

基于论文题目的知识驱动关系 分析方法研究*

——以信息与知识传播研究领域为例

张运良

(1. 中国科学技术信息研究所 北京 100038;
2. 富媒体数字出版内容组织与知识服务重点实验室 北京 100038)

摘要:[研究目的] 基于论文题名快速梳理特定研究领域的概念,并确定概念所代表的知识之间驱动关系,有助于科学了解研究领域状况,为进一步寻找研究方向提供可能。[研究方法] 以信息与知识传播研究领域为例,收集整理题名中包含“知识”研究论文元数据,在 Stanford Parser 的短语结构分析结果基础上,对驱动关系进行人工审定,并做边界调整、拆分和规范化,得到概念间的真实驱动关系数据,定义若干分析指标,以上述数据为基础构建加权有向图,并基于复杂网络分析和可视化技术分析展示在知识驱动方面具有不同特点的概念及驱动路径。[研究结论] 该研究借鉴和发展共词分析方法,提出并实现了一种能够揭示研究领域概念之间有方向的知识驱动关系的方法,利用本方法可以发现不同概念在某一领域的驱动强度大小、特定的驱动路径和若干稳定或脆弱的概念团组。

关键词:知识驱动关系;论文题名;短语结构分析;复杂网络分析

中图分类号:G353.1

文献标识码:A

文章编号:1002-1965(2022)01-0170-07

引用格式:张运良. 基于论文题目的知识驱动关系分析方法研究[J]. 情报杂志, 2022, 41(1): 170-176.

Research on the Analytical Method for Knowledge-Driven Relationship Based on Title of Papers —— Taking Information and Knowledge Dissemination Domain as an Example

Zhang Yunliang

(1. Institute of Scientific & Technical Information of China, Beijing 100038;
2. Key Laboratory of Rich-Media Knowledge Organization and Service of Digital Publishing Content, Beijing 100038)

Abstract: [Research purpose] Based on the title of the papers, the concepts in a specific domain can be sorted out quickly, and the driven relationship between the knowledge represented by the concepts is determined, which is helpful to understand the status of the domain, and provides the possibility for further research approach exploring of the domain. [Research method] Taking information and knowledge dissemination domain as an example, the metadata of research papers containing "knowledge(in Chinese)" in the title was collected and sorted out. Based on the results of phrase structure analysis with Stanford Parser, the driven relationship was manually examined, and the concept boundary was adjusted, relationship are split and standardized, so as to obtain the real driving data between concepts, define several analysis indexes, build a weighted directed graph based on the data above, and analyze and demonstrate the knowledge-driven relationship based on complex network analysis and visualization technology. [Research conclusion] The Co-word analysis method is referenced and improved, and a method that can reveal the knowledge-driven relationship between subject concepts, which is a stronger and more accurate association, is put forward and implemented. Using this method, we can find the driven relationships between different concepts in a

收稿日期:2021-05-24

修回日期:2021-07-09

基金项目:中国科学技术信息研究所重点work项目“多模态知识图谱构建及应用研究”(编号:ZD2021-11);中国工程科技知识中心建设项目“知识组织体系建设”(编号:CKCEST-2021-2-6);中国科学技术信息研究所创新基金项目“知识驱动的评论类情报甄别方法研究”(编号:MS2020-05)的研究成果之一。

作者简介:张运良,男,1979年生,博士,研究员,研究方向:知识工程、情报理论方法。

certain domain, the specific driven paths and some concept groups with stable or fragile relationship.

Key words: knowledge-driven relationship; title of papers; phrase structure analysis; complex network analysis

知识是人类在实践中认识客观世界的成果,是人类智慧的结晶^[1]。在从基础的数据升华到智慧的复杂过程中,研究者提出多种模型,知识都在其中占据一席之地^[2-3],而且都处于从数据和信息发展到智慧的关键位置。同时,知识也是人类改造客观世界的有力工具,在推动人类发展和社会进步的过程中具有重要的作用,当前人工智能如火如荼,但以数据驱动的人工智能发展面临瓶颈,将知识驱动与数据驱动结合发展第三代人工智能的新思路被提出^[4],以知识驱动的检索、问答、甄别等应用日益普遍,知识重要性进一步凸显。

知识驱动关系是特定研究领域重要概念之间的一种关系,这种关系体现了概念之间有方向的影响程度。研究驱动关系影响的强度与路径,有助于进一步布局关键和高潜力研究方向,科学有效的引导领域发展。近年来,情报工程正向着更高的分析精准度方向发展^[5],本文是在这个方向上的一个尝试,将定量和定性相结合,形成系统了解和梳理不同研究领域概念之间驱动关系的体系方法。

为了研究和验证相关分析方法,本文选择信息与知识传播研究领域作为实例,为了降低问题难度,本文用论文题名代替全文进行分析。在信息与知识传播研究领域“知识”及相关概念是最为重要的研究内容之一,对于揭示本研究领域及相关研究领域概念之间驱动关系具有较为充分的代表性。深入理解和分析论文文本内容是最为理想的方法,但是一方面由于数据量大导致处理难度较大,另一方面无关的论述也可能为关键知识驱动关系的确定带来干扰。为此本文基于题名进行研究,其合理性在于“题名是科技论文的必要组成部分,它要求用最简洁、恰当的词语来准确表达论文的中心内容”^[6],是全文分析的非常好的替代。本文基于采集的信息与知识传播研究领域题名数据,经短语结构分析后,以介词短语和动词短语为切入点,抽取领域概念和概念间知识驱动关系,并利用图分析为主的方法对本研究领域中的知识驱动情况进行分析。

1 相关研究

目前,尚未见直接研究知识驱动关系分析的文献,但具有相关的分析基础。驱动关系是相关关系的特例,更是共现关系的特例,原有基于共现统计的共词分析等方法可以作为分析的基础;但是驱动关系要体现方向,这是目前共词分析尚不能解决的,本文基于图分析对文本内容进行深入分析。共词分析和图分析方法

相关研究都有重要的借鉴意义。

1.1 共词分析方法 共词分析是文献计量的常用分析方法,通过分析在同一个文本主体中的款目(单词或名词短语对,以后简称为词对)共同出现的形式,以发现科学领域的学科结构的一种分析方法^[7]。共词分析方法基于这样的一个基本认识:某一词对在同一篇文献中出现的次数越多,则代表这两个词之间的关系越紧密。进而扩大范围,统计一组文献的主题词两两之间在同一篇文献出现的频率,便可形成一个由这些词对关联所组成的共词网络,网络内词对之间的距离便可以反映对应概念之间的关联关系。这种数值化的直观表示方法已经被广泛应用到自然科学和社会科学领域^[8-9],在研究领域知识揭示上发挥了重要的作用^[10]。当前,共词分析已经发展形成较为完善的流程,包括 SPSS、Ucinet、CiteSpace 等在内的系列工具,可以有效支撑各个领域的基本分析、聚类分析和可视化展示^[11]。

1.2 图分析方法 将相关主题、概念或者因素纳入到一个图的结构来进行研究,会对领域形成更为直观的认识。图分析主要研究连通性、可遍行性、可平面性、匹配性质、染色性质、某些特殊结构、特殊点子集等性质^[12]。实际上共词分析也需要建立图结构,但是一般为无向图。在本领域中具有代表性的图分析对象包括概念地图、认知地图和因果图。概念地图是利用概念以及概念之间的关系表示关于某个主题的结构化知识的一种图示方法。概念地图用节点表示概念,用连接线和连接语表示概念之间的关系;节点和连线按照宽泛概念在上、具体概念在下的顺序,形成等级结构^[13]。概念地图有时候有方向,但一般揭示的是定性的关系,没有对关系进行量化。认知地图是一种认知映射工具,是一种使领域概念和概念之间关系同时显示的可视化表达,是一个由节点构成的有向图,通过因果链接,建立起多个实体之间的关系。也有研究将其泛化,而将因果图(Causal Map)作为其一个特例,研究者通过深度访谈确定了“关于低碳经济下武钢可持续发展认知地图”中重要概念以及影响的方向和权重的正负,并通过调查问卷方法求得具体的权重数值,为决策提供了支撑^[14]。驱动关系构成的图类似于认知地图。总之图分析,分析对象涵盖无向图和有向图,但是有向图包含更为丰富的信息,有助于进一步揭示事物内在的规律。

2 数据准备

2.1 采集、清洗 本文研究的数据采集自万方数据

(采集日期为 2021 年 3 月 18 日),由于只分析题名,因此检索时在“题名”字段中用“知识”作为检索词,同时将网站定义的学科分类限定在“文化、科学、教育、体育”大类下的“信息与知识传播”类目,采集数据共 31 187 条。

由于题名本身长度和命名规则各异,为了将驱动关系提取出来,首先用“驱动”来筛选全部数据记录,得到结果 55 条,作为候选集合的一部分,但是数据量较小,大量隐含的驱动关系没有体现出来。通过对题名进行观察,发现大多数情况下是复杂的短语,可以先进性短语结构分析。本文选取 Stanford parser^[15] 进行分析,并重点关注其中的介词 P,经统计共出现 10 276 条,这是因为一个句子中可能出现多个介词,如“关于信息服务向知识服务转变的深度思考”的分析结果如图 1 所示,该题名包含“关于”和“向”两个介词。对全部介词进行统计分析,发现所有的介词共 56 种,其中出现 10 次以上的有 24 种,如表 1 所示。其中有一个英文介词“of”看上去比较突兀,因此返回到原始题名分析,发现题名中包含了“Web of Science”或“Web of Knowledge”,体现的是研究的数据源。其中“基于”“对”“以”“从”“为”等属于频次较高并且体现驱动关系可能性较大的介词,以此筛选出 5 288 条题名,然后

对介词后的近邻的名词短语 NP 进行抽取,部分示例如表 2 所示。但是存在包含这些介词但未体现驱动关系的情况,部分示例如表 3 所示。

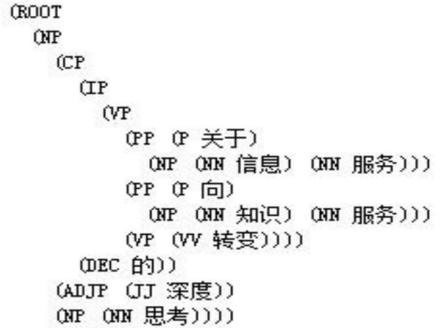


图 1 包含双介词的题名短语结构分析结果示例

表 1 题名中出现的高频介词统计

序号	介词	词频	序号	介词	词频	序号	介词	词频
1	基于	3637	9	到	133	17	据	28
2	在	1433	10	论	91	18	由	28
3	对	1334	11	向	89	19	作为	22
4	与	1181	12	用	73	20	of	18
5	以	947	13	自	33	21	面向	18
6	从	479	14	和	30	22	按	13
7	关于	367	15	于	29	23	通过	13
8	为	152	16	给	28	24	当	12

表 2 部分知识驱动关系抽取示例

序号	题名	短语结构分析结果	抽取的驱动关系
1	基于平衡记分卡的知识链管理	(ROOT (NP (DNP (PP (P 基于) (NP (NN 平衡) (NN 记分卡))) (DEG 的)) (NP (NN 知识链) (NN 管理))))	平衡记分卡→知识链管理
2	基于知识管理的图书馆服务方式研究	(ROOT (IP (VP (PP (P 基于) (NP (DNP (NP (NN 知识) (NN 管理))) (DEG 的)) (NP (NN 图书馆) (NN 服务) (NN 方式)))) (VP (VV 研究))))	知识管理→图书馆服务方式
3	知识管理对图书馆虚拟参考咨询服务的理论阐释	(ROOT (IP (NP (NN 知识) (NN 管理)) (VP (PP (P 对) (NP (DNP (NP (NP (NP (NN 图书馆)) (ADJP (JJ 虚拟)) (NP (NN 参考))) (NP (NN 咨询) (NN 服务))) (DEG 的)) (NP (NN 理论)))) (VP (VV 阐释))))	知识管理→图书馆虚拟参考咨询服务
4	以知识经济理论重塑参考咨询	(ROOT (IP (VP (PP (P 以) (NP (NP (NN 知识) (NN 经济)) (NP (NN 理论)))) (VP (VV 重塑) (NP (NN 参考) (NN 咨询))))	知识经济理论→参考咨询
5	从知识经济论高校图书馆的资源管理	(ROOT (IP (VP (PP (P 从) (NP (NN 知识) (NN 经济))) (VP (VV 论) (NP (DNP (NP (NN 高校) (NN 图书馆)) (DEG 的)) (NP (NN 资源) (NN 管理))))	知识经济→高校图书馆资源管理
6	图书馆如何为知识经济服务	(ROOT (IP (NP (NN 图书馆)) (VP (ADVP (AD 如何)) (PP (P 为) (NP (NN 知识) (NN 经济))) (VP (VV 服务))))	图书馆→知识经济

表 3 部分包含介词但未体现驱动关系的题名示例

序号	题名	介词	短语结构分析结果
1	发展机构知识库的若干问题探讨——基于机构知识库调研的思考	基于	(ROOT (IP (NP (DNP (NP (NP (NN 发展) (NN 机构)) (NP (NN 知识库))) (DEG 的)) (QP (CD 若干) (NP (NN 问题))) (VP (VV 探讨)) (PRN (PU ——) (NP (DNP (PP (P 基于) (NP (NN 机构) (NN 知识库) (NN 调研))) (DEG 的)) (NP (NN 思考))))
2	图书馆为弱势群体提供知识援助:市场经济体制下的“悖论”	为	(ROOT (UCP (IP (NP (NN 图书馆)) (VP (PP (P 为) (NP (ADJP (JJ 弱势)) (NP (NN 群体)))) (VP (VV 提供) (NP (NN 知识) (NN 援助)))) (PU :) (NP (DNP (LCP (NP (NN 市场) (NN 经济) (NN 体制)) (LC 下)) (DEG 的)) (NP (NN " ") (NN 悖论) (NN "))))

续表3 部分包含介词但未体现驱动关系的题名示例

序号	题名	介词	短语结构分析结果
3	从文本知识源中挖掘概念格的形式分析	从	(ROOT (IP (NP (CP (IP (VP (PP (P 从) (LCP (NP (NN 文本) (NN 知识源)) (LC 中))) (VP (VV 挖掘) (NP (NN 概念) (NN 格)))))) (DEC 的)) (NP (NN 形式)) (VP (VV 分析))))))
4	如何对图书馆进行知识重组	对	(ROOT (IP (VP (ADVP (AD 如何)) (PP (P 对) (NP (NN 图书馆))) (VP (VV 进行) (NP (NN 知识) (NN 重组))))))

2.2 自动筛选处理、人工审校 在上述5288条记录基础上进行人工校对处理,主要的处理策略包括规范化、拆分、调整边界和删除。其中记录规范化主要是去除复杂的修饰成分,虽然这些修饰成分实际上可能体现了作者的创新,在其他的分析中可能有重要的作用,但这里需要从本质上进行归纳,部分归纳示例如表4所示。记录拆分主要针对并列的概念,拆分过程中实际上也包含了规范化的处理,部分示例如表5所示。调整边界主要针对因短语结构分析错误引起的驱动关系识别错误,部分示例如表6所示。删除处理主要对于关系提取错误无法调整或者规范化前后关系前后概念皆不包含“知识”的情况,如“泛在知识环境下基于

用户需求的高校图书馆服务模式研究”提取出“用户需求→高校图书馆服务模式”的驱动关系应删除。

表4 部分规范化的驱动关系示例

原始驱动关系	规范化后的驱动关系
平衡记分卡→知识链管理	平衡记分卡→知识管理
知识管理→图书馆服务方式	知识管理→图书馆服务
知识管理→图书馆虚拟 参考咨询服务	知识管理→参考咨询

表5 部分拆分的驱动关系示例

原始驱动关系	拆分后的驱动关系
共引与共词分析→国内移动商务核心知识识别	共引分析→核心知识识别 共词分析→核心知识识别
机构知识库→知识分析及可视化	机构知识库→知识分析 机构知识库→知识可视化

表6 部分驱动关系边界调整示例

题名	短语结构分析结果	提取的驱动关系	调整边界后的驱动关系	题名	短语结构分析结果	提取的驱动关系	调整边界后的驱动关系
基于本体的注塑模改模知识表达与推理研究	(ROOT (IP (VP (PP (P 基于) (NP (DNP (NP (NN 本体)) (DEG 的)) (NP (NN 注塑模)))) (VP (VV 改模) (NP (NN 知识) (NN 表达)) (CC 与) (NP (NN 推理) (NN 研究))))))	本体的注塑模→知识表达与推理研究	本体→知识表达 本体→知识推理	基于工作流的知识地图及其在企业知识管理中的应用	(ROOT (IP (VP (PP (P 基于) (NP (NP (NP (NP (NN 工作) (NN 流的)) (NP (NN 知识) (NN 地图))) (CC 及) (NP (NP (PN 其)) (PP (P 在) (NP (NN 企业))) (NP (NN 知识) (NN 管理)))))) (VP (VV 中的) (NP (NN 应用))))))	工作流的知识地图→其在企业管理	工作流→知识地图 知识地图→知识管理

2.3 待分析数据集情况说明 以上数据记录经过人工审核,确认在3928个题名中,共标识出4124条驱动关系,每条驱动关系包含两个概念节点,至少其中一个概念节点中包含“知识”,以此作为后续分析的数据基础,严格来讲,这些节点应该是词语而不是概念,但是本文假设通过规范化能够将词语形成概念,只是不同情况下规范化程度不同,不影响本文的分析方法研究。

出度、入度等概念,并进一步定义驱动强度、绝对驱动强度、相对驱动强度、纯主动驱动概念、纯被动驱动概念、驱动链条、驱动关系强度,强驱动链条、驱动回路等概念,如下。

定义1:驱动强度即该概念作为主体驱动其他概念的强度,分为绝对驱动强度和相对驱动强度。

定义2:绝对驱动强度以该概念在有向图的中出度的单调增函数定义,本文简单定义为该概念在有向图中的出度。

定义3:相对驱动强度以该概念在有向图的中出度的单调增函数和入度的单调减函数定义,本文简单定义为有向图中出度减去入度的差。

3 分析研究

3.1 相关指标定义 为了更好的分析驱动关系,特别对部分指标定义做说明。本文引入图分析中的度、

定义 4:纯主动驱动概念即有向图中入度为 0 的概念。

定义 5:纯被动驱动概念即有向图中出度为 0 的概念。

定义 6:驱动链条即在驱动关系形成的有向图中,由两个或者两个以上概念形成的节点及相关的边形成的具有单一方向的路径。

定义 7:驱动关系强度即对特定两个概念从驱动主体到驱动客体之间驱动实例的个数。

定义 8:强驱动链条即链条上任意驱动关系强度都达到一定阈值以上的驱动链条。

定义 9:驱动回路即形成了闭环的驱动链条。

3.2 “知识”相关的核心概念统计分析 依据绝对驱动强度排名 TOP20 概念如表 7 所示,由于最后两个概念强度相同,实际上有 21 个概念。依据相对驱动强度排名 TOP20 概念如表 8 所示,由于驱动强度相同原因,相对驱动强度最高的概念有 22 个,相对驱动强度最低的概念有 21 个。找出纯主动驱动概念依据其出度排名,纯被动驱动概念依据其入度排名, TOP20 概念如表 9 所示,由于出度或入度相同原因,纯主动驱动概念有 21 个,纯被动驱动概念有 22 个。

表 7 绝对驱动强度 TOP20 概念

Table with 6 columns: 序号, 概念, 绝对驱动强度, 序号, 概念, 绝对驱动强度. Lists 20 concepts and their absolute driving strengths.

表 8 相对驱动强度 TOP20 概念

Table with 8 columns: 序号, 高驱动强度概念, 出度, 入度, 相对驱动强度, 低驱动强度概念, 出度, 入度, 相对驱动强度. Lists 20 concepts and their relative driving strengths.

续表 8 相对驱动强度 TOP20 概念

Table with 8 columns: 序号, 高驱动强度概念, 出度, 入度, 相对驱动强度, 低驱动强度概念, 出度, 入度, 相对驱动强度. Continuation of Table 8, listing concepts 16-22.

表 9 纯主动驱动概念和纯被动驱动概念 TOP20

Table with 5 columns: 序号, 纯主动驱动概念, 出度, 纯被动驱动概念, 入度. Lists 20 concepts categorized as pure active or pure passive drivers.

Table with 5 columns: 序号, 概念, 出度, 概念, 入度. Lists 20 concepts with their out-degree and in-degree values.

图 2 出度和入度 TOP20 节点关联

依据上述驱动关系形成的复杂网络进行分析,提取出度和入度前 20 名(其中出度第 20 名有两个并列)并做标识,如图 2 所示,可以观察到在出入度前 20 名的概念中,重复的概念有 8 个,并且其排序也有所差异,说明这 8 个概念在驱动其他概念,同时也被其他概念驱动,其中“知识管理”和“知识图谱”的出度大于入度,相对驱动强度更高,而其余 6 个概念,出度小于入

度,相对驱动强度较低。

3.3 核心驱动链条研究 为了更好的观察不同概念之间的驱动关系,找到核心驱动链条,使用 Graphviz 和 PyGraphviz 绘图,采用椭圆表示概念节点,椭圆的高度用节点度的对数加上一个常数(这里取 0.5)定义。两个节点之间的边的宽度用两个节点之间实际产生驱动关系数量的平方根来定义。考察图需要选择特定的驱动关系强度阈值,首先设定阈值为 10 的驱动关系及相关概念进行展示分析,可知在本领域最核心驱动主体是“知识管理”,其次是“知识图谱”“本体”,最后是“DSpace”,“知识管理”驱动了“图书馆管理”“图书馆服务”“信息服务”等,但是“本体”也是“知识管理”的驱动主体,并且同时也驱动了“知识组织”“知识库”和“知识检索”。“知识管理”驱动“知识服务”,并和“知识服务”一起驱动“图书馆”和“档案管理”,当然关于“知识服务”还有“知识构建”“用户需求”“大数据”等驱动主体。“知识图谱”的唯一驱动主体是“CiteSpace”,可见这里的知识图谱仍然是“Knowledge Mapping”的含义,这是图情领域在过去对知识图谱使用最多的内涵,而非近年来更广泛使用的“Knowledge Graph”,图中还有一条独立的驱动链条,即“DSpace”对“机构知识库”的驱动。

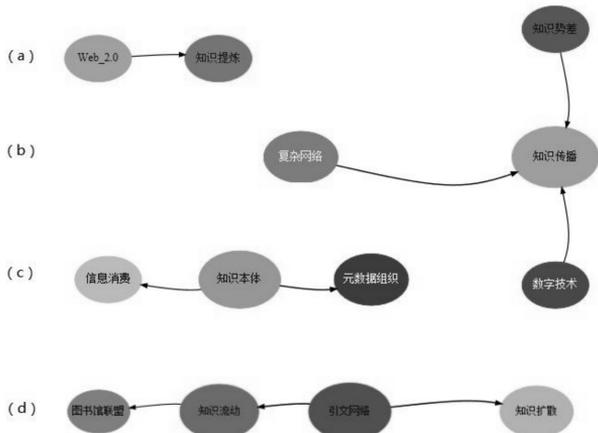


图3 部分未与最大连通子图连通的子图示例

把驱动关系强度阈值降低为 5,大体结构不变,只是增加了更多的驱动关系细节。如“文献→知识发现→知识服务→……”,“关联数据→知识发现→知识服务→……”,“关联数据→知识组织→知识服务→……”,“主题图→知识组织→知识服务→……”等链条,“知识图谱”的驱动主体增加了“VOSviewer”,同时增加了“SNS”对“知识社区”的驱动。进一步降低驱动关系强度阈值至 2 时,由于边过多,整体可读性降低,但仍然可以通过局部放大的方式进行观察,“知识图谱”这一概念通过对“档案管理”的驱动,与“知识管理”所属的最大连通子图(这里连通性依据有向图的底图考察)连接在一起,但是产生了更多的相对独立

的驱动链条,一些子图式样如图 3 所示。

3.4 驱动回路研究 仅保留驱动回路,可以发现,核心结构仍然由“知识管理”“知识服务”和“图书馆”“数字图书馆”构成,由于“数字图书馆”属于“图书馆”的下位概念,所以实际上是由“知识管理”“知识服务”和“图书馆”构成了稳定的三角结构,还可以提取出多个驱动回路。但是这些回路仍主要通过核心的驱动三角回路发挥作用,在与“知识管理”关联的概念中,存在“开放获取↔机构知识库↔学科服务”回路,此外还有“分类↔知识发现↔知识图谱↔计量分析”回路,以及“知识共享↔教育”“Wiki↔知识组织”两条较短的回路。如果把驱动关系强度阈值逐步提高到 3,如图 4 所示,发现唯一的驱动回路变成了“知识管理↔知识服务↔图书馆”,尽管各边驱动强度不同,但是这是目前在研究领域内最重要的回路。

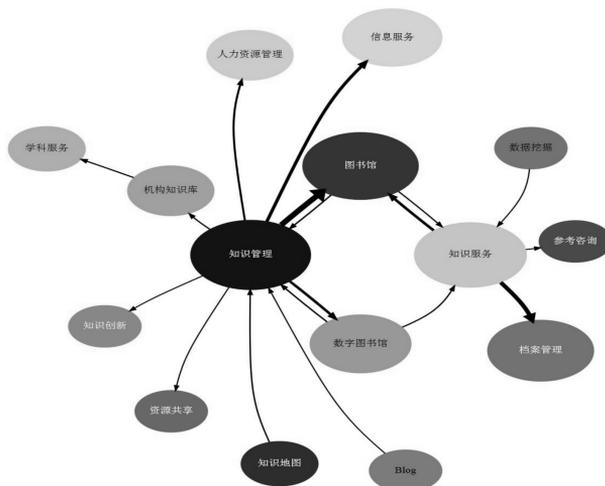


图4 驱动回路(驱动关系强度阈值3)

3.5 与共词分析方法比较 共词分析方法的典型分析结果包括共词矩阵、高频词聚类分析树状图、类团分析(包括凝聚子图)、多维尺度分析(包括战略坐标图)等等,揭示了领域内高度关联的概念,但是重要程度的分析主要基于概念本身的频率和共现频率,而知识驱动关系分析则在共现基础上,明确了概念之间关系的方向。在共词分析中,如果两组概念的共现次数是一样的,则对其解读是一样的,在知识驱动关系中还要区分是概念 A 影响概念 B,还是概念 B 影响概念 A,或者两者皆有,则两组关系也可能大相径庭,以此为视角能够得到一系列新的观察结果。如在本文的数据中,“语义网-知识服务”和“知识管理-人力资源管理”均共现 9 次,但是前者 9 次均是“语义网”驱动“知识服务”,“语义网”是一个纯主动驱动概念,后者中有 8 次是“知识管理”驱动“人力资源管理”,还有 1 次是“人力资源管理”驱动“知识管理”,体现了两个概念相互影响,但是以“知识管理”驱动“人力资源管理”为主的情况。当然由于共词分析不需要进行方向的分析,

实现相对容易,而本文分析方法需要引入短语结构分析,实现起来相对困难,也正是基于此,本文采用了以题名代表论文的简化研究方法,减少分析的数据量和可能带来的干扰。

4 结论与展望

本文在对论文题名短语结构分析的基础上,通过将概念之间的驱动关系表示为加权有向图,提出了相关指标及分析方法,并在信息与知识传播研究领域进行实证研究。本文发展了共词分析方法,本文能够观察到一些独特的关联和性质,如:a.不同概念在某一领域的驱动强度大小,又可以从绝对驱动强度和相对驱动强度来分别进行分析;b.特定的驱动路径,包括强驱动路径和驱动回路,为进一步干预和影响相关工作提供了可参考的依据;c.若干具有稳定或脆弱驱动关系的概念团组。从驱动强度角度对领域研究全貌进行了解,有助于寻找重点攻关的研究内容和潜在的研究方向

本文研究还存在一定的改进空间。首先是概念的归并问题,作者在拟定题名的时候往往使用不同的术语,或者加以限定修饰,需要加以归并,概念归并的程度和准确性在一定程度上会对领域概念驱动强度大小、驱动链条和驱动回路的发现、整体连通性等的观察研究产生影响,但不影响本方法的有效性。其次是题名的代表性问题,尽管绝大多数题名代表了论文的核心内容,但是仍有一些题名过于简单,或题名中概念存在歧义,这些可能还需要通过对摘要和全文的理解才能实现。最后是驱动关系的发现和确认,本文采用了基于部分介词和动词为核心的分析方法,但是在具体处理上,仍然需要人工干预。

本文中对知识驱动关系强度的确定是等权的,未来可以对其类型进行区分并设置不同权重,以得到更为准确的分析结果。本文分析高频概念间的驱动关系并未考虑时间因素,未来可以进一步研究驱动网络的

时间演化、萌芽状态驱动关系的发现、各不连通子图之间的潜在可能驱动关系预测。

参考文献

- [1] 《中国情报学百科全书》编委会. 中国情报学百科全书[M]. 北京:中国大百科全书出版社, 2010.
- [2] 刘嘉. 网络信息资源的组织:从信息组织到知识组织[M]. 北京:北京图书馆出版社, 2002.
- [3] 杜元清. 论信息环境设计[D]. 北京:北京大学, 2009.
- [4] 张钹,朱军,苏航. 迈向第三代人工智能[J]. 中国科学:信息科学, 2020, 50(9):7-28.
- [5] 刘琦岩. 情报工程须统合精益分析和高阶分析[J]. 情报工程, 2018, 4(6):1.
- [6] 李胜春. 科技论文标题的实例分析[J]. 重庆大学学报(社会科学版), 2001(4):66-68.
- [7] 图书馆·情报与文献学名词审定委员会. 图书馆·情报与文献学名词[M/OL]. [2021-05-19]. <https://baike.baidu.com/item/共词分析/53304287>.
- [8] 杨莲莲,杨之音,杨朝峰. 基于共词分析的微生物学植物学领域研究热点分析[J]. 情报工程, 2016, 2(4):96-103.
- [9] 黄丽娜,黄璐,邵晓. 基于共词分析的中国互联网政策变迁:历史、逻辑与未来[J]. 情报杂志, 2019, 38(5):83-91, 70.
- [10] 钟伟金,李佳. 共词分析法研究(一)——共词分析的过程与方式[J]. 情报杂志, 2008, 27(5):70-72.
- [11] 唐果媛,张薇. 基于共词分析法的学科主题演化研究进展与分析[J]. 图书情报工作, 2015, 59(5):128-136.
- [12] 高随祥. 图论与网络流理论[M]. 北京:高等教育出版社, 2009.
- [13] 马费成,郝金星. 概念地图在知识表示和知识评价中的应用(I)——概念地图的基本内涵[J]. 中国图书馆学报, 2006(3):5-9.
- [14] 张凌. 基于认知地图的隐性知识表达与共享[M]. 武汉:武汉大学出版社, 2011.
- [15] Levy R, Manning C. Is it harder to parse Chinese, or the Chinese Treebank? [C] // ACL. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics; vol. 1. Stroudsburg: Association for Computational Linguistics, 2003: 439-446.

(责编/校对:王育英)