

基于 T-LDA2vec 的高校图书馆信息素质教育 主题挖掘与演化分析

王智迪

(吉林动画学院图书馆, 长春 130012)

摘要: 为了深入研究高校图书馆信息素质教育领域的发展趋势和演化过程, 收集了 1998—2023 年的 1 606 篇相关文献, 经过数据清洗和预处理后, 构建 T-LDA2vec 混合模型, 用于主题建模和文本分析。在时间趋势上, 揭示高校图书馆信息素质教育领域存在学术繁荣期和学术调整期两个关键时期。在学术繁荣期相关文献数量迅速增长, 而学术调整期文献数量急剧下降, 反映该领域正在经历学术调整。继而, 运用 T-LDA2vec 模型进行主题挖掘, 确定每个时期的最佳主题数量, 并将主题划分为高校教育评估、师资培养、情报管理、网络化图书馆服务、高校图书馆素质教育服务、心理素质与教育表现、地方信息化与课程发展、数字化图书馆员培养八大类别。结果表明, 计算不同时间段内各主题强度, 并通过交互式条形图描述热点主题。研究发现, 一些主题在不同时期内保持较高的强度, 表明它们在相关文献中具有重要影响力。师资培养、教学改革及高校图书馆资源创新与服务质量等主题在不同时期内维持了较高的强度。通过主题演化分析, 揭示了不同时期内主题之间的关联和演化过程, 指出高校图书馆信息素质教育研究的关注焦点逐渐从基础服务向资源创新、知识管理和在线教育等领域演化。该研究有助于更全面地理解该领域的研究动态, 为未来的研究方向和政策制定提供有益的参考, 同时也为文本分析方法的应用提供了有益的方法论示范。

关键词: 文本挖掘; LDA; word2vec; 高校图书馆; 信息素质教育

中图分类号: G250 **文献标志码:** A **文章编号:** 1671-1807(2024)02-0102-09

高校图书馆一直扮演着知识传授和信息资源管理的重要角色。随着信息时代的到来, 高校图书馆的使命不仅仅局限于提供图书和文献资源, 还包括培养学生的信息素质, 帮助他们适应不断变化的信息环境。教育部新版《普通高等学校图书馆规程》中明确规定: “图书馆应重视开展信息素质教育, 采用现代教育技术, 加强信息素质课程体系建设”^[1]。2018 年教育部办公厅印发《贯彻落实〈高校思想政治工作质量提升工程实施纲要〉部内分工方案》的第八条“不断深化服务育人”中提出: “在图书资料服务中, 开展信息素质教育, 引导师生尊重和保护知识产权, 维护信息安全”^[2]。2023 年 9 月全国高校信息素质教育研讨会召开, 会议以“数智时代背景下的高校图书馆信息素质教育”为主题, 就“信息素养教学能力提升新路径”“人工智能时代的数据素养教育”“泛信息素养教育改革与实践”等方面^[3]进行探讨。因此, 高校图书馆信息素质教育领

域的研究变得至关重要, 它涉及如何有效地将信息技能、信息资源管理和信息素质融入高等教育课程中, 以满足学生在信息时代的需求。

我国学者针对高校图书馆信息素质教育领域已开展了多方位、深层次的探索, 近年来, 对主题内容和现状分析的探索成为研究热点。目前, 已有文献中对研究主题的分析方法主要有: ①文献计量法。储文静^[4]采用文献题录信息统计分析工具 SA-TI4.0 进行文献计量分析, 研究高校图书馆信息素养教育在教育模式、教育实践、教育路径等方面取得的成果。②调研法。王莲^[5]通过网络调研法对美国与国内比较有影响力的开放式信息素质教育网站的内容、功能、风格等方面进行调研。③内容分析法。肖烨和汤曼^[6]利用 SWOT 分析方法对高校图书馆信息素质教育进行机会、威胁、优势、劣势分析。

已有文献从文献计量法、调研法、内容分析法

收稿日期: 2023-11-01

基金项目: 2023 年度吉林动画学院科学研究项目(党建与思想政治工作专项)重点项目(KY23DZ03)

作者简介: 王智迪(1990—), 男, 吉林长春人, 吉林动画学院图书馆, 馆员, 硕士研究生, 研究方向为文本数据挖掘与文献计量学。

等五个方面为高校图书馆信息素质教育研究的现状分析方面提供了重要的理论与实践参考。但高校图书馆信息素质教育主题挖掘与现状分析方面的已有文献还存在三个问题,一是从整体发文量来看,文献较少;二是发表在核心期刊的文献数量较少;三是近年来对此方面研究的文献较少,多为 10 年前的文献,这也导致所应用的研究方法较为传统。随着数据挖掘技术的兴起,机器学习算法与对各领域有着深度交叉与应用。所以,为了深入了解这一领域的发展趋势和变化,本文采用文本挖掘技术,分析了 1998—2023 年的大量相关文献。通过对文献数据的清洗和处理,构建了 T-LDA2vec 混合模型,用于主题建模和文本分析。这一研究方法不仅有助于挖掘关键主题,还能够揭示不同时间段内主题的强度和演化关系。

1 数据来源与研究方法

1.1 数据来源

样本数据源自中国知网(China National Knowledge Infrastructure, CNKI)中文期刊数据库,设置主题检索为“高校图书馆”和“信息素质教育”,检索时间从 1998—2023 年。检索结果获得 1 606 篇与相关主题文献。将检索的 1 606 篇论文的标题、作者、研究机构、关键词、摘要、发表时间等数据以自定义中的“xls”格式导出,最终将格式转化为“xlsx”。

按年份将样本数据分类,统计每年文献数量,绘制高校图书馆信息素质教育研究的年发文量趋势图(图 1)。从图 1 年发文量可看出其整体变化可以分为两个阶段,第一阶段 1998—2007 年为“学术繁荣期”,年发文量从一个较低的初始值开始,呈现快速增长的趋势;第二阶段 2007—2023 年为“学术调整期”,从 2007 年开始,年发文量开始急速下降,尽管在这期间发文量有所反弹,但总体趋势仍是下降,这意味着该领域正在经历一个学术上的调整阶段。

1.2 研究方法

本文所应用的 T-LDA2vec 是一种混合模型,“T-LDA”是对 LDA 引入时间序列,代表“Time-Topic Latent Dirichlet Allocation”,LAD2vec 结合了 Latent Dirichlet Allocation(LDA)和 word2vec 这两种流行的自然语言处理模型。

LDA 是一种用于文本主题建模的机器学习方法,“在文本挖掘中常用于识别大规模、非结构化文本集或语料库中的隐藏主题信息”^[7],由文档-主题-词构成的三层贝叶斯概率模型(图 2)。其中, M 为所采集的文本数, α 和 β 分别用于控制文档-主题分

布和主题-词分布的先验分布, θ 为文档级别的变量, Z 和 W 为单词级别变量。简而言之, LDA 模型就是通过一篇文章中的每个词以概率的形式抽取某主题,再从这个主题中抽取某特征词,并不断遍历此过程,对于文档中的每个词可以使用文档-主题分布和主题-词分布来计算其生成概率: $p(\text{词语} | \text{文档}) = \sum_{\text{主题}} p(\text{词语} | \text{主题}) \times p(\text{主题} | \text{文档})$

word2vec 一种学习词嵌入(word embedding)的神经网络模型,它通过将词汇映射到高维空间中的向量来捕获词汇之间的语义关系,将 LDA 和 word2vec 结合在一起,可以在文档和单词级别捕获主题和语义信息。word2vec 分为 CBOW 和 Skip-gram 两种模型,本文应用的是 CBOW 模型,它“在训练的过程中,使用上下文的词向量中,从而达到学习语义信息的目的”^[9]。CBOW 通过梯度下降算法对模型的参数进行优化,最大化给定上下文单词预测中心单词的概率,即最大化对数似然函数,计算公式为: $\max L = \ln P(w_i | C)$,整体流程为:输入层-隐藏层-输出层-函数优化(图 3),其中, w_i 是中心单词, C 是上下文。本文的参数设置:嵌入向量的维度 N 为 100 (vector_size=100),上下文窗口大小 c 为 5(window=

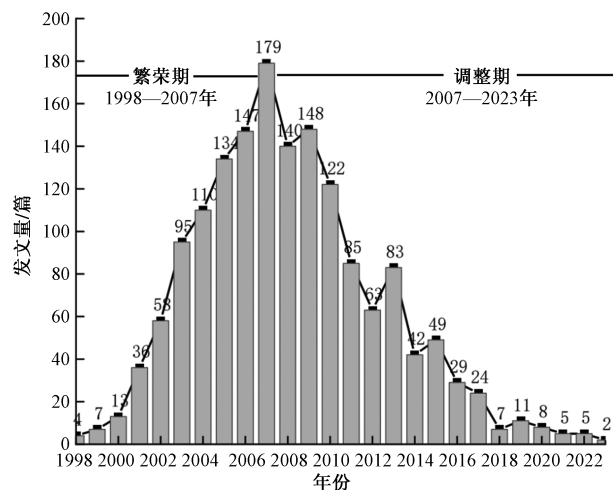


图 1 1998—2023 年发文量趋势

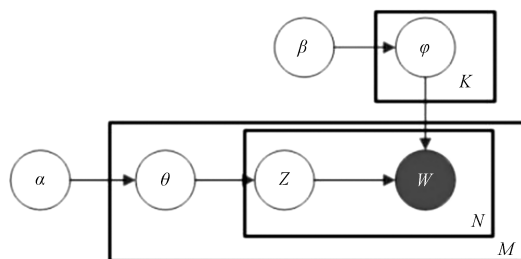


图 2 LDA 主题模型^[8]

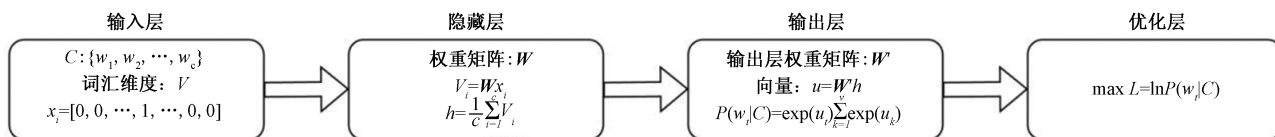


图3 CBOW 流程

5),忽略出现次数少于2的单词($\text{min_count}=2$),使用4个工作线程进行训练($\text{workers}=4$)。

2 研究设计与数据预处理

2.1 研究设计

2.1.1 困惑度计算

困惑度一般用于衡量自然语言处理中训练模型的优劣,其值越低,说明模型对数据的拟合越好。计算公式如下:

式中: D 是文档集; M 是文档的数量; W_d 是第 d 个文档的词语向量; N_d 是第 d 个文档的词语数量, $p(W_d)$ 是文档的似然函数。

$$\text{Perplexity}(D) = \exp\left(-\frac{\sum_{d=1}^M \ln W_d}{\sum_{d=1}^M N_d}\right) \quad (1)$$

2.1.2 主题强度计算

主题强度反映了一个主题在特定文档集合或时间段内的重要性或影响力。一个主题的主题强度越高,表示这个主题在文档集合中越重要,文档集合中的文档越有可能与这个主题相关。主题强度 S_t 计算公式表示为

$$S_t = \sum_{d=1}^D \sum_{w=1}^W \theta_{d,t} \phi_{t,w} n_{d,w} \quad (2)$$

式中: $\theta_{d,t}$ 为主题 t 在文档 d 中的重要性; $\phi_{t,w}$ 为单词 w 在主题 t 下的重要性; $n_{d,w}$ 为单词 w 在文档 d 中出现的频次。

2.1.3 主题演化计算

主题演化是基于生成的主题及其下的关键词与权重生成的。权重通过LDA模型概率分布计算所得。文档-主题概率分布函数表示为 $P(\theta_d | \alpha) \sim \text{Dirichlet}(\alpha)$, θ_d 为文档 d 中每个主题的分布, α 为参数。主题-词汇分布表示为 $P(\phi_k | \beta) \sim \text{Dirichlet}(\beta)$,其中, ϕ_k 为主题 k 中每个词汇的分布, β 为参数。权重的计算涉及从这些分布中估计每个词汇在每个主题中的权重值。

2.2 数据预处理与研究思路

在进行主题建模之前,首先需要对文本数据进行预处理。预处理步骤包括中文分词、停用词去除和同义词替换。中文分词使用Python的jieba库进

行处理,停用词列表是综合了《百度停用词表》《哈工大停用词表》《四川大学机器智能实验室停用词库》等主流中文停用词表来去除无意义的词语。同义词替换是为了将语义相近但表达不同的词语统一为一个词语,以提高模型的准确性。

在数据预处理基础上,对清洗后的“干净数据”进行主题建模与可视化分析。所以,本文的整体思路可以分为数据采集、数据清洗、主题建模、数据分析等四个步骤,研究思路图如4所示。

3 主题分析

3.1 主题挖掘分析

在进行T-LDA2vec主题建模时,需确定主题数量,“困惑度能够衡量主题模型预测样本的精确程度,是目前确定最优主题数使用最多的方法”^[10]。根据图1对“繁荣期”与“调整期”的划分,定义1998—2007年和2007—2023年两个时间窗口,并分别计算困惑度。将主题数区间设为 $[2, 20]$,取500特征词,迭代200次后,两个时间区的困惑度趋势如图5所示。可以看出,1998—2007年,从困惑度在9~18个主题之间相对稳定,在9~13个主题间,困惑度略有下降。2007—2023年,困惑度在5~13个主题之间相对较低并稳定,从第13个主题开始,困惑度开始持续上升。基于上述,确定1998—2007年的最佳主题12个,2007—2023年为13个。

根据困惑度确定最优主题数后,训练word2vec模型,按照两个时区创建LDA主题模型,考虑词频权重的重要性,绘制各主题下的权重值前10的关键词,并对每个主题进行人工命名,如表1所示。

由表1可知,1998—2007年的主题与关键词进行归类,可以分为以下类别:高校教育评估与体系(Topic 1)、师资培养和教学改革(Topic 2, Topic 6, Topic 11)、情报管理与知识结构(Topic 3)、网络化图书馆服务与读者需求(Topic 4)、高校图书馆素质教育服务(Topic 5, Topic 10, Topic 12)、心理素质与教育表现(Topic 7)、地方信息化与课程发展(Topic 8)、数字化图书馆员培养(Topic 9)。对2007—2023年的主题与关键词进行归类,可以分为以下类别:教育体系建设和在线教育(Topic 1,

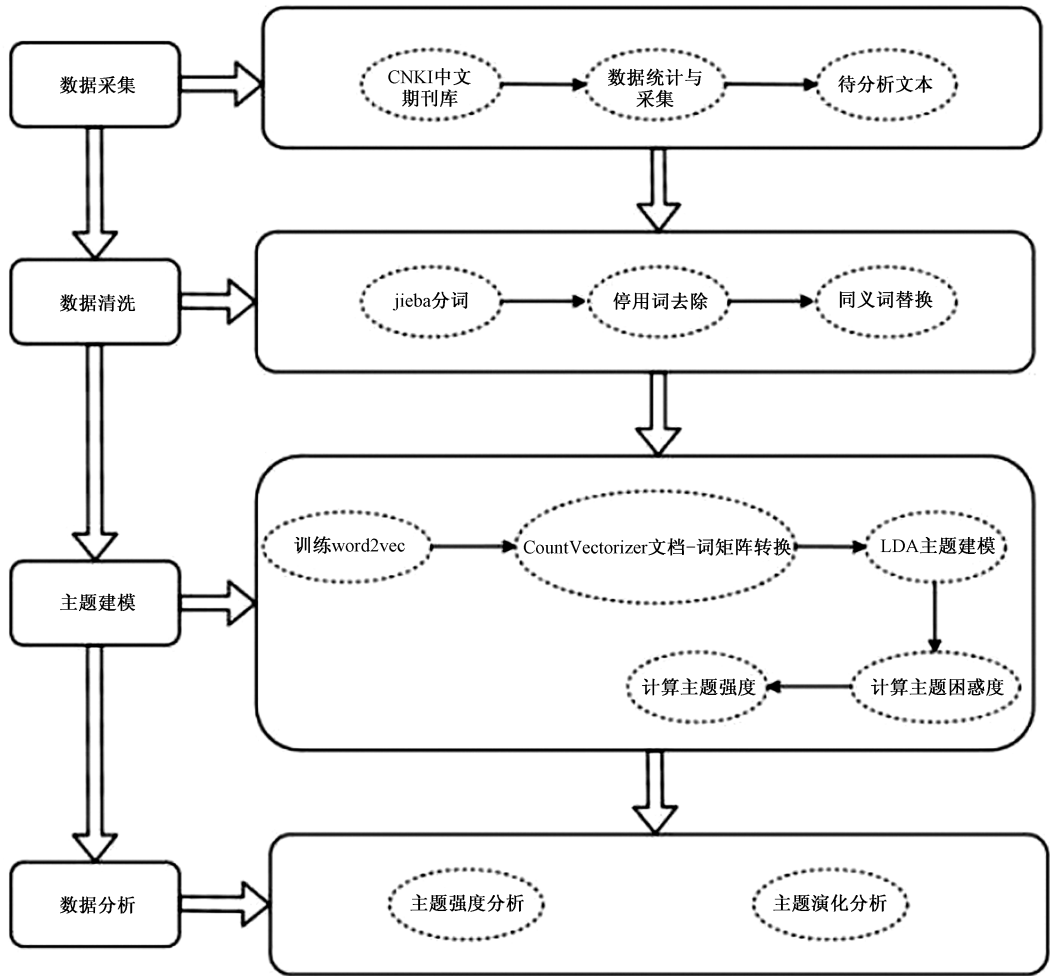


图 4 研究思路

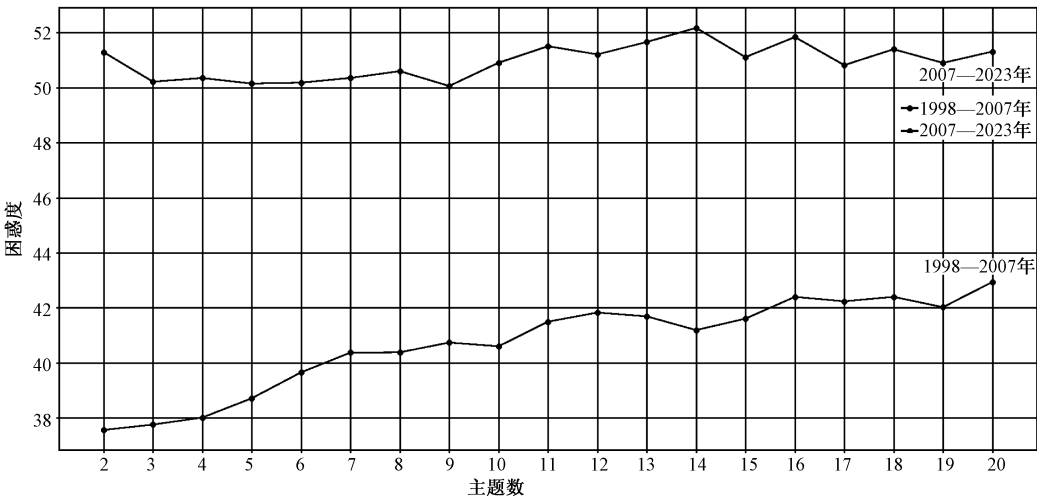


图 5 不同主题数的困惑度值

Topic 2)、学科嵌入式服务和高校图书馆资源创新 (Topic 3, Topic 4)、素质教育在高校图书馆中的角色和素质教育与高校图书馆服务 (Topic 5, Topic 10)、图书馆阅读推广和情报资源管理 (Topic 6)、课程改革和

信息检索教学 (Topic 7)、学术文献检索与学术诚信教育 (Topic 8)、素养教育与信息化发展 (Topic 9)、教育成果与高校信息教育体系 (Topic 11, Topic 12)、图书馆员素质培训和协作管理 (Topic 13)。

表1 不同时区下的主题标签与关键词

主题	关键词(权重)
	1998—2007年
Topic 1: 高校教育评估与体系	体系(27.78), 校园文化(14.13), 评估(11.43), 目标(10.97), 高校(8.39), 教学(7.83), 标准(6.73), 特有(6.26), 指标体系(5.3), 教学内容(4.02)
Topic 2: 师资培养和教学改革	师资(0.08), 切入点(0.08), 校园(0.08), 教学改革(0.08), 综述(0.08), 当今社会(0.08), 系统工程(0.08), 职业(0.08), 针对性(0.08) 统计(0.08)
Topic 3: 情报管理与知识结构	情报(27.6), 知识结构(12.47), 文检课(9.83), 情报信息(8.1), 重视(5.2), 管理(5.18), 入馆(4.38), 领域(3.6), 索引(1.76), 在线(0.94)
Topic 4: 网络化图书馆服务与读者需求	网络(126.53), 环境(74.8), 读者(74.35), 服务(51.95), 文化(29.25), 信息技术(21.5), 革命(16.96), 咨询(16.59), 需求(15.16), 网络化(12.25)
Topic 5: 高校图书馆素质教育服务	图书馆(820.28), 高校(548.93), 素质教育(511.03), 信息(290.37), 教育(280.62), 职能(156.66), 服务(153.57), 文献(114.78), 学生(114.24), 资源(72.16)
Topic 6: 师资发展与教学创新	师资(0.08), 校园(0.08), 切入点(0.08), 教学改革(0.08), 当今社会(0.08), 综述(0.08), 职业(0.08), 系统工程(0.08), 统计(0.08), 服务水平(0.08)
Topic 7: 心理素质与教育表现	心理素质(6.16), 心理(5.36), 创建(3.57), 多方面(2.67), 积极开展(0.08), 表现(0.08), 道德素质(0.08), 课程体系(0.08), 咨询服务(0.08), 培训(0.08)
Topic 8: 地方信息化与课程发展	信息检索(23.6), 地方(16.6), 课程(9.94), 民族(9.72), 全国(8.9), 信息化(8.24), 工委(8.0), 委员会(6.24), 研讨会(5.57), 少数民族(5.3)
Topic 9: 数字化图书馆员培养	图书馆员(71.97), 馆员(45.83), 数字化(26.12), 数字(21.75), 农业院校(7.03), 增加(4.46), 事业(4.43), 国家(3.59), 人才培养(2.79), 教师(2.41)
Topic 10: 高校图书馆与素质教育服务	信息(3463.61), 素质教育(3048.85), 图书馆(1540.29), 高校(1510.49), 大学生(1349.69), 教育(277.74), 能力(137.33), 内容(132.16), 素养(89.88), 必要性(84.91)
Topic 11: 师资发展和校园改革	师资(0.08), 校园(0.08), 切入点(0.08), 教学改革(0.08), 当今社会(0.08), 综述(0.08), 职业(0.08), 系统工程(0.08), 统计(0.08), 针对性(0.08)
Topic 12: 高校图书馆创新与素质教育	素质教育(121.17), 图书馆(106.5), 高校(101.0), 人才队伍(92.53), 信息(73.35), 教学(71.72), 文献检索(67.36), 评价(59.99), 社会(54.77), 创新(45.26)
2007—2023年	
Topic 1: 国家一流教育体系建设	国家(24.75), 一流(8.87), 培养教育(6.19), 积极探索(3.74), 教材(1.06), 持续(1.05), 背景(0.98), 经验(0.97), 详细(0.96), 加快(0.9)
Topic 2: 在线教育平台和用户调查分析	在线(278.23), 用户(136.05), 调查(108.4), 平台(47.7), 国内(21.29), 媒介(19.37), 形式(17.67), 形式(17.67), 在线教育(16.86), 媒体(15.24)
Topic 3: 学科嵌入式服务与创新实践	学科(129.72), 嵌入式(85.83), 服务(49.32), 信息(28.6), 模式(18.67), 创新(17.11), 特色(7.72), 素质教育(5.13), 医学生(4.9)
Topic 4: 高校图书馆资源创新与服务质量	服务(361.62), 图书馆(279.19), 高校(162.65), 资源(92.57), 知识(70.83), 文化(62.98), 创新(54.41), 地方(45.03), 信息(42.13), 科研(34.86)
Topic 5: 素质教育在高校图书馆中的角色	信息(982.96), 高校(790.27), 素质教育(779.46), 图书馆(777.44), 教育(533.99), 网络(178.07), 模式(171.05), 职能(153.07), 环境(131.87), 读者(130.05)
Topic 6: 图书馆阅读推广和情报资源管理	阅读(29.09), 图书馆(18.08), 推广(15.49), 情报(13.02), 农业(8.64), 领域(8.05), 在线教育(7.75), 高质量(6.8), 核心(6.44), 索引(5.98)
Topic 7: 课程改革与信息检索教学	课程(108.43), 信息检索(78.17), 教学(40.21), 改革(34.16), 教学改革(18.87), 高校(16.42), 教师(14.35), 拓展(12.05), 文检课(11.9), 本科(11.81)
Topic 8: 学术文献检索与学术诚信教育	文献检索(140.49), 论文(27.43), 学术(20.52), 图书馆学(11.62), 历程(10.1), 心理素质(8.93), 抄袭(8.9), 诚信教育(7.88), 综述(7.44), 美国大学(6.18)
Topic 9: 素养教育与信息化发展	素养(268.33), 教育(155.07), 创新(133.95), 信息化(65.91), 社会(61.63), 人才队伍(56.51), 信息(50.25), 学习(47.1), 功能(43.03), 互联网(25.99)
Topic 10: 素质教育与高校图书馆服务	信息(3506.11), 素质教育(3354.92), 图书馆(1826.08), 高校(1815.02), 大学生(1548.92), 能力(110.72), 内容(109.49), 教育(106.97), 学生(98.2), 社会(85.38)
Topic 11: 江西教育成果与借鉴经验	江西(11.9), 成果(1.06), 若干(0.97), 问卷调查(0.08), 高等学校(0.08), 推进(0.08), 国内外(0.08), 历程(0.08), 借鉴(0.08), 网络(0.08)
Topic 12: 高校信息教育体系和研究管理	信息(378.62), 素质教育(308.03), 高校(231.56), 图书馆(169.46), 教学(84.04), 新生(71.18), 体系(67.74), 研究生(53.62), 电子(51.44), 教育(50.23)
Topic 13: 图书馆员素质培训与协作管理	图书馆员(147.12), 馆员(124.01), 素质教育(118.32), 信息(111.41), 合作(91.97), 图书馆(74.19), 高校(64.27), 培训(61.5), 管理(39.02), 教师(37.66)

根据表 1 的主题类别与关键词发现多处重复与意义相近的数据,所以对其进行分类与统计,最终得到如表 2 所示的八大主题类别。

3.2 主题强度分析

基于模型输出结果,分别计算 1998—2007 年和 2007—2023 年的主题强度,绘制主题强度的交互式条形图(图 6)来可视化主题在不同时间段内的强度变化,强度较高的主题为热点主题。

由图 6 可知,1998—2007 年强度大于 0.5 的主

表 2 1998—2023 年主题类别统计

主题类别	主题标签(Topic)
教育体系与教育改革	1998—2007 年:高校教育评估与体系(Topic 1)、师资培养和教学改革(Topic 2, Topic 6, Topic 11)
	2007—2023 年:教育体系建设和在线教育(Topic 1, Topic 2)
知识管理与信息服务	1998—2007 年:情报管理与知识结构(Topic 3)
	2007—2023 年:学科嵌入式服务和高校图书馆资源创新(Topic 3, Topic 4)、图书馆阅读推广和情报资源管理(Topic 6)
高校图书馆服务与素质教育	1998—2007 年:网络化图书馆服务与读者需求(Topic 4)、高校图书馆素质教育服务(Topic 5, Topic 10, Topic 12)
	2007—2023 年:素质教育的角色与高校图书馆服务(Topic 5, Topic 10)
教育改革与课程发展	1998—2007 年:地方信息化与课程发展(Topic 8)
	2007—2023 年:课程改革和信息检索教学(Topic 7, Topic 8)
数字化图书馆与信息化发展	1998—2007 年:数字化图书馆员培养(Topic 9)
	2007—2023 年:素养教育与信息化发展(Topic 9)
心理素质与教育表现	1998—2007 年:心理素质与教育表现(Topic 7)
教育成果与信息教育体系	2007—2023 年:教育成果与高校信息教育体系(Topic 11, Topic 12)
图书馆员培训与协作管理	2007—2023 年:图书馆员素质培训和协作管理(Topic 13)

题有“Topic 2:师资培养和教学改革”“Topic 8:地方信息化与课程发展”。高校信息素质教育在发展阶段时师资力量还较为落后,“单一的教学方式、陈旧的教材、落后的师资制约了信息素质教育的有效开展”“在师资队伍建设方面,有些学校至今没有专职文献检索课教师”“师资队伍已远不能适应 21 世纪对大学生信息素质的要求”^[11],急需培养一批具有专业背景、技能熟练的教师队伍,施行教学改革,高校图书馆“要改变以往只以考试成绩评价大学生信息素养水平的传统方式,在学生信息素养水平的评价上,应探索新的评价方式”^[12]。此外,全国与地方的高校图书馆信息素质教育不平衡,信息化水平差距较为明显,地方高校课程发展缓慢,“尤其是经济欠发达地区和不发达地区,经济的落后导致教育的有效供给不足,区域间的经济、文化、信息化程度的不平衡发展导致较大的地区差异、贫富不均等,使得地方性本科院校信息素质教育发展缓慢”^[13]。

2007—2023 年强度大于 0.5 的主题有“Topic 4:高校图书馆资源创新与服务质量”“Topic 9:素养教育与信息化发展”“Topic 10:素质教育与高校图书馆服务”。经过前期发展,高校图书馆信息素质教育从资源匮乏、师资落后向资源创新、图书馆服务创新转变。“信息化的社会发展趋势使人们更加依赖数字化技术进行文献信息的检索和利用”^[14],多元化的数字资源是高校图书馆馆藏资源的主要组成部分,资源创新也是对数字资源的深入拓展。高校图书馆对大学生的信息素质教育是通过组织相关活动或开设文献检索课、信息素质课程实现,信息素质教育领域的服务创新主要由高校图书馆职能驱动,创新类型主要涉及课程改革、新生入馆教

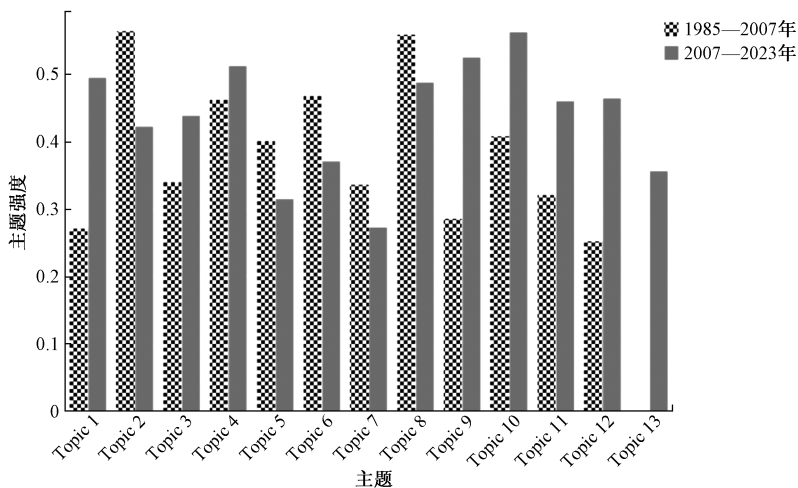


图 6 主题强度交互式条形图

育、体系构建、信息素质培训、嵌入式课程服务等五类^[15]。

3.3 主题演化分析

研究基于 TLDA-2vec 模型分析了 1998—2023 年高校图书馆信息素质教育领域的主题演化。研究方法上,结合表 1 的关键词权重和表 2 的主题分类,构建了两个时期(1998—2007 年和 2007—2023 年)的桑基图(图 7 和图 8),直观地展示了主题间的演化路径。图中线条的粗细代表了主题间关系的紧密程度,反映不同主题和关键词之间的权重差异。

研究结果表明,在两个时间段内,素质教育、信息、高校和图书馆始终是研究的核心内容。这些主题不仅具有广泛的分布和强烈的延续性,而且在不同时间窗口下保持了高度的相似性。其中,高校图书馆服务与素质教育是最为紧密相连的主题。

值得注意的是,随着信息素质教育研究的深入和数字技术的快速发展,研究焦点和主题类别发生了显著变化。1998—2007 年,研究主要集中在高校图书馆服务和素质教育上;然而,进入 2007—2023

年,这一研究重点比例逐渐减少,知识管理和信息服务成为新的关注点。此外,主题和关键词的焦点也由高校图书馆素质教育服务转向素质教育在图书馆服务中的作用,以及情报管理和知识结构方面的探讨。

总体来看,从 1998—2023 年的主题演化趋势揭示了从传统的高校教育和信息素质教育向情报管理、知识结构、在线教育的转型。图书馆的角色也随之从单纯的服务提供者转变为参与素质教育和提供高质量服务的重要参与者。同时,随着技术的发展,数字技术在高校图书馆信息素质教育中的作用日益凸显,预示着情报与知识管理将成为未来研究的重要方向。

4 结语与展望

本文利用 TLDA-2vec 模型深入分析了 1998—2007 年和 2007—2023 年两个时期内高校图书馆信息素质教育的主题演化和强度变化。通过交互式条形图和桑基图的绘制可视化主题强度趋势和演化过程,同时也揭示了关键词的变化趋势。

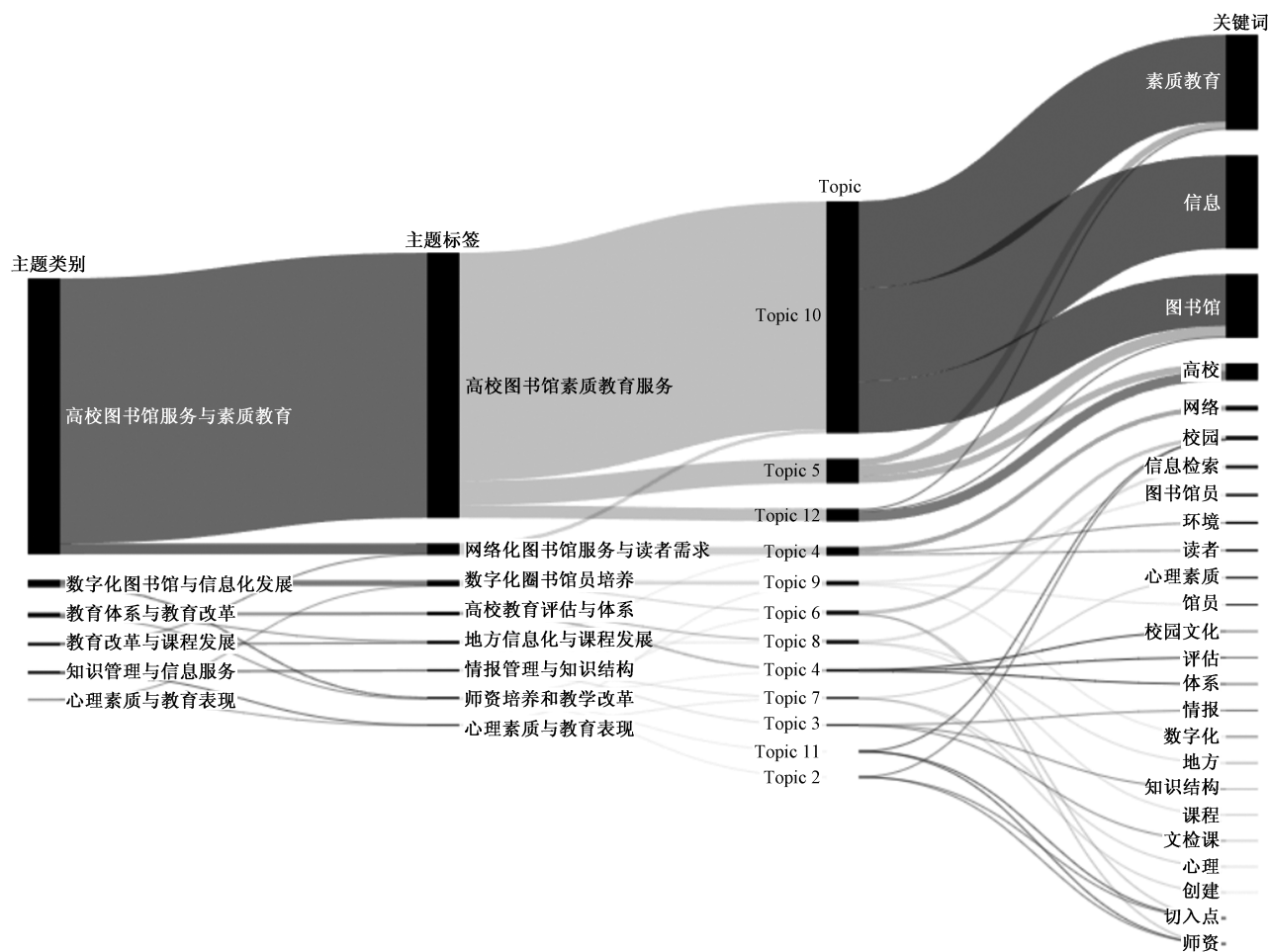


图 7 1998—2007 年主题演化

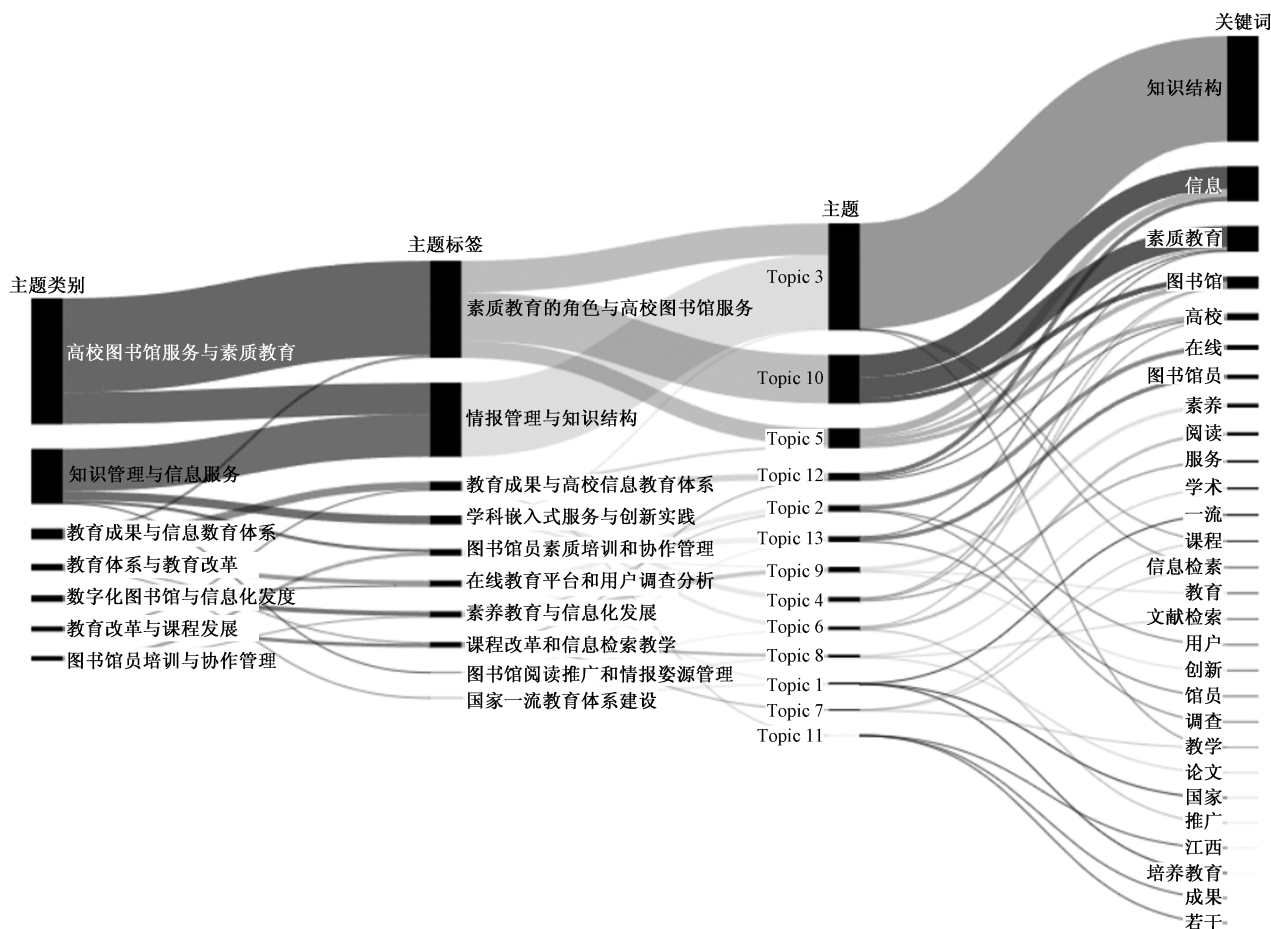


图 8 2007—2023 年主题演化

1998—2007 年,高校图书馆信息素质教育面临了严峻的挑战,包括师资培养和教学改革问题,以及地方信息化与课程发展的不均衡情况。这一时期的研究焦点主要聚焦在提升师资水平和解决地方性差异问题上。

2007—2023 年,高校图书馆信息素质教育经历了资源创新、服务质量提升以及素质教育与信息化发展的阶段。数字技术的飞速发展导致图书馆的角色发生了根本性的变化,从过去的基本服务提供者演变为积极参与素质教育并提供高品质服务的机构。研究兴趣逐渐转向知识管理、情报管理、知识结构以及在线教育等领域。

综上所述,高校图书馆信息素质教育经历了不同历史时期的深刻变革和持续发展,从传统向现代的转变,从基础服务到高品质服务的演进。未来,可以期待更多关于情报与知识管理在高校图书馆信息素质教育中的研究,以适应不断变化的教育和技术环境。这些研究将有助于推动高校图书馆信息素质教育的进一步发展,为学生提供更加优质的信息素质教育服务。

参考文献

- [1] 教育部关于印发《普通高等学校 图书馆规程》的通知[EB/OL]. (2016-01-04)[2023-11-11]. http://www.moe.gov.cn/srcsite/A08/moe_736/s3886/201601/t20160120_228487.html.
- [2] 教育部办公厅关于印发《贯彻落实〈高校思想政治工作质量提升工程实施纲要〉部内分工方案》的通知[EB/OL]. (2018-01-12)[2023-11-11]. http://www.moe.gov.cn/srcsite/A12/s7060/201802/t20180201_326325.html.
- [3] 教育部高等学校图书情报工作指导委员会. 2023 年全国高校信息素质教育研讨会在沈阳召开[EB/OL]. (2023-09-28)[2023-11-11]. <http://www.scal.edu.cn/zxdt/202309280218>.
- [4] 储文静. 高校图书馆信息素养教育探究——基于可视化文献计量统计分析[J]. 办公室业务, 2023(4): 115-118.
- [5] 王莲. 中美开放式信息素质教育网站比较研究[J]. 图书情报工作, 2014(3): 89-95.
- [6] 肖烨, 汤曼. 我国高校图书馆信息素质教育的 SWOT 分析[J]. 图书馆建设, 2010(5): 101-104.
- [7] 孙瑞英, 陈宜泓. 基于 LDA 主题模型的国内智慧阅读研究热点及发展导向研判[J]. 图书馆建设, 2023(3): 84.
- [8] 刘金岭, 钱升华. 文本数据挖掘与 python 应用[M]. 北

- 京：清华大学出版社，2021：69.
- [9] 王刚，郭蕴，王晨. 人工智能技术丛书 自然语言处理基础教程[M]. 北京：机械工业出版社，2022：133.
- [10] 阮霁阳. 数字政府建设影响因素研究——基于127份政策文件的大数据分析[J]. 西南民族大学学报(人文社会科学版)，2022(4)：187.
- [11] 谢艳. 试论高校信息素质教育的现状及应对措施[J]. 教育与职业，2006(8)：95.
- [12] 周建平，肖本罗. 科学发展观与大学生信息素养培养[J]. 商场现代化，2007(7)：377.
- [13] 徐益，陈建英. 地方性本科院校信息素质教育的需求与提升策略[J]. 高校图书馆工作，2011(5)：65.
- [14] 衣建东，徐亚以. 高品质的信息素质教育和阅读推广服务助力高职教育高质量发展[J]. 河南图书馆学刊，2020(12)：74.
- [15] 孙金娟，郑建明，孙红蕾，等. 高校图书馆服务创新实践进展与生成逻辑研究[J]. 图书馆学研究，2022(11)：23-24.

Topic Mining and Evolution Analysis of Information Literacy Education in University Libraries Based on T-LDA2vec

WANG Zhidi

(Jilin Animation Institute Library, Changchun 130000, China)

Abstract: In order to delve into the development trends and evolution processes in the field of information literacy education in university libraries, a total of 1 606 relevant documents from 1998 to 2023 were collected, and after data cleaning and preprocessing, a T-LDA2vec hybrid model was constructed for topic modeling and text analysis. In terms of time trends, two key periods were revealed in the field of information literacy education in university libraries: the academic prosperity period and the academic adjustment period. During the prosperity period, the number of relevant documents rapidly increased, while during the adjustment period, the number of documents sharp declined, reflecting that the field was undergoing academic adjustments. Subsequently, the T-LDA2vec model was used to perform topic mining and determine the optimal number of topics for each period, which were classified into eight categories, including university education evaluation, teacher training, intelligence management, networked library services, information literacy education services in university libraries, psychological qualities and educational performance, local informatization and curriculum development, and digital librarian training. The strength of each topic within different time periods was calculated and hot topics were described through interactive bar charts. The study found that some topics maintained high intensity during different periods, indicating their important influence in relevant documents. Topics such as teacher training, teaching reform, and innovation and quality of university library resources maintained high intensity during different periods. Through topic evolution analysis, the correlation and evolution process of topics during different periods were revealed, pointing out that the focus of research on information literacy education in university libraries has gradually shifted from basic services to areas such as resource innovation, knowledge management, and online education. A comprehensive understanding of the research trends in this field and a useful reference for future research directions and policy development are provided. Additionally, a beneficial methodological demonstration for the application of text analysis methods is put forward.

Keywords: text mining; LDA; word2vec; university library; information literacy education