基于知识图谱嵌入的阿尔茨海默病药物重定位研究

摘要

阿尔茨海默病是一种起病隐匿、多因素、进行性神经退行性疾病，痴呆表现为主要特征，给社会带来巨大医疗负担，但目前还没有特效药物。然而，传统的药物开发存在成本高周期长等问题，且药物安全性需要大量的时间验证，而药物重定位能够极大的缓解上面问题。本文采用知识图谱嵌入研究阿尔茨海默病的药物重定位。首先，利用 4 种知识嵌入模型对知识图谱进行表示学习；其次，使用多种评估指标评估了知识图谱嵌入模型的性能和学习到的嵌入向量的质量；最后，利用知识图谱嵌入模型进行链接预测得出 14 种治疗阿尔茨海默病的候选药物。除此之外，我们还通过查阅文献的方法证明了本文的研究方法能够有效的预测治疗阿尔茨海默病的药物，为研究人员提供了新的研究方法。本文的源代码可以从 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

1. 引言

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]。中国阿尔茨海默病报告 2021 显示我国 60 岁及以上人群中有 983 万例 AD 患者[3]，并且另一份研究报告称，我国 AD 患者的治疗费到 2050 年将高达 18871.8 亿美元[4]，这充分说明了 AD 给社会带来了巨大的经济负担。因此，AD 的治疗药物开发势在必行。

然而，成功研发一款新药至少花费 26 亿美元[5]和 10 年时间[6]，需要海量的金钱和时间成本。药物重定位，又可以称为“老药新用”，具体是指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。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。近年来，药物重定位发展迅速，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。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[6]。

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 都非常巨大，如药物再利用知识图谱（Drug Repurposing Knowledge Graph，DRKG）[7]包含 97238 个实体和 5874261 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[8]、DistMult[9]、ComplEx[10]和 RotatE[11]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，本质上就是使用

KGE 模型进行“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知识图谱进行药物重定位的方法。Zeng 等[12]建立了一个 1500 万个三元组的综合知识图谱，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[13]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知识图谱，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[14]。

Wang 等[6]提出了一种基于知识图谱的深度学习方法进行 AD 药物重定位。首先，利用 DistMult 学习了预先构建的阳性药物靶点对知识图谱的实体和关系的嵌入表示，然后利用一个 Conv-Conv 模块来提取药物-靶点对的特征，提取到的特征被传入到一个全连接网络进行二分类，最终通过载脂蛋白 E 作为靶点寻找治疗 AD 的药物。Nian 等[1]从文献中构建一个知识图谱，利用 TransE、DistMult 和 ComplEx 预测有助于 AD 治疗或预防的候选物质，以研究 AD 与化学物质、药物和膳食补充剂之间的关系，进而确定预防或延缓神经退行性进展的机会。

本文采用 KGE 模型研究了 AD 药物重定位。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 种经典的知识图谱嵌入评估指标评估了 4 种 KGE 模型；然后，在整个知识图谱上重新训练 KGE 模型，并利用多种嵌入向量分析手段评估模型学习到的嵌入向量的质量；最终，根据前面 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 种治疗 AD 的候选药物。

2. 方法

2.1. 数据

DRKG[7]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知识图谱，包括来自 DrugBank、Hetionet、GNBR、String、IntAct 和 DGIdb 等六个现有数据库的信息，以及从最近发表的 Covid19 出版物（截止到 2020 年 3 月 22 日）中收集的数据（被标记为 bioarx）。它有属于 13 种实体类型的 97238 个实体；以及属于 107 种关系类型的 5874261 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“DRUGBANK”是数据源名，“treats”是关系名，“Compound”是头实体类型，“Disease”是尾实体类型。

2.2. 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，本文仅研究和对比了四种经典且具有线性时间复杂度的 KGE 模型，即 TransE[8]、DistMult[9]、ComplEx[10]、RotatE[11]。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[11]。

2.2.1. TransE 模型基本原理

TransE[8]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$\mathbb{R}^d$，d 是向量空间的维度。关系 r 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 ≈ 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[11]。TransE 的评分函数为：

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

2.2.2. DistMult 模型的基本原理

DistMult[9]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$\mathbb{R}^d$，评分函数为：

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

其中，diag(r)是关系 r 的对角矩阵。

2.2.3. ComplEx 模型的基本原理

由于 DistMult[9]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[10]将向量空间从实数域扩展到复数域，极大的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$\mathbb{C}^d$，评分函数为：

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

其中，Real(∗)表示复数的实部，$\bar{t}$表示 t 的共轭。

2.2.4. RotatE 模型的基本原理

受到 TransE 和欧拉恒等式的启发，RotatE[11]将头实体和尾实体映射到复数向量空间，即当 $h, t \in \mathbb{C}^d, r \in \mathbb{C}^d, |r_i| = 1$，将关系 r 建模为从头实体 h 到尾实体 t 的逐元素旋转。RotatE 模型能够捕获对称、反对称、反转和组成四种类型关系，评分函数为：

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

其中，∘表示哈达玛积。

2.2.5. 优化

本文使用最大间隔方法训练模型，以最小化正确三元组的排名[8]，其损失函数如下：

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

其中，$\gamma>0$ 是正负例三元组得分的间隔距离。T是正例三元组集合，T^{-}是负三元组的集合，它是通过破坏原有三元组中的实体和关系得到的[15]：

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

2.3. KGE 模型的评估

2.3.1. 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。常使用三种经典指标来评估 KGE 模型的性能：正确实体评分函数的平均排名（Mean Rank，MR）[8]，正确实体评分函数的平均倒数排名（Mean Reciprocal Rank，MRR）[11]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）[8]。

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，T 表示需要评估的三元组集合，那么 MR 具体的计算方法为：

$$MR=\frac{1}{2|T|}\sum_{(h,r,t)\in T}rank_h+rank_t \tag{7}$$

MRR 具体计算方法为：

$$MRR=\frac{1}{2|T|}\sum_{(h,r,t)\in T}\frac{1}{rank_h}+\frac{1}{rank_t} \tag{8}$$

Hits@N 被计算为

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

其中如果条件为真，I[*]等于 1，否则等于 0。从式（7）、（8）、（9）可知，对于相同的 T，MR 值越小，代表正确实体的排名越靠前，说明 KGE 模型预测越精确；MRR 和 Hits@N 值越大，代表正确实体的排名越靠前，说明 KGE 模型预测越精确。

2.3.2. 嵌入评估

由于 DRKG 结合了来自不同数据源的信息，本文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。具体来说，我们希望 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。

首先采用 t 分布随机近邻嵌入（T-distributed Stochastic Neighbor Embedding，t-SNE）[16]将关系嵌入向量进行降维并可视化。DRKG 有 7 个数据来源，相同数据来源的关系在 2 维可视化图中倾向聚集在一起，因此相同数据来源的关系越分散，表明 KGE 模型效果越好。

由于本文的研究对象中有 97238 个实体，数量众多，直接利用 t-SNE 降维和可视化处理可能会引入大量噪声。因此首先使用主成分分析将实体嵌入向量降到 30 维[16]，然后再利用 t-SNE 将其将降到 2 维空间并进行可视化。DRKG 有 13 种实体类型，如果相同类型的实体聚集在一起，表明 KGE 模型效果越好。

然后，本文使用余弦相似性来计算 DRKG 的关系嵌入向量对之间的相似度，并通过对比相似度分布的直方图来评估各种 KGE 模型。如果不同关系的相似度越低，表明 KGE 模型的效果越好。

2.4. AD 药物重定位

本文使用上述 KGE 模型进行链接预测来寻找治疗 AD 的候选药物。图 1 展示了如何使用 TransE 模型进行 AD 药物重定位的基本原理。药物实体、AD 实体和治疗关系都被表示成了嵌入空间中的向量，现存治疗其他疾病的药物用黄色胶囊表示，是头实体 h；绿色和红色胶囊表示模型推测的初始候选治疗 AD 药物，是头实体平移后的结果向量 $h + r$；TransE 仅仅选择离 AD 实体距离最近的 N 种药物作为最终推荐药物，因此两个绿色胶囊表示的药物就是重定位得出的药物。

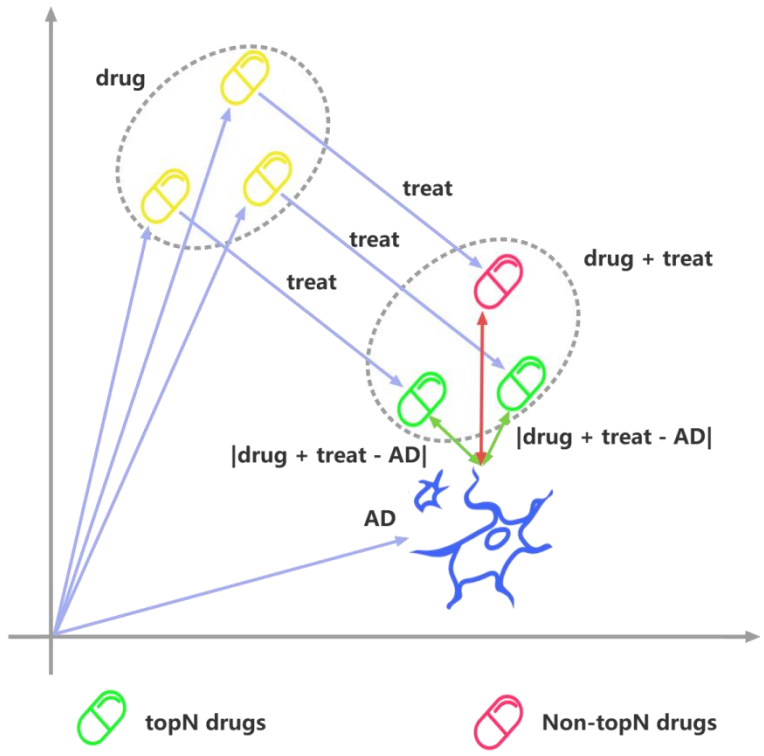

Figure 1 AD drug repurposing using TransE.

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（分子量 ≥250 道尔顿，共 8104 个），它们是头实体集合。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 三种，其中“treats”、“T”、“CtD”分别是 DrugBank 数据库、GNBR 数据库、Hetionet 数据库的治疗关系。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 三种，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 $3 \times 3 \times 8104 = 72936$ 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 的药物作为初始 AD 的治疗药物，N 根据 KGE 模型在测试集上的 MR 指标结果选择。

2.5. 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 5286834 个、293713 个和 293714 个。

综合上面 5 个指标（MR、MRR、Hits@1、Hits@3、Hits@10）的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），所有模型的训练批次大小 batch_size 和负采样大小 neg_sample_size 分别固定为 4096 和 256，从{0.01,0.05,0.1}中选择学习率 lr；由于 RotatE 模型实体维度是超参数嵌入维度 hidden_dim 的 2 倍，因此将其嵌入维度固定为 200，从{200,400}中选择其他模型的嵌入维度 hidden_dim；从{6,12,18}中选择 TransE_l1、TransE_l2 和 RotatE 的超参数γ，从{50,125,200}中选择 DistMult、

ComplEx 的超参数γ。

本文的实验是利用 Zheng 等[17]开发 DGL-KE 工具包实现的。

3. 结果

3.1. 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 The traditional evaluation results of the KGE model. The best results are in **bold** and the second best results are in underline.

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**

各个模型最佳配置是：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，本文仅选择 TransE_l1、TransE_l2、ComplEx 和 RotatE 模型，利用上面列出的超参数，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。

3.2. KGE 模型的嵌入评估

图 2A、2B、2C、2D 分别展示了 TransE_l1、TransE_l2、ComplEx 和 RotatE 的关系嵌入 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。每幅子图中的箭头指示了 AD 药物重定位 3 种治疗关系（“treats”、“T”、“CtD”）。从图 2A、2B 和 2C 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，如代表 DGIdb 数据源的橙色点；而 RotatE 的关系嵌入向量广泛的分布在 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。

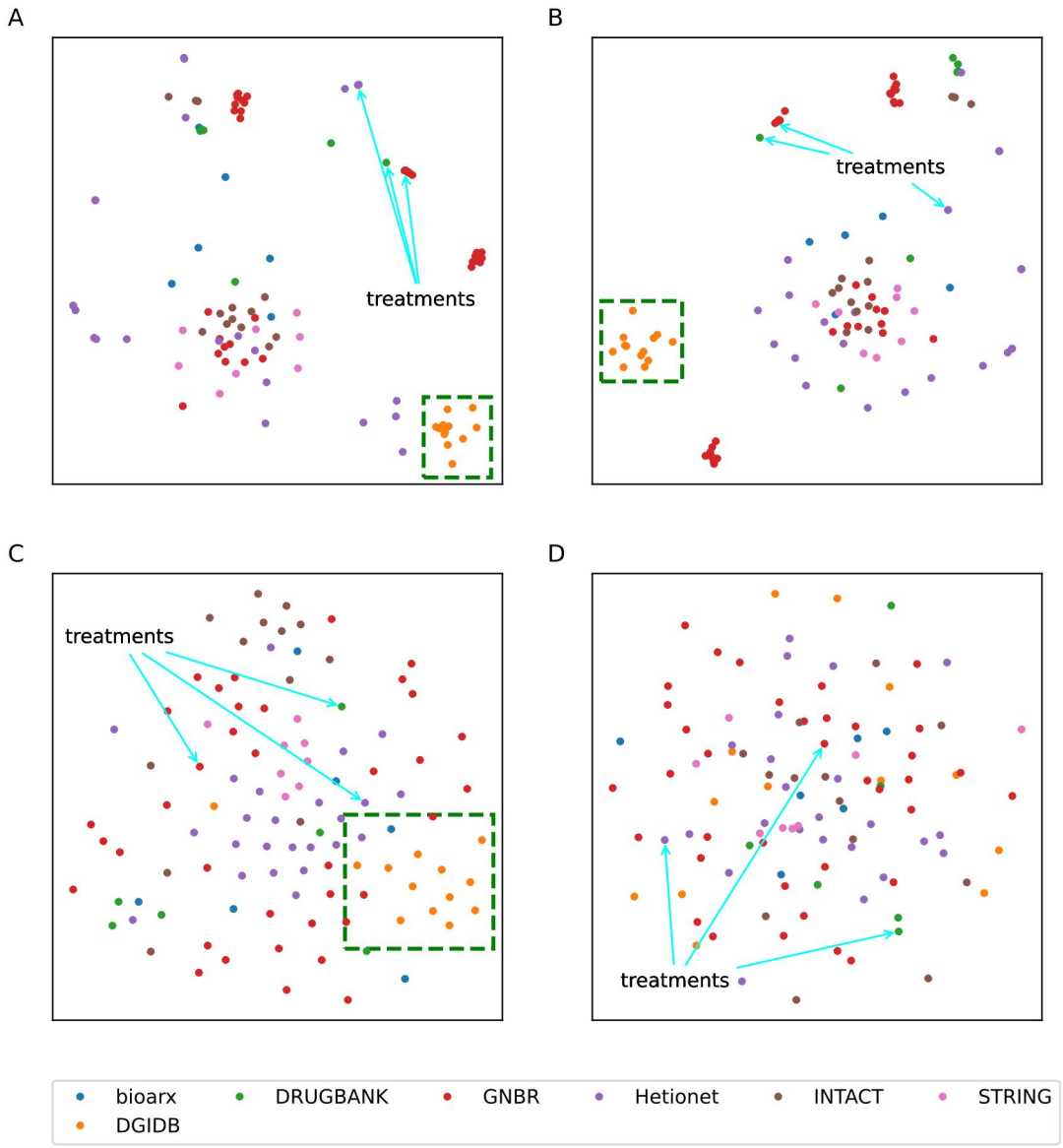

Figure 2 Distribution of relation embeddings in 2D euclidean space for 4 models. Subgraphs A, B, C and D are the results of TransE_l1, TransE_l2, ComplEx and RotatE respectively.

图 3A、3B、3C、3D 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。所有模型都可以观察到相同类别的实体正如期望的那样聚集到一起，其中 TransE_l1 和 RotatE 的结果要好于另外 2 个模型。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，而在 ComplEx 的 2D 空间中两种实体还有较大距离。4 个模型都将自 Hetionet 数据源的 AD 实体和另外两种 AD 实体区分开了。总体上各个模型学习到了实体类型信息。

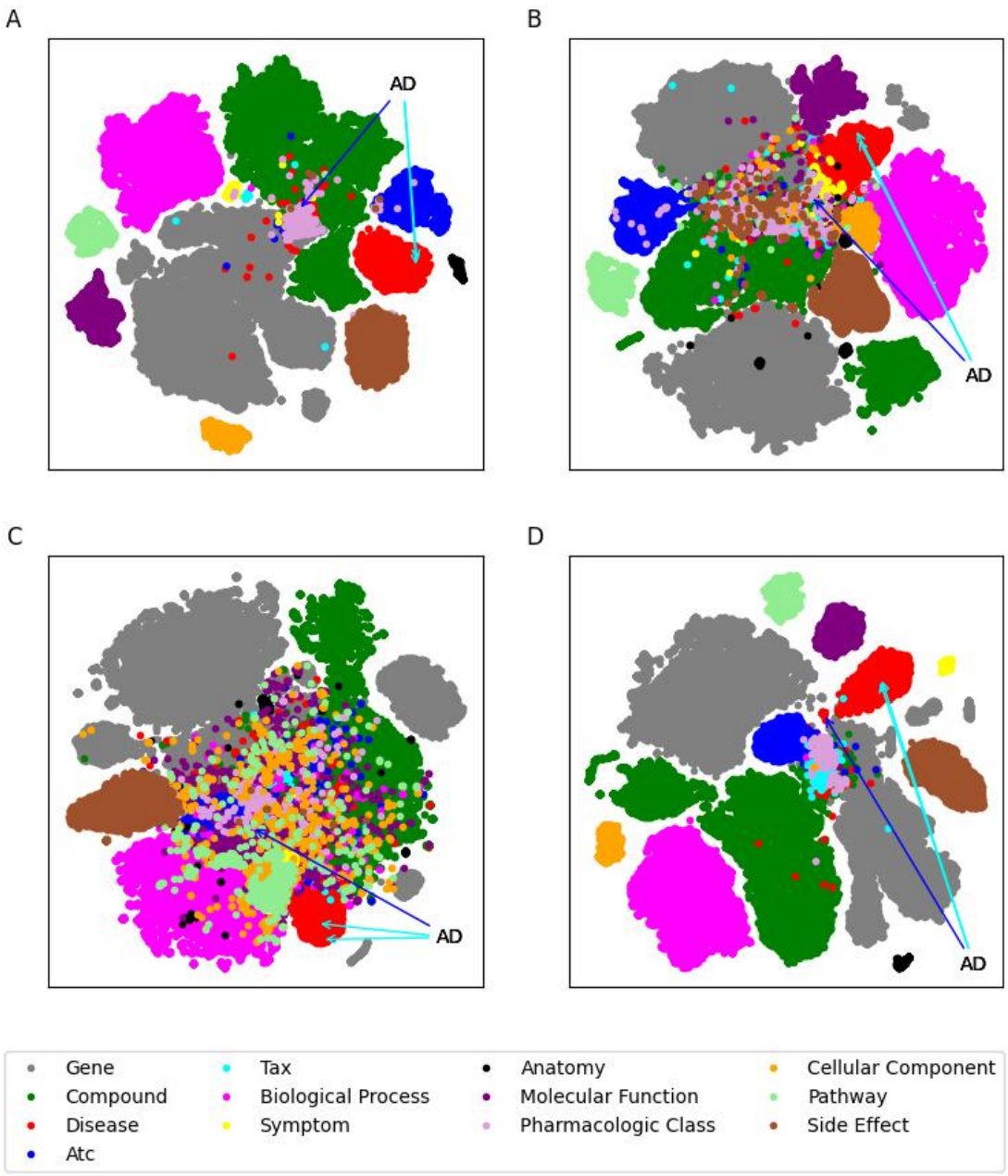

Figure 3 Distribution of entity embeddings in 2D euclidean space for 4 models. Subgraphs A, B, C and D are the results of TransE_l1, TransE_l2, ComplEx and RotatE respectively.

图 4A、4B、4C、4D 显示了 TransE_l1、TransE_l2、ComplEx 和 RotatE 的基于嵌入的不同关系类型之间的成对余弦相似度的分布直方图。对于 TransE_l1，相似度值分布在[-0.873, 0.977]范围内，且有 0.899%相似度大于 0.90 的关系对；TransE_l2 与 TransE_l1 类似，也存在着 4.992%相似度大于 0.50 的关系对。ComplEx 模型的相似度值分布在[-0.208, 0.908]范围内，存在 1.199%相似度大于 0.50 的关系对，相比而言，RotatE 模型的相似度值整体都较小，仅分布在[-0.241, 0.233]范围内。

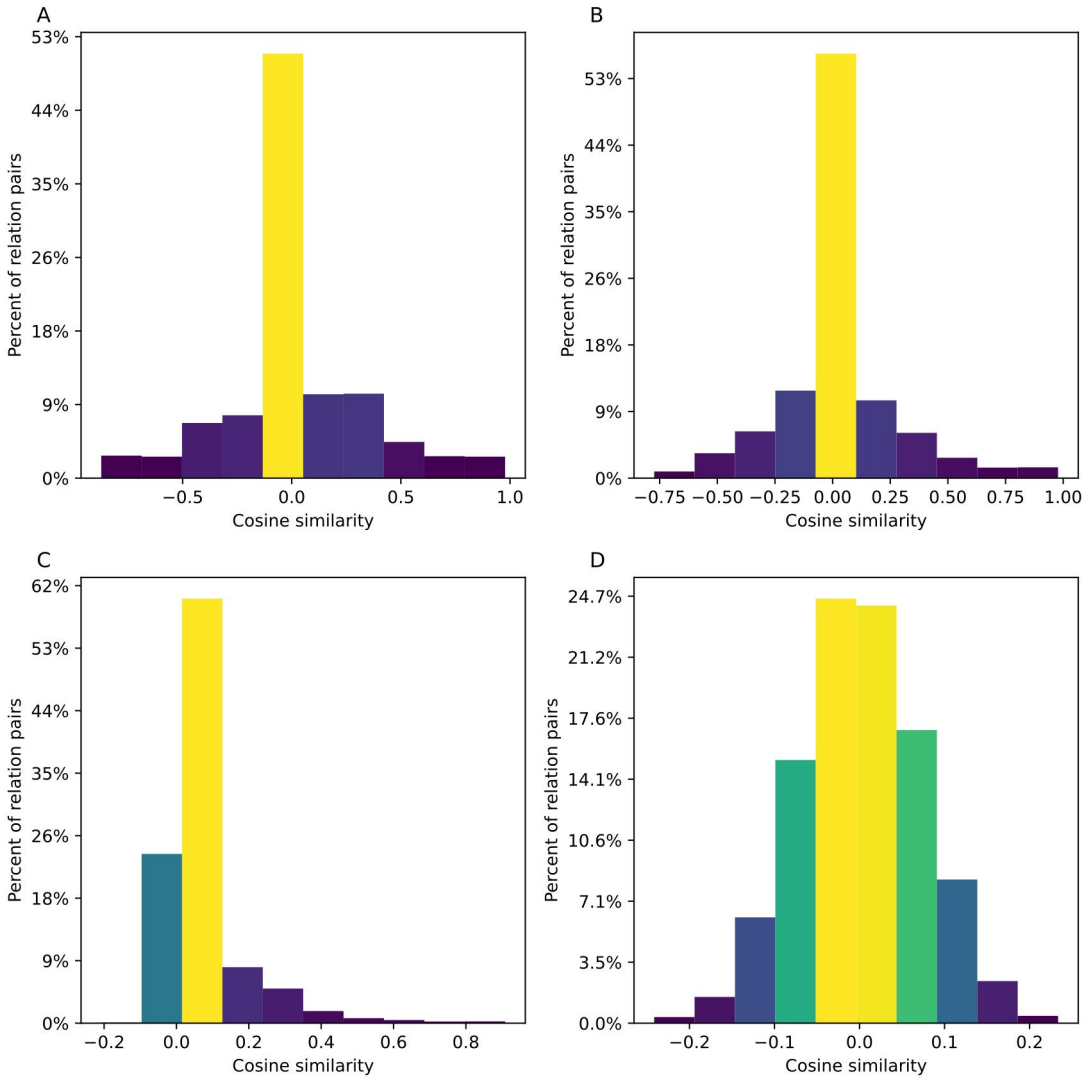

Figure 4 Histogram of cosine similarity between relations for 4 models. Subgraphs A, B, C and D are the results of TransE_l1, TransE_l2, ComplEx and RotatE respectively.

对于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，都与各个模型对应的关系对余弦相似度最高分值都有很大差距，其中 ComplEx 的结果差距最大；对于 ComplEx 模型还有一个惊喜的结果，DRUGBANK::treats::Compound:Disease 和 GNBR::T::Compound:Disease 组成的关系对的余弦相似度为 0.554，远远超过了上面的 ComplEx 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。

综合上面的 KGE 的经典评估和嵌入评估结果，我们将使用 RotatE 模型作为 AD 药物重定位的最终模型。

3.3. AD 药物重定位

我们使用 RotatE 模型进行 AD 药物重定位，由于 RotatE 模型的 MR 指标结果是 63.51，所以将得分前 50 的药物作为候选药物。

候选药物列表中，得分前 10 中只有第 9 名的药物不是 DRKG 原有治疗 AD 的药物，初步证明了该方法的有效性。然后排除了 DRKG 中原有的治疗 AD 的药物，剩余了 17 种药物，药物和找到的支撑文献显示在表 2 中，其中得分排名在 23 名的西布曲明退市，因此最终剩余 16 种候选药物。

Table 2 Candidate drugs obtained from 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[18].
11	Haloperidol	Haloperidol inactivates AMPK and reduces tau phosphorylation in a tau mouse model of Alzheimer's disease[19].
13	Capsaicin	In Alzheimer's disease, capsaicin reduces neurodegeneration and memory impairment[20].
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)[21].
17	Estradiol	Mounting evidence indicates that the neurosteroid estradiol (17 β -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[22].
18	Glucose	Specifically, decreased O-GlcNAcylation levels by glucose deficiency alter mitochondrial functions and together contribute to Alzheimer's disease pathogenesis[23].
20	Disulfiram	Identification of disulfiram as a secretase-modulating compound with beneficial effects on Alzheimer's disease hallmarks[24].
21	Adenosine	Emerging evidence suggests adenosine G protein-coupled receptors (GPCRs) are promising therapeutic targets for Alzheimer's disease[25].
23	Sibutramine	In October 2010, Sibutramine was withdrawn from Canadian and U.S.
29	Paroxetine	Paroxetine ameliorates prodromal emotional dysfunction and late-onset memory deficit in Alzheimer's disease mice[26].
31	Cocaine	None.

39	Paclitaxel	In addition to NSAIDs, an anticancer drug, paclitaxel, has considerable potential as an AD treatment[27].
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[28].
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[29].

4. 讨论与结论

通过比较 KGE 模型经典评估的结果，我们能得出以下结论。DistMult 受限于只能建模对称关系，各项指标都没有最优和次优结果；TransE 的 MR 指标达到了最优结果，但是受限于只能建模一对一的关系，无法在其他指标上达到最优和次优结果；对于 MR 指标，RotatE 和 ComplEx 呈现出截然不同的结果，RotaE 接近于 TransE 取得的最优结果，但是 ComplEx 取得了最差结果，这可能是因为 RotaE 相较于 ComplEx 多捕获了组成关系；对于 MRR 和 Hits@N 两种指标，RotatE 和 ComplEx 各取得了 2 次最优和次优结果，且最优和次优结果也非常接近，充分说明将嵌入向量空间由实数域转换到复数域的必要性。

通过观察关系嵌入 2D 空间的可视化图，我们发现 RotatE 比另外几种更好的整合 DRKG 的关系信息，没有出现比较明显的关系聚集现象，这表明 RotatE 能够将 DRKG 各个数据源的信息很好的映射到一个嵌入向量空间中。从实体嵌入 2D 空间的可视化图中，我们发现 TransE_l1 和 RotatE 能够很好的学习到实体类型信息，但是 ComplEx 无法较好的划分不同类型的实体，甚至对于语义比较相近的映射到 MESH ID 空间的 2 种 AD 实体无法像另外 3 个模型将其映射到接近于一点。通过关系嵌入向量余弦相似度的实验，我们发现 RotatE 能够很好的区分出关系的差异，但是 TransE 无法达到很好的效果，进一步表明复数向量空间的重要性。对于 ComplEx 模型还有一个惊喜的结果，DRUGBANK::treats::Compound:Disease 和 GNBR::T::Compound:Disease 组成的关系对的余弦相似度为 0.554，远远超过了 ComplEx 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，表明 ComplEx 很好的学习了治疗关系的语义相同点和不同点。综上，我们认为 DRKG 不同数据源的信息被 RotatE 很好的整合到了一起，并且生成了有意义的实体和关系嵌入向量，能够有效的进行 AD 药物重定位。

RotatE 得分前 10 的药物只有第 9 名的药物不是 DRKG 已有治疗 AD 的药物，因此可以认为 RotatE 很好的拟合了 DRKG 知识图谱。通过寻找候选治疗药物的支撑文献，我们认为 RotatE 能够很好的完成药物重定位的任务。

由于 DRKG 没有将所有的疾病都映射到统一的 ID 空间，如“Disease::DOID:10652”，这

对药物重定位的效果产生了一定的影响。在构建 KG 时，有必要将同类型的实体映射到一个统一的 ID 空间，这对 KGE 模型学习嵌入向量有很大的帮助。

本文采用 RotatE 对 AD 进行了药物重定位。具体来说，使用 4 种 KGE 模型在 DRKG 上学习实体和关系的嵌入向量，通过多种评估手段评估了 4 种 KGE 模型的性能，最终选择了 RotatE 作为最终的药物重定位模型。选择了得分前 50 的药物作为推荐药物，剔除了 DRKG 原有治疗 AD 的药物，我们最终得到了 16 种治疗 AD 的候选药物，其中 12 种药物找到了支撑文献。

未来，我们将研究更多种类的 KGE 模型在药物重定位中的应用；也将研究实体对齐技术，来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. https://doi.org/10.1186/s12859-022-04934-1.

[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].Molecular & Cellular Proteomics,2016,15(2):409-25. https://doi.org/10.1074/mcp.R115.053330.

[3] Ren RJ,Yin P,Wang ZH,et al.China Alzheimer disease report 2021[J].Journal of Diagnostics Concepts & Practice(诊断学理论与实践),2021,20(04):317-337. https://doi.org/10.16150/j.1671-2870.2021.04.001.

[4] Jia JP,Wei CB,Chen SQ,et al.The cost of Alzheimer's disease in China and re-estimation of costs worldwide[J].Alzheimer's & Dementia,2018,14(4):483-491. https://doi.org/10.1016/j.jalz.2017.12.006.

[5] Avorn J.The $2.6 billion pill–methodologic and policy considerations[J].New England Journal of Medicine,2015,372(20):1877-1879. https://doi.org/10.1056/NEJMp1500848.

[6] 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].Applied Intelligence,2022,52(1): 846–857. https://doi.org/10.1007/s10489-021-02454-8.

[7] Ioannidis VN,Song X,Manchanda S,et al.DRKG - drug repurposing knowledge graph for Covid-19[J]. https://github.com/gnn4dr/DRKG/,2020.

[8] Bordes A,Usunier N,Garcia-Duran A,et al.Translating embeddings for modeling multi-relational data[C]//Advances in Neural Information Processing Systems.Curran Associates, Inc.,2013,26. https://proceedings.neurips.cc/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf.

[9] Yang BS,Yih S,He XD,et al.Embedding entities and relations for learning and inference in knowledge bases[C]//Proceedings of ICLR.2015. http://arxiv.org/abs/1412.6575.

[10] Trouillon T,Welbl J,Riedel S,et al.Complex embeddings for simple link

prediction[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning.JMLR.org,2016,48:2071-2080. https://arxiv.org/abs/1606.06357.

[11] Sun ZQ,Deng ZH,Nie JY, et al. RotatE: knowledge graph embedding by relational rotation in complex space[C]//Proceedings of ICLR. 2019. https://openreview.net/forum?id=HkgEQnRqYQ.

[12] Zeng XX,Song X,Ma TF,et al.Repurpose open data to discover therapeutics for COVID-19 using deep learning[J].Journal of proteome research,2020,19(11):4624-4636. https://doi.org/10.1021/acs.jproteome.0c00316.

[13] Zhang R,Hristovski D,Schutte D,et al.Drug repurposing for COVID-19 via knowledge graph completion[J].Journal of Biomedical Informatics,2021,115(1):103696. https://doi.org/10.1016/j.jbi.2021.103696.

[14] 李宗贤.基于知识图谱的帕金森病药物重定位[J].信息技术与信息化,2022,No.268(07):28-32. https://doi.org/10.3969/j.issn.1672-9528.2022.07.006.

[15] 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. https://aclanthology.org/D18-2024/.

[16] Maaten LVD,Hinton G.Visualizing data using t-SNE[J].Journal of Machine Learning Research,2008,9(86):2579-2605. http://jmlr.org/papers/v9/vandermaaten08a.html.

[17] Zheng Da,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, NY, USA:Association for Computing Machinery,2020:739–748. https://arxiv.org/abs/2004.08532.

[18] Sliwinska S,Jeziorek M.The role of nutrition in Alzheimer's disease[J].Roczniki Panstwowego Zakladu Higieny,2021,72(1):29-39. https://doi.org/10.32394/rpzh.2021.0154.

[19] 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].Alzheimer's & dementia,2016,2(2):121-130. https://doi.org/10.1016/j.trci.2016.05.003.

[20] Pasierski M,Szulczyk B.Beneficial effects of capsaicin in disorders of the central nervous system[J].Molecules,2022,27(8):2484. https://doi.org/10.3390/molecules27082484.

[21] Zu GX,Sun KY,Li L,et al.Mechanism of quercetin therapeutic targets for Alzheimer disease and type 2 diabetes mellitus[J].Scientific reports,2021,11(1):22959. https://doi.org/10.1038/s41598-021-02248-5.

[22] Sahab-Negah S,Hajali V,Moradi HR,et al.The impact of estradiol on neurogenesis and cognitive functions in Alzheimer's disease[J]. Cellular and molecular neurobiology,2020,40(3):283-299. https://doi.org/10.1007/s10571-019-00733-0.

[23] 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 regeneration research,2023,18(4):779-783. https://doi.org/10.4103/1673-5374.354515.

[24] 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].Scientific Reports,2018,8(1):1329. https://doi.org/10.1038/s41598-018-19577-7.

[25] Trinh PNH,Baltos JA,Hellyer SD,et al.Adenosine receptor signalling in Alzheimer's disease[J].Purinergic signal,2022,18(3):359-381. https://doi.org/10.1007/s11302-022-09883-1.

[26] Ai PH,Chen S,Liu XD,et al.Paroxetine ameliorates prodromal emotional dysfunction and late-onset memory deficit in Alzheimer's disease mice[J].Translational Neurodegeneration,2020,9(1):18. https://doi.org/10.1186/s40035-020-00194-2.

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

[28] Baraka A,ElGhotny S.Study of the effect of inhibiting galanin in Alzheimer's disease induced in rats[J].European Journal of Pharmacology,2010,641(2):123-127. https://doi.org/10.1016/j.ejphar.2010.05.030.

[29] 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. https://doi.org/10.1371/journal.pone.0021660.