

知识图谱分析法

一、概念

(一) 定义

知识图谱 (Knowledge Mapping, 在图书情报界也称为知识域可视化或知识领域映射地图, 科学知识图谱), 是显示知识发展进程与结构关系的一系列各种不同的图形, 用可视化技术描述知识资源及其载体, 挖掘、分析、构建、绘制和显示知识之间的相互关系。它通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合, 并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的的现代理论。它把复杂的知识领域通过数据挖掘、信息处理、知识计量和图形绘制而显示出来, 揭示知识领域的动态发展规律, 为学科研究提供切实的、有价值的参考。

知识图谱本质上是一种语义网络。其结点代表实体 (entity) 或者概念 (concept), 边代表实体/概念之间的各种语义关系。它可以用可视化技术描述人类随时间拥有的知识资源及其载体, 绘制、挖掘、分析和显示科学知识以及它们之间的相互联系, 在组织内创造知识共享的环境以促进科学技术研究的合作与深入。

(二) 特点

1. 用户搜索次数越多, 范围越广;
2. 赋予字串新的意义, 而不只是单纯的字串;
3. 融合了所有的学科, 以便于用户搜索时的连贯性;
4. 把与关键词相关的知识体系系统化地展示给用户;
5. 从整个互联网汲取有用的信息让用户能够获得更多相关的公共资源;
6. 为用户找出更加准确的信息, 作出更全面的总结并提供更有深度相关的信息。



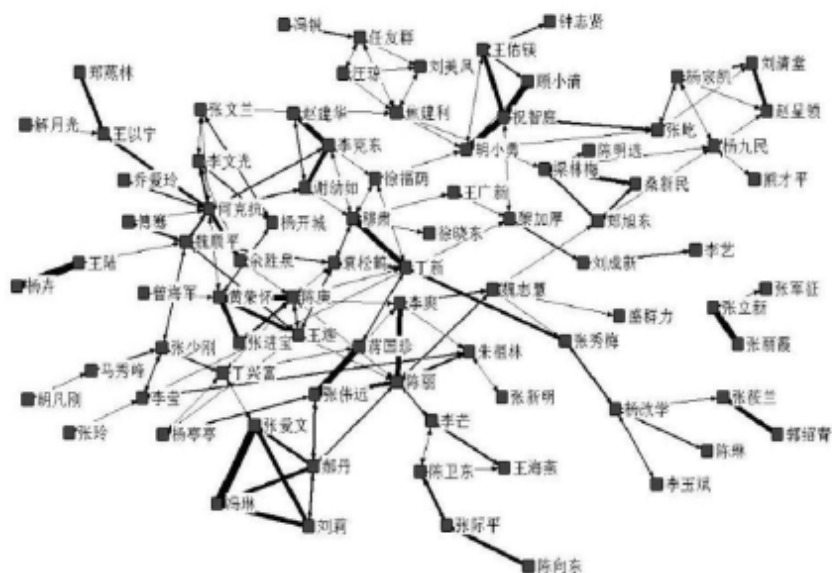


图 1 我国教育技术学者合著网络知识图谱

二、知识图谱描述的对象及应用

(一) 知识图谱描述的对象

- 1.从事科学技术活动和作为知识载体的人，包括科学家、技术家、项目组、实践团体、或某一知识领域共同体。
- 2.学科领域内显性或编码化的知识（作者、机构、专利、期刊和其他出版物等）之间的关系。
- 3.过程或方法，包括研究问题和解决问题的过程或方法、组织的业务流程以及相关的知识投入等。
- 4.某一学科主要研究领域之间的内部联系，各研究领域之间的知识输入与知识输出。

(二) 知识图谱的应用

知识图谱主要应用于组织知识管理、线上虚拟社区管理、教育学习等方面。目前应用较多的在科研、教学、社会学，科研用于文献、专利的结构分析；学科动态、社会网络、领域发展分析。

表 1 知识图谱在国内国外的应用

学科	国内	国外
管理学	社区及社区服务、生态可持续发展、中小企业管理、德尔菲法、管理学、技术创新、企业知识共享、专利研究	知识管理、科技人才资源、政府绩效管理、区域创新系统、先进制造技术、管理学、ICT领域、技术创新、虚拟社区、ERP、企业风险管理、管理伦理、技术预见、创新管理、创新系统、企业知识共享、专利信息
经济学	数量经济学、教育经济学、金融学、经济学、物流学	信息经济学、物流研究、会计学、旅游研究、社会资本研究、创业研究

图书管理学	图书情报学、档案学、情报学、图书馆学、知识图谱、Citespace、期刊评价、参考咨询、生命周期理论、搜索引擎、公共与高校图书馆、信息可视化、《情报科学》、《图书馆理论与实践》、开放存取、引文分析	图书情报学、情报学、图书馆学、竞争情报、医学情报、网络搜索引擎、电子文件管理、信息资源管理、h指数、知识组织、网络计量学、文献计量学、知识计量学、知识发现、信息政策、公共信息资源管理、信息素养、期刊影响因子、加菲尔德、引文分析、科学知识图谱
科学学	《中国科技期刊研究》、科学学、科学基金管理、科技政策	同行评议专家、科学计量学家、科学计量学、科技政策、科学传播学、科学学、《科学哲学》
教育学	教育技术学、教育研究、《高等教育研究》、高等教育、独立学院、民办高等教育	E-learning、创业教育、创业新大学、二语习得领域、高等工程教育、远程教育、高等教育、大学战略管理
其他学科	自然辩证法、农史学科、力学、制浆造纸、极地研究、《体育科学》、体育科学、脑认知研究、循证医学	自然辩证法、农史学科、力学、制浆造纸、极地研究、《体育科学》、体育科学、脑认知研究、循证医学

三、绘制工具与绘制方法

(一) 绘制工具

对于知识图谱的绘制工具可分为专用工具如Bibexcel、HiteCite、CiteSpace、NWB Tools和TD A；通用工具如统计分析软件如SPSS和SAS，词频分析工具Word Smith和Word Cat，社交网络分析工具如Pajek和Ucinet，可视化工具如VOSviewer、Net Draw等。

表 2 知识图谱绘制工具

软件名称	软件简介	功能描述
CiteSpace	陈超美博士开发的专门用于科学知识图谱绘制的免费软件。国内使用最多知识图谱绘制软件	可用于追踪研究领域热点和发展趋势，了解研究领域的研究前沿及演进关键路径，重要的文献、作者及机构。可用于对ISI、CSSCI和CNKI等多种文献数据库进行分析。
Bibexcel	瑞典科学计量学家Persoon开发的科学计量学软件，用于科学研究免费软件	具有文献计量分析、引文分析、共引分析、耦合分析、聚类分析和数据可视化等功能。可用于分析ISI的SCI、SSCI和A&HCI文献数据库。
HistCite	Eugene Garfield等人于2001年开发的科学文献引文链接分析和可视化系统，免费软件	可对ISI的SCI、SSCI和SA&HCI等文献数据库的引文数据进行计量分析，生成文献、作者和期刊的引文矩阵和实时动态引文编年图。直观的反映文献之间的引用关系、主题的宗谱关系、作者历史传承关系、科学知识发展演进等。
SPSS	大型统计分析软件，商用软件	具有完整的数据输入、编辑、统计分析、报表、图形绘制等功能。常用于多元统计分析、数据挖掘和数据可视化。

TDA	Thomson Data Analyzer (TDA) 是Thomson集团基于VantagePoint开发文献分析工具, 商用软件	具有去重、分段等数据预处理功能; 可形成共现矩阵、因子矩阵等多种分析矩阵; 可使用Pearson、Cosine等多种算法进行数据标准化; 可进行知识图谱可视化展示。
Sci2 Tools	印第安纳大学开发的用于研究科学结构的模块化工具	可从时间、空间、主题、网络分析和可视化等多角度, 分析个体、局部和整体水平的知识单元。
ColPal Red	Gradnada大学开发的共词单元文献分析软件, 商用软件	结构分析, 在主题网络中展现知识(词语及其关系); 战略分析, 通过中心度和密度, 在主题网络中为主题定位; 动态分析, 分析主题网络演变, 鉴定主题路径和分支。
Leydesdorff系类软件	阿姆斯特丹大学Leydesdorff开发的这对文献计量的小程序集合	处理共词分析、耦合分析、共引分析等知识单元体系。使用“层叠图”实现可视化知识的静态布局和动态变化。
Word Smith	词频分析软件	可将文本中单词出现频率排序和找出单词的搭配词组。
NWB Tools	印第安纳大学开发的对大规模知识网络进行建模、分析和可视化工具	数据预处理; 构建共引、共词、耦合等多种网络; 可用多种方法进行网络分析; 可进行可视化展示。
Ucinet NetDraw	Ucinet是社会网络分析工具, 包括网络可视化工具Net Draw	用于处理多种关系数据, 可通过节点属性对节点的颜色、形状和大小等进行设置。用于社交网络分析和网络可视化。
Pajek	来自斯洛文尼亚的分析大型网络的社会网络分析免费软件	Pajek基于图论、网络分析和可视化技术, 主要用于大型网络分解, 网络关系展示, 科研作者合作网络图谱的绘制。
VOSviewer	荷兰莱顿大学开发的文献可视化分析工具	使用基于VOS聚类技术实现知识单元可视化工具。突出特点可视化能力强, 适合于大规模样本数据。四种视图浏览: 标签视图、密度视图、聚类视图和分散视图。

(二) 绘制方法

1. 文献计量方法

科学知识图谱属于科学计量学, 因此必然文献计量学的方法, 主要包括:

(1) 共词分析: 属于内容分析法的一种。它的原理主要是对一组词两两统计它们在同一篇文献中出现的次数, 以此为基础对这些词进行聚类分析, 生成共词文献簇, 进而分析这些词所代表的学科和主题的结构变化。利用共词分析法及其相关的可视化方法可以进行深入的主题分析, 系统而直观地了解学科结构和发展状况, 并进行学科发展预测。

(2) 引文分析方法: 利用各种数学、统计学方法和比较、归纳、抽象、概括等逻辑方法, 对科学期刊、论文、著者等各种分析对象的引用与被引用现象进行分析, 以便揭示其数量特征和内在规律的一种文献计量分析方法。

(3) 链接分析方法: 利用图计算、拓扑学和文献计量学等方法, 对网络连接文档、自身属性、连接对象、连接网络等进行分析。链接分析涉及文档包括: 页面、目录、域名和站点。在理论上, 连接分析与文献计量学中的引文分析有高度的相似。链接分析运用拓扑学知识通过分析链接网络来研究网络结构, 结合社会网络分析可以分析研究和绘制网络信息知识图谱, 展示网络信息、知识分布结构和演化规律等。

(4) 共引分析方法: 1973年斯莫尔等人就提出把文献共引分析作为计量文献之间关系的一种方法。共引 (co-citation) 又称被共引, 即两篇文献同时被一篇或多篇文献说引用, 同时把共同引用这两篇文献的文献数称为共引强度 (或共引频率), 共引强度越大这两篇文献关系越密切。在共引图谱中, 点表示文献, 当相关文献对的共被引强度等于或大于某个阈值时, 两点就被连接起来。共引分析多用于作者共引分析和期刊共引分析。

(5) 耦合分析方法: 与共引分析相对应的是耦合分析。几篇文献具有相同的参考文献就形成了文献耦合关系。具有相同参考文献的文献数称为耦合强度。耦合分析包括文献耦合分析、期刊耦合分析、作者耦合分析、学科耦合分析等, 分别表示文献、期刊、作者、学科之间具有主题和内容相似性, 可做相关文献分析、作者群体分析和科学演化分析等的依据。

(6) 词频分析方法: 是齐普夫定律为理论基础进行文献内容分析方法。词频分析可分为标题关键词词频分析、摘要词频分析、内容词频分析、引文词频分析和混合词频分析等。词频分析大量应用于科学前沿主题领域和发展趋势等研究。

2. 统计分析方法

科学知识图谱构建实用的统计分析方法主要是多元统计分析。多元统计分析是经典统计学的分支, 在多个对象或指标相互关联的情况下分析其统计规律。“维度降低技术”是多元统计分析的一个特征, 从几何学看这个过程是讲高维空间的目标投影到低维空间。主要的其中包括:

(1) 因子分析 (主成分分析): 用少数几个因子来描述许多指标之间的关系, 即将较密切的几个变量归为同一类, 每一类变量成为一个因子, 以较少的几个因子来反应原资料的大部分信息。

(2) 多维尺度分析 (MDS): 通过低维空间展示作者 (文献) 之间联系, 并利用平面距离来反映作者 (文献) 之间的相似度。多维尺度分析的图形显示结果更加直观和形象, 因子分析则更容易确定各个学术群体的边界和数目, 因此需要同时借助因子分析的结果, 进行知识图谱的绘制。

3. 数据挖掘方法

数据挖掘是指从大量的数据中通过算法提取、挖掘未知的、有价值的模式或规律等知识的复杂过程。科学知识图谱的绘制使用了很多数据挖掘方法, 常用的方法有聚类分析、数据可视化和社会网络分析等。

(1) 聚类分析: 将物理或抽象的对象集合分成相似的对象类的过程。簇是数据对象的集合, 同一个簇中的对象彼此相似, 而不同的簇彼此相异。文献聚类分析是聚类分析技术在引文分析中的具体应用。处理方法是文献通过分词、去停用词等方法转化为词向量, 并将每个词条赋予不同的权重, 这样一篇文献就可以由词条权重值组成的特征向量来表示, 所有文献将组成了特征向量空间模型, 在该模型中使用聚类分析技术进行引文分析。

(2) 数据可视化: 也成为信息可视化是指将抽象数据用图形图像等可视化形式表示出来, 以利于分析数据、发现规律和支持决策。常用的可视化算法有: ①自组织特征映射网络SOM (Self-organizing Feature Map) 是一种基于神经网络的算法, 它通过把高维数据映射到低维空间进行聚类, 并保持一定的拓扑有序性。②寻径网络图谱PFNET (PathFinder Network) 是对不同的概念或实体间联系的相似或差异程度进行评估, 应用图论中原理和方法生成的一类特殊的网状模型。

(3) 社会网络分析 (Social Network Analysis) 也称为结构分析, 是将社会结构界定为一个网络, 这个网络由成员之间的联系进行连接。社会网络分析聚焦于成员之间的联系而非个体特征, 并把共同体视为“个体的共同体”, 即视为人们在日常生活中所建立、维护并应用的个人关系的网络。社会网络分析方法被证明可以成功的研究科学合作网络和互联网络所得到的可视化网络, 并被用于展示科学计量学的合作网络结构与发展。

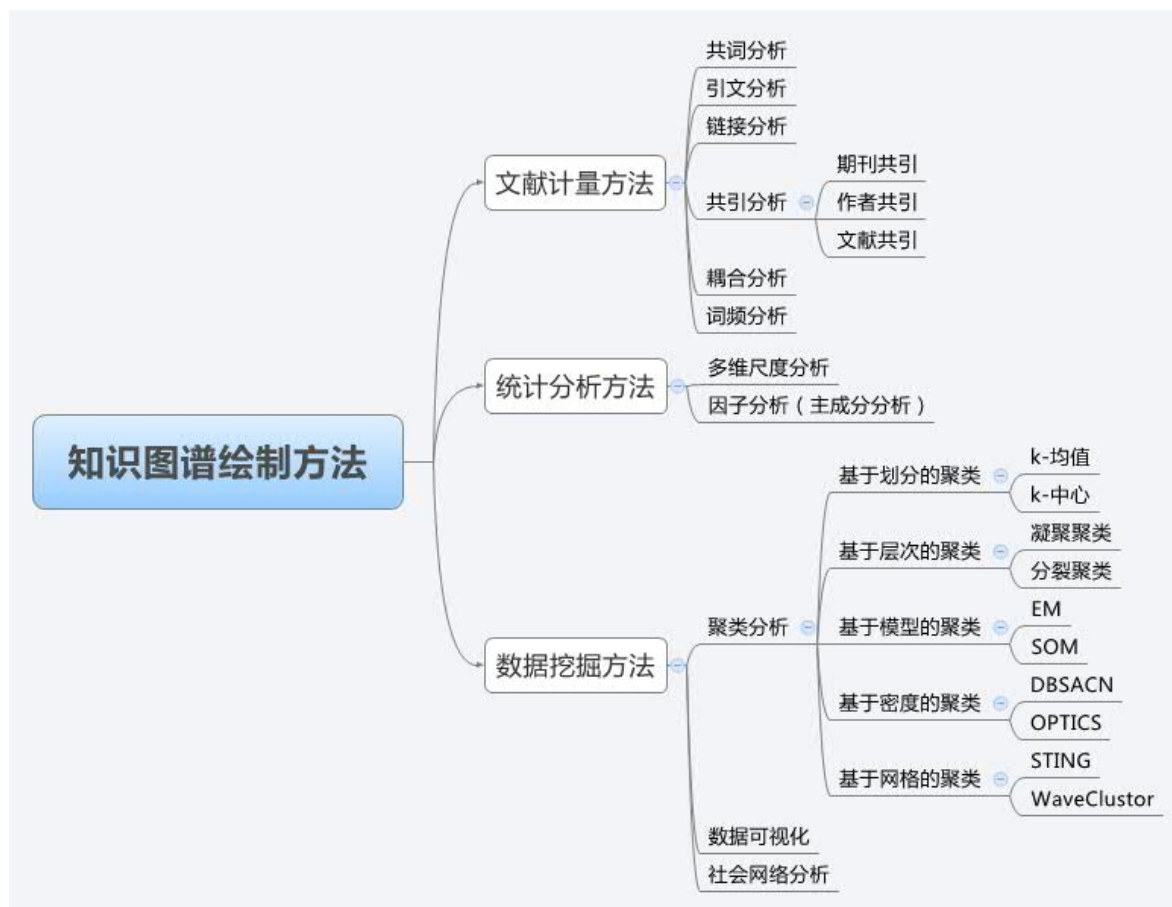


图 2 知识图谱绘制方法

四、操作步骤

科学知识图谱绘制过程可分为8个步骤：


- 1.样本数据检索（对期刊数据CNKI WOS、Science Direct或网络数据库Google Scholar、CitSee r等）；
- 2.数据预处理（分词、去停用词、去重、勘误等）；
- 3.选择知识单元（作者、关键词、机构、期刊和文献）；
- 4.构建单元关系（引文分析、共现分析、共引分析、耦合分析、词频分析等）；
- 5.数据标准化（Cosine、Jaccard、Equivalence、Association Strength等）；
- 6.数据分析（因子分析、多维尺度分析、自组织映射图、寻径网络图谱、聚类分析和潜在语义分析等）
- 7.知识可视化（几何图、主题河图、星团图、冲积图、地形图等）；
- 8.图谱解读（历时分析、突变分析、空间分析、网络分析、地理分布、浏览查询、放大缩小、过滤关联等）。








图 3 知识图谱绘制步骤


五、案例

 近十年我国电子书包研究热点与发展趋势_基于共词矩阵的知识图谱分析_王佑镁.pdf (http://lcell.bnu.edu.cn/UserFiles/73080/File/近十年我国电子书包研究热点与发展趋势_基于共词矩阵的知识图谱分析_王佑镁.pdf)

 国际创业型大学研究主流理论与热点图谱分析_易高峰.pdf (http://lcell.bnu.edu.cn/UserFiles/73080/File/国际创业型大学研究主流理论与热点图谱分析_易高峰.pdf)

 我国教师教育研究的文献计量分析_2000_2012年_李泮泮.pdf (http://lcell.bnu.edu.cn/UserFiles/73080/File/我国教师教育研究的文献计量分析_2000_2012年_李泮泮.pdf)

 我国教育技术学者合著网络知识图谱构建研究_陈瑜林.pdf (http://lcell.bnu.edu.cn/UserFiles/73080/File/我国教育技术学者合著网络知识图谱构建研究_陈瑜林.pdf)

 Analysis on the Technology Front of Workflow Based on KnowledgeMapping.pdf (<http://lcell.bnu.edu.cn/UserFiles/73128/PDF/Analysis on the Technology Front of Workflow Based on KnowledgeMapping.pdf>)



讨论区 (<http://lcell.bnu.edu.cn/do/bbsPost?action=detail&postId=24304>)



六、测试



知识图谱分析法练习测试 (<http://lcell.bnu.edu.cn/do/testTask?action=visitTestFromLa&laid=5933&testId=841>)

七、参考文献

- [1]王佑镁,陈慧斌.近十年我国电子书包研究热点与发展趋势--基于共词矩阵的知识图谱分析[J].中国电化教育,2014,(05).
- [2]易高峰.国际创业型大学研究主流理论与热点图谱分析[J].清华大学教育研究,2009,(10).
- [3]李泮泮.我国教师教育研究的文献计量分析(2000-2012年)[J].教师教育研究,2014,(5).
- [4]陈瑜林.我国教育技术学者合著网络知识图谱构建研究[J].远程教育杂志,2012,(6).
- [5]杨思洛,韩瑞珍.知识图谱研究现状及趋势的可视化分析[J].情报资料工作,2012,(4).
- [6]白文倩,李文昊.国际教育技术学学术群体知识图谱构建与分析[J].中国电化教育,2013,(6).
- [7]杨思洛,韩瑞珍.国外知识图谱绘制的方法与工具分析[J].图书情报知识,2012,(6).
- [8]陈悦,刘则渊,陈劲等.科学知识图谱的发展历程[J].科学学研究,2008,(3).
- [9]胡芳槐.基于多种数据源的中文知识图谱构建方法研究[D].华东理工大学,2015.
- [10]伍国华,李克东.知识可视化教学应用的元分析研究[J].电化教育研究,2011,(12).
- [11]叶平浩.国内高等职业教育研究热点与前沿的可视化分析[J].中国高教研究,2012,(9).
- [12]赵慧臣,王玥,张舒予.可视化分析方法在我国教育研究领域中的元分析[J].现代教育技术,2014,(1).
- [13]曹树金,吴育冰,韦景竹,马翠嫦.知识图谱研究的脉络、流派与趋势——基于SSCI与CSSCI期刊论文的计量与可视化[J].中国图书馆学报,2015,(5).

学习元评论 0条



聪明如你,不妨在这发表你的看法与心得~



登录之后可以发表学习元评论。

友情链接



(<http://slp.bnu.edu.cn>)
智慧学伴
(<http://slp.bnu.edu.cn>)



(<http://aic-fe.bnu.edu.cn>)
未来教育高精尖
创新中心 (<http://aic-fe.bnu.edu.cn>)



(<http://mllab.bnu.edu.cn>)
“移动学习”
联合实验室
(<http://mllab.bnu.edu.cn>)

关于我们

联系我们

快速入门

意见反馈

学习元平台 (www.etc.edu.cn) 由移动学习教育部-中国移动联合实验室开发和维护为国家教育体制改革试点项目 (04-207-283) “面向农村的教育开放学院” 提供知识分享、协作交流、教师培训等专业服务。

© 2017 北京师范大学 -  (<http://creativecommons.org/licenses/by-nc-sa/4.0/>)

