

ZHU Xuesong, QU Enrui, ZHU Yufeng, QUAN Yuan. Construction and Application of a Ternary Relationship Prediction Model for Microgravity Biological Knowledge Graph (in Chinese). *Chinese Journal of Space Science*, 2025, 45(2): 493–505. CSTR:32142.14.cjss.2024-0167. DOI:10.11728/cjss2025.02.2024-0167

微重力生物学知识图谱三元关系 预测模型构建及应用

朱学松 曲恩锐 朱玉锋 全源

(华中农业大学信息学院 武汉 430070)

摘要 空间微重力环境会对航天员的生理和心理健康造成重大影响, 包括骨骼密度下降、肌肉萎缩、心血管功能改变等。这些问题对于实现长期空间居住和深空探测构成重大障碍。为应对这些问题, 研究通过整合微重力生物学知识图谱 (Microgravity Biomedical Knowledge Graph, MBKG) 和药物重定位知识图谱 (Drug Repurposing Knowledge Graph, DRKG), 构建一个全面的知识图谱, 覆盖广泛的疾病、药物和基因实体及实体之间的复杂关系。在此基础上, 研究训练并使用新的三元关系预测模型异质因果元路径图神经网络 (Heterogeneous Causal Meta path Graph Neural Network, HCMGNN) 获取预测结果。结果表明, 与传统的知识图谱二元链路预测相比, 研究提出的三元预测方法在提高基因和药物预测准确率方面具有明显优势。研究结论强调三元关系模型在探测基因-药物-疾病三元关系预测方面的有效性和潜力, 不仅为未来空间探测中航天员的生理和心理健康、药物再利用研究提供了新的方法和研究思路, 而且为药物再利用领域开辟了新的视角。

关键词 空间微重力, 老药新用, 知识图谱, 三元关系预测, 图神经网络

中图分类号 V7

Construction and Application of a Ternary Relationship Prediction Model for Microgravity Biological Knowledge Graph

ZHU Xuesong QU Enrui ZHU Yufeng QUAN Yuan

(College of Information, Huazhong Agricultural University, Wuhan 430070)

Abstract With the advancement of science and technology, the demand for space exploration has become particularly urgent. However, the microgravity environment in space has negative impacts on the physiological and psychological health of astronauts, including decreased bone density, muscle atrophy, and changes in cardiovascular function. These challenges pose significant barriers to the realization of long-term space habitation and deep space exploration. To address these challenges, this study integrates Microgravity Biomedical Knowledge Graphs (MBKG) and Drug Repurposing Knowledge Graphs

2024-11-19 收到原稿, 2025-01-17 收到修定稿

©The Author(s) 2025. This is an open access article under the CC-BY 4.0 License
(<https://creativecommons.org/licenses/by/4.0/>)

(DRKG) to construct a comprehensive knowledge graph that covers a wide range of diseases, drugs, and genes, as well as the complex relationships between entities. Based on this, the study trains and uses a new ternary relationship prediction model, Heterogeneous Causal Meta path Graph Neural Network (HCMGNN), to obtain prediction results. The results show that compared with traditional binary link prediction in knowledge graphs, the ternary prediction method proposed in this study has a significant advantage in improving the accuracy of gene and drug predictions. The study concludes that the ternary relationship model is effective and has the potential to explore the prediction of gene-drug-disease ternary relationships, provide new methods and research ideas for the physiological and psychological health of astronauts in future space exploration and drug repurposing research, and opening up new perspectives in the field of drug repurposing.

Key words Space microgravity, Drug repurposing, Knowledge graph, Ternary relationship prediction, Graph neural network

0 引言

随着地球自然资源的逐渐枯竭及环境问题的日益严峻,人类对空间资源的依赖日益增加^[1].矿产资源的枯竭、能源危机以及环境恶化迫使人类寻求新的生存和发展空间,推动了对外层空间的探测和开发^[2].此外,空间通信技术的迅猛发展,例如卫星通信和深空网络建设,为地球与空间的信息传输提供了前所未有的便捷和速度,进一步证明了空间探测的必要性^[3].

然而,空间环境中的微重力条件对航天员的生理及心理健康带来了不小的挑战.生理上,航天员在长期微重力环境中会经历肌肉萎缩和骨质疏松等现象,这些生理变化不仅影响航天员的身体健康,还可能导致任务执行能力的下降^[4].因此,针对这些生理性疾病的治疗策略显得尤为重要.在空间环境中,长时间的失重和隔离将会对航天员的心理稳定性造成严重威胁,使航天员产生负性心理反应.航天飞行中的负性心理反应包括精力衰减、心境和思维混乱、睡眠问题、节律紊乱和人际关系匮乏等,以上问题可能导致航天员产生焦虑、抑郁情绪,甚至造成飞行后的人格变化^[5].这些影响不仅威胁到航天员的健康,也对航天任务的成功率和安全性构成了潜在的威胁^[6].综上,深入研究微重力环境对航天员的具体影响,对于保障航天员的健康和确保航天任务的顺利进行具有至关重要的意义^[7].

在现代药物研发领域,面临的挑战包括高昂的研发成本、漫长的研发周期和较低的成功率.为应对这些挑战,知识图谱作为一种先进的数据处理技术,通过构建药物与疾病、药物与其他药物之间的关联网

络,能有效地预测药物的作用机制和潜在的副作用^[8].并且知识图谱能够整合异构生物医学数据,提供结构化关系和语义关系,这对于药物再利用和不良药物反应预测等应用非常有帮助^[9].特别是在针对微重力环境下的生理治疗策略方面,知识图谱的应用能够为药物的再利用提供新的视角和思路.

现有研究多集中于建立例如药物与疾病、疾病与基因等二元关系预测模型,但是针对多元实体例如药物、疾病、基因之间复杂相互作用的研究则相对匮乏^[10].本研究计划利用并扩展现有的微重力生物学知识图谱(MBKG),构建更为精确的基因-药物-疾病三元关系预测模型,以提高在微重力环境下的基因和药物预测精确度.项目将整合和分析大量的生物医学数据,并利用最新的图神经网络技术,深入理解基因、药物和疾病之间的复杂相互作用,期望发现可用于治疗的再利用药物.这种方法不仅有助于解决空间环境下的健康问题,还为药物发现和疾病治疗开辟了新的研究路径.通过计算机辅助技术和大数据分析,本研究有望加速微重力环境下的药物研究进程,降低成本和风险,同时提升研究效率和准确性.这不仅能保障航天员的健康,还能为地面上的医疗研究和实践提供参考.

1 材料与方法

研究的整体流程可以分为5个如下关键步骤.

步骤1 知识图谱的构建,收集、整理并融合了MBKG和DRKG两个知识图谱中基因、药物和疾病相关的数据,形成了一个新的微重力相关的知识

图谱.

步骤 2 使用 TransE 模型训练知识图谱, 获取实体的低维嵌入文件, 并通过其结果判断融合图谱的质量.

步骤 3 基于融合后的知识图谱, 构建基因-药物-疾病三元关系.

步骤 4 基于构建的三元关系, 结合 TransE 模型生成的实体低维嵌入文件, 计算与三元关系相关实体的特征向量.

步骤 5 基于三元关系预测模型 HCMGNN 预测新的三元关系.

通过这 5 个步骤, 本研究旨在利用并扩展已有的 MBKG, 构建更为精确的基因-药物-疾病三元关系预测模型, 以提高在微重力环境下的药物预测精确度, 并为未来的研究提供理论基础和数据支持. 具体流程如图 1 所示.

1.1 材料来源

在构建性能优异的知识图谱过程中, 数据的质量与数量起着至关重要的作用. 为创建一个高质量的知识图谱, 需要收集和整合大量多样化的数据, 确保能够全面覆盖广泛的生物学实体及其相互关系. 收集整理并使用两个主要的数据, 即 MBKG 和 DRKG. 通过对这两个知识图谱的收集、整合和加工处理, 逐步构建一个针对于空间微重力条件、功能完善、涵盖面广、可用性高的生物学知识图谱.

1.1.1 微重力生物学知识图谱

Zheng 等^[1] 通过整合 GNBR, DDInter, STITCH, DrugBank, TCMi 等重要数据库中的信息, 创建了一个涵盖基因、化合物、疾病的 MBKG, 旨在促进适用于微重力环境的抗骨质疏松药物的预测, 解决航天员在空间环境中面临的骨质流失等问题. MBKG 包含 134796 个生物学实体、3395273 个三元组.

1.1.2 药物重定位知识图谱

Ioannidis 等^[12] 整合了 DrugBank, Hetionet, GNBR, String, IntAct 和 DGIdb 6 个数据库的信息, 创建了一个涵盖基因、化合物、疾病、生物过程、副作用及症状的全面生物学知识图谱 DRKG. DRKG 是一个综合型生物医药知识图谱, 涉及人类基因、化合物、生物过程、药物副作用、疾病和症状 6 个方面. DRKG 从 6 个公开的大型医药数据库和 2200 万篇医学文献中挖掘数据, 并进行整理和规范化. DRKG 知识图谱包含实体数 97238 个, 分为 13 种实体类型; 三元组数目 5874261 个, 分为 107 种关系类型. DRKG 在药物重定位领域展现出卓越的表现, 成功识别了 41 种可能对抗新冠病毒的药物, 基于 DRKG 的新研究也呈现出更高的可信度和更强的解释性.

这里使用了 DRKG 中的部分数据来构建知识图谱, 包括 39220 个基因实体、24313 个药物实体和 5103 个疾病实体, 以及 93 种不同类型的关系, 总计 4155757 个不同的三元组.

1.2 融合知识图谱

1.2.1 转换实体和关系名称

在构建知识图谱的过程中, 确保实体唯一性是关键步骤. 这要求每个实体在图谱中有且仅有一个表示. 实体的重复或多重命名会在后续的知识图谱嵌入过程中引发问题, 导致同一实体对应多个嵌入向量, 从而影响知识图谱的准确性和效能. 因此, 需要将相同类型的实体以相同的命名方式编码, 研究中化学物质和疾病使用医学主题词表的唯一标识符 (MeSHID), 基因使用 EntrezID 表示, 蛋白质均表示为基因形式. 同时, 将实体格式统一为 DRKG 中的实体格式 (见表 1).

由于 MBKG 中部分数据使用了汉语拼音或自主命名, 这类实体构成的三元组数量少于 20 且无法找

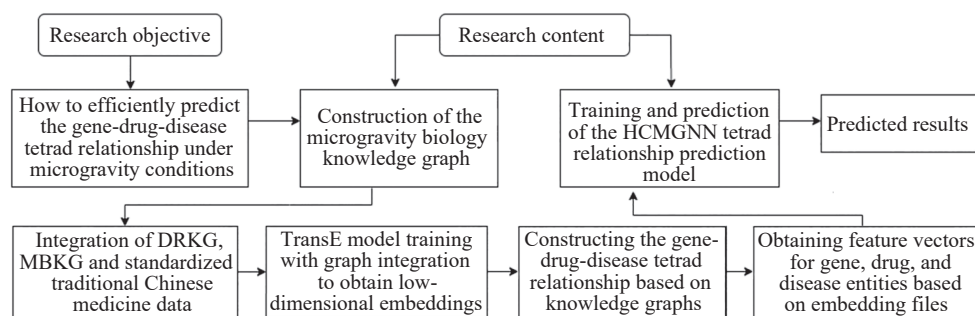


图 1 研究内容流程

Fig. 1 Flowchart of research content

到对应的 MeSH ID 或 Entrez ID, 因此在数据处理时删除了这类数据. MBKG 包含 42 种关系, 并对关系进行了简化, 这里手动将其与 DRKG 的关系进行对齐 (见表 2).

1.2.2 数据的合并与去重

在融合 MBKG 和 DRKG 的过程中, 合并这两个图谱中的标准化三元组是关键步骤. 实体与关系对齐后发现, 来源于两个知识图谱的数据中有 34.15% 相同实体, 78.25% 相同关系, 这说明两个图谱在数据一致性、准确性、覆盖面和补充性方面具有高度重叠和互补性, 能够相互验证并丰富最终的知识图谱.

合并过程中, 遵循细粒度优先于粗粒度的原则, 确保信息的精确性和实用性. 这意味着, 存在多个表示同一实体对关系的三元组时, 应优先保留描述更为详尽 (细粒度) 的三元组, 而移除较为泛泛 (粗粒

度) 的表述. 此外, 关系稀疏性也是一个重要考虑因素. 对于那些关系实例数量极少 (少于 20 个) 的三元组, 应予以删除. 这是因为这些稀疏的关系在后续的嵌入学习中不太可能提供有价值的信息, 而且可能对学习过程的效率和结果的质量产生负面影响. 完成这些步骤后, 最终形成的将是一个大规模、去冗余、内容丰富且精准的 MBKG.

1.3 知识图谱嵌入学习

1.3.1 TransE 模型

知识图谱嵌入是一种关键的技术, 用于将知识图谱中的实体和关系转化为低维的向量表示, 这一过程对于后续的机器学习任务和知识发现至关重要. 在这个转化过程中, 每个实体和关系被编码为向量空间中的点, 这些点的相对位置能够反映实体之间的关系和属性.

研究选择 TransE 模型作为嵌入学习的方法. TransE 模型是一种简洁有效的知识图谱嵌入方法, 基于这样一个假设: 如果一个三元组 (头实体 h , 关系 r , 尾实体 t) 是正确的, 则头实体向量与关系向量的和应该接近尾实体向量, 即 $h+r \approx t$. 如果这个三元组是错误的, 即真正的尾实体应该是 t' 而不是 t , 则有 $h+r \approx t'$. 在这个模型中, 实体和关系被视为向量

表 1 实体格式统一示例
Table 1 Examples of entities formatting uniformity

数据原格式	统一后格式
ENTREZ:174	Gene::174
MESH:C000123	Compound::MESH::C000123
MESH:D011471	Disease::MESH::D011471

表 2 部分 MBKG 关系补全
Table 2 Alignment of MBKG relations

MBKG关系	对齐	缩写
chemical:gene:Z	GNBR::Z::Compound:Gene	Z
gene:disease:Y	GNBR::Y::Gene:Disease	Y
gene:disease:X	GNBR::X::Gene:Disease	X
gene:gene:W	GNBR::W::Gene:Gene	W
gene:gene:V+	GNBR::V+::Gene:Gene	V+
gene:disease:Ud	GNBR::Ud::Gene:Disease	Ud
gene:disease:U	GNBR::U::Gene:Disease	U
gene:disease:Te	GNBR::Te::Gene:Disease	Te
chemical:disease:T	GNBR::T::Compound:Disease	T
chemical:disease:Sa	GNBR::Sa::Compound:Disease	Sa
gene:gene:Rg	GNBR::Rg::Gene:Gene	Rg
gene:gene:Q	GNBR::Q::Gene:Gene	Q
chemical:disease:Pr	GNBR::Pr::Compound:Disease	Pr
chemical:disease:Pa	GNBR::Pa::Compound:Disease	Pa
chemical:gene:O	GNBR::O::Compound:Gene	O
...	...	...

空间中的点和向量, 通过优化这些向量来最小化头实体和尾实体向量之和与实际尾实体向量之间的距离, 即模型训练的目标就是最小化正确三元组的得分和错误三元组的得分之间的差距. 具体来说, 对于一个正确的三元组 (h, r, t) , 其得分为 $s(h, r, t)$; 对于错误的三元组 (h, r, t') , 其得分为 $s(h, r, t')$. 得分的描述有两种, 一种是 L1 范数 $s(h, r, t) = |h + r - t|$, 另一种是 L2 范数 $s(h, r, t) = (h + r - t)^2$. 从数学角度来讲, 模型训练的目标就是最小化损失函数:

$$\delta_{\text{Loss}} = \sum_{(h, r, t) \in K} \sum_{(h', r, t') \in K'} [\gamma + s(h, r, t) - s(h' + r, t')]_+, \quad (1)$$

$$K' = \{(h', r, t) | h' \in E\} \cup \{(h, r, t') \in E\}. \quad (2)$$

其中, K 表示正确的三元组集合, K' 表示错误的三元组集合, $[\]_+$ 表示取正数部分, γ 为常数, E 表示实体空间. 确切地说, TransE 模型通过将实体和关系映射到同一向量空间内, 并利用向量算术来表现三元组的关系, 从而实现知识图谱的嵌入. 在这个模型中, 正确三元组 (头实体+关系 $\approx$ 尾实体) 的向量表示应该在空间上更接近, 而不正确三元组则在向量空间中相隔较远. 这种表示方式使得当头实体向量加上关系向量后, 结果应该尽可能接近尾实体向量, 从而在向量空间中形成一种直观的结构化表示.

这种模型设计允许机器学习算法有效地识别并利用这些向量表示, 进而支持复杂的知识推理和数据分析任务. 通过这样的方法, TransE 能够支持链路预测、实体识别和关系抽取等多种知识图谱相关的下游任务, 为进一步的分析和决策提供数据驱动的见解.

1.3.2 嵌入训练

采用亚马逊开源知识图谱嵌入库 (Deep Graph Library - Knowledge Embedding, DGL-KE) 进行知识图谱嵌入, 因为 DGL-KE 使用的是图神经网络 (Graph Neural Networks, GNN), 能够有效地捕捉实体之间以及实体与关系之间的复杂交互和上下文信息, 进而提高嵌入的质量和准确性. DGL-KE 的一个显著优势是其对多图形处理器 (Graphics Processing Unit, GPU) 并行计算的支持, 这一特性可以显著提高嵌入学习的速度, 尤其是处理大规模知识图谱时. 这种并行加速能力使得 DGL-KE 成为处理庞大知识图谱的理想工具, 能够有效地缩短模型训练时间, 加快研究进程^[8]. 在本课题的实践中, 将整合后的知识

图谱三元组数据集分割为训练集、验证集和测试集, 按照 9:0.5:0.5 的比例进行分割. 这样的分割比例确保了足够大的训练集来学习模型, 同时提供了合理的验证集和测试集来评估模型的泛化能力和预测性能. 通过这种方法, 可以确保嵌入学习过程的稳健性, 并有效评估模型对未见数据的推理能力.

1.4 三元关系预测

1.4.1 三元关系构建

基于融合后知识图谱中的二元关系, 构建新的三元关系, 为后续的三元关系预测奠定基础. 具体方法如下: 从知识图谱中筛选出基因-疾病、基因-药物和药物-疾病的二元关系, 通过共享实体 (如基因和药物), 将这些二元关系整合成新的三元关系. 为了节约内存空间, 从 0 开始对不同的实体进行编号.

通过这种基于二元关系构建三元关系的方法, 能够系统地揭示潜在的生物医学关联, 为进一步的三元关系预测提供重要的数据支持.

1.4.2 异质因果元路径图神经网络模型

HCMGNN 是一种新型的图神经网络架构, 其结合了异质图和元路径分析的概念, 专注于模拟和解析不同类型的实体及其复杂交互中的因果关系. 这种网络设计充分考虑了实体类型的多样性和实体间交互的复杂性, 使得网络不仅能处理标准的节点和边信息, 还能深入理解和利用节点间的潜在因果关系.

HCMGNN 模型的关键在于如何处理这些复杂的数据结构. 首先, 模型采用多层感知机 (Multilayer Perceptron, MLP) 对各种类型的节点进行特征转换, 确保了输入特征的维度统一且信息丰富. 接着, 通过定义和利用元路径来构建表示不同生物学过程的因果图, HCMGNN 能够在这些子图上应用图神经网络技术, 有效地传递和更新节点信息. 在子图分析过程中, HCMGNN 模型利用图神经网络内部的信息传递机制来更新每个节点的表示, 确保了模型能够捕捉节点间复杂的依赖关系. 最终, 模型通过注意力机制整合不同子图产生的信息, 对不同来源的信息赋予不同的权重, 实现了信息的有效融合. 通过这种方式, HCMGNN 模型能够生成高质量的节点嵌入, 这些嵌入随后可以用于各种生物医学预测任务, 如药物再利用、基因-疾病-药物关联预测等, 以提高这些任务的准确性和可靠性^[13].

与传统的图神经网络相比, HCMGNN 中, 网络首先识别并利用不同类型的节点和边, 这些节点和边

代表了数据中的不同实体和实体间的关系, 并通过引入元路径的概念, 网络能够通过预定义的路径模式, 即元路径, 来捕捉实体间复杂的因果关系, 这种网络能更精确地表达和处理图数据的异质性和因果关系.

在实际应用中, HCMGNN 的引入极大地增强了模型处理复杂数据结构的能力. 例如, 在推荐系统中, 传统的图神经网络可能只侧重于用户和物品之间的直接交互, 而忽视了上下文因素如时间、地点等的影响. HCMGNN 通过构建包括用户、物品、上下文在内的多元路径, 不仅能捕捉用户和物品的直接关系, 还能考虑到上下文的因果影响, 从而实现更为准确的推荐^[14].

此外, 相比普通的图神经网络, HCMGNN 在三元组预测上展示了显著的优势. 在处理如社交网络分析、知识图谱补全等任务时, HCMGNN 不仅能够准确预测实体间是否存在链接, 更能预测链接的类型和方向, 这是因为其通过元路径分析深入挖掘了节点间的潜在因果关系. 这种深度的因果关系挖掘, 使得网络在预测准确度和模型解释性方面都有很大的提升^[13]. 利用 HCMGNN 模型处理的异构图数据包含基因、药物和疾病的生物学实体.

1.4.3 五折交叉验证与独立性验证

五折交叉验证的基本思想是将原始数据集分成 5 个相等大小的子集 (或折叠), 其中 4 个子集用于训练模型, 而剩余 1 个子集用于测试模型. 这个过程重复 5 次, 每次选择不同的一个子集作为测试集, 其余的作为训练集^[15]. 这种方法能够有效地利用数据, 减少过拟合风险, 并提供模型性能的可靠估计^[16].

独立性验证通常用于估计模型的泛化能力和在未见数据上的表现, 其基本思想是将原始数据集分成训练集和独立验证集, 训练集用于训练模型, 而独立验证集用于测试模型. 在模型训练过程中, 独立验证集从未被使用过, 因此可以提供一个对模型性能的无偏估计^[17].

从数学角度来讲, 模型验证的目标就是计算评价指标函数 E_{HITS} 和平均倒数排名 (Mean Reciprocal Ranking, MRR) R_{MRR} . 此外, 使用 Q_{NDCG} 衡量预测结果的排序质量. 计算 Q_{NDCG} 时, 归一化折扣累积增益, 考虑到了排名顺序和相关性权重, 给予排名靠前且相关性高的项更高权重. Q_{IDCG} 是理想状态下的最大可能 Q_{DCG} , 用于将 Q_{DCG} 归一化, 以确保数值处于 $[0, 1]$. 计算公式如下:

$$E_{\text{HITS}-N} = \frac{1}{|S|} \sum_{i=1}^{|S|} \prod (R_i \leq n), \quad (3)$$

$$R_{\text{MRR}} = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{1}{R_i}, \quad (4)$$

$$Q_{\text{NDCG}} = \frac{Q_{\text{DCG}}}{Q_{\text{IDCG}}}, \quad (5)$$

$$Q_{\text{DCG}} = \sum_{i=1}^p \frac{2^{R_{\text{result}}^i} - 1}{\text{lb}(i+1)}, \quad (6)$$

$$Q_{\text{IDCG}} = \sum_{i=1}^{|N_{\text{REL}}|} \frac{2^{R_{\text{result}}^i} - 1}{\text{lb}(i+1)}. \quad (7)$$

其中, $|S|$ 表示样本集的大小, 即测试集中的样本数量; 1 为指示函数, 如果预测的排名在前 N 位内, 则返回 1, 否则返回 0; R_i 表示样本中目标项的排名; p 表示在 Q_{DCG} 计算中考虑的位置数; R_{result}^i 表示在位置 i 处的结果的相关等级, 通常是一个整数值, 反映该项与查询的相关性; N_{REL} 表示模型验证中衡量排序质量的关键参数, 定义了每个样本的理想相关项基数, 为 Q_{NDCG} 的归一化提供理论依据. 通过引入 N_{REL} , 模型能够区分“部分相关但排序不佳”与“完全无关”的预测结果, 从而更精准地评估排序算法的性能.

1.4.4 链路预测

将提取出的疾病实体的嵌入向量作为输入, 以三元关系中所有的疾病和药物作为预测对象, 计算每个基因-药物-疾病三元关系的合并嵌入向量 $\mathbf{h}$, 其表达式为

$$\mathbf{h} = [\mathbf{G} \cdot \mathbf{C} \cdot \mathbf{D}], \quad (8)$$

其中, $\mathbf{G}$ 表示基因 (Gene) 实体的嵌入矩阵, $\mathbf{C}$ 表示药物 (Compound) 实体的嵌入矩阵, $\mathbf{D}$ 表示疾病 (Disease) 实体的嵌入矩阵, $[\]$ 表示合并向量. 计算三元组存在的第 i 个样本预测分数, 有

$$s_i = \sigma(\mathbf{W} \cdot \mathbf{h} + b). \quad (9)$$

其中, σ 表示激活函数, $\mathbf{W}$ 表示线性变换的权重矩阵, b 表示线性变换的偏置项.

2 结果与分析

2.1 知识图谱融合结果

通过整合标准化后的三元组, 成功构建了大规模 MBKG, 如图 2 所示. 融合后知识图谱包含 157586

个生物学实体、104 个关系和 7338119 个三元组。其中包含基因实体和药物实体的三元组有 872016 个, 涵盖了 39 种细分关系; 包含基因实体和疾病实体的三元组有 804808 个, 涵盖了 15 种细分关系; 包含疾病实体和药物实体的三元组有 675952 个, 涵盖了 12 种细分关系。

融合后的 MBKG 包含了丰富的信息, 并对实体间的关系进行详细分类, 这有助于构建三元关系, 更有助于在知识图谱训练中获得更丰富的嵌入, 更准确地获取实体之间存在的关系, 从而更准确地预测与所选微重力疾病相关的三元关系。

2.2 TransE 嵌入训练结果

嵌入训练的结果中, HITS- N 是用于评估信息检索或推荐系统中的排名性能的指标之一。其衡量的是在给定的查询下, 模型返回的结果中是否包含了正确答案 (Ground Truth) 在前 N 个位置上。评价指标函数 $E_{\text{HITS-}N}$ 的计算方式是统计有多少查询的正确答案在前 N 个位置上, 并将这个数量除以总查询数量, 以得到一个百分比。如果目标项在前 N 位内, 则记为 1, 否则为 0, 再计算所有样本的平均值, 数字越接近 1, 则表示模型训练结果越优异^[18]。MRR 是衡量模型性能的一个常用指标, 其通过计算所有正确三元组的倒数排名的平均值来评估模型。MRR 的取值范围在 0~1, 越接近 1 表示模型性能越好, 而越接近 0 则表示模型性能较差^[19]。MR (Mean Rank) 通常指在信息检索或推荐系统中, 针对每个查询返回的结果集合, 计算其中正确答案在排名中的平均位置, MR 值越小, 表明模型性能越好^[18]。融合图谱的 TransE 嵌入训练结果见表 3。其中, 平均 $E_{\text{HITS-10}} > 0.9$, 本研究认为该融合图谱的质量尚可。

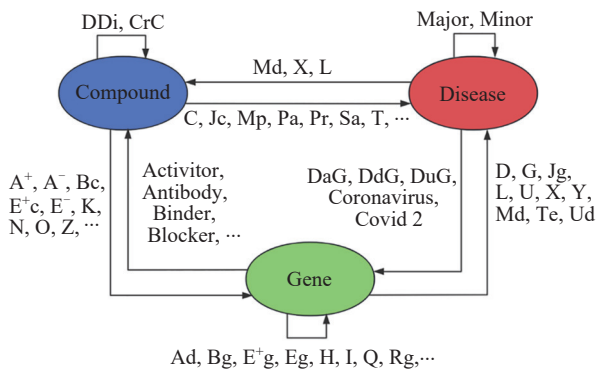


图 2 融合知识图谱

Fig. 2 Fused knowledge graph

2.3 三元关系预测结果

2.3.1 三元关系实体特征向量分析

基于融合后的知识图谱, 研究构建基因-药物-疾病三元关系, 共计 89312 条。TransE 模型训练融合图谱后生成对应实体的低维嵌入文件 TRAIN_TransE_l2_entity.npy, 其中包含了图谱中所有实体的低维嵌入向量。因为嵌入文件中部分实体无法构建三元关系, 所以本研究基于构建完成的三元关系, 从嵌入文件中筛选出基因、药物和疾病所对应的向量, 分别用于计算对应的特征向量。

在计算实体特征向量时, 研究基于余弦相似度和欧氏距离分别进行计算。余弦相似度衡量了两个向量之间夹角的余弦值, 即方向是否相似。在低维空间中, 当实体向量的方向相似时, 余弦相似度会很高。欧氏距离衡量了两个向量之间的直线距离, 考虑了向量之间的长度差异。在低维空间中, 如果两个实体向量之间的距离很小, 这通常意味着之间的关系更紧密, 而距离更大则意味着关系更弱。由于 TransE 模型在训练时的目标是最小化关系三元组之间的差距, 因此使用欧氏距离来衡量实体向量之间的差异可能更能反映出这种关系的差异性。而余弦相似度可能无法很好地捕捉到这种差异, 因为其主要关注向量之间的方向而不是长度。本研究最终采用基于欧氏距离计算特征向量的结果^[20]。

2.3.2 五折交叉验证与独立性检验结果

针对 HCMGNN 模型, 研究选择用五折交叉验证评价模型训练结果, 同时以独立性检验结果佐证。其中将 $E_{\text{HITS-1}}$, $E_{\text{HITS-3}}$, $E_{\text{HITS-5}}$, $Q_{\text{NDCG-1}}$, $Q_{\text{NDCG-3}}$, $Q_{\text{NDCG-5}}$, R_{MRR} 7 个评价指标作为衡量基于 MBKG 的三元关系预测模型预测效果的指标。五折交叉验证结果如表 4 所示, 其中 $E_{\text{HITS-5}}$ 结果为 0.981 表示在五折交叉验证中, 有 98.10% 的正确答案出现在前 5 个推荐结果中; $Q_{\text{NDCG-5}}$ 结果为 0.886 表示在五折交叉

表 3 TransE 模型嵌入训练结果

Table 3 Results of TransE model embedding training

评价指标	结果
Average MRR	0.737
Average MR	6.67
Average HITS-10	0.640
Average HITS-3	0.804
Average HITS-10	0.911

验证中, 模型推荐的前 5 个结果的排序质量高, 接近理想排序; MRR 结果为 0.857 表示在独立测试中, 正确答案平均排名的倒数是 0.857, 即模型通常能够在推荐结果的前 1.17(即 $1/0.857$) 个位置找到正确答案, 预测准确性高. 独立性检验结果列于表 4, 其结果与五折交叉验证结果相近, 这表明模型具有很好的预测能力和泛化性能. 由于五折交叉验证模型的稳定性, 而独立测试则验证模型在真实世界应用中的有效性.

将五折交叉验证结果与 TransE 结果对比发现, 五折交叉验证结果的 Average HITS 和 Average MRR 均高于 TransE 结果, 即结果更加优异. 这证明在本研究中, 选用 HCMGNN 进行三元预测可以得到不低于传统的基于知识图谱二元预测的结果. 此外, 研究将五折交叉验证与独立性检验的过程, 以 Epoch 次数为横轴, Average HITS, Average NDCG, Average MRR 三个评价指标为纵轴, 利用 Python 中的 matplotlib 包给出折线图 3 和图 4 显示模型基于提供数据的学习过程.

2.3.3 链路预测结果

基于 MBKG 的三元关系预测模型, 本研究以骨质疏松症 Osteoporosis, MeshID:D010024, 三元关系编号 5746 为例, 预测与该疾病相关的基因和药物, 得到打分前 10 的基因-药物-疾病三元关系 (见表 5). 检索各大生物学数据库发现, 结果显示前 10 个预测中得分排名 1, 3, 4, 5, 6, 9 和 10 的三元关系对应三个实体已被证实存在两两相互作用. 在排名 2, 7 和 8 的三元关系中, 目前的研究没有直接探讨 CYP19 A1 和 Genistein, NR3 C1 和 Testosterone, AMH 和 Dexamethasone 之间的直接相互作用.

骨质疏松症是一种常见的骨骼疾病, 其特征在干

骨密度和骨质量显著降低, 导致骨骼变得脆弱和易发生骨折. 这种疾病主要影响老年人, 特别是绝经后的女性, 由于激素变化会加速骨质流失. 骨质疏松症的主要病理特征包括骨小梁结构的退化和骨皮质的变薄, 导致骨骼微结构的破坏和整体骨强度的减弱. 常见的骨折部位包括髋部、脊椎和前臂, 其骨折不仅会导致严重的疼痛, 还可能引起功能丧失和生活质量的显著下降. 由于骨质疏松症早期通常没有明显症状, 常被称为“沉默的疾病”.

研究表明, BGLAP 基因的多态性与骨质疏松症有关. 在土耳其绝经后女性中研究发现, BGLAP/Col1 A1 CCTT 和 ER1/CALCR CCTT 联合基因型在骨质疏松症中的风险较高, 而 ER1/CALCR TC-CC 联合基因型对骨质疏松症有保护作用^[21]. 而 Genistein 是一种大豆异黄酮, 被证明对骨质疏松症有保护作用. 研究显示, Genistein 可通过增加骨形成标志物和减少骨吸收标志物来提高骨密度. 每天 54 mg 的 Genistein 可显著增加绝经后女性的骨密度, 尤其是在腰椎和股骨部位^[22]. 另一项研究表明, Genistein 可以通过调节 RANKL/OPG 比率来抑制类风湿关节炎患者的骨吸收, 从而促进骨形成. 同时, Genistein 可导致 BGLAP mRNA 表达的增加, 并且在多种相互作用中均有表现^[23].

PPAR γ 基因的多态性同样与骨质疏松症的风险有关. 在土耳其绝经后女性中研究发现, PPAR γ Pro12 Ala 和 C161 T 多态性与骨质疏松症的风险相关, Pro12 Pro 基因型的患者腰椎骨密度较低^[24]. 此外, Genistein 同时激活了两种不同的转录因子 ERs 和 PPAR γ , 这二者在骨生成和脂肪生成中具有相反的作用. Genistein 对骨生成和脂肪生成的生物

表 4 五折交叉验证结果与独立性验证结果
Table 4 Results of 5-fold cross-validation and the results of independence validation

五折交叉验证结果		独立性验证结果	
评价指标	结果	评价指标	结果
Average HITS-1	0.761	Average HITS-1	0.865
Average HITS-3	0.944	Average HITS-3	0.980
Average HITS-5	0.981	Average HITS-5	0.990
Average NDCG-1	0.761	Average NDCG-1	0.865
Average NDCG-3	0.872	Average NDCG-3	0.872
Average NDCG-5	0.886	Average NDCG-5	0.934
Average MRR	0.857	Average MRR	0.921

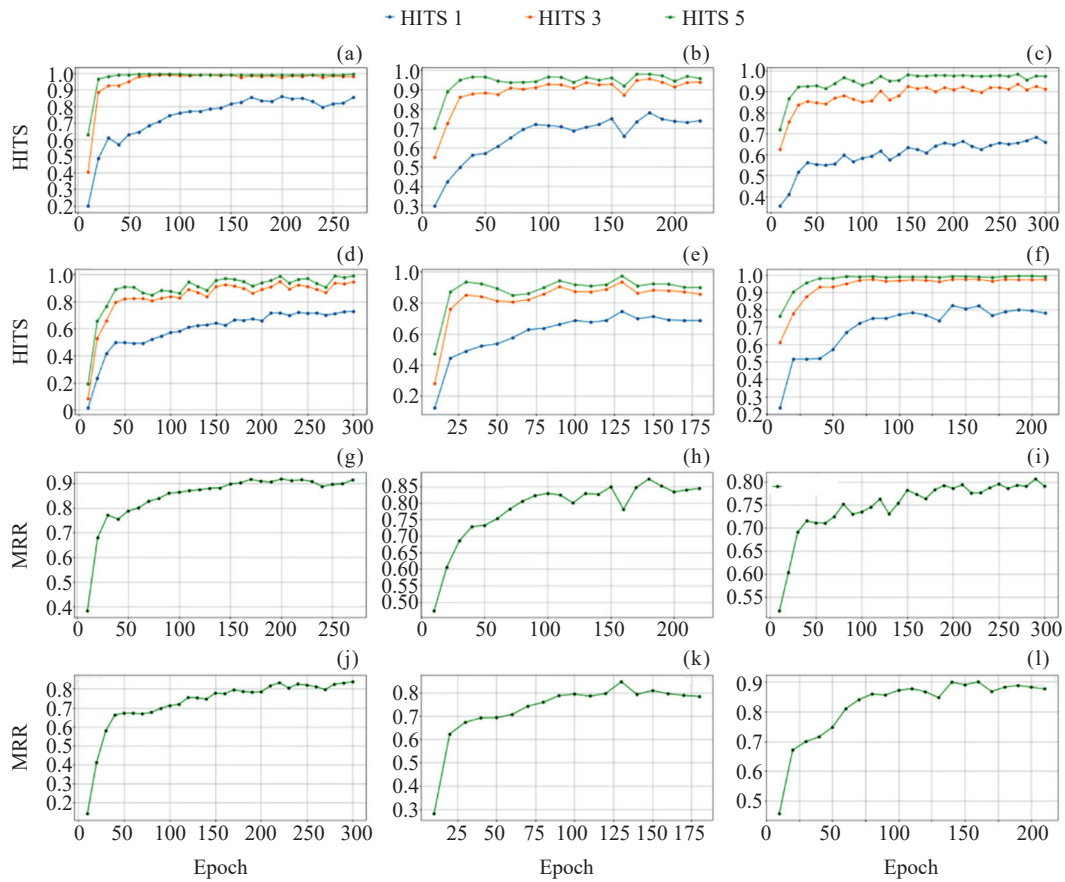


图 3 五折交叉验证和独立性验证 E_{HITS-N} 与 Average MRR 结果折线. (a) 独立性验证 E_{HITS-N} 结果折线, (b)~(f) 五折交叉验证 E_{HITS-N} 结果折线. (g) 独立性验证 Average MRR 结果折线, (h)~(l) 五折交叉验证 Average MRR 结果折线

Fig. 3 Line graphs of E_{HITS-N} and Average MRR Results for 5-fold cross-validation and independence validation.
(a) Independence validation E_{HITS-N} result line graph, (b)~(f) five-fold cross-validation E_{HITS-N} result line graphs. (g) independence validation Average MRR result line graph, (h)~(l) five-fold cross-validation Average MRR result line graphs

学效应由 ERs 和 PPAR γ 的激活平衡决定^[25]. 文献 [26] 的研究表明, 在微重力响应中, PPAR γ 2 高度表达, 因此本文模型预测出在微重力环境下, Genistein 通过对 PPAR γ 基因的作用来影响骨质疏松是可靠的.

TGF- β 1 是一种多功能细胞因子, 在细胞增殖、分化和免疫调节中发挥重要作用. 文献 [27] 的研究表明, 在微重力环境下, 在胎儿成骨细胞中, TGF- β 1 会差异表达, 并且在两周的微重力环境暴露后, 成骨细胞球状体呈现出骨特异性形态^[27]. 研究结果表明, Genistein 可以通过多种机制调节 TGF- β 1 的表达. 例如, Genistein 能够抑制由 Galactosamine 诱导的 TGF- β 1 mRNA 和蛋白的表达, 这表明其在调节 TGF- β 1 信号通路中具有抑制作用. 此外, Genis-

tein 还可以通过影响雌激素受体途径, 增加 TGF- β 1 的合成, 特别是在生殖系统中. 关于骨质疏松症, TGF- β 1 在骨代谢中发挥关键作用, 其能够促进成骨细胞的分化 and 功能, 同时抑制破骨细胞的活性, 从而维护骨骼健康^[28].

ESR1 基因编码雌激素受体 α (ER α), 其突变与乳腺癌耐药性相关. Lasofoxifene 是一种第三代选择性雌激素受体调节剂 (SERM), 被研究用于治疗骨质疏松症和乳腺癌. 在一些临床试验中, Lasofoxifene 显示出对具有 ESR1 突变的 ER+乳腺癌患者的显著疗效, 能够减少癌症转移并改善无进展生存期^[29].

ESR1 基因多态性 (例如 PvuII 和 XbaI 位点) 与骨矿物质密度 (BMD) 和骨折风险密切相关. 有研究表明, 许多细胞类型在微重力下培养时会形成三维聚

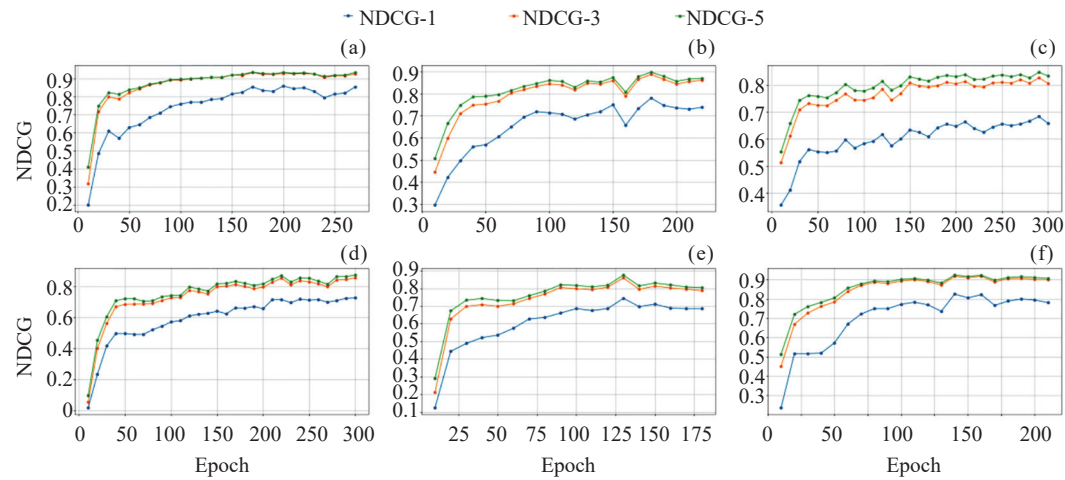


图 4 五折交叉验证和独立性验证 Average NDCG 结果折线. (a) 独立性验证 Average NDCG 结果折线, (b)~(f) 为五折交叉验证 Average NDCG 结果折线

Fig. 4 Line graphs of Average NDCG results for 5-fold cross-validation and independence validation. (a) Independence validation Average NDCG result line graphs, (b)~(f) the five-fold cross-validation Average NDCG result line graphs. The horizontal scale shows the change in Epcho, the vertical scale shows the change in Average NDCG

表 5 预测得分前 10 的基因-药物-疾病三元关系
Table 5 Gene-compound-disease ternary relationships with Top 10 predicted scores

Rank	Gene_id	Gene_name	Compound_id	Compound_name	Score
1	Gene::632	BGLAP	Compound::DB01645	Genistein	-1.96359
2	Gene::1588	CYP19 A1	Compound::DB01645	Genistein	-2.00393
3	Gene::5468	PPAR γ	Compound::DB01645	Genistein	-2.06026
4	Gene::7040	TGF- β 1	Compound::DB01645	Genistein	-2.06260
5	Gene::2099	ESR1	Compound::DB06202	Lasofoxifene	-2.13473
6	Gene::5745	PTH1 R	Compound::DB06285	Teriparatide	-2.16564
7	Gene::2908	NR3 C1	Compound::DB00624	Testosterone	-2.19646
8	Gene::268	AMH	Compound::DB01234	Dexamethasone	-2.41574
9	Gene::2099	ESR1	Compound::DB06249	Arzoxifene	-2.42039
10	Gene::5745	PTH1 R	Compound::DB01234	Dexamethasone	-2.43293

集体 (多细胞球体 (MCS)), 在微重力环境下的乳腺癌 MCS 中, ESR1 表达上调, Goetz 等^[30]的研究发现, 携带 ESR1 某些基因型的个体在骨密度和骨折风险上存在显著差异, 提示 ESR1 基因在骨质疏松症的遗传易感性中起重要作用. Lasofoxifene 作为一种 SERM, 被证明对骨质疏松症患者具有显著的治疗效果, 能够增加骨密度并减少骨折风险. 临床研究表明, Lasofoxifene 在治疗绝经后女性的骨质疏松症方面表现出色, 显著改善了骨骼健康^[31].

PTH1 R (甲状旁腺激素 1 型受体) 基因多态性

与骨密度 (BMD) 以及骨折风险密切相关. 研究发现, PTH1 R 基因变异可能会影响 Teriparatide 治疗的效果, 因为该基因在调节骨矿物质密度和骨代谢过程中起重要作用. 同时, PTH1 R 在骨髓中的高表达与某些药物治疗反应有关, 包括使用 Dexamethasone 治疗的多发性骨髓瘤患者. 在这些患者中, PTH1 R 表达水平越高, 生存期越长^[32]. 研究表明, PTH1 R 基因的某些变异可能与低骨密度和增加的骨折风险相关, 从而提示该基因在骨质疏松症的遗传易感性中发挥作用^[33]. Teriparatide 是一种重组人甲

状旁腺激素 (PTH 1-34), 其通过增加骨形成来治疗骨质疏松症. 临床研究表明, Teriparatide 能够显著提高骨密度并减少骨折风险, 特别是在骨质疏松症患者中表现出色^[34].

Arzoxifene 作为一种 SERM, 在骨质疏松症的治疗中表现出显著效果. 临床研究表明, Arzoxifene 能够显著提高绝经后女性的骨密度, 并减少骨折风险, 特别是在骨密度低下和骨折高风险的患者中表现出良好的治疗效果^[35]. 与之相反, Dexamethasone 是一种合成的肾上腺皮质激素 (类固醇), 具有强效的抗炎和免疫抑制作用. 但是长期使用 Dexamethasone 会导致骨质疏松症, 因其抑制骨形成并增加骨吸收. 研究表明, Dexamethasone 通过与骨代谢相关基因的相互作用, 影响骨密度和骨折风险^[32].

3 讨论

随着全球环境恶化和能源消耗的增长, 空间探测变得越来越重要. 然而微重力环境对航天员的生理和心理健康带来了挑战, 包括骨质流失、肌肉萎缩等问题. 本文研究融合 MBKG 和 DRKG, 构建基因-药物-疾病三元关系数据, 训练 HCMGNN, 并展示了其在基因-药物-疾病三元关系预测中的有效性. HCMGNN 模型在处理复杂异构数据和识别潜在因果关系方面表现出色, 并且在五折交叉验证和独立性验证中均展示了较高的预测准确性和可靠性, 显著优于传统的知识图谱嵌入方法 TransE.

在具体案例分析中, HCMGNN 模型在预测骨质疏松症 Osteoporosis 相关的基因-药物-疾病三元关系方面表现出色. 以 BGLAP 基因和 Genistein 药物为例, 这一组合在本研究的预测中排名最高. 这一发现与现有研究一致, Genistein 通过调节 RANKL/OPG 比率来促进骨形成, 从而在骨质疏松症的防治中发挥作用. 其他如 PPAR γ 和 TGF- β 1 等基因也显示出与特定药物的潜在相互作用, 进一步验证了模型的可靠性和实用性.

除骨质疏松外, 肌肉萎缩和心血管疾病也是微重力环境中对航天员健康影响的重要生理性疾病. 肌肉萎缩主要是由于缺乏重力荷载和运动刺激, 导致肌肉蛋白质合成减少和分解增加. 心血管疾病是由于微重力导致血液分布的改变, 心脏负荷减少, 从而引发心脏结构和功能的变化. 在这些疾病的研究中,

HCMGNN 模型的应用前景广阔. 通过构建基因-药物-疾病的三元关系, HCMGNN 模型能够识别与肌肉萎缩和心血管疾病相关的潜在药物. 例如, HCMGNN 模型可以预测特定基因 (例如 MSTN 和 IGF1) 与药物 (如米托坦和洛卡特普) 之间的相互作用, 从而为肌肉萎缩和心血管疾病的治疗提供新的药物再利用策略. 此外, HCMGNN 模型的灵活性使其能够整合多种生物医学数据, 为未来的研究提供更全面的理论基础和数据支持.

尽管 HCMGNN 模型表现优异, 但仍存在一些局限性, 数据集的完整性和准确性直接影响其预测效果. MBKG 和 DRKG 虽然提供了丰富的生物医学数据, 但数据的来源和质量不一, 可能导致预测的偏差. 此外, HCMGNN 模型的复杂性增加了计算成本, 在实际应用中可能面临挑战.

未来的研究可以从多个方面进行优化和扩展. 首先, 在微重力环境中, 航天员经历的骨流失主要是由于机械负荷缺失以及激素调节的改变. 现有地面治疗骨质疏松的药物如双膦酸盐, 在微重力条件下未能显示出有效的防治效果. 通过对现有文献的调研, 潜在的治疗空间骨质疏松的药物包括新型合成生长因子 (如 FGF-2 或 BMPs), 这些因子在动物模型中显示出促进骨形成的前景^[36]. 同时, 药物再利用分析显示, 一些抗炎药物可能通过改善骨重塑过程来减缓骨流失. 基于这些观察, 结合适当的营养补充, 如维生素 D 和钙的搭配使用, 可能为微重力环境下的骨质疏松治疗提供新的思路 and 方向^[37]. 未来的研究应进一步验证这些药物在微重力条件下的实际效果, 为航天员的健康提供切实可行的解决方案. 其次, 心理问题可能是未来长期航天任务的最大阻碍, 在空间失重条件下的隔离密闭空间内, 乘组内的个体之间必须建立并维持一种有效且稳定的相互关系, 才能减缓心理问题的影响^[38]. 如何进行临床医学选拔、心理选拔和航天特殊环境耐力选拔是未来值得深入探讨的方向. 最后, 可以尝试加入更先进的神经网络算法, 如图卷积网络 (Graph Convolutional Network, GCN), 进一步提高模型的预测准确性.

研究基于 MBKG 的三元关系预测模型在探索基因-药物-疾病三元关系预测方面的有效性和潜力, 为未来空间探测中航天员的生理和心理健康以及药物再利用研究提供了新的方法和思路. 研究结果不仅为微重力环境下药物筛选提供了参考, 也为知识图谱和

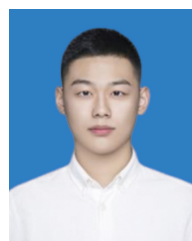
图神经网络在生物医学中的应用展示了新的方向. 未来, 应继续优化模型, 扩展知识图谱的覆盖范围, 提高预测的准确性和应用价值, 以期在实际应用中取得更大的突破. 通过不断完善模型和引入更多数据, 有望进一步提升对复杂生物医学问题的理解和解决能力, 为空间探测和地面医疗带来更多实质性进展. 这不仅有助于保障航天员的健康, 也为更广泛的医学研究和临床应用提供了新的视角和工具.

参考文献

- [1] KIM D H, LIN S C. Human capital and natural resource dependence[J]. *Structural Change and Economic Dynamics*, 2017, **40**: 92-102
- [2] DREES J M, HEUGENS P P M A R. Synthesizing and extending resource dependence theory: a meta-analysis[J]. *Journal of Management*, 2013, **39**(6): 1666-1698
- [3] FANKHAUSER T, WANG Q, GERLICHER A, *et al*. Resource dependency processing in web scaling frameworks[J]. *IEEE Transactions on Services Computing*, 2018, **11**(1): 155-168
- [4] DICKERSON B L, SOWINSKI R, KREIDER R B, *et al*. Impacts of microgravity on amino acid metabolism during spaceflight[J]. *Experimental Biology and Medicine*, 2023, **248**(5): 380-393
- [5] OLUWAFEMI F A, ABDELBAKI R, LAI J C Y, *et al*. A review of astronaut mental health in manned missions: Potential interventions for cognitive and mental health challenges[J]. *Life Sciences in Space Research*, 2021, **28**: 26-31
- [6] SPRUGNOLI G, CAGLE Y D, SANTARNECCHI E. Microgravity and cosmic radiations during space exploration as a window into neurodegeneration on earth[J]. *JAMA Neurology*, 2020, **77**(2): 157-158
- [7] VAN OMBERGEN A, DEMERTZI A, TOMILOVSKAYA E, *et al*. The effect of spaceflight and microgravity on the human brain[J]. *Journal of Neurology*, 2017, **264**(1): 18-22
- [8] ZHENG D, SONG X, MA C, *et al*. DGL-KE: training knowledge graph embeddings at scale[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020
- [9] SANG S T, YANG Z H, LIU X X, *et al*. GrEDeL: a knowledge graph embedding based method for drug discovery from biomedical literatures[J]. *IEEE Access*, 2018, **7**: 8404-8415
- [10] FARRUGIA L, AZZOPARDI L M, DEBATTISTA J, *et al*. Predicting drug-drug interactions using knowledge graphs[OL]. arXiv preprint arXiv: 2308.04172, 2023
- [11] ZHENG Y H, PAN G J, QUAN Y, *et al*. Construction of microgravity biological knowledge graph and its applications in anti-osteoporosis drug prediction, Life Sciences in Space Research, Volume 41, 2024, Pages 64-73, ISSN 2214-5524.
- [12] Ioannidis, V. N. *et al*. Drkg-drug repurposing knowledge graph for covid-19. GitHub <https://github.com/gnm4dr/DRKG> (2020). Accessed 01 Jan 2022.
- [13] ZHU G H, ZHU Z N, CHEN H Y, *et al*. HAGNN: hybrid aggregation for heterogeneous graph neural networks[OL]. arXiv preprint arXiv: 2307.01636, 2023
- [14] LIANG T, LIU J. Meta-path generation online for heterogeneous network embedding[C]//Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN). Glasgow, UK: IEEE, 2020
- [15] RODRIGUEZ J D, PEREZ A, LOZANO J A. Sensitivity analysis of k-fold cross validation in prediction error estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(3): 569-575
- [16] POHJANKUKKA J, PAHIKKALA T, NEVALAINEN P, *et al*. Estimating the prediction performance of spatial models via spatial k-fold cross validation[J]. *International Journal of Geographical Information Science*, 2017, **31**(10): 2001-2019
- [17] BERGMIEIR C, HYNDMAN R J, KOO B. A note on the validity of cross-validation for evaluating autoregressive time series prediction[J]. *Computational Statistics & Data Analysis*, 2018, **120**: 70-83
- [18] SAHA S, ROY D, MITRA M. On modifying evaluation measures to deal with ties in ranked lists[C]//Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries. Cologne, Germany: IEEE, 2022
- [19] GOEL S, KUMAR R, KUMAR M, *et al*. An efficient page ranking approach based on vector norms using sNorm(p) algorithm[J]. *Information Processing & Management*, 2019, **56**(3): 1053-1066
- [20] SADEGHI A, GRAUX D, YAZDI H S, *et al*. MDE: multiple distance embeddings for link prediction in knowledge graphs[OL]. arXiv preprint arXiv: 1905.10702, 2019
- [21] TURAL S, KARA N, ALAYLI G, *et al*. Association between osteoporosis and polymorphisms of the bone Gla protein, estrogen receptor 1, collagen 1-A1 and calcitonin receptor genes in Turkish postmenopausal women[J]. *Gene*, 2013, **515**(1): 167-172
- [22] SEHMISCH S, UFFENORDE J, MAEHLMEYER S, *et al*. Evaluation of bone quality and quantity in osteoporotic mice—the effects of genistein and equol[J]. *Phytomedicine*, 2010, **17**(6): 424-430
- [23] Da Silva F R P, Vasconcelos A C C G, Casimiro G S, *et al*. Quantitative assessment of the association between polymorphisms in osteoprotegerin gene and risk of low bone mineral density[J]. *International Archives of Medicine*, 2015: 8(169)
- [24] ŞİRİN Ö K, AYDOĞAN H Y, UYAR M, *et al*. PPAR γ Pro12Ala and C161T polymorphisms, but not PPAR α L162V, are associated with osteoporosis risk in Turkish

- postmenopausal women[J]. *Istanbul Journal of Pharmacy*, 2019, **49**(1): 14-19
- [25] SCHOLTYSEK C, KATZENBEISSER J, FU H, *et al*. PPAR β/δ governs Wnt signaling and bone turnover[J]. *Nature Medicine*, 2013, **19**(5): 608-613
- [26] ZAYZAFOON M, GATHINGS W E, MCDONALD J M. Modeled microgravity inhibits osteogenic differentiation of human mesenchymal stem cells and increases Adipogenesis[J]. *Endocrinology*, 2004, **145**(5): 2421-2432
- [27] MANN V, GRIMM D, CORYDON T J, *et al*. Changes in human foetal osteoblasts exposed to the random positioning machine and bone construct tissue engineering[J]. *International Journal of Molecular Sciences*, 2019, **20**(6): 1357
- [28] JURZAK M, ADAMCZYK K, ANTONCZAK P, *et al*. Evaluation of genistein ability to modulate CTGF mRNA/protein expression, genes expression of TGF β isoforms and expression of selected genes regulating cell cycle in keloid fibroblasts *in vitro*[J]. *Acta Poloniae Pharmaceutica*, 2014, **71**(6): 972-986
- [29] Plourde P V, SCHWARTZBERG L S, GREENE G L, *et al*. . An open-label, randomized, multi-center phase 2 study evaluating the activity of lasofoxifene relative to fulvestrant for the treatment of postmenopausal women with locally advanced or Metastatic ER+/HER2- Breast Cancer (MBC) with an ESR1 mutation[C]//Cancer Research, San Antonio: San Antonio Breast Cancer Symposium, 2019
- [30] GOETZ M P, GAL-YAM E, STOVER D, *et al*. Abstract P5-05-04: Estrogen Receptor 1 (ESR1) mutations in circulating tumor DNA (ctDNA) from patients with ER+/HER2-metastatic Breast Cancer (mBC) treated with lasofoxifene or fulvestrant in the ELAINE 1 study[J]. *Cancer Research*, 2023, **83**(S5): P5-05-04
- [31] SEFRIQUI D, PERDRIX A, SARAFAN-VASSEUR N, *et al*. Short report: monitoring ESR1 mutations by circulating tumor DNA in aromatase inhibitor resistant metastatic breast cancer[J]. *International Journal of Cancer*, 2015, **137**(10): 2513-2519
- [32] NOORDIN S, GLOWACKI J. Parathyroid hormone and its receptor gene polymorphisms: implications in osteoporosis and in fracture healing[J]. *Rheumatology International*, 2016, **36**(1): 1-6
- [33] STYRKARSDOTTIR U, THORLEIFSSON G, GUDJONSSON S A, *et al*. Sequence variants in the PTCH1 gene associate with spine bone mineral density and osteoporotic fractures[J]. *Nature Communications*, 2016, **7**(1): 10129
- [34] NAKAMURA T, SUGIMOTO T, NAKANO T, *et al*. Randomized Teriparatide [Human Parathyroid Hormone (PTH) 1-34] Once-Weekly Efficacy Research (TOWER) trial for examining the reduction in new vertebral fractures in subjects with primary osteoporosis and high fracture risk[J]. *The Journal of Clinical Endocrinology & Metabolism*, 2012, **97**(9): 3097-3106
- [35] CUMMINGS S R, MCCLUNG M, REGINSTER J Y, *et al*. Arzoxifene for prevention of fractures and invasive breast cancer in postmenopausal women[J]. *Journal of Bone and Mineral Research*, 2011, **26**(2): 397-404
- [36] VICO L, HARGENS A. Skeletal changes during and after spaceflight[J]. *Nature Reviews Rheumatology*, 2018, **14**(2): 229-245
- [37] SMITH S M, HEER M A, SHACKELFORD L C, *et al*. Benefits for bone from resistance exercise and nutrition in long-duration spaceflight[J]. *Journal of Applied Physiology*, 2012, **112**(1): 105-113
- [38] CARTER J A, BUCKEY J C, GREENHALGH L, *et al*. An interactive media program for managing psychosocial problems on long-duration spaceflights[J]. *Aviation, Space, and Environmental Medicine*, 2005, **76**(S6): B213-B223

作者简介



朱学松 (共同第一作者) 男, 2004年5月出生于山西省大同市, 现就读于华中农业大学信息学院生物信息系, 主要研究方向为空间疾病致病机制解析、药物发现等。

E-mail: 1474452250@qq.com



曲恩锐 (共同第一作者) 男, 2004年11月出生于内蒙古自治区包头市, 现就读于华中农业大学信息学院生物信息系, 主要研究方向为空间疾病致病机制解析、药物发现、数学建模等。

E-mail: 3315690090@qq.com



朱玉锋 男, 2001年11月出生于江苏省泰州市, 毕业于华中农业大学信息学院生物信息系, 主要研究方向为空间生物医学、药物筛选等。

E-mail: neverbo1101@outlook.com



全源 (通信作者) 女, 1990年7月出生于广西壮族自治区柳州市, 现为华中农业大学信息学院副研究员, 博士生导师, 主要研究方向为空间生物信息学、空间医学、药物发现等。

E-mail: quanyuan@mail.hzau.edu.cn