

基于复杂网络 AECOPD 证候分布 及中药组合研究^{*}

胡金亮, 李素云, 张兴红, 余海滨^{**}, 崔国荣, 李 真

(河南中医学院第一附属医院 郑州 450000)

摘 要 基于临床医疗科研信息共享系统, 利用复杂网络开展河南中医学院第一附属医院慢阻肺基地慢性阻塞性肺疾病急性加重期(AECOPD)证候分布及中药组合研究, 分析研究河南中医学院第一附属医院慢阻肺基地 AECOPD 诊疗常见证候分布规律及中药核心药组组合特点。方法是利用临床医疗科研信息共享系统, 开展河南中医学院第一附属医院 AECOPD 数据研究(2012 年 8 月-2015 年 1 月 3 000 份格式化电子病历), 通过 ETL 软件进行数据处理, 构建数据库, 运用复杂网络进行数据挖掘, 获取河南中医学院第一附属医院 AECOPD 数据常见证候分布及中药药组组合研究资料。挖掘结果是常见证候分布为: 痰热壅肺 41.94%、痰浊阻肺 22.97%、痰湿阻肺 10.30%、气阴两虚 6.31%、痰瘀阻肺 5.72%、肺肾气虚 4.01%、肺脾气虚 1.15%、痰蒙神窍 1.15%。中药组合明显分 2 大组合, 每一组合有核心中药组和关联中药组。基于以上结果分析, 临床医疗科研信息共享系统, 有很好的联想分析归类作用, 不仅可以分析老中医专家个人的及单个医院单个病种的临床治疗用药规律, 而且可用于大数据(全国性中医临床研究基地单病种及多病种的数据汇总)的分析研究, 为中医的临床科研研究提供一个很好的平台。

关键词 临床医疗科研信息共享系统 慢阻肺急性加重期 证候分布 中药组合 复杂网络

doi:10.11842/wst.2015.06.026 中图分类号: R259 文献标识码: A

当今大规模生产、分享和应用数据的时代正在开启。大数据的来临使人类有机会和条件在较多的领域和较深入的层次获得和使用全面数据、完整数据和系统数据, 深入探索现实世界的规律^[1]。近年来, 智能化的信息科技手段大量应用于生物信息学领域和临床医学领域, 基于数据库的知识发现(Knowledge Discovery in Database, KDD)及相应的数据挖掘(Data Mining)理论和技术成为近年来研究的热点, 尤其数据挖掘应用于中医学的交叉研究已取得了不少成果^[2-7]。

在 20 世纪即将结束之际, 对复杂网络的科学探

索发生重要的转变, 复杂网络从规范化网络、随机网络研究演变发展而来, 从实际网络的特征结合无向网络的静态特征(联合度分布、度度相关性、聚集系数分布、聚一度相关性、介一度相关性、节点度中心性、网络度中心性等)、有向网络的静态特征、加权网络静态特征及其它静态特征的综合研究, 基于节点关联性的方法、最短路径的方法、模拟流的方法及其它分析方法等节点挖掘方法及社团挖掘方法, 研究复杂网络的机制模型、复杂网络上的传播动力学、复杂网络的混沌同步、复杂网络中的搜索、复杂网络的博弈等。以其复杂的网络结构形式研究复杂的系统, 利用其研究成果上升到现实社会和自然界的复杂环境、空间研究, 探索其中的奥秘, 其理

收稿日期 2015-03-15

修回日期 2015-05-04

^{*} 国家中医药管理局 2012 中医药行业科研专项(HY12079675)“全国中医医疗与临床科研信息共享推广与应用研究, 负责人 李真。

^{**} 通讯作者 余海滨, 副主任医师, 主要研究方向: 基于数据挖掘中医诊疗研究。

论不再局限于数学领域。从物理学到生物学的众多学科中掀起了复杂网络的热潮,网络中“物以类聚,人以群分”的特性,可能形成一种内聚的结构^[8]。复杂网络的研究为我们中医复杂的概念、机理、辨证、诊治规律的研究提供了新视角、新方法。本研究对象为河南中医学院第一附属医院临床慢性阻塞性肺疾病急性加重期(Acute Exacerbation Chronic Obstructive Pulmonary Disease, AECOPD)住院患者格式化电子病历数据(2012年8月至2015年1月3 000余患者数据),利用中医临床医疗科研信息一体化系统(临床智能系统)进行信息数据处理、分析研究。

1 复杂网络理论中医应用研究

1.1 复杂网络理论中医病证症药基础理论研究

“中医复杂关系分析系统”就是复杂网络在中医临床知识发现中的应用。利用丰富的中医临床数据,从复杂网络的角度对中医证候的基本规律进行分析研究,尝试形成对证候学研究的复杂网络特性解释。证候特征研究,如不同地区疾病证候的分布特点^[9];病-证关系研究,如某疾病的不同时候、不是并发症的证候分布及相兼规律研究;病证结合的疾病分类特征研究;证候诊断标准研究;证候疗效判定标准研究;临床症状、症状组合、体征、检查指标与证候及证候变化的关系研究;证候与预后的关系研究;证候基础理论研究及基础理论验证研究。疾病特征研究,从疾病不同病程、年龄段并发症与合并病进行比较分析,研究其症状,在中医复杂网络特性分析的基础上,利用相应的社团分析方法进行研究,复方药物配伍规律(如君臣佐使)的层次性进行分析研究,发现核心和一般共性的复方药物配伍知识和症状表现组合知识^[10,11]。

1.2 复杂网络理论名老中医临床诊疗应用研究

基于现代信息技术“中医复杂关系分析系统”,开展名老中医诊疗经验的传承研究,提高名老中医经验传承研究的效率及水平。建立面向分析的名老中医临床信息数据仓库。利用结构化名老中医临床诊疗信息采集系统,实时、动态、全面采集名老中医临床诊疗信息,在保留原始病历文本的情况下,将其转化为可分析的结构化数据;利用多维检索分析查询系统,从病-症-证-治-方-药等各个角度对名老中医经验进行分析,对名老中医辨证、选方、用

药、疗效等方面的关系进行分析,迅速发现显性知识;通过复杂网络方法等多种数据分析挖掘方法,对名老中医经验中的隐性知识进行发现,梳理名老中医经验的辨证特点、核心处方、加减变化规律,进行多位名老中医经验间的比较研究,对较成熟的临证经验开展验证研究^[12,13]。

2 基于复杂网络的AECOPD证候分布研究

2.1 数据的收集、预处理

数据的收集预处理主要是针对结构化的电子病历的数据及 HIS、LIS、PACKS 等数据,通过 ETL 导入 ODS 相应的表中。进行结构化数据的转换、核查后再用 ETL 导入细节数据库。结构化信息进行进一步规范,从内容上讲,一般包括症状类信息、检查类信息、诊断类信息、治法方药类信息等建立规则库。处理模式上,包括拆分、转化及补充上下位关系等多方面。构建 7 大规则库:①诊断类信息的预处理;②治则治法整理;③症状类数据整理;④方剂类信息的预处理;⑤药物类的预处理;⑥检查类信息的预处理;⑦针灸穴位类的整理等。

2.2 数据挖掘

基于复杂网络河南中医学院第一附属医院患者慢性阻塞性肺疾病急性加重证候分布研究见图 1。基于复杂网络河南中医学院第一附属医院慢阻肺病人慢性阻塞性肺疾病急性加重证候分布数据见表 1。证候分布(选取 1% 以上的证候):痰热壅肺 41.94%、痰浊阻肺 22.97%、气阴两虚 6.31%、肺肾

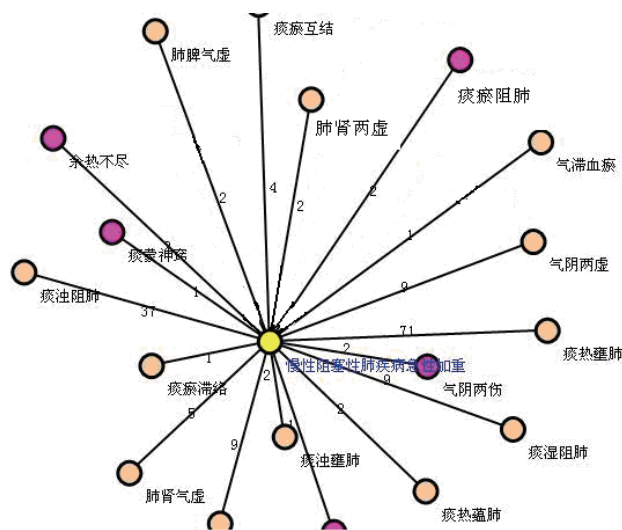


图 1 河南中医学院第一附属医院慢阻肺患者慢性阻塞性肺疾病急性加重证候分布特点

表 1 河南中医学院第一附属医院慢阻肺急性加重期
证候分布数据

节点一	节点二	频度	MISCR
肺脾气虚	慢性阻塞性肺疾病急性加重	2	5.577 0
肺肾两虚	慢性阻塞性肺疾病急性加重	2	8.015 9
肺肾气虚	慢性阻塞性肺疾病急性加重	5	20.052 7
气阴两伤	慢性阻塞性肺疾病急性加重	2	7.198 6
气阴两虚	慢性阻塞性肺疾病急性加重	9	47.125 7
气滞血瘀	慢性阻塞性肺疾病急性加重	1	3.489 1
痰蒙清窍	慢性阻塞性肺疾病急性加重	1	1.869 4
痰蒙神窍	慢性阻塞性肺疾病急性加重	1	1.294 0
痰热蕴肺	慢性阻塞性肺疾病急性加重	2	6.128 0
痰热壅肺	慢性阻塞性肺疾病急性加重	71	765.960 9
痰湿阻肺	慢性阻塞性肺疾病急性加重	9	48.611 4
痰湿壅肺	慢性阻塞性肺疾病急性加重	9	49.467 7
痰浊蕴肺	慢性阻塞性肺疾病急性加重	1	2.371 0
痰浊阻肺	慢性阻塞性肺疾病急性加重	37	270.374 3
痰浊壅肺	慢性阻塞性肺疾病急性加重	2	6.128 0
痰瘀互结	慢性阻塞性肺疾病急性加重	4	12.414 0
痰瘀内阻	慢性阻塞性肺疾病急性加重	2	6.856 9
痰瘀滞络	慢性阻塞性肺疾病急性加重	1	3.489 1
痰瘀阻肺	慢性阻塞性肺疾病急性加重	3	9.251 7

气虚 4.01%、痰瘀阻肺 5.72%、肺脾气虚 1.15%、痰湿阻肺 10.30%、痰蒙神窍 1.15%。

从以上结果看出,慢性阻塞性肺疾病急性加重期证候频度从大到小排序,痰热壅肺证最多,其次是痰浊阻肺证,依次为气阴两虚证、肺肾气虚证、痰瘀阻肺证、痰蒙神窍证等,证候与慢性阻塞性肺疾病急性加重 MISCR 值大小同频度大小相关,说明 AECOPD 期实证痰热壅肺证、痰浊阻肺证为多、虚证以气阴两虚证、肺肾气虚证为多,此病多瘀,久者则痰瘀阻肺,虚实兼证。危重期多伴痰蒙神窍证,治疗上要注意辨证治疗,治法补虚泻实,醒神开窍等。以上数据为河南中医学院第一附属医院慢阻肺基地慢阻肺病人数据挖掘结果。不是专门的流行病学调查中的发病率。

中药药组组合分析见图 2。中药药组组合结果见表 2。

从以上结果看出,红花、紫草、皂刺、赤芍、冰片等配伍几率最大,说明活血化瘀组频次最多,考虑肺病科中药离子导入治疗用药多,效果好,它是肺病科中医特色疗法。其次,另一核心组合药物分布是川贝、太子参、白术、地龙、款冬花等说明

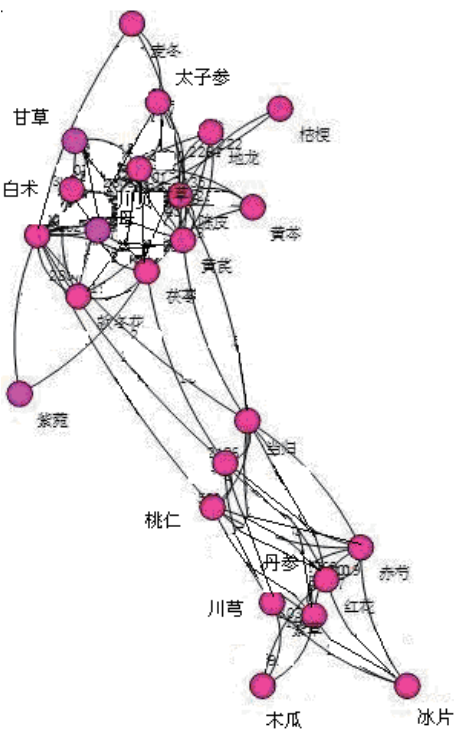


图 2 中药药组组合

AECOPD 补虚邪实止咳平喘,多联合桔梗、紫菀、黄芪、茯苓等相关药物组,揭示了河南中医学院第一附属医院慢阻肺临床研究基地治疗 AECOPD 治疗用药的中药配伍规律。

3 讨论

以上结果说明河南中医学院第一附属医院慢阻肺临床基地临床所见慢性阻塞性肺疾病急性加重证候多为痰、瘀、虚,慢阻肺易合并肺性脑病,外感证如风寒束肺、风热袭肺等有,但少见,故治疗上多偏于治痰、治瘀、治虚、调理神智等。

中药组合分析从图中及数据结果中看出, AECOPD 的临床中药应用特点大致皂角刺、紫草、赤芍、红花、冰片、木瓜为一组配伍的几率最大,为核心中药组,多与川芎、当归、丹参、木瓜、冰片相关药物组等配伍联合,这些药多为中医临床特色疗法中用药配伍,突出理气活血化痰通络的特点,另外一组以太子参、白术、川贝、甘草、陈皮、半夏、款冬花、地龙、蝉蜕等为核心的药物组,其次与麦冬、桔梗、紫菀、麦冬、黄芪、黄芩、茯苓等配伍,多为中药治疗 AECOPD 口服用药,突出补虚(阴阳气血五脏)化痰,二者组合治疗补虚化痰化瘀。从中药组图中看出,红花、紫草、皂角刺等活血化瘀核心药物组紧密度更高,其次

表 2 中药药组组合结果

(续表 2)

节点一	节点二	频度	MISCR
白术	陈皮	3 149	29.265 5
陈皮	太子参	2 680	18.847 2
陈皮	麦冬	1 950	9.388 5
赤芍	皂角刺	3 019	125.783 2
赤芍	当归	2 930	57.885 4
赤芍	紫草	2 769	139.285 7
赤芍	冰片	2 012	73.588 9
川贝母	甘草	4 784	56.388 4
川贝母	白术	3 768	35.827 9
川贝母	法半夏	3 720	37.259 4
川贝母	太子参	3 556	31.461 5
川贝母	款冬花	2 807	23.805 2
川贝母	桔梗	2 701	19.073 9
川贝母	紫菀	2 397	16.906 7
川芎	赤芍	2 320	34.892 3
川芎	皂角刺	2 291	45.675 4
川芎	当归	2 126	18.371 0
川芎	紫草	1 947	40.246 9
丹参	赤芍	2 783	58.634 8
丹参	皂角刺	2 627	67.037 3
丹参	川芎	2 560	37.917 2
丹参	当归	2 378	25.028 8
当归	紫草	2 263	52.072 2
当归	陈皮	1 983	7.4830 8
地龙	川贝母	3 327	27.097 3
地龙	陈皮	2 264	11.949 5
地龙	太子参	1 781	9.736 4
法半夏	陈皮	3 232	34.195 9
法半夏	白术	2 388	22.097 3
法半夏	款冬花	1 838	16.188 8
甘草	陈皮	3 678	35.858 9
甘草	白术	2 602	20.342 7
甘草	法半夏	2 486	19.174 1
甘草	款冬花	2 001	14.869 6
甘草	地龙	1 890	8.565 6
甘草	太子参	1 813	7.189 3
红花	皂角刺	3 267	239.486 2
红花	紫草	3 073	285.903 7
红花	赤芍	3 010	137.835 0
红花	丹参	2 791	88.882 8
红花	当归	2 606	61.214 0
红花	川芎	2 293	50.630 1
红花	冰片	2 244	153.232 3

红花	木瓜	1 963	124.114 3
黄芩	川贝母	3 456	30.279 3
黄芩	陈皮	2 377	13.786 7
黄芩	甘草	2 135	12.308 8
黄芩	茯苓	1 820	6.765 1
黄芪	川贝母	4 075	39.129 6
黄芪	陈皮	3 481	34.133 5
黄芪	茯苓	3 263	30.733 4
黄芪	白术	2 608	22.999 1
黄芪	甘草	2 469	15.006 5
黄芪	太子参	2 310	16.696 7
黄芪	当归	2 115	12.560 2
黄芪	地龙	1 868	9.285 3
黄芪	法半夏	1 833	8.629 6
桔梗	陈皮	2 221	14.841 0
苦杏仁	川贝母	1 922	11.892 7
款冬花	茯苓	1 982	12.820 6
款冬花	陈皮	1 959	11.324 2
麦冬	川贝母	2 875	21.500 9
木瓜	皂角刺	1 915	104.087 6
木瓜	紫草	1 849	133.919 2
前胡	川贝母	1 981	12.168 0
太子参	茯苓	2 198	11.366 2
太子参	麦冬	2 181	21.986 9
皂角刺	当归	2 670	59.554 7
皂角刺	冰片	2 204	131.060 2
炙款冬花	川贝母	2 163	13.028 6
紫草	皂角刺	3 036	249.085 8
紫草	丹参	2 423	75.506 2
紫草	冰片	2 134	169.687 2
紫苏子	川贝母	3 252	28.139 5
紫苏子	茯苓	2 476	19.000 8
紫苏子	陈皮	2 379	15.417 3
紫苏子	法半夏	2 014	16.190 3
紫苏子	黄芪	1 939	11.547 6
紫苏子	白术	1 907	12.720 7
紫苏子	甘草	1 808	8.338 2
紫菀	茯苓	1 803	11.007 7
茯苓	川贝母	5 208	64.546 2
茯苓	陈皮	4 315	51.374 3
茯苓	甘草	3 483	33.289 9
茯苓	白术	3 394	40.055 7
茯苓	法半夏	3 037	31.016 1
茯苓	地龙	2 145	11.109 5

(续表 2)

茯苓	川芎	1 838	7.910 1
----	----	-------	---------

太子参等核心药组的紧密度较高,红花与皂角刺的MISCR最高,其次是红花与紫草,说明这些药物的高频次应用,川芎、当归、赤芍在活血化瘀核心药组和太子参、川贝等核心药组两组之间,说明活血化瘀药物在慢阻肺急性加重期治疗中所占比重最大,用药最多,另外说明川芎、当归、赤芍不仅用于中医特色疗法外治疗法,同时用于中药内服。以上解释了中药复杂关系的配伍规律。

“大数据”(Big Data)不仅是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。而且在于对这些含有意义的数据进行专业化处理。提高对数据的“加工能力”,通过“加工”实现数据的“增值”。大数据与一般数据最本质的区别体现在采集来源以及应用方向上。从以下几点说明:①一般数据诠释宏观、整体的情况,大数据可以分析微观、个体的情况。②一般数据挖掘方式,采集方法,内容分类,采集标准等都已存在既有规则,方法论完整;大数据挖掘为新鲜事物,还没有形成清晰的方法、路径以及评判标准。③一般数据来源于阶段性的、针对性的评估,其采样过程可能有系统误差;大数据来源于过程性的、即时性的行为与现象记录,第三方、技术型的观察采样的方式误差较小。④一般数据分析所需要的人才、专业技能以及设施设备都较为普通,易获得;大数据挖掘需要的人才、专业技能以及设施设备要求较高,并且从业者需要有创新意识与挖掘数据的灵感而不是按部就班者,这样的人才十分稀缺。

大数据与一般数据有联系。大数据是一般数据的动态的、连续的、系统信息资料的总和,没有一般数据就不会有大数据,本研究对象2012年8月至2015年1月3 000余患者信息资料不是“大数据”,是一般数据,随着以后全国共享数据中心的完善,可以成为大数据中的一部分。

“中医复杂关系分析系统”就是复杂网络在中医临床知识发现中的应用。通过对大规模中医临床数据进行网络化建模,从复杂网络的角度对中医学的基本规律如辨证论治的临床表现(四诊信息)共性和个体性的组织特性、尝试形成对中医基础理论的复杂网络特性解释。在中医复杂网络特性分析的基础上,利用相应的社团分析方法进行核心和一般共性的知识的发现研究。揭示中医临床数据中的核心共性知识,从而加深对中医临床诊疗规律的认识和理解。同时基于古方及当代临床复方数据的分析表明,中医药理论指导下的复方配伍过程具有无尺度复杂网络现象。这对中医药理论如复方配伍、药物相互作用以及药性理论等的研究提供了实证基础,为进行中医特色的科学研究提供了方法学启发^[1]。尽管本研究数据有3 000例只代表河南中医学院第一附属医院慢阻肺基地慢阻肺急性加重期的临床研究,不是临床流行病学研究结果,但随着更多的这样数据的加入,借助我们今天基地的临床科研共享系统数据中心,也会有我们中医大数据,通过中医大数据挖掘出我们中医诊疗规律,指导临床。目前全国有多家临床基地开展共享系统,前期基础已做好,但后期的工作和工作升华仍需要我们继续努力,再接再厉。

参考文献

1 Viktor Mayer-Schonberger, Kenneth Cukiercompile. Big data a revolution that will transform how we live, work, and think. Sheng Yangyan, Zhou Tao translation. Hangzhou: Zhejiang People's Publishing House, 2013:1-247.

2 胡金亮,张润顺,周雪忠,等.基于临床医疗科研信息共享系统中医病-证-中药关联分析方法的研究.世界科学技术-中医药现代化,2013,15(5):905-909.

3 胡金亮,李素云,余海滨,等.基于临床医疗科研信息共享系统AECOPD证症相关性分析研究.世界科学技术-中医药现代化,2013,15(7):1596-1599.

4 余海滨,李素云,胡金亮,等.中医临床医疗科研信息共享系统的智能整合及应用.世界科学技术-中医药现代化,2013,15(6):1480-1482.

5 胡金亮,李建生,王至婉,等.基于自适应模糊推理系统的慢性阻塞性肺疾病急性加重期证候特征研究.中医杂志,2010,51(8):733-735.

6 Hu Jinliang, Yue Caiqing, Li Jiansheng, et al. Research on Clinical Diagnostic Models of IPF Syndromes in TCM Based on dynamic Kohonen Network and Decision Tree. GCIS2010. Wuhan: Academic Press, 2010:107-110.

7 Hu Jinliang, Yue Caiqing, Li Jiansheng, et al. Clinical research of TCM common syndrome for Chronic pulmonary heart disease Based on dynamic Kohonen Network and Decision Tree. ICCME2011. Hangzhou: Academic Press, 2011:469-473.

8 汪小帆,李翔,陈关荣.复杂网络理论及其应用.北京:清华大学出版社, 2005:23-89.

- 9 Li S, Zhang Z Q, Wu L J, *et al.* Understanding ZHENG in traditional Chinese medicine in the context of neuro–endocrine–immunenetwork. *IET Syst Biol*, 2007, 1(1):51–60.
- 10 Li S, Zhang B, Jiang D, *et al.* Herb network construction and co–module analysis for uncovering the combination rule of traditional Chinese herbal formulae. *BMC Bioinformatics*, 2010, 11(S11):S6
- 11 周雪忠,刘保延,王映辉,等.复方药物配伍的复杂网络方法研究.中国中医药信息杂志,2008,15(11):98–100.
- 12 刘保延,周雪忠,张润顺,等.“医疗与临床科研信息共享系统”中医电子病历系统信息基本要求.中国数字医学,2012,7(10):57–60.
- 13 胡镜清,刘保延,王永炎.中医临床个体诊疗信息特征与数据挖掘技术应用分析.世界科学技术–中医药现代化,2004,6(1):14–16.

Study on Syndrome Distribution and Chinese Medicine Composition of AECOPD Based on Complex Networks

Hu Jinliang, Li Suyun, Zhang Xinghong, Yu Haibin, Cui Guorong, Li Zhen

(The First Affiliated Hospital of Henan College of Traditional Chinese Medicine, Zhengzhou 450000, China)

Abstract: Based on clinical research and information–sharing system, this article was aimed to study both the syndrome distribution and Chinese medicine composition of acute exacerbation of chronic obstructive pulmonary disease (AECOPD) in the COPD base of our hospital using the complex networks. The analysis was made on the common rules of syndrome distribution and features of key Chinese medicine compositions in the diagnosis and treatment of AECOPD. The clinical research and information–sharing system was used on the AECOPD data study (3000 hospital electronic medical records from August 2012 to January 2015). Data was processed by ETL software. The database was constructed. And the complex networks were used in the data mining. The results showed that the distribution of AECOPD common syndromes was heat–phlegm obstructing lung of 41.94%, phlegm–turbid obstructing lung of 22.97%, phlegm–damp obstructing lung of 10.30%, *qi–yin* deficiency of 6.31%, phlegm–stasis obstructing lung of 5.72%, lung–kidney *qi*–deficiency of 4.01%, lung–spleen *qi*–deficiency of 1.15%, and phlegm obstructing orifices of 1.15%. Chinese medicine compositions were obviously divided into two large groups, which were the core Chinese medicine group and the associated Chinese medicine group. The research results illustrated that the clinical research and information–sharing system had good analysis and classification effects, which not only analyzing the medication laws of senior Chinese medicine experts and single disease treatment by a hospital, but also using in the large data analysis (data collection of single or multiple diseases on Chinese medicine clinical research nationally). It provided a good scientific research platform for Chinese medicine.

Keywords: Clinical research and information–sharing system, acute exacerbation of chronic obstructive pulmonary disease, syndrome distribution, Chinese medicine composition, complex networks

(责任编辑 :曹新伟 张志华, 责任译审 :王 晶)