基于知识图谱嵌入的阿尔茨海默病药物重定位研究

卢艳峰,杨思瀚,莫鸿仪,侯凤贞*

(中国药科大学理学院,医药大数据与人工智能研究院,南京 211198)

摘 要 阿尔茨海默病(Alzheimer's disease,AD)给社会带来了巨大的医疗和经济负担,寻找和发现其治疗药物有着重大的研究意义。本研究采用知识图谱嵌入在公开的药物再利用知识图谱(drug repurposing knowledge graph,DRKG)上研究了AD的药物重定位。首先,利用4种知识图谱嵌入模型,即TransE、DistMult、ComplEx和RotatE在DRKG上学习实体和关系的嵌入向量;随后使用3种经典的知识图谱评估指标评估和比较了这些模型的性能和学习到的嵌入向量的质量;根据评估比较的结果,选择利用RotatE模型进行链接预测,确定了16种有可能用于AD治疗的药物,其中谷胱甘肽、氟哌啶醇、辣椒素、槲皮素、雌二醇、葡萄糖、双硫仑、腺苷、帕罗西汀、紫杉醇、格列本脲、阿米替林已被前人的研究证实对于AD有潜在的治疗作用。研究结果表明,基于知识图谱嵌入的药物重定位研究有望为AD药物发现提供新的思路和方法,RotatE模型可以有效地整合DRKG的多源信息,进而很好地完成了AD药物重定位任务。本研究的源代码可以从https://github.com/LuYF-Lemon-love/AD-KGE获得。

关键词 药物重定位;阿尔茨海默病;知识图谱;知识图谱嵌入;知识图谱补全

中图分类号 TP391;R971 **文献标志码** A **文章编号** 1000-5048(2023)03-0344-11

doi:10.11665/j.issn.1000-5048.2023040305

引用本文 卢艳峰,杨思瀚,莫鸿仪,等.基于知识图谱嵌入的阿尔茨海默病药物重定位研究[J].中国药科大学学报,2023,54(3):344-354.

Cite this article as: LU Yanfeng, YANG Sihan, MO Hongyi, *et al*. Drug repurposing for Alzheimer's disease using knowledge graph embedding[J]. *J China Pharm Univ*, 2023, 54(3): 344-354.

Drug repurposing for Alzheimer's disease using knowledge graph embedding

LU Yanfeng, YANG Sihan, MO Hongyi, HOU Fengzhen*

Institute of Medical Big Data and Artificial Intelligence, School of Science, China Pharmaceutical University, Nanjing 211198, China

Abstract Alzheimer's disease (AD) has brought to us huge medical and economic burdens, and so discovery of its therapeutic drugs is of great significance. In this paper, we utilized knowledge graph embedding (KGE) models to explore drug repurposing for AD on the publicly available drug repurposing knowledge graph (DRKG). Specifically, we applied four KGE models, namely TransE, DistMult, ComplEx, and RotatE, to learn the embedding vectors of entities and relations on DRKG. By using three classical knowledge graph evaluation metrics, we then evaluated and compared the performance of these models as well as the quality of the learned embedded vectors. Based on our results, we selected the RotatE model for link prediction and identified 16 drugs that might be repurposed for the treatment of AD. Previous studies have confirmed the potential therapeutic effects of 12 drugs against AD, i.e., glutathione, haloperidol, capsaicin, quercetin, estradiol, glucose, disulfire, adenosine, paroxetine, paclitaxel, glybride and amitriptyline. Our study demonstrates that drug repurposing based on KGE may provide new ideas and methods for AD drug discovery. Moreover, the RotatE model effectively integrates multi-source information of DRKG, enabling promising AD drug repurposing. The source code of this paper is available at https://github.com/LuYF-Lemon-love/AD-KGE.

Key words drug repurposing; Alzheimer's disease; knowledge graph; knowledge graph embedding; knowledge graph completion

收稿日期 2023-04-03 * **通信作者** Tel:025-86185121 E-mail:houfz@cpu.edu.cn

阿尔茨海默病（Alzheimer's disease，AD）是一种常见的神经退行性疾病，无法治愈且不可逆转[1]，其特征是伴有神经精神症状的渐进性严重痴呆[2]。据报道，2021年我国60岁以上人群中有983万例AD患者[3]；且另一份研究报告称，到2050年，我国AD患者的治疗费用将高达18 871.8亿美元[4]。AD对社会经济造成了巨大的负担，开发AD的治疗药物势在必行。

然而，研发一款新药往往用时漫长、耗资巨大[5]。药物重定位，又可以称为“老药新用”，指的是从获批准的临床药物中发现新适用的病症或新用途的方法[6]。该方法具有低成本、高效率的特点，在突发性疾病和罕见病方面优势更为突出[7]。近年来，药物重定位得到了迅速发展，领域内已经出现了很多用于探索药物和疾病之间关系的方法[8]。其中，知识图谱（knowledge graph，KG）就是实现药物重定位的一个重要举措[9]。

KG是一种基于拓扑结构图存储知识的数据库。知识中的具体事物和抽象概念在KG中被表示为实体，实体之间的联系被表示为关系，进而知识被表示成格式为（头实体，关系，尾实体）的三元组。KG是一个由大量的三元组组成的有向图结构，图中的节点表示实体，边表示实体间的关系。

然而，许多KG规模巨大，如药物重定位知识图谱（drug repurposing knowledge graph，DRKG）[10]包含97 238个实体和5 874 261个三元组。因此，常采用知识图谱嵌入（knowledge graph embedding，KGE）技术将实体和关系表示成低维稠密向量，进而将KG建模成低维向量空间。在过去几年中，研究人员提出了很多KGE模型，如TransE[11]、DistMult[12]、ComplEx[13]和RotatE[14]等，来学习实体和关系嵌入向量。KGE模型能够利用各自对应的模型假设进行链接预测进而推测三元组中缺失的实体。因此使用KG进行药物重定位研究，关键就在于如何选择并训练一个合适的KGE模型，然后基于它进行“疾病”实体和“药物”实体之间缺失关系的预测。

近年来，研究人员提出了很多利用KG进行药物重定位的方法。Zeng等[15]建立了一个1 500万个三元组的综合KG，包括药物、基因、疾病、药物副作用4种实体以及它们之间的39种关系，然后利用RotatE学习实体和关系的表示，进而确定了41种针对COVID-19的治疗药物。Zhang等[16]提出了一种基于神经网络和文献发现的方法，首先利用PubMed和其他专注COVID-19的研究文献构建了一个生物医学KG，然后利用多种KGE模型预测COVID-19的候选治疗药物，并利用发现模式解释了KGE预测的合理性。目前也有研究人员利用KGE模型研究帕金森病的药物重定位，并取得了不错的效果[17]。

Wang等[9]首次在基于KG的AD药物重定位中做出了有意义的尝试。他们构建了一个包含4种实体（药物、靶标、酶、载体）和1种关系类型（药物-靶点）的药物靶点KG，然后利用DistMult学习了实体和关系的嵌入表示，最终通过载脂蛋白E作为靶点寻找治疗AD的药物。但是，该KG只利用了单一的药物靶点信息，忽略了大量对AD药物重定位有用的其他信息，如基因、药物副作用和症状等。Nian等[1]则从文献中利用规则和BERT分类器保留了与AD相关的三元组，构建了一个KG，然后利用TransE、DistMult和ComplEx模型来研究AD与化学物质、药物和膳食补充剂之间的关系。Nian等[1]的工作因为聚焦于AD，而忽略了很多其他疾病的信息。但是AD患者也可能会出现一些精神症状，而其他精神疾病的实体也许会对AD的药物重定位很有帮助。

因此，本研究采用KGE模型，在DRKG上研究了AD药物重定位。图1展示了基于KGE进行AD药物重定位研究的工作流程。首先，利用多种KGE模型（TransE、DistMult、ComplEx和RotatE）在DRKG上学习实体和关系的嵌入向量，通过3种经典的KG评估指标评估了4种KGE模型；然后，在整个KG上重新训练KGE模型，并利用多种嵌入向量分析手段评估了模型学习到的嵌入向量的质量；最终，根据KGE模型的评估结果选择RotatE作为最终的药物重定位模型，找到了16种有可能用于AD治疗的药物，并通过查找支撑文献证明了研究方法的有效性。

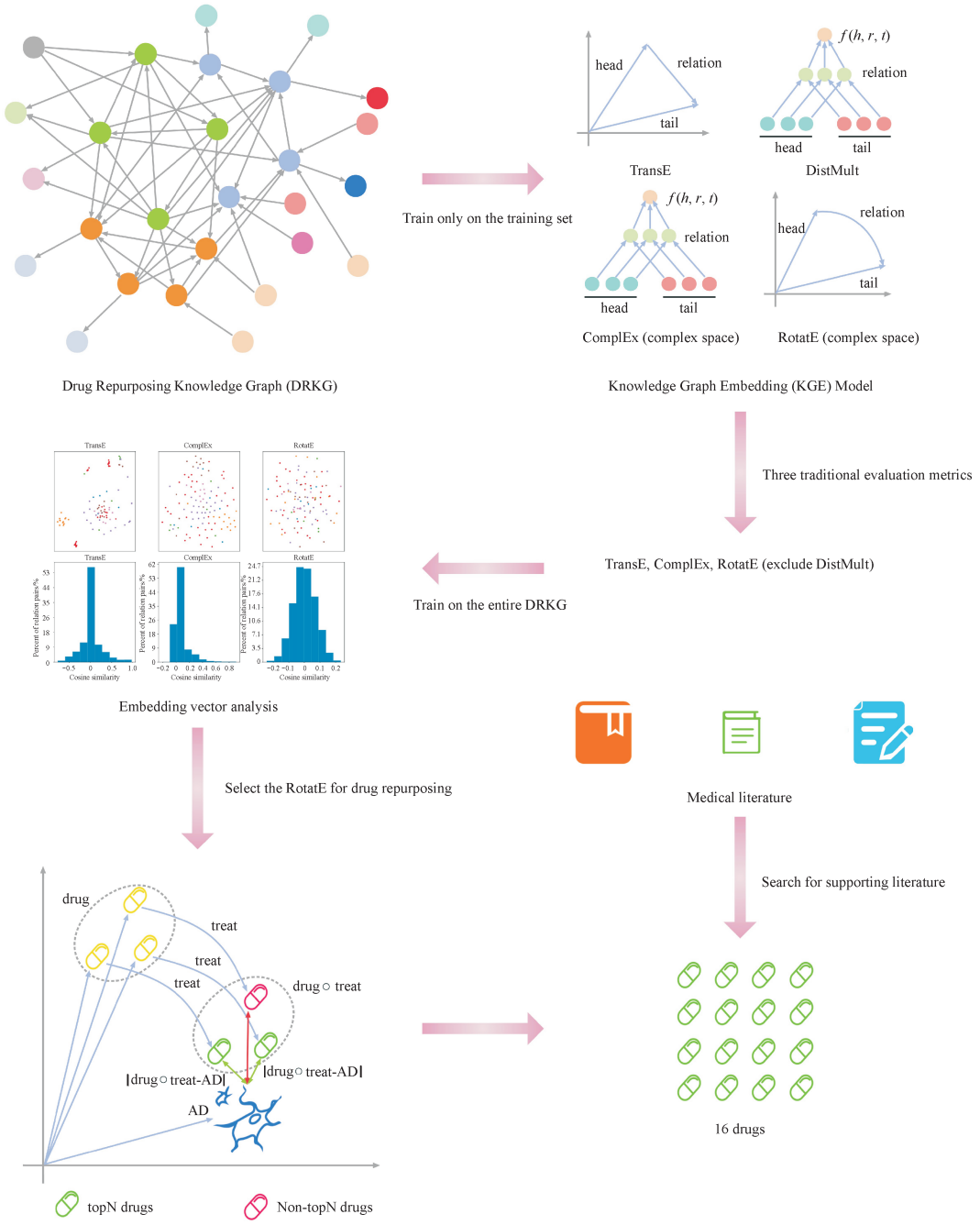

Figure 1 Diagram illustrating the workflow of our approach

1 方 法

1.1 数 据

DRKG[10]是一个涉及基因、药物、疾病、生物过程、副作用和症状的综合生物KG,包括来自DrugBank、Hetionet、GNBR、String、IntAct和DGIdb等6个现有数据库的信息,以及从最近发表的COVID-19出版物(截至2020年3月22日)中收集的数据(后文标记为bioarx数据库)。它有属于13种实体类型的97 238个实体;以及属于107种关系类型的5 874 261个三元组。DRKG使用“实体类型::ID”的格式表示一个实体,如“Disease::MESH:D000544”,其中“Disease”是实体类型,“MESH:D000544”是ID;使用“数据源名::关系名::头实体类型:尾实体类型”的格式表示关系,如“DRUGBANK::treats::Compound:Disease”,其中“DRUG-

BANK”是数据源名，“treats”是关系名，“Compound”是头实体类型，“Disease”是尾实体类型。

1.2　KGE模型基本原理

为了实现在DRKG上学习实体和关系的嵌入向量，考虑到算力限制，仅研究和对比了4种经典且具有线性时间复杂度的KGE模型，即TransE[11]、DistMult[12]、ComplEx[13]、RotatE[14]。首先描述一下常见的符号，使用h表示头实体，r表示关系，t表示尾实体；h,r,t分别表示相应的嵌入向量。在利用KGE模型来推断现有KG的缺失关系，从而达到补全KG的任务中，KG通常被标记为T，是一组格式为(h,r,t)三元组的集合，其中h，$t\in$E，$r\in$R，E是KG的实体集合，R是KG的关系集合。KGE模型一般都具有一个度量(h,r,t)成立概率的评分函数，该评分函数是特定KGE模型对KG的建模假设[14]。

1.2.1　TransE模型基本原理

TransE[11]是一个代表性的平移模型，它假设实体和关系属于同一向量空间R^d，d是向量空间的维度。关系被建模为实体向量的平移，如果三元组(h,r,t)成立，那么$h+r\approx t$，即t应该是$h+r$最近的实体向量；如果不成立，$h+r$应该远离t。TransE只能建模1对1的关系类型；但是从另一种关系分类角度，它能捕获反对称、反转和组成三种关系但不能捕获对称关系[14]。TransE的评分函数如公式(1)所示。

$$f(h,r,t)=-\left\|h+r-t\right\|_{L_1/L_2} \tag{1}$$

如公式(1)所示，TransE依据距离函数(L_1范数和L_2范数)选择的不同有两个变体分别为TransE_l1和TransE_l2。

1.2.2　DistMult模型的基本原理

DistMult[12]是一个双线性模型，它为每一种关系提供了一个对角矩阵来建模实体之间的交互进而捕获KG的潜在语义。DistMult也假设实体和关系属于同一向量空间R^d，其评分函数如公式(2)所示。

$$f(h,r,t)=h^{\mathrm{T}}\mathrm{diag}(r)t \tag{2}$$

其中，diag(r)是关系r的对角矩阵。

1.2.3　ComplEx模型的基本原理

由于DistMult[12]使用的是对角矩阵，因此仅仅能捕获对称关系。为了捕获反对称和反转关系，ComplEx[13]将向量空间从实数域扩展到复数域，极大地提升了模型的表现能力。ComplEx假设实体和关系属于同一复数向量空间C^d，其评分函数如公式(3)所示。

$$f(h,r,t)=\mathrm{Re}(h^{\mathrm{T}}\mathrm{diag}(r)\bar{t}) \tag{3}$$

其中，Re表示复数的实部，$\bar{t}$表示t的共轭。

1.2.4　RotatE模型的基本原理

受到TransE和欧拉恒等式的启发，RotatE[14]将头实体和尾实体映射到复数向量空间，即当$h,t,r\in C^d$，$|r_i|=1$，将关系建模为从头实体到尾实体的逐元素旋转。RotatE模型能够捕获对称、反对称、反转和组成4种类型关系，其评分函数如公式(4)所示。

$$f(h,r,t)=-\left\|h\circ r-t\right\|^2 \tag{4}$$

其中，$\circ$表示哈达玛积。

1.2.5　优化

本研究使用最大间隔方法训练模型，以最小化正确三元组的排名[11]，其损失函数如公式(5)所示。

$$L=\sum_{(h,r,t)\in\mathbf{T}}\sum_{(h',r,t')\in\mathbf{T}^-}\max(0,\gamma-f(h,r,t)+f(h',r,t')) \tag{5}$$

其中，$\gamma>0$是正负例三元组得分的间隔距离。$\mathbf{T}$是正例三元组集合，$\mathbf{T}^-$是负例三元组的集合，如公式(6)所示，它是通过破坏原有三元组中的实体和关系得到的[18]。

$$\mathbf{T}^-=\mathbf{E}\times\mathbf{R}\times\mathbf{E}-\mathbf{T} \tag{6}$$

1.3　KGE模型的评估

1.3.1　经典评估

KGE模型可以通过链接预测技术预测KG中缺失的三元组，即给定$(h,r,?)$预测缺失的尾实体t，或者给定$(?,r,t)$预测缺失的头实体h。可以通过链接预测给出正确实体的排名。常使用3种经典指标来评估链接预测的性能：正确实体评分函数的平均排名(mean rank，MR)[11]，正确实体评分函数的平均倒数排名(mean reciprocal rank，MRR)[14]和正确实体评分函数的前N的比例即前N命中率Hits@N($N=1,3,10$)[11]。

如果用rank_h和rank_t分别表示预测正确头实体和尾实体的排名，$\mathbf{T}$表示需要评估的三元组集合，那么MR、MRR和Hits@N的具体的计算方法分别如公式(7)(8)和(9)所示。

$$\mathrm{MR}=\frac{1}{2|\mathbf{T}|}\sum_{(h,r,t)\in\mathbf{T}}\mathrm{rank}_h+\mathrm{rank}_t \tag{7}$$

$$\mathrm{MRR}=\frac{1}{2|\mathbf{T}|}\sum_{(h,r,t)\in\mathbf{T}}\frac{1}{\mathrm{rank}_h}+\frac{1}{\mathrm{rank}_t} \tag{8}$$

$$\mathrm{Hits@N}=\frac{1}{2|\mathbf{T}|}\sum_{(h,r,t)\in\mathbf{T}}\mathrm{I}\left[\mathrm{rank}_h\le N\right]+\mathrm{I}\left[\mathrm{rank}_t\le N\right] \tag{9}$$

在公式(9)中，如果条件为真，I[*]等于1，否则等于0。从式公式(7)(8)和(9)可知，对于相同

的T,MR越小,代表正确实体的排名越靠前,说明链接预测越精确;MRR和Hits@N越大,代表正确实体的排名越靠前,说明链接预测越精确。

1.3.2 嵌入评估 由于DRKG结合了来自不同数据源的信息,本研究通过嵌入评估来定性验证KGE模型是否生成了有意义的实体和关系嵌入。理想的情况是,KGE模型能够学习到不同关系嵌入向量的差异之处和相同类型实体的相似之处。

本研究首先采用t分布随机近邻嵌入(t-distributed stochastic neighbor embedding, t-SNE)[19]将关系嵌入向量进行降维并可视化。DRKG共有来源于7个数据库的107种关系类型,如果相同数据来源的关系向量在可视化图中越分散,就说明KGE模型越能学习到不同关系嵌入向量的差异之处,即使它们来源于同一数据库。进一步地,如公式(10)所示,本研究还使用了余弦相似度来计算DRKG的关系嵌入向量对之间的相似性,并通过对比相似度分布的直方图来评估各种KGE模型。不同关系嵌入向量的相似度越低,表示KGE模型越能捕捉到不同关系嵌入向量的差异。使用这样的KGE模型进行链接预测的效果也就越好。

$$\text{similarity} = \cos(\theta) = \frac{a \cdot b}{\|a\|\|b\|} = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n}(a_i)^2} \times \sqrt{\sum_{i=1}^{n}(b_i)^2}} \quad (10)$$

在公式(10)中,a_i和b_i分别表示向量a和b的第i个分量,余弦相似度的取值范围为[−1,1],−1表示两个向量方向相反,1表示方向相同,0表示相互独立。

接下来使用主成分分析将实体嵌入向量降到30维[19],并利用t-SNE将其将降到2维空间进行可视化。使用主成分分析的原因在于本研究对象中共有97 238个实体,数量众多,若直接利用t-SNE降维和可视化,可能会引入大量噪声。DRKG一共有13种实体类型,相同类型的实体在可视化图中越聚集,KGE模型对实体嵌入的效果就越好。

1.4 AD药物重定位

使用KGE模型做药物重定位时,将DrugBank中被FDA批准的药物作为候选药物(相对分子质量≥250,共8 104个),它们构成了头实体集合。选择DRKG中所有治疗关系作为链接预测的关系,共有DRUGBANK::treats::Compound:Disease,GNBR::T::Compound:Disease,Hetionet::CtD::Compound:Disease 3种,其中treats、T、CtD分别是DrugBank数据库、GNBR数据库、Hetionet数据库中的治疗关系。选择DRKG中全部AD实体作为尾实体集合,共有Disease::DOID:10652,Disease::MESH:C536599,Disease::MESH:D000544 3种,其中Disease::DOID:10652是来自Hetionet数据源的AD实体,Disease::MESH:C536599和Disease::MESH:D000544是被映射到MESH ID的AD实体(其中Disease::MESH:C536599是无神经纤维缠结AD的实体)。将上面实体和关系集合进行格式为(h,r,t)排列组合(总共8 104 × 3 × 3 = 72 936种可能),然后计算所有组合评分函数的得分,最后选择得分前N的药物作为AD的治疗药物,其中N的值取决于不同KGE模型在测试集上的MR指标结果。

1.5 实验设置

将DRKG的三元组按照90%、5%、5%的比例划分为训练集、验证集和测试集,分别为5 286 834个、293 713个和293 714个。

综合5个经典的KGE评估指标(即MR、MRR、Hits@1、Hits@3、Hits@10)的综合表现,在验证集上利用网格搜索所有模型的超参数(TransE_l1、TransE_l2、DistMult、ComplEx和RotatE)。所有模型的训练批处理大小和每个正例三元组使用的负例三元组的数量分别固定为4 096和256,学习率(learning rate, lr)则都从{0.01,0.05,0.1}中选择。由于RotatE模型实体维度是超参数嵌入维度(the embedding dimension, hidden_dim)的2倍,本研究选择将RotatE模型的hidden_dim固定为200,其他模型的hidden_dim则从{200, 400}中选择。对于超参数γ,TransE_l1、TransE_l2和RotatE从{6,12,18}中选择,而DistMult、ComplEx模型则从{50,125,200}中进行选择。

研究利用Zheng等[20]开发DGL-KE工具包实现。

2 结 果

2.1 KGE模型的经典评估

表1列出了在KG补全任务中,4种KGE模型在测试集上的结果。如表1所示,对于MR指标,TransE两种变体分别取得了最优结果60.83和次

优结果 62.64；对于 MRR 指标，ComplEx 取得了最优结果 0.621，RotatE 次之为 0.614；对于 Hits@1 指标，ComplEx 取得了最优结果为 0.537，RotatE 次之为 0.515；对于 Hits@3 和 Hits@10，RotatE 取得了最优结果分别为 0.681 和 0.780，ComplEx 取得了次优结果分别为 0.673 和 0.768。而 DistMult 在 3 种指标上都没有取得最优和次优结果。

Table 1　Traditional evaluation results of the KGE models

Model	MRR	MR	Hits@1	Hits@3	Hits@10
TransE_l1	0.530	62.64	0.412	0.606	0.740
TransE_l2	0.437	**60.83**	0.302	0.515	0.693
DistMult	0.484	105.55	0.401	0.515	0.643
ComplEx	**0.621**	112.74	**0.537**	0.673	0.768
RotatE	0.614	63.51	0.515	**0.681**	**0.780**

The best results are in **bold** and the second best results are in underline

各个模型超参数的最佳配置是：对于 TransE_l1，hidden_dim=400，γ=18，lr=0.05；对于 TransE_l2，hidden_dim=400，γ=12，lr=0.1；对于 DistMult，hidden_dim=400，γ=50，lr=0.1；对于 ComplEx，hidden_dim=400，γ=50，lr=0.1；对于 RotatE，hidden_dim=200，γ=18，lr=0.05。

鉴于 DistMult 模型在经典评估中并不出色的表现，本研究仅选择 TransE_l1、TransE_l2、ComplEx 和 RotatE 模型，利用最佳超参数，重新在整个 DRKG 上进行训练，并进一步进行模型的嵌入评估和 AD 药物重定位。

2.2　KGE 模型的嵌入评估

图 2 分别展示了 TransE_l1、TransE_l2、ComplEx 和 RotatE 的关系嵌入向量在 2D 空间的可视化图。图中每一个圆点代表 DRKG 中一种关系类型，因此共有 107 个圆点；相同颜色的圆点代表关系来自相同的 DRKG 中相同的数据库。从图 2-A、2-B 和 2-C 中可以看出，TransE_l1、TransE_l2 和 ComplEx 的关系嵌入向量出现不同程度的同数据源聚集现象，如虚线框中标注出来的代表 DGIdb 数据源的橙色点；而 RotatE 的关系嵌入向量广泛地分布在 2D 的空间中，即便来自相同源数据集的关系都没有出现聚集的现象，可以说，RotatE 更好地学习到了各个关系本身的差异，受数据源的影响较小。

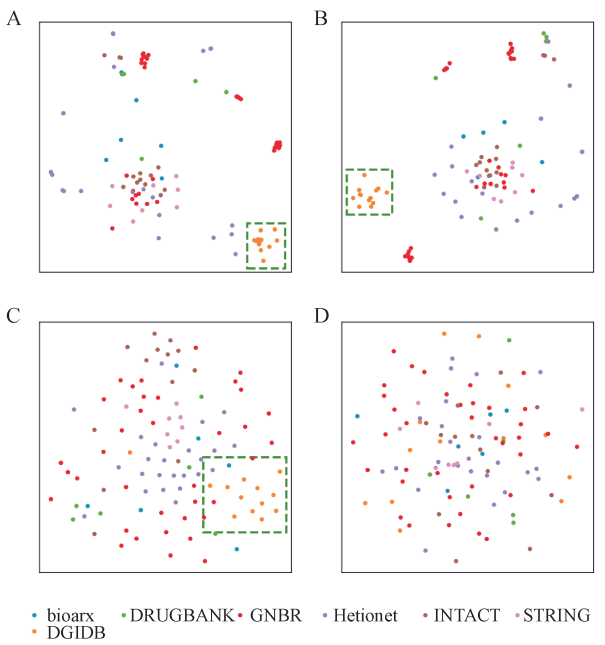

Figure 2　Distribution of relation embeddings in 2D euclidean space for 4 models

A: TransE_l1 embeddings; B: TransE_l2 embeddings; C: ComplEx embeddings; D: RotatE embeddings

图 3 显示了 TransE_l1、TransE_l2、ComplEx 和 RotatE 的不同关系嵌入向量对之间的余弦相似度分布直方图。对于 TransE_l1，相似度分布在[-0.873，0.977]范围内，其中约有 7% 相似度大于 0.50 的关系对；TransE_l2 与 TransE_l1 类似，也存在着 5% 相似度大于 0.50 的关系对。ComplEx 模型的相似度分布在[-0.208，0.908]范围内，存在 1% 相似度大于 0.50 的关系对。相比而言，RotatE 模型的相似度整体都较小，分布在[-0.241，0.233]的范围内。进一步地，本研究考察了包含并且只包含一种治疗关系的嵌入向量对之间余弦相似度的最大值，TransE_l1 为 0.917，TransE_l2 为 0.841，ComplEx 为 0.225，RotatE 为 0.180。这就说明对于 TransE_l1 和 TransE_l2，存在着与治疗关系非常相似的其他类型的关系向量，这很可能会干扰链接预测的结果。而对于 RotatE 模型，治疗关系向量与其他类型的关系向量之间的相似度最高也仅为 0.180，说明治疗关系与其他类型的关系有着极小的相似性，在链接预测时，不易受到其他关系类型的影响。

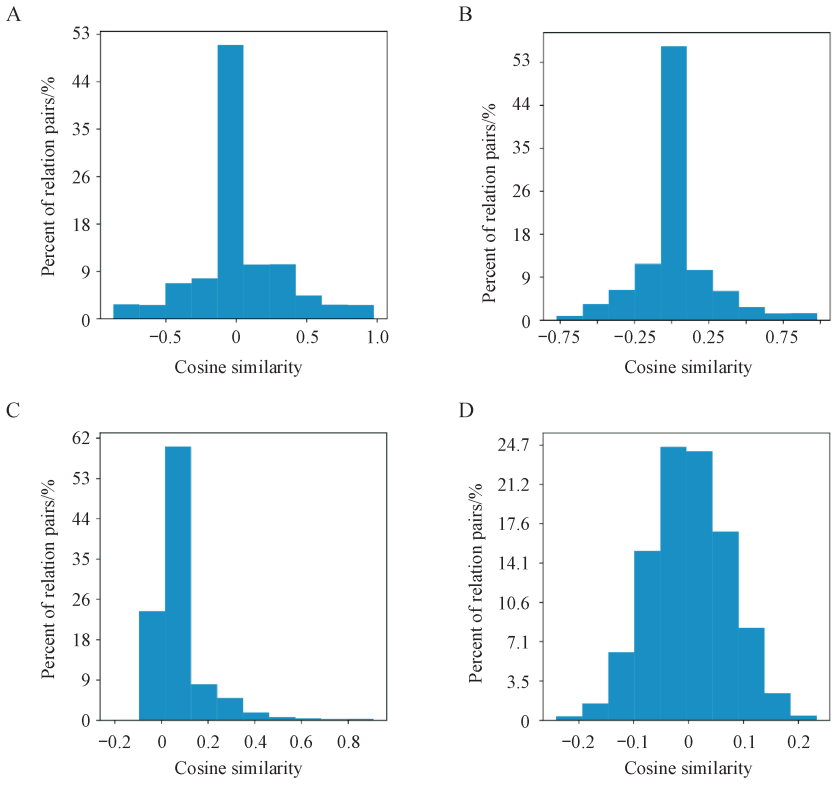

Figure 3 Histogram of cosine similarity between relations for 4 models

A: TransE_l1 embeddings; B: TransE_l2 embeddings; C: ComplEx embeddings; D: RotatE embeddings

图4是TransE_l1、TransE_l2、ComplEx和RotatE的实体嵌入2D空间的可视化图，每一个圆点代表了一个实体，不同的颜色代表不同的实体类型。用蓝色和蓝绿色箭头指出了药物重定位3个AD实体，蓝色箭头指向的是Disease::DOID:10652实体，它是来自Hetionet数据源的AD实体。从图4中可以看到，在所有模型中，相同类别的实体正如期望的那样聚集到了一起，其中TransE_l1和RotatE的结果要优于另外2个模型。4个模型都将来自Hetionet数据源的AD实体和来自MESH ID空间中两种AD实体区分开来。2种MESH ID空间的AD实体在TransE_l1、TransE_l2和RotatE的2D空间中距离很近，但在ComplEx的2D空间中这两种实体还有较大距离。

2.3 AD药物重定位

综合KGE的经典评估和嵌入评估结果，本研究使用RotatE模型作为AD药物重定位的最终模型。在得分前10的药物列表中，只有第9名的药物没有被DRKG标注为对AD疾病实体有治疗关系，说明该方法能够正确表达DRKG中原有的三元组。

由于RotatE的MR指标结果是63.51，因此将得分前50、且没有被DRKG标注为对AD疾病实体有治疗关系的药物作为重定位得到的AD候选药物。考虑到其中得分排名在第23的西布曲明已退市[21]，因此最终确定了16种药物作为本研究的结果。表2列出了这些药物在RotatE模型中排名、在DRKG中的名称以及文献中提及的该药物与AD的关系。从表2中可以看到，其中的12种药物，即谷胱甘肽、氟哌啶醇、辣椒素、槲皮素、雌二醇、葡萄糖、双硫仑、腺苷、帕罗西汀、紫杉醇、格列本脲、阿米替林已被前人的研究证实对于AD有潜在的治疗作用，也从另一方面证实了本研究所训练的RotatE模型的正确性和有效性。而对于另4种药物，即可卡因、胆固醇、星形孢菌素、可的松，暂未发现对AD有直接治疗作用的报道。

3 讨 论

本研究利用KGE模型研究了AD的药物重定位。先采用4种不同的KGE模型来学习DRKG的

Figure 4　Distribution of entity embeddings in 2D euclidean space for 4 models

A: TransE_l1 embeddings; B: TransE_l2 embeddings; C: ComplEx embeddings; D: RotatE embeddings

实体和关系的嵌入向量表示，通过比较确定使用 RotatE 模型基于链接预测技术发现 AD 的治疗药物。研究结果表明，RotatE 能够有效整合 DRKG 的多源信息，完成 AD 药物重定位任务：共确定了 16 种可重新利用的药物，其中 12 种已被前人研究证实对于 AD 的治疗有着潜在的积极意义。

本研究使用的数据集是涉及 13 种实体和 107 种关系、包含 5 874 261 个三元组的 DRKG。相比于仅利用单一药物靶点相互作用或利用单一疾病相关三元组而构建的 KG，DRKG 包含了各种各样的生物信息，会使 AD 重定位的结果更加全面。Wang 等[9]基于单一药物靶点相互作用构建的 KG 并进行 AD 药物重定位，得到的候选药物是诸如锌、铜、银、氯化锌、醋酸锌、硫酸锌等金属或金属化合物。类似地，相比于 Nian 等[1]的研究结果，本研究的预测结果包括了不少治疗其他精神疾病的药物，如排名第 11 位的氟哌啶醇可用于治疗精神分裂症[34]、排名第 29 位的帕罗西汀可用于治疗重度抑郁症和恐慌症等[35]、排名第 48 位的阿米替林是一种抗抑郁药[36]。这些结果都表明，在药物重定位研究中，更应该使用大型多实体类型和多关系类型的 KG。

但是，在对诸如 DRKG 这种包含多实体类型和多关系类型的 KG 进行嵌入时，模型的训练与选择却是一个挑战。本研究通过使用多种 KGE 模型并使用多种评估方法来综合比较，发现与 Nian 等[1]的实验结果类似，DistMult 在经典评估实验结果中也表现不佳。这一点可能是因为 DistMult 仅能捕获对称关系，无法建模 DRKG 中的非对称关系（如 3 种治疗关系就是非对称关系）。而 RotatE 能很好地整合 DRKG 中来自多个数据源的信息，避免不同数据源的三元组集合相互独立而影响

Table 2 Candidate drugs obtained via drug repurposing

Rank	Drug name	Literature support
9	Glutathione	The beneficial effect of many nutrients on the course of AD has been demonstrated. These include: glutathione, polyphenols, curcumin, coenzyme Q10, vitamins B6, B12, folic acid, unsaturated fatty acids, lecithin, UA, caffeine and some probiotic bacteria[22]
11	Haloperidol	Haloperidol inactivates AMPK and reduces tau phosphorylation in a tau mouse model of Alzheimer′s disease[23]
13	Capsaicin	In Alzheimer′s disease, capsaicin reduces neurodegeneration and memory impairment[24]
16	Quercetin	Quercetin has demonstrated antioxidant, anti-inflammatory, hypoglycemic, and hypolipidemic activities, suggesting therapeutic potential against type 2 diabetes mellitus (T2DM) and Alzheimer′s disease (AD)[25]
17	Estradiol	Mounting evidence indicates that the neurosteroid estradiol (17β-estradiol) plays a supporting role in neurogenesis, neuronal activity, and synaptic plasticity of AD. This effect may provide preventive and/or therapeutic approaches for AD[26]
18	Glucose	Specifically, decreased O-GlcNAcylation levels by glucose deficiency alter mitochondrial functions and together contribute to Alzheimer′s disease pathogenesis[27]
20	Disulfiram	Identification of disulfiram as a secretase-modulating compound with beneficial effects on Alzheimer′s disease hallmarks[28]
21	Adenosine	Emerging evidence suggests adenosine G protein-coupled receptors (GPCRs) are promising therapeutic targets for Alzheimer′s disease[29]
23	Sibutramine	In October 2010, Sibutramine was withdrawn from U.S.[21]
29	Paroxetine	Paroxetine ameliorates prodromal emotional dysfunction and late-onset memory deficit in Alzheimer′s disease mice[30]
31	Cocaine	None
39	Paclitaxel	In addition to NSAIDs, an anticancer drug, paclitaxel, has considerable potential as an AD treatment[31]
41	Cholesterol	None
43	Glyburide	Our findings suggest that a pharmacologic approach to inhibit galanin in the brain, either by glibenclamide or pioglitazone might dramatically improve symptoms in Alzheimer′s disease[32]
44	Staurosporine	None
46	Cortisone	None
48	Amitriptyline	These results indicate that amitriptyline has significant beneficial actions in aged and damaged AD brains and that it shows promise as a tolerable novel therapeutic for the treatment of AD[33]

′None′ indicates no supporting literature found to date

AD药物重定位的效果。在RotatE得分前10的候选药物列表中，只有第9名的药物没有被DRKG标注为与AD疾病实体有治疗关系，可以认为RotatE很好地拟合了DRKG中的治疗关系三元组。16种候选药物中除了4种排名较靠后(31、41、44、46)的药物，其余药物都被文献证实可能是AD的潜在药物。虽然我们暂未发现可卡因、星形孢菌素和可的松对AD有直接治疗作用的文献报道，但是可卡因能够增加大脑中的可卡因-苯丙胺调节转录肽的表达水平，而这种肽能够缓解AD的临床症状[37]；星形孢菌素能够显著降低tau蛋白的磷酸化[38]；可的松也出现在了Nian等[1]通过预防关系进行AD药物重定位的结果中，并且可的松在他们的结果中取得了第1名的位置。这些结果表明，能够捕获对称、反对称、反转和组成4种类型关系的RotatE可以有效地整合DRKG的多源信息，进而很好地完成AD药物重定位任务。

本研究结果表明，基于大型的多实体类型和多关系类型的KG，如DRKG，进行药物重定位研究，有着可观的应用场景，可为药物研发人员提供有意义的参考信息。不过，DRKG没有将所有的疾病都映射到统一的ID空间，这可能会对药物重定位的效果产生一定的影响。未来，我们将研究实体对齐技术，以实现将多种数据源的实体映射到统一的命名空间中，进而使得KGE模型能学习到更好的嵌入向量。

References

[1] Nian Y, Hu XY, Zhang R, *et al*. Mining on Alzheimer′s diseases related knowledge graph to identity potential AD-related semantic triples for drug repurposing[J]. *BMC Bioinformatics*, 2022, **23**(Suppl 6): 407.

[2] Moya-Alvarado G, Gershoni-Emek N, Perlson E, *et al*. Neurodegeneration and Alzheimer′s disease (AD). What can proteomics

tell us about the Alzheimer′s brain[J]? *Mol Cell Proteomics*, 2016, **15**(2): 409-425.

[3] Ren RJ,Yin P,Wang ZH,*et al*.China Alzheimer′s disease report 2021[J].*J Diagn Concept Pract* (诊断学理论与实践),2021,**20**(4):317-337.

[4] Jia JP, Wei CB, Chen SQ, *et al*. The cost of Alzheimer′s disease in China and re-estimation of costs worldwide[J]. *Alzheimers Dement*, 2018, **14**(4): 483-491.

[5] Avorn J. The $2.6 billion pill: methodologic and policy considerations[J]. *N Engl J Med*, 2015, **372**(20): 1877-1879.

[6] Zhang YS, Yang ZJ, Bao XF, *et al*. Progress of clinical research on drug repurposing for Alzheimer′s disease[J]. *Chin J Med Chem* (中国药物化学杂志), 2022, **32**(5): 372-389.

[7] Wang CC, Li W, Shi ZX. Research progress on new use of old drugs[J]. *World Clin Drugs* (世界临床药物), 2021, **42**(8): 699-704.

[8] Zhang W, Gu F, Fu YK, *et al*. Progress in research on drug repositioning in new drug research and development[J]. *Anim Husb Vet Med* (畜牧与兽医), 2021, **53**(12): 123-127.

[9] Wang SD, Du ZZ, Ding M, *et al*. KG-DTI: a knowledge graph based deep learning method for drug-target interaction predictions and Alzheimer′s disease drug repositions[J]. *Appl Intell*, 2022, **52**(1): 846-857.

[10] Ioannidis VN. DRKG - drug repurposing knowledge graph for Covid-19[EB/OL].(2021-07-12)[2023-03-31].https://github.com/gnn4dr/DRKG/.

[11] Bordes A, Usunier N, Garcia-Durán A, *et al*. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. New York: ACM, 2013: 2787-2795.

[12] Yang BS,Yih WT,He XD,*et al*.Embedding entities and relations for learning and inference in knowledge bases[J]. *arXiv*, 2015: 1412.6575.

[13] Trouillon T,Welbl J,Riedel S,*et al*.Complex embeddings for simple link prediction[J].*arXiv*,2016:1606.06357.

[14] Sun ZQ,Deng ZH,Nie JY,*et al*.RotatE: knowledge graph embedding by relational rotation in complex space[J]. *arXiv*, 2019: 1902.10197.

[15] Zeng XX, Song X, Ma TF, *et al*. Repurpose open data to discover therapeutics for COVID-19 using deep learning[J]. *J Proteome Res*, 2020, **19**(11): 4624-4636.

[16] Zhang R, Hristovski D, Schutte D, *et al*. Drug repurposing for COVID-19 via knowledge graph completion[J]. *J Biomed Inform*, 2021, **115**: 103696.

[17] Li ZX. Repositioning drugs for Parkinson′s disease based on knowledge graph[J]. *Inf Technol Informatization* (信息技术与信息化), 2022(7): 28-32.

[18] Han X,Cao SL,Lv X,*et al*.OpenKE: an open toolkit for knowledge embedding[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations.Brussels,Belgium:Association for Computational Linguistics,2018:139-144.

[19] Maaten LVD, Hinton G. Visualizing data using t-SNE[J]. *J Machine Learn Res*,2008,**9**(86):2579-2605.

[20] Zheng Da, Song X, Ma C, *et al*. DGL-KE: training knowledge graph embeddings at scale[J].*arXiv*,2020:2004.08532.

[21] U. S. Food and Drug Administration.FDA drug safety communication: FDA recommends against the continued use of Meridia (sibutramine) [EB/OL]. (2018-02-06) [2023-04-03]. https://www.fda. gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-recommends-against-continued-use-meridia-sibutramine.

[22] Śliwińska S, Jeziorek M. The role of nutrition in Alzheimer′s disease[J]. *Rocz Panstw Zakl Hig*, 2021, **72**(1): 29-39.

[23] Koppel J, Jimenez H, Adrien L, *et al*. Haloperidol inactivates AMPK and reduces tau phosphorylation in a tau mouse model of Alzheimer′s disease[J]. *Alzheimers Dement*, 2016, **2**(2): 121-130.

[24] Pasierski M, Szulczyk B. Beneficial effects of capsaicin in disorders of the central nervous system[J]. *Molecules*, 2022, **27**(8): 2484.

[25] Zu GX, Sun KY, Li L, *et al*. Mechanism of quercetin therapeutic targets for Alzheimer disease and type 2 diabetes mellitus[J]. *Sci Rep*, 2021, **11**(1): 22959.

[26] Sahab-Negah S, Hajali V, Moradi HR, *et al*. The impact of estradiol on neurogenesis and cognitive functions in Alzheimer′s disease[J]. *Cell Mol Neurobiol*, 2020, **40**(3): 283-299.

[27] Huang CW, Rust NC, Wu HF, *et al*. Altered O-GlcNAcylation and mitochondrial dysfunction, a molecular link between brain glucose dysregulation and sporadic Alzheimer′s disease[J]. *Neural Regen Res*, 2023, **18**(4): 779-783.

[28] Reinhardt S, Stoye N, Luderer M, *et al*. Identification of disulfiram as a secretase-modulating compound with beneficial effects on Alzheimer′s disease hallmarks[J]. *Sci Rep*, 2018, **8**(1): 1329.

[29] Trinh PNH, Baltos JA, Hellyer SD, *et al*. Adenosine receptor signalling in Alzheimer′s disease[J]. *Purinergic Signal*, 2022, **18**(3): 359-381.

[30] Ai PH, Chen S, Liu XD, *et al*. Paroxetine ameliorates prodromal emotional dysfunction and late-onset memory deficit in Alzheimer′s disease mice[J]. *Transl Neurodegener*, 2020, **9**(1): 18.

[31] Lehrer S,Rheinstein PH.Transspinal delivery of drugs by transdermal patch back-of-neck for Alzheimer′s disease: a new route of administration[J].*Discov Med*,2019,**27**(146):37-43.

[32] Baraka A, ElGhotny S. Study of the effect of inhibiting galanin in Alzheimer′s disease induced in rats[J]. *Eur J Pharmacol*, 2010, **641**(2/3): 123-127.

[33] Chadwick W, Mitchell N, Caroll J, *et al*. Amitriptyline-mediated cognitive enhancement in aged 3 × Tg Alzheimer′s disease mice is associated with neurogenesis and neurotrophic activity

[J]. *PLoS One*, 2011, **6**(6): e21660.

[34] Feng JM. Clinical effect analysis of quetiapine combined with haloperidol in the treatment of schizophrenia in acute stage[J]. *Med Forum* (基层医学论坛), 2022, **26**(20): 37-39.

[35] Jiang YL. Research progress of Paroxetine combined with other therapies in the treatment of Major Depressive Disorder(MDD) [J]. *Chin J Conval Med* (中国疗养医学), 2021, **30**(9): 919-923.

[36] Kim L. A brief review of the pharmacology of amitriptyline and clinical outcomes in treating fibromyalgia[J]. *Biomedicines*, 2017, **5**(2): 24.

[37] Liu SC, Fu Q, Peng QH, *et al*. Research progress on the role and mechanism of CART peptide in central nervous system[J]. *J Nanchang Univ Med Sci* (南昌大学学报 医学版), 2022, **62**(5): 76-80.

[38] Gu GJ, Wu D, Lund H, *et al*. Elevated MARK2-dependent phosphorylation of tau in Alzheimer's disease[J]. *J Alzheimers Dis*, 2013, **33**(3): 699-713.

［**专家介绍**］侯凤贞，博士，教授，美国哈佛大学访问学者，江苏省“青蓝工程”优秀青年骨干教师。近年来，相继主持和参与多项国家/省级自然科学基金项目，主持多个横向课题。正在开展的研究主要集中在两个方面：一是通过对各种生物医学信号（如心电、脑电、功能磁共振信号）的分析来挖掘生理系统的内在机制，从而为临床应用，如疾病诊断、健康监测等提供参考；二是探索人工智能在大健康领域的应用场景，如药物重定位、睡眠的科学评估、心脏病的精准预测以及老年痴呆症的及早诊断等。以第一作者或通信作者身份在*Sleep*，*Progress in Neuropsychopharmacology & Biological Psychiatry*，*Sleep Medicine*，*Frontiers in Neuroscience*等国际学术期刊上发表研究论文20余篇。