

# 基于知识图谱嵌入的阿尔茨海默病药物重定位

## 摘要：

TODO

**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

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<sup>[1]</sup>。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。伴随着深度学习的发展，各种类型的 KG 已经被建立，如 WikiData<sup>[2]</sup>、Freebase<sup>[3]</sup>、DBpedia<sup>[4]</sup>、YAGO<sup>[5]</sup>及 WordNet<sup>[6]</sup>等经典知识图谱。

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）<sup>[7]</sup>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实值向量的技术，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，包括但不限于 TransE<sup>[8]</sup>、TransR<sup>[9]</sup>、RESCAL<sup>[10]</sup>、DistMult<sup>[11]</sup>、Complex<sup>[12]</sup>、RotatE<sup>[13]</sup>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<sup>[14]</sup>，其特征是伴有神经精神症状的渐进性严重痴呆<sup>[15]</sup>。中国阿尔茨海默病报告 2021 显示我国 60 岁及以上人群中 983 万例 AD 患者<sup>[16]</sup>。并且另一份研究报告称，我国 AD 患者的 2015 年治疗费用为 1 677.4 亿美元，到 2050 年将高达 18871.8 亿美元<sup>[17]</sup>，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<sup>[18-19]</sup>。药物重定位是指将现有药物的适应症拓宽到其他疾病，从而大大节省成本并缩短新药开发周期。药物重定位利用了同一个分子的代谢途径可以导致不同疾病的事实，因此一些药物可以治疗不同的疾病<sup>[7]</sup>。

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 等人<sup>[20]</sup>建立了一个 1500 万个三元组的综合知识图谱，包括药物、疾病、蛋白质、基因、代谢途径和表达等多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<sup>[21]</sup>等人提出了一种基于神经网络和文献发现的方法。首先利用 PubMed 和其他以 COVID-19 为重点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<sup>[22]</sup>。

Wang<sup>[23]</sup>等人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，然后利用一个 Conv-Conv 模块来提取药物-靶点对的特征，之后通过一个全连接的神经网络进行药物靶点计算，最终通过载脂蛋白 E 寻找治疗 AD 的药物。Nian<sup>[24]</sup>等人通过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 治疗或预防的候选者，以研究 AD 与化学品，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。

## 参考文献

- [1] Lin Y K, Shen S Q, Liu Z Y, et al. Neural relation extraction with selective attention over instances[C]. Proceedings of ACL, 2124–2133, 2016.
- [2] Vrandečić D, Krotzsch M. WikiData: a free collaborative knowledgebase[C].

Communications of the ACM, 2014, 57(10): 78-85.

[3] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]. Proceedings of KDD, 1247-1250, 2008.

[4] Auer S, Bizer C, Kobilarov G, et al. DBpedia: A nucleus for a Web of open data[C]. Proceedings of ISWC, 722-735, 2007.

[5] Hoffart J, Suchanek F M, Berberich K, et al. YAGO2: A spatially and temporally enhanced knowledge base from wikipedia[J]. Artificial Intelligence, 2013, 194: 28-61.

[6] Miller G A. WordNet: a lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.

[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]. Proceedings of NIPS, 2787-2795, 2013.

[9] Lin Y K, Liu Z Y, Sun M S, et al. Learning entity and relation embeddings for knowledge graph completion[C]. Proceedings of AAAI, 2181-2187, 2015.

[10] Nickel M, Tresp V, Kriegel H P. A Three-way model for collective learning on multi-relational data[C]. Proceedings of ICML, 809-816, 2011.

[11] Yang B S, Yih W T, He X D, et al. Embedding entities and relations for learning and inference in knowledge bases[C]. Proceedings of ICLR, 1-13, 2015.

[12] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C]. Proceedings of ICML, 2071-2080, 2016.

[13] Sun Z Q, Deng Z H, Nie J Y, et al. RotatE: Knowledge graph embedding by relational rotation in complex space[C]. Proceedings of ICLR, 2019.

[14] Nian Y, Hu X, Zhang R, et al. Mining on Alzheimer's diseases related knowledge graph to identify potential AD-related semantic triples for drug repurposing[J]. BMC Bioinformatics. 2022 Sep 30;23(Suppl 6):407.

[15] 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 Feb;15(2):409-25.

- [16] 任汝静,殷鹏,王志会, et al.中国阿尔茨海默病报告 2021[J].诊断学理论与实践,2021,20(04):317-337.
- [17] Jia, J., Wei, C., Chen, S., et al. The cost of Alzheimer's disease in China and re-estimation of costs worldwide[J]. *Alzheimer's & Dementia*, 2018, 14: 483-491.  
<https://doi.org/10.1016/j.jalz.2017.12.006>
- [18] Avorn J. The \$2.6 billion pill—Methodologic and policy considerations[J]. *New England Journal of Medicine*, 2015, 372(20): 1877-1879.
- [19] Wang, S., Du, Z., 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: 846–857.
- [20] Zeng X, Song X, Ma T, et al. Repurpose Open Data to Discover Therapeutics for COVID-19 Using Deep Learning[J]. *J Proteome Res*. 2020 Nov 6;19(11):4624-4636 .
- [21] Rui Zhang and Dimitar Hristovski and Dalton Schutte and Andrej Kastrin and Marcelo Fiszman and Halil Kilicoglu. Drug repurposing for COVID-19 via knowledge graph completion[J]. *Journal of Biomedical Informatics*. 2021
- [22] 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(07):28-32.
- [23] Wang, S., Du, Z., 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, 846–857.
- [24] Nian Y, Hu X, 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 Sep 30;23(Suppl 6):407.