

基于关联规则的心系基础证与舌脉表现相关性初探^{*}

杨 涛, 吴承玉[△]

(南京中医药大学基础医学院, 江苏南京 210046)

摘要:目的 运用 Microsoft 关联规则探讨心系基础证与舌脉表现的相关性。方法 利用 Microsoft 数据挖掘平台,建立心系基础证与舌脉表现的关联规则模型,并对 1741 份心系病案资料进行分析。结果 生成 294 条规则,其中涉及心气虚证 158 条,心血虚证 1 条,心阴虚证 8 条,心阳虚证 4 条,心火亢盛证 4 条,心血瘀阻证 113 条,痰蒙心神证 6 条;规则能够反映基础证的舌脉特点及基础证之间的相关性。结论 运用关联规则算法能够探明心系基础证与舌脉表现之间的关系,但算法设置及规则整理尚存在不足之处,在今后的研究中,要尝试多算法的联合使用,同时要设计符合中医药数据特点的新算法,从而为五脏系统数字化辨证体系服务。

关键词: 心系;基础证;舌脉表现;关联规则

中图分类号: R256.2

文献标志码: A

文章编号: 1000-2723(2013)04-0044-03

历代医家在长期的医疗实践中提出多种辨证方法,如八纲辨证、气血津液辨证、脏腑辨证、六经辨证、卫气营血、三焦辨证等,由于产生于不同的时代背景下,其理论特点和适用范围不尽相同,彼此间尚存在名实异同,甚至相互矛盾之处。鉴于此,吴承玉教授经过数十年的研究与实践,提出“以五脏为病位核心,按病性分类立证”的五脏系统辨证体系,研究由单一病位与病性构成的基础证,将其有机组合后可以执简驭繁地把握灵活多变的证^[1]。然而,四诊采集的信息具有主观性、模糊性等特点,加之病位、病性特征与基础证之间的关系错综复杂,阻碍了五脏系统辨证体系的推广及应用^[2]。

关联规则是描述数据库中数据项之间潜在关系的规则,即根据一个事务中某些项的出现,可推导出另一些项在同一事务中也出现。Microsoft 关联算法是优先关联族算法,即 FP-Growth 算法^[3]。本文利用 Microsoft 关联规则算法,建立心系基础证与舌脉表现的关联规则模型,并对 1741 份心系病案资料进行分析,用以探讨心系基础证与舌脉表现的相关性。

1 资料及方法

1.1 数据准备

项目团队在开展心系基础证^[3]的研究中,通过文献调研、流行性学调查及专家咨询等方法,确定了 8 类心系基础证。搜集了江苏省中医院病案室 2000-2010 年存档的心系病证的医案 1 394 份,古今医案及临床报道的典型医案 347 份,合计 1 741 份,其中涉及心气虚证 621 份,心血虚证 70 份,心阴虚证 184 份,心阳虚证 121 份,心血瘀阻证 548 份,心火亢盛证 53 份,痰蒙心神证 112 份,瘀阻脑络证 32 份。利用 SQL Server 2005 建立病案数据库,以基础证、舌神、舌形、舌态、舌色、苔色、苔质、脉象为字段,将病案录入数据库。

1.2 方法

利用 Microsoft 数据挖掘平台建立关联规则模型,以舌神、舌形、舌态、舌色、苔色、苔质、脉象为输入项,基础证为预测项,设定算法最小支持度 MINIMUM_SUPPORT=0.03(生成规则之前必须包含项集的事例的最小概率),最小概率 MINIMUM_PROBABILITY=0.4(指定规则为“真”的最小概率)。

2 结果

规则中的“概率”反应规则发生的可能性,例如一条规则中有 A 的条件下会有 B(A⇒B)的概率是

^{*} 基金项目:江苏省自然科学基金课题(NO:BK2009429);江苏省研究生创新计划项目(NO:CXZZ13_0607)

收稿日期:2013-07-15

作者简介:杨涛(1987~),男,江苏南京人,博士研究生,研究方向:心系病位特征及基础证的研究。

[△]通信作者:吴承玉,E-mail:chengyu720@yahoo.com.cn

指 $\text{Probability}(A \Rightarrow B) = \text{Probability}(B|A) = \text{Support}(A, B) / \text{Support}(A)$; “重要性”用于度量规则的有用性, 值越高则意味着规则越有价值, 定义如下: $\text{Importance}(A \Rightarrow B) = \log(\text{Probability}(B|A) / \text{Probability}(B|not A))$, 从定义可知, 如果该值为 0 表示 A 和 B 没有关联性, 正值表示一旦拥有 A 则再拥有 B 的概率会增长, 负值表示一旦拥有 A 则再拥有 B 的概率会降低。

2.1 心系基础证与舌脉表现的关联规则

算法共产生 294 条规则, 其中涉及心气虚证 158 条, 心血虚证 1 条, 心阴虚证 8 条, 心阳虚证 4 条, 心火亢盛证 4 条, 心血瘀阻证 113 条, 痰蒙心神证 6 条, 瘀阻脑络证 0 条。将规则按照“重要性”从高到低排列, 如表 1 所示。

表 1 心系基础证与舌脉的关联规则

编号	概率	重要性	规则
1	0.50	1.23	脉象=数、滑, 苔质=黄、腻=>心火亢盛证
2	1.00	1.07	舌态=短缩, 舌色=红=>痰蒙心神证
3	0.40	1.03	舌色=淡白, 苔质=少=>心血虚证
4	0.67	0.97	舌态=短缩, 苔质=腻=>痰蒙心神证
.....			
292	0.42	0.07	舌形=胖大=>心气虚证
293	0.40	0.07	脉象=细、沉, 弦=>心气虚证
294	0.40	0.06	苔质=少, 脉象=细=>心气虚证

2.2 同一基础证的关联规则

将与同一基础证相关的规则进行整理, 可以看出基础证与舌脉的相关性, 以痰蒙心神证为例, 从表 2 可以看出, 痰蒙心神证病人的舌脉表现: 舌体短缩, 舌质红、淡红或淡白, 舌苔厚腻, 脉弦、滑、数。

表 2 痰蒙心神证相关规则

概率	重要性	规则
1.00	1.07	舌态=短缩, 舌色=红=>痰蒙心神证
0.67	0.97	舌态=短缩, 苔质=腻=>痰蒙心神证
0.67	0.97	舌态=短缩=>痰蒙心神证
0.50	0.90	脉象=数, 滑, 苔质=腻=>痰蒙心神证
0.50	0.89	苔质=厚, 腻, 舌色=淡红=>痰蒙心神证
0.50	0.89	舌色=淡白, 脉象=滑, 弦=>痰蒙心神证

2.3 不同基础证舌脉表现之间的关系

通过模型的依赖关系网络, 可以直观地看出基础证之间的相关性, 如图 1 所示, 心血瘀阻证和心气虚证都可能表现出舌绛、苔厚, 脉沉、细、涩、弦。

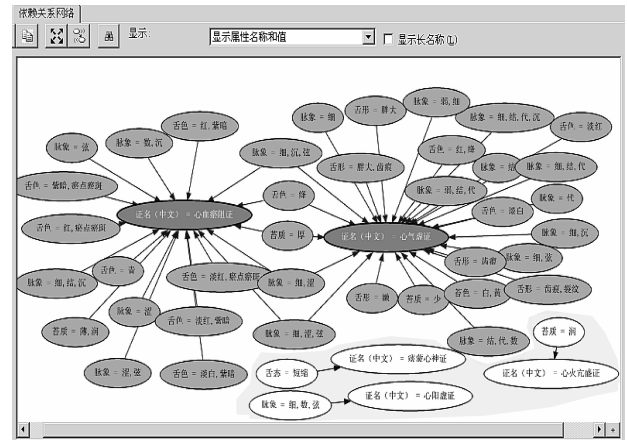


图 1 依赖关系网络图

3 分析及讨论

3.1 规则分布问题讨论

算法产生了 294 条规则, 其中心气虚证 158 条, 而瘀阻脑络证 0 条, 造成分布不规律的原因如下: 第一, 原始病案资料分布不均匀, 其中心气虚证 621 份, 而瘀阻脑络证 32 份; 第二, 算法参数 $\text{MINIMUM_SUPPORT}=0.03$, 则项集的最小数目应该为 52.23(样本数 1741×0.03), 由于瘀阻脑络证样本数小于 52.23, 故无法产生相应的规则。可以适当下调参数, 以便产生与瘀阻脑络证相关的规则, 然而, 下调参数之后, 与各个基础证相关的规则都会增加, 这会增加规则整理及分析的难度。

3.2 规则整理及分析中的问题讨论

利用关联规则算法产生了大量规则, 其中包含已有知识, 同时也蕴含着隐式知识, 需要挖掘人员根据专业知识进行筛选, 从中提取出有用的规则。本文中与心气虚证有关的规则有 158 条, 这就需要对规则进行合并、筛选, 从中提取最能反映心气虚证的舌脉表现; 与痰蒙心神证有关的规则仅为 6 条, 通过分析可以看出: 痰蒙心神证病人往往舌体短缩, 舌质红、淡红或淡白, 舌苔厚腻, 脉弦、滑、数。其结果与中医理论基本一致。

3.3 关联规则算法局限性分析

关联规则算法在中医药数据挖掘中应用较多,

然而关联规则是一种基于概率的算法,数据中某一项集出现的概率越高,其规则的支持度及置信度也越大,对于大样本随机事件而言,可以产生较为准确的关联规则,但是,对于中医病案数据而言却存在一定的问题:首先,疾病的发生具有地域性,因而样本中疾病及基础证的分布具有偏态性;其次,若某一疾病或基础证非常罕见(假如它在一万份病案中只出现 10 次),那么即使其中蕴含的信息非常有价值,也很难产生相应的关联规则;最后,假设某一症状在每个病案中出现了,如淡红舌,那么就会产生与淡红舌相关的关联规则,但这些规则的信息量却很小。

4 小结

本文利用关联规则算法探讨心系基础证与舌脉的相关性,在一定程度上揭示了心系基础证与舌脉之间的关系,但尚存在参数设置、规则整理分析等方面的不足。由于中医药数据具有其特殊性,单

一的数据挖掘方法较难提炼其中的精华,所以可以利用多种工具、结合多种方法进行分析;此外,在合理应用现有算法的基础上,根据中医药数据的特点,设计符合中医药数据挖掘的新算法,最终为构建五脏系统数字化辨证体系服务。

参考文献

- [1] 吴承玉,徐征,骆文斌,等. 五脏系统病位特征与基础证的研究[J]. 南京中医药大学学报,2011,27(3):201-203.
- [2] 杨涛,吴承玉. 应用模糊数学探讨心系病位、病性特征与基础证的相关性 [J]. 南京中医药大学学报,2013,29(3):208-210.
- [3] 陆楠,王喆,周春光. 基于 FP-tree 频集模式的 FP-Growth 算法对关联规则挖掘的影响 [J]. 吉林大学学报(理学版),2003,02:180-185.
- [4] 吴承玉,丁以艳,吴承艳,等. 心系病位特征与基础证的研究[J]. 南京中医药大学学报,2012,28(1):1-2.

(编辑:岳胜难)

(上接第 43 页)

五诊:T:37.6℃,神情清爽,咳嗽渐平,已无胸部紧闷感,仍胸痛,咯铁锈色痰,口苦口干,大便 1 次,已不甚臭秽,舌苔转淡黄腻,脉滑,至数减少。查体右肺中上部罗音减少。治以清热化痰,化瘀散痛。

处方:芦根 25g,薏苡仁 30g,冬瓜仁 15g,桃仁 10g,瓜蒌 25g,黄连 10g,法夏 15g,枳壳 15g,桔梗 10g,连翘 15g,甘草 10g,1 剂。

六诊:T:37℃,胸痛减少,仍有铁锈痰,舌上腻苔减少,脉稍滑。肺部罗音甚少。以千金苇茎汤加味连服五日。症状消失,右肺呼吸音恢复正常。再复查血常规:WBC:8.6×10⁹/L、N:76%、L:23%、EO:10%。X 线胸透提示右肺中上叶病灶基本吸收。

评析:本案缘起于劳动后受寒,外感未及时宣解。病 3 日后,头疼身楚,咳引胸痛,渴饮,是表卫不解,肺热闭郁,且已灼伤津液。以桑菊饮加石膏、枳实等效罔。壮热不已,咳反加剧,且添胸憋,口臭,溲短,大便不行,加瓜蒌、大黄,冀图宣白通腑。虽得大

便 1 次,但壮热不为所降,胸秩更觉紧闷,气粗,腹胀,苔黄厚,是表里郁闭,邪热壅盛。三诊转用麻黄、豆豉宣散,枳桔开上,联合萎贝、黄连、杏仁、硝黄是仿小陷胸汤合宣白承气汤,攻下阳明热结,解除痰热结胸,更益银花、鱼腥草清泄热毒,得大便畅下,澀然汗出,身热随之而降,胸闷得以宽释。此表里郁闭解散,气机得以流动所致。反思于二诊,即应着手宣开表郁,桑菊清轻力薄不足以宣动郁甚散之表卫,枳实乃治热扰胸膈,与此案病位相左,唯麻、豉之辛散借石膏之辛寒,相配有越婢汤意,发泄肌表热郁方胜一筹,再借通腑泄毒,合力奏效。三诊若再犹豫不决,肺之热不仅移于腑,还必逆传于心包,貽误病人。

参考文献

- [1] 朱虹江. 陈乔林治疗外感热病经验[J]. 光明中医,2000,15(91):26-27.

(编辑:徐建平)