

分类号：
学号：20222108022

密级：公开
单位代码：10759

石河子大学

硕士学位论文



基于知识图谱双曲嵌入的运动损伤 康复辅助研究及应用

学位申请人	林龙鑫
指导教师	秦怀斌 教授 齐全 副教授
申请学位类别	专业硕士
专业名称	电子信息
研究领域	计算机技术
所在学院	信息科学与技术学院

中国·新疆·石河子

2025年6月

分类号：
学号：20222108022

密级：公开
单位代码：10759

石河子大学

硕士学位论文



基于知识图谱双曲嵌入的运动损伤 康复辅助研究及应用

学位申请人	林龙鑫
指导教师	秦怀斌 教授 齐全 副教授
申请学位类别	专业硕士
专业名称	电子信息
研究领域	计算机技术
所在学院	信息科学与技术学院

中国·新疆·石河子

2025 年 6 月

**Research and Application of Sports Injury Rehabilitation Aid
based on Hyperbolic Embedding of Knowledge Graph**

A Dissertation Submitted to

Shihezi University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Engineering

By

Lin Long-Xin

(Computer Technology)

Dissertation Supervisor: Prof. Qin Huai-bin & A/Prof. Qi Quan

June, 2025

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：林龙鑫

时间：2025年5月26日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：林龙鑫

时间：2025年5月26日

导师签名：秦松斌

时间：2025年5月26日

摘要

随着运动参与人数的增加，运动损伤的发生率逐年上升，如何有效地进行康复治疗已经成为一个亟待解决的课题。传统的康复方法往往依赖于经验，缺乏系统性和科学依据，导致康复效果不理想。本研究致力于解决运动损伤康复领域中的信息整合与决策支持问题，通过构建运动损伤知识图谱和应用双曲几何嵌入模型，开发一套基于知识图谱的智能康复辅助系统，旨在提升康复治疗的科学性和个性化水平，为运动损伤患者提供更为精准和高效的康复方案。

(1) 本研究通过对运动损伤相关文献的广泛调研，构建运动损伤本体，明确运动损伤的类型、病因、症状及其对应的康复措施。这一过程主要为知识图谱的构建提供理论基础，并为后续的数据处理和模型设计奠定基础。通过对运动损伤知识的系统化整理，研究实现对运动损伤与康复治疗之间复杂关系的有效识别和表示。

(2) 提出一种基于双曲几何的嵌入算法，优化现有知识图谱嵌入方法在高维数据处理中的局限性。双曲几何嵌入模型能够在低维空间中有效捕捉到运动损伤与康复措施之间的潜在关联，进而实现链路预测。该模型的核心优势在于其能有效处理复杂的非欧几里得数据结构，适应运动损伤领域中多样化和动态变化的特征。通过多种对比实验，验证了该模型在链路预测任务中的优越性。实验结果显示，本文的模型在自建数据集上进行低维度嵌入的 MRR 指标为 0.503、Hits@1 指标为 0.459、Hits@3 指标为 0.520、Hits@10 指标为 0.591，与若干基线模型对比均为最佳。

(3) 设计并实现运动损伤康复辅助系统，系统集成知识图谱和双曲几何嵌入模型，能够对运动损伤数据进行管理、推理与辅助决策。系统的设计充分考虑用户需求，确保其具备实际操作价值。通过该系统，康复治疗师可以快速获取针对性的康复建议，帮助患者制定个性化的康复方案，从而提高康复效率，缩短康复周期，降低二次受伤的风险。

本研究的创新之处在于将双曲几何嵌入技术与知识图谱相结合，为运动康复领域开辟新的研究方向。通过系统化的知识管理和智能化的决策支持，研究不仅提升了运动损伤康复的科学性和有效性，也为未来的研究提供新的思路和方法。未来的研究可进一步优化模型，扩展知识图谱的应用范围，并结合深度学习等先进技术，推动智能医疗的发展。

关键词：知识图谱；双曲几何；庞加莱球；链接预测

Abstract

With the increasing number of participants in sports, the incidence of sports injuries has been rising year by year. How to effectively carry out rehabilitation treatment has become an urgent issue. Traditional rehabilitation methods often rely on experience, lacking systematic and scientific basis, leading to unsatisfactory rehabilitation outcomes. This study aims to address the information integration and decision support issues in the field of sports injury rehabilitation. By constructing a sports injury knowledge graph and applying a hyperbolic geometry embedding model, an intelligent rehabilitation assistance system based on the knowledge graph is proposed. The goal is to enhance the scientific and personalized level of rehabilitation treatment, with the aim of providing more precise and efficient rehabilitation plans for sports injury patients.

(1) The study conducted a comprehensive review of literature on sports injuries to construct a sports injury ontology, defining injury types, causes, symptoms, and corresponding rehabilitation measures. This process not only establishes a theoretical foundation for building the knowledge graph but also lays a solid groundwork for subsequent data processing and model design. Through systematic organization of knowledge on sports injuries, the study effectively identifies and represents the complex relationships between injuries and rehabilitation treatments.

(2) The study innovatively proposes an embedding algorithm based on hyperbolic geometry to optimize the limitations of existing knowledge graph embedding methods in high-dimensional data processing. Hyperbolic geometric embedding model can effectively capture the potential correlation between sports injuries and rehabilitation measures in low-dimensional space, and then realize link prediction. The core advantage of this model is its ability to process complex non-Euclidean data structures and adapt to diverse and dynamic characteristics in the field of sports injuries. The superiority of this model in the task of link prediction is verified by a variety of comparative experiments. Experimental results show that the MRR index of low-dimensional embedding in the proposed model on the self-built dataset is 0.503, Hits@1 is 0.459, Hits@3 is 0.520, and Hits@10 is 0.591, which is the best compared with several baseline models.

(3) The study designs and implements a sports injury rehabilitation support system that integrates the knowledge graph and hyperbolic embedding model. This system enables data management, reasoning, and decision support for sports injury rehabilitation. The system's design prioritizes user needs to ensure practical applicability. Through this system, rehabilitation therapists can quickly access targeted rehabilitation recommendations, helping patients develop personalized rehabilitation plans, improving rehabilitation efficiency, shortening recovery times, and reducing the risk of secondary injuries.

The innovation of this study lies in the integration of hyperbolic geometry embedding techniques with

knowledge graphs, opening new research avenues in the field of sports rehabilitation. By combining systematic knowledge management with intelligent decision support, the study enhances the scientific and effective aspects of sports injury rehabilitation while providing new perspectives and methodologies for future research. Future work may focus on further optimizing the model, expanding the application scope of the knowledge graph, and integrating advanced technologies such as deep learning to drive the development of intelligent healthcare.

Key words: Knowledge graph; Hyperbolic geometry; Poincare ball; Link prediction

目 录

摘要	I
Abstract	II
第 1 章 绪论	1
1.1 研究背景与意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 国内外研究现状	3
1.2.1 国内运动损伤康复医学发展历程	3
1.2.2 医学知识图谱构建的研究现状	5
1.2.3 链路预测模型的研究现状	6
1.2.4 文献评述	9
1.3 研究内容与技术路线	9
1.3.1 研究内容	9
1.3.2 技术路线	10
1.4 本文组织架构	11
第 2 章 相关技术及理论介绍	12
2.1 本体构建与数据处理技术	12
2.1.1 知识图谱的本体构建	12
2.1.2 数据预处理	13
2.2 知识抽取与图谱构建	13
2.2.1 知识抽取模型	13
2.2.2 知识图谱构建与可视化	14
2.2.3 知识图谱双曲嵌入模型	15
2.2.4 图谱嵌入模型评估指标	16
2.3 先验知识	16
2.3.1 基于知识图谱嵌入的链接预测	17
2.3.2 庞加莱球模型	17
2.3.3 LTE-KGE 框架	18
2.4 本章小结	19

第 3 章	基于双曲几何的嵌入模型 MrpHKGE 研究	20
3.1	引言	20
3.2	双曲嵌入模型	21
3.2.1	双曲空间中的嵌入	21
3.2.2	双曲空间中的几何变换	21
3.2.3	双曲注意力机制	22
3.2.4	得分函数	23
3.3	实验设计	24
3.3.1	实验环境介绍	24
3.3.2	实验设置	24
3.4	实验结果分析	25
3.4.1	低维度下的模型表现	25
3.4.2	高维度下的模型表现	27
3.5	本章小结	29
第 4 章	运动损伤图谱构建和链接预测模型研究	30
4.1	引言	30
4.2	数据获取与预处理	31
4.2.1	数据获取与本体构建	31
4.2.2	数据处理	32
4.2.3	实体对齐	32
4.3	运动损伤知识图谱构建	33
4.4	运动损伤链接预测	34
4.4.1	链接预测模型的设计与实现	34
4.4.2	LTE-MrpHKGE 模型构建	35
4.4.3	低维度下的模型表现	36
4.5	面向运动损伤知识图谱的康复辅助模型	38
4.5.1	康复方案生成	38
4.5.2	损伤机制分析	40
4.5.3	运动损伤康复辅助模型结果示例与有效性验证	41
4.6	本章小结	43
第 5 章	运动损伤康复辅助系统设计与实现	44
5.1	需求分析	44
5.1.1	功能需求分析	44

5.1.2 非功能需求分析	45
5.2 系统概要设计	46
5.2.1 系统总体框架	46
5.2.2 系统模块设计	47
5.3 系统详细设计	48
5.4 系统实现	51
5.4.1 系统开发工具及软硬件环境	51
5.4.2 系统界面展示	51
5.5 系统测试	55
5.6 本章小结	56
第6章 结论与展望	57
6.1 结论	57
6.2 展望	58
参考文献	59
致谢	64
作者简介	65

第1章 绪论

1.1 研究背景与意义

1.1.1 研究背景

随着社会经济的发展和人们生活水平的提高,体育运动和健身活动在中国的普及程度日益增高。然而,随之而来的运动损伤问题也日益突出,尤其在运动员群体中更为常见。所有专业或非专业运动员在运动的过程中都承受着各种软组织、骨、韧带、跟腱和神经损伤的风险,这些损伤往往会因创伤或者重复性应力作用引起^[1-3]。在2022年北京冬奥会期间,据国家奥委会和北京的医务人员报告显示,来自91个国家的2,897名专业运动员就有326人受伤,运动损伤的发生率高达11.3%,其中滑冰运动的损伤率最高受伤人数占受伤运动员的31.9%,运动损伤的部位主要集中在四肢,常见的损伤部位包括膝盖(占总损伤的22.1%)、腰椎/下背部(占10.4%)和腿部(占9.2%)^[4]。这些损伤可能涉及肌肉、骨骼、关节、韧带和肌腱等多种组织。运动员在高强度的运动过程中,肌肉和关节承受着巨大的压力,一旦发生损伤,不仅会影响训练和比赛,更重要的是可能对未来的职业生涯造成长期影响^[5]。因此,如何快速有效地对运动损伤进行治疗和康复,成为运动医学的重要研究课题。

目前的运动损伤康复方式往往依赖于医生和物理治疗师的经验,对运动员的个体化差异考虑不足,康复过程较为漫长且效果不稳定。运动损伤康复过程中,通常采用的方法包括物理治疗、药物治疗和手术治疗。然而,由于个体差异和损伤类型的多样性,这些方法的效果往往难以预测,且康复周期较长,容易出现旧伤复发或并发症。例如,肌肉拉伤的康复通常需要数周甚至数月时间,关节扭伤的康复可能需要更长的时间进行关节稳定性和功能恢复。而骨折和严重的韧带损伤,则可能需要手术治疗和长时间的康复训练^[6]。这些传统方法在面对复杂的运动损伤时,存在治疗效果不确定、康复周期长、复发风险高等问题。如图1-1所示为一般运动损伤康复处方的制定流程。

在现代科技迅猛发展的背景下,如何利用先进的信息技术手段,提高运动损伤康复的科学性和个性化水平,成为一个值得深入研究的方向。知识图谱(Knowledge Graph)作为一种新兴的数据组织和知识管理技术,通过整合多源异构数据,构建全面的医学知识体系,在运动损伤的康复辅助中展现出广阔的应用前景。知识图谱是一种以图形结构表示知识的技术,通过节点和边的关系展示实体之间的关联。它不仅能有效地组织和管

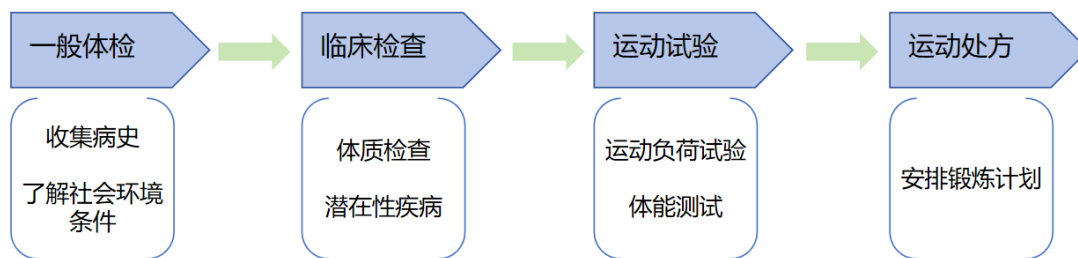


图 1-1 康复处方制定流程

Fig. 1-1 Rehabilitation prescription formulation process

理海量数据，还能揭示数据背后的复杂关系，从而为数据分析和应用提供强大的支持。在医疗领域，知识图谱已得到广泛应用，例如构建狭窄医学领域的专家系统、辅助疾病的预测和诊断、辅助进行临床决策等。目前常见的公共知识图谱如表 1-1 所示：

表 1-1 常见的公共知识图谱

Tab. 1-1 Common public knowledge graph

知识图谱	数据来源	贡献机构	图谱特点
WN18RR	WordNet	斯坦福大学	用于知识图谱补全任务的数据集
FB15K	Freebase	Facebook 人工智能研究院	用于关系抽取和实体链接任务的常用基准数据集
UMLS	MeSH、SNOMED CT、LOINC	美国国立医学图书馆	制定了统一医学语言系统这一卫生信息标准
CMeKG	ICD、ATC、SNOMED、MeSH	北京大学计算语言学研究所	基于大规模医学文本数据，以人机结合的方式研发的中文医学知识图谱

1.1.2 研究意义

知识图谱在医学领域已经有着诸多研究与应用，这也推动了研究人员构建更多更细致的医学知识图谱，构建细致的医学知识图谱是各医学领域智慧医疗应用的迫切需求。而构建好的知识图谱只能算是给 KOA 领域中的知识进行了标准化的存储，要想将其应用到各种下游任务中，还需要知识图谱嵌入 (Knowledge Graph Embedding, KGE) 技术的支撑，本文所使用的双曲嵌入 (Hyperbolic Embeddings) 就是近年来先进的一项知识图谱嵌入技术，相比于基于欧氏空间的嵌入模型在处理复杂、层次化的知识图谱时往往难以捕捉深层次的结构信息，双曲嵌入技术利用双曲空间的几何特性，能够更有效地表示具有层次结构的知识图谱，因此本文着重使用双曲几何的相关方法进行知识图谱的嵌入，并将该嵌入模型用于运动损伤知识图谱的链路预测任务。链路预测是一种基于图结构的

数据挖掘技术,旨在通过已有的图结构信息和节点属性预测出可能存在的但尚未被观测到的关系,从而改善知识图谱的完整性,挖掘出有用的信息。它既可以做到对隐藏关系的可能性预测,也可以实现对未来关系的推测。链路预测任务研究不仅在理论上具有重要意义,也在实际应用中具有重大的应用前景,特别在康复医学这种知识密集型领域,该技术有望为挖掘隐含的运动损伤与康复治疗之间的关系带来巨大的帮助和指导。

综上所述,本文将以运动损伤相关医书、典籍,通过完整的知识图谱构建流程,建立一套完整的运动损伤知识图谱,有助于深入挖掘康复医学领域中的复杂联系,辅助于康复处方决策过程。利用知识图谱链路预测技术提高知识图谱的完整性。通过知识图谱和链路预测技术结合,可以实现快速传播和共享最新的运动损伤康复医学研究成果和临床经验,帮助医生和康复师提升专业水平,提高整体医疗服务质量。

1.2 国内外研究现状

许多运动损伤患者在经历临床治疗后因为不够重视而不积极参与康复治疗,导致后期运动功能得不到恢复,甚至有的产生严重的功能障碍,影响生活质量,由此可见康复治疗在受伤患者的整个治疗过程中起着非常大的作用。康复处方是通过康复治疗师根据患者的实际病情、手术情况和康复需求,制定出的一系列康复计划和指导。康复处方的目的是帮助患者恢复或改善功能,提高生活质量,并促进康复过程的有效进行。同时,由于我国康复医学领域起步较晚,相对于这方面起步早的国家在康复工作模式上不够完善,在多数骨科病房中,仍普遍存在“重手术、轻康复”的现象,且康复工作模式尚未广泛采用欧美等发达国家普遍实施的多学科协作(Multi-disciplinary Team, MDT)模式^[7],近年来,利用科学技术手段辅助康复医学已成为提升患者康复效果、加速康复进程的重要途径^[8-10]。通过结合人工智能、物联网、大数据分析等先进技术,运动损伤康复医学正在从传统的经验导向转向数据驱动和个性化定制^[11,12]。

1.2.1 国内运动损伤康复医学发展历程

2004年,北京体育大学首次将运动康复的基础理论与技术引入我国,用于专业运动员肌肉骨骼系统疾病的防治。运动康复以运动疗法为核心,通过个性化运动处方进行系统化干预,结合主动肌筋膜释放技术、关节动态松动术等先进方法,并辅以声、光、电、磁、针灸等物理因子治疗,逐步形成了一套科学的多维度干预体系。这一技术的引入,为我国运动损伤康复医学的发展奠定了理论和实践基础。

我国运动康复行业发展起步较晚,2008年北京奥运会的成功举办直接促进了中国体育事业进入了新的阶段,同时也开启了运动康复医学发展的序幕。奥运会期间,运动

损伤康复的意识和需求逐渐显现，许多运动康复服务由国外专家提供。这一时期，虽然国内运动康复领域尚未成熟，但国际先进的康复理念和技术开始传入，为后续发展奠定了基础。

自 2012 年起，中国运动康复医学逐步进入起步阶段。国内开始出现一些专业的运动损伤康复机构，开始为运动员和体育爱好者提供专业的康复服务。特别是在 2014 至 2016 年期间，运动康复领域的企业集中涌现，市场需求显著增加。随着专业机构和企业的涌现，运动康复服务的质量和覆盖面也得到了提升，推动了运动康复医学的进一步发展。

2018 年，中国的运动康复机构数量突破 100 家，标志着运动康复医学进入了产业化推进阶段。这一时期，运动康复从职业运动员的小众市场逐步向广大体育爱好者的广大市场扩展。2019 年，运动康复产业联盟成立，为行业规范和资源整合提供了平台，进一步推动了产业的健康发展。在此阶段，运动康复服务的供给和需求均显著提升。现代康复技术，如生物力学分析、运动训练监控、功能性电刺激等，在国内得到了广泛应用，提升了康复效果。同时，康复服务的对象也从专业运动员扩展到普通体育爱好者，覆盖面和影响力大幅增加。

自 2020 年起，数字化与智能化技术的应用开始成为我国运动康复医学发展的重要推动力。特别是新冠疫情期间，传统的线下康复服务受限，许多运动康复机构纷纷转向线上平台，提供远程医疗和在线咨询。这一时期，智能穿戴设备、运动监控 APP、虚拟现实^[13]等数字化工具的广泛应用，使运动康复逐渐进入精准化和个性化的新阶段。远程医疗支持和虚拟康复训练的出现，突破了传统线下治疗的空间限制，帮助更多患者实现康复目标。疫情催生的线上服务模式，不仅提升了行业的服务效率，也为运动康复行业的未来发展提供了新的方向，使得数字化和智能化成为推动行业持续发展的核心力量。

随着全民健康意识的不断提升以及运动参与水平的显著增加，公众对肌肉骨骼系统疾病的运动康复需求持续增长。为满足这一需求，截至 2022 年，全国已有 53 所高校开设运动康复专业，平均每年每所高校培养毕业生约 100 人。同时，国家级和重点区域医疗中心也陆续开设了专业化的运动康复服务体系。尽管我国在运动康复专业人才培养和服务体系建设方面取得了显著进展，但相较于庞大的肌肉骨骼疾病患者群体，供需之间的矛盾依然突出，如何通过创新和变革推动运动康复服务供给端的发展，已成为亟待解决的核心问题。

1.2.2 医学知识图谱构建的研究现状

随着知识图谱的快速发展,医疗知识图谱同样是一个研究热点^[14]。国外方面,有面向基因工程的 Gene Ontology^[15]、面向药物工程的 DrugBank^[16] 以及面向蛋白质工程的 UniProt^[17] 等。国内方面,有涵盖医疗领域全科的中文医学知识图谱 CMeKG^[18]、中文一体化医学语言系统 CUMLS^[19] 和针对中医药领域的中医药知识图谱^[20],都为后续使用计算的方法解决该领域内固有的问题提供了更多可能性。此外,还有大量用于构建医疗知识图谱的方法研究, Murali 等人^[21] 对近年来基于电子病理的医学知识图谱构建方法做出了总结; Mohamed 等人^[22] 利用现有的药物靶点知识图谱,将药物靶点相互作用预测视为知识图谱补全任务,构建了一个知识图谱嵌入语义模型来预测新的药物靶点相互作用; Han 等人^[23] 通过定义语义模式和语法来识别电子病历中的医疗信息以构建医学知识图谱。

知识抽取是知识图谱构建的关键步骤,知识抽取的效果直接影响知识图谱数据质量,对一系列下游任务的实现效率都有显著影响,知识抽取包括实体识别和关系抽取,实体识别是从文本中提取出命名实体(如人名、药名、疾病名等),关系抽取是从文本中提取出实体之间的关系。Nikola 等人^[24] 通过比较传统和现代机器学习方法在生物医学关系提取中的表现,展示了 Transformer 模型在构建知识图谱中的优越性; Ke 等人^[25] 提出了一种改进的 BiLSTM-CRF,以此进行中文电子病历医学实体识别及知识图谱关系分析; He 等人^[26] 提出了一种使用深度学习技术的自动实体和关系提取方法; Roy 等人^[27] 提出了一种将医学知识融入 BERT 以提取临床关系方法; Qi 等人^[28] 提出了一种面向中医药说明书的知识增强型医学关系提取方法; Harnoune 等人^[29] 提出了一种基于 BERT 的临床知识提取方法,可以用于生物医学知识图谱的构建与分析; Hu 等人^[30] 创新性的采用大语言模型 chatGPT 来进行零样本的临床实体识别,并取得了不错的效果,且随着大模型的发展,近年来越来越多的研究者尝试将大语言模型应用于医学领域的知识抽取任务^[31-33]。

根据知识抽取步骤中提取的实体和关系,构建知识图谱,并存储于流行的数据库中,由于知识图谱是一种典型的基于图的数据结构,使用传统的数据库进行存储比较困难,而图数据库在处理复杂图关系时有着独特的优势,能够实现高效的关系查询,故将图数据库作为本项目构建的知识图谱存储的载体。目前,主流图数据库主要有 Neo4j^[34]、Nebula、GraphDB 等。

医学领域中的文本数据通常包含大量的噪声和无关信息,因此,数据清洗和预处理是医学知识图谱构建中的关键技术之一。数据清洗包括去除无关的符号、格式化不一致的部分以及删除冗余信息等步骤,而数据预处理则涉及分词、去停用词、词性标注等任

务。为了提高后续实体识别和关系抽取的准确性，近年来，许多研究采用自然语言处理（NLP）技术进行文本的清洗和预处理。基于深度学习的序列标注模型 BiLSTM-CRF^[35] 被广泛应用于医学实体识别中，能够有效识别医学术语、疾病名称、药物名称等实体，并显著提高实体识别的精度^[36]；BioBERT^[37] 是专为生物医学文本预训练的 BERT 模型，其通过在大量生物医学领域的文献上进行预训练，显著提高了医学实体识别、关系抽取等任务的性能。

随着医学研究的深入，医学领域知识图谱开始涉及到的多模态数据（如医学影像、基因组数据、临床试验数据等）。如何有效地融合这些多模态数据，构建更加全面、精准的医学知识图谱，已成为当前研究的热点之一。近年来，基于深度学习的多模态学习方法被广泛应用于医学领域。Mei 等人^[38] 提出了一种通过多模态时空融合生成医疗报告方法，通过结合医学影像和电子病历数据，构建了一个集成的医学知识图谱。除了医学影像，基因组数据也是医学知识图谱构建中的重要数据源，Chandak 等人^[39] 通过将基因数据转化为图结构，模型能够更好地捕捉基因与疾病之间的复杂关系，进而提高知识图谱的准确性和完整性实现精确医疗。Liu 等人^[40] 提出了跨模态学习融合多模态内容的嵌入模型，该模型通过将不同模态的数据映射到一个共享的语义空间中，从而有效地解决了不同模态数据之间的语义差异。Park 等人^[41] 提出了一种用于医学表征学习的图文多模态预训练模型，为理解电子健康记录数据的基本模式打开了一扇新的大门。

1.2.3 链路预测模型的研究现状

链路预测（Link Prediction, LP）是预测网络中节点之间未来可能形成的新连接的方法，它利用网络结构和历史连接信息，揭示节点间潜在的关联关系，可以做到填补网络中缺失的连接或预测未来可能形成的连接。链路预测方法主要分为基于图结构的方法、基于节点属性的方法以及基于机器学习和深度学习的方法。基于图结构的方法包括共现模式、路径依赖等，这些方法通过分析节点之间的邻近度、共同邻居数量以及路径长度等特征来预测新的连接。基于节点属性的方法则利用节点的固有属性信息，例如节点的标签、特征向量等来进行预测。近年来，机器学习和深度学习方法的兴起为链路预测提供了新的技术手段，通过训练模型学习图结构特征和节点属性，实现更高效和精准的预测。

链路预测在知识图谱中的应用具有重要的实践价值。知识图谱是以图的形式表示知识结构化数据，其中节点代表实体，边代表实体之间的关系。由于构建知识图谱的数据来源多样且不完全，导致图谱中往往存在大量的缺失关系。链路预测技术可以通过预测这些可能存在但未被明确标识的关系来大幅提高知识图谱的完整性和准确性。此外，链路预测还可以评估和校正已有关系的准确性，通过分析图谱中的关系是否合理，发现