



中药注射剂不良反应信号自动预警算法的研究与实现

庄严¹, 谢邦铁¹, 谢雁鸣^{2*}, 翁盛鑫¹

(1. 解放军海军总医院, 北京 100048;

2. 中国中医科学院 中医临床基础医学研究所, 北京 100700)

[摘要] 目的:为更加快速准确的从大量的不良反应数据中发现中药注射剂不良反应的预警信号提供分析依据和自动化方法。方法:构建中药注射剂不良反应监测系统,集成多种不良反应信号挖掘算法并自动产生预警分析报告,同时对算法进行了多方面的扩展以达到更好的预警分析效果。结果:经数据验证,系统可以提供与实际情况相符合的预警结果。结论:建立的系统具有良好的可扩展性,可以为多种来源的不良反应数据提供预警分析报告,为领域专家提供决策依据。

[关键词] 药品不良反应;药品不良反应检测;数据挖掘算法;报告率失衡测量法

中药制剂在我国应用广泛。安全、合理地使用中药,尤其是上市中药注射剂的安全有效应用是事关国计民生的大事。如何做好中成药上市后不良反应监测工作既是国家药品监督管理工作的重要组成部分,也是中药现代化的重要内容。目前世界上用于发现 ADR 信号、进行 ADR 监测最基本的方法是自发呈报系统(spontaneous reporting system, SRS)。将国际先进的不良反应信号检测算法应用于中药注射剂的监测中,对采集的不良反应数据进行自动分析和挖掘,及时有效的发现药品不良反应并主动预警,可以为政府药品监督部门制定应对措施提供更多的时间缓冲,对有效预防和避免药害事件具有重要的意义。

1 自动预警算法的选择

报告率失衡测量法(measures of disproportionality),是目前药物不良事件信号检测中研究最多,应用最广泛的方法。如果某种药物确实会导致某种不良事件,则该不良事件相关报告在目标药物的所有不良事件报告中所占的比例会明显高于在其他所有药物不良事件报告中所占的比例,即所谓“失衡”。具体信号检测算法有多种,大体可分为3类:①传统频数法,如报告比值比法(reporting odds ratio, ROR)、比例报告比法(proportional reporting ratio, PRR),及卡方检验法(Chi-square test)。②现代贝叶斯法,包括贝叶斯置信传播神

经网络(Bayesian confidence propagation neural network, BCPNN)、经验贝叶斯伽玛泊松缩减(empirical Bayes gamma poisson shrinker, GPS)及多项伽玛泊松缩减法(multi-item gamma poisson shrinker, MGPS)。③综合法,整合多种方法检测结果,综合判断是否为信号^[1]。

报告率失衡测量法是基于四格表的监测方法,具体算法内容可参考相关文献。其中常用监测算法各具优缺点:ROR由荷兰药物警戒中心首先提出,其计算方法简单易用,但需要积累一定量的数据,且四格表中分母为零则ROR值无法计算;PRR由Evans等首先用于英国黄卡系统不良事件信号检测的尝试,其计算方法简单直观,也不用考虑分母问题,但其需要完善的地方更多,比如一种药物某个强烈信号可能会减小该药其他信号的重要性等;BCPNN是用于描述变量之间相互依赖性关联的概率分布,该网络能够描述各属性子集之间有条件的相互独立,主要基于信息成分(information component, IC)及其可信区间来评价药物与不良事件之间的联系强度,具有对样本先验信息利用较为充分、灵敏度高的特点,但其实现较为复杂。理论和实践证明没有一种方法可以完全被另一种方法所替代,因此,综合采用多种方法进行中药注射剂不良反应的监测是最为行之有效的方法。

2 自动预警算法的实现

2.1 实现步骤 本研究构建的中药注射剂不良反应自动预警系统实现来自SRS数据的自动预警和来自临床信息系统采集形成的不良反应数据仓库的自动预警,这2种方式对应的监测算法实现没有区别,但在系统中的处理步骤略有不同,具体内容如下。

步骤一:数据拆分。由于自发报告中的数据大部分来自医护人员手工填写,因此实现SRS中ADR信号检测与自动预警,首先需提高ADR信息的利用率,要对怀疑药品对应的

[稿件编号] 20120715029

[基金项目] 国家“重大新药创制”科技重大专项(2009ZX09502-030)“中药上市后评价关键技术研究”

[通信作者] *谢雁鸣,研究员,博士生导师,研究方向为中医临床评价方法研究, Tel: (010) 64014411-3302, E-mail: zhinanb2012@yahoo.com.cn

[作者简介] 庄严,工程师,硕士,研究方向为数据仓库与系统开发, Tel:15601374620, E-mail:joyear2008@yahoo.com.cn



不良反应事件进行数据拆分,生成对应的药物-不良反应事件组合^[2]。假定 1 份报告中有 2 个怀疑药物,有 2 种不良事件,则拆分后会生成 4 条记录。为提高自动化程度,减少人为误差,系统会根据分隔符自动对怀疑药品和怀疑药品字段进行拆分。对于不良反应数据仓库的自动预警,在临床信息系统不良反应数据采集系统的设计中充分考虑了数据拆分问题,已经按照需要自动生成了药物-不良反应事件组合。

步骤二:数据清理与提取。信号的稳定性,很大程度上决定于数据库中报告的数量和质量。提高报告的质量,需要对拆分后的数据进行清洗和标准化,首先要删除重复、缺失信息过多或有明显逻辑错误的数据条目,对数据进行充分清理。其次对数据中出现的药物名称和不良反应事件名称,要严格按照药物通用名称字典和药物不良反应事件术语集进行清理和编码。系统建立了药品及不良反应术语的标准化对照表,只要将拆分后的数据与系统中标准的术语系统进行对照并保存入库,即可自动完成标准化转换,为数据处理做好准备。

步骤三:数据处理。数据清理与标准化之后首先从已拆分数据表中读取在某一时段发生的某目标药物与某目标不良事件,运用比值失衡测量法计算四格表中 A, B, D 的值并代入预警算法中计算出相应变量的值并存储,最后利用这些数据进行分析报表和预警。系统的核心逻辑利用数据库存储过程和视图来实现,以保证良好的扩展性。系统首先构建一个主存储过程进行参数的接收和不同算法的调用。不同算法由不同存储过程实现,新增算法只要编写相应算法存储过程并在主存储过程中安排调用即可。算法存储过程主要是利用传入参数和不良反应数据填写统计临时表并利用算法公式进行计算并展示和预警。

2.2 结果展示 系统实现界面见图 1。系统提供了多项参数配置,其中药品名称和不良反应事件可以多选,分隔符用于将报告中并列的多项名称自动分割,警戒值可以用于调整自动预警的灵敏度,分组和排列用于监测算法的扩展,系统可以接收来自 SRS 的不良反应数据和设计好的临床不良反应电子数据库中的数据。

某药品 2005 年至 2009 年不良反应报告预警分析结果见图 2~4。数据来自国家药监局不良反应中心,药品名称使用数字 0,1,2... 代替。2 号药品至过敏样反应,3 种不良反应监测算法基本趋势完全一致。从趋势图中可见,其可信区间范围较窄,说明估计精度较高,且 PRR 和 ROR 2 种算法从 2007 年第 2 季度开始出现预警信号,2007 年第 4 季度后连续出现预警信号,经查 2008 年,2009 年增加多例不良反应报告,与实际相符,这段时间内 PRR 和 ROR 值介于 1 和 2 之间,说明该药品相对于其他药品发生该不良反应的风险差别并不明显。BCPNN 算法 2007 年以后的数据虽然没有完全超过预警线但大部分也位于预警临界点,也属于待重点分析的范围。如果在 2007 年第 4 季度不良反应预警信号发生时

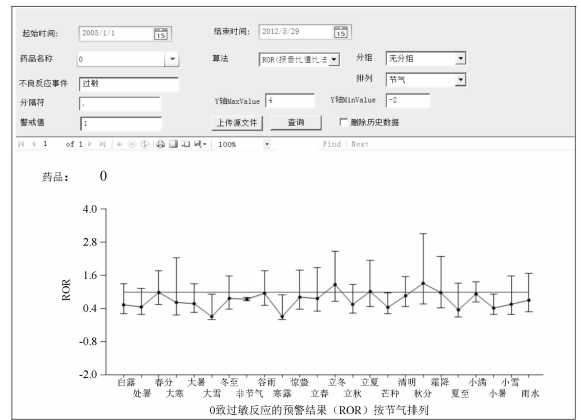


图 1 程序界面
Fig. 1 Program interface

进行及时的关注、监控与预防则能很大程度上减少 2008 年以来不良反应事件的发生。

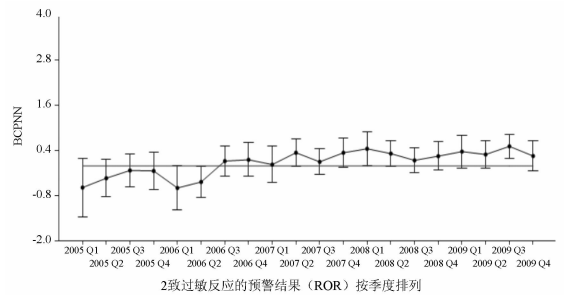


图 2 2 号药品过敏样反应 BCPNN 算法预警结果
Fig. 2 BCPNN algorithm warning results of allergy-like reaction on the 2nd drug

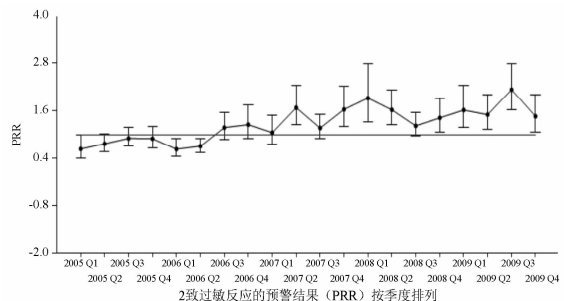


图 3 2 号药品过敏样反应 PRR 算法预警结果
Fig. 3 PRR algorithm warning results of allergy-like reaction on the 2nd drug

3 自动预警算法的扩展

目前药品不良反应的自动预警算法相对较少且检验效能不高,需要不断的扩展和完善,算法扩展主要分为 2 个方面:一是新的更高灵敏度和特异度的数据挖掘算法的发现和

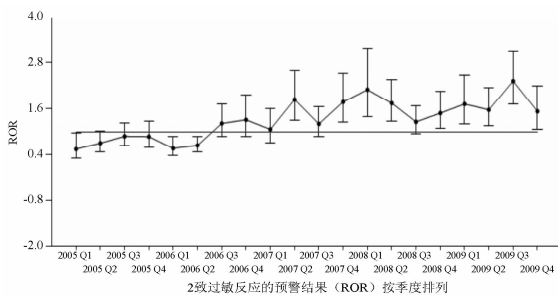


图4 2号药品过敏样反应 ROR 算法预警结果
Fig. 4 ROR algorithm warning results of allergy-like reaction on the 2nd drug

引入,比如随机森林算法和贝叶斯 Logistic 回归算法,根据其算法的特点和在相关领域应用的效果可能会成为不良反应监测预警的更有效的方法^[3]。二是在算法基础上对药品不良反应发生的相关影响因素的综合筛选和分析,因为除药物之间会产生交互作用外,药物与性别、体重、地域、节气以及机体生理生化指标等因素间亦会产生交互作用,因此数据挖掘算法还可以这些相关因素进行对比分析和报告,从而为预警分析报告提供更多的依据和内容。

系统对算法的扩展利用存储过程扩展来实现。新的算法只要编写相应的存储过程并在主存储过程中注册和传递相关参数即可。对于影响因素的扩展利用传入参数进实现,地域、性别、节气等因素的扩展只要定义相应参数并利用参数从数据库中取得相关数据代入分组和排列的计算中即可。系统中的分组指的是可以将不良反应数据按预设的参数进行分报表展示,排列指的是报表横坐标的内容。

4 问题与展望

本研究构建的中药注射剂不良反应检测系统无论是应用于来自于 SRS 中的以文件形式存储的不良反应数据,还是应用于本课题建立的基于临床信息采集的不良反应数据仓库,都取得了较好的效果,特别是在大量不良反应数据的预警分析上可以给研究人员带来更好的辅助效果。然而,应该看到不良反应信号挖掘尚不完善,虽然已有多种方法被尝试引入到该领域且不同方法间可以相互验证比较,但缺乏金标准,尤其是随着药物及其相互作用所产生 ADR 的发生率较高以及人们对用药安全问题的不断重视,更加需要探索更高灵敏度及特异度的数据挖掘算法及其对相关影响因素进行研究^[4]。因此,一方面要继续扩展系统的功能,不断纳入新算法、新方法,多角度综合分析各种方法产生的预警数据,提高系统自动化程度;另一方面也不能忽视人的作用,所有的预警数据和分析报表最终还需要专家的参与和专业的判断才能发挥更大的作用。

[参考文献]

[1] 王永炎,吕爱平,谢雁鸣. 中药上市后临床评价关键技术[M]. 北京:人民卫生出版社,2010;281.
[2] 陈文戈,李婵娟,江静,等. 基于 BCPNN 法的药品不良反应信号检测与自动预警技术研究[J]. 计算机应用研究,2009(4):1394.
[3] 罗宝章,钱轶峰,叶小飞,等. 药物不良反应信号检测方法的现状与展望[J]. 药学服务与研究,2009,9(4):255.
[4] 钱轶峰,罗宝章,叶小飞,等. 检测联合用药不良反应信号的数据挖掘方法[J]. 中国卫生统计,2010,27(1):31.

Research and implementation of adverse reaction automatic warning algorithm on Chinese medicine injection

ZHUANG Yan¹, XIE Bang-tie¹, XIE Yan-ming^{2*}, WENG Sheng-xin¹

(1. The PLA Navy General Hospital, Beijing 100048, China;

2. Institute of Basic Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China)

[Abstract] **Objective:** To provide analytical data and automatic method for more quickly and accurately discovering warning signals of adverse reactions of traditional Chinese medicine(TCM) injections from a large number of adverse reaction data. **Method:** Constructed the TCM injection adverse reaction monitoring system, integrated of multiple mining algorithms of warning signals of adverse reactions and automatically generate early warning analysis, and extended the algorithm to achieve better effect of warning and analysis. **Result:** By data validation, the system can provide consistent warning results with the actual situation. **Conclusion:** The established system has good scalability, it can provide warning analysis of adverse reaction data from multiple sources, and provide a basis for decision making to experts in the field.

[Key words] adverse drug reactions; adverse drug reactions monitor; data mining algorithm; measures of disproportionality

doi:10.4268/cjcm20121805