

R语言在直观教学中的应用与可视化分析

邓文丽, 鄢田云, 段金林, 杨 闰, 郭 霞, 柳敏焯, 李泽滔, 刘春莉, 袁馨琪, 周 懿,
罗 艺, 杨浩宇

成都信息工程大学应用数学学院, 四川 成都

收稿日期: 2025年1月27日; 录用日期: 2025年2月19日; 发布日期: 2025年2月28日

摘 要

随着数据科学的发展, R语言在高等教育中的直观教学应用日益增加。本文通过Python爬虫从知网获取了79篇关于R语言直观教学的文献, 筛选后保留了76篇用于数据分析, 进行了数据预处理, 包括删除与主题无关的文献、提取关键词并按学科分类。接着, 利用Excel透视表饼图和R语言词云图进行可视化分析, 研究R语言在各学科的应用及关键词热点。最后, 运用LDA模型分析文献摘要, 识别R语言直观教学领域的四个研究热点。结果表明, R语言不仅提升了教学质量, 还推动了教学方法的创新, 对未来学术研究与教学实践具有深远的影响。

关键词

R语言, Python, 直观教学, 可视化, LDA主题模型

Application and Visualization Analysis of R Language in Intuitive Teaching and Learning

Wenli Deng, Tianyun Yan, Jinlin Duan, Run Yang, Xia Guo, Minyang Liu, Zetao Li,
Chunli Liu, Xinqi Yuan, Yi Zhou, Yi Luo, Haoyu Yang

College of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Jan. 27th, 2025; accepted: Feb. 19th, 2025; published: Feb. 28th, 2025

Abstract

With the development of data science, the application of R language for intuitive teaching in higher education is increasing. In this paper, 79 pieces of literature on intuitive teaching and learning of R language were obtained from CNKI through Python crawler, and 76 were retained for data analysis after screening. Data preprocessing was carried out, including deleting literature unrelated to the topic,

文章引用: 邓文丽, 鄢田云, 段金林, 杨闰, 郭霞, 柳敏焯, 李泽滔, 刘春莉, 袁馨琪, 周懿, 罗艺, 杨浩宇. R语言在直观教学中的应用与可视化分析[J]. 统计学与应用, 2025, 14(2): 150-163. DOI: 10.12677/sa.2025.142044

extracting keywords and categorizing them by discipline. Then, Excel pivot table pie charts and R language word cloud maps were used for visual analysis to study the application of R language in various disciplines and keyword hotspots. Finally, the LDA model was used to analyze the literature abstracts and identify four research hotspots in the field of R language intuitive teaching and learning. The results show that R language not only improves the quality of teaching, but also promotes the innovation of teaching methods, which has a far-reaching impact on future academic research and teaching practice.

Keywords

R Language, Python, Intuitive Teaching, Visualization, LDA Topic Models

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着科技的快速发展,计算机辅助教学逐渐成为现代教育的重要组成部分。传统教学中,教师常依赖抽象语言和静态图像,难以激发学生的兴趣。引入计算机图形技术,特别是在数学、物理和工程等学科中,显著提升了教学质量。计算机生成的图形能够动态展示复杂数据和概念,增强学生的参与感。R语言作为开源统计计算和数据可视化工具,因其简单性和强大的功能,广泛应用于教育领域[1],能够创建高质量的图表,并允许教育者定制开发个性化的教学方案。尽管已有大量文献探讨R语言在教学中的应用,但系统性评价仍缺乏。本文拟对2012~2024年间R语言在国内高校课程直观教学的研究进行文献回顾,挖掘主要应用情况和研究热点,以便为传统课程引入R语言提供建议。研究显示,直观教学法在高等数学、材料力学和机械专业等多个学科中得到广泛应用,国外高校如斯坦福和哈佛也将R语言作为核心教学工具,设计生动的案例以激发学生的兴趣。因此,R语言在直观教学中的应用日益受到重视,教育工作者不断探索与创新,以提高教学效果,培养掌握R语言的复合型人才,支持各行业的数据分析与决策。

2. 数据获取与预处理

2.1. 数据来源

在CNKI数据库中,以“R语言教学”为关键词进行数据爬虫[2],预期爬取79篇文献,实际获得78篇,最后一篇因故障需手动补充。文献的主要内容由编号、题目、作者、单位、发表日期、文献来源、专辑、专题、来源库、引用、下载、关键词、摘要以及URL-网址14列构成。

2.2. 数据预处理

从CNKI中获取的数据不能直接用来进行数据可视化分析,需要经过进一步的处理才能够用于后续的分析,本文中对于文献数据的预处理包括对文献的筛选、文献的一级学科分类、MySQL数据库导入三个部分。

2.2.1. R教学数据的筛选

在使用Python爬取数据时,程序在获取第78条文献时意外停止。通过在CNKI上以“R语言教学”为关键词检索,共找到79条相关文献,但最后一条《R语言在统计学教学中的运用》未成功导入。手动

补充后，数据文件中无空值。

部分获取的文献与研究主题不符，如图书来源的《R 语言》和《语言学统计研究方法》，因此将这些不符合要求的文献删除，最终保留有效数据 76 篇。

2.2.2. R 教学数据的一级学科分类

按照《中华人民共和国学科分类与代码》将所得文献数据分为数学、生物学、统计学、教育学、医学、计算机、地理、经济学、资源科学以及化学 10 门学科。

通过专题中所包含的学科进行分类，如果专题中存在一种或多种学科，则按出现的先后和其所属的一级学科进行分类，分类标准如表 1 所示。

Table 1. Examples of topic-based categorization of first-level disciplines

表 1. 基于专题的一级学科分类示例

专题	一级学科
自然地理学和测绘学、高等教育、计算机软件及计算机应用	地理
资源科学、水产和渔业、高等教育	资源科学
数学、职业教育	数学
生物学、植物保护、高等教育	生物
教育理论与教育管理、计算机软件及计算机应用	教育学
医学教育与医学边缘学科、高等教育	医学

专题中没有学科内容的，按文献的关键词进行分类，如《R 语言与生物统计学结合的教学实践探索》中的关键词有 R 语言、生物统计、教学改革、统计软件 and 数据处理，综合来讲该文献属于生物学科。《基于 R 语言 Shiny 包的概率统计交互可视化教学研究与实践》的关键词有抽样分布、Shiny 和动态网页，综合为计算机学科。按照上述方法将剩余的文献中专题为空值的文献分类，结果如表 2 所示。

Table 2. Keyword-based classification of first-level disciplines

表 2. 基于关键词的一级学科分类

题目	关键词	一级学科
如何在计算机 R 语言教学中实现数据可视化的应用研究	R 语言教学、实现、数据可视化、应用	计算机
谈 R 语言在辅助教学方面的应用	R 语言、概率统计、辅助教学	数学
基于 OBE 模式的三重考核制《R 语言基础》课程线上教学探究	R 语言基础、OBE 教学模式、线上教学；信息化教学	教育学
R 语言 animation 包的辅助教学探索	R 语言、animation 包、概率统计、辅助教学	数学
浅议本科教学中 R 语言的教学模式	R 语言、编程训练、教学模式	教育学
教育教学中 R 语言运用的优势	无	教育学
R 语言在经济数学模型教学中的应用	财政收入、时间序列模型、R 语言、回归分析	经济学

部分文献的学科相关度很高，文献量又很少，采用将其归为一类的方式进行分类。将《R 语言在经济数学模型教学中的应用》《“商业分析学”课程教学创新策略探析——基于 R 语言实证研学模式》《基于 R 语言的“金融时间序列分析”课程教学改革与实践》《“数理金融”课程教学改革与 R 语言应用》

《口试考核在 R 语言与金融数据分析课程教学中的应用探索》五篇关于经济学和金融学这两门密切相关的学科统一归为经济学的类别。

2.2.3. 数据库导入

爬取并保存在 Excel 中的数据年份格式为“yyyy/mm/dd”，我们分析数据只需要准确到年份即可，因此先将表中的年份数据统一修改为“yyyy”格式。将文献摘要数据以“content”为数值键保存在 Excel 文件中，用来作为 LDA 主题分析的原始数据。

然后建立数据库来管理数据资料。在数据库表中，以 Title 作为主键，数据类型为 VARCHAR(100)；以 Author、Year、Cited、Subject 为键的数据类型分别是 VARCHAR(45)、INT、INT、VARCHAR(45)，导入 MySQL 数据库中，创建的 MySQL 表结构如表 3 所示。

Table 3. Object list
表 3. 对象列表

序号	键名	中文名称	数据类型
1	Title	文献题目	VARCHAR(100)
2	Author	作者	VARCHAR(45)
3	Year	年份	INT
4	Cited	引用量	INT
5	Subject	学科	VARCHAR(45)

3. R 语言在直观教学情况的可视化分析

本章主要统计各基本学科采用 R 语言教学的占比情况，并进行可视化分析。对目前 CNKI 中关于 R 语言直观教学的相关文献研究，主要分析文献引用量情况以及文献的热点关键词，并对 R 语言直观教学在各学科中的研究现状进行可视化分析。

3.1. 各学科相关文献的占比和引用量分析

3.1.1. 各学科占比分析

使用 Excel 中的数据透视表，绘制 R 语言直观教学相关文献中一级学科所占总比的饼图，如图 1 所示。

分析图 1 中的饼图，数学、生物和统计学是占比最高的学科，其中数学以 25% 的占比位居首位，显示出其在研究和教学中的重要性。生物学以 17.11% 排名第二，反映了 R 语言在生物统计和生物信息学中的应用情况。统计学占比 14.47%，强调了 R 语言在数据科学中的核心地位。

计算机科学、医学和教育学的综合占比总数为 24%，显示了 R 语言在多学科交叉应用中的广泛性。这些领域利用 R 语言进行机器学习、流行病学研究和学习数据分析等。

地理学、经济学和资源科学的占比比较小，表明这些领域对 R 语言的需求尚未充分开发。尽管如此，随着研究的深入，这些领域的需求可能会增加。化学的占比最低，仅为 1.32%，主要由于传统化学更依赖于实验。

整体来看，R 语言作为一种强大的统计和图形工具[3]，学科中的应用显示了其广泛的适用性和在学术研究中的核心作用。随着数据科学的发展，R 语言在未来的学术和教学中将扮演越来越重要的角色。各学科之间的比例关系反映了科学研究领域的多样性和相互关联性。

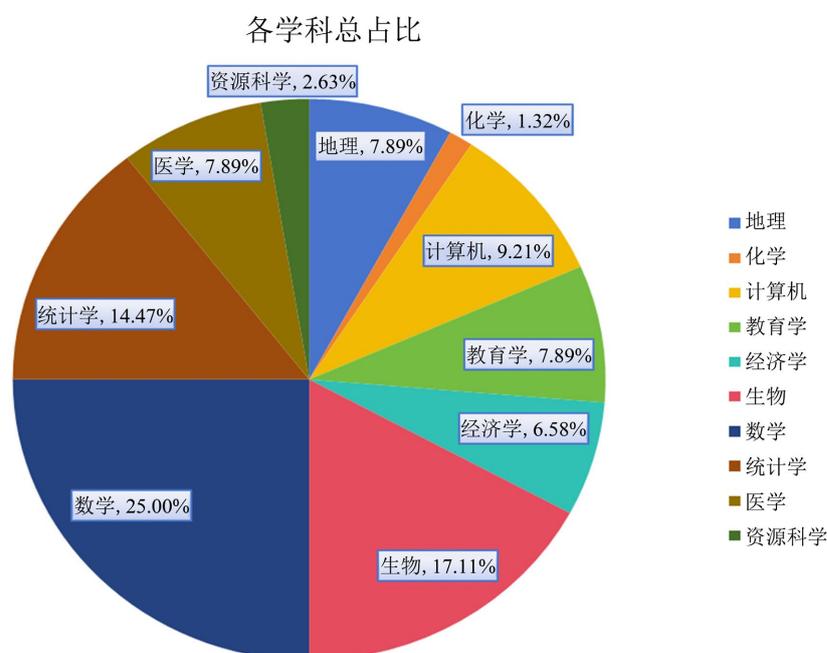


Figure 1. Pie chart of percentage of each discipline
图 1. 各学科占比饼图

3.1.2. 引用量分析

文献的引用量是评价学术研究影响力的重要指标[4]。高引用量通常意味着该文献在学术界具有较高的认可度和关注度，可能对相关领域的发展产生了显著影响。因此，对引用量的统计分析可以揭示在该学科中 R 语言直观教学的研究现状和运用情况。首先查询各学科引用量前三的文献及相关信息，然后对其进行教学应用情况分析，最后从跨学科教学项目、在线资源和 MOOCs、教育研究和评估这三方面拓展 R 语言的教育应用。

1) 各学科引用量前三的文献及文献量分析

通过 MySQL 窗口函数查询语句先选出在各学科中引用量前三的文献及相关信息，相关信息包括文献名称、作者、发表年份、引用量、所属学科和在该学科中的引用排名，再通过总和和分析影响力。

代码部分：

```
//以学科(Subject)为分组，引用量(Cited)从大到小排序，通过窗口函数//ROW_NUMBER()查询出各学科中引用量前三的文献
SELECT *
FROM
  (SELECT
    *,
    ROW_NUMBER() over
    (partition by `Subject` order by `Cited` DESC) rn
    FROM `papers`) AS r
WHERE rn<=3;
```

为了将 MySQL 筛选出来的数据可视化，把得到的数据以 CSV 的格式存储，然后通过 R 语言中的“ggplot2”和“dplyr”包做堆积图展示。各学科引用量前三的文献数据堆积图如图 2 所示。

各学科文献引用量

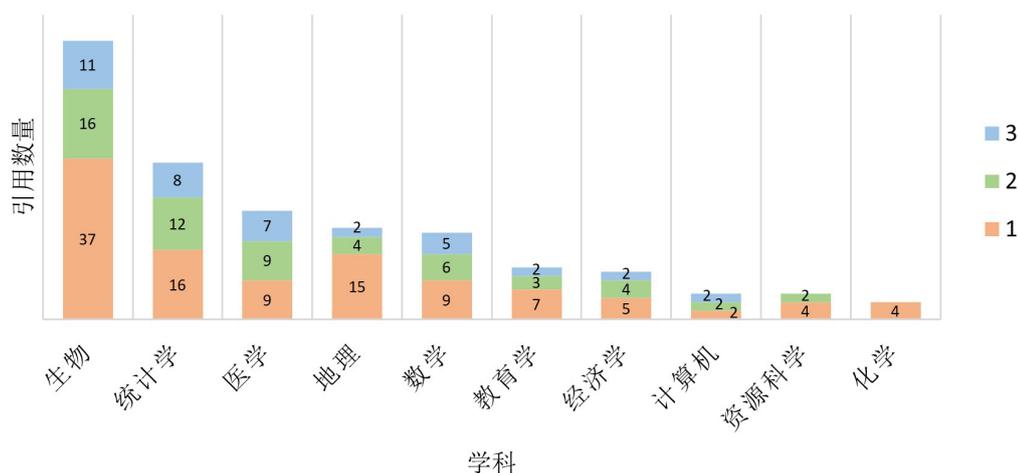


Figure 2. Stacked diagram
图 2. 堆积图

根据图 2 所给的数据, 对 R 语言在各个学科中的教学应用情况进行更具体和深入的分析。以下是对各个学科中 R 语言引用情况的详细总结。

在生物学中, 《浅谈 R 语言在生物统计学教学中的应用》具有最高的引用量(37 次), 这表明 R 语言在生物统计学中占据核心地位。在统计学中, R 语言自然是核心工具, 文献《基于 R 语言的统计教学应用初探》有 16 次引用, 表明 R 语言在教授现代统计方法中的广泛应用和重要性。在医学领域中, 对 R 语言的关注集中在统计分析教学上, 《医学本科生 R 语言教学初探》和《医学统计学教学中利用 R 语言进行描述性统计分析》均有 9 次引用, 这反映了医学研究中对数据分析工具的高需求, 以及 R 语言在处理医学数据上的适用性。在地理学领域中, 《R 语言可视化的优势及其在空间统计教学中的应用》引用量达 15 次, 显示了 R 语言在地理学中特别是空间数据分析的重要性。在数学领域中, 《R 语言在数理统计教学中的应用及延伸》有 9 次引用, 强调了 R 语言在提高数学和统计课程教学质量中的作用[5]。在教育领域中, R 语言被用来探讨教学方法和评估教学效果, 《浅谈 R 语言在高等学校教学当中的重要性》有 7 次引用, 表明了 R 语言在教育研究中作为数据分析和结果可视化工具的潜力。在经济学领域中, R 语言主要被用于金融时间序列分析和数理金融课程, 《基于 R 语言的〈金融时间序列分析〉课程教学改革与实践》有 5 次引用, 突出了 R 语言在现代经济学教学中的应用, 尤其是在金融数据分析中。在计算机科学中, 计算机科学领域中, R 语言主要用于数据可视化和数据分析课程, 《如何在计算机 R 语言教学中实现数据可视化的应用研究》有 2 次引用, 反映了计算机科学领域对数据分析技能的需求。在资源科学中, 尽管引用量不如其他学科, R 语言在资源科学中的应用表明了其在环境数据分析和资源评估中的潜力。在化学中, 虽然引用量最少, 但是 R 语言在化学数据分析、化学实验结果的可视化以及化学教育中的潜在应用, 未来可能成为一个重要的研究方向。

2) 拓展 R 语言的教育应用

跨学科教学项目: 鉴于 R 语言在不同学科中的广泛应用, 学校和教育机构可以考虑开展跨学科的 R 语言教学项目。例如, 结合计算机科学、生物学和统计学的课程, 通过项目基础教学来增强学生的实际操作能力和解决问题的能力。

在线资源[6]和 MOOCs: 随着在线教育的兴起, 高校和教育机构可以开发关于 R 语言的在线课程和开放式在线课程, 这不仅可以扩大学生对 R 语言学习的接触面, 还可以提高学科间的互动和合作。

教育研究和评估：R 语言的强大数据处理和可视化功能使其成为教育研究中的一个宝贵工具。研究人员可以利用 R 语言来评估教学方法的效果，分析学生表现数据，从而更科学地指导教学方法和教育政策的制定。

3.1.3. 小结

R 语言作为教学工具的广泛影响：从上述分析可以看出，R 语言在多个学科中都展现出了其作为教学工具的巨大潜力。通过实际案例和应用，R 语言不仅提高了学术教育的质量，还促进了新教学方法的发展，尤其是在数据分析和统计可视化方面。

未来展望：随着科学研究和教育的不断发展，预计 R 语言的应用将进一步扩展到更多学科领域中。同时，对 R 语言教学内容的持续更新和改进，将是提高教学质量和学生学习成效的关键。

综上所述，R 语言在学术和教育领域的应用显著，未来有望在全球范围内继续扩大其影响力，成为一个不可或缺的教学和研究工具。

3.2. 关键词分析

在文献计量学[6]中，学科领域的研究热点一般采用高频关键词进行表征。关键词频次越高，其所代表的研究内容越受关注。关注度较高的研究方向一般预示着该领域未来的发展趋势，有助于研究学者从整体上把握领域发展动态。

本节通过统计文献关键词及其词频，将结果导入 Excel，并使用 R 语言中的“jiebaRD”、“jiebaR”、“wordcloud2”包生成词云图，以分析 R 语言直观教学的研究热点。由于选取的文献均包含“R 语言”这一关键词，因此在统计时可将其排除，从而得到表 4 中的文献关键词及表 5 中的关键词频次。接着，使用 R 语言对统计的关键词频次进行词云图分析，如图 3 所示。

Table 4. Literature keywords

表 4. 文献关键词

文献标题	文献关键词
1	医学统计学；课堂教学改革；R 语言；生物医学工程专业
2	医学统计学；大数据；R 语言
3	统计学教学；统计软件；R 语言
4	统计学教学；R 语言；现代回归分析

Table 5. Keyword word frequency

表 5. 关键词词频

文献中出现的关键词	对应关键词词频
医学统计学	2
统计学教学	2
课堂教学改革	1
生物医学工程专业	1
统计软件	1
大数据	1
现代回归分析	1

教学策略：词汇如“案例教学”和“直观教学”显示了教学方法在使用 R 语言教育过程中的关键角色。这表明教育者在授课时倾向于采用具体案例和直观材料来提高教学效果[8]。

4. 基于摘要的 LDA 主题发现

文档的主题可以理解为语料库中一组具有相似语境或语义的词汇集合。摘要是作者提炼出来的文献的内容概括，对文献的摘要建立主题模型，可以识别摘要中蕴含的主题，挖掘摘要中隐藏的信息，即可以实现以一种非监督学习方式对摘要中隐含的语义结构进行聚类的效果，从而分析出我国 R 语言在直观教学领域的集中关注点及其相关特征词。

4.1. LDA 模型介绍

4.1.1. LDA 模型的基本原理

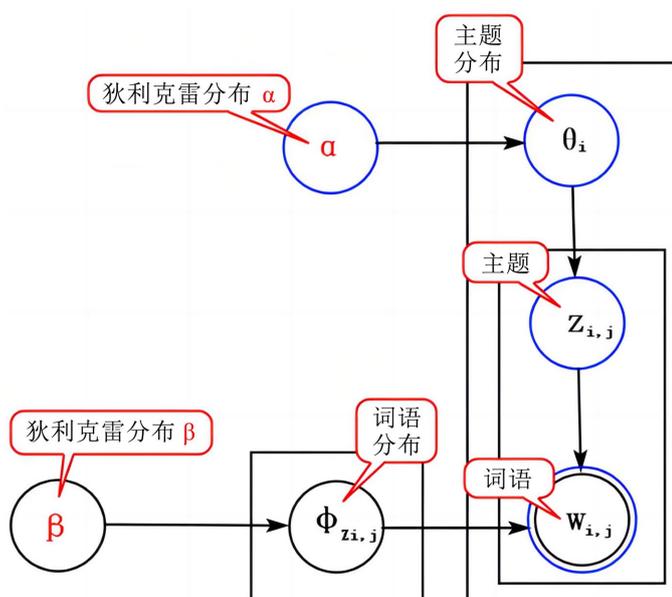


Figure 4. LDA model
图 4. LDA 模型

LDA 模型的基本思想是将文档表示为多个主题的混合物，每个主题又是由单词的分布组成。具体来说，给定一个文档集 D ，LDA 模型的目标是推断出每个文档的主题分布和每个主题的单词分布。LDA 模型通过以下步骤进行建模。

1) 定义符号与假设

D : 文档集。

N : 文档中单词的数量。

K : 主题的数量。

α : 狄利克雷先验分布的参数，用于文档的主题分布。

β : 狄利克雷先验分布的参数，用于主题的单词分布。

η : 主题分布参数 β 的先验分布参数。

2) 生成过程

对于文档集 D 中每个文档 d :

① 从狄利克雷分布中抽取一个主题分布 $\theta_d \sim \text{Dir}(\alpha)$ 。

② 对于文档中的每个单词位置 n :

从主题分布中抽取一个主题 $z_{d,n} \sim \text{Multinomial}(\theta_d)$;

从主题对应的单词分布中抽取一个单词 $w_{d,n} \sim \text{Multinomial}(\beta_{z_{d,n}})$ 。

通过上述过程, LDA 模型能够生成每个文档的主题分布和每个主题的单词分布。这一过程可以用图形模型表示, 如图 4 所示。

公式(1)和(2)分别表示 LDA 模型的似然函数和后验分布:

$$P(w, z | \alpha, \beta) = \prod_{d=1}^D \left(\int_{\theta_d} p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{d,n}} p(z_{d,n} | \theta_d) p(w_{d,n} | z_{d,n}, \beta) \right) d\theta_d \right) \quad (1)$$

$$p(\theta, \beta | w, \alpha, \beta) \propto p(\theta | \alpha) p(\beta | \eta) \prod_{d=1}^D \left(\prod_{n=1}^{N_d} p(w_{d,n} | z_{d,n}, \beta) p(z_{d,n} | \theta_d) \right) \quad (2)$$

4.1.2. LDA 模型在文献摘要上的应用

LDA 模型在文献摘要中的应用主要体现在识别和分析文档集中的潜在主题结构[9]。这种应用在科学研究中非常普遍, 尤其是在大规模文献分析、科研热点识别等方面。这里描述了 LDA 模型在文献摘要上的具体使用方法和步骤。

1) 数据预处理

在应用 LDA 模型之前, 首先需要对文献摘要数据进行预处理。预处理步骤包括:

- ① 数据清洗: 去除文档中的噪音和无关信息, 如标点符号、停用词等。
- ② 分词处理: 将文档摘要分割成单个单词或词组。
- ③ 词频统计: 计算每个词在文档中的出现频率, 构建词袋模型。

2) 模型训练

在数据预处理完成后, 接下来就是训练 LDA 模型。LDA 模型训练过程包括以下三个步骤:

- ① 确定主题数: 选择合适的主题数 K 。可以通过困惑度指标或交叉验证来确定最优主题数。
- ② 初始化参数: 设置 LDA 模型的超参数 α 和 β , 通常使用经验值或通过调参确定。
- ③ 迭代训练: 使用变分推断或吉布斯采样方法, 迭代更新模型参数, 直到模型收敛。

3) 主题识别与分析

训练完成后, LDA 模型会输出每个文档的主题分布和每个主题的词分布。通过分析这些输出, 可以进行以下三方面的工作:

- ① 主题识别: 识别文档集中的主要主题, 每个主题由一组高频词组成。
- ② 主题可视化: 使用图表等方式展示主题及其变化趋势, 便于理解和解释。
- ③ 热点分析: 根据主题分布, 识别文献集中的研究热点和发展趋势。例如, 通过分析科研论文的摘要, 可以发现当前的研究热点和未来的发展方向。

总之, LDA 模型在文献摘要中的应用能够帮助研究者快速理解文献集中的主要主题和研究热点, 为科学研究提供有力支持[10]。

4.2. R 语言直观教学的热点主题识别

4.2.1. 主题数量的选取

当 LDA 模型的结果显示每个主题都是互相隔开的, 这通常意味着模型成功地识别出了文档集中的不

同主题，并且每个主题都有其独特的词汇分布，理论上这种现象是模型效果好的一个标志。因此，通过对比选取不同数量的主题时，LDA 模型结果中主题分布情况选出合适的主题数量进行选取[11]。

根据分析，在选取主题数量为 4 个时每个主题都是互相隔开的，选取 3 个主题时会漏掉一个主题，选取 5 个时有 2 个主题重合度较高。因此，选取 4 个主题数量的模型效果良好，所以在主题数量上选择 4 个作分析。

4.2.2. 主题分析

在 Python 中使用 sklearn-LDA 模型对知网上所有有关 R 语言直观教学的中文文献摘要建立 LDA 模型，模型图如图 5 所示，选取 4 个主题以及和主题最相关的 10 个词汇，如表 6 所示。

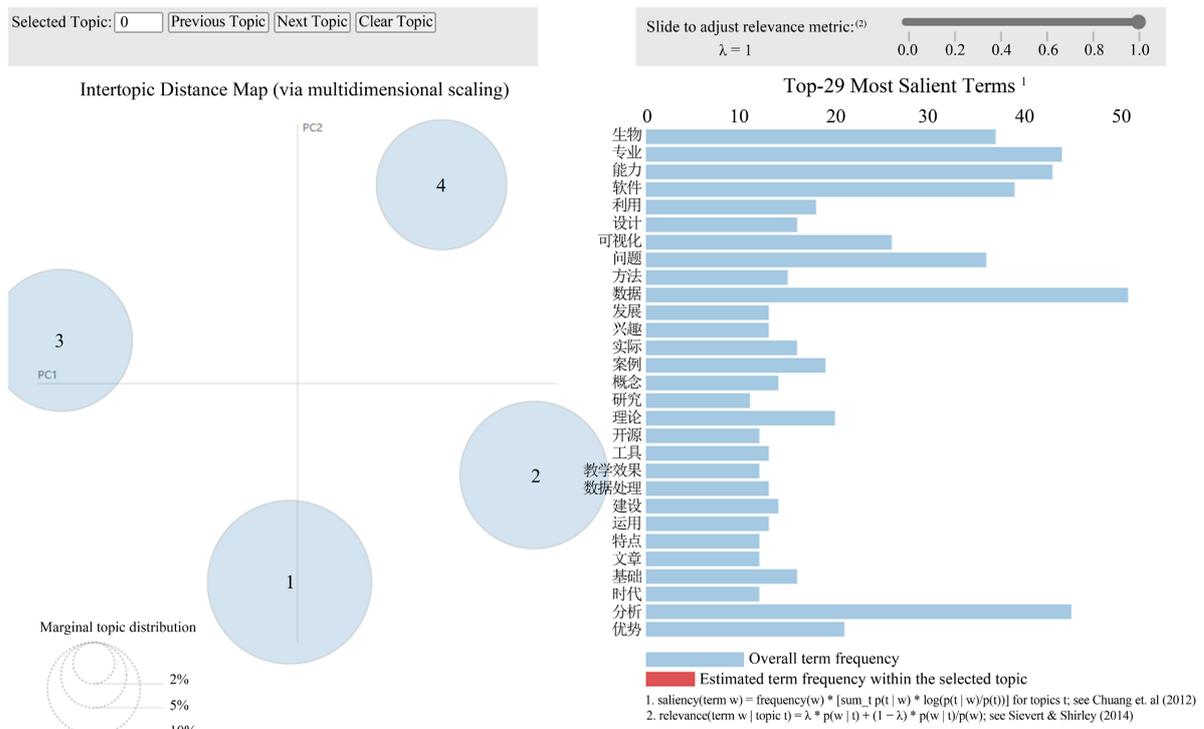


Figure 5. LDA model diagram for the abstract of literature related to intuitive teaching and learning in R language
图 5. R 语言直观教学相关文献摘要的 LDA 模型图

Table 6. LDA theme distribution
表 6. LDA 主题分布

主题编号	与主题最相关的词汇
1	问题，专业，设计，数据，方法，发展，分析，工具，运用，数据处理
2	生物，专业，能力，理论，基础，特点，数据，教学效果，概念，建设
3	数据，软件，可视化，分析，案例，理论，优势，开源，文章，建设
4	能力，利用，分析，实际，兴趣，问题，软件，概念，案例，研究

图 5 展示了根据知网上所有有关 R 语言直观教学的中文文献建立 LDA 模型的结果。表 6 中列出了文献中最重要的 4 个主题以及与主题相关的词汇。

1) 专业性的数据分析

主题描述：该主题关注数据分析的方法、设计与利用。数据分析在教学中具有独特的优势，通过实战操作、方法设计和数据处理等过程，使学生深入了解数据科学的专业性。这种教学方法帮助学生掌握数据分析的基本技能，并在实际应用中发挥作用。相关词汇包括：问题、专业、设计、数据、方法、发展、分析、工具、运用、数据处理。

2) 生物数据分析

主题描述：该主题强调生物学与数据分析的结合，聚焦于生物数据处理的能力和工具的应用。数据分析在生物信息学领域拥有许多特定的软件包和工具，可以帮助学生在生物领域应用数据分析技能，理解生物数据的特点，并从事相关研究。相关词汇包括：生物、专业、能力、理论、基础、特点、数据、教学效果、概念、建设。

3) 数据分析与可视化

主题描述：该主题突出数据、软件、可视化和分析，体现了数据分析和可视化领域的重要作用。数据分析提供了丰富的数据分析工具和可视化包，可以帮助学生直观地理解数据，从而提升分析能力和解决问题的能力。相关词汇包括：数据、软件、可视化、分析、案例、理论、优势、开源、文章、建设。

4) 实践与教学效果

主题描述：该主题强调实际操作和教学效果。数据分析教学中的实践环节，包括代码编写、数据处理和分析，可以让学生获得实用技能并提高教学效果。这种教学方法培养学生解决问题的能力，让他们在实践中巩固所学知识。相关词汇包括：能力、利用、分析、实际、兴趣、问题、软件、概念、案例、研究。

这些主题展现了 R 语言在数据分析与教学中的多方面作用。通过专业性、实践性和理论基础相结合的教学方式，R 语言可以帮助学生掌握数据分析技能，拓展其专业领域，为职业发展打下坚实基础。

5. R 语言在直观教学中的优势

在当今数据驱动的时代，统计学与数据分析已成为各学科领域的重要基础。随着大数据、人工智能和机器学习技术的迅猛发展，数据分析的需求与日俱增。作为一种专为统计计算和数据分析设计的编程语言，R 语言在学术界和商业界都占据了重要地位。由于其强大的数据分析与统计功能、丰富的包生态系统和专用功能以及开源特性与全球协作，R 语言在直观教学中展现出了显著的优势，成为许多高等院校数据科学和统计学课程中的首选工具。

1) 强大的数据分析与统计功能

R 语言自 20 世纪 90 年代初面世以来，一直被专门设计用于统计计算和数据可视化，因此它在统计学领域得到了广泛应用。与传统的统计分析软件，如 SPSS、Matlab 相比，R 语言具有显著的优势。首先，R 语言是开源的，这意味着教师和学生可以自由使用，并根据需求进行定制化开发。其次，R 语言拥有丰富的扩展包和强大的统计计算功能，支持从基本的数据处理到复杂的统计分析和建模任务[12]。

此外，R 语言的易用性也使其在教学中占据了重要地位。虽然其他编程语言(如 C、C++、Java 或 Fortran)也能进行数据分析，但它们通常在学习曲线和开发效率上较为复杂。与此不同，R 语言通过简洁的语法和强大的统计库，使学生能更快地掌握数据分析技能，进而应用于实际问题的解决。因此，在数据分析和统计计算课程中，R 语言已成为备受推崇的教学工具。

2) 丰富的包生态系统和专用功能

R 语言在数据分析、可视化以及统计建模方面的独特优势，主要体现在其丰富的包生态系统和专门为统计计算设计的功能。R 语言内建了大量便捷的数据处理和分析工具。例如，dplyr 包提供了简单易用

的数据清洗函数, 帮助学生快速有效地对数据进行预处理; 而 `ggplot2` 包则让数据可视化变得异常简单, 同时能够生成专业级别的图表, 提升教学质量。

对于高等教育中的统计课程, R 语言的应用尤为突出, 尤其是在涉及复杂统计模型的场景下。例如, R 语言为教育测量中的项目反应理论(IRT)提供了强大的支持, 诸如 `ltm`、`mirt` 和 `IRTShiny` 等包能够帮助学生处理学生能力、试题难度等复杂问题。在多元统计分析、混合效应模型、因子分析和聚类分析等领域, R 语言也提供了丰富的工具和方法, 帮助学生更好地理解和掌握复杂的统计理论和应用。

3) 开源特性与全球协作

R 语言的开源特性使得它在全球范围内吸引了大量的统计学家和数据科学家的参与与贡献。许多新的统计方法和包的出现, 都是基于全球学术界和研究者的合作和共享。在教学中, 这种开放的资源共享模式使得学生能够接触到最新的统计工具和方法, 并与全球的科研社区保持紧密联系。这不仅拓宽了学生的视野, 也培养了他们的批判性思维和解决问题的能力[13]。

综上所述, R 语言在统计计算和数据分析领域所展现的强大功能、丰富的资源和灵活的定制化能力, 使其成为高等教育中进行统计教学的理想工具。无论是从学习的便捷性、计算的高效性, 还是从全球协作的开源优势来看, R 语言都提供了一个理想的环境, 帮助学生掌握现代数据分析和统计计算的核心技能。

6. 结束语

R 语言作为一种强大的数据处理和统计分析工具, 在高等教育中的应用具有重要意义。其在直观教学中的广泛应用, 不仅提高了教学质量, 还促进了教学方法的创新。随着数据科学的不断发展, 预计 R 语言的应用将进一步扩展到更多的学科领域, 对未来的学术研究和教学实践产生深远的影响。因此, 推动 R 语言在教育中的应用和推广, 对于提升教学效果和培养数据分析人才具有重要意义。

致 谢

成都信息工程大学应用数学学院鄢田云博士/副教授是本论文的指导教师和通讯教师, 感谢鄢老师对本论文涉及的大创项目系统开发、论文撰写的悉心指导和修改建议, 感谢成都信息工程大学教务处对本论文成果来源的大学生创新创业训练计划项目及论文版面费的支持。

基金项目

四川省大学生创新创业训练计划项目(S202410621117), 成都信息工程大学创新创业训练计划项目(202410621326), 四川省大学生创新创业训练计划项目(S202410621115), 成都信息工程大学创新创业训练计划项目(202410621323)。成都信息工程大学 2024 年本科教学工程项目“数智项目”(在线实践平台数据可视化的数智实验教学模式创新)。

参考文献

- [1] 王欢, 汤强. 教育教学中 R 语言运用的优势[J]. 百科知识, 2022(12): 75-76.
- [2] 于平. 基于大数据的深度学习网络爬虫算法在信息搜集与处理中的应用[J]. 科技资讯, 2024, 22(16): 55-57.
- [3] 汪国海. 大数据背景下 R 语言在生物统计学图形可视化中的应用[J]. 数字技术与应用, 2023, 41(1): 10-12.
- [4] 刘展, 刘元亮. 工科课程论文参考文献引用情况分析[J]. 高等建筑教育, 2022, 31(3): 28-33.
- [5] 池召艳. 教育现代化形势下 R 语言教学的探索与实践——以湖北文理学院信息与计算专业为例[J]. 湖北文理学院学报, 2022, 43(2): 85-88.
- [6] 李志强, 刘建平, 王雪峰. 文献计量研究报告指南与清单的制定与展望[J]. 中国医学前沿杂志(电子版), 2024, 16(10): 92-93.

-
- [7] 杨永明, 李霄. 直观可视化教学方法在现代高等教育课堂教学中的应用[J]. 高教学刊, 2022, 8(3): 94-98.
- [8] 郭洪飞, 赵敏, 李荣彪, 等. 基于知识图谱的新质生产力研究文献计量分析[J]. 材料导报, 2024, 38(12): 58-67.
- [9] 夏巍, 范宇辰. 基于 LDA 主题模型的中国学前融合教育研究主题分析及热点特征研判[J]. 渭南师范学院学报, 2024, 39(3): 80-87.
- [10] 袁毅, 孟盈. 基于 BERT-LDA 的国外 LIS 领域学科交叉研究演化分析与前沿主题识别[J]. 数字图书馆论坛, 2024, 20(9): 1-15.
- [11] 陈建垒, 王纯. 政策关注、测量与经济效益——基于区域协同发展政策文本的 LDA 建模[J]. 统计研究, 2024, 41(9): 32-43.
- [12] 高天书. 常用金融建模软件: 施用意义、使用难点与适用场域——以 SPSS、MATLAB、R 语言为例[J]. 金融理论与教学, 2020(4): 48-53.
- [13] 陆婧, 张兆远. R 语言与 Python 在高等教育统计计算课程教学中的应用对比研究[J]. 高教学刊, 2024, 10(8): 58-62.