

# 知识图谱嵌入技术综述

## 摘要

知识图谱在近年来获得了极大的发展，已经成为极其热门的研究领域。为了顺利操作知识图谱三元组，人们提出了知识图谱嵌入技术的研究方向。知识图谱嵌入也可以理解为知识图谱表示学习，它将每个三元组的实体和关系都嵌入到一个低维向量空间中，同时保存知识图谱三元组中的原有结构，最后得到实体和关系潜在的特征向量。我们可以用这些得到的嵌入，辅助进行各种应用。

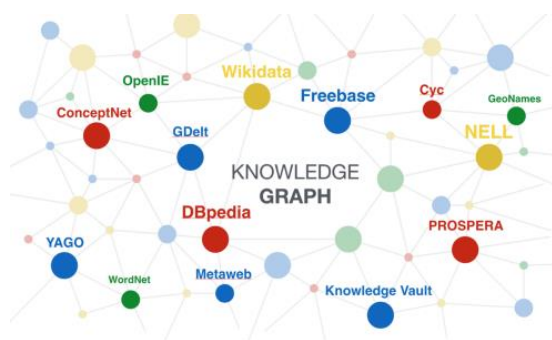
从最开始的 RESCAL 双线性模型，到 Trans 系列模型，再到如今发展起来的神经网络模型，知识图谱嵌入技术的发展日新月异。在文中，我首先介绍了一下知识图谱及其嵌入的发展，然后将各种模型分成三类进行介绍，最后对其发展进行了展望。

## 1 背景介绍

随着人工智能的发展，人们对于机器认知智能的需求越来越看重，而推理是认知智能的重要组成部分，自然语言处理技术也很难解决推理问题。专家们考虑到，我们可能需要通过大量的知识库来进行推理查询，于是，知识工程应运而生。

而如今的知识图谱的概念是由谷歌于 2012 年正式提出的，它承接了知识工程的发展，在 2013 年以后就开始在业界普及。知识图谱以结构化的形式描述客观世界中概念、实体及其之间的关系，它被认为是从感知智能通往认知智能的重要基石。本质上，知识图谱就是一种叫做语义网络的知识库，它将知识库表示为有向图，其中的结点和边分别表示为实体和实体之间的关系。

近年来，知识图谱的构建和应用得到了迅速的发展。人们创建了大量的知识图谱，如 Freebase、DBpedia、Yago 和 Nell。但是这种结构很难操作，为此，研究者们提出了一种新的研究方向，也就是知识图谱嵌入。通过这项操作，人们成功将结点和边嵌入到低维向量表示，并将它们应用到了各种实际情况之中，从语义搜索[11][12]、智能问答[13][14]，到可视化决策支持和推荐系统[15][16][17]等。



图一 知识图谱

而且，知识图谱嵌入技术不仅能将实体和关系转化为连续的向量空间，从而简化操作，同时也能保留知识图谱的原有结构，能够很好的将得到的嵌入进一步应用到各种任务中，起到完美辅助的作用。如 KG 补全、关系提取、实体分类和实体解析。

现在的知识图谱嵌入技术，大致可以分为以下几类[3][4]：(1) 平移距离模型，(2) 双线性模型，(3) 神经网络模型。

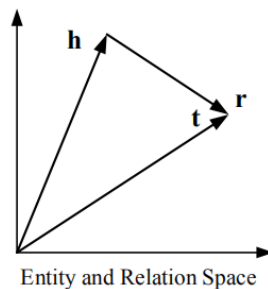
## 2 技术发展

### 2.1 平移距离模型

#### 2.1.1 TransE 模型

在平移距离模型中,最具有代表的就是 TransE 模型[20]了。2013 年,Mikolov 等人提出了 word2vec 模型[1],并通过它发现了词向量空间中存在平移不变现象。Border 等人就是通过平移不变的启发,提出了 TransE 模型。

对于每一个三元组  $(h, r, t)$ ——其中  $h$  和  $t$  代表头实体和尾实体, $r$  代表两者之间的关系,它们对应的向量满足  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ 。即把实体和关系表示在同一个空间中,把关系向量  $\mathbf{r}$  看成是头实体向量  $\mathbf{h}$  和尾实体向量  $\mathbf{t}$  之间的平移。



图二 TransE 模型

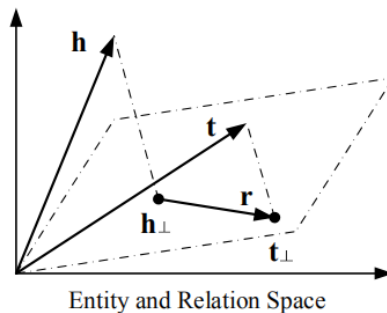
不过虽然 TransE 模型计算复杂度很低,在大规模稀疏知识库上也有很不错的性能和扩展性,但是 TransE 模型对于处理复杂关系有些捉襟见肘。比如在一个头实体可能对应多个尾实体,一个尾实体也可能对应多个头实体等的情况下,可能完全不同的实体会被表示成非常接近的向量。也就是它对处理一对多、多对一和多对多的关系拥有很明显的局限性。

#### 2.1.2 其他拓展模型

为了解决 TransE 模型遗留下的问题,许多模型提出了自己的解决方法。

TransH 模型[22]认为一个实体在不同的关系下应该拥有不同的向量表示。于是它先把实体向量通过一个投影向量  $\mathbf{w}_r$  投影到关系  $r$  对应的超平面上,投影到超平面上的实体向量分别用  $\mathbf{h}_\perp$  和  $\mathbf{t}_\perp$  表示。

$$\mathbf{h}_\perp = \mathbf{h} - \mathbf{w}_r^\top \mathbf{h} \mathbf{w}_r, \quad \mathbf{t}_\perp = \mathbf{t} - \mathbf{w}_r^\top \mathbf{t} \mathbf{w}_r.$$



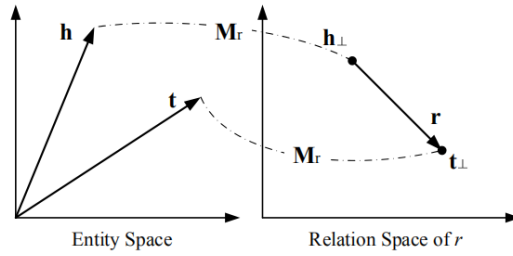
图三 TransH 模型

然后再通过平移不变性质处理,  $\mathbf{h}_\perp + \mathbf{r} \approx \mathbf{t}_\perp$ 。

不过 TransH 模型仍然假设实体和关系处于相同的语义空间中。后面提出的 TransR 模型[23]则认为，一个实体是多种属性的综合体，不同关系关注实体的不同属性。TransR 模型认为不同的关系拥有不同的语义空间，对每个三元组，首先应该将实体投影到对应的关系空间中，然后再建立从头实体到尾实体的平移关系。

TransR 会为每个关系  $r$  定义一个投影关系  $M_r$ ，然后将实体向量通过  $M_r$  投影到关系  $r$  的子空间。表示如下：

$$\mathbf{h}_{\perp} = \mathbf{M}_r \mathbf{h}, \quad \mathbf{t}_{\perp} = \mathbf{M}_r \mathbf{t}.$$



图四 TransR 模型

除此之外，还有很多类似的 Trans 系列的模型，它们都在各种方面对早期的模型进行改进。如 TransD 模型，TranSparse 模型，TransM 模型，TorusE 模型[24]等。但他们的核心，仍然是早期的平移不变性。

## 2.2 双线性模型

### 2.2.1 RESCAL 模型

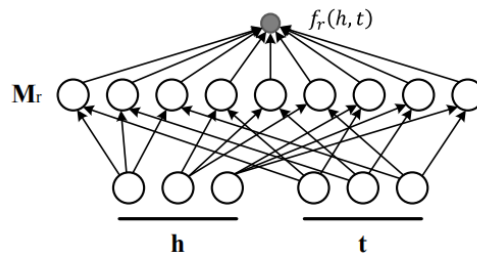
RESCAL 模型[18]将每个关系都表示为一个矩阵，使用  $M_r$  表示：

$$\mathbf{M}_r = \sum_i \pi_r^i \mathbf{u}_i \mathbf{v}_i^T$$

该矩阵对潜在因素之间的成对交互进行了建模。同时，它把三元组  $(h, r, t)$  的评分函数定义为一个双线性函数，

$$f_r(h, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t} = \sum_{i=0}^{d-1} \sum_{j=0}^{d-1} [\mathbf{M}_r]_{ij} \cdot [\mathbf{h}]_i \cdot [\mathbf{t}]_j$$

这也是双线性模型的来由。它使用张量分解方法，能够通过分解的潜在组成部分执行集体学习。

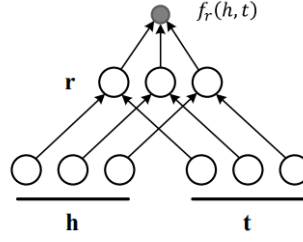


图五 RESCAL 模型

### 2.2.2 其他双线性模型

RESCAL 使用张量积，这些张量积捕获了丰富的交互作用，但是需要大量参数来建立关系模型，计算复杂度较高。

为了克服这些缺点，DistMult 模型[20]通过将  $M_r$  限制为对角矩阵来简化 RESCAL。对于每一个关系  $r$ ，它会引入一个向量嵌入  $\mathbf{r}$ ，并要求  $M_r = \text{diag}(\mathbf{r})$ 。



图六 DistMult 模型

评分函数只捕获沿同一维度的  $h$  和  $t$  分量之间的成对交互作用。

$$f_r(h, t) = \mathbf{h}^\top \text{diag}(\mathbf{r}) \mathbf{t} = \sum_{i=0}^{d-1} [\mathbf{r}]_i \cdot [\mathbf{h}]_i \cdot [\mathbf{t}]_i.$$

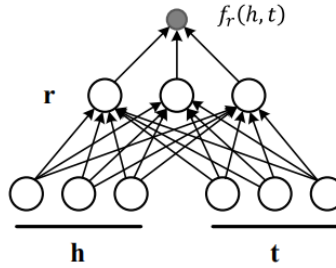
然而，这种过度简化的模型只能处理对称的关系，这显然对一般的知识图谱是不能完全适用的。

HoIE 模型[19]则将 RESCAL 模型的表达能力和 DistMult 模型的效率和简单性相结合。使用循环相关操作将实体表示形式组成  $\mathbf{h} \star \mathbf{t}$ ，

$$[\mathbf{h} \star \mathbf{t}]_i = \sum_{k=0}^{d-1} [\mathbf{h}]_k \cdot [\mathbf{t}]_{(k+i) \bmod d}.$$

然后将组合后的向量与关系表示相结合，进行评分。

$$f_r(h, t) = \mathbf{r}^\top (\mathbf{h} \star \mathbf{t}) = \sum_{i=0}^{d-1} [\mathbf{r}]_i \sum_{k=0}^{d-1} [\mathbf{h}]_k \cdot [\mathbf{t}]_{(k+i) \bmod d}.$$



图七 HoIE 模型

CompIEEx 模型[21]通过引入复值嵌入来扩展 DistMult，以便更好地对非对称关系进行建模。 $h, t, r$  的嵌入于是被映射到了复空间中，评分函数被设置为

$$f_r(h, t) = \text{Re}(\mathbf{h}^\top \text{diag}(\mathbf{r}) \bar{\mathbf{t}}) = \text{Re}\left(\sum_{i=0}^{d-1} [\mathbf{r}]_i \cdot [\mathbf{h}]_i \cdot [\bar{\mathbf{t}}]_i\right),$$

最近的研究表明，每个 CompIEEx 都有一个等价的 HoIE，反之，在嵌入上施加共轭对称 CompIEEx 包含 HoIE 的特殊情况。

ANALOGY 模型[1]扩展了 RESCAL，它对实体和关系的类比属性进行建模，作者称这个模型是包括 DistMult 在内的很多模型的统一。它遵循 RESCAL 模型并使用双线性评分函数

$$f_r(h, t) = \mathbf{h}^\top \mathbf{M}_r \mathbf{t},$$

为了对类比结构进行建模，它进一步要求关系线性映射  $M_r$  是正规的和符合交换

律，即要满足：

$$\text{normality: } \mathbf{M}_r \mathbf{M}_r^\top = \mathbf{M}_r^\top \mathbf{M}_r, \forall r \in \mathbf{R};$$

$$\text{commutativity: } \mathbf{M}_r \mathbf{M}_{r'} = \mathbf{M}_{r'} \mathbf{M}_r, \forall r, r' \in \mathbf{R}.$$

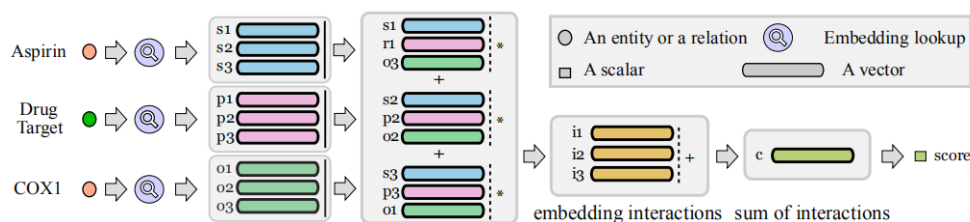
ANALOGY 模型将关系表示为矩阵，这些矩阵可以同时对角化成一组稀疏的准对角矩阵。作者进行了探索，DistMult、HoIE、CompIE<sub>x</sub> 等方法都可以归为 ANALOGY 模型的特例。类比推断对于知识库的完成非常有用。

最近还有 Simple 模型[9]和 TriModel 模型[10]。

Simple 是对 CP（规范多联体）的简单增强，从而可以独立地学习每个实体的两个嵌入。Simple 的复杂度随着嵌入的大小线性增长。通过 Simple 学习的嵌入是可以解释的，并且可以通过权重绑定将某些类型的背景知识并入这些嵌入中。

TriModel 模型的提出是用于进行药物靶标相互作用预测，它与 DistMult 模型有些类似，在一种特殊情况下，DistMult 模型可以看成是 TriModel 模型的特里。其训练过程分为多个阶段。

首先，KGE 模型使用随机噪声对嵌入向量进行初始化。然后，使用一个评分函数对模型进行打分。最后，通过得分函数计算损失误差，并进行训练。通过 AMSGrad 等优化器将误差回传并更新初始嵌入，更新后的嵌入对真实事实给出更高的分数，对伪事实给出较低的分数。通过将这一过程不断迭代，使得所学习的嵌入为可能的真假事实提供最佳的可能评分。



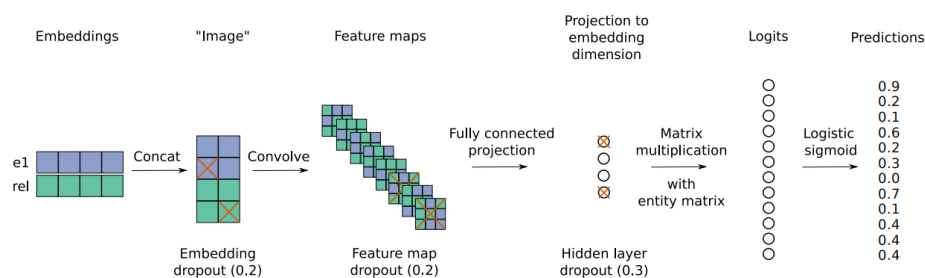
图八 TriModel 模型的评分函数流程

## 2.3 神经网络模型

### 2.3.1 ConvE 模型

由于现在出现的数据集有测试集泄露问题，就是测试集中的三元组可以通过翻转训练集中的三元组得到，即  $(h, r, t) = (t, r, h)$ 。这会导致一个很简单的基于规则的模型就可以在这些数据集上获得非常好的效果。

ConvE 模型[5]，是基于 CNN 的嵌入模型。NLP 中的 CNN 通常使用 Conv1D，就是把嵌入拼接起来进行卷积，而 ConvE 模型用的是 Conv2D，把多个向量堆叠成矩阵，然后就可以像图像一样用二维卷积核来抽取嵌入之间的联系。同时对于处理数据泄露过后的测试集也有很好的效果。



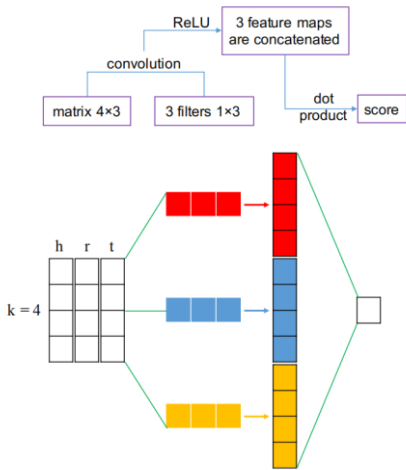
图九 ConvE 模型

在 ConvE 模型中，首先将实体和关系初始化嵌入后，再进行级联和重塑（步骤 1、2）；然后将得到的矩阵用作卷积层的输入（步骤 3）；将得到的特征映射张量矢量化并投影到  $k$  维空间（步骤 4），并与所有候选对象嵌入匹配（步骤 5）。

值得一提的是，与传统模型对三元组关系  $(s, r, o)$  打分的 1-1 scoring 模式不同，ConvE 以实体关系对  $(s, r)$  作为输入，同时对所有实体  $o$  进行打分，即 1-N scoring。这种方式极大加快了计算速度。

### 2.3.2 ConvKB 模型

ConvKB 模型[6]应用卷积神经网络研究了同一维度之间的全局关系，从而使其推广了基于平移的嵌入模型中的过渡特性。

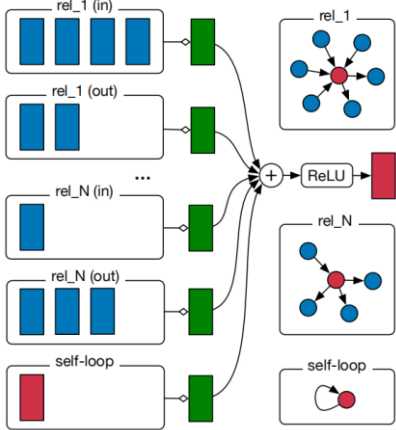


图十 ConvKB 模型涉及的过程

如图所示，设嵌入大小  $k$  为 4，用 3 个不同的滤波器对初始嵌入进行卷积运算，使用 ReLU 作为激活函数。之后将 3 个特征映射拼接起来，使用目标函数优化。

### 2.3.3 R-GCN

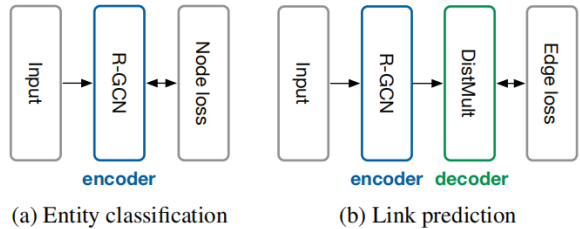
R-GCN 模型[7]基于图的神经网络，是将图卷积网络[8]应用到关系数据的扩展。它对每个实体的邻域应用卷积运算，并为其分配相等的权重。在链接预测和实体分类都取得了非常不错的结果。



图十一 用于计算 R-GCN 模型中单个图形节点/实体（红色）的更新的图表。



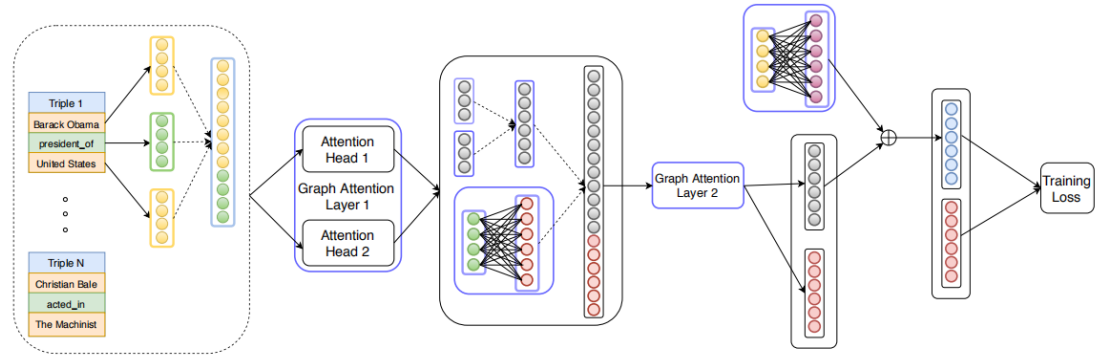
如图所示，从相邻节点(深蓝色)收集激活( $d$  维向量)，然后对每种关系类型分别进行转换(包括内边和出边)。所得到的表示(绿色)以(归一化)和的形式累积，并通过激活函数(如 ReLU)传递。每个节点的更新可以与整个图中的共享参数并行计算。



图十二 实体分类和链接预测的 R-GCN 模型描述

### 2.3.4 基于注意力机制的模型

Nathani 等人[2]在 2019 年 ACL 会议上提出了一种基于知识图谱的注意力网络嵌入方法，该方法在注意力机制中考虑了结点和边的特征。作者提出了一个广义图注意力模型，学习新的基于图关注的嵌入，该嵌入专门针对知识图谱上的关系预测，可捕获任何给定实体的多跳邻域中的实体和关系特征。架构是一个编码器-解码器模型，而广义图注意力模型和 ConvKB 模型分别充当编码器和解码器的角色。



图十三 广义图注意力模型的端到端结构

可以从图中看出，作者采用了多头注意力架构，在最后一层中，使用了类似于残差网络的思想，保留最开始输入的嵌入。这个模型表现出了非常好的性能。

## 3 总结

为了更好的使用知识图谱为各个领域服务，知识图谱嵌入技术应运而生。从早期的 Trans 系列模型，双线性模型，到后来兴起的神经网络模型，可以说，每一次的提升，都会促进知识图谱的应用。无论是医疗、互联网金融还是推荐系统等，都焕发出新的生机。

但即使如此，现有的模型仍然有一些缺点。它们要么仅关注实体特征，要么单独考虑每个三元组，而没有考虑它们之间的关系，又或者无法太好的解决分层图的情况，无法太好的捕获实体之间的高级主题等。但是，庆幸还有问题留给我们。

## 参考文献

- [1] Liu H, Wu Y, Yang Y. Analogical Inference for Multi-Relational Embeddings. ICML'17 Proceedings of the 34th International Conference on Machine Learning(ICML) - Volume 70. 2017. Pages 2168-2178.
- [2]Nathani D, Chauhan J, Sharma C, et al. Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics(ACL), 2019, pages 4710 - 4723.
- [3]Wang Q, Mao Z, Wang B, et al. Knowledge Graph Embedding: A Survey of Approaches and Applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, PP(99):1-1.
- [4] Zhang, Yongqi, Yao, Quanming, Dai, Wenyuan, et al. AutoKGE: Searching Scoring Functions for Knowledge Graph Embedding[J]. 2019.
- [5]Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In Thirty-Second AAAI Conference on Artificial Intelligence (AAAI), 2018.
- [6]Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Phung. 2018. A novel embedding model for knowledge base completion based on convolutional neural network. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), 2018, volume 2, pages 327 - 333.
- [7]Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In European Semantic Web Conference (ESWC), 2018, pages 593 - 607.
- [8] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations (ICLR), 2017.
- [9]M. Kazemi and D. Poole. Simple embedding for link prediction in knowledge graphs. In NeurIPS, 2018.
- [10]Mohamed S K, Novacek V, Nounu A, et al. Discovering Protein Drug Targets Using Knowledge Graph Embeddings. In Bioinformatics, 2019.
- [11] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on freebase from question-answer pairs. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2013, pages 1533 - 1544.
- [12] Jonathan Berant and Percy Liang. 2014. Semantic parsing via paraphrasing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.
- [13]Yuanzhe Zhang, Kang Liu, Shizhu He, Guoliang Ji, Zhanyi Liu, Hua Wu, and Jun Zhao. 2016. Question answering over knowledge base with



- neural attention combining global knowledge information. arXiv preprint arXiv:1606.00979, 2016
- [14] Dennis Diefenbach, Kamal Singh, and Pierre Maret. 2018. Wdaqua-corel: a question answering service for rdf knowledge bases. In Companion of the The Web Conference 2018 on The Web Conference (WWW), 2018, pages 1087–1091. International World Wide Web Conferences Steering Committee.
- [15] Xiang Wang, DingxianWang, Canran Xu, Xiangnan He, Yixin Cao, TatSeng Chua. Explainable Reasoning over Knowledge Graphs for Recommendation. Proceedings of the AAAI Conference on Artificial Intelligence(AAAI). 2019 .
- [16] Wang X, He X, Cao Y, et al. KGAT: Knowledge Graph Attention Network for Recommendation. In KDD. 2019.
- [17] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, Tat-Seng Chua. Unifying Knowledge Graph Learning and Recommendation: Towards a Better Understanding of User Preferences. In WWW. 2019. Pages 151–161.
- [18] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In Proceedings of the 28th International Conference on International Conference on Machine Learning, (ICML), 2011.
- [19] Maximilian Nickel, Lorenzo Rosasco, Tomaso A Poggio, et al. 2016. Holographic embeddings of knowledge graphs. In (AAAI), 2016, volume 2, pages 3–2.
- [20] Antoine Bordes, Nicolas Usunier, Alberto GarciaDuran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems (NIPS), 2013, pages 2787–2795.
- [21] Theo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In International Conference on Machine Learning (ICML), 2016, pages 2071–2080.
- [22] Z. Wang, J. Zhang, J. Feng, and Z. Chen, Knowledge graph embedding by translating on hyperplanes, in Proc. 28th AAAI Conf. Artif. Intell., 2014, pp. 1112–1119.
- [23] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, Learning entity and relation embeddings for knowledge graph completion, in Proc. 29th AAAI Conf. Artif. Intell., 2015, pp. 2181–2187.
- [24] Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J. Smola, and Le Song. TorusE: Knowledge Graph Embedding on a Lie Group. In Proceedings of the AAAI Conference on Artificial Intelligence(AAAI). 2018