

概率论知识图谱构建的思考与实践

曹黎侠*, 蒋梦想, 王楠

西安工业大学理学院 陕西 西安 中国

摘要 概率论是统计学、人工智能 (AI) 等学科的基础, 由于传统的知识载体多以线性文本形式呈现, 知识之间的关联关系被隐藏, 因而无法直观地展现知识体系的全貌。为满足 AI 等领域的研究者学习与应用概率论体系结构的需求, 概率论知识图谱的构建具有重要的现实意义。鉴于当前概率论知识图谱的构建, 尚缺乏深层逻辑关联挖掘, 本文围绕概率论知识图谱的构建展开研究, 通过分析概率论知识体系的结构特征, 提出一套完整的知识图谱构建方法, 涵盖知识抽取、表示、融合与存储等关键步骤, 并结合具体示例验证方法的有效性。最后探讨概率论知识图谱在自主学习与智能教学等领域的应用策略, 为相关领域的知识工程实践提供理论参考与技术支撑。

关键词 概率论; 知识图谱; 知识表示; 知识融合

文章编号 056-2026-3850

Thinking and Practice on the Construction of Probability Theory Knowledge Graph

Lixia Cao*, Mengxiang Jiang, Nan Wang

School of Science, Xi'an University of Technology, Shaanxi 710021, China

Abstract Probability theory serves as the foundation for disciplines such as statistics and artificial intelligence (AI). Due to the fact that traditional knowledge carriers are mostly presented in linear text form, the interrelationships between knowledge are hidden, making it impossible to visually present the entire knowledge system. To meet the needs of researchers in fields like AI for learning and applying the probabilistic theory framework, the construction of a probabilistic theory knowledge graph holds significant practical significance. Given the current lack of deep logical relationship mining in the construction of probabilistic theory knowledge graphs, this paper conducts research on the construction of probabilistic theory knowledge graphs. By analyzing the structural characteristics of the probabilistic theory knowledge system, a complete knowledge graph construction method is proposed, covering key steps such as knowledge extraction, representation, fusion, and storage. The method's effectiveness is verified through specific examples. Finally, the application strategies of probabilistic theory knowledge graphs in autonomous learning and intelligent teaching are discussed, providing theoretical references and technical support for knowledge engineering practices in related fields.

收稿日期: 2026-01-20 录用日期: 2026-04-25

通讯作者: 曹黎侠; 单位: 西安工业大学理学院 陕西 西安

基金项目: 西安工业大学教学改革研究揭榜挂帅项目 (编号: 25JBGS05): 交叉融合视角下数学专业知识图谱的构建及应用策略研究

Keywords: Probability Theory; Knowledge Graph; Knowledge Representation; Knowledge Fusion

1 引言

1.1 研究背景

在信息爆炸的时代，海量的非结构化与半结构化数据中蕴含着丰富的知识，但传统的知识组织方式难以满足智能化、精准化的知识服务需求。知识图谱（Knowledge Graph）的出现为解决这一问题提供了有效途径^[1]。自2012年以来，KG逐渐在向教育、医疗、金融等垂直领域渗透。

概率论的理论与方法是统计学、机器学习、人工智能、金融工程等学科的基础。然而，概率论的知识体系存在概念抽象、定理逻辑链条长、知识点交叉关联紧密等特点，传统的概率论知识载体如教材、学术论文、百科词条等多以线性文本形式呈现，知识之间的关联关系被隐藏，无法直观地展现知识体系的全貌^[2]。因此，构建概率论知识图谱，将分散的概率论知识要素结构化、网络化，对于促进概率论知识的传播、学习与应用具有重要的现实意义。

从技术层面来看，自然语言处理技术的快速发展为KG的自动化构建提供了支撑。命名实体识别、关系抽取、实体链接等技术的成熟，使得从大规模文本数据中抽取概率论相关知识成为可能。在此背景下，探索一套针对性的知识图谱构建方法，具有重要的理论价值与实践意义。

1.2 文献综述

国内外学者对数学领域知识图谱的研究已取得一定成果。在国外，Chen et al. (2020)^[3]利用本体论方法构建了高等数学知识图谱，实现了知识点的智能关联与习题推荐；Zhang & Li (2022)^[4]将知识图谱与自然语言处理结合，

完成了统计学概念的自动抽取与关系建模。在国内，王磊等(2021)^[5]构建了线性代数知识图谱，验证了其在教学辅助中的有效性；刘芳等(2023)^[6]针对概率论与数理统计课程，设计了基于知识图谱的在线学习平台，但该研究仅聚焦于课程知识点的简单关联，未深入探讨概率论知识的逻辑推导关系与应用场景的融合。

总体而言，现有研究多侧重数学知识图谱的教学应用，对概率论知识体系的深层逻辑关联挖掘、动态构建方法及跨领域应用策略的研究仍较为欠缺。因此，本文从构建原理、步骤、应用等维度系统探讨概率论知识图谱的构建，弥补现有研究的不足。

2 概率论知识图谱构建的方法

2.1 构建原理

概率论知识图谱的构建以本体论为理论基础，结合知识抽取、知识融合、知识存储等技术，实现从非结构化/半结构化的概率论知识源中提取核心要素，并建立要素间的关联。其核心原理可概括为3点：

(1) 本体建模

定义概率论知识的核心概念（如“随机变量”“条件概率”）、概念属性（如“离散型随机变量的取值范围”）与概念间关系（如“包含”“推导”“应用于”），形成标准化的知识框架；

(2) 知识表示

采用RDF（资源描述框架）或Neo4j图数据库的三元组（主体-谓词-客体）形式表示知识，例如（条件概率，推导自，古典概型）、（正态分布，属于，连续型随机变量分布）。

(3) 迭代优化

由于概率论知识具有层次性与拓展性，构建过程需结合领域专家反馈与实际应用需求，持续补充知识节点、修正关系模型。

2.2 构建步骤

概率论知识图谱的构建分为知识源获取、知识抽取、知识融合、知识存储、知识可视化五个核心步骤，具体流程如下：

·Step1: 知识源获取

选取权威的概率论知识载体作为数据源，主要包括：

经典教材：《概率论与数理统计》（浙大版）^[7]、《概率论基础》（李贤平版）^[8] 等高校主流教材；

学术文献：IEEE Xplore、CNKI 中关于概率论理论与应用的核心论文；

行业标准：金融、人工智能领域中基于概率论的技术规范与应用案例。

对获取的知识源进行预处理，如文本分词、公式解析、章节结构化拆分，为后续知识抽取做准备。

·Step2: 知识抽取

采用人工抽取 + 机器辅助的方式，从知识源中提取三类核心要素：

实体抽取：识别概率论中的概念（如“贝叶斯公式”）、定理（如“大数定律”）、方法（如“最大似然估计”）、应用场景（如“风险评估”）等实体；

关系抽取：挖掘实体间的逻辑关系，包括“推导关系”（如“二项分布推导自伯努利试验”）、“包含关系”（如“泊松分布属于离散型分布”）、“应用关系”（如“马尔可夫链应用于自然语言处理”）；

属性抽取：提取实体的特征属性，如“正态分布的均值 $\mu \in \mathbb{R}$ ，方差 $\sigma^2 > 0$ ”“古典概型

的样本空间有限且等可能”。

对于公式类知识，采用 LaTeX 解析与符号映射的方式，将数学公式转化为知识图谱可识别的属性信息。

·Step3: 知识融合

对抽取的知识进行去重、消歧与整合，解决多源知识的冲突问题：

实体消歧：统一同一概念的不同表述，如“条件概率公式”与“乘法定理”本质为同一实体，需合并为单一节点；

关系标准化：将自然语言描述的关系（如“由……得到”）统一为预定义的关系类型（如“推导自”）；

属性补全：结合领域专家知识，补充缺失的实体属性，如为“随机过程”添加“时间依赖性”“状态空间”等属性。

·Step4: 知识存储

根据知识的类型与应用需求，选择合适的存储方式：

对于结构化的三元组知识，采用 Neo4j 图数据库存储，支持高效的图查询与关系遍历；

对于公式、文本描述等非结构化知识，采用 MySQL 关系型数据库存储，并通过实体 ID 与图数据库关联；

对于大规模的概率论知识图谱，可采用分布式图数据库（如 JanusGraph）实现海量数据的存储与并行计算。

·Step5: 知识可视化

利用可视化工具（如 Neo4j）将知识图谱以图形化形式展示，通过节点大小、颜色区分实体的重要性与类型，通过边的样式区分关系类型，实现概率论知识体系的直观呈现。

2.3 构建示例

以“概率论与其各章”模块为例，展示知识图谱的构建过程：

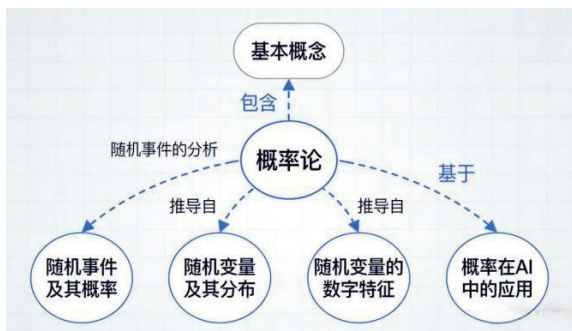


图 1. 概率论首层知识图谱

(1) 实体提取

核心实体包括“概率论”“基本概念”“随机事件及其概率”“随机变量及其分布”“随机变量的数字特征”“概率在 AI 中的应用”；

(2) 关系提取

建立关系，如（概率论，包含，随机事件）、（随机变量，推导自，随机事件）、（数字特征，推导自，随机变量）、（概率在 AI 中的应用，基于，概率的计算）等；

(3) 属性提取

为“概率论的概念”添加属性“不确定性现象”、“统计规律性”，为“随机事件及其分布”添加属性“随机试验的取值及取该值概率的大小”；

(4) 存储与可视化

将上述三元组存入 Neo4j，可视化后可清晰地看到从概率论的基本概念到数字特征的逻辑推导链，以及各概念的核心属性。

该子模块知识图谱（见图 1）能够直观呈现概率论与其各章的内在联系，帮助学习者理解“从基础概型到复杂公式”的推导过程。

3 知识图谱在教育领域的应用策略

3.1 教育教学应用

概率论知识图谱在教育领域的核心价值是实现个性化学习与智能教学，具体应用策

略包括：

(1) 知识点关联导航

在线学习平台中，概率论知识点抽象、逻辑性强、前后依赖紧密，传统线性呈现方式易造成知识碎片化、逻辑不连贯、应用不落地。如果将概率论知识图谱嵌入学习平台，以知识点为节点、以逻辑关系为边，实现点击概念 --- 查看前置知识 --- 查看后续知识 --- 直达应用案例的关联导航，可帮助学习者构建系统化的知识网络，提升学习效率与理解深度，解决学习者“不知道该先学什么、不知道能延伸什么、不知道如何用”的痛点。

在此，以“概率论知识图谱关联导航”为例，说明该功能实现的步骤：

·Step1 触发点击，学习者在学习页面点击任意高亮概念，如“条件概率”，系统调用图谱接口；

·Step2 中心展示，界面中央显示当前知识点的定义、公式、讲解要点；

·Step3 关联展示，左侧自动呈现前置知识，如随机事件、概率定义、古典概型；右侧呈现后续知识，如全概率公式、贝叶斯公式、随机变量独立性；底部显示应用案例，如疾病筛查、信用风险评估、机器故障预测；

·Step4 跳转导航，点击任意关联知识点，中心自动切换并刷新其关联链；点击应用案例，直接进入场景解析页；点击习题可进入针对性训练，实现“点即通、点即学、点即用”。

知识图谱关联导航实现的示意图见图 2。按照上述步骤，点击“条件概率”，其知识点关联导航界面示意图见图 3。

该功能实现能够把离散知识点整合成可导航、可拓展、可应用的知识网络，让在线学习从“被动接收”变为“主动建构”，进一步提升个性化与智能化水平，为数理类课程在线学

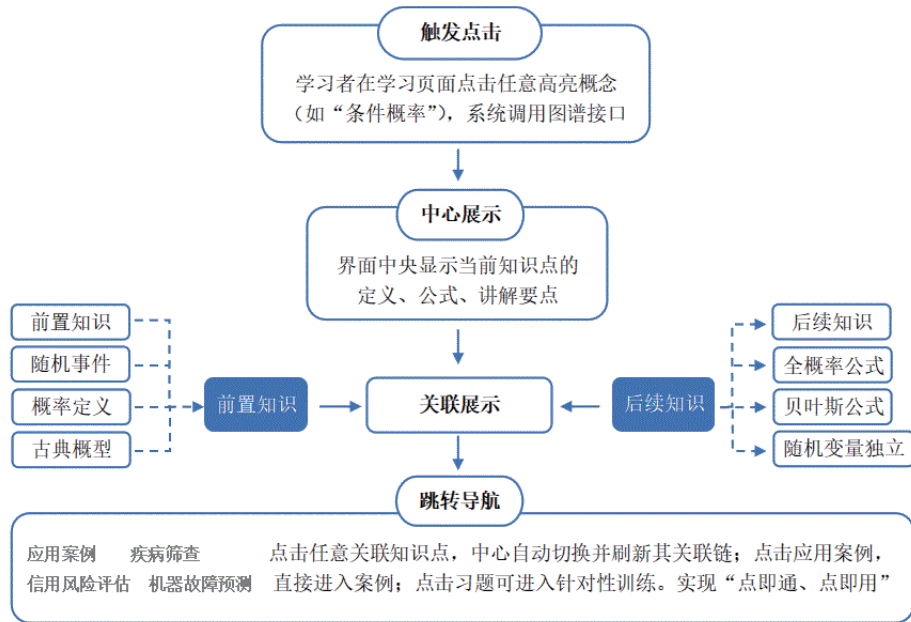


图 2. 知识图谱关联导航实现的示意图



图 3. 知识点关联导航界面示意图

习提供可复制的范式。

(2) 智能习题推荐

基于知识图谱的节点关联，根据学习者的薄弱知识点，推荐对应的基础、综合与应用习题，实现针对性的学习训练。具体操作的基本原理为：以知识点节点为核心、以关联关系为依据，能够精准定位学习者薄弱点，自动推送基础题、综合题、应用题三级习题，实现“缺什么、推什么、练什么”的个性化训练模式，有效提升学习效率与知识掌握程度。

与传统推荐相比，基于知识图谱的推荐具

备三大优势，一是可解释，知道推题原因；二是可溯源，能定位到最底层薄弱点；三是可递进，从基础到综合再到应用形成训练闭环。

智能习题推荐流程共分为 5 个步骤，自上而下依次执行：

Step1 答题行为采集，学习者完成习题或测验，平台自动记录对错、耗时、错误步骤，形成原始学习数据；

Step2 薄弱知识点定位，系统通过知识图谱匹配错误题目对应的知识点节点，判断薄弱点；

Step3 知识关联推理，前置基础知识点、递进综合知识点、应用场景知识点；

Step4 三级习题智能推荐，根据薄弱点自动推送三类习题：基础题，巩固薄弱知识点本身；综合题，关联前后知识点的串联题型；应用题，结合真实场景的实践题型；

Step5 效果评估与循环优化，完成训练后重新评估掌握程度，若仍薄弱则继续推荐，



图 4. 智能习题推荐示意图



图 5. 基于知识图谱教学内容设计流程示意图

若已掌握则进入下一知识点, 形成持续提升的闭环。

其流程示意图见图 4。该模式不仅提升了习题训练的有效性, 也推动在线学习向个性化、精准化、智能化发展, 可为数学、统计学、工科基础课等知识点密集型课程提供可复制的应用范式。

(3) 教学内容设计

在概率论与数理统计的教学中, 知识体系抽象、逻辑链长、模块跨度大, 传统线性备课容易出现知识点割裂、重点不突出、理论与应用脱节等问题。如果将概率论知识图谱引入教学内容设计, 能够帮助教师以可视化方式整体把握知识结构, 清晰呈现知识点之间的依赖、

递进、综合与应用关系，从而科学调整教学顺序、优化教学重点、强化实践应用，让教学更具系统性、逻辑性与实用性。基于知识图谱的教学内容设计流程见图 5。

3.2 科研创新应用

在概率论相关的科研中，知识图谱可助力理论挖掘与跨领域研究，具体策略包括先是文献关联分析，接着理论创新推导，随后科研趋势预测。各步的内涵及流程如下：

(1) 文献关联分析

将科研文献中的理论方法、实验数据与知识图谱的实体关联，可通过图谱快速检索某一概率方法的相关研究成果、应用领域与待解决问题。基于概率论知识图谱的文献关联分析应用策略流程见图 6。

通过构建完善的概率论科研知识图谱，科研人员无需逐篇研读海量文献，仅通过检索核

心实体，即可快速梳理某一概率方法的发展脉络、历代研究成果、主流应用领域，同时精准定位该理论当前尚未解决的科研问题、存在的技术瓶颈与研究空白，极大提升文献调研与前期研究的效率。

(2) 理论创新推导

基于知识图谱的逻辑关系，挖掘潜在的理论关联，如通过“马尔可夫链”与“深度学习”节点关联，探索概率模型在神经网络中的新应用。

基于知识图谱构建的结构化逻辑关系网络，能够将概率论各类基础理论、算法模型与其他学科的技术框架、研究方法进行实体关联与语义匹配，打破人工科研难以捕捉跨领域关联的局限。

依托知识图谱可精准建立“马尔可夫链”随机演化理论与“深度学习”技术体系的节点

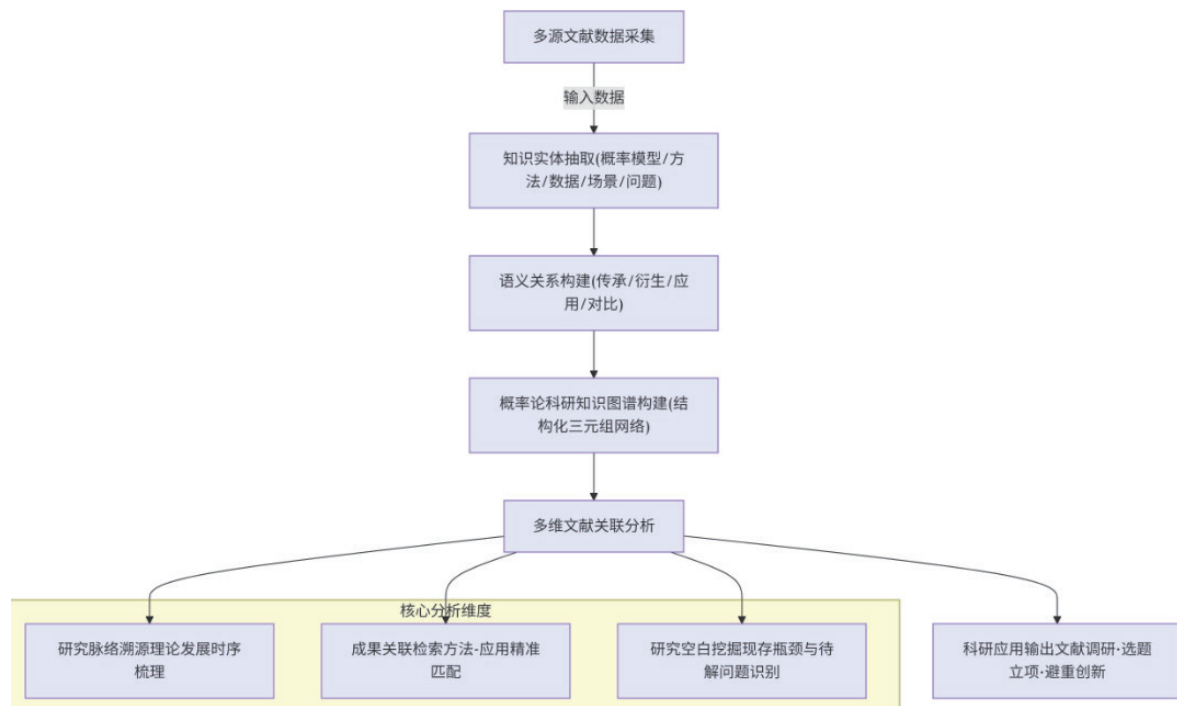


图 6 文献关联分析应用策略示意图

关联,系统梳理二者的适配逻辑与现有融合研究基础。基于这一结构化关联关系,科研人员可针对性探索各类概率模型在神经网络训练、模型构建、性能优化中的全新应用方式,弥补传统确定性神经网络算法稳定性弱、不确定性量化能力不足的缺陷,有效推动概率论理论与人工智能技术的深度融合,催生跨学科原创性理论成果。

(3) 科研趋势预测

结合知识图谱的实体热度与关系新增量,分析概率论在人工智能、生物医药等领域的研究趋势,为科研选题提供参考。

依托知识图谱具备的动态量化统计与时序分析能力,可实现概率论领域科研趋势的科学化、精准化预测,有效弥补传统人工经验研判主观性强、滞后性明显的短板。

知识图谱可实时监测领域内实体热度、语义关系新增量、跨学科关联频次等核心数据,剖析概率论在人工智能、生物医药、工程建模、金融风控等重点领域的渗透态势与发展潜力,精准识别当下热点研究方向、衰退研究领域及前沿空白赛道。通过研判概率模型在智能算法不确定性推理、生物医药数据统计分析等场景的应用增速与研究热度变化,能够清晰预判学科未来的核心发展方向。

4 结论

概率论知识图谱的构建需以本体论为基础,结合人工与机器的知识抽取方式,通过知识融合解决多源数据的冲突问题,再采用图数据库实现高效存储与可视化;

该知识图谱在教育、科研等领域具有重要应用价值,可通过知识点导航、科研文献分析、模型构建等策略,实现概率论知识的系统化利用与智能化应用。

知识点导航、科研文献分析、科研趋势的预测等策略,为科研人员将概率论知识图谱应用于创新性教学和科技创新提供了可视化的操作流程,有助于提升 AI 时代的创新性教学和交叉研究的效率。

概率论知识图谱的构建是对传统概率论知识体系的数字化重构,不仅能够降低知识学习与研究的门槛,还能为跨领域的概率应用提供结构化的知识支撑,随着人工智能技术的发展,其应用场景与价值将进一步拓展。

参考文献

- [1] 唐好选,郑世博,于游等. 科研数据知识图谱赋能学科交叉研究数智赋能[J]. 中国高校科技. 2026(51):48-55.
- [2] 池红梅. 知识图谱赋能概率论与数理统计课程教学的实践与思考[J]. 中国多媒体与网络教学学报, 2025(2):81-83.
- [3] Chen Y, Liu J, Wang H. Construction of Higher Mathematics Knowledge Graph Based on Ontology[J]. Journal of Intelligent & Fuzzy Systems, 2020, 39(5): 6891-6902.
- [4] Zhang L, Li M. Knowledge Extraction of Statistical Concepts Based on Natural Language Processing[J]. IEEE Access, 2022, 10: 112345-112357.
- [5] 王磊,张萌,李娟. 线性代数知识图谱的构建与教学应用研究[J]. 高等数学研究, 2021, 24(6): 102-105.
- [6] 刘芳,赵阳,陈明. 基于知识图谱的概率论与数理统计在线学习平台设计[J]. 计算机工程与应用, 2023, 59(8): 287-294.
- [7] 盛骤,谢式千,潘承毅. 概率论与数理统计(第7版)[M]. 北京:高等教育出版社,2020.
- [8] 李贤平. 概率论基础(第3版)[M]. 北京:高等教育出版社,2010.