# 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、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

## 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，早在2015年，开发一种新药就需要花费26亿美元和10-12年[5-6]。药物重定位技术能够将现有药物的适应症拓宽到其他疾病，从而大大节省成本并缩短新药开发周期。

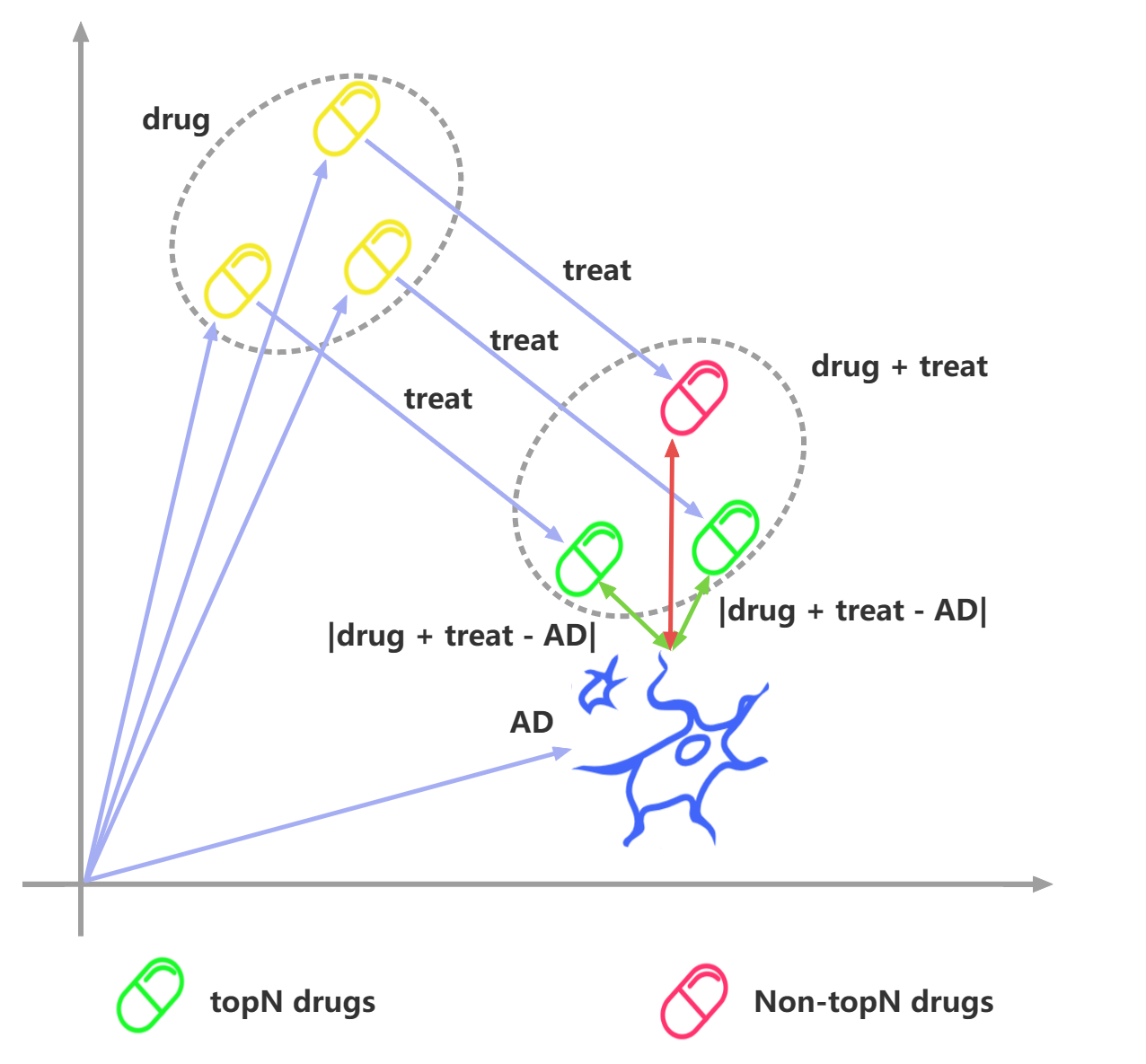
知识图谱（Knowledge Graph，KG）是一种基于拓扑结构图存储知识的数据库，主要用于网页搜索、问答系统、推荐任务[7]。知识中的具体事物和抽象概念在KG中被表示为实体，实体之间的联系被表示为关系，进而知识被表示成格式为（头实体，关系，尾实体）的三元组。KG是一个由大量的三元组组成的有向图结构，图中的节点表示实体，边表示实体间的关系。

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）[8]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[9]。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[9]。因此，如何表示KG是一个近些年热门的领域。

KG通常被标记为，是一组格式为三元组的集合，其中，，是KG的实体集合，是KG的关系集合。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，进而KG被建模成低维向量空间，在这个向量空间内，头实体和尾实体是一个单独的向量，每个关系是头实体和尾实体间的一个运算操作。

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，本文仅介绍和利用一些线性时间复杂度算法，如TransE[10]、DistMult[11]、ComplEx[12]、RotatE[13]。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，KGE模型能够被用于药物重定位研究。

TransE[10]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，关系被建模为实体向量的平移，如果三元组成立，那么，即应该是最近的实体向量；如果不成立，应该远离。图1展示了如何使用TransE模型进行AD药物重定位。药物实体、AD实体和治疗关系都被表示成了嵌入空间中的一个向量，现存治疗其他疾病的药物用黄色胶囊表示，是头实体；绿色和红色胶囊表示模型推测的初始候选治疗AD药物，是；TransE模型仅仅选择离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.

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[14]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[15]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[16]。

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模型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价指标评估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，通过链接预测技术寻找治疗AD的候选药物。

## 方法

### 数据

DRKG[8]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。DRKG使用“实体类型::ID”的格式表示实体，如Disease::MESH:D000544，其中Disease是实体类型，MESH:D000544是MESH 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。

### 知识图谱嵌入模型

#### 基本原理

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成立的概率的评分函数。

TransE[10]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，是向量空间的维度。关系被建模为实体向量的平移，如果三元组成立，那么，即应该是最近的实体向量；如果不成立，应该远离。TransE的评分函数是，它只能建模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[13]。

DistMult[11]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。它也假设实体和关系属于同一向量空间，评分函数为。

由于DistMult[11]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[12]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。它假设实体和关系属于同一复数向量空间，评分函数为。

受到TransE和欧拉恒等式的启发，RotatE[13]将头实体和尾实体映射到复数向量空间，即当 ，，，将关系r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、反转和组成四种类型关系，评分函数为。

#### 优化

在后续的实验中，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[10]，损失函数如下：

其中，是正负例三元组得分的间隔距离。是正例三元组集合，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：

### KGE模型的评估

#### 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，即给定预测缺失的尾实体，或者给定预测缺失的头实体h。可以通过链接预测给出正确实体的排名。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[10]），正确实体评分函数的平均倒数排名（Mean Reciprocal Rank，MRR[13]）和正确实体评分函数的前N的比例即前N命中率Hits@N[10]（N = 1，3，10）。MRR和Hits@N都是越高越好，MR越低越好。

如果用和分别表示预测正确头实体和尾实体的排名，表示需要评估的三元组集合，那么MR具体的计算方法为：

MRR具体计算方法为：

Hits@N 被计算为

其中如果条件为真，等于1，否则等于0。

#### 嵌入评估

由于DRKG结合了来自不同数据源的信息，我们希望验证KGE模型可以生成有意义的实体和关系嵌入。

采用t-SNE[17]将关系嵌入向量降维和可视化处理；由于实体的数量众多，直接利用t-SNE降维和可视化处理会引入很多噪声，因此首先使用主成分分析（Principal component analysis， PCA）将其降维到30，然后再利用t-SNE将其将降维到2D空间并进行可视化处理；除此之外，还使用余弦相似性计算关系嵌入向量对的相似度，并输出了相似性得分分布的直方图、关系对余弦相似性最高分值、余弦相似性分值最高的治疗关系。

为了进一步分析KGE模型学习到的关系嵌入向量的差异性，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，用链接预测找到关系下最可能成立的前10尾实体，然后对关系重复上述工作，并计算关系和的前10尾实体集合的Jaccard相似系数[8]。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。Jaccard相似系数计算方法如下：

其中表示集合的基数，， 是两个关系对应的尾实体集合。

### 药物重定位

在药物重定位任务中，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），它们是头实体列表；选择DRKG中所有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），其中DRUGBANK::treats、GNBR::T、Hetionet::CtD分别是DrugBank数据库、GNBR数据库、Hetionet数据库的治疗关系；选择DRKG中所有AD实体作为尾实体列表（Disease::DOID:10652，Disease::MESH:C536599，Disease::MESH:D000544），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；将上面实体和关系列表进行格式为排列组合（总共种可能），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模型的MR指标选择。

### 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，从{50,125,200}中选择DistMult、ComplEx的超参数。

本文的实验是利用Zheng等[18]开发DGL-KE工具包实现的。

## 结果

### KGE模型的经典评估

实验比较4种KGE模型在知识图谱补全任务中的性能，表1列出了KGE模型测试集的结果。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为0.621，RotatE模型次之为0.614。对于Hits@1指标，ComplEx模型取得了最优结果为0.537，RotatE模型次之为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。DistMult模型3种指标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** |

表2列出了KGE的最优超参数，考虑到表1的结果，排除了DistMult模型，重新在整个DRKG上训练了其他模型用于本文后面的嵌入评估和药物重定位，选择表2的超参数作为模型的参数。

**Table 2** The optimal hyperparameters of the KGE model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | batch\_size | neg\_sample\_size | hidden\_dim |  | 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 |

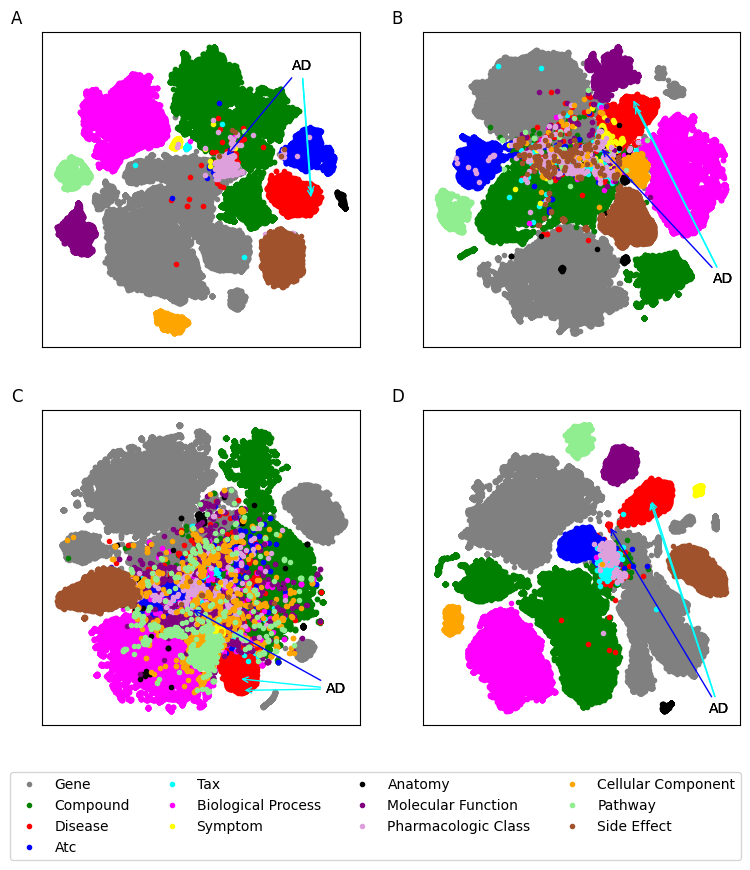
### KGE模型的嵌入评估

图2是TransE\_l1、TransE\_l2、ComplEx和RotatE的关系嵌入2D空间的可视化图，用不同颜色显示关系来自的数据来源，用蓝绿色箭头指出了药物重定位3种治疗关系。TransE\_l1 和TransE\_l2模型的关系嵌入向量出现同数据源聚集的现象，尤其GNBR数据源的关系； ComplEx和RotatE模型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，因此，ComplEx和RotatE模型学习到了各个关系的差异。



**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3是TransE\_l1、TransE\_l2、ComplEx和RotatE的实体嵌入2D空间的可视化图，用不同的颜色表示不同的实体类型，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，但是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4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性的详细分布。RotatE的相似度区间最小（小于0.25）。ComplEx和RotatE取得了相似的结果，但是ComplEx有少量高相似度长尾值。TransE\_l1和TransE\_l2模型的相似度区间非常大（接近于1.00），而且有很多相似性得分接近于的1.00的关系对。



**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.

表3列出4个模型关系对余弦相似性最高分值、余弦相似性分值最高的治疗关系。ComplEx、TransE\_l1和TransE\_l2的关系对相似性最高值都高于0.9，而RotaE的最高值0.2335小于0.25，与图4的结果相互印证。4个模型的余弦相似性分值最高的治疗关系与各个模型关系对余弦相似性最高分值都有很大差距，其中ComplEx的结果差距最大。其中ComplEx模型相似性得分最高的治疗关系所在的关系对正好是DRUGBANK::treats::Compound:Disease和GNBR::T::Compound:Disease组成的关系对。ComplEx和RotatE能够很好的将治疗关系和其他类型的关系区分开。

**Table 3** Cosine similarity between relations based on their embeddings for 4 models.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Highest similarity score | Highest similarity score of the three treatment relations | Name of the highest similarity score treatment relation |
| TransE\_l1 | 0.9774 | 0.9175 | GNBR::T::Compound:Disease |
| TransE\_l2 | 0.9778 | 0.8410 | GNBR::T::Compound:Disease |
| ComplEx | 0.9083 | 0.5538 | DRUGBANK::treats::Compound:Disease, GNBR::T::Compound:Disease |
| RotatE | 0.2335 | 0.1798 | Hetionet::CtD::Compound:Disease |

表4列出了4个模型关系对基于链接预测相似性最高分值、基于链接预测相似性分值最高的治疗关系。4个模型的基于链接预测相似度最高的治疗关系与各个模型关系对基于链接预测相似度最高分值都有很大差距。其中ComplEx和RotatE模型相似性得分最高的治疗关系所在的关系对正好是DRUGBANK::treats::Compound:Disease和GNBR::T::Compound:Disease组成的关系对。4个模型能够很好的将治疗关系和其他类型的关系区分开。

Table 4 Jaccard similarity coefficient between relations based on their link recommendation similarity for 4 models .

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Highest score | Highest score of the three treatment relations | Name of the highest score treatment relation |
| TransE\_l1 | 0.7266 | 0.3986 | Hetionet::CtD::Compound:Disease |
| TransE\_l2 | 0.8446 | 0.5345 | GNBR::T::Compound:Disease |
| ComplEx | 0.4037 | 0.1164 | DRUGBANK::treats::Compound:Disease, GNBR::T::Compound:Disease |
| RotatE | 0.6095 | 0.2401 | DRUGBANK::treats::Compound:Disease,  GNBR::T::Compound:Disease |

### AD 药物重定位

使用上面重新训练的TransE\_l1、TransE\_l2、ComplEx和RotatE进行药物重定位，考虑上面几种模型的MR结果（TransE\_l1：62.64，TransE\_l2：60.83，ComplEx：112.74，RotatE：63.51），TransE\_l1、TransE\_l2、ComplEx和RotatE分别选择得分前50、前50、前100、前50的药物作为初步的候选药物。

表5列出各个模型得分前10且不是DRKG已有的治疗AD的药物列表。TransE\_l1、TransE\_l2、ComplEx和RotatE得分前10的药物分别只有1种、4种、1种、1种不是DRKG已有治疗AD的药物，且所有排名都大于5。7种候选药物仅仅只有胆固醇没有找到治疗AD的支撑文献。

**Table 5** The Top10 recommended results of each model which are not DRKG's existing drugs for the treatment of AD. Drugs with supporting literature are shown in **bold**.

|  |  |  |
| --- | --- | --- |
| Model | Ranking of predicted results in the model | Drug name |
| TransE\_l1 | 9 | **Quercetin** |
| TransE\_l2 | 6 | Cholesterol |
| 7 | **Glucose** |
| 8 | **Glutathione** |
| 10 | **Cisplatin** |
| ComplEx | 8 | **Paroxetine** |
| RotatE | 9 | **Glutathione** |

各个模型的候选药物列表排除了DRKG已有治疗AD的药物，TransE\_l1、TransE\_l2、ComplEx和RotatE分别剩余了19种、24种、34种、17种治疗药物。

将上面4个模型得到的候选药物集合求交集，结果显示在表6中。

**Table 6** Overlap of prediction results of four models. Drugs with supporting literature are shown in bold.

|  |  |
| --- | --- |
| List of drugs for set intersection | Drug name |
| TransE\_l1\_top50 and TransE\_l2\_top50 | **Estradiol**, **Testosterone**, **Glucose**, Chlorpromazine, **Quercetin**, Cholesterol, **Verapamil**, **Glutathione**, Morphine, Cocaine, **Clozapine** |
| TransE\_l1\_top50 and ComplEx\_top100 | **Clozapine**, **Verapamil**, **Methotrexate** |
| TransE\_l1\_top50 and RotatE\_top50 | **Estradiol**, **Glucose**, **Quercetin**, Cholesterol, **Glutathione**, **Capsaicin**, Cocaine |
| TransE\_l2\_top50 and ComplEx\_top100 | Methylprednisolone, **Clozapine**, **Verapamil** |
| TransE\_l2\_top50 and RotatE\_top50 | **Estradiol**, **Glucose**, **Quercetin**, Cholesterol, **Glutathione**, Cocaine, **Paclitaxel**, **Haloperidol** |
| ComplEx\_top100 and RotatE\_top50 | **Paroxetine**,**Glyburide** |
| TransE\_l1\_top50 and TransE\_l2\_top50 and ComplEx\_top100 | **Clozapine**, **Verapamil** |
| TransE\_l1\_top50 and TransE\_l2\_top50 and RotatE\_top50 | **Estradiol**, **Glucose**, **Quercetin**, Cholesterol, **Glutathione**, Cocaine |
| TransE\_l1\_top50 and ComplEx\_top100 and rotatE\_top50 | None |
| TransE\_l2\_top50 and ComplEx\_top100 and rotatE\_top50 | None |

表7列出了上述药物重定位实验中有支撑文献的药物。

**Table 7** Drugs with supporting literature in drug repositioning experiment.

|  |  |
| --- | --- |
| Drug name | Literature support |
| Glucose | Specifically, decreased O-GlcNAcylation levels by glucose deficiency alter mitochondrial functions and together contribute to Alzheimer's disease pathogenesis[19]. |
| 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[20]. |
| 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]. |
| 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[22]. |
| Verapamil | Verapamil Prevents Development of Cognitive Impairment in an Aged Mouse Model of Sporadic Alzheimer’s Disease[23]. |
| Clozapine | Clozapine Improves Memory Impairment and Reduces A beta Level in the Tg-APPswe/PS1dE9 Mouse Model of Alzheimer's Disease[24]. |
| Testosterone | Animal models demonstrated that testosterone (T) exerted a neuroprotective effect reducing the production of amyloid-beta (Aβ), improving synaptic signaling, and counteracting neuronal death[25]. |
| Methotrexate | Anti-inflammatory methotrexate treatment reduced the incidence of Alzheimer's disease in high-risk individuals[26]. |
| Capsaicin | In Alzheimer's disease, capsaicin reduces neurodegeneration and memory impairment[27]. |
| Haloperidol | Haloperidol inactivates AMPK and reduces tau phosphorylation in a tau mouse model of Alzheimer's disease[28]. |
| Paclitaxel | In addition to NSAIDs, an anticancer drug, paclitaxel, has considerable potential as an AD treatment[29]. |
| 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[30]. |
| Paroxetine | Paroxetine ameliorates prodromal emotional dysfunction and late-onset memory deficit in Alzheimer's disease mice[31]. |
| Cisplatin | Cisplatin Inhibits the Formation of a Reactive Intermediate during Copper-Catalyzed Oxidation of Amyloid beta Peptide[32]. |

## 讨论与结论

通过比较KGE模型的经典评估，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。

通过关系嵌入2D空间的可视化图，我们发现ComplEx和RotatE比TransE的两种变体更好的整合DRKG的关系信息，没有出现比较明显的聚集现象。这表明这两种模型能够将DRKG各个数据源的信息很好的映射到一个嵌入向量空间中。通过实体嵌入2D空间的可视化图，TransE\_l1和RotatE模型能够很好的学习到实体类型信息，但是ComplEx模型无法较好的划分不同的实体，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种模型将其映射到接近于一点。通过关系嵌入向量余弦相似性的实验，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，表明复数向量空间的重要性。在计算基于链接预测相似性的实验中，ComplEx 和 RotatE相似性得分最高的治疗关系所在的关系对正好是 DRUGBANK::treats::Compound:Disease 和 GNBR::T::Compound:Disease组成的关系对，且总体上所有关系对的链接预测相似性都很低，表明ComplEx 和 RotatE很好的学习了治疗关系的语义相同点和不同点。虽然TransE模型和另外两种模型有些许差距，但是我们仍旧认为DRKG不同数据源的信息被4个KGE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，能够有效的进行AD药物重定位。

TransE\_l1、TransE\_l2、ComplEx 和 RotatE 得分前10的药物分别只有 1 种、4 种、1 种、1 种不是 DRKG已有治疗AD的药物，且所有排名都大于 5，我们认为4个模型都很好的拟合了DRKG知识图谱，TransE\_l2模型相对于其他3种模型拟合性较差。通过寻找候选治疗药物的支撑文献，我们认为KGE模型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模型的学习有很大的帮助。

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4个模型作为最终的药物重定位模型，并分析4个KGE模型学习到的嵌入向量的质量。通过多种KGE模型的重叠情况，我们找到了多种治疗AD的候选药物。

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；我们也将研究实体融合技术，来将多种数据源的实体和关系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. 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.
8. Ioannidis VN,Song X,Manchanda S,et al.DRKG - drug repurposing knowledge graph for Covid-19[J]. <https://github.com/gnn4dr/DRKG/,>2020.
9. 刘知远,韩旭,孙茂松.知识图谱与深度学习[M].北京:清华大学出版社,2020:9.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
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. Ahmed HA,Ismael S,Mirzahosseini G,et al.Verapamil Prevents Development of Cognitive Impairment in an Aged Mouse Model of Sporadic Alzheimer’s Disease[J]. Molecular Neurobiology,2021,58(7):3374–3387. https://doi.org/10.1007/s12035-021-02350-9.
24. Choi Y,Jeong HJ,Liu QF,et al.Clozapine Improves Memory Impairment and Reduces Aβ Level in the Tg-APPswe/PS1dE9 Mouse Model of Alzheimer’s Disease[J]. Molecular Neurobiology,2017,54(1):450–460. https://doi.org/10.1007/s12035-015-9636-x.
25. Bianchi VE.Impact of Testosterone on Alzheimer's Disease[J]. World Journal of Nens Health,2022,40(2):243-256. https://doi.org/10.5534/wjmh.210175.
26. Lindbohm JV,Mars N,Sipilae, PN,et al.Immune system-wide Mendelian randomization and triangulation analyses support autoimmunity as a modifiable component in dementia-causing diseases[J].nature aging,2022,2(10):956–972. https://doi.org/10.1038/s43587-022-00293-x.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. Walke GR,Rapole S,Kulkarni PP.Cisplatin inhibits the formation of a reactive intermediate during copper-catalyzed oxidation of amyloid β peptide[J].Inorganic Chemistry,2014,53(19):10003-10005. https://doi.org/10.1021/ic5007764.