

思路与方法

运用复杂系统熵聚类技术探讨高忠英治疗慢性呼吸系统疾病经验

史成和¹ 王秀娟² 赵宇昊² 陈建新³ 刘仁权³

(1 北京大学第三医院中医科, 北京市海淀区花园北路 49号, 100191;

2 首都医科大学中医药学院中药教研室; 3 北京中医药大学信息中心)

摘要 目的 应用数据挖掘技术研究高忠英教授治疗慢性呼吸系统常见疾病用药经验。方法 收集高忠英教授治疗慢性呼吸系统常见疾病病历 389份, 对其处方药物进行数据挖掘研究, 单味药研究采用频次统计法, 药对之间的关联使用改进的互信息, 多味药之间的关联使用复杂系统熵聚类技术。结果 高忠英教授治疗慢性呼吸系统常见疾病时使用频率较高的药物有 28味, 使用频次均在 30次以上; 常用药对 47个, 其关联系数均在 0.05以上; 常用药物核心组合 34个。结论 应用数据挖掘技术研究名老中医经验, 得出结果与临床实际相符, 对总结名老中医经验具有重要意义。

关键词 复杂系统熵聚类; 名医经验; 数据挖掘; 高忠英

首都医科大学高忠英教授系全国老中医药专家学术经验继承工作指导老师, 从事中医临床、教学、科研工作 50余年, 研习医理, 精心临证, 创立“五脏补益法”, 突破了传统的理论概念, 延伸与拓展了补法的寓义, 使其在临证时具有独特的辨证思维与用药特色, 处方精当, 疗效显著。本研究利用复杂系统熵聚类技术, 对高忠英教授的处方数据进行挖掘, 探讨其治疗慢性呼吸系统常见疾病的处方用药规律。

1 资料

1.1 纳入及排除标准

纳入标准: ①研究类型为慢性呼吸系统疾病; ②研究对象为符合临床诊断标准的慢性呼吸系统病患者, 性别、年龄、病例来源不限, 诊断标准参考公认的临床诊断标准。排除标准: ①无明确诊断的慢性呼吸系统疾病; ②兼有其他系统病证的病例。

1.2 处方资料

研究病历均来源于高忠英教授 2008年 3月~2009年 7月在北京同仁堂大栅栏门诊部、东四十条门诊部及鼓楼中医医院名医馆门诊。纳入研究共 389首处方, 所治慢性呼吸系统疾病 8种, 其中治疗慢性支气管炎 141首、过敏性鼻炎 95首、过敏性哮喘 92首、肺癌术后化疗期 22首、肺纤维化 26首、支气管扩张 6首、肺心病 5首、自发性气胸 2首。

2 方法

2.1 处方数据库的建立

由具有副主任医师及以上高级职称的高忠英教授

师承继承人 3人为核心的研究小组对处方进行分类整理。为了分析挖掘的要求, 首先要对结构化的信息进行整理。对病历中的症状、体征、检查信息形成最终术语、字段名(值)统一的信息; 对诊断、证候及治法按照教科书的标准进行统一、归纳或拆分成有相对独立含义的证素, 最终形成证候、治法词段库; 对方中所用中药进行归纳并明确其类别、功能、性味与归经。中药名以现行 21世纪规划教材为准。

以 Access平台为基础, 建立适用于高忠英教授临床诊疗信息采集的结构化病历模板。将患者信息按照国家统一标准病例书写格式全部录入数据库。详细记录门诊病历资料, 并将 389份病历的临床数据, 录入 Epidata数据库平台进行熵聚类等方法对数据库的处理, 形成病历数据库。

2.2 数据挖掘方法

从单味药、药对和多味药 3个层面对本研究的名老中医的处方数据库进行多层面和多维度的挖掘, 充分挖掘治疗慢性呼吸系统疾病的临床经验。单味药采用频次统计法, 药对之间的关联使用改进的互信息^[1], 相比于传统的互信息, 改进后的互信息能够区分正相关和负相关, 使得药对之间表达更多的正相关的信息。其核心思想是对不同时在一个处方上出现过的药对进行相关性的惩罚, 这样既不损失原本正相关的药对之间的信息, 又剔除了负相关的药对, 保证了挖掘的有效性。改进的互信息的核心公式如下。

$$\begin{aligned} & \text{燮} \text{ 牀} + \text{燮} \text{ 牀} + \text{燮} \text{ 牀} \text{ 牀} \\ & \text{燮} \text{ 牀} - \text{燮} \text{ 牀} - \text{燮} \text{ 牀} \text{ 牀} \\ & \text{燮} \text{ 牀} \end{aligned}$$

爬牀牀牀 ≥ 牀

爬牀牀牀 < 牀

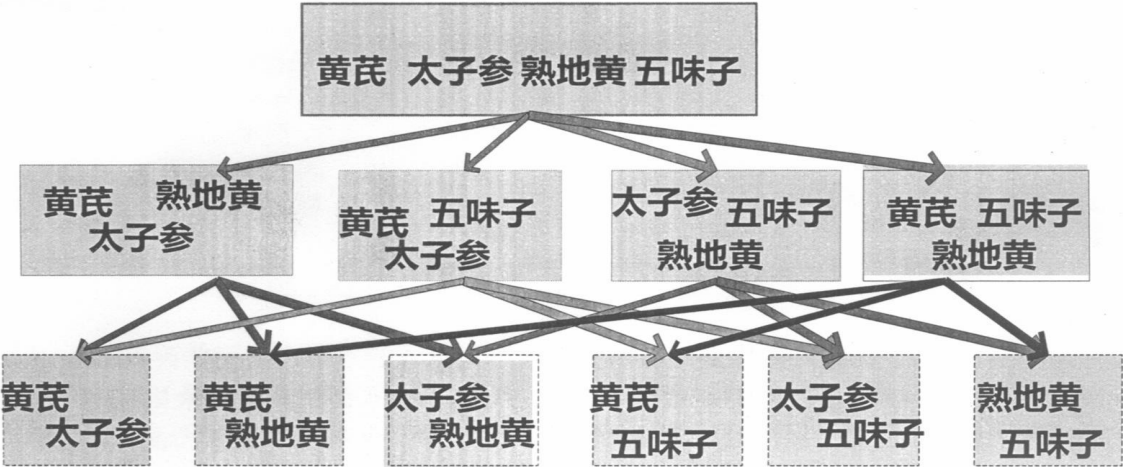


图 1 多味药之间复杂的关联使用复杂系统熵聚类技术核心技术示例

多味药之间复杂的关联使用复杂系统熵聚类,在改进的互信息的前提下,定义变量之间的强相关性,提出了聚类收敛的原则,利用互信息的良好数学性质,从两两正相关推导出三三相关,再此基础上继续聚类,同时类的个数在下降,直到收敛。其核心思路如图 1所示。

3 结果

3 1 高忠英教授治疗慢性呼吸系统疾病常用药物

通过频次统计,得到单味药物使用频率较高的药物 28味,使用频次均在 30次以上,具体药物见表 1。

表 1 高忠英教授治疗慢性呼吸系统疾病单味药使用频次			
药物	频次	药物	频次
太子参	284	杏仁	55
黄芪	270	黄芩	54
熟地黄	269	麦冬	52
五味子	265	款冬花	46
紫菀	222	补骨脂	45
桑白皮	178	麻黄	43
葶苈子	136	醋柴胡	41
瓜蒌	131	干姜	39
半夏	91	桂枝	34
桔梗	91	牛蒡子	33
生地黄	76	甘草	32
川贝母	64	鱼腥草	32
百部	56	冬瓜子	31
射干	55	乌梅	31

3 2 高忠英教授治疗慢性呼吸系统疾病常用药对

应用改进的互信息法,分析得到了 47个常用药对,其关联系数均在 0.05以上,其中改进的互信息值在 0.1 以上的药对见表 2。

3 3 高忠英教授治疗慢性呼吸系统疾病常用药物核心组合

应用复杂系统熵聚类方法,得到药物核心组合 34个,其中重要核心组合 9个,分别为①苍耳子、辛夷、醋

柴胡、五味子;②半夏、川贝母、款冬花、射干;③款冬花、射干、紫苏子、葶苈子;④射干、紫苏子、葶苈子、紫菀;⑤苍耳子、麻黄、辛夷;⑥醋柴胡、防风、乌梅;⑦胆南星、瓜蒌、黄芩;⑧熟地黄、黄芪、五味子;⑨熟地黄、太子参、五味子。

表 2 高忠英教授治疗慢性呼吸系统疾病常用药对及其互信息值

药对名称	改进的互信息值
苍耳子—辛夷	0.20823
补骨脂—芡实	0.1568
射干—葶苈子	0.14528
麻黄—杏仁	0.1384
桔梗—葶苈子	0.13568
茯苓—桂枝	0.12965
半夏—橘红	0.1181
补骨脂—葶苈子	0.11559
熟地黄—五味子	0.10894
桔梗—木蝴蝶	0.10519
桔梗—牛蒡子	0.10188

4 讨论

本研究采用 Epidata数据库平台进行熵聚类等方法进行研究,主要从 3个层面挖掘隐含在数据库背后的处方用药规律,第一个层面是单味药的,主要通过频次法实现;第二个层面是药对的分析,主要通过互信息,关联规则等,探讨治疗慢性呼吸系统疾病常用药对,初步揭示处方用药规律;最后一个层面是通过复杂系统熵聚类、聚类分析等,进行多个药物之间的组合规律分析。这 3个层面相互印证,互为补充,共同揭示高忠英教授的处方用药规律。

在单味药使用频次中,频次 30以上的药物多为教材中补肺益气及止咳平喘化痰药,其中太子参、黄芪甘温益气实卫固表,直补脾肺已虚之气;熟地黄补肾填精

以化气,补下以充上,兼以壮水润肺,济上源之虚燥;五味子、紫菀敛肺润燥,以平虚燥咳嗽,桑白皮清热降逆化痰止咳。葶苈子是泻肺平喘的代表药,也是治疗慢性呼吸系统常见疾病时出现喘证时必不可少的药物;瓜蒌是清化热痰的代表药,半夏是燥湿化痰的代表药,有痰必咳喘,咳喘多有痰,故瓜蒌、半夏是治疗慢性呼吸系统常见疾病必选药物。需要注意的是,高忠英教授创立“五脏补益法”,提出治疗肺病使用补肺汤的观点,而频次统计结果中最高的太子参,其次的黄芪、熟地黄、五味子、紫菀、桑白皮,恰好是补肺汤的重要组成药物(补肺汤的原方组成是桑白皮、地黄、人参、紫菀、黄芪、五味子)。这也证实了临床数据统计与理论和临床实际应用是一致的。

在药对使用中,苍耳子与辛夷的配伍应用改进的互信息值为 0.20823,数值最高,是用于过敏性鼻炎的有效药物组合。补骨脂与芡实改进的互信息值为 0.1568 的组合是高忠英教授的用药特色,用于慢性呼吸系统常见病证时特别是虚证的药物组合。麻黄可宣肺平喘,杏仁为止咳平喘的常用药物,麻杏的配伍为治疗咳喘病证的经典组合。通过数据统计,这两个药物组合的改进的互信息值为 0.1384 与临床实际相符。

在药物使用重要核心组合中,苍耳子、辛夷、醋柴胡、五味子为治疗过敏性鼻炎的核心组合;半夏、川贝母、款冬花、射干是治疗咳喘病证的核心组合。紫苏子、葶苈子、射干、或者款冬花、或者紫菀,本药物组合的配伍意义是治疗咳喘病证时视其痰多或咳重而设的加减配伍。黄芪、熟地黄、太子参、五味子是补肺汤的重要组成部分,因此,也是核心组合中比较明确的组合之一。

另外,如苍耳、麻黄、辛夷是治疗过敏性鼻炎时通畅鼻窍,特别是鼻眼痒证明显时而经常使用的一组药物,醋柴胡、防风、乌梅是一组抗过敏药物的组合,胆南星、黄芩与瓜蒌是热痰病证时必选的清化热痰的中药。

方剂的数据往往存在着离散度高,非线性等特点,传统统计和数据挖掘方法在针对这种数据时容易得到假阳性结果。复杂系统熵聚类是最近被提出的一种专门针对中医临床四诊信息和中药方剂数据特点的技术^[1],它是一类无监督的数据挖掘方法,针对分级的变量(如四诊信息,中药等)之间的非线性关系,在离散的信息中尽量集中有用的信息,剔除无效的信息,自组织对变量进行聚类,自动确定类中变量的个数。同时,此技术能够实现一个变量在不同的类里面出现,符合中医药的基本理论。此技术在中医临床和中药新药处方发现等已经进行了探索,基本符合临床实际^[2-3]。本研究表明,数据统计与临床应用符合,进一步证实应用熵聚类的方法研究名老中医经验有着实际意义。

参考文献

- [1]陈建新,陈静,西广成,等.复杂系统熵聚堆算法及其在中风病临床的应用[J].系统仿真学报,2008,20(15):4155—4162
- [2]王阶,邢雁伟,陈建新,等.复杂系统熵聚堆方法对1069例冠心病心绞痛证候要素提取和证组组合规律研究[J].中国中医基础医学杂志,2008,14(3):211—213
- [3]唐仕欢,陈建新,杨洪军,等.基于复杂系统熵聚类方法的中药新药处方发现研究思路[J].世界科学技术——中医药现代化,2009,11(2):225—228

收稿日期:2009-11-23 修回日期:2009-12-10

编辑:崔京艳]

脉平片临床应用有奖征文活动通知

脉平片是治疗冠心病心绞痛、高脂血症、脑血管疾病的药物,具有扩冠、防栓、调脂、保肝的作用,是国家基本医疗保险目录品种,在临床上获得了广泛的应用。为进一步总结和评估脉平片的临床疗效,交流临床用药经验,指导临床合理用药,中医杂志社与甘肃独一味生物制药股份有限公司联合举办“脉平片临床应用有奖征文活动”。

1. 征文内容: (1)脉平片治疗冠心病心绞痛的临床疗效和安全性观察; (2)脉平片治疗高脂血症的疗效和安全性观察; (3)脉平片治疗脑血管疾病的疗效和安全性观察; (4)脉平片与其他药物联合应用临床经验及疗效观察; (5)脉平片药物经济学研究; (6)脉平片临床应用新经验和体会等。2. 征文要求: (1)征文应符合《中医杂志》稿约规范(参照《中医杂志》稿约),未在公开发行的期刊上发表过; (2)附中、英文标题及500左右字中文摘要; (3)标注中、英文关键词; (4)请附打印稿及电子文档(Word文档); (5)请注明作者姓名、单位、职称、地址、邮编、电子信箱及联系电话。3. 评奖方法: (1)由全国有关专家组成脉平片征文活动评委会,召开论文评审会,评奖时间为2010年4月,将于2010年5月在《中医杂志》公布评选结果; (2)获一等奖论文将推荐在《中医杂志》正刊发表; (3)优秀论文收入2010年《中医杂志》增刊。4. 奖励办法: 设一等奖3名,二等奖7名,三等奖20名,分别奖励奖金3000元、2000元、1000元。奖励仅限第一作者。凡符合征文要求的论文作者都将获得纪念品。5. 征文截止日期: 2010年4月10日

投稿地址:北京市东直门内南小街16号《中医杂志》编辑部20号编辑收 请在信封上注明“脉平片征文”字样。电话:(010)64035632 传真:(010)64050201 邮编:100700 电子信箱:jtm@163.com

Copyright © 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net
本次活动的最终解释权归甘肃独一味生物制药股份有限公司和中医杂志社。