DOI:10.13530/j.cnki.jlis.166005

多维度视角下学科主题演化可视化分析方法研究
——以我国图书情报领域大数据研究为例*

刘自强　王效岳　白如江

摘　要　探测、识别某学科领域研究主题的演化过程并进行可视化分析，对于掌握研究现状和发展趋势具有重要意义。学科主题演化是一个复杂过程，存在多种变量，如主题强度、结构和内容等，目前研究主要以单一维度进行可视化分析，信息负荷过大，存在感知局限性。本文提出多维度视角下学科主题演化可视化分析方法：通过人工标注方法对关键词进行语义角色分类，利用 Fast Unfolding 算法识别出具有语义特征的学科主题；利用余弦相似度计算公式计算学科主题相似度判定演化关系；构建多维度学科主题演化分析模型，并设计了三种创新性的科学知识图谱，进行学科主题强度、结构和内容三个维度的可视化分析，通过相互作用可以帮助快速消化、理解信息和精炼分析结果，有效地分析学科主题演化的复杂过程。通过对我国图书情报领域近 10 年大数据研究的实证分析，证明该方法具有可行性和有效性。图 9。表 4。参考文献 37。

关键词　学科主题演化　语义角色标注　社区发现算法　可视化

分类号　G301

Research on Visualization Analysis Method of Discipline Topics Evolution from the Perspective of Multi-Dimensions: A Case Study of the Big Data in the Field of Library and Information Science in China

LIU Ziqiang, WANG Xiaoyue & BAI Rujiang

ABSTRACT

Detection and identification of the evolution of research topics in a discipline has important significance for researchers to grasp its research status and development trend. Visual analysis can show the relationship between themes based on topics recognition, help users to enhance their perception and cognition, and to find useful information quickly in a field on the research status, research hotspots and development trends, and to digest, understand and effectively analyze vast amounts of information. However, discipline topics evolution is a complex process and there are many variables, such as the intensity, structure and content of topics. The single dimension visualization analysis causes the information overload, leading to three problems: perceptive limitations, cognitive limitations, and performance limitations.

* **本文系国家社会科学基金项目“未来新兴科学研究前沿识别研究”(编号:16BTQ083)的研究成果之一。**(This article is an outcome of the project “Research on the Identification of Emerging Scientific Research Frontier in the Future”(No. 16BTQ083) supported by National Social Science Foundation of China.)

通信作者:王效岳, Email: sdutcspace@163.com, ORCID: 0000-0002-7100-7758 (Correspondence should be addressed to WANG Xiaoyue, Email: sdutcspace@163.com, ORCID: 0000-0002-7100-7758)

This paper presents a visualization analysis method of discipline topics evolution from a multidimensional perspective: using the artificial annotation method to make semantic role classification of keywords, using Fast Unfolding algorithm recognition with the semantic features to identify the topics; using cosine similarity to calculate the formula of similarity between topics evolution; constructing evolution analysis model of multidimensional discipline topics, and designing three innovative scientific knowledge map by using JavaScript and Web front-end visualization technology to analyze the visualization of the intensity, structure and content of the topics. Through the interaction of the core areas of research questions, the evolution path and trend of research methods and key technologies of the topics, and the macro evolution trend, meso evolution process and microscopic evolution details can be revealed. It can effectively help to quickly digest and understand information and refine the topic evolution analysis results to reveal the complex process of topics evolution.

The experiment on 'big data study in library and intelligence field in the past 10 years' proves the visualization analysis method proposed in this paper could effectively demonstrate the complicated process of topics evolution in a discipline. Compared with the other visualization analysis methods, the proposed method based on topic strength, structure and internal basic knowledge unit evolution could visualize and analyze the core research points, the primary research methods, the key technology topics evolution path and trends and the evolution trend from the macro-angle, the evolution process from meso-angle and the evolution particulars from the micro angle, so as to better analyze the complicated topics evolution process. The proposed method is applicable only to scientific papers including key words and needs to be expanded to include other data resources.

The practical significance lies in: 1) Proposing the dynamic topics identification method to provide reference to related research; 2) Proposing a multidimensional topics evolution analysis model and an innovative visualization method which can be used to analyze the topics evolution rules and discover scientific and technological knowledge, etc; 3) The findings can be used in scientific and research management to support decision-making, raise research efficiency and help to promote the scientific and technological innovation in related fields. 9 figs. 4 tabs. 37 refs.

KEY WORDS

Discipline topics evolution. Semantic role labeling. Community discovery algorithm. Visualization.

0 引言

随着党的"十八大"提出"创新驱动发展战略",科技信息服务机构、科研人员加大了科技情报分析的研究力度,力求通过分析科技文献特别是学术论文,借助自然语言处理(Natural Language Processing,NLP)、科学计量和可视化方法充分挖掘学科主题之间的演化关系,揭示学科研究主题的内在联系、总体结构特征、演化路径等。评估特定学科领域的研究现状、研究热点、研究前沿和发展趋势,可以有效地提高科研效率,合理配置科技资源,辅助科研决策和促进相关领域开展科技创新[1]。

学科主题演化分析发展至今已经有半个世纪的历史,研究人员对某一研究领域或某一学科的主题演化分析不再只停留于文本、数据层面的处理分析上,而是逐渐进入可视化层面,通过可视化技术将某学科领域的研究现状、研究热点、研究前沿和发展趋势形象直观地展现出

来。可视化分析可以在学科主题识别基础上展现主题之间的关系，帮助人们更准确地把握信息的脉搏，增强用户的洞察力和认知，帮助其快速消化、理解信息，有效地分析海量信息。

目前，学科主题演化可视化分析方法将所有主题词同等看待，而实际科技文献中的主题词有的代表研究问题，有的代表研究方法，将所有主题词同样看待会导致主题演化分析准确度的下降。而且目前可视化方法主要基于某一可视化软件（如 UciNet、SciMAT、CiteSpace、SPSS 等）进行单一维度的主题演化可视化分析，比如利用 UciNet 生成社会网络图能够展示主题分布及其联系，但是无法展示主题强度演化等情况；利用 SciMAT 生成战略坐标图能够展示主题间联系、结构和演化路径，但是无法展示内部主题词的微观变化情况以满足用户的细粒度信息需求。

为了弥补目前研究中存在的不足，本文提出创新性的学科主题演化可视化分析方法，构建多维度学科主题演化分析模型，利用 Web 前端可视化技术研究设计与之契合的科学知识图谱，实现该模型的可视化分析。该方法能多维度、准确有效地揭示学科主题的复杂演化过程。

1　相关研究

1.1　动态主题识别

如何高效地从海量科技文献中识别出隐含的主题及其演化趋势一直是情报学的研究重点，国内外专家学者提出了大量算法、模型。与本文主题相关的主要有两类：动态主题模型和网络社区演化模型。

2003 年，Blei 等提出 LDA 模型发现文本中的主题[2]，可以基于统计概率层面表达词间语义层次关系，但是不能解释主题演化情况；为了弥补不足，2006 年，Blei 等又提出了动态主题模型，可以处理具有时间戳记的文档数据集，实现动态主题识别与追踪[3]。2015 年，祝娜等提出基于 LDA 模型的科技创新主题演化路径识别方法，能够分析某学科领域创新主题的动态演化过程[4]。基于主题模型的学科主题演化分析方法，能在一定程度上改善引文分析方法的时滞性问题，处理多种类型数据，适应性较好，而且促进主题演化分析可以实现更深层次的语义分析，深入揭示学科研究主题的微观发展动态；不足之处在于揭示的是统计概率层面的语义关系，不能充分满足科技创新中的深度语义信息需求。

网络社区（Network Community）或网络簇结构（Network Cluster）是复杂网络最普遍和最重要的拓扑结构属性之一[5]，社区探测（Community Detection）是复杂网络研究中的一个重要分支。2009 年，Wallace 等人研究发现社区发现算法在科研主题识别上具有天然的优势[6]。2013 年，程齐凯等研究发现学术文献中的关键词共词网络内也存在社区现象，共词网络内的社区可以表征学科研究主题，而且共词网络中社区的演化在一定程度上揭示了学科研究主题的演化发展过程[7]。基于社区发现算法的网络社区演化模型，相较于词频、共词分析方法，能够准确有效地发现学科研究复杂网络中的隐含社区（主题），不足是将关键词同等看待。实际上文献中的关键词具有不同的语义角色，比如同一关键词在一篇文章中代表主要研究问题，在另一篇文章中可能代表研究方法，将关键词同等看待会影响学科主题演化分析的准确性。

动态主题识别方法和文献处理技术息息相关，Nasukawa 等对文献处理技术的进展进行概括总结[8]，具体见表 1。

结合表 1 分析可知，目前动态主题识别方法主要基于 NLP、文献计量学和可视化等技术。随着计算机技术的快速发展，信息可视化技术逐渐成熟。2003 年，美国国家科学院提出科学知识图谱的概念[9]，随后引起众多专家学者的重视，并产生了大量研究成果。特别是在学科主题演化分析研究领域，提出众多可视化分析方法，并以此为基础，研究开发了相应的科学知识图谱绘制工具，促进了学科主题演化分析

表1 文献处理技术进展

功能	目的	技术	数据表示	自然语言处理	输出
文献查找	找出特定主题的相关数据	信息检索	字符串,关键词	关键词抽取(转换为原型)	一组文献
文献组织	主题概述	聚类、分类	关键词集(矢量空间模型)	关键词分布分析	文献集(簇)
知识挖掘	从内容中抽取感兴趣的信息	NLP,数据挖掘,可视化	语义概念	语义分析、意图分析	提炼过的信息(趋势模式,关联分析等)

研究广泛开展,具有重要的理论与实践意义。因此,利用可视化技术构建学科知识图谱进行学科主题演化分析,已成为比较新颖且可行的研究思路。

1.2 主题演化可视化

目前,相关专家学者提出了众多主题演化可视化方法:2003 年,Morris 等基于文献耦合聚类方法识别研究前沿主题,并采用创新性的时间线图谱方法分析和展现研究前沿主题的演化情况[10]。2004 年,Garfield 基于直接引用网络分析,研究设计了可视化图谱,用来分析某一个知识领域主要研究主题的历史演化过程[11]。

2004 年,陈超美提出了一种新的分析某知识领域演化情况的可视化分析方法,并基于 Java 语言研究开发了知识图谱绘制软件 Citespace I,具有时序分割、同被引聚类、寻径网络、时序网络可视化分析等功能[12];2006 年,陈超美针对前期研究中的不足,又开发了 Citespace II,并对其基本原理进行了细致阐释,新增了 N-Gram 术语提取、突发检测、中介核心性、异构网络分析等功能,能够更加有效地展示某学科领域研究主题的演进历程,拥有良好的可视化效果[13]。

2008 年,Rosvall 等借鉴地理学领域的冲积图(Alluvial Diagram)提出一种社区演化可视化分析方法,能够直观展示学科主题结构的演化过程[14]。2013 年,王晓光等改进 Rosvall 等的方法,研究开发了学科主题演化可视化分析软件 Neviewer,能够以冲积图、赋色网络图揭示学科主题演化的宏观过程和微观细节[15]。

2011 年,Cobo 等提出了一种可用于探测、量化和可视化分析某研究领域演变过程的方法,并通过对模糊集理论领域的实证研究验证了该方法的有效性[16];2012 年,Cobo 等又研究开发了知识图谱绘制工具 SciMAT,并详细介绍其基本原理和算法,该方法能够通过密度、中心度指标分析主题词间的关联强度和主题演化能力,并且能够识别主题演化路径[17]。

2011 年,微软亚洲研究院的研究人员提出一种能够分析多个主题演化关系的文本可视化分析方法 TextFlow,在海量文本分析中引入主题合并和分裂的理念,能让人利用直观的流式图形迅速把握海量信息的发展脉络[18]。2015 年,Gad 等基于文本中的高频词提出动态时序主题演化可视化系统,可以充分展示内部基本知识单元的动态演化过程,并通过奥巴马演讲文本、报纸文本等进行实证研究,验证了其有效性[19]。

学科主题演化可视化分析方法和相应科学知识图谱绘制软件的广泛传播,促进了相关应用研究的展开,众多专家学者利用以上方法与软件工具进行某学科领域的主题演化分析。2011 年,游毅等通过对 2000—2009 年我国信息生命周期领域的期刊论文的关键词进行处理,利用多维尺度和战略坐标图谱,分析我国信息生命周期领域的研究现状并对未来发展趋势进行预测[20]。2012 年,薛调等利用 CitespaceII 软件的主题演化图谱,分析国内图书馆学科知识服务领域演进路径、研究热点与前沿[21]。2014 年,李长玲等利用共词网络图分析知识网络研究领域的主题演化情况[22]。2014 年,孙静等利

用 NEViewer 软件处理 22 种学报类医学期刊题录数据，分析了医学领域科研主题的演化情况[23]。2016 年，祝娜等利用 SciMAT 软件进行 3D 打印领域知识演化路径的构建研究[24]。

笔者对目前主要学科主题演化可视化分析图谱、绘制软件工具进行对比分析，具体见表 2。

表 2　主要学科主题演化分析可视化软件工具对比

名称	主要绘制软件	优点	缺点
共词网络图	UciNet NetDraw	可以通过节点、连线展示词间关系，发现核心和边缘主题词	学科主题演化趋势展示不足
CiteSpace 演化图	CiteSpace	美观、色彩丰富，可以展示主题的时间演化趋势	主题词间关系及其内部各主题词的权重不能很好地展示
战略坐标图	SciMAT	通过向心度和密度为参数，有效展示主题间的联系与相互关联	不能展示不同主题间的内部联系，不能很好地展示演化趋势
多维尺度图	SPSS	能够表现出主题词间的亲疏、相似关系，反映主题内容的整体结构	无法确定主题的边界与数目；不能展示不同主题间的关系，不能展示演化趋势
主题演化冲积图	NEViewer	能够展示主题结构、内容的复杂演化过程	不能充分展示内部基本知识单元的演化趋势

综上所述，学科主题演化是一个复杂的过程，存在多种变量，比如主题强度、结构和内容等，如果以单一类型可视化图谱进行分析会造成信息负荷过大，存在感知局限性、认知局限性和性能局限性等问题，所以从多个维度进行学科主题演化可视化分析有助于提高分析效果。

为了改进目前学科主题演化可视化分析方法的不足，本文设计了一种基于学科主题强度、结构和内容（内部基本知识单元）多维的学科主题演化分析模型，并利用 JavaScript 语言的 Web 前端可视化技术，分别研究设计了与之相契合的创新性可视化图谱。通过对近 10 年我国图书情报领域的大数据研究论文进行实验分析，验证该模型的有效性。

2　多维度学科主题演化可视化分析方法

本文提出一种多维度视角下的学科主题演化可视化分析方法，分三个步骤完成。

第一步，学科主题识别。利用人工标注的方法对核心关键词进行语义角色标注，人工训练数据集，然后构建具有语义角色特征的共词网络，利用 Fast Unfolding 社区发现算法识别出具有语义角色特征的学科主题。

第二步，学科主题相似度计算。利用余弦相似度算法计算相邻子时期的主题相似度，判定学科主题演化关系。

第三步，多维度学科主题演化可视化分析。基于本文提出的多维度学科主题演化分析模型，利用 JavaScript 语言的 Web 前端可视化技术，分别构建学科主题强度、结构和内容（内部基本知识单元）三个维度的主题演化可视化图谱，进行多维学科主题演化分析。

2.1　学科主题识别

学术文献中的关键词共词网络内也存在社区现象，共词网络内的社区可以表征某一学科领域的研究主题，将社区发现方法用于学科研究主题识别，不仅是可行的，更是一种非常理想的思路，而且共词网络中社区的演化在一定程度上揭示了学科研究主题的演化发展过程[25]。因此，本文利用 Fast Unfolding 算法探测、识别不同语义角色关键词共词网络中的社区，进而识别学科主题。

(1) 核心关键词确定

根据分析目标,确定数据库,构建检索式,获取相应学科领域的学术文献,然后进行时间划分,Time Line 方法和固定时间窗口是两种常见的划分方法[26-28],本文按照固定时间窗口(1年)进行时间划分。然后,根据 Donohue 提出的高、低频关键词界分公式[29]确定学科主题演化分析所需要的核心关键词。高、低频关键词界分公式为:

$$T=\frac{1}{2}(-1+\sqrt{1+8I_1}) \quad (1)$$

其中,T 是高频关键词的边界,即词频大于 T 的关键词为核心关键词;I_1 代表关键词词频为 1 的个数。

(2) 关键词语义角色分类与共词网络构建

为了提高语义角色标注的准确性,本文采用人工标注的方法,结合题名和摘要内容对核心关键词(工作量适中,准确度高于无监督机器学习方法)进行语义角色标注,人工训练数据集,实现关键词的语义角色分类。比如:k_i [character] 表示第 i 个关键词的语义角色标注为 [character],character ∈ {P,M,T};P,M,T 分别代表核心研究问题(Problem),主要研究方法(Means),关键技术与工具(Technology and Tools)。

通过语义角色标注不仅可以帮助用户更好地理解关键词的语义信息,还能帮助计算机构建具有语义特征的共词网络。根据关键词的语义角色,在每个时间窗口下,构建具有语义角色特征的共词网络;然后基于共现矩阵构建不同语义角色的共词网络 N_{tn} [character](每个子时期 t 内含有 P,M,T 三个语义角色的共词网络),基本节点为 K [character]。其中,N_{tn} 代表 t 时间窗口内共词网络集合中的第 n 个共词网络;N_{tn} [character] 代表 t 时间窗口内第 n 个语义角色为 [character] 的共词网络。

(3) 基于 Fast Unfolding 算法的主题识别

基于复杂网络分析软件 Gephi 的社区识别模块,利用 Fast Unfolding 社区发现算法[30]找出每个时间窗口内的共词网络社区,识别具有语义特征的主题,主要细分为原始划分、模块度优化、社区聚合和社区(主题)识别四个步骤。

Fast Unfolding 算法是基于模块度优化(Modularity Optimization)的启发式方法,对于复杂网络进行社区发现的效果优异,模块度(Modularity)指标可以用来衡量社区发现算法的优劣,模块度越大意味着社区发现的效果越好,模块度的计算公式为[31]:

$$Q=\sum_i(e_{ij}-a_j^2)=Tre-\|e^2\| \quad (2)$$

其中,Q 代表模块度;e_{ij} 代表复杂网络中连接 i 和 j 两个不同社区的边数占总边数的比例(总边数是原始网络的边数);$a_i=\sum_j e_{ij}$ 代表对称矩阵中每行或列各元素之和,即与第 i 个社区中的节点相连的边占总边数的比例;Tre 代表对阵矩阵的矩形单元;$\|e^2\|$ 代表对称矩阵 e 中所有元素之和。

2.2 演化关系判定

本文通过计算学科主题相似度来判定学科主题演化关系,采用点积余弦相似度算法计算学科主题相似度。首先构造向量空间模型(Vector Space Model,VSM),即通过向量的方式来表征学科主题,其中学科主题由若干关键词(Keyword)组成,先将主题表示为 $Topic=\{k_1,k_2,k_3\ldots\ldots k_n\}$,进一步可以被描述为一系列关键词的词频向量 $Topic\ Vector=\{weight_1,weight_2,weight_3\ldots\ldots weight_n\}$,由 n 个关键词组成,每个词都有一个权重 Weight(由关键词词频表示);然后计算两个向量的余弦相似度(介于 0 和 1 之间,值越大表示两个主题越相似)。主题相似度计算公式为:

$$Sim(Topic_i,Topic_j)=\cos\theta=\frac{\sum_{k=1}^{n}w_k(Topic_i)\times w_k(Topic_j)}{\sum_{k=1}^{n}w_k^2(Topic_i)\times\left(\sum_{k=1}^{n}w_k^2(Topic_j)\right)} \quad (3)$$

其中，分子表示两个向量的点乘积，分母表示两个向量的模的积，权重 w 由关键词词频表示。

2.3　多维度学科主题演化分析模型

本文认为影响学科主题演化分析的主要因素有四个：时间、主题强度、主题结构、主题内容（内部基本知识单元）。其中，时间因素是基础，在学科主题演化分析中，必须加上时间维度，才能准确表达出主题强度、结构和内容的演化过程。主题强度、结构和内容因素是主体，是学科主题演化分析的三个主要维度，综合测度这三个因素可以对学科主题演化过程进行全面、系统、准确的分析，揭示学科主题的生命周期动态演化全过程。基于上述影响因素，本文提出多维度视角下的学科主题演化分析模型，下面对其进行详细说明。

2.3.1　主题强度

主题强度（Topic Intensity），是指学科主题所拥有的关注度、研究热度，可通过主题词频次、发文量、被引量[32-33]等指标进行测度。本文通过主题内部关键词总频次表征学科主题强度。

由于本文识别出的学科社区（主题）是由一组关键词组成，因此将主题强度定义为主题内部关键词的总频次，计算公式为：

$$TI=\sum_{i=1}^{n}k_i\ [character]_f \tag{4}$$

其中，TI 代表主题强度（Topic Intensity）；$k_i\ [character]_f$ 代表关键词 $k_i\ [character]$ 的频次（frequency），$\sum_{i=1}^{n}k_i\ [character]_f$ 代表主题内部关键词的频次之和，通过计算各个时间段不同主题内部关键词的总频次，以表征各个主题的强度。

学科主题强度演化过程定义如下。

定义 1：学科主题强度演化具有“惯性”与“相关性”，即学科主题强度时间序列变化发展具有延续性并且是相互联系的，一定时期内存在可预测的发展变化规律。

定义 2：某学科领域同一学科主题在不同的时间段（t—t+1）具有不同的强度，不同时刻的主题强度值可以构成时间序列，会发生上升、持平、下降等演化过程。

2.3.2　主题结构

主题结构（Topic Structure），是指学科主题内各个部分之间的联系、层级、分布和相互影响，可通过度中心性（Degree Centrality）、向心度（Centrality）、密度（Density）等指标进行测度[34-35]。

由于本文识别出的学科主题是由关键词共词网络社区表示，因此通过主题内部关键词的度中心性来测度，参考度中心性测量公式[34]，定义本文中主题结构表征公式为：

$$TS=\sum_{i=1}^{n}C_D(K_i),\ C_D(K_i)=\sum_{i=1}^{g}k_{ij}\ [character]\ (i\neq j) \tag{5}$$

其中，$C_D(K_i)$ 代表关键词 i 的度中心性，$\sum_{i=1}^{g}k_{ij}\ (i\neq j)$ 代表计算某社区内部关键词 i 与其他 $g-1$ 个 j 关键词（$i\neq j$，排除 i 与自身的联系）之间的直接联系的数量；$\sum_{i=1}^{n}C_D(K_i)$ 代表计算主题内部关键词的标准化度中心性之和，以直观测度学科主题结构。

2007 年，Palla 等在《自然》（*Science*）杂志上发表文章探索社群（复杂网络）演化过程，将复杂网络演化过程分为新生、消亡、合并、分裂、增长和收缩六种[36]。借鉴 Palla 等的定义，鉴于本文识别出的学科主题具有明显的语义角色特征，将学科主题结构演化过程定义如下。

定义 3：研究问题主题演化：代表核心研究问题（Problem）的学科主题结构在不同的时间段（t—t+1）会发生新生、消亡、合并、分裂、增长和收缩六种演化过程。

定义 4：技术与工具主题演化：代表技术与工具（Technology and Tools）的学科主题结构在不同的时间段（t—t+1）会发生新生、消亡、合并、分裂、增长和收缩六种演化过程。

定义 5：研究方法主题演化：代表研究方法（Means）的学科主题结构在不同的时间段（t—t

+1)会发生新生、消亡、合并、分裂、增长和收缩六种演化过程。

2.3.3 主题内容

主题内容(Topic Content),是指学科主题的内部基本知识单元,可通过主题词、关键词等进行表示。本文通过关键词表征学科主题的内部基本知识单元。

由于本文识别出的学科主题内部基本知识单元是由关键词表示,因此,定义本文中主题内容表征公式为:

$$TC=\{k_1\ [character],k_2\ [character],k_3\ [character]\}\cdots\cdots k_n\ [character] \quad (6)$$

其中,k_n [character]表示主题中第n个关键词,character∈{P,M,T}。

学科主题内容演化,是学科主题整体结构演化分析的进一步深化,即学科核心研究问题、主要研究方法和关键技术主题内部基本知识单元(关键词)在不同时刻所从属的主题的动态变化情况。学科主题内容演化过程定义如下。

定义6:同一语义角色关键词构成的主题演化过程:同一语义角色关键词构成的学科主题内部基本知识单元在不同的时间段(t—t+1),会发生新生、消亡、转移、稳定等动态变化情况,即内部基本知识单元从属主题的动态变化。

定义7:不同语义角色关键词构成的主题演化过程:不同语义角色关键词构成的学科主题内部基本知识单元在不同的时间段(t—t+1),会发生核心研究问题刺激产生了创新性的研究方法或技术;研究方法被应用于解决某核心研究问题等演化过程。

3 多维度学科主题演化可视化设计

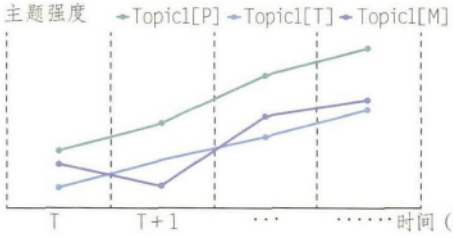
目前学科主题演化可视化研究,主要试图通过某一种可视化方案展示学科主题演化的复杂过程,存在感知局限性、信息负荷过大、展示不够深入等问题。

本文构建了多维度学科主题演化分析模型,以"描述"学科主题生命周期动态演化的复杂过程,并利用JavaScript语言的Web前端可视化技术,从学科主题强度、结构和内容三个维度对学科主题演化过程进行可视化设计,尝试剖析学科主题演化的复杂过程,以期提高学科主题演化可视化分析的深度和准确度。

3.1 主题强度演化可视化设计

利用主题强度计算公式(4)计算具有演化关系的学科主题强度*TI*,构建学科主题强度时间序列。时间序列是将某种统计指标的数值按时间先后顺序排列所形成的数列,通过编制和分析时间序列,根据其所反映出来的发展过程、方向和趋势进行类推或者延伸,可以预测下一段时间或以后若干年内可能达到的水平[37]。

为了展示不同语义角色的学科主题强度演化的过程,本文研究设计了主题强度演化折线图,可以用来展示学科主题演化复杂过程中的主题强度演化过程及其趋势,主题强度演化折线图基本元素设计如图1所示。



注:横坐标代表时间;纵坐标代表主题强度;折线颜色代表主题的语义角色,绿色代表研究问题主题(Problem),蓝色代表技术与工具主题(Technology and Tools),紫色代表研究方法主题(Means)。

图1 学科主题强度时间序列演化图设计

与目前研究中提出的学科主题强度演化图相比,本文中主题强度演化折线图主要包含三种颜色的折线,分别表示研究问题、技术与工具和研究方法三种不同语义角色的学科主题,同一颜色折线具体节点形状不同,以区分同一语义角色的不同主题,能够更加直观、细致、分层次地展示某学科领域研究主题的强度变化趋势。

3.2　主题结构演化可视化设计

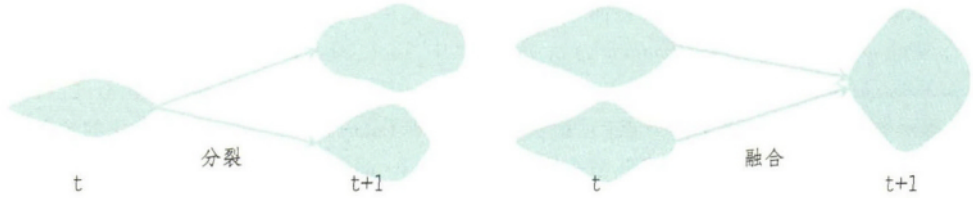
利用学科主题结构测度公式(5)，分别计算各个学科主题结构数值 *TS*，然后进行数据转换，以便为 JavaScript 语言“理解”，为学科主题结构演化可视化奠定数据基础。数据转换是学科主题结构演化可视化非常重要的一步，即将模型数据 *TS* 转换为可视化数据，学科主题结构数据转换过程中的度中心性数据转换如图 2 所示。

```
学科主题结构演化图代码节选：度中心性数据转换
series : [
    {
    "name" : "[character] 语义角色 ",
    "data" : [
                {
    "name" : "Topic_i [character] ",
    "weight" :  TS = Σ_{i=1}^{n} C_D(K_i) , //Topic i[character] 的 TS 值
                }
            ]
```

图 2　学科主题结构演化图谱设计代码片段

完成数据转换后，首先通过函数 *legend*：{ *data*：［‘研究问题’，‘研究方法’，‘研究技术与工具’］}定义学科主题结构演化图的语义角色。然后通过函数 *xAxis*：{ *type*：‘*time*’，*boundaryGap*：［*t*，*t* + 1］}定义图谱横坐标轴的时间维度。

以 t—t + 1 时间窗口代表核心研究问题的学科主题结构演化的六种基本演化过程为例，进行学科主题结构演化可视化设计，具体代码过长不再列举，结果如图 3 所示。



注：图 3 中气泡代表主题；气泡大小代表主题整体结构（由内部关键词的标准化度中心性测度）；连线代表主题演化方向；颜色代表主题的语义角色，绿色代表研究问题主题（Problem）。

图 3　学科主题结构演化过程设计：以核心研究问题主题分裂、融合过程为例

如果给定 t 时间段的学科主题 $Topic_i$ 和 t + 1 时间段的学科主题 $Topic_j$ 相似度 $sim(Topic_i, Topic_j) > 阈值_{TS}$，认为 $Topic_i$ 和 $Topic_j$ 具有学科主题结构演化关系，令 $Color(Topic_j) = Color(Topic_i)$，$ArrowLine(Topic_i, Topic_j) = 阈值_s$。其中，$Topic_i$ 和 $Topic_j$ 表示相邻子时期的学科主题；$sim(Topic_i, Topic_j)$ 代表两者相似度值；$Color(Topic_j)$ 和 $Color(Topic_i)$ 代表学科主题气泡的颜色；$ArrowLine(Topic_i, Topic_j)$ 代表两个主题连线的粗细，由相似度大小确定。

与目前研究中的学科主题结构演化可视化图相比，本文提出的学科主题结构演化可视化方案，除了能够直观展示学科主题结构的新生、消亡、合并、分裂、增长和收缩六种演化过程，还可以单独展示某学科领域研究问题、研究方法等主题的结构演化情况，能够满足细粒度、针对性的学科主题结构演化可视化分析需求。

3.3 主题内容演化可视化设计

利用学科主题结构测度公式(6)，分别计算各个学科主题内容 *TC*，然后进行数据规范化处理，以便为 JavaScript 语言“理解”，为学科主题内容演化可视化奠定数据基础。数据转换同样是学科主题内容演化可视化非常重要的一步，即模型数据 *TC* 转换为可视化数据，数据转换过程如图4所示。

```
主题内容演化图谱设计代码片段：节点与连接数据转换
var data = {
'nodes': [
  {name: "K1[character]"},
  {name: " K2[character]"},
  {name:" K3[character]"},
  {name: "Kn[character]"},
],// TC = {k1[character],k2[character],k3[character].....kn[character]},通过数据
规范化处理，将 TC、频次数据分别转换为节点数据和连接数据。
  'links': [
  {source: K1[character], target: K1[character], value:频次},
  {source: K2[character], target: K2[character], value: 频次},
  {source: K3[character], target: K3[character], value: 频次},
  {source: Kn[character], target: Kn[character], value: 频次},
  ]
};
```

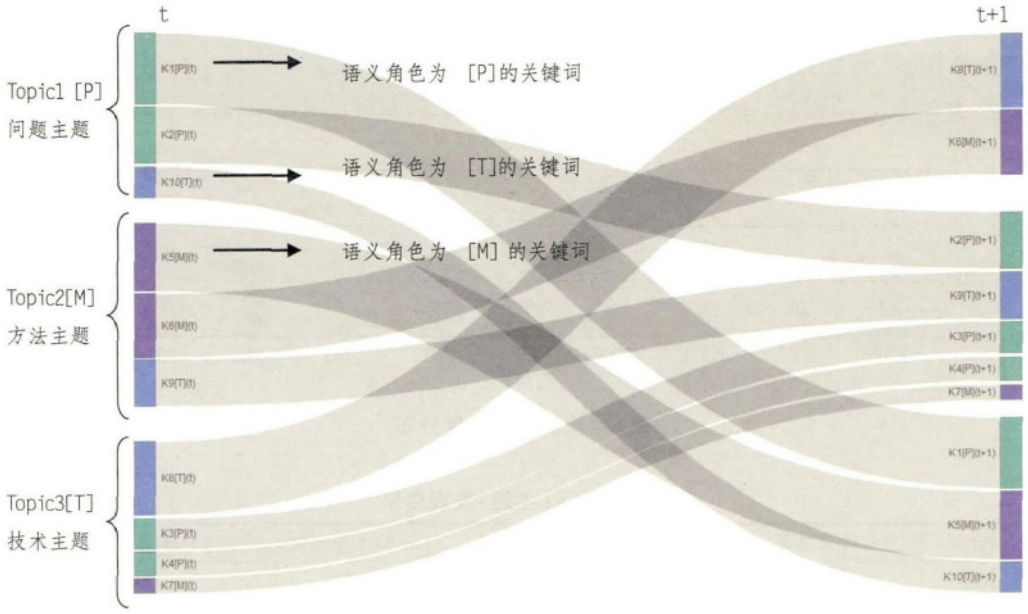
图4 主题内容演化图谱设计代码片段：节点与连接数据转换

完成数据转换后，定义学科内容演化图谱布局，通过 *nodeWidth*(*i*)、*nodePadding*(*j*)、*size*([*width*,*height*])、*nodes*(*data. nodes*)、*links*(*data. links*)等函数分别定义图谱中的节点(元素块)宽度、节点(元素块)高度、图谱宽高、节点数组(TC 数据)、连接数组。

以 t—t + 1 时间窗口代表核心研究问题、主要研究方法和关键技术与工具的学科主题 Topic1 [P]、Topic2 [M] 和 Topic3 [T] 的演化情况为例，进行学科主题内容演化可视化设计说明，如图5所示。

如果给定 t 时间段 $Topic_i$ 和 t + 1 时间段 $Topic_j$ 相似度 $sim(Topic_i, Topic_j) > 阈值_{TC}$，认为 $Topic_i$ 和 $Topic_j$ 具有学科主题内容演化关系，令 $Size(k_i) = frequency(k_i)$，$Color(k_j) = Color(k_i)$。其中，$Topic_i$ 和 $Topic_j$ 表示相邻子时期的学科主题；$sim(Topic_i, Topic_j)$ 代表两者相似度值；$Color(k_j)$ 和 $Color(k_i)$ 代表关键词元素块的颜色；$Size(k_i) = frequency(k_i)$ 关键词元素块的大小，由其频次确定。

与目前研究中的学科主题内容演化可视化图相比，本文提出的学科主题内容演化可视化方案，除了能够直观展示学科主题内部基本知识单元的动态变化，还可以单独展示学科核心研究问题、主要研究方法和关键技术主题内部基本知识单元在不同时刻所从属主题的动态变化情况，能够满足深度学科主题演化可视化分析的需求。



注：元素块代表关键词；灰色连线代表关键词所从属主题的变化情况；由左至右代表时间 t 的增长，为演化图谱增加时间维度；元素块大小代表关键词的权重（频次），即在主题中的重要程度；元素块的聚集代表学科主题，即学科主题由权重不同、语义角色相同的关键词构成；元素块颜色代表关键词的语义角色，绿色代表研究问题［P］，蓝色代表技术与工具［T］，紫色代表研究方法［M］。

图 5　学科主题内容演化基本图谱元素设计

4　实证研究

4.1　数据源

为验证本文提出的学科主题演化可视化分析方法的可行性和有效性，选择近 10 年我国图书情报领域大数据研究的相关论文进行实证研究，分析我国图书情报领域大数据研究的核心研究问题、主要研究方法和关键技术主题的演化情况。选择《中国学术期刊全文数据库》为检索数据库并限定检索范围为期刊论文；确定检索式，以“大数据”为检索词进行全文检索，时间跨度确定为 2006 年 1 月 1 日至 2015 年 12 月 31 日，共计 10 年。选择图书情报与数字图书馆学科分组，共检索到 4 909 篇期刊论文。按照年度将文献划分到不同的时间窗口，文献数量年度分布如图 6 所示。2006—2011 年发文量变化趋势平稳，年均发文量 79.5 篇；2012 年以后发文量快速增长，反映出从 2012 年开始，我国图书情报领域大数据研究热度快速提高。

4.2　学科主题识别与演化关系判定

利用本文提出的方法，对论文数据进行处理（所有论文由检索词“大数据”检索得到，因此不单独分析关键词“大数据”），具体步骤包括：关键词抽取与核心关键词选定，时间切片，核心关键词语义角色标注，基于语义角色的共词网络构建，基于 Fast Unfolding 社区发现算法的语义社区（主题）识别。社区（主题）发现结果见表 3。

2006 至 2015 年共识别出 107 个学科主题，然后基于 Python 的 Gensim 工具包进行相似度计算，具体采用余弦相似度计算公式（3）进行计算，以判定演化关系，并将主题相似度计算结果

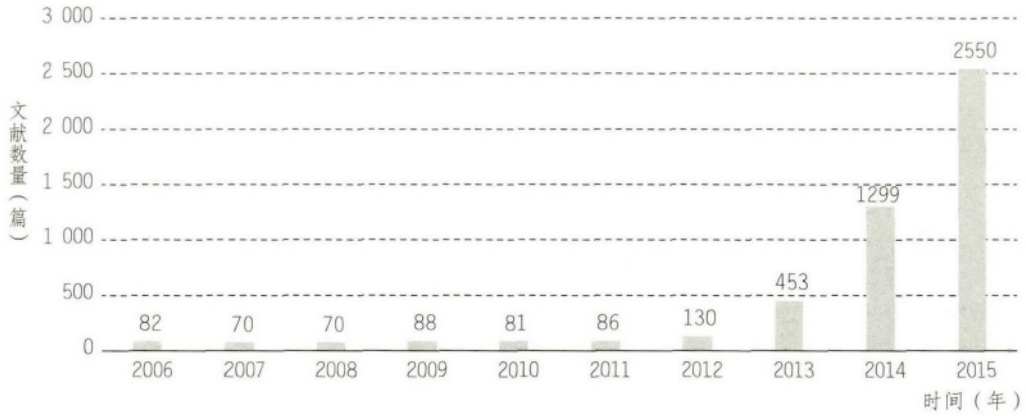


图6 论文数量年度分布

表3 社区(主题)发现结果

时间	关键词总数	核心关键词数	主题数	平均主题词数	平均模块度 Q(modularity)
2006	234	78	6	13	0.597
2007	199	72	7	10	0.683
2008	211	76	6	13	0.677
2009	275	86	7	12	0.681
2010	244	82	5	16	0.669
2011	267	85	7	12	0.672
2012	481	113	10	11	0.689
2013	1 410	189	14	13	0.558
2014	2 912	269	18	15	0.641
2015	4 774	337	27	12	0.653

写入 Excel 文件，以待分析使用。

相似度大于阈值判定具有演化关系，其中学科主题强度演化关系的主题相似度阈值设定为0.71，学科主题整体结构演化关系的主题相似度阈值设定为0.69，学科内容演化关系的主题相似度阈值设定为0.73。

4.3 多维度学科主题演化可视化结果与分析

基于 JavaScript 语言的 Web 前端可视化技术，进行可视化展示，具体计算过程以及算法代码不再一一列举。可视化结果如图7、图8和图9所示。展示了我国图书情报领域近10年大数据研究的发展演化现状，分别从学科主题强度、学科主题整体结构和学科主题内容三个层面进行描述，每个层面下又详细显示了核心研究问题、主要研究方法和关键技术主题的演化路径、脉络。总体来说，近10年我国图书情报领域的大数据研究具有以下特征。①从学科主题强度维度来说，大数据研究主题强度普遍快速提高；②从学科主题整体结构维度来说，新生主题不断涌现，众多研究主题处于不断成长、分裂、交叉融合的过程中；③从学科主题内容维度来说，新生关键词不断涌现，并且关键词所从属的主题不断变化。

以云计算主题为例，进行深入分析，云计算主题识别结果见表4，由结果可知，云计算主题最早是在2009年出现的，即我国图书情报领域相关学者2009年才开始针对云计算展开研究。

表 4　云计算主题识别结果

社区(主题)	关键词	社区(主题)	关键词	社区(主题)	关键词
Topic2009 – 3 [t]	云计算 [t]	Topic2012 – 1 [t]	云计算 [t]	Topic2014 – 1 [t]	云计算 [t]
Topic2009 – 3 [t]	云存储 [t]	Topic2012 – 1 [t]	云存储 [t]	Topic2014 – 1 [t]	非结构化 [t]
Topic2009 – 3 [t]	服务集成 [t]	Topic2012 – 1 [t]	弹性云计算 [t]	Topic2014 – 1 [t]	手机定位 [t]
Topic2009 – 3 [t]	SaaS [t]	Topic2012 – 1 [t]	虚拟化 [t]	Topic2014 – 1 [t]	物联网 [t]
Topic2009 – 3 [t]	Web2.0 [t]	Topic2012 – 1 [t]	动态迁移 [t]	Topic2014 – 1 [t]	网络技术 [t]
Topic2009 – 3 [t]	SOA [t]	Topic2012 – 1 [t]	云服务 [t]	Topic2014 – 1 [t]	数据挖掘 [t]
Topic2010 – 2 [t]	云计算 [t]	Topic2012 – 1 [t]	MapReduce [t]	Topic2014 – 2 [t]	云计算 [t]
Topic2010 – 2 [t]	计算机技术 [t]	Topic2012 – 1 [t]	Hadoop [t]	Topic2014 – 2 [t]	混合云 [t]
Topic2010 – 2 [t]	网络技术 [t]	Topic2013 – 1 [t]	云计算 [t]	Topic2014 – 2 [t]	分析即服务 [t]
Topic2010 – 2 [t]	共建共享 [t]	Topic2013 – 1 [t]	MapReduce [t]	Topic2014 – 2 [t]	国防科技 [t]
Topic2010 – 2 [t]	云存储 [t]	Topic2013 – 1 [t]	Hadoop [t]	Topic2014 – 2 [t]	虚拟机 [t]
Topic2010 – 2 [t]	虚拟化 [t]	Topic2013 – 1 [t]	物联网 [t]	Topic2014 – 2 [t]	数据分析 [t]
Topic2010 – 2 [t]	分布式 [t]	Topic2013 – 1 [t]	移动阅读 [t]	Topic2014 – 2 [t]	物流信息 [t]
Topic2011 – 1 [t]	云计算 [t]	Topic2013 – 1 [t]	信息技术 [t]	Topic2015 – 1 [t]	云计算 [t]
Topic2011 – 1 [t]	共建共享 [t]	Topic2013 – 1 [t]	数据存储 [t]	Topic2015 – 1 [t]	数据分析 [t]
Topic2011 – 1 [t]	云存储 [t]	Topic2013 – 3 [t]	云计算 [t]	Topic2015 – 1 [t]	云存储 [t]
Topic2011 – 1 [t]	动态供应 [t]	Topic2013 – 3 [t]	竞争情报系统 [t]	Topic2015 – 1 [t]	云服务 [t]
Topic2011 – 1 [t]	信息技术 [t]	Topic2013 – 3 [t]	Hadoop [t]	Topic2015 – 1 [t]	MapReduce [t]
Topic2011 – 1 [t]	移动终端 [t]	Topic2013 – 3 [t]	数据库建设 [t]	Topic2015 – 1 [t]	Hadoop [t]
Topic2011 – 1 [t]	信息系统 [t]	Topic2013 – 3 [t]	整合模型 [t]	Topic2015 – 1 [t]	体系架构 [t]

(1) 主题强度演化分析

在图 7 中，选取主题热度前 10 位的主题，其中云计算 [t] 主题(由 Topic2009 – 3 [t] →Topic2010 – 2 [t] →Topic2011 – 1 [t] →Topic2012 – 1 [t] →Topic2013 – 1 [t] →Topic2014 – 1 [t] →Topic2015 – 1 [t] 构成)。可以比较直观地看出，云计算主题 2009 年才开始出现，主题强度呈快速上升趋势，而且在代表技术与工具的学科主题中上升趋势最快，说明云计算主题是我国图书情报领域大数据研究中涉及的关键技术，比如：数据处理及服务技术、数据集成和分布式存储等技术。主要研究如何充分利用云计算、云存储、分布式、虚拟化技术等促进我国图书馆建设，云计算主题下竞争情报系统、虚拟机和整合模型等技术在企业危机预警、国防科技建设等方面的应用实践也是研究重点。

(2) 主题结构演化分析

在图 8 中，Topic2009 – 3 [t] →Topic2010 – 2 [t] →Topic2011 – 1 [t] →Topic2012 – 1 [t] →Topic2013 – 1 [t] →Topic2014 – 1 [t] →Topic2015 – 1 [t] 和 Topic2009 – 3 [t] →Topic2010 – 2 [t] →Topic2011 – 1 [t] →Topic2012 – 1 [t] →Topic2013 – 3 [t] →Topic2014 – 2 [t] →Topic2015 – 1 [t] (具体内部关键词见表 4) 代表了云计算主题结构的两条演化方向、路径，可以比较直观地看出云计算主题结构正处于分裂、成

长阶段，不断有新的研究内容融入云计算主题中，说明我国图书情报领域大数据研究中的云计算主题发展势头良好。而且图中云计算主题的连线有逐渐变粗的趋势，代表云计算主题的相似度有变大的趋势，说明其主题结构在分裂、成长的同时，逐渐形成特定的主题结构，比如云计算[T]、云存储[T]、云服务[T]等关键词逐渐成为支撑云计算主题结构的核心节点。

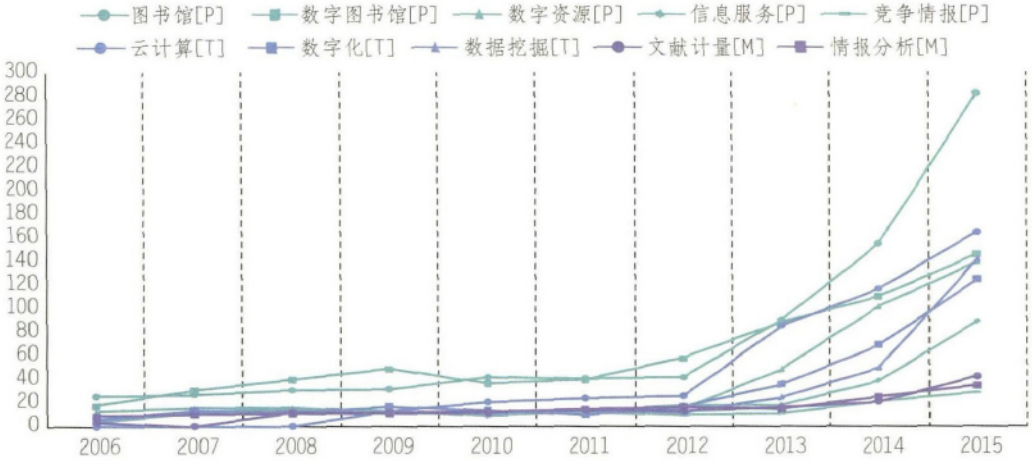


图7 大数据研究主题强度演化图

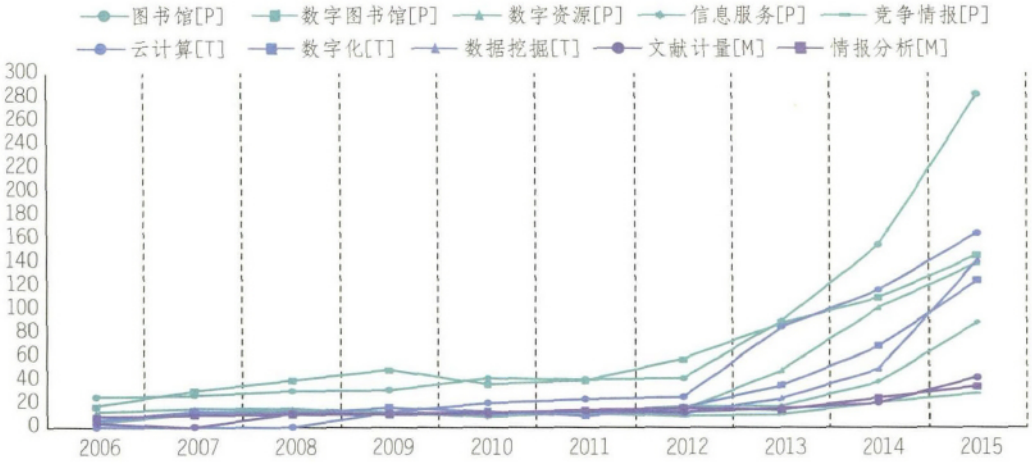
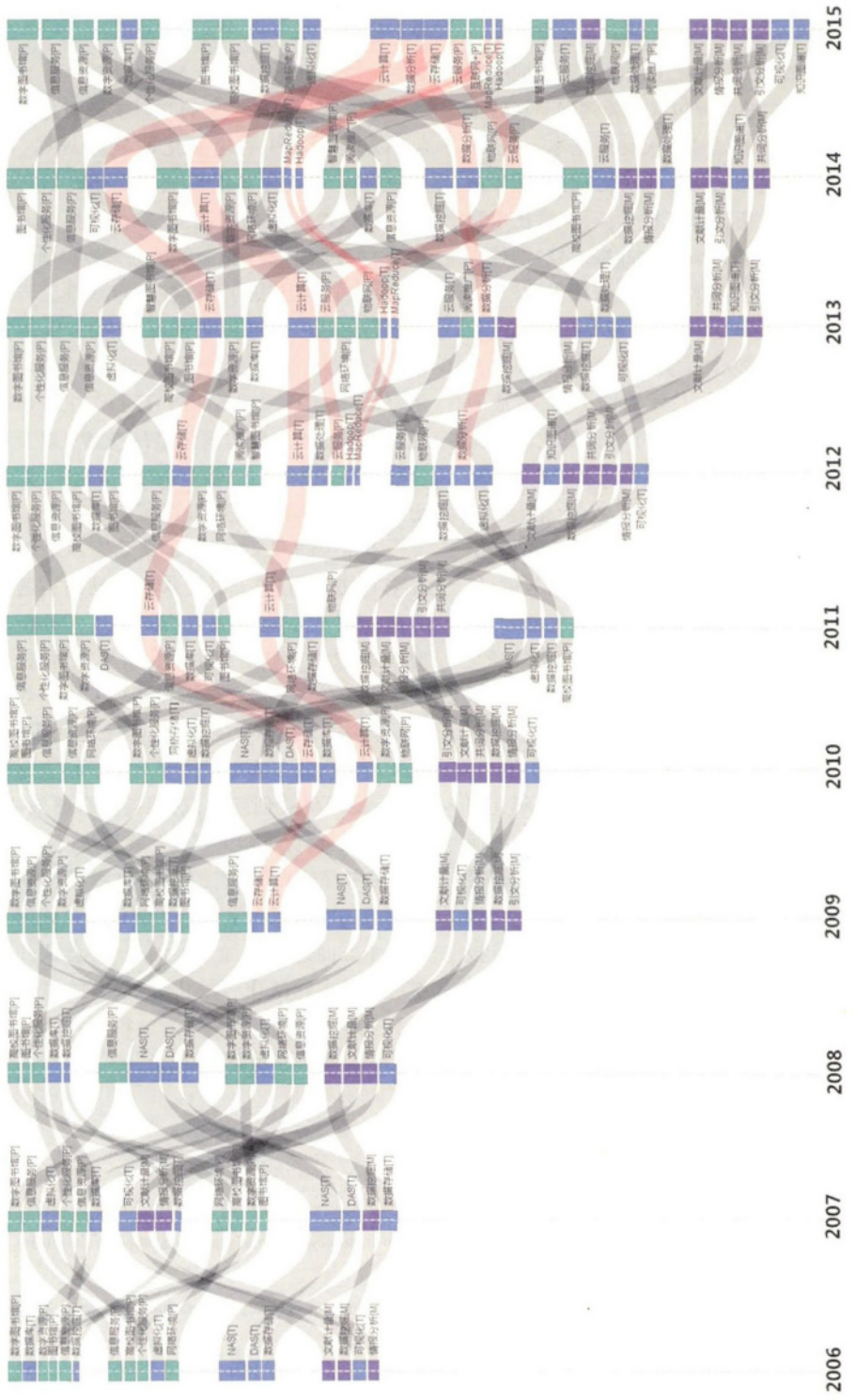


图8 大数据研究主题整体结构演化图

(3) 主题内容演化分析

为了更好地展示云计算主题内部基本知识单元的流动、演化情况，利用不同语义角色关键词构成的学科主题内容演化图谱进行云计算主题内容演化分析，以有效、微观地分析云计算主题内部涉及的主要技术及其需要处理、解决的问题。

图9中构建了2006—2015年大数据研究主题内容演化图。其中，对2015年云计算主题的演化路径、脉络以粉红色进行特殊标注，可以较为简洁、明显地展示出云计算主题从产生至今内


2015
2014
2013
2012
2011
2010
2009
2008
2007
2006

部关键词的“流动”情况，即内部基本知识单元——关键词在不同时间段所从属主题的变化情况。从整体来看，云计算主题内部关键词变化明显，不同时刻主题内部包含的关键词一直处于动态变化过程中，但是核心关键词基本不变，比如云计算、云存储和云服务等关键词；从具体内容来看，云计算是指云计算技术，云计算技术是目前信息技术（Information Technology，IT）的核心技术，主要包括海量数据的处理与分析，以及云计算为基础的云存储、分布式处理等技术手段。

我国图书情报领域大数据研究主要涉及大数据时代背景下云计算技术对图书馆、企业服务模式和流程的影响。2009 年云计算主题首次出现，主要研究如何结合云存储、云计算、SaaS、SOA 等技术，促进图书馆的数字化建设。2010—2012 年是对 2009 云计算主题的继承、丰富，研究大数据时代图书馆面临的机遇与挑战，大数据处理、挖掘、分析会发生的变化，如何利用云计算技术从中挖掘、分析潜在的、有效的、有价值的信息、知识，提供知识服务、个性化服务，并预测了大数据时代与云计算技术息息相关的图书馆的变革与发展。2013—2015 年，云计算主题发生了主题的分裂、融合，除了研究图书馆领域相关内容，逐渐涉及云计算技术如何转变企业竞争情报、经济管理、物流信息管理等方面的管理模式、服务理念、服务形态等。

可以预测在接下来几年中，我国图书情报领域会进一步在图书馆建设、企业危机预警和经济管理等领域展开基于云计算技术的实践研究，进一步深化大数据研究内涵。

5 结语

本文提出了多维度的学科主题演化分析模型，以分析学科主题演化的复杂规律，利用 JavaScript 语言的 Web 前端可视化技术研究设计了创新性可视化图谱，以实现该模型。通过对近 10 年我国图书情报领域的大数据研究的实验，验证了该方法的可行性和有效性。与目前的学科主题演化可视化分析方法相比，本文提出的方法，更加具有针对性，能够可视化地展示、分析某学科领域核心研究问题、主要研究方法和关键技术主题的演化路径、趋势，以及宏观演化趋势、中观演化过程和微观演化细节，可以更好地分析学科主题演化的复杂过程。

本研究的意义在于：①从动态主题识别角度来看，提出了基于语义角色的动态主题识别方法，为主题识别相关研究提供了一种新的研究视角；②从主题演化可视化角度来看，提出了多维度学科主题演化分析模型以及创新性的可视化方法，丰富了主题演化可视化分析方法。本文所提出的方法可以用于分析学科主题演化规律，发现科技创新知识，能提高科研效率，辅助科研决策，促进相关领域开展科技创新。

在接下来的研究工作中，将继续在学科主题发现的准确度和学科主题演化可视化的直观性两个方面进行更加深入的研究，尝试开发多维度学科主题演化可视化分析软件，以期提高学科主题演化分析水平，为学科主题演化分析相关研究提供新思路、新视角。

参考文献

[1] 王效岳，白如江. 海量网络学术文献自动分类技术研究[M]. 北京：人民出版社，2015：40－42.（Wang Xiaoyue，Bai Rujiang. The research of automatic classification technology of mass online academic literatures[M]. Beijing：People's Publishing House，2015：40－42.）

[2] Blei D M，Ng A Y，Jordan M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research，2003（3）：993－1022.

[3] Blei D M，Lafferty J. Dynamic topic models[C]//Proceedings of the 23rd International Conference on Machine Learning. NewYork：ACM，2006：113－120.

[4] 祝娜，王效岳，杨京，等. 基于 LDA 的科技创新主题语义识别研究[J]. 图书情报工作，2015(14)：126－134.(Zhu Na，Wang Xiaoyue，Yang Jing，et al. Semantic recognition of technological innovation theme based on LDA[J]. Library and Information Service，2015(14)：126－134.)

[5] 杨博，刘大有，金弟，等. 复杂网络聚类方法[J]. 软件学报，2009，20(1)：54－66.(Yang Bo，Liu Dayou，Jin Di，et al. Complex network clustering algorithms[J]. Journal of Software，2009，20(1)：54－66.)

[6] Wallace M L，Gingras Y，Duhon R. A new approach for detecting scientific specialties from raw co-citation networks[J]. Journal of the American Society for Information Science and Technology，2009，60(2)：240－246.

[7] 程齐凯，王晓光. 一种基于共词网络社区的科研主题演化分析框架[J]. 图书情报工作，2013，57(8)：91－96.(Cheng Qikai，Wang Xiaoguang. A new research frame for analyzing the evolution of research topics based on co－word network communities[J]. Library and Information Service，2013，57(8)：91－96.)

[8] Nasukawa T，Nagano T. Text analysis and knowledge mining system[J]. IBM Systems Journal，2001(20)：967－984.

[9] 陈悦，刘泽渊. 科学知识图谱的发展历程[J]. 科学学研究，2008，26(3)：449－460.(Chen Yue，Liu Zeyuan. History and theory of mapping knowledge domains[J]. Studies in Science of Science，2008，26(3)：449－460.)

[10] Morris S A，Yen G，Wu Z，et al. Timeline visualization of research fronts[J]. Journal of American Society for Information Science and Technology，2003，54(5)：413－422.

[11] Garfield E. Historiographic mapping of knowledge domains literature[J]. Journal of Information Science，2004，30(2)：119－145.

[12] Chen C M. Searching for intellectual turning points：progressive knowledge domain visualization[J]. Proceedings of the National Academy of Sciences of the United States of America(PNAS)，2004(1)：5303－5310.

[13] Chen C M. CiteSpace II：detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology，2006，57(3)：359－377.

[14] Rosvall M，Bergstrom C T. Mapping change in large networks[J]. PLoS One，2010，5(1)：e8694.

[15] 王晓光，程齐凯. 基于 NEViewer 的学科主题演化可视化分析[J]. 情报学报，2013，32(9)：900－911.(Wang Xiaoguang，Cheng Qikai. Analysis on evolution of research topics in a discipline based on NEviewer[J]. Journal of the China Society for Scientific and Technical Information，2013，32(9)：900－911.)

[16] Cobo M J，López－Herrera A G，Herrera－Viedma E，et al. An approach for detecting，quantifying，and visualizing the evolution of a research field：apractical application to the fuzzy sets theory field[J]. Journal of Informetrics，2011，5(1)：146－166.

[17] Cobo M J，López－Herrera A G，Herrera－Viedma E，et al. SciMAT：a new science mapping analysis software tool[J]. Journal of the American Society for Information Science and Technology，2012，63(8)：1609－1630.

[18] Cui W W，Liu S X，Li T，et al. Text flow：towards better understanding of evolving topics in text[J]. Transactions on Visualization and Computer Graphics，2011，17(12).

[19] Gad S，Javed W，Ghani S，et al. Theme delta：dynamic segmentations over temporal topic models[J]. Transactions on Visualization and Computer Graphics，2015，21(5)：672－685.

[20] 游毅，索传军. 国内信息生命周期研究主题与趋势分析——基于关键词共词分析与知识图谱[J]. 情报理论与实践，2011(10)：17－21.(You Yi，Suo Chuanjun. Research on topic and trend analysis of domestic information life cycle—based on co－word analysis and knowledge mapping[J]. Information Studies：Theory & Application，2011(10)：17－21.)

[21] 薛调. 国内图书馆学科知识服务领域演进路径、研究热点与前沿的可视化分析[J]. 图书情报工作，2012，26(15)：9－14.(Xue Diao. Visualization analysis of evolution path，research hotspots and frontiers of subject knowledge service in domestic libraries[J]. Library and Information Service，2012，26(15)：9－14.)

[22] 李长玲，刘非凡，魏绪秋. 基于3－mode 网络的领域主题演化规律分析——以知识网络研究领域为例[J]. 情报理论与实践，2014，34(12)：104－110.(Li Changling，Liu Feifan，Wei Xuqiu. Analysis of evolution law of

domain subject based on 3 - mode network: a case study of knowledge network [J]. Information Studies: Theory & Application, 2014, 34(12): 104 - 110.)

[23] 孙静,齐成凯,张雯. 基于 NEViewer 的医学科研主题演化可视化分析 [J]. 中华医学图书情报杂志,2014(10): 56 - 60. (Sun Jing, Qi Chengkai, Zhang Wen. NEViewer-based visual analysis of medical scientific research topics evolution [J]. Chinese Journal of Medical Library and Information Science, 2014(10): 56 - 60.)

[24] 祝娜,王芳. 基于主题关联的知识演化路径识别研究——以3D 打印领域为例 [J]. 图书情报工作,2016,60(5): 101 - 109. (Zhu Na, Wang Fang. Identification of knowledge evolutionary path based on topic relevance: taking the case of 3D printing field [J]. Library and Information Service, 2016, 60(5): 101 - 109.)

[25] 白如江,冷伏海. k - clique 社区知识创新演化方法研究 [J]. 图书情报工作,2013,57(17): 94 - 99. (Bai Rujiang, Leng Fuhai. Knowledge innovational evolution analysis based on k - clique community network [J]. Library and Information Service, 2013, 57(17): 94 - 99.)

[26] 吴斌,王柏,杨胜畸. 基于事件的社会网络演化分析框架 [J]. 软件学报,2011,22(7): 1488 - 1502. (Wu Bin, Wang Bai, Yang Shengqi. Framework for tracking the event - based evolution in social networks [J]. Journal of Software, 2011, 22(7): 1488 - 1502.)

[27] 钱铁云,李青,许承瑜. 面向科技主题发展分段的社区核心圈技术 [J]. 计算机科学与探索,2010,4(2): 170 - 179. (Qian Tieyun, Li Qing, Xu Chengyu. A core group method for segmenting the life cycle of scientific topics [J]. Journal of Frontiers of Computer Science and Technology, 2010, 4(2): 170 - 179.)

[28] Sun J M, Padimitriou S, Faioutsos C, et al. Graph scope: parameter - free mining of large time-evoiving graphs [C] // Proceedings of Knowledge Discovery in Databases: KDD. New York: ACM, 2007: 687 - 696.

[29] Donohue J C. Understanding scientific literature [J]. Education, 1974, 14(1): 75 - 76.

[30] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of communities in large networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 30(2): 155 - 168.

[31] Newman M E J, Girvan M. Finding and evaluating community structure in networks [J]. Physical Review, 2004, 69(2): 108 - 113.

[32] 黄鲁成,唐月强,吴菲菲,等. 基于文献多属性测度的新兴主题识别方法研究 [J]. 科学学与科学技术管理,2015(2): 34 - 43. (Huang Lucheng, Tang Yueqiang, Wu Feifei, et al. Research on identification of emerging topics based on muti - attribute measurement of literature [J]. Science of Science and Management of S. &T. 2015(2): 34 - 43.)

[33] 范云满,马建霞. 基于 LDA 与新兴主题特征分析的新兴主题探测研究 [J]. 情报学报,2014,33(7): 698 - 711. (Fan Yunman, Ma Jianxia. Detection of emerging topics based on LDA and feature analysis of emerging topics [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(7): 698 - 711.)

[34] 戴维·诺克,杨松. 社会网络分析 [M]. 上海: 上海人民出版社,2012: 103 - 104. (Knock D, Yang Song. Social network analysis [M]. Shanghai: Shanghai People's Publishing House, 2012: 103 - 104.)

[35] Callon M, Courtial J P, Laville F. Co - word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry [J]. Scientometrics, 1991, 22(1): 155 - 205.

[36] Palla G, Barabási A L, Vicsek T. Quantifying social group evolution [J]. Nature, 2007, 446(7136): 664 - 667.

[37] 张美英,何杰. 时间序列预测模型研究综述 [J]. 数学的实践与认识,2011,41(18): 189 - 195. (Zhang Meiying, He Jie. Summary on time series forecasting model [J]. Mathematics in Practice and Theory, 2011, 41(18): 189 - 195.)

刘自强　山东理工大学科技信息研究所硕士研究生。山东 淄博 255049。

王效岳　山东理工大学科技信息研究所教授。山东 淄博 255049。

白如江　山东理工大学科技信息研究所副教授。山东 淄博 255049。

(收稿日期: 2016 - 06 - 03; 修回日期: 2016 - 07 - 29)