

Arrowsmith 在预测中药副作用研究中的思路探讨

高宏杰 张华敏 李敬华

(中国中医科学院中医药信息研究所,北京,100700)

摘要 美国芝加哥大学情报学教授 Don. R. Swanson 创立的情报学知识发现方法是揭示非相关文献及学科间隐秘联系的强有力的工具。本文简要介绍了闭合式的知识发现方法及其实现工具 Arrowsmith 软件系统,并对其在预测中药不良反应研究中的思路进行探讨。

关键词 Don R. Swanson; Arrowsmith; 中药; 不良反应

Thoughts on Using Arrowsmith to Predict Adverse Effects of Chinese Herb

Gao Hongjie, Zhang Huamin, Li Jinghua

(Institute of Medical Informatics, China Academy of Traditional Chinese Medical Sciences, Beijing 100700, China)

Abstract The methodology on discovering intelligence information founded by professor Don. R. Swanson in Chicago University is a powerful tool to reveal the non-relevant documents and hidden links between disciplines. This article briefly introduces the method of closed types of knowledge discovery process as well as the Arrowsmith software, and also explores its application in the prediction of adverse effects of Chinese herb.

Key Words Don R. Swanson; Arrowsmith; Chinese herb; Adverse effects

doi:10.3969/j.issn.1673-7202.2013.05.032

随着科学技术的飞速发展以及计算机网络技术的出现,各行业各领域的信息量都在呈指数增长,不同领域不同学科之间的交流变得越来越困难,出现了“知识分裂”现象。一个专业领域的信息对另一个专业领域是有价值的,它们之间一定存在着隐含的有价值的关联,目前国内外一些情报人员对揭示大量科学文献间的隐含关联进行了深入的研究,其中具有开拓性和代表性的是美国芝加哥大学的情报学教授 Swanson DR,他创立了基于非相关文献的知识发现方法。本文将主要介绍闭合式非相关文献知识发现方法及其研究设计出的人机交互系统——Arrowsmith 软件,并探讨其在药物不良反应预测中的研究思路。

1 闭合式非相关文献知识发现方法

1986 年,Don. R. Swanson 提出的基于非相关文献的知识发现方法,揭示了蕴含于公开发表的文献中尚未被人们认识或发觉的知识片段间的逻辑联系,提出知识假设,以便科学研究人员进一步加以证实^[1]。Swanson 的知识发现过程分为两个步骤:形成假设的过程和检验假设的过程。我们将形成假设的过程称为开放式的非相关文献知识发现方法,而检验假设的过程称为闭合式的非相关文献知识发现方法^[2]。闭合式的知识发现方法以 A 和 C 为出发点,研究人员需要努力寻找共同的中介词 B,如图 1 所示。A 和 C 的联系越

多,所做的假设越有价值。Swanson 的网络版知识发现工具 Arrowsmith 使用的是闭合式方法,Arrowsmith 系统可以免费使用,可以登录 <http://kiwi.uchicago.edu> 或 <http://Arrowsmith.psych.uic.edu>,后面的网站装载的 Arrowsmith 是最新版本,不仅可以应用于标题词,还可以应用于文摘和主题词,并且 Arrowsmith 与 PubMed 实现了紧密结合,使 Arrowsmith 系统的应用变得更加容易、快捷和灵活^[3]。Arrowsmith 可以应用的领域包括:预测药物的不良反应;确定细胞生物反应的调节代谢机制;建议新的治疗方案;确定导致疾病的可能危险因素;确定潜在的可代替人体实验的动物模型等^[4]。

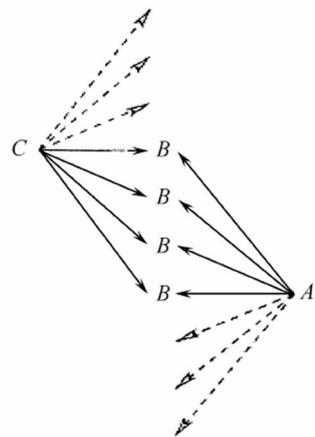


图 1 闭合式知识发现过程

基金项目:中国中医科学院第四批自主选题“基于中医药英文主题词在 Arrowsmith 系统的知识发现研究”(编号:ZZ040304)

通信作者:张华敏,女,硕士,中国中医科学院中医药信息研究所副研究员

2 Arrowsmith 系统在中药不良反应预测研究中的应用

以某中药名称的相关文献为 A, 以某种疾病的相关文献 C, 运用 Arrowsmith 的功能找出在 A、C 文献集中共同出现的词语列表 B, 通过禁用词表、语义、频次等过滤功能确定词语列表 B, 然后运用相关专业知对 AB、BC 的文献内容进行分析, 直至找出新的、有价值的信息, 具体思路如下:

2.1 某中药名称相关文献 A 的形成 在中国中医药期刊文献数据库和中国生物医学文献数据库中检索中药药理文献, 提取中药药理文献的主题词并按照降序排列, 由于在中药药理文献中出现的中药名称主题词频次越高, 说明此类中药研究的较多, 则运用 Arrowsmith 软件系统发现非相关文献间的隐秘联系可能性越大, 所以在主题词词频统计数据中截取位于前 10% 的中药名称, 将中药名称翻译成英文, 依次进行研究。

2.2 以某种疾病的相关文献 C 的形成 将中国中医药期刊文献数据库和中国生物医学文献数据库中中药药理文献中出现的所有主题词去除中药名称集合 A 后, 提取集合 C 中与中药不良反应有关的主题词, 翻译成英文主题词。

3 Arrowsmith 软件系统的运用

3.1 3 个数据库中未共现于任意一篇文章的主题词对的产生 将中药名称集合 A 中主题词分别与疾病名称集合 C 中的主题词用 and 连接, 在中国中医药期刊文献数据库、中国生物医学文献数据库以及 Pubmed 数据库中查找是否有两类主题词不产生共现的文章, 即该文章中两类主题词并没有同时出现在主题词标签部分, 表明未有对此两类主题词进行研究的文章, 也就是将要运用 Arrowsmith 软件系统进行闭合式知识发现的关键问题, 此两类主题词为将运用 Arrowsmith 系统进行知识发现的主题词词对。随机抽取几个主题词词对运用 Arrowsmith 软件系统进行闭合式知识发现研究, 寻找 A→B←C 的关系, 并结合专业知识对词对进行筛选。若未发现其隐形联系, 则继续随机抽取。

3.2 对中药不良反应进行预测分析 寻找到 A、B、C 的关系后, 在文献数据库中检索并阅读文献集合 A、B、C 的文献, 通过对文献的内容分析可对中药不良反应进行预测, 并初步探讨其不良反应的作用机制。其研究结果尚有待于中药试验加以证明。

4 Arrowsmith 系统在中医药领域的应用现状

在中医药领域, 由于中医药文言文和中医药术语语言本身的独特特点, 无论是对于开放式非相关文献知识发现方法的应用还是对于 Arrowsmith 系统的应用都较少, 且大多停留在应用可行性的验证上及对如何

研制适合于中医药领域特点的 Arrowsmith 系统的思考。邵运峰等^[5]主要介绍了 Arrowsmith 系统, 并对在中医药领域研制非相关文献知识发现软件系统提出了 3 点关键性的技术。李文林等^[6]利用 Arrowsmith 系统探讨当归与痛经的相关性, 发现当归部分有效成分对痛经有潜在的治疗作用, 另一部分则不利于痛经的治疗。李文林等^[7]运用 Arrowsmith 系统探讨人工培养的蛹虫草中的虫草素对肠黏膜免疫屏障功能损伤保护作用的可能机制^[12]。其基于文献的知识发现形成的知识关联结果不仅能为实验室试验或临床观察提供可证实或否定的数据, 也能为研究者提供新的研究思路。

5 小结

Swanson 教授的闭合式知识发现方法的作用并不是为科研人员直接产生一个新的假设, 而是为科学研究提供新的线索, 从文献中验证此假设的可行性, 从而为科技创新研究提供一个合理的研究思路, 并且该科学假设还有赖于科研实践的最终验证。该方法不仅使科学研究少走弯路, 而且可大大减少医学研究的投资, 对医学实践有很好的指导作用。但是, 该方法仍然存在一些不足, 而且该不足在中医药领域表现得尤为突出。中医药文献中包含其特有的中医学学术语, 它既不同于西方医学的术语, 又不同于英文文献的表述特点, 且尚没有类似 Arrowsmith 系统这样的适用于中医药文献特点的中文软件来指导科研人员的研究发现, 如果能开发出这样的软件系统将是非常有意义的事情, 也将为中医药情报研究领域的发展翻开新的一页。

参考文献

- [1] HuangWei, NakamoriY, Wang Shouyang, et al. Mining scientific literature to predict new relationships [J]. Intelligen DataAnalysis, 2005 (9): 219-23.
- [2] Weeber, M., Klein, H. and de Jong - van den Berg, LT. Using Concepts in Literature - Based Discovery: Simulating Swanson's Raynaud - Fish Oil and Migraine - Magnesium Discoveries [J]. J Am Soc Inf Sci Technol, 2001, 52 (7): 548 - 557.
- [3] 马明, 武夷山. Don R. Swanson 的情报学学术成就的方法论意义与启示 [J]. 情报学报, 2003, 22 (3): 259 - 266.
- [4] 许建阳, 马明, 王发强. Swanson 的非相关文献知识发现法对医学发展的思考 [J]. 医学与哲学, 2003, 24 (8): 21 - 23.
- [5] 邵运峰, 翁捷. 非相关文献知识发现法在中医研究中的应用 [J]. 江西中医学院学报, 2005, 17 (3): 8 - 9.
- [6] 李文林, 等. 利用知识发现工具 Arrowsmith 探讨当归与痛经的相关性 [J]. 中华医学图书情报杂志, 2008, 17 (4): 7 - 11.
- [7] 李文林, 潘苏华, 关洪月. 基于 Arrowsmith 探讨虫草素与肠黏膜屏障功能稳定的潜在相关性 [J]. 医学信息学, 2009, 22 (12): 2684 - 2687.