

# 关键词网络在药物不良反应领域资源聚合与知识发现中的应用研究\*

曹玉莹 王 伟

毕 强

李 梅

王丽伟

(吉林大学公共卫生学院  
长春 130021)(吉林大学管理学院  
长春 130021)(河北医科大学图书馆  
石家庄 050017)(吉林大学公共卫生学院  
长春 130021)

**[摘要]** 基于万方医学网和维普数据库的文献数据, 获取文献关键词、构建关键词网络, 将心血管药物不良反应本体中的关系嵌入到关键词网络中, 对网络进行可视化显示, 列举案例验证基于关键词网络的领域资源聚合与知识发现方法在药物不良反应领域的可行性。

**[关键词]** 关键词网络; 资源聚合; 知识发现; 药物不良反应; 本体

**[中图分类号]** R-056      **[文献标识码]** A      **[DOI]** 10.3969/j.issn.1673-6036.2018.02.016

**Application Study on Keywords Network in Resources Aggregation and Knowledge Discovery in Adverse Drugs Reaction Do-**

**main** CAO Yu-ying, WANG Wei, School of Public Health of Jilin University, Changchun 130021, China; BI Qiang, School of Management of Jilin University, Changchun 130021, China; LI Mei, Hebei Medical University Library, Shijiazhuang 050017, China; WANG Li-wei, School of Public Health of Jilin University, Changchun 130021, China

**[Abstract]** Based on literature data from Wanfang Med Online and VIP database, the paper grabs keywords of literature to establish the keywords network, in which the relation in cardiovascular adverse drugs reaction ontology is imbedded, visualizes the network, and lists cases to verify the feasibility of keywords network-based domain resources aggregation and knowledge discovery methods in the adverse drugs reaction domain.

**[Keywords]** Keywords network; Resources aggregation; Knowledge discovery; Adverse drugs reaction; Ontology

## 1 引言

近年来药物的使用量不断增加, 由此带来的不良反应也逐渐增多, 给人们带来巨大的健康代价和

经济损失<sup>[1]</sup>, 同时由药物引起的不良反应越来越受到人们的重视, 如何有效地发现未知的不良反应成为一个亟待解决的问题。关键词网络是基于关键词的共现而构建, 即将同一篇文献中的关键词两两连接, 不同的论文间通过相同的关键词构成连接, 如此构成关键词共现网络<sup>[2]</sup>。关键词网络方法是网络分析法的一种, 主要用于发现意料之外的或隐藏的生物医学关联, 在生物医学领域知识发现中发挥着重要作用<sup>[3-6]</sup>。本研究以心血管不良反应领域的知识发现研究为例, 验证基于关键词网络的知识聚合和知识发现方法在药物领域应用的可行性。

**[修回日期]** 2017-10-31

**[作者简介]** 曹玉莹, 硕士研究生, 发表论文 4 篇; 通讯作者: 王丽伟, 副教授。

**[基金项目]** 国家自然科学基金项目 (项目编号: 71273111)。

## 2 数据来源、方法及实现步骤

### 2.1 数据来源

2.1.1 文献来源 本文主要对关键词网络方法应用进行探讨, 仅选取 2012 年数据做一例证。编写 JAVA 程序, 提取万方医学网中所有文献的关键词, 从维普数据库中筛选出万方医学网未收录的医药卫生类期刊, 利用程序将其关键词提取出来, 作为补充。从万方医学网中共抽取 689 种期刊中 645 776 篇文章的 334 246 个关键词; 从维普数据库中补充了 127 种期刊中 66 196 篇文章的 46 766 个关键词。

2.1.2 我国心血管类药物不良反应本体 在另一项研究中构建了我国心血管类药物不良反应本体<sup>[7]</sup>, 该本体主要是基于我国当前使用的心血管类药物说明书中不良反应术语, 基于美国不良反应本体 (the Ontology of Adverse Events)<sup>[8]</sup>中现有的关系及概念进行构建, 可以在当前世界上最综合的生物学医学本体知识库中免费获取。我国心血管类药物不良反应本体包括 282 种药物, 按治疗用途可以分为 9 类: 钙通道阻滞药、降血压药、抗动脉粥样硬化药、周围血管扩张药、抗心绞痛药、抗心律失常药、抗血小板活性药、强心药及其他心血管系统药, 按化学成分可以分成 11 个大类, 共有 1 529 个中文不良反应术语。

### 2.2 方法

本研究利用关键词网络构建的方法, 将不同文献关键词构成一个庞大的网络, 将本体知识聚合到网络中, 对心血管药物不良反应进行知识发现, 进而提供知识服务。基于关键词网络的药物不良反应知识聚合与发现方法流程, 见图 1。图中文本指的是来自资源数据库收录期刊中的每篇文章。关键词是指各期刊对于文献所要求的关键词字段中的每个关键词。为实现跨库信息整合, 领域本体中的关系嵌入到网络中, 本研究应用的是药物不良反应本体集。为验证本研究结果的准确性, “已有文献”被引入进来, 即已在国内外期刊中公开发表的文献。

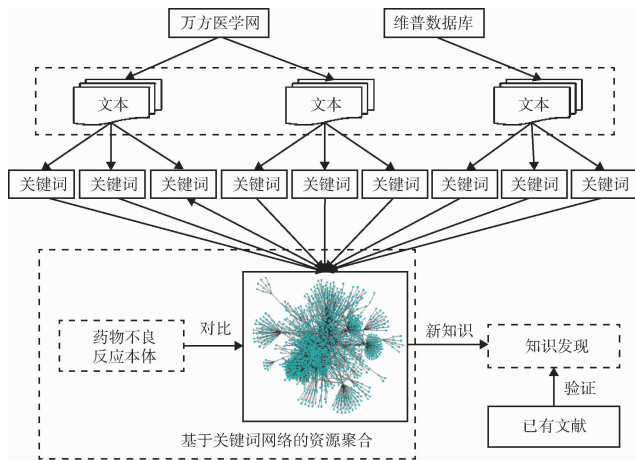


图 1 基于关键词网络的领域知识聚合与知识发现方法流程

### 2.3 实现步骤

2.3.1 关键词网络构建 以万方医学网与维普数据库中的文献作为数据来源, 将文献提取形成文本格式。提取文献的关键词字段, 将来自同一篇文章的关键词两两成对, 构建关键词对, 最后利用所有关键词对构建关键词网络。在关键词网络中, 每个关键词作为网络中一个结点, 关键词之间的联系为边, 每对关键词的共现次数作为边的权重, 共现次数越多, 边的权重越大。

2.3.2 基于本体构建领域关键词子网络 为避免遗漏相关结果, 首先将我国心血管类药物不良反应本体收录的药物名称中各种剂型去掉, 然后在关键词网络中检索带有这些药物名称的子网络, 即心血管类药物子网络。为提高知识发现效果, 再将本体集中的药物术语作为第 1 级网络结点, 与之直接相连的结点作为第 2 级网络结点, 与第 2 级结点直接相连的作为第 3 级网络结点, 构建 3 级网络, 同时将这些药物的不良反应相关知识聚合到关键词网络中, 如不良反应的概念、药物与不良反应的关系、各种不良反应之间的关系等。最后利用 Cytoscape 3.2.1 版本对网络进行可视化展示。

2.3.3 领域知识发现 在心血管类药物的子网络中, 去掉第 2 级与第 3 级结点中的不良反应术语, 筛选剩余关键词中的疾病术语, 可以得到以下假设: 与药物直接相关的第 2 级结点中的疾病术语, 以及

与药物间接相关的第 3 级结点中的疾病术语,可能是药物新的适应症,或者是说明书中未记载的而实际用药中发现的不良反应;第 2 级结点中得到的假设权重高于第 3 级结点。本研究中以替米沙坦、苯磺酸左旋氨氯地平和多巴胺为例进行分析。以上数据处理过程除特殊说明外均通过 Perl 编程实现。

## 3 结果

### 3.1 关键词网络构建

基于万方医学网与维普数据库中医学文献的关键词所构建的关键词网络含有 380 684 个结点,共 3 446 697 对关键词,即 3 446 697 个边。心血管类药物子网络共包括 111 种心血管类药物,7 223 个结点,13 148 个关键词对(边),关键词对共现次数最多为 81 次。将心血管类药物不良反应本体知识聚合在该子网络中后,82 种心血管类药物共聚合 1 163 种不良反应术语及概念。为清晰地显示心血管类药物子网络,截取共现次数  $\geq 3$  的关键词对,构成高共现关键词子网络,该网络包括 327 个结点,1 398 个关键词对(边)。完整的心血管类药物高共现关键词子网络,见图 2。根据边的权重进行可视化的一部分心血管类药物的高共现关键词子网络,可以从整体上对网络进行观察。

### 3.2 心血管药物不良反应知识发现

首先对心血管药物高共现关键词子网络中的第 2 级结点进行文献验证,将心血管药物高共现关键词子网络中 1、2 级结点间的联系(1 398 个关键词对)分别作为主题字段,在中国知网中进行检索,其中有 574 个关键词对能够查到文献依据,证实第 2 级结点均为心血管药物(第 1 级结点)的适应症或说明书中未记载的不良反应,即在第 2 级结点中有 41.06% 的结点可以得到验证。第 1 个案例分析中,以心血管类药物子网络为例,在去除心血管类药物子网络中第 2 级和第 3 级结点中已知的不良反应术语后,第 3 级还剩下 1 437 个结点,第 3 级剩

下 48 405 个结点。以第 1 级结点中的替米沙坦为例,检索与其直接相关的第 2 级结点有 54 个,其中包括“胰岛素抵抗”、“替米沙坦”与“胰岛素抵抗”共现次数为 9,见图 3。在去除不良反应后,药物与疾病还可能为适应症的关系,因此假设替米沙坦对于改善胰岛素抵抗具有一定作用,这一假设在文献中得到证实,蒋秋明等人通过临床试验证实替米沙坦发挥较好的降压作用的同时,还可改善糖尿病合并高血压患者的胰岛素抵抗情况<sup>[9]</sup>,周建敏等人通过实验观察替米沙坦对高血压合并代谢综合征患者的胰岛素抵抗情况的影响,且替米沙坦治疗组治疗前后结果具有统计学意义,证实替米沙坦在治疗高血压的同时,能够改善患者胰岛素抵抗情况<sup>[10]</sup>。

第 2 个案例分析中,使用心血管类药物的高共现关键词子网络,以苯磺酸左旋氨氯地平为例,见图 4。在第 1 级结点中检索苯磺酸左旋氨氯地平,得到与其直接相关的第 2 级结点 1 个,即左心室肥大,苯磺酸左旋氨氯地平与左心室肥大这对关键词共现的次数为 3,即边的权重为 3;以左心室肥大检索第 3 级结点,得到 1 个结点,即原发性高血压,左心室肥大与原发性高血压这对关键词共现的次数为 6,即边的权重为 6。通过网络分析,得到苯磺酸左旋氨氯地平—左心室肥大—原发性高血压这一关键词链,通过聚合的不良信息可以排除左心室肥大为苯磺酸左旋氨氯地平的不良反应,揭示苯磺酸左旋氨氯地平的适应症为左心室肥大,而原发性高血压是左心室肥大的根本原因。

另外在网络中发现多巴胺的使用可能引起静脉炎,见图 5。以多巴胺作为第 1 级结点,与其相连的第 2 级结点有 48 个,包括酚妥拉明、心力衰竭等代表多巴胺联合应用治疗疾病的药物和其适应症的关键词,此外还包括静脉炎这一关键词,两者共现次数为 4,静脉炎不是多巴胺的适应症,因此其可能是多巴胺引起的不良反应,多巴胺和静脉炎的关系也在相关研究文献中得到了证实<sup>[11]</sup>。



现中的作用。将药物不良反应本体中的知识聚合到心血管类药物子网络中,进而在该网络中的药物-疾病对中排除药物的不良反应,所获得的药物-疾病对可能存在治疗药物与适应症或新不良反应的关系。在第 1 个案例研究中发现胰岛素抵抗可能为替米沙坦的新适应症。作为一种心血管类药物,替米沙坦主要用于高血压的治疗。胰岛素抵抗与替米沙坦共现的次数为 9,说明在所分析文献样本中万方医学网与维普数据库中有 9 篇相关研究文献,也说明替米沙坦在治疗胰岛素抵抗方面的研究仍处于起始阶段,这可能是替米沙坦的适应症研究的一个新方向。在第 2 个案例研究中,通过苯磺酸左旋氨氯地平-左心室肥大-原发性高血压这一知识链,揭示心血管药物所治疗疾病的发病原因。最后一个案例得到多巴胺和静脉炎这一关系对,显示静脉炎可能是多巴胺的一个新的不良反应。

#### 4.4 知识发现

随机选择药物不良反应领域中的 3 种药物进行案例分析,而关键词网络还含有大量关键词与关键词对,直接相连的关键词对与间接相关的词对为知识发现提供大量数据基础,将本体知识聚合到关键词网络中,在网络分析中发挥本体的推理功能,为大规模的知识发现提供可能。另外间接相关的词对可能提示药物与疾病或者药物与药物之间的新关联,这些关联为药物领域的实验科学家进行进一步研究提供科学假设。

#### 5 结语

本研究证实基于关键词网络进行知识聚合与知识发现方法的可行性,通过可视化图形显示关键词网络,在实例中发现胰岛素抵抗可能为替米沙坦新的适应症,通过苯磺酸左旋氨氯地平-左心室肥大-原发性高血压这一知识链,揭示出心血管药物所治疗疾病的发病原因,发现多巴胺可引起静脉炎这一不良反应。研究结果表明基于关键词网络进行知识聚合与知识发现的方法,将具有语义的知识聚合

到关键词网络中,可以提高图书情报领域的知识服务能力与利用率。

#### 参考文献

- 1 Wester K, Jonsson A K, Spigset O, et al. Incidence of Fatal Adverse Drug Reactions: a population based study [J]. *Br J Clin Pharmacol*, 2008, 65 (4): 573-579.
- 2 耿志杰, 王文鼎. 关键词同现网络结构研究 [J]. *情报杂志*, 2010, 29 (2): 14-16.
- 3 Yildirim M A, Goh K-I, Cusick M E, et al. Drug-target network [J]. *Nature Biotechnology*, 2007, 25 (10): 1119-1126.
- 4 Sun K, Buchan N, Larminie C, et al. The Integrated Disease Network [J]. *Integrative Biology*, 2014, 6 (11): 1069-1079.
- 5 Wang L, Liu H, Chute C G, et al. Cancer Based Pharmacogenomics Network Supported with Scientific Evidences: from the view of drug repurposing [J]. *BioData Mining*, 2015, 8 (1): 9.
- 6 Cheng F, Liu C, Jiang J, et al. Prediction of Drug-target Interactions and Drug Repositioning via Network-based Inference [J]. *PLoS Computational Biology*, 2012, 8 (5): e1002503.
- 7 Ontology of Cardiovascular Drug Adverse Events [EB/OL]. [2017-08-07]. <http://bioportal.bioontology.org/ontologies/OCVDAE>.
- 8 He Y, Sarntivijai S, Lin Y, et al. OAE: the ontology of adverse events [J]. *J Biomed Semantics*, 2014, (5): 29.
- 9 蒋秋明, 张熙洋, 王斌, 等. 替米沙坦对 2 型糖尿病患者血糖及胰岛功能影响的临床观察 [J]. *中华临床医师杂志 (电子版)*, 2013, 7 (14): 6663-6665.
- 10 周建敏, 郭文建, 于荣强. 替米沙坦对高血压伴代谢综合征患者血脂及胰岛素抵抗的影响 [J]. *中国医药*, 2013, 8 (9): 1222-1223.
- 11 苏严琳. 多巴胺致静脉炎的原因分析及护理进展 [J]. *中西医结合护理 (中英文)*, 2016, 2 (4): 167-170.
- 12 Chen C. *Mapping Scientific Frontiers: the quest knowledge visualization* [M]. Berlin: Springer Verlag, 2003.
- 13 刘芳. *信息可视化技术及应用研究* [D]. 杭州: 浙江大学, 2013.