

DOI: 10.13288/j.11-2166/r.2014.02.007

应用数据挖掘技术分析张代钊对肺癌病机的认识

崔慧娟^{*}, 邹超², 王红岩²

(1. 中日友好医院中西医结合肿瘤科, 北京市朝阳区樱花东路 2 号, 100029; 2. 北京中医药大学)

基金项目: 国家中医药管理局 2010 年全国名老中医药专家传承工作室建设项目

摘要 应用数据挖掘技术, 建立张代钊教授临床诊疗信息采集电子信息化模板。采集 145 例肺癌患者门诊病历 324 诊次, 确定每个病例的证候类型, 形成病历数据库。通过无尺度网络提供的网络图对病例进行分析, 发现气滞血瘀是肺癌的中心病机, 痰湿热毒、阴虚、气虚是中间病变, 各脏腑功能失衡是最初的病理基础。肺癌患者以虚证为主, 气虚证候为 39.1%, 阴虚证候为 23.2%, 脾虚证候为 6.6%; 实证中痰湿证候为 10.1%, 血瘀证候为 8.9%, 气滞证候为 7.0%, 痰热证候为 5.1%。

关键词 肺癌; 数据挖掘; 病机; 张代钊

数据挖掘是由统计学、人工智能、数据库、可视化技术等多个领域相融合而形成的一个交叉学科, 可以辅助医务人员快速准确地诊断、预测、监控和确定最优的治疗方案。中日友好医院张代钊教授是首批全国老中医药专家学术经验继承工作指导老师, 从事中医及中西医结合防治肿瘤工作 50 余年, 在中医中药治疗肺癌方面有丰富经验。本研究对张代钊教授治疗肺癌的病例资料进行数据处理, 应用数据挖掘技术进行分析。

1 研究对象

1.1 临床资料

建立临床诊疗信息采集电子信息化模板, 采集并录入自 2008 年 10 月至 2011 年 8 月在中日友好医院中西医结合肿瘤内科由张代钊诊治的 145 例肺癌患者门诊病历 324 诊次。

1.2 诊断标准

参照《现代肿瘤学》^[1]中原发性支气管肺癌的诊断标准。

1.3 病历纳入标准

患者姓名、性别、年龄等基本信息完整, 有完整的中医药治疗处方, 至少有一次复诊。初诊病历有主诉、现病史、刻下症、诊断、治法及用药情况, 复诊病历有体现治疗效果描述等。

2 研究方法

2.1 资料处理

首先根据病历完善患者资料, 确定每个病例的证候类型。证候类型标准参照《中医内科学》^[2]中肺癌的中医辨证, 包括气虚痰湿证、阴虚内热证、气阴两虚证、气滞血瘀证、热毒炽盛证。在此基础上, 根据中国中医科学院中医药信息研究所编写的《中医药主题词表》和《中医药标准表》结合患者症状, 对证候类型的复合证型进行分解, 形成单一证型。部分病例在就诊时已由张代钊教授确定证候类型, 未确定证型的病例由张代钊教授继承人参考以上标准结合病历提供的四诊资料确定, 并取得张代钊教授认可。

2.2 建立模板数据

采集模板包括门诊病例基本信息界面, 病史采集界面, 舌、脉诊界面, 中西医诊断界面, 治则治法界面, 中药处方界面等。数据录入由专人负责, 按病历采集的数据项逐项录入, 包括患者基本信息、主诉、现病史、既往史、个人史、过敏史等, 以及中药处方、医嘱等信息。

2.3 质量控制

录入员录入病例资料后, 由中国中医科学院中医药信息研究所课题组专门质控人员审核, 保证重要的临床信息全部结构化, 如主症、舌脉、重要的检查、辨证要点、中医诊断 (包括中医疾病诊断和中医证候诊断)、西医诊断、治法、主方、用

* 通讯作者: cuihj1963@sina.com, 13911835018

药、剂量等。再将全部文档病历从“名老中医临床信息采集系统”中导出，打印成纸质文档，由中国中医科学院中医药信息研究所课题组复审，并对患者编码、诊次、性别、年龄、诊断、治法等信息进行审核，保证患者一人一码，上述信息无误后，形成病历数据库，导入数据仓库以备分析。

2.4 数据整理及挖掘

数据整理过程由专人负责,并由中国中医科学院中医药信息研究所课题组质控人员审核后,由课题组专人用数据的抽取—转换—装载(Extraction—Transformation—loading, ETL)软件,导入数据仓库中,形成用于分析的“最终术语”及“字段名及值统一”信息。根据基于图论的无尺度网络分析数据挖掘方法^[3-4],应用中国中医科学院开发的中医复杂关系分析系统 V2.0 软件进行证型关系分析。

3 结果

3.1 一般情况

患者共 145 例, 其中男性 81 例占 55.86%; 女性 64 例占 44.14%。年龄 37 ~ 84 岁, 平均年龄 (59.16 ± 12.27) 岁。病理诊断: 小细胞癌 18 例占 12.41%; 非小细胞癌共 127 例, 其中腺癌 61 例占 42.07%, 鳞状细胞癌 41 例占 28.28%, 肺泡癌 6 例占 4.14%, 其他 19 例占 13.10%。

3.2 辨证分型

145 例肺癌患者证候类型频次共计 527，气虚证为 206 (39.1%)，其中脾气虚 72 (13.7%)，肺气虚 71 (13.5%)；阴虚证为 122 (23.2%)，其中肺阴虚 43 (8.2%)；脾虚证为 35 (6.6%)。实证中痰湿证为 53 (10.1%)，血瘀证 47 (8.9%)，气滞证 37 (7.0%)，痰热证 27 (5.1%)。按照虚实辨证，虚证共为 68.9%，实证为 31.1%，可见肺癌患者以虚证为主。

3.3 关联性分析

3.3.1 关联密切证候类型间的关联度 图1示, 阴虚证与气虚证互见频率为58, 气滞证与血瘀证互见频率为37。

3.3.2 所有证型关联性分析 图2示,外围证候为阴虚(包括肺阴虚、肾阴虚、胃阴虚)、气虚(包括脾气虚、肺气虚、脾虚)、肾阳虚和热证(包括胃热、肺热、痰热、湿热、肝阳上亢);第二层为气虚、阴虚、湿阻、脾胃不和;中心为气滞

血瘀。提示气滞血瘀是肺癌的中心病机，痰湿热毒、阴虚、气虚是中间病变，各脏腑功能失衡是最初的病理基础。

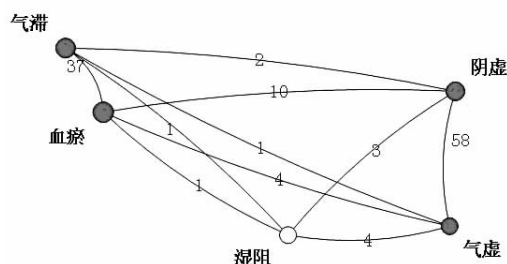


图 1 145 例肺癌患者关联密切证候类型间的关联度

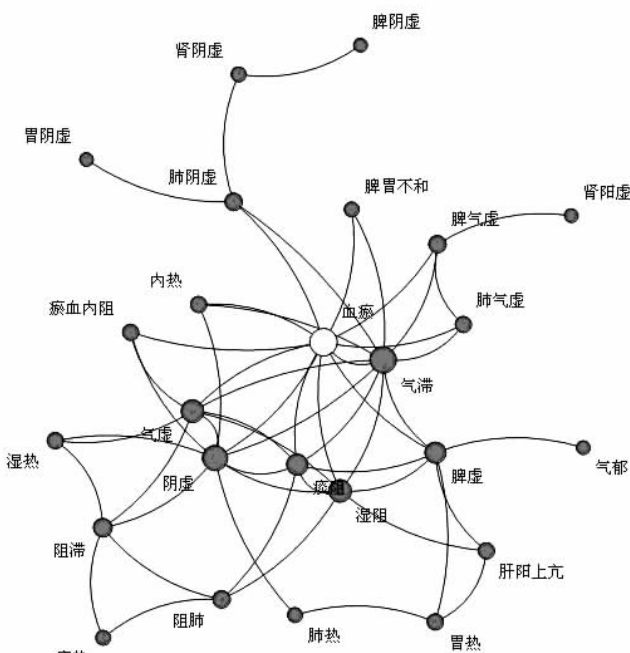


图2 145例肺癌患者肺癌病机的无尺度网络结构

4 讨论

数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中提取出隐含在其中的可信、新颖、人们事先不知道但又是潜在有用的信息的高级处理过程^[5]。近年来,数据挖掘被用于对名医经验的研究^[6-7],研究结果基本反映了名医临床经验和学术思想。

数据挖掘有许多实施技术，其中无尺度网络作为复杂系统研究的一种实现现象和方法，用于研究复杂现象和复杂系统。无尺度网络是一种具有节点度幂律分布现象的复杂网络。把每个患者每次的证候作为节点，连接某两个不同证候的边权重表示这两种证候出现的频度及其相互关系，这样就构建了网络。随着患者的增加，网络的节点和边权重会逐步增加。网络中心是出现最多的事件，周边则是出

现较少的事件。数据关联技术用于研究数据库中存在的一个或多个变量的取值之间可能存在的某种规律性^[8]。在中医证候的研究中,由大量的中医症状、舌脉表现组成的数据库相当繁杂,关联规则可以找出相关联的各个数据。当某些症状总是同时出现时,可以从中找出某种证型规律^[9],也能够用来探讨容易同时出现的证候之间的相关性,反映疾病层面的指标和证候的关联。若关联规则加入时间因素,还可以体现证候的动态演变。数据挖掘技术在建立新型中医诊断系统的研究^[10]、中医证候诊断标准规范化研究^[11]中得到运用。

本研究从数据挖掘得到的资料来看,肺癌患者以虚证为主,为 68.9%,实证为 31.1%。无尺度网络结构图显示,气滞血瘀是肺癌的中心病机,痰湿热毒、阴虚、气虚是中间病变,各脏腑功能失衡是最初的病理基础。反过来讲,最初的病因为各个脏腑的功能失调,由此导致痰湿热毒、阴虚、气虚,其结果是气滞血瘀。

张代钊教授认为^[12],肺癌发病原因主要与正气虚损有关,正气虚贯穿肺癌的整个发病过程,尤以中晚期肺癌最为明显。正气内虚、脏腑阴阳失调是肺癌发病的基础。正气不足,气血阴阳失衡,脏腑功能失调,使机体抗病能力下降,邪气乘虚而入。邪气入内,留滞不去,阻于胸中,宣降失常,气机不畅,气滞血瘀,阻塞脉络,津液输布不利,聚而为痰,痰瘀胶结,从而形成肿块。早期肺癌患者往往有气虚表现,多见乏力;中期部分患者出现阴虚合并气虚,兼有内热症状,部分患者出现痰湿阻肺的表现;晚期瘀毒渐显,伴有疼痛等症状,肺阴虚也渐渐发展至肾阴虚,导致肺肾阴虚,最终出现气血双亏、阴阳俱虚。

气虚、阴虚及气阴两虚为肺癌的主要分型。肺为娇脏,不耐寒热,喜润恶燥,易为燥伤。烟毒秽气,邪热耗津,房劳伤肾,均可致肺阴不足,故肺脏受病,必有伤阴。《理虚元鉴》曰“阴虚之证统于肺,阳虚之证统于脾。”《医门法律·虚劳门》亦言“阴虚者,十常八九,阳虚者,十之一二”。阴伤化热,内外邪热,更伤津液,津亏血少,脉道失充血行郁滞而致血瘀;阴伤化热,亦炼津为痰,痰阻气滞,气滞血瘀,日久成积。痰阻血瘀,郁而

化热,更伤阴液,互为因果,形成恶性循环。同时肺癌患者大多屡经手术、放疗、化疗,手术中失血、化疗中剧烈呕吐、利尿均可致津血亏乏,加重阴伤,而放射治疗更是“大热峻剂”^[13],耗伤人体阴液。因此,肺癌患者多阴虚、多气虚。

综上所述,数据挖掘技术得出的结论基本反映了张代钊教授对肺癌病机的认识,也基本与中医学文献记载一致。

参考文献

- [1] 汤钊猷. 现代肿瘤学 [M]. 3 版. 上海: 复旦大学出版社, 2011: 1078.
- [2] 田德禄. 中医内科学 [M]. 北京: 人民卫生出版社, 2002: 382.
- [3] 刘保延, 胡镜清, 谢雁鸣, 等. 中医药学现代个体诊疗体系建立的构想与研究 [J]. 世界科学技术—中药现代化, 2003, 5(1): 1-6.
- [4] 刘保延, 李平. 立足科技前沿构建现代化的中医临床研究技术平台 [J]. 亚太传统医药, 2005, 1(1): 45-48.
- [5] Jiawei Han, Micheline Kamber. 数据挖掘概念与技术 [M]. 范明, 孟小峰, 译. 北京: 电子工业出版社, 2001: 3-5.
- [6] 史成和, 王秀娟, 赵宇昊, 等. 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 [J]. 中医杂志, 2010, 51(3): 268-270.
- [7] 刘铁钢, 陈建新, 张望, 等. 基于复杂系统熵方法的孔一光教授治疗月经病经验数据挖掘研究 [J]. 中医杂志, 2013, 54(6): 481-483.
- [8] 劳飞. 数据挖掘技术在医学中的应用 [J]. 实用心脑血管病杂志, 2010, 18(6): 823.
- [9] 龚燕冰, 倪青, 王永炎. 中医证候研究的现代方法学述评(一): 中医证候数据挖掘技术 [J]. 北京中医药大学学报, 2006, 29(12): 797-801.
- [10] 龚德平, 高颖, 唐涛. 基于数据挖掘的数字化中医诊断系统 [J]. 中国医学影像技术, 2003, 19(204): 132-134.
- [11] 戴霞, 郭伟星. 中医证候诊断标准规范化研究概况 [J]. 中医杂志, 2011, 52(2): 168-171.
- [12] 崔慧娟, 张培宇. 张代钊教授治疗肺癌经验 [J]. 中日友好医院学报, 2011, 25(1): 57-58.
- [13] 赵炜. 李佩文教授治疗肺癌经验介绍 [J]. 新中医, 2003, 35(1): 9-10.

Professor ZHANG Daizhao's Knowledge of the Pathogenesis of Lung Cancer Based on Data Mining Technology

CUI Huijuan¹, ZOU Chao², WANG Hongyan²

(1. China-Japan Friendship Hospital, Beijing 100029; 2. Beijing University of Chinese Medicine)

ABSTRACT The clinical information gathering electronic template of Professor ZHANG Daizhao was established. Totally 145 cases of lung cancer were collected and the pattern types were determined to establish a medical record database. Qi stagnation blood stasis is the main pathogenesis, phlegm-damp-heat-toxin, yin deficiency and qi deficiency are the intermediate lesions and organ function imbalance is the initial pathologic basis of lung cancer according to the network diagram provided by the scale-free networks. The lung cancer patients were mainly with deficiency pattern and qi deficiency pattern, yin deficiency pattern and spleen deficiency pattern accounted for 39.2%, 23.2% and 6.6% respectively. There were 10.1% phlegm damp pattern, 8.9% blood stasis pattern, 7.0% qi stagnation pattern and 5.1% phlegm heat pattern in the repletion patterns.

Keywords data mining; lung cancer; pathogenesis; ZHANG Daizhao

(收稿日期: 2013-03-21; 修回日期: 2013-09-23)

编辑: 焦爽

中国第八届中医、中西医结合暨非药物疗法防治 心脑血管病、糖尿病、肿瘤高级论坛通知

由世界中医药学会联合会国际合作部、《世界中医药》杂志社联合主办的“中国第八届中医、中西医结合暨非药物疗法防治心脑血管病、糖尿病、肿瘤高级论坛”将于 2014 年 5 月 24 至 27 日在杭州召开。

1. 会议征文与研讨内容

(1) 中医药治疗心脑血管病的理论研究及临床经验; (2) 中医、中西医结合治疗心脑血管病的新疗法、新进展; (3) 中医药治疗心脑血管病的实验研究; (4) 针灸推拿等非药物疗法治疗心脑血管病的体会; (5) 中医、中西医结合治疗糖尿病的理论研究及临床进展; (6) 糖尿病、肿瘤及心脑血管病的中医诊疗规范; (7) 中医药、中西医结合治疗肿瘤的基础及临床研究; (8) 中医、中西医结合抗肿瘤的理论与临床探讨; (9) 名老中医经验和民间中医验方防治心脑血管病、糖尿病、肿瘤经验总结; (10) 抗肿瘤中药的研究与开发; (11) 肿瘤止痛及晚期肿瘤患者生存质量与养生方法研究; (12) 中医非药物疗法的基础研究和理论研究探讨; (13) 中医非药物疗法在临床中的应用研究与总结; (14) 针灸在各科临床的经验总结; (15) 推拿按摩及其他理疗方法在各科临床的运用; (16) 非药物康复治疗经验总结; (17) 中风后遗症的中医特色治疗经验; (18) 老年痴呆症、顽固性失眠、血管性头痛的论治精神及特色疗法; (19) 名老中医临床经验总结; (20) 中医成才经验介绍; (21) 临床应用经方、成方、验方的体会; (22) 中医养生保健、药膳食疗方法介绍及理论研究; (23) 中医药教学经验介绍; (24) 中国医学史研究; (25) 中药、民族药研究及其现代应用; (26) 复方制剂的开发与研究; (27) 民族医经典药方及常用方剂的整理研究; (28) 民族药品种和炮制研究; (29) 民族医特色疗法的研究与应用; (30) 民族医诊脉方法研究。

2. 征文要求: (1) 论文要求论点明确、论据充分、结构严密、层次分明、文字精炼,具有科学性、创新性、实用性; (2) 征文请附医院或主管单位介绍信; (3) 征文截止日期为 2013 年 5 月 10 日; (4) 每位作者只限投一篇论文(要求 2 000 字以内); (5) 文稿中的计量单位均须采用法定计量单位; (6) 文题下署作者姓名,作者单位名称、地址、邮政编码、电话请写在文稿末尾,请在投稿信封左下角注明“会议”字样; (9) 征文经专家评审后,录用文章将刊登在《世界中医药》杂志 2014 年增刊(第一、第二作者分别授予国家继续教育学分),并邀请参加会议,入选论文安排在大会交流,参加会议者颁发证书并授予国家 II 类继续教育 10 学分。投稿邮箱: E-mail: 13901318403@139.com。

3. 会议日程: 2013 年 5 月 24 日报到; 2013 年 5 月 25 至 5 月 26 日专题演讲、大会交流、分组交流、参观考察; 2013 年 5 月 27 日中午 12 点代表离会。

4. 会议地址: 浙江省军区涌金饭店, 杭州市上城区劳动路 127 号。

5. 注册费用: 每位代表会务费 1280 元(含餐费、场地租用费、专家讲课费、纪念品、资料、会议合影、参观交通费等); 提前 2 个月交费者优惠 10%。

6. 地址与会议回执: 北京市朝阳区小营路 19 号财富嘉园 A 座 301 室, 世界中医药学会联合会国际合作部《世界中医药》杂志社, 邮编: 100101, 电话: (010) 58239055; 传真: (010) 58239066, 手机: 13901318403, 联系人: 王炳岐、赵和爱、徐晖。