

基于知识图谱嵌入的阿尔茨海默病药物重定位

摘要

阿尔茨海默病是一种起病隐匿的进行性发展的神经退行性疾病，痴呆表现为主要特征，给社会带来巨大医疗负担，因此迫切需要有效的治疗药物。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，评估了知识图谱嵌入模型在测试集的性能，选择 RotatE 模型作为药物重定位模型；最后，利用知识图谱嵌入模型进行链接预测得出 10 种治疗阿尔茨海默病的候选药物。除此之外，我们通过对实体和关系的嵌入分析和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，为研究人员提供了全新的思路。本文的源代码可以从 <https://github.com/LuYF-Lemon-love/AD-KGE> 处获得。

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

知识图谱（Knowledge Graph, KG）是一种基于图结构存储知识的数据库，主要用于网页搜索、问答系统、推荐任务^[1]。知识中的具体事物和抽象概念在 KG 中被表示为实体（entity），实体之间的联系被表示为关系（relation），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，KG 中的节点（node）代表上面的实体，边（edge）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）^[2]包含 97238 个实体和 5874261 个三元组。如此巨大的 KG，直接像关系数据库一样进行检索和多步推理非常耗时，无法达到现代应用的要求。而且知识图谱具有非常严重的长尾现象，有很多实体与其他实体间仅仅具有很少的关系，对于这些长尾实体，往往很难理解其含义，进而影响对其的推理。因此，如何表示 KG 并进而将其与深度学习联合是一个近些年热门的领域。

知识图谱嵌入（Knowledge Graph Embedding, KGE）是一种将实体和关系表示成低维稠密向量的技术，在这个向量空间内，语义相似的对象之间的距离很近。在过去几年中，研究人员提出了很多 KGE 模型来学习实体和关系嵌入向量，包括但不限于 TransE^[3]、TransR^[4]、RESCAL^[5]、DistMult^[6]、ComplEx^[7]、RotatE^[8]等。

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^[9]，其特征是伴有神经精神症状的渐进性严重痴呆^[10]。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^[11]，并且另一份研究报告称，我国 AD 患者的 2015 年治疗费用为 1677.4 亿美元，到 2050 年将高达 18871.8 亿美元^[12]，这充分说明了 AD 给社会带来了巨大的经济负担。因此，AD 的治疗药物开发迫在眉睫。然而，早在 2015 年，开发一种新药就需要花费 26 亿美元和 10-12 年^[13-14]。药物重定位是指将现

有药物的适应症拓宽到其他疾病，从而能大大节省成本并缩短新药开发周期。

最近，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^[15]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、蛋白质、基因 4 种实体以及它们之间的 39 种关系，然后利用 RotatE 模型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知识图谱，然后利用多种 KGE 模型预测候选药物，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 研究帕金森病的药物重定位，并取得了不错的效果^[17]。

Wang 等^[14]提出了一种基于知识图谱的深度学习方用于 AD 药物重定位。首先，利用 DistMult 学习了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^[9]从文献中构建一个知识图谱，利用 TransE、DistMult 和 ComplEx 学习并预测有助于 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价指标选择了 RotatE 模型作为我们最终的药物重定位模型；然后，在整个知识图谱上训练 RotatE 模型，并利用多种嵌入向量分析手段评价我们学习到的嵌入向量的质量；最终，通过链接预测技术找到 10 种治疗 AD 的候选药物。

2. 方法

2.1.1. 数据

DRKG^[2]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 出版物中收集的数据。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。此外它还提供了一个 TransE₁₂ 的预训练 KGE 模型。

2.1.2. 模型

在本小节中，我们首先介绍一些 KG 和 KGE 数学符号，这些符号将应用于本研究的其他部分。表 1 列出了文章后续部分的概念和定义。

Table 1 The Notation and Definition of KG and KGE.

Name	Notation	Definition
Knowledge Graph	T	A set of triplets in the form (h,r,t)
Entity Set	E	The set of entities

Relation Set	\mathbf{R}	The set of relations
Entity Vectors	\mathbf{h}, \mathbf{t}	The entity vectors for head entity \mathbf{h} and tail entity \mathbf{t}
Relation Vector	\mathbf{r}	The vector for relation \mathbf{r}
Scoring Function	$\mathbf{f}(\mathbf{h}, \mathbf{r}, \mathbf{t})$	The scoring function for the triple $(\mathbf{h}, \mathbf{r}, \mathbf{t})$
Loss Function	\mathcal{L}	The loss function of the model

KG 一般被标记为 \mathbf{T} ，是一组格式为 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ 三元组的集合，其中 $\mathbf{h}, \mathbf{t} \in \mathbf{E}$, $\mathbf{r} \in \mathbf{R}$, \mathbf{E} 是 KG 中的实体集合， \mathbf{R} 是 KG 中的关系集合。KGE 模型旨在将 KG 建模为低维向量空间。在这个向量空间中，头实体 \mathbf{h} 和尾实体 \mathbf{t} 是一个单独的向量，每个关系 \mathbf{r} 是头实体 \mathbf{h} 和尾实体 \mathbf{t} 间的一个运算操作。

KGE 是一个通过概率推断现有 KG 缺失关系进而补全 KG 的任务。近几年研究人员已经提出了很多 KGE 模型，它们都具有一个度量 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ 成立的概率的评分函数。表 2 列出了一些 KGE 模型的评分函数。

Table 2 The brief introduction of some typical KGE models.

Models	Entity	Relation	Score Function	Complexity
TransE	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$- \mathbf{h} + \mathbf{r} - \mathbf{t} _{1/2}$	$O(d)$
TransR	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{k \times d}$	$- \mathbf{M}_r \mathbf{h} + \mathbf{r} - \mathbf{M}_r \mathbf{t} _{1/2}$	$O(d^2)$
RESICAL	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_r \in \mathbb{R}^{d \times d}$	$\mathbf{h}^T \mathbf{M}_r \mathbf{t}$	$O(d^2)$
DistMult	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{h}^T \text{diag}(\mathbf{r}) \mathbf{t}$	$O(d)$
ComplEx	$\mathbf{h}, \mathbf{t} \in \mathbb{C}^d$	$\mathbf{r} \in \mathbb{C}^d$	$\text{Real}(\mathbf{h}^T \text{diag}(\mathbf{r}) \bar{\mathbf{t}})$	$O(d)$
RotatE	$\mathbf{h}, \mathbf{t} \in \mathbb{C}^d$	$\mathbf{r} \in \mathbb{C}^d, r_i = 1$	$- \mathbf{h} \circ \mathbf{r} - \mathbf{t} ^2$	$O(d)$

TransE 是一个具有代表性的平移模型，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 \mathbf{r} 被建模为实体向量的平移，如果三元组 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ 成立，那么 $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ ，即 \mathbf{t} 应该是 $\mathbf{h} + \mathbf{r}$ 最近的实体向量；如果不成立， $\mathbf{h} + \mathbf{r}$ 应该远离 \mathbf{t} 。

TransE 只能建模 1 对 1 的关系类型，TransR 通过从实体向量空间 $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$ 分离出关系的向量空间 $\mathbf{r} \in \mathbb{R}^k$ 来解决这个问题。TransR 通过学习一个关系投影矩阵 $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ 将实体投影到不同的关系向量空间，进而捕获了不同关系的差异。然而 TransR 为了获得更强的表示能力，也将复杂度从 $O(d)$ 提升为 $O(d^2)$ ，导致训练时间急剧增加，因此我们并没有采取 TransR 作为药物重定位的候选模型。

RESICAL 是一个双线性模型，它为每一种关系提供了一个矩阵来建模实体之间的交互进而捕获 KG 的潜在语义。该模型也像 TransR 一样，复杂度为 $O(d^2)$ ，这也是它的扩展模型主要改进的方向，我们也没有使用它作为药物重定位的候选模型。

DistMult 通过将 \mathbf{M}_r 换成对角方阵简化了 RESICAL 模型，使得复杂度变成了 $O(d)$ ，但是该

模型的表现能力相较于 RESCAL 模型也大大减弱，无法达到理想的结果。

由于 DistMult 使用的是对角方阵，因此仅仅能捕获对称关系。为了捕获反对称关系，Complex 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。

受到 TransE 和欧拉恒等式的启发，RotatE 将头实体和尾实体映射到复数向量空间，即 $h, t \in \mathbb{C}^d$ ，将关系建模为从头部实体 h 到尾实体 t 的旋转。RotatE 模型能够捕获对称、反对称、反转和组成四种类型关系，因此它最大限度的表示出 KG 的语义和结构信息。

在后续的实验，我们都是使用最大间隔方法训练模型，这将最小化正确三元组的排名，损失函数如下：

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

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：

$$T^- = E \times R \times E - T$$

2.1.3. 评估指标

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 。我们可以通过链接预测给出需要评估三元组的排名。为了评估 KGE 模型的性能，我们使用三种经典指标：正确实体评分函数的平均排名（Mean Rank, MR），正确实体评分函数的平均倒数排名（Mean Reciprocal Rank, MRR）和正确实体评分函数的前 N 的比例即前 N 命中率 Hits@N（ $N = 1, 3, 10$ ）。MRR 和 Hits@N 都是越高越好，MR 越低越好。

如果我们用 rank_h 和 rank_t 分别表示预测缺失头实体和尾实体三元组的排名， T 表示需要评估的三元组集合，那么 MR 具体的计算方法为：

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

MRR 具体计算方法为：

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

Hits@N 被计算为

$$\text{Hits@N} = \frac{1}{2|T|} \sum_{(h, r, t) \in T} I[\text{rank}_h \leq N] + I[\text{rank}_t \leq N]$$

其中如果条件为真， $I[*]$ 等于 1，否则等于 0。

2.1.4. 实验设置

我们将 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。

我们综合上面 5 个指标的表现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 。

在药物重定位任务中，选择全部 DRKG 中的 AD 实体作为头实体（Disease::DOID:10652，Disease::MESH:C536599，Disease::MESH:D000544），选择 DRKG 中的治疗关系作为链接预测的关系（DRUGBANK::treats::Compound:Disease，GNBR::T::Compound:Disease，Hetionet::CtD::Compound:Disease），将 Drugbank 中所有被 FDA 批准的药物作为候选药物（分子量 ≥ 250 ，共 8104 个），通过预测缺失尾实体的方法选择评分前 50 的药物作为初步 AD 的治疗药物。

3. 结果

3.1.1. KGE 模型的对比

实验比较 4 种 KGE 模型在知识图谱补全任务中的性能。对于 MR 指标，TransE 模型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 模型取得了最优结果 0.621，RotatE 模型次之 0.614。对于 Hits@1 指标，ComplEx 模型取得了最优结果 0.537，RotatE 模型次之 0.515。对于 Hits@3 和 Hits@10，RotatE 模型取得了最优结果分别为 0.681、0.780，ComplEx 模型取得了次优结果分别为 0.673、0.768。表 3 列出了 KGE 模型测试集的结果。

Table 3 The 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	<u>62.64</u>	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	<u>0.673</u>	<u>0.768</u>
RotatE	<u>0.614</u>	63.51	<u>0.515</u>	0.681	0.780

通过比较 KGE 模型的评估结果，我们能得出以下结论。DistMult 模型受限于只能建模对

称关系，因此各项指标都没有最优和次优结果。TransE 模型的 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理想的水平。对于 MR 指标 RotatE 模型和 ComplEx 呈现出截然不同的结果，RotaE 模型接近于 TransE 模型取得的最优结果，但是 ComplEx 模型取得了最差结果，这可能是因为 RotaE 相较于 ComplEx 多捕获了反转和组成两类关系。对于 MRR 和 Hits@N 两种指标，RotatE 和 ComplEx 模型各取得了 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。考虑到 MR 结果，我们最终选择 RotatE 作为我们最终的药物重定位模型。

Table 3 The optimal hyperparameters of the KGE model.

Model	batch_size	neg_sample_size	hidden_dim	γ	lr
TransE_l1	4096	256	400	18	0.05
TransE_l2	4096	256	400	12	0.1
DistMult	4096	256	400	50	0.1
ComplEx	4096	256	400	50	0.1
RotatE	4096	256	200	18	0.05

表 4 列出了 KGE 的超参数，我们重新在整个 DRKG 上训练了 RotatE 模型，选择表 4 中 RotatE 的超参数作为模型的参数。

3.1.2. RotatE 模型的嵌入分析

由于 DRKG 结合了来自不同数据源的信息，我们希望验证使用知识图嵌入技术可以生成有意义的实体和关系嵌入。我们采用 t-SNE 将关系嵌入向量降维和可视化处理。由于实体的数量众多，且 RotatE 实体维度是 400，直接利用 t-SNE 降维和可视化处理会引入很多噪声，因此我们首先使用 PCA 将其降维到 30，然后再利用 t-SNE 将其降维到 2D 空间并可视化处理。我们还使用余弦距离计算关系嵌入向量间的相似度，并输出了最相似的 10 对关系和相似度得分分布的直方图。

图 1 是关系 2D 空间的可视化图，可以发现关系广泛的分布在 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，可以得出 RotatE 模型学习到了各个关系间差异的结论，这对于我们后面的药物重定位是很有意义的。

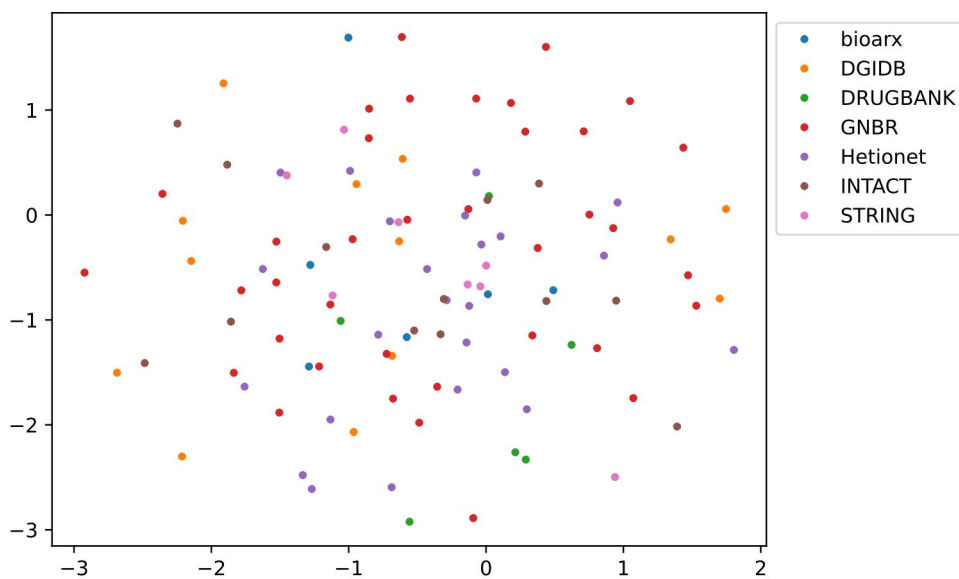


Figure 1 Distribution of relation embeddings in 2D euclidean space

图 2 显示了基于嵌入的不同关系类型之间的成对余弦相似性的详细分布。可以发现绝大多数关系嵌入向量都有很小的相似度，而且最相似的也只有 0.2。

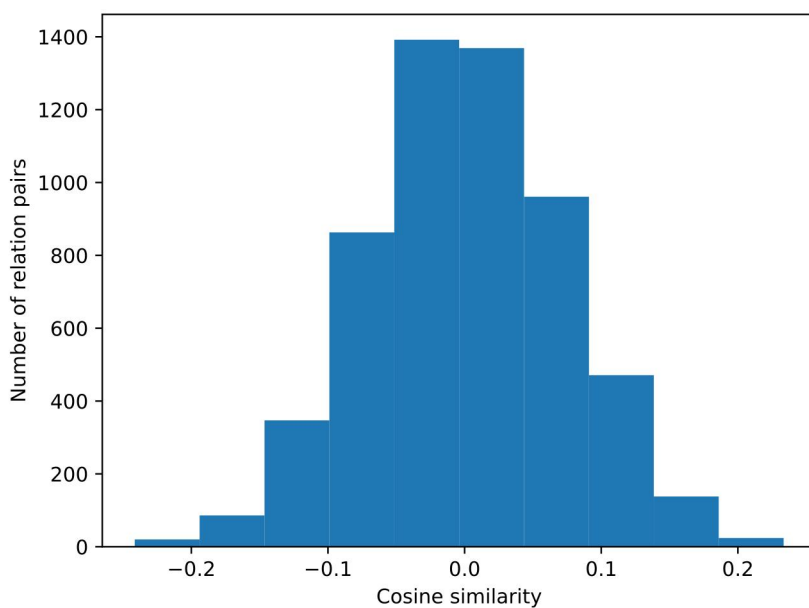


Figure 2 Histogram of cosine similarity between relations

表 5 列出关系间余弦相似度最高的 10 对关系，可以发现所有关系对的相似得分都 <0.25 ，与图 2 的结果相互得到印证。而且我们选择的三种治疗关系并不在其中，即 RotatE 模型能够很好的将治疗关系和其他类型的关系区分开，这对于我们的药物重定位非常重要。

Table 5 Top 10 most similar relation pairs based on their cosine similarity.

Relation 1	Relation 2	Cosine score
Hetionet::DaG::Disease:Gene	INTACT::ASSOCIATION::Compound:Gene	0.2335
DRUGBANK::x-atc::Compound:Atc	GNBR::in_tax::Gene:Tax	0.2331
GNBR::L::Gene:Disease	Hetionet::DpS::Disease:Symptom	0.2269
DGIDB::BLOCKER::Gene:Compound	GNBR::J::Gene:Disease	0.2269
DGIDB::AGONIST::Gene:Compound	INTACT::ASSOCIATION::Compound:Gene	0.2226
DRUGBANK::enzyme::Compound:Gene	GNBR::Ud::Gene:Disease	0.2192
Hetionet::DpS::Disease:Symptom	Hetionet::CbG::Compound:Gene	0.2172
GNBR::N::Compound:Gene	Hetionet::GcG::Gene:Gene	0.2151
GNBR::W::Gene:Gene	Hetionet::CdG::Compound:Gene	0.2138
GNBR::Z::Compound:Gene	GNBR::in_tax::Gene:Tax	0.2115

图 3 是实体 2D 空间的可视化图，不同的颜色表示不同的实体类型，观察到相同类别的实体正如我们期望的那样聚集到一起。

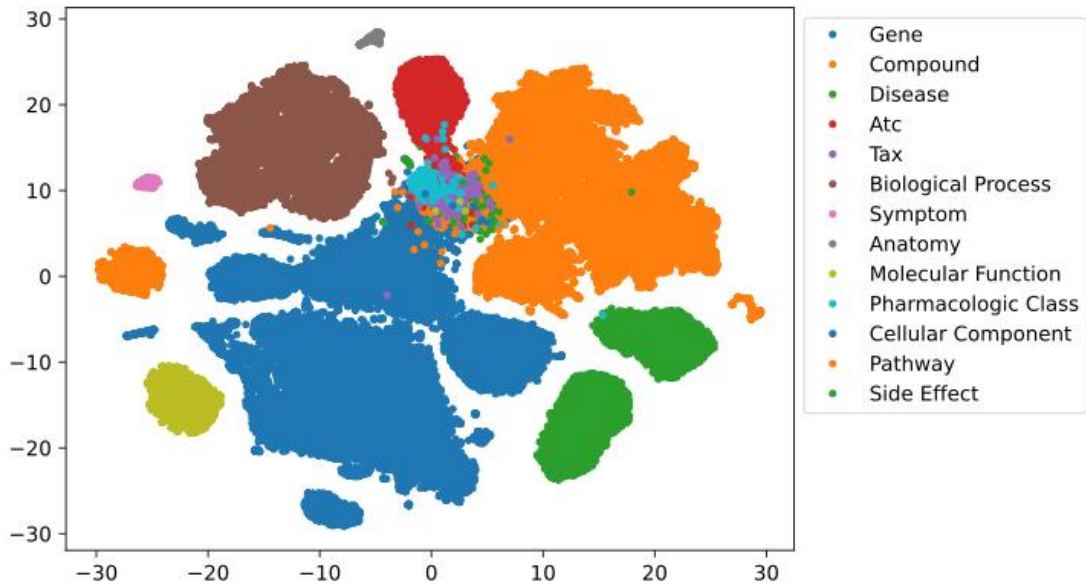


Figure 3 Distribution of entity embeddings in 2D euclidean space

为了进一步分析 RotatE 模型学习到的关系嵌入向量的差异性，我们还分析了不同关系类型之间在链接预测时的相似程度，对于种子头实体 h_i ，用链接预测找到关系 r_j 下最可能成立的前 10 尾实体，然后对关系 r_j 重复上述工作，并计算关系 r_j 和 r_j' 的前十尾实体集合的 Jaccard 相似系数（Jaccard similarity coefficient）。我们选择了 100 个随机头实体种子，然后报告了所有关系对的平均相似度，表 6 列出了基于链接预测相似度的前十关系对。Jaccard 相似系数计算如下：

$$J_{E_1, E_2} = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|}$$

其中 $|*|$ 表示集合的基数, E_1, E_2 是两个关系对应的尾实体集合。

Table 6 Top 10 most similar relation pairs based on their link recommendation similarity

Relation 1	Relation 2	Jaccard score
INTACT::COLOCALIZATION::Gene:Gene	INTACT::DIRECT INTERACTION::Gene:Gene	0.570
Hetionet::GiG::Gene:Gene	INTACT::PHYSICAL ASSOCIATION::Gene:Gene	0.566
INTACT::PHYSICAL ASSOCIATION::Gene:Gene	INTACT::COLOCALIZATION::Gene:Gene	0.564
INTACT::PHYSICAL ASSOCIATION::Gene:Gene	INTACT::DIRECT INTERACTION::Gene:Gene	0.558
bioarx::DrugHumGen:Compound:Gene	DRUGBANK::target::Compound:Gene	0.545
GNBR::E::Compound:Gene	GNBR::E+::Compound:Gene	0.534
bioarx::DrugHumGen:Compound:Gene	Hetionet::CbG::Compound:Gene	0.507
Hetionet::GiG::Gene:Gene	INTACT::COLOCALIZATION::Gene:Gene	0.505
Hetionet::GiG::Gene:Gene	INTACT::DIRECT INTERACTION::Gene:Gene	0.501
GNBR::E+::Gene:Gene	GNBR::E::Gene:Gene	0.490

可以发现我们的三种治疗关系依旧没有出现在前 10 的位置, 更加说明 RotatE 模型能够将治疗关系能够与其他关系区分出来。

通过上面对 RotatE 模型学习到的实体和关系嵌入向量的分析结果, 我们认为 DRKG 不同数据源的信息被药物重定位模型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。

3.1.3. AD 药物重定位

我们使用最终的 RotatE 模型进行 AD 药物重定位, 由于 RotatE 模型的 MR 是 63.51, 所以将评分前 50 的作为初步的候选药物。初步的候选药物列表中, 评分前 10 中只有第 9 名的药物不是 DRKG 已经存在的治疗 AD 的药物, 证明该方法的有效性。然后我们排除了 DRKG 中原有的治疗 AD 的药物, 最终剩余了 17 种治疗药物。

我们将 DRKG 提供的 TransE_I2 预训练模型重复上面的步骤, 排除了 DRKG 中已经存在的治疗 AD 的药物, 最终剩余 25 种治疗药物, 可以发现 RotatE 模型的效果要好于 TransE 模型。我们将上面两个模型得到的候选药物求交集, 最终获得 10 种治疗 AD 的治疗药物, 结果显示在表 7 中。

Table 7 Candidate drugs obtained from drug repurposing.

Rank in RotatE	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] .
21	Adenosine	Emerging evidence suggests adenosine G protein-coupled receptors (GPCRs) are promising therapeutic targets for Alzheimer's disease ^[24] .
31	Cocaine	None.
39	Paclitaxel	In addition to NSAIDs, an anticancer drug, paclitaxel, has considerable potential as an AD treatment ^[25] .
41	Cholesterol	None.

4. 结论

本文采用 KGE 模型对 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，通过评价指标选择了 RotatE 模型作为最终的药物重定位模型。实验结果表明，RotatE 模型能够预测出合理的 AD 治疗药物，能够为研究人员进一步探索 AD 的治疗药物提供研究思路。

5. 参考文献

[1] Lin YK, Shen SQ, Liu ZY, et al. Neural relation extraction with selective attention over instances[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational

- Linguistics,2016:2124-2133.<https://aclanthology.org/P16-1200>.
- [2] Ioannidis VN,Song X,Manchanda S,et al.DRKG - Drug Repurposing Knowledge Graph for Covid-19[J].<https://github.com/gnn4dr/DRKG/>,2020.
- [3] 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>.
- [4] Lin YK,Liu ZY,Sun MS,et al.Learning entity and relation embeddings for knowledge graph completion[C]//Proceedings of the AAAI Conference on Artificial Intelligence.Palo Alto,California,USA:AAAI Press,2015,29(1).<https://doi.org/10.1609/aaai.v29i1.9491>.
- [5] Maximilian N,Volker T,Hans-Peter K.A Three-way model for collective learning on multi-relational data[C]//Proceedings of the 28th International Conference on Machine Learning (ICML-11).New York,NY,USA:ACM,2011:809-816.https://icml.cc/Conferences/2011/papers/438_icmlpaper.pdf.
- [6] 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>.
- [7] 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>.
- [8] 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>.
- [9] 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>.
- [10] 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>.
- [11] 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>.
- [12] 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>.
- [13] 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>.
- [14] 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>.

- [15] 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>.
- [16] 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>.
- [17] 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*, 2022, No.268(07):28-32. <https://doi.org/10.3969/j.issn.1672-9528.2022.07.006>.
- [18] Sliwinska S, Jeziorek M. The role of nutrition in Alzheimer's disease[J]. *Roczniki Panstwowej Zakładu 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] 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>.
- [25] 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.