

# 地理信息系统与健康相关 SCI 论文主题词聚类分析\*

范 婷 赵玉虹

(中国医科大学 沈阳 110122)

**[摘要]** 对 Web of Science<sup>TM</sup> 核心合集数据库中地理信息系统 (GIS) 与健康相关的 SCI 论文进行计量学分析, 统计高频主题词在同一文献中的出现频率并生成共现矩阵, 使用 SPSS 对矩阵进行聚类分析, 通过对各类文献内容分析进而获得该领域 4 个主要研究热点。

**[关键词]** 地理信息系统; 共现矩阵; 聚类分析

**[中图分类号]** R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2016.07.015

**Cluster Analysis of Subject Headings in SCI Papers on the Geographic Information System and Health** FAN Ting, ZHAO Yu-hong, China Medical University, Shenyang 110122, China

**[Abstract]** The paper makes a metric analysis of SCI papers on the Geographic Information System (GIS) and health in the database of Web of Science<sup>TM</sup> Core Collection, makes statistics of the occurrence frequency of high-frequency subject headings in the same literature and generates a co-occurrence matrix. Upon cluster analysis of the matrix with SPSS and analysis of various literature contents, it obtains four main research hotspots.

**[Keywords]** Geographic Information System (GIS); Word co-occurrence matrix; Cluster analysis

## 1 引言

采用文献计量学中的共词分析法对 2010-2016 年地理信息系统与健康相关的 SCI 论文研究领域结构和内部联系进行分析。共词分析法认为若两个能够表达文献主题的词条在一篇文献中同时出现, 则

两者之间具有一定的共现关系, 共现次数越多则关系越强<sup>[1]</sup>。对于一组文献, 将两两出现的主题词构建成共词网络, 网络间节点的远近表示了主题词的亲疏关系, 将文献群中的主题词进行聚类分析, 聚集在一起的若干主题词就构成一个研究主题领域<sup>[1-2]</sup>。

## 2 资料与方法

### 2.1 资料来源

所下载的数据来源于 Web of Science<sup>TM</sup> 核心合集数据库, 共词矩阵处理使用文献题录信息统计分析工具 SATI 实现<sup>[3]</sup>, 聚类分析通过 SPSS 19.0 软件实现。

**[收稿日期]** 2016-04-01

**[作者简介]** 范婷, 讲师, 发表论著 10 余篇; 通讯作者: 赵玉虹, 教授。

**[基金项目]** 2014 年辽宁省教育厅科学研究一般项目 (项目编号: L2014291)。

2.2 方法

地理信息系统可以为医疗健康产业提供科学技术支撑，成为辅助卫生研究，健康教育以及卫生项目计划、监测和评估的有利工具。本文以 Geographic Information System 与 Health 作为主题词，时间跨度 2010 – 2016 年，共计检索出 1 131 篇论文。使用 SATI 软件对检索结果进行处理，选择频次 ≥ 50 次的主题词作为高频主题词，生成主题词共现矩阵，

对同一文献中高频主题词的出现频率两两统计，使用 SPSS 19.0 对共现矩阵聚类分析，观察主题词之间的紧密程度并进行总结，分析出该领域研究热点。

3 结果

3.1 高频主题词的词频分布（表 1）

在所有关键词中，将出现频率 ≥ 50 次的主题词作为高频主题词，共有 10 个主题词。

表 1 地理信息系统与健康相关 SCI 论文高频主题词频数分布

| 序号 | 主题词                      | 频数  | 序号 | 主题词                                      | 频数 |
|----|--------------------------|-----|----|------------------------------------------|----|
| 1  | 健康（Healeh）               | 183 | 6  | 地理信息系统（Geographis Information – Systems） | 62 |
| 2  | 美国（United – States）      | 101 | 7  | 护理；照料（Care）                              | 62 |
| 3  | 死亡率（Mortality）           | 82  | 8  | 暴露（Exposure）                             | 56 |
| 4  | 风险（Risk）                 | 82  | 9  | 儿童（Children）                             | 55 |
| 5  | 运动量（Physical – activity） | 77  | 10 | 影响（Impact）                               | 54 |

3.2 高频主题词的共词聚类结果

如图 1 所示，纵轴的数字代表主题词编号，横轴数字代表主题词两两之间的距离，如果两个词在很短的距离聚集在一起，说明这两个词之间的关系紧密。如 2 号 United – States 和 3 号 Mortallity 连接的横轴很短，说明生存率相关研究在美国开展较多。通过 SCI 数据库中对主题词检索，验证了分类结果的正确性。经聚类图分析，结果显示 2010 – 2016 年，SCI 文章中关于该主题的研究热点主要包括 4 个方面。

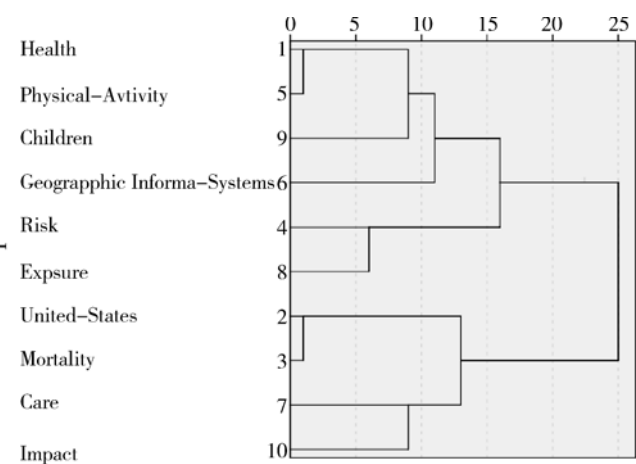


图 1 高频主题词聚类树形图

4 讨论

4.1 通过地理信息系统对儿童的运动量及健康展开的研究

Dowda M 等使用地理信息系统开展了针对 12 年级女孩家附近 75 英里范围内商业运动设施与心肺功能的相关性研究，通过混合回归模型得出结论：该范围内存在商业运动设施对提高被研究对象的心肺功能存在显著相关<sup>[4]</sup>。Cutumisu N 等人依据地理信息系统选取加拿大埃德蒙顿地区的 2 879 名调查对象进行逻辑回归预测，通过对健康情况、自我效能、运动量及该地区建筑环境指标的研究表明女性、老年人、受教育程度高的人群进行体育活动的可能性相对较小，通过干预整合为不同层次人群提供相应的运动及娱乐设施可促进该社区的整体活动水平<sup>[5]</sup>。De Guzman 等使用基于地理信息系统的多维模型试图确定距离公共交通的远近和居住密度评估了该社区的低收入人群的健康程度、心理健康状况、健康相关的局限性，实验结果表明两者之间没有明显关联<sup>[6]</sup>。Oluyomi AO 等将地理信息系统用于肥胖干预项目的研究，实验得出地理信息系统可以为调查对象、项目评估人员、政策制定者提供有效

的研究和实践依据,对社区健康研究以及社区居民健康评估方面有重要意义<sup>[7]</sup>。Meeredy G 等认为街区的互通性与青少年参加体育活动方面具有密切关系,为此针对 6~10 年级 8 535 名学生进行调查问卷,采用多元回归分析方法进行评估协变量之间的关联,结果表明两者之间没有明显联系<sup>[8]</sup>。

#### 4.2 美国将地理信息系统广泛应用于生存率方面的研究

Rereddy SK 等人认为医疗服务机构的空间分布不均严重影响癌症死亡率,该研究抽取 2010-2012 年患有颈部、头部癌症晚期患者进行地理信息系统定位,非裔聚居区由于缺少筛查,癌症晚期患者比重高,严重限制治疗方案、疾病恢复和生存率<sup>[9]</sup>。Smith ND 等人假设膀胱癌死亡率受环境和社会经济差异影响,使用 Getis-Ord Gi\* 方法统计从 1950-2007 年的膀胱癌患者死亡数据,使用地理空间加权回归分析,得出相关性一致假设成立<sup>[10]</sup>。MacDonald LA 等采用地理信息系统辅助电话访谈的报告法开展大数据异构采集,以职业为指标进行统计,研究表明职业可作为社会和环境替代因素用于人群健康领域的大数据研究<sup>[11]</sup>。Rutten LJF 等认为互联网是获取健康信息的重要来源,通过地理信息系统对 2009 年 8 个地区的人群随机调查,卡方检验和 T 检验进行回归分析,发现美国农村非裔人口使用互联网寻求癌症信息率仅 20%<sup>[12]</sup>。Brochu PJ 等认为城市经济发展程度和空气污染有直接关系,设计基于地理信息系统的 spatio-temporal 模型用来预测美国东北部地区的 PM 值,实验发现低发达地区空气污染程度相对较低<sup>[13]</sup>。

#### 4.3 地理信息系统针对暴露与风险方面的研究

Liang CP 等使用蒙特卡罗模型和连续性指标模型评估水产品养殖区砷暴露对地下水质量的影响,结论显示地理信息集成技术可以为渔业管理者提供有效的管理策略,通过控制鱼塘选址,降低危害人类健康的潜在风险<sup>[14]</sup>。Milla dos Santos 等依据地理信息系统对巴西圣保罗地区由于空气污染引发的健康风险进行评估,检测 2007-2011 年空气中 SO<sub>2</sub>、

NO<sub>2</sub>、O<sub>3</sub>、CO、PM<sub>10</sub> 的平均最高浓度结合该地区的人群密度,发现圣保罗市区及周边地区健康风险系数较高,表明将模糊逻辑法结合地理信息系统用于健康风险评估及暴露指数具有较高的科研价值<sup>[15]</sup>。Van-Hieu V 等对越南海防市城市道路交通颗粒物污染引发的死亡率和疾病率进行了风险评估,以集成模型和地理信息系统(Integration of Modelling and GIS)建模对交通状况和颗粒物浓度进行估计,结果表明 2007 年该市由于悬浮颗粒物致死人员达 1 200 名,两项指标显著相关<sup>[16]</sup>。Tseng Eva 等针对台湾 43 227 名医疗服务行业公务人员心血管死亡率与细颗粒物(PM<sub>2.5</sub>)的相关性进行了长达 18 年(1989-2008 年)的队列研究。通过 COX 模型调整混杂因素后,结果显示 1 992 名死亡人数中仅有 230 人死于心血管疾病,从而论证 PM<sub>2.5</sub> 值与心血管疾病死亡率无显著相关<sup>[17]</sup>。

#### 4.4 地理信息系统对护理与影响方面的能效分析

Sudhof L 等依据地理信息系统对卢旺达乡村地区应急产科展开研究,统计分析 2009 年卢旺达社区健康中心登记的乡镇地区新生儿数量、8 个健康中心和地区医院登记的临床数据,研究发现距离医院更远的地区剖腹产率更低,但该地区产后婴儿死亡率及子宫破裂的发生率相对较高<sup>[18]</sup>。Angier H 等对 2011 年 52 家诊所的电子病例进行研究,选取问诊记录超过 1 次的 0~64 岁患者进行横断面分析。研究发现地理信息系统可以显示患者保险费率地区差异和年龄差异,证明应用地理信息系统能够有效识别地区间差异指标<sup>[19]</sup>。Fortney JC 等应用地理信息系统研究地理位置对身心健康问题服务的影响,探索美国境内印第安人使用的传统康复治疗有关生物医学服务的应用。通过对西南部和北部平原两个部落包括道路交通距离、海拔及边界在内的地理信息进行评估,结果显示存在心理健康问题的人群不受地理位置影响<sup>[20]</sup>。Ross JM 等对 2009-2012 年乌干达地区初级卫生诊所进行了横断面研究,探讨地理障碍是否影响成人肺结核患病率,应用地理信息系统结合道路、土地覆盖率、住宅与健康中心的距离、个人的出行时间等指标评估相关性,结果

显示距离和出行时间与肺结核均不相关<sup>[21]</sup>。

## 5 结语

近年来除了上述研究以外, 仍有学者提出了许多全新的观点, 由于这类研究的主题词频次较低, 故而没有纳入本次研究中, 如 Brawner BM 等认为不同地区人群艾滋病免疫功能受到社会环境和社会结构的影响, 种族和地理信息系统结合的映射模型可有效探索多层次复杂因素在艾滋病高速传播地区<sup>[22]</sup>。这些研究仍然很有价值, 需要去挖掘、发现和思考。

## 参考文献

- 1 刘启元, 叶鹰. 文献题录信息挖掘技术方法及其软件 SATI 的实现——以中外图书情报学为例 [J]. 信息资源管理学报, 2012, (1): 50-58.
- 2 钟伟金, 李佳, 杨兴. 共词分析法研究 (三) ——共词聚类分析法的原理与特点 [J]. 情报杂志, 2008, (7): 118-120.
- 3 宋伟, 顾志良, 景凡伟, 等. 地理信息系统 (GIS) 在医疗和公共卫生领域的应用 [J]. 解放军医院管理杂志, 2010, (10): 981-982.
- 4 Dowda M, Pfeiffer KA, Lobelo F, et al. Cardiorespiratory Fitness and Proximity to Commercial Physical Activity Facilities Among 12th Grade Girls [J]. Journal of Adolescent Health, 2012, 50 (5): 497-502.
- 5 Cutumisu N, Spence JC. Sport Fields as Potential Catalysts for Physical Activity in the Neighbourhood [J]. International Journal of Environmental Research and Public Health, 2012, 9 (1): 294-314.
- 6 DeGuzman PB, Merwin EI, Bourguignon C I. Population Density, Distance to Public Transportation, and Health of Women in Low-Income Neighborhoods [J]. Public Health Nursing, 2013, 30 (6): 478-490.
- 7 Oluyomi AO, Byars A, Byrd-Williams C, et al. The Utility of Geographical Information Systems (GIS) in Systems-Oriented Obesity Intervention Projects: The Selection of Comparable Study Sites for a Quasi-Experimental Intervention Design - TX CORD [J]. Childhood Obesity, 2015, 11 (1): 58-70.
- 8 Mecredy G, Pickett W, Janssen I. Street Connectivity is Negatively Associated with Physical Activity in Canadian Youth [J]. International Journal of Environmental Research and Public Health, 2011, 8 (8): 3333-3350.
- 9 Rereddy SK, Jordan DR, Moore CE. Dying to be Screened: Exploring the Unequal Burden of Head and Neck Cancer in Health Provider Shortage Areas [J]. Journal of Cancer Education, 2015, 30 (3): 490-496.
- 10 Smith ND, Prasad SM, Patel AR, et al. Bladder Cancer Mortality in the United States: a geographic and temporal analysis of socioeconomic and environmental factors [J]. Journal of Urology, 2016, 195 (2): 290-296.
- 11 MacDonald LA, Pulley LV, Hein MJ, et al. Methods and Feasibility of Collecting Occupational Data for a Large Population-based Cohort Study in the United States: the reasons for geographic and racial differences in stroke study [J]. BMC Public Health, 2014, 14 (1): 1-12.
- 12 Rutten LJF, Hesse BW, Moser RP, et al. Socioeconomic and Geographic Disparities in Health Information Seeking and Internet Use in Puerto Rico [J]. Journal of Medical Internet Research, 2010, 14 (4): 199-210.
- 13 Brochu PJ, Yanosk JD, Paciorek CJ, et al. Particulate Air Pollution and Socioeconomic Position in Rural and Urban Areas of the Northeastern United States [J]. American Journal of Public Health, 2011, 101 (1): S224-S230.
- 14 Liang CP, Jang CS, Liu CW, et al. An Integrated GIS-Based Approach in Assessing Carcinogenic Risks via Food-Chain Exposure in Arsenic-Affected Groundwater Areas [J]. Environmental Toxicology, 2010, 25 (2): 113-123.
- 15 Milla dos SAP, Passuello A, Schuhmacher M, et al. A Support Tool for Air Pollution Health Risk Management in Emerging Countries: a case in Brazil [J]. Human and Ecological Risk Assessment, 2014, 20 (5): 406-424.
- 16 Van-Hieu Vu, Xuan-Quynh Le, Ngoc-Ho Pham, et al. Application of GIS and Modelling in Health Risk Assessment for Urban Road Mobility [J]. Environmental Science and Pollution Research, 2013, 20 (8): 5138-5149.
- 17 Tseng Eva, Ho WC, Lin Meng-Hung, et al. Chronic Exposure to Particulate Matter and Risk of Cardiovascular Mortality: cohort study from Taiwan 1993 [J]. BMC Public Health, 2015, (15): 936.

(下转第 81 页)

阶段,还存在诸多问题,嵌入式馆员嵌入用户环境的深度服务,也不是一朝一夕能够完成的,必须长期不断地积累更多的经验,总结发现的难点问题加以解决,及时改进嵌入式服务模式,才能逐步渗透和完善嵌入式服务,更深入地融入用户需求与工作环境中,为不同层次、类型的用户提供多层次的信息服务,在探索和实践过程中体现出嵌入式馆员自身的创造性和知识潜能。

### 参考文献

- 1 吴延辉. 新世纪临床医学图书馆员发展方向 [J]. 医学信息学杂志, 2009, 30 (12): 86-88.
- 2 谭丹丹, 刘金涛. 对学科化服务背景下嵌入式图书馆服务的思考: 定位、关键步骤及挑战 [J]. 图书馆杂志, 2012, 31 (2): 51-55.
- 3 罗亚泓. 国内外嵌入式图书馆服务研究及主要观点 [J]. 图书情报工作, 2014, 58 (6): 123-128.
- 4 程容溪. 医院图书馆嵌入式服务探讨 [J]. 医学信息学杂志, 2013, 34 (10): 71-74.
- 5 周晓杰, 刘海听, 张春杨. 我国图书馆嵌入式服务研究述评 [J]. 图书馆学研究, 2012, (12): 19-22.
- 6 初景利. 学科馆员对嵌入式学科服务的认知与解析 [J]. 图书情报研究, 2012, 5 (3): 1-8.
- 7 林燕. 嵌入式馆员——图书馆用户服务新理念 [J]. 图书馆学研究 (理论版), 2010, (1): 71-74.
- 8 初景利, 孔青青, 栾冠楠. 嵌入式学科服务研究进展 [J]. 图书情报工作, 2013, 57 (22): 11-17.
- 9 姚敏, 王璇, 尹智, 等. 新时代图书馆员角色定位及重塑 [J]. 医学信息学杂志, 2014, 35 (2): 75-78.
- 10 黄岩. 泛在知识环境下医院图书馆嵌入式学科服务 [J]. 中华医学图书情报杂志, 2013, 22 (5): 25-27.
- 11 Sudhof L, Amoroso C, Barebwanuwe P, et al. Local Use of Geographic Information Systems to Improve Data Utilisation and Health Services: mapping caesarean section coverage in rural Rwanda [J]. Tropical Medicine & International Health, 2013, 18 (1): 18-26.
- 12 Angier H, Likumahuwa S, Finnegan S. Using Geographic Information Systems (GIS) to Identify Communities in Need of Health Insurance Outreach: an OCHIN Practice-based Research Network (PBRN) report [J]. Journal of the American Board of Family Medicine, 2014, 27 (6): 804-810.
- 13 Fortney JC, Kaufman CE, Pollio DE, et al. Geographical Access and the Substitution of Traditional Healing for Bio-medical Services in 2 American Indian Tribes [J]. Medical Care, 2012, 50 (10): 877-884.
- 14 Ross JM, Cattamanchi A, Miller CR, et al. Investigating Barriers to Tuberculosis Evaluation in Uganda Using Geographic Information Systems [J]. American Journal of Tropical Medicine and Hygiene, 2015, 93 (4): 733-738.
- 15 Brawner BM, Reason JL, Goodman BA, et al. Multilevel Drivers of Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome Among Black Philadelphians Exploration Using Community Ethnography and Geographic Information Systems [J]. Nursing Research, 2015, 64 (2): 100-110.

(上接第 69 页)

欢迎订阅 欢迎赐稿