

关联规则技术在儿童安全用药中的应用*

胡珊珊

厉玉鹏

(济宁医学院医学信息工程学院 日照 276826)

(日照市人民医院 日照 276826)

〔摘要〕 介绍数据挖掘及关联规则的基本概念,采用 Apriori 算法对儿童肺炎用药数据进行挖掘分析,总结总体用药、高频用药、用药组合和用药关联规则,为儿科临床专家在科研和疾病诊治方面提供科学依据,有助于提高儿童用药的安全性和规范性。

〔关键词〕 数据挖掘; 关联规则; 儿童安全用药; Apriori 算法; 儿童肺炎

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.10.016

Application of Association Rule Technology in the Safe Medication for Children HU Shanshan, Medical Information Engineering School, Jining Medical University, Rizhao 276826, China; LI Yupeng, People's Hospital of Rizhao, Rizhao 276826, China

〔Abstract〕 The paper introduces the basic concepts of data mining and association rule, mines and analyzes the medication data of pneumonia in children by adopting the Apriori algorithm, summarizes overall medication, high-frequency medication, medication combination, and the association rule of medication, provides pediatric clinical specialists with scientific basis in the aspects of scientific study and disease diagnosis, which is conducive to the safety and standardization of safe medication for children.

〔Keywords〕 data mining; association rule; safe medication for children; Apriori algorithm; pneumonia in children

1 引言

随着我国社会的发展以及医药卫生事业的不断进步,儿童安全用药问题受到越来越多的社会关注。儿童由于体质原因,药物代谢酶的分泌和肝肾

功能等与成人相比均存在较大差异,且儿童机体器官及系统功能发育尚不完善,对药物的敏感性和耐受性也与成人不同,因此一些成人使用的药物,儿童使用可能发生严重的不良反应^[1]。《2016 年儿童用药安全调查报告白皮书》显示中国 14 岁以下儿童药物不良反应发生率是成人的两倍,新生儿达到 4 倍^[2]。因此儿童安全用药是亟待解决的社会问题。

医院积累海量的医疗数据资源,这些医疗大数据是医学领域临床诊治和科学研究的重要资源。因此采用科学实用的数据挖掘技术,在海量医疗大数据中发现有价值的规则,揭示数据内部隐藏的知识,为专家科研和临床诊治方面提供科学依据,提高临床诊疗水平,具有重大意义和广阔发展前景^[3]。本文在研究关联规则数据挖掘技术的基础上,以儿童肺炎的用药医嘱记录为研究对象,将关联规则的 Apriori 算法用于儿童肺炎安全用药的研

〔收稿日期〕 2018-07-09

〔作者简介〕 胡珊珊,硕士,讲师,发表论文 8 篇,参编教材 3 部。

〔基金项目〕 济宁医学院校级科研计划项目“大数据时代针对儿童安全用药的数据挖掘技术研究”(项目编号:JY2015KJ020);济宁医学院教师科研扶持基金“面向基层医院的健康医疗大数据挖掘和应用研究”(项目编号:JY2017KJ 056)。

究, 汇总该病的高频用药, 探寻其中的用药规律, 为儿童肺炎用药决策提供支持, 降低儿童不合理用药以及发生药物不良反应的概率, 提高儿童用药的安全性和规范性。

2 关联规则分析

2.1 概述

关联规则是数据挖掘中一个重要的研究分支, 用于在海量数据中发现数据项之间隐藏的关联和相关联系, 帮助人们做出正确的决策^[4]。Apriori 算法是关联规则技术的核心算法, 使用 Apriori 算法进行数据挖掘主要分为两个阶段: 第 1 阶段从挖掘的数据集中找出所有的频繁项集, 第 2 阶段再从频繁项集中产生关联规则^[5-7]。

2.2 关联规则基本概念

2.2.1 频繁项集 假设 A 是由项组成的集合, 称 A 为项集, 若项集 A 包含 k 个项, 则称 A 为 k 项集。项集 A 在事务数据库 D 中出现的次数与事务数据库 D 中总事务个数的比值, 称为项集的支持度。如果项集的支持度大于预先设定的最小支持度阈值, 则称该项集为频繁项集。

2.2.2. 关联规则 是形如 $X \rightarrow Y$ 的逻辑蕴含式, 其中 X 为先决条件, Y 为关联结果。通常可以用 3 个指标——支持度、置信度、提升度来衡量关联规则的有效性。(1) 支持度 (Support)。表示项集 X 和 Y 在总项集中同时出现的概率。公式为 $\text{Support}(X \rightarrow Y) = P(X \cup Y)$ 。(2) 置信度 (Confidence)。表示在先决条件 X 发生的条件下关联结果 Y 发生的概率。公式为 $\text{Confidence}(X \rightarrow Y) = P(Y | X) = P(X \cup Y) / P(X)$ 。(3) 提升度 (Lift)。表示在发生 X 的条件下发生 Y 的概率与不考虑 X 的条件下发生 Y 的概率之比。公式为 $\text{Lift}(X \rightarrow Y) = P(Y | X) / P(Y) = P(X \cup Y) / (P(X)P(Y))$ 。

2.3 儿童肺炎用药关联规则分析

在儿童肺炎用药关联规则挖掘中支持度是指药品 A 和药品 B 同时出现的概率, 反映规则的重要性

和普遍性; 置信度是指使用药品 A 的情况下使用药品 B 的概率, 反映规则的准确性和可靠性; 提升度反映使用药品 A 对使用药品 B 的影响程度。若 $\text{Lift}(A \rightarrow B) > 1$, 说明药品 A 和药品 B 是有关联的, 若使用药品 A 则也有可能使用药品 B ; 若 $\text{Lift}(A \rightarrow B) = 1$, 说明药品 A 和药品 B 相互独立, 即是否使用药品 A 对药品 B 的使用无影响; 若 $\text{Lift}(A \rightarrow B) < 1$, 说明药品 A 和药品 B 是排斥的, 若使用药品 A 则不太可能使用药品 B 。满足最小支持度和最小置信度阈值的规则 $A \rightarrow B$, 称为强关联规则, 如果 $\text{Lift}(A \rightarrow B) > 1$, 则规则 $A \rightarrow B$ 是有效的强关联规则。本文中研究的是有效的强关联规则。

3 儿童肺炎用药关联规则挖掘

3.1 儿童肺炎概述

是儿童期尤其是婴幼儿期重要的常见病, 是我国住院儿童死亡的首要原因^[8]。儿童肺炎是由病原体感染或吸入羊水、油类或过敏反应等因素所引起的肺部炎症, 主要临床表现为发热、咳嗽、气促、呼吸困难和肺部啰音等^[9]。如果治疗不及时, 患儿可能出现气道梗阻、呼吸困难等症状, 严重影响患儿的正常生活, 甚至可能威胁患儿的生命健康^[10]。

3.2 资料来源

选取 2017 年 5 月 - 2018 年 5 月在日照市某市级医院接受治疗的儿童肺炎患者 856 例。其中男 448 例, 女 408 例, 年龄 2 月 ~ 13 岁, 临床上主要有咳嗽、咳痰、发热、气促、肺部啰音等症状。

3.3 数据预处理

数据挖掘中数据的质量是保证挖掘结果准确性至关重要的一步^[11]。因此数据预处理是数据挖掘的重要前提, 主要包括数据清理、集成、规约和转换^[12]。本文主要对儿童肺炎用药数据进行预处理, 去除注射用水、0.9% 氯化钠注射液、5% 葡萄糖注射液等无特殊治疗作用的溶媒。对原始用药数据中涉及隐私的属性进行屏蔽处理, 如患者姓名、住址、联系方式等。然后在用药数据中选择与数据挖

掘相关的有效字段，如患者的住院号、使用的药品名、用药时间及用药总量等重要字段，删除不必要的、与数据挖掘过程无关的字段值。在清理过程中重点处理不同用药时间的用药记录重复冗余问题并对缺失数据和噪声数据进行处理。参考《中国药典》对医嘱记录中的药品名称进行统一规范处理，方便数据挖掘对数据的识别，在此基础上汇总儿童肺炎医嘱记录中的相关药品。最后进行数据转换，将医嘱记录中出现的药品名从 1 开始编号，为关联规则分析提供适合的数据模型^[13]，可以在很大程度上提高数据挖掘的质量和效率，减少实际挖掘需要的时间。

3.4 关联规则挖掘过程

使用 R 语言对儿童肺炎用药数据进行挖掘分析。首先将经过预处理后的用药数据保存为 R 语言需要的 csv 格式，将转换后的数据集导入 R 中，为数据挖掘做好数据准备。在挖掘过程中利用 summary（）函数对儿童肺炎用药数据进行汇总，利用 arules 包中的 eclat（）函数和 apriori（）函数分别产生频繁项集和关联规则，其格式为^[14]：

```
eclat (data, parameter = NULL, control = NULL)
apriori (data, parameter = NULL, appearance = NULL,
control = NULL)
```

其中 data 为要挖掘的数据集；parameter 设置支持度、置信度、项集所含项数的最大值或最小值和输出结果类型等参数；appearance 可限定先决条件和关联结果中包含哪些项集；control 用来控制函数性能。在参数的设置上，通过设置不同的支持度和置信度阈值进行多次尝试，对儿童肺炎用药数据进行挖掘分析以得到有效的治疗该病的高频用药以及药品之间的关联规则。通过 inspect（）函数查看频繁项集和关联规则的详细信息。

4 结果分析

4.1 总体用药

通过 summary（）函数进行汇总，856 例肺炎患儿在住院期间使用药品（不包括溶媒）品种数最

少 1 种，最多 10 种。第 1 个 4 分位数（Q1）是 4，表明 25% 的患者使用不超过 4 种药品；中位数是 5，表明 50% 的患者使用药品不超过 5 种；第 3 个 4 分位数（Q3）是 7，表明 75% 的患者使用药品不超过 7 种；均值为 5.449，表明所有患者平均使用 5.449 种药品；众数为 4，表明肺炎患儿中使用 4 种药品的最多。通过汇总数据可以看出虽然儿童肺炎的病情比较复杂，用药品种数相对较多，但临床对于肺炎住院患儿的治疗比较谨慎，符合临床诊疗指南。

4.2 单一用药

通过 arules 包中的 eclat（）函数进行分析，设置最小支持度 15%，项数为 1，得到频繁 1 项集。使用 inspect（）函数查看，按支持度降序排序，得到前 10 位儿童肺炎常用药，见表 1。吸入用布地奈德混悬液主要有抗感染、抗炎、减轻气道炎症、平喘的作用。热毒宁注射液是一种中成药，主要有抗病毒、抗炎和退烧的作用。盐酸氨溴索葡萄糖注射液主要有止咳、祛痰的作用。硫酸特布他林雾化液主要有扩张气管、平喘的作用。可见抗感染、祛痰、抗炎、平喘是肺炎患儿的主要治疗措施。

表 1 儿童肺炎常用药

药品名称	支持度（%）
吸入用布地奈德混悬液	77.57
热毒宁注射液	69.16
盐酸氨溴索葡萄糖注射液	69.16
硫酸特布他林雾化液	57.00
维生素 C 注射液	42.99
孟鲁司特钠咀嚼片	37.38
注射用乳糖酸阿奇霉素	26.17
注射用盐酸溴己新	25.23
注射用美洛西林钠	19.63
注射用哌拉西林钠他唑巴坦	17.76

4.3 合并用药

通过 arules 包中的 eclat（）函数进行分析，设置最小支持度 25%，项数为 2，得到频繁 2 项集。使用 inspect（）函数查看，按支持度降序排序，得到前 10 位儿童肺炎常用药对，见表 2。盐酸氨溴索

葡萄糖注射液和吸入用布地奈德混悬液两种组合药物在儿童肺炎治疗过程中应用最广泛，吸入用布地奈德混悬液和硫酸特布他林雾化液次之，注射用美洛西林钠和注射用哌拉西林钠他唑巴坦在临床中应用较少，表中未显示。盐酸氨溴索葡萄糖注射液和吸入用布地奈德混悬液是止咳祛痰与抗炎平喘的治疗方案，吸入用布地奈德混悬液和硫酸特布他林雾化液是抗炎与扩张气管的治疗方案，热毒宁注射液和吸入用布地奈德混悬液是抗病毒感染与抗炎的治疗方案，热毒宁注射液和盐酸氨溴索葡萄糖是抗病毒感染与止咳祛痰的治疗方案。可见肺炎患儿的病原体主要还是病毒，治疗上无明显的抗病毒药物，主要使用中药热毒宁抗病毒，以祛痰、雾化为主。另外与单一用药相比，随着联合用药品种数的增加，相同组合用药的频次减少，表明临床对儿童肺炎的治疗原则基本相同，但具体方案会根据实际情况有所变化。

4.4 用药关联规则

通过 arules 包中的 apriori () 函数进行分析，设置最小支持度 30%，最小置信度 60%，通过 inspect () 函数查看，按提升度降序排序，得到前 10 位儿童肺炎主要药物的关联规则，见表 3。通过挖

掘出的关联规则可以看出儿童肺炎用药中药品间的典型相关联系。硫酸特布他林雾化液和吸入用布地奈德混悬液两种组合药物在儿童肺炎治疗过程中关联最广，两者联合应用可以起到抗炎和扩张气管的作用。热毒宁注射液联合吸入用布地奈德混悬液和硫酸特布他林雾化液应用广泛，3 者联合应用可以起到抗病毒感染和对气道的抗炎及扩张气管的作用。盐酸氨溴索葡萄糖注射液联合吸入用布地奈德混悬液和硫酸特布他林雾化液应用次之，说明在雾化基础上联合止咳祛痰药是儿童肺炎的主要治疗措施。

表 2 儿童肺炎常用药对

药品名称	支持度(%)
盐酸氨溴索葡萄糖注射液，吸入用布地奈德混悬液	59.81
吸入用布地奈德混悬液，硫酸特布他林雾化液	57.01
热毒宁注射液，吸入用布地奈德混悬液	54.21
热毒宁注射液，盐酸氨溴索葡萄糖	46.73
盐酸氨溴索葡萄糖注射液，硫酸特布他林雾化液	43.93
热毒宁注射液，硫酸特布他林雾化液	41.12
盐酸氨溴索葡萄糖注射液，维生素 C 注射液	34.58
维生素 C 注射液，吸入用布地奈德混悬液	31.78
孟鲁司特钠咀嚼片，吸入用布地奈德混悬液	31.78
热毒宁注射液，孟鲁司特钠咀嚼片	28.97

表 3 儿童肺炎用药关联规则

关联规则			支持度（%）	置信度（%）	提升度
热毒宁注射液吸入用布地奈德混悬液	→	硫酸特布他林雾化液	41.12	75.86	1.3307
硫酸特布他林雾化液	→	吸入用布地奈德混悬液	57.01	100	1.2892
吸入用布地奈德混悬液	→	硫酸特布他林雾化液	57.01	73.49	1.2892
盐酸氨溴索葡萄糖注射液硫酸特布他林雾化液	→	吸入用布地奈德混悬液	43.93	100	1.2892
热毒宁注射液硫酸特布他林雾化液	→	吸入用布地奈德混悬液	41.12	100	1.2892
盐酸氨溴索葡萄糖注射液吸入用布地奈德混悬液	→	硫酸特布他林雾化液	43.93	73.44	1.2882
维生素 C 注射液	→	盐酸氨溴索葡萄糖注射液	34.58	80.43	1.1630
盐酸氨溴索葡萄糖注射液	→	吸入用布地奈德混悬	59.81	86.49	1.1149
吸入用布地奈德混悬液	→	盐酸氨溴索葡萄糖注射液	59.81	77.11	1.1149
硫酸特布他林雾化液	→	盐酸氨溴索葡萄糖注射液	43.93	77.05	1.1141

5 结语

儿童由于特殊的生理特征,其用药与成人有很大不同,存在更大的安全隐患。随着医疗信息化的发展,如何利用智能的信息处理技术分析挖掘儿童用药的特点和规律,将其用于指导临床实践,提高临床疗效,减少不合理用药,成为医药工作者迫切需要解决的重要问题。本文采用关联规则技术的 Apriori 算法对儿童肺炎用药进行挖掘分析,得出治疗该病的高频用药、用药组合和用药关联规则,为儿科临床专家研究儿童肺炎的治疗提供有临床价值的参考依据,做出有益的探索。但本研究中用到的数据资料来源比较单一且医嘱案例数据偏少,还有待于进一步扩充完善。在规则评价方面,只用到支持度、置信度和提升度,下一步将研究引入其他评价指标以提高挖掘的质量和可靠性。

参考文献

- 1 罗艳萍. 儿童用药安全性和有效性影响因素分析 [J]. 临床合理用药杂志, 2016, 9 (19): 123-124.
- 2 郭时民. 儿童安全用药问题不容乐观 [J]. 江苏卫生保健, 2016, 18 (23): 50-51
- 3 曾勇. 关联规则在脑科电子病历挖掘中的应用 [J]. 医学信息学杂志, 2014, 35 (10): 55-58.
- 4 张宇. 数据挖掘中关联规则的研究与应用 [D]. 长春:

长春工业大学, 2015.

- 5 翟运开, 武戈. 基于电子病历信息大数据挖掘的患者就医行为分析 [J]. 医学信息学杂志, 2017, 38 (7): 12-17.
- 6 朱习军, 陈亚楠, 董国华. Apriori 改进算法在哮喘病案数据挖掘中的应用 [J]. 徐州工程学院学报 (自然科学版), 2015, 30 (3): 8-14.
- 7 吴锐东, 张卯红, 董静, 等. 基于 Apriori 算法的病案首页数据分析 [J]. 中国数字医学, 2018, 13 (1): 79-81.
- 8 陈志敏, 张园园. 儿童肺炎病原体的实验室检查方法 [J]. 中华妇幼临床医学杂志 (电子版), 2014, 10 (4): 448-450.
- 9 沈茹, 李艳君, 林丽佳. Th17 细胞在儿童肺炎中的研究进展 [J]. 分子诊断与治疗杂志, 2017, 9 (3): 211-214.
- 10 王喜鹊, 杨晓庆. 中西医联合治疗小儿支原体肺炎的临床疗效探究 [J]. 临床医学研究与实践, 2016, 1 (22): 53-54.
- 11 黄文金, 姚明龙, 叶云金, 等. 关联规则在中医儿科学证素诊断中的应用 [J]. 中医药临床杂志, 2015, 27 (3): 329-331.
- 12 李准, 严少彪, 余颖, 等. 关联规则在儿童门诊信息分析中的应用 [J]. 医学信息学杂志, 2016, 37 (8): 76-78.
- 13 李准, 冯思佳, 杨美洁, 等. 关联规则技术在冠心病电子病历中的应用 [J]. 医学信息学杂志, 2015, 36 (1): 58-62.
- 14 黄文, 于正林. 数据挖掘: R 语言实战 [M]. 北京: 电子工业出版社, 2014.

《医学信息学杂志》版权声明

(1) 作者所投稿件无“抄袭”、“剽窃”、“一稿两投或多投”等学术不端行为,对于署名无异议,不涉及保密与知识产权的侵权等问题,文责自负。对于因上述问题引起的一切法律纠纷,完全由全体署名作者负责,无需编辑部承担连带责任。(2) 来稿刊用后,该稿包括印刷出版和电子出版在内的出版权、复制权、发行权、汇编权、翻译权及信息网络传播权已经转让给《医学信息学杂志》编辑部。除以纸载体形式出版外,本刊有权以光盘、网络期刊等其他方式刊登文稿,本刊已加入万方数据“数字化期刊群”、重庆维普“中文科技期刊数据库”、清华同方“中国期刊全文数据库”、中邮阅读网。(3) 作者著作权使用费与本刊稿酬一次性给付,不再另行发放。作者如不同意文章入编,投稿时敬请说明。

《医学信息学杂志》编辑部