何从浩如烟海的处方中提取出规律性的知识和信息加以应用,是方剂领域研究的重点和热点^[1-2]。文献分析发现,古今不少医家在临证用药时,每治一种病常拟定或选择一方作为基本方,有兼症、变症,则以基本方进行加减变通。其中典型如《伤寒论》载方 113 首,所用药物共 89 种,但作为众多方剂之基础的方剂只有 8 首:桂枝汤、麻黄汤、泻心汤、大承气汤、苓桂术甘汤、栀子豉汤、四逆汤、小柴胡汤。半数以上的伤寒方都是通过这基本方加减化裁而形成的^[3]。与此类似,清初名医叶天士^[4]治病常遵从一定标准,即主药数味不甚换,其换者一、二味。其咳嗽门用沙参、天花粉、川贝母、桑叶四味尤多;又如清代名医曹仁伯^[4]在谈及处方经验时也曾说:“凡治病,每病必须揣定一主方,从此前后左右,轻重疾徐,化为八法,即有八方。如其脾阳或脾气虚即以理中汤为基本方,有变证、兼证即随症加减变通……”

方剂中由基本方衍化复方的种种现象表明,通过基本方解决疾病的共性,在此基础上通过加减变通解决疾病的个性不仅有确切的历史依据,也是当代临床医家常用的配伍之法^[5-7],因此,在名医经验的分析挖掘中,以基本方为切入点研究其用药特色无疑是一个有效的途径。

名老中医干祖望教授为中医耳鼻喉科奠基者之一。在执医 60 余年的教学和临床生涯中,形成了独具一格的辨证论治体系,为中医耳鼻喉科的逐步完善和发展做出了巨大贡献,在中医界享有很高的声誉。多年来,干老撰写了大量文章、专著,其中也积累了大量处方。为了从其病案处方中分析挖掘其诊疗思想与用药特色,为耳鼻喉科的中医诊疗提供客

观的参考,我们用模糊聚类方法其病案处方中含有的基本方进行了分析挖掘。

1 方法

1.1 数据来源 全部数据均来自干祖望所著的耳鼻喉科专著及《中国期刊全文数据库》等相关期刊文献中公开的病案处方。经查重、规范,共获得耳鸣耳聋方面的处方 101 例。

1.2 方法学原理 首先利用无监督聚类方法,对全部处方按照药物的相似度进行聚类,将全部处方分成若干类别后,抽取每一类中均含有的最大药物组合作为该类的核心药群,其次,根据最大频繁模式增长算法,通过设置一定的支持度,利用计算机抽取每一类中出现的最大药物组合,将这些药物与核心药群合并,形成每一类的核心药群,以此核心药群为中心,对全部处方再次进行无监督聚类。待形成稳定的聚类结果后,将形成的核心药群与方剂数据库中的基本方进行匹配,根据核心药群与基本方的药物构成、方剂位置、方药功效等参数,识别核心药群中含有的基本方结构,从而分析名老中医的处方思路。

1.2.1 方剂聚类 首先对所有方剂样本进行硬聚类即 kmeans 聚类,根据药物构成的相似度提供多种聚类结果,如可将全部方剂聚成 3,4,5,7 类或者是其他类别数,然后对各种聚类结果进行比较,依据背景知识选择类群之间相似度最低的结果作为最优聚类结果,为下一步模糊聚类做准备。

假设共分成了 c 类,每一类的样本方差是:

$$D_c^{(j)} = \frac{\sum_{k=1}^{n_j} \|x_k^{(j)} - m_j\|^2}{n_j} \quad \text{式(1)}$$

其中 j 表示第 j 类,是第 j 类的中心(即样本均值),是第 j 类的样本数,则 c 类样本方差总和是

$$\sum_{j=1}^c D_c^{(j)}。$$

定义聚类结果评价准则如下：

$$T_1 = \frac{\sum_{j=1}^c D_c^{(j)}}{j_b} \quad \text{式(2)}$$

对于每一个类别数的 kmeans 聚类结果,挑选以上阈值最小的那个作为最终的聚类结果。

1.2.2 药群识别与确定 针对上述基础聚类结果,设定一个阈值,如可设定为处方样本数的一半,即支持度为 50%。对于每一类群中的处方,利用频繁项集增长方法,获得每一类群中若干药物组合达到支持度的最大频繁项集。对于其中的单味药,若其符合最小支持度,尽管其并不在最大药物组合中,但由于用药频次较高,仍将其加入最大药物组合中,由此得到计算机识别的核心药群。如一类 41 个处方构成的药群中,桔梗、贝母组合用了 39 次;桔梗、贝母、钩藤组合用了 28 次;桔梗、贝母、钩藤、生甘草组合用了 24 次;桔梗、贝母、钩藤、生甘草、法半夏、陈皮组合用了 22 次;桔梗、贝母、钩藤、生甘草、法半夏、陈皮、黄芩组合用了 21 次,则桔梗、贝母、钩藤、生甘草法、半夏、陈皮、黄芩组合均入核心方。

在此基础上,桔梗、贝母、钩藤、生甘草、法半夏、陈皮、黄芩、百部组合用了 18 次,18 次虽不过半数,但在这一类 41 方中,百部用了 27 次,则百部亦入核心方。

桔梗、贝母、钩藤、生甘草、法半夏、陈皮、黄芩、百部、桑白皮组合用了 8 次,但桑白皮在 41 方中用了 21 次,桑白皮亦入核心方。

桔梗、贝母、钩藤、生甘草、半夏、陈皮、黄芩、百部、桑白皮、牛蒡子组合用了 6 次,而 41 方中牛蒡子只用了 18 次,则牛蒡子不入核心方。依此类推。

具体实现路径如下：

设 $\{x, i = 1, 2, \dots, n\}$ 是个样本的组成的样本集合, c 为基础聚类的类别数目,以基础聚类的各类中心作为初始中心 $m_i, i = 1, 2, \dots, c, \mu_j(x_i)$ 是第 i 个样本对于第 j 类的隶属度函数。用隶属度函数定义的聚类损失函数可以写为：

$$J_f = \sum_{j=1}^c \sum_{i=1}^n [\mu_j(x_i)]^b \|x_i - m_j\| \quad \text{式(3)}$$

其中 $b = 1$, 是一个可以控制聚类结果的模糊程度的常数。

在不同的隶属度定义方法下,最小化式(1)的损失函数,就得到不同的模糊聚类方法。其中最聚代表的就是模糊 C 均值方法,它要求一个样本对于

各个聚类的隶属度之和为 1,即

$$\sum_{j=1}^c \mu_j(x_i) = 1, i = 1, 2, \dots, c \quad \text{式(4)}$$

在条件式(2)下求式(1)的极小值,令 J_f 对 m_i 和 $\mu_j(x_i)$ 的偏导数为 0,可得必要条件：

$$m_j = \frac{\sum_{i=1}^n [\mu_j(x_i)]^b x_i}{\sum_{i=1}^n [\mu_j(x_i)]^b}, j = 1, 2, \dots, c \quad \text{式(5)}$$

$$\mu_j(x_i) = \frac{(1/\|x_i - m_j\|^2)^{1/(b-1)}}{\sum_{k=1}^c (1/\|x_i - m_j\|^2)^{1/(b-1)}}, i = 1, 2, \dots, n; j = 1, 2, \dots, c \quad \text{式(6)}$$

用迭代方式求解式(3)和式(4),就是模糊 C 均值算法。在这里,算法步骤如下：

①以基础聚类的结果作为模糊聚类的类别数 c ,设定参数 b 。

②用欧式距离计算作为初始隶属度值,

$$\mu_j(x_i)_0 = \frac{\sum_{k=1}^c \|x - m\|}{\|x - m\|}; \quad \text{式(7)}$$

③重复下面的运算,直到各个样本的隶属度值稳定:用当前的聚类中心根据式(4)计算隶属度函数;用当前的隶属度函数按式(3)更新计算各类聚类中心。

当算法收敛时,就得到了各类的聚类中心和各个样本对于各类的隶属度值,从而完成了聚类划分。然后对模糊结果进一步去模糊化,选出隶属度最大的类别作为该样本的最终类别,将模糊分类转化为确定性分类。

1.3 基本方结构识别 作为组方因子,尽管目前关于基本方的纳入标准尚未有定论,但根据文献报道^[5-6],一般是以《伤寒论》中那些针对单一病机、用药精当、配伍关系明确严谨、疗效确切的独立方剂为主。考虑到数据挖掘方法对样本量的需求,本次研究中,我们将《伤寒论》及《中医方剂学》中具有明确配伍关系的方剂为基础建立了基本方数据库,以此为参照,借鉴基于传递闭包的摄动思想^[8],对方剂各味药的药效、剂量等指标进行模糊化处理后,通过方药模糊相似矩阵与其传递闭包矩阵中元素的对应关系,利用计算机对病案处方与经方的相似性进行匹配,并通过处方药物的药效、剂量、位置等方面的权重设置,通过匹配率(为 0~1 之间的有理数)来推断方剂中蕴含的经方结构。具体匹配过程如下：

一般的,首选匹配条件为:考虑药物的君臣佐使,考虑药物的剂量,君臣药完全匹配,佐使药的匹配率为 p ,所有的药物剂量所在模糊区完全匹配,以此时给出的经方结构为首选;若首选条件下没有相关方剂结构出现,则进入匹配条件2:即考虑药物的君臣佐使,不考虑药物的剂量,君臣药完全匹配,佐使药的匹配率为 p 。再次,则进入匹配条件3:即不考虑药物的君臣佐使,不考虑药物的剂量,只判断药物是否存在,匹配率为 p 。

2 结果

全部数据在.net环境下构建的名医处方分析挖掘平台上完成。通过数据规范、导入、聚类分析、基本方结构匹配,对干祖望教授101例治疗耳鸣耳聋的处方共聚类形成9类,经与数据库中经典方剂匹配,形成的核心药群及基本方识别结果见表1。

3 结果分析

聚类分析可以将一组数据按照本身的内在规律较合理地分为几类,通过比较各事物之间的性质,将性质相近的归为一类,性质差别较大的归入不同类,缩小主观判断所造成的误差,用这种方法可将一些观察对象依据某些特征加以归类。传统的硬划分聚类,按照特征函数的0和1的二值属性,将样本对象严格地划分到某个类中。这种类别划分的局限是分明的,具有非此即彼的性质。在实际应用中的样本对象,其形态和类属方面存在着中介性,具有亦此亦比的性质。为了表达样本类属的中介性,得到各个

样本隶属于各个类别的不确定程度,可以在聚类分析引入模糊集理论,利用模糊的软划分来解决问题,即模糊聚类分析。模糊聚类使每一个样本与各个聚类中心都有一个隶属关系,并用 $[0,1]$ 闭区间的值表示样本对各个类的隶属度^[9]。本文即在基础聚类的基础上,通过模糊聚类计算干老病案中的每一例处方分别隶属于每个核心方的隶属度,得到每例处方最有潜力类别,进而将不同的处方划分到不同的类别中,然后通过与已知配伍关系的基础方对比分析,对干老病案中蕴含的核心方药结构进行了识别。

据报道^[8-9],干老的学术思想源于东垣学派。他致力于脾胃学说研究,提出七窍以脾为本的观点,即“诸窍空清统于土”,认为头面清窍的视、听、嗅、味完全靠清阳之气来濡养温通,一旦升清降浊的机能发生障碍,清阳不升,浊阴不降,五官就会被浊阴之气弥漫笼罩,致清窍不清,耳不能闻、舌不能味、鼻不能嗅。因此耳鼻咽喉在其生命活动中所需要的生气,不论是耳之肾、鼻之肺、喉之肺其生命之源都在脾胃。具体到耳病,常体现于脾虚湿浊内困。脾虚则运化水湿功能失常,湿浊内生,邪毒滞留于耳窍。另一方面,脾虚,输布水谷精微失职,气血生化之源不足,失于濡养而功能虚弱。湿浊犯于耳壳则流脓,湿浊停聚于耳道或耳壳则为湿疹;湿浊痹阻耳窍脉络,蒙蔽清窍则症见耳鸣、耳聋、眩晕等。在宗脾胃学说的前提下,干老将健脾利湿、醒脾燥湿、健脾制

表1 干祖望病案核心药群与核心方数据挖掘识别结果

核心药群	含有该药群的 处方数量	匹配基本方	匹配率
知母 10 g,川黄柏 3 g,龟甲 10 g,熟地黄 10 g,山药 10 g,茯苓 10 g,丹皮 6 g,泽泻 6 g,山萸肉 6 g	11	大补阴丸 知柏地黄汤	0.654 0.752
黄芪 15 g,党参 15 g,当归 10 g,升麻 6 g,柴胡 12 g,白芍 6 g,山药 10 g,茯苓 10 g,石菖蒲 3 g,葛根 6 g	21	补中益气汤	0.823
升麻 3 g,葛根 6 g,石菖蒲 3 g,路路通 10 g,党参 10 g,白术 6 g,茯苓 10 g,山药 10 g,甘草 3 g	24	四君子汤	0.638
党参 10 g,白术 6 g,茯苓 10 g,熟地黄 10 g,当归 10 g,川芎 6 g,白芍 6 g,麦冬 10 g,甘草 3 g	16	八珍汤	0.784
地龙 10 g,黄芪 15 g,当归 10 g,赤芍 6 g,泽兰 6 g,红花 6 g,桃仁 10 g,胆南星 3 g,天竺黄 6 g,竹茹 10 g	8	补阳还五汤	0.846
当归 10 g,赤芍 6 g,桃仁 10 g,红花 6 g,泽兰 6 g,陈皮 6 g,法半夏 6 g,白术 6 g,茯苓 6 g	4	桃红四物汤 二陈汤	0.621 0.647
生地黄 10 g,茅根 10 g,竹叶 6 g,灯芯 2 g,当归 10 g,丹皮 6 g,甘草 6 g	6	导赤散	0.723
大黄 6 g,黄连 3 g,黄芩 3 g,栀子 10 g,银花 10 g,连翘 6 g,赤芍 6 g	4	金匱泻心汤 银翘散	0.856 0.618
猪苓 10 g,茯苓 10 g,泽泻 6 g,肉桂 6 g,白术 6 g,升麻 6 g,柴胡 6 g	7	五苓散	0.847

注:匹配率为0~1的有理数。如经方权重设置为0.3,功效权重设置为0.5,位置权重设置为0.1,剂量权重设置为0.1,一般的,其四者之和必须为1,与基本方匹配率较高的为最优结果。

痰、升提清阳、培土生金等一系列治则灵活运用到耳病的辨证施治中,时临难症,从脾胃论治,每获良效。

从本文对干老 101 个处方的聚类及其与基本方的匹配识别结果来看,在挖掘形成的 10 类大的药群中,共识别出 12 个基本方结构。其中 11 个处方中含有大补阴丸与知柏地黄汤的结构,21 个处方中含有补中益气汤的组方结构,24 个处方中含有四君子汤的结构,16 个处方中含有八珍汤的结构,8 个处方中含有补阳还五汤的结构,4 个处方中含有桃红四物汤与二陈汤的结构,6 个处方中含有导赤散的组合,4 个处方中含有金匱泻心汤与银翘散的结构。从功效来看,具有调理脾胃、补益气血功效的基本方为主的方剂结构在全部处方中所占比例较大(占全部处方数量的 60%)。这种结果无疑与干老从脾胃论治耳病,以益气升清、培土生津为主辨证施治的主要思想是一致的。

4 讨论

名医病案表述各异,繁简不一,相对而言,方剂的信息差异相对较小,尽管处方特点不同,但其组方遵循的共同规则理、法、方、药的高度统一^[12]。所谓方从法出,法随证出,方剂可以说是集理、法、方、药于一体的最终疗效载体。每一首方剂的组成既要考虑病情的需要,在辨证立法的基础上既要选择合适的药物,又要考虑各药物间配伍关系,使之达到方之既成,能使药各全其性,亦能使药各失其性的妙用。因此,以方剂为突破口分析总结名医经验特色,能更客观地挖掘名老中医理法方药的特点和规律^[13-14]。

本文首先采用基础聚类的方法对收集的干祖望教授的 101 个病案处方从整体上进行分类,有利于从群体上辨析其组方特征,然后基于模糊推理的方法对其不同类群中蕴含的基本方进行了匹配与识别,为从基本方体现的配伍规律方面探析干教授诊治耳鸣耳聋的用药特色与组方思路提供了参考。当然,由于数据挖掘用于方剂配伍规律的分析尚处于探索阶段,加之本次分析采用的样本数量有限,难免会对研究结果的可信度有所影响。加之目前对中医数据挖掘结果的分析依据依然是以传统中医理论为主,因此尚无法直接确认本研究对于名医配伍规律发现的有效性。为此,今后的研究中,我们将在进一步扩增样本量的前提下,在名老中医处方配伍规律的分析中增加基本方和症状群关系的研究,并

在条件许可的前提下,尽可能结合其他统计分析、实验和临床研究的多维度交叉验证,以进一步增强论证的强度,从而把数据挖掘方法用于名医经验分析的研究引向深入。

[参考文献]

- [1] 武红莉,荆志伟,王忠. 中药复方优化设计方法研究进展[J]. 中国实验方剂学杂志,2010,16(13):226.
- [2] 胡志帅,曹慧,马金刚,等. 计算机科学在中药信息化建设中的应用进展[J]. 中国实验方剂学杂志,2011,17(13):263.
- [3] 李敬华,尹爱宁,崔蒙,等. 中医方剂“基本方”理论及“方剂树形分析工具”的开发设计[J]. 中国中医药信息杂志,2008,15(10):89.
- [4] 孟景春. 治病宜用基本方加减化裁[J]. 江苏中医杂志,1981,2(2):6.
- [5] 干祖望. 答刘存华-谈谈基本方后的加减法问题[J]. 中国中西医结合耳鼻喉科杂志,1997,5(3):147.
- [6] 李认书,蒋永光,金桂花. 方剂中基本方的研究意义[J]. 四川中医,2001,19(8):77.
- [7] 吴嘉瑞,张冰,叶恩培. 基于数据库分析的颜正华消化系统用药总体规律研究[J]. 中国实验方剂学杂志,2012,18(2):242.
- [8] 程鹤,耿双军. 基于摄动思想的传递闭聚类法的分析与应用[J]. 兰州交通大学学报:自然科学版,2007,26(4):11.
- [9] 高彦伟,郑文瑞,韩燕,等. 复方配伍的摄动模糊聚类方法[J]. 生物数学学报,2005,20(2):219.
- [10] 任一军,赵白玉. 耳鼻喉疾病从脾胃论——干祖望学术思想介绍[J]. 中医药研究,1994(1):5.
- [11] 张世中,严道南. 干祖望运用仲景方治疗耳鼻喉科疾病小结[J]. 中医耳鼻喉科学研究杂志,2006,5(3):41.
- [12] 黄允瑜,陈慕芝,郑光,等. 基于文本挖掘技术的肥胖和高血脂症处方规律研究[J]. 中国实验方剂学杂志,2011,17(9):236.
- [13] 秦莉花,王小云,陈晓阳. 基于关联规则绝经综合征睡眠问题的研究[J]. 中国实验方剂学杂志,2010,16(5):214.
- [14] 阙文博,滕晶. 基于中医传承辅助系统的治疗失眠方剂组方规律分析[J]. 中国实验方剂学杂志,2012,18(12):1.

[责任编辑 邹晓翠]