

# SparseMult: 基于稀疏关系矩阵的知识图谱张量分解模型

谢志文, 刘进\*

武汉大学计算机学院, 湖北武汉, 430072

{xiezhwen, jiuliu}@whu.edu.cn

**摘要** 链接预测又称为知识图谱补全, 目的是预测知识图谱中缺失的三元组。基于张量分解的模型是解决链接预测任务的一种简单而有效方法。本文提出了一种新的基于稀疏关系矩阵的知识图谱张量分解模型 SparseMult 来解决链接预测问题。本文将该模型应用到 CCKS2021 表型-药物-分子多层次知识图谱的链接预测任务中进行了实验, 实验结果表明本文提出的 SparseMult 模型效果优于很多先进的模型。

**关键词:** 知识图谱 · 链接预测 · 张量分解.

## 1 引言

知识图谱中包含了丰富的结构化的知识, 可以为很多下游自然语言处理任务提供支撑。但是, 知识图谱通常是不完整的, 实体之间有很多缺失的关系。链接预测又称为知识图谱补全, 目的是预测知识图谱中缺失的三元组。基于张量分解的模型是解决链接预测任务的一种简单而有效方法。很多先进的链接预测模型, 比如 Rescal [2]、DistMult [5] 等, 都是基于张量分解的。其中, Rescal 模型参数量很大, 难以扩展到大规模知识图谱上。DistMult 对关系矩阵进行了简化, 使用对角矩阵来表示关系, 从而极大降低了参数量。但是, DistMult 无法处理非对称的关系, 过度简化的关系表示限制了模型的性能。

针对以上问题, 我们提出了一个新的基于稀疏关系矩阵的张量分解模型: SparseMult。Rescal [2] 模型中参数量大的主要原因在于每个关系矩阵具有  $O(d^2)$  的空间复杂度。DistMult 通过将关系矩阵简化为关系向量, 把关系参数的空间复杂度缩减到  $O(d)$ 。与 DistMult [5] 不同, 为了降低参数量, 我们使用一个稀疏的分块对角矩阵来表示关系, 将每个关系矩阵空间复

---

\* 通讯作者

杂度降低到  $O(kd)$ , 其中  $d$  是矩阵的维度,  $k$  是稀疏的分块对角矩阵中每个小方阵的维度。从而大大降低了模型的参数量, 并且解决了 DistMult 存在的表达能力不足, 无法处理非对称关系的问题。

我们参加了 CCKS 2021: 表型-药物-分子多层次知识图谱的链接预测的评测任务<sup>1</sup>, 将我们提出的 SparseMult 模型应用到该知识图谱中。我们在 CCKS2021 表型-药物-分子多层次知识图谱上进行了大量实验, 实验结果表明 SparseMult 模型的效果不仅优于 Rescal 和 DistMult 模型, 并且超越了目前的很多 SOTA 模型, 比如 ComplEx [4], RotatE[3]、Tucker[1] 等。

## 2 方法

### 2.1 SparseMult 模型

张量分解是知识图谱链接预测的主流方法之一。传统的张量分解模型存在参数量大, 难以扩展到大规模知识图谱的问题, 比如, Rescal 模型。DistMult 对关系矩阵进行了简化, 使用对角矩阵来表示关系, 从而极大降低了参数量。但是 DistMult 无法处理非对称的关系, 过度简化的关系表示限制了模型的性能。

本文提出了一个新的基于稀疏关系矩阵的张量分解模型: SparseMult。SparseMult 的核心思想是将关系矩阵表示成一个稀疏的分块对角矩阵, 即:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & 0 & 0 & 0 \\ 0 & \mathbf{R}_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{R}_{ss} \end{bmatrix} \quad (1)$$

其中,  $\mathbf{R} \in \mathbb{R}^{d \times d}$ ,  $\mathbf{R}_{ii} \in \mathbb{R}^{k \times k}$  是分块对角矩阵  $\mathbf{R}$  中的第  $i$  个方阵,  $k$  是每个方阵的维度,  $s = \frac{d}{k}$  为  $\mathbf{R}$  中方阵的数量。

基于公式 (1) 中的分块对角矩阵的关系表示, 对于给定的一个三元组  $(h, r, t)$ , 三元组的评分函数可以表示为:

$$f(h, r, t) = \mathbf{h}^T \mathbf{R} \mathbf{t} \quad (2)$$

其中,  $\mathbf{h}$  和  $\mathbf{t}$  分别是头实体和尾实体的向量表示。

<sup>1</sup> [https://www.biendata.xyz/competition/ccks\\_2021\\_kg\\_link\\_prediction/](https://www.biendata.xyz/competition/ccks_2021_kg_link_prediction/)

$$\begin{array}{ccc}
 \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{bmatrix} & 
 \begin{bmatrix} r_{11} & 0 & 0 & 0 \\ 0 & r_{22} & 0 & 0 \\ 0 & 0 & r_{33} & 0 \\ 0 & 0 & 0 & r_{44} \end{bmatrix} & 
 \begin{bmatrix} r_{11} & r_{12} & 0 & 0 \\ r_{21} & r_{22} & 0 & 0 \\ 0 & 0 & r_{33} & r_{34} \\ 0 & 0 & r_{43} & r_{44} \end{bmatrix} \\
 \text{Rescal} & \text{DistMult} & \text{SparseMult}
 \end{array}$$

图 1. 不同模型的关系表示对比。

由于稀疏的分块对角矩阵中每个方阵是相互独立的，无法学习不同方阵之间的交互特征。为了解决这个问题，我们增加了一个转换矩阵整合不同的方阵学习到的特征，从而进一步增强模型的性能。最终的评分函数为：

$$f(h, r, t) = \mathbf{h}^T \mathbf{R} \mathbf{W} t \quad (3)$$

其中  $\mathbf{W} \in \mathbb{R}^{d \times d}$  是可训练的参数。

## 2.2 模型训练

给定一个三元组  $(h, r, t)$ ，我们使用 softmax 计算尾实体的概率：

$$P(t|h, r) = \frac{\exp(f(h, r, t))}{\sum_{t' \in \mathcal{E}} \exp(f(h, r, t'))} \quad (4)$$

其中， $\mathcal{E}$  是所有的实体的集合。然后通过最小化负对数损失函数对模型参数进行优化，损失函数计算公式为：

$$\mathcal{L} = -\log(P(t|h, r)) \quad (5)$$

对于每个三元组  $(h, r, t)$  我们构造了反转的三元组  $(t, r\_reverse, h)$ ，当需要预测头实体时，即给定  $query = (?, r, t)$ ，预测  $h$  时，我们使用它的反转形式预测，即预测  $(t, r\_reverse, ?)$  的尾实体。

## 2.3 模型分析

本文提出的 SparseMult 模型使用稀疏的分块对角矩阵来表示关系，不仅可以减少参数量，并且不会过度简化模型。图 1 对比了不同模型的关系矩阵，Rescal 的参数量较大，具有  $O(d^2)$  的空间复杂度，而 DistMult 使用对角矩阵表示关系，空间复杂度为  $D(d)$ 。但是这也限制了 DistMult 对知识图谱的建模能力，使它无法处理非对称的关系。SparseMult 的分块对角矩阵可

以有效的解决这些问题。由公式 (1) 可知, 对于每个关系, 其分块对角矩阵  $\mathbf{R}$  中参数量为:

$$O(k \times k \times s) = O(k \times k \times d/k) = O(kd) \quad (6)$$

当  $k$  较小时, 可以有效降低参数数量, 并不损害模型对知识图谱的建模能力。表 1 中对比了三种模型的空间复杂度。

表 1. 模型复杂度对比。| $R$ | 表示关系的数量, | $E$ | 表示实体的数量。

模型	空间复杂度
Rescal	$ R d^2 +  E d$
DistMult	$ R d +  E d$
SparseMult	$ R kd +  E d + d^2$

相比于 Rescal 和 DistMult 等经典的张量分解模型, SparseMult 具有以下优势:

- 参数量小, 可扩展性强。当  $k$  较小时参数量不会随着关系数量急剧增加。通过合理设置  $k$  的值, SparseMult 模型可以扩展到大规模知识图谱上。
- 模型表达能力更强。相比于 DistMult, 我们的 SparseMult 模型可以处理非对称的关系, 具有更强的表达能力。
- 模型性能更好。我们在 CCKS2021 表型-药物-分子多层次知识图谱上进行了大量实验, 实验结果表明 SparseMult 模型的效果优于 Rescal 和 DistMult 模型, 并且超越了目前的很多 SOTA 模型, 比如 RotatE, Tucker 等。

## 3 实验

### 3.1 数据

实验使用的数据集为 CCKS 2021: 表型-药物-分子多层次知识图谱的链接预测任务的评测数据集<sup>2</sup>。表型-药物-分子多层次知识图谱依据表型(疾病、症状)、药物、基因等及其之间的关系来构建的知识图谱, 其包含了大量实体及其关系数据, 可以为致病机理和药理作用机制的研究提供一定支持。

<sup>2</sup> [https://www.biendata.xyz/competition/ccks\\_2021\\_kg\\_link\\_prediction/](https://www.biendata.xyz/competition/ccks_2021_kg_link_prediction/)

总共包含 80000 左右实体, 1200000 左右的三元组, 训练集与测试集的比例约 9 : 1。本文使用 Top10 结果的 MRR(Mean reciprocal rank, 平均倒数排名) 值作为评测指标, 使用 MRR@10 表示。

### 3.2 实验设置

我们使用深度学习框架 Pytorch 实现模型, 并在 NVIDIA RTX 2080Ti 上进行实验。使用 AdamW 作为优化器训练模型参数, 初始学习率设置为 0.001。实体和关系的 embedding 的维度  $d = 3200$ 。Batch size 设置为 1024, Dropout 设置为 0.5。为了加快模型收敛, 我们使用 Batch Normalization 进行正则化。

### 3.3 实验结果

表 2. 在验证集上的实验结果。

模型	维度	MRR@10
Rescal	$d = 200$	0.149
TuckER	$d = 200$	0.152
SparseMult	$d = 200$	0.158
DistMult	$d = 1000$	0.161
ComplEx	$d = 1000$	0.163
RotatE	$d = 1000$	0.169
SparseMult	$d = 1000$	0.178
SparseMult	$d = 3200$	0.194

实验结果如表 2 所示, 我们跟一些基线模型进行了对比, 包括 Rescal, TuckER, SparseMult, ComplEx 和 RotatE 等。为了保证对比的公平性, 我们在相同的维度的情况下进行了对比。由于 Rescal 和 TuckER 模型的参数量随着维度增加急剧膨胀, 为了方便实验我们设置  $d = 200$  进行了对比。相比与 Rescal 和 TuckER, 我们提出的 SparseMult 在 MRR@10 上取得了较好的效果。对于其他的基线模型, 我们设置了  $d = 1000$  进行对比。与 DistMult, ComplEx 和 RotatE 相比, SparseMult 也取得了更好的结果。以上实验结果验证了本文提出的模型的有效性。

由表 2 不同维度的对比我们可以看到, 随着维度的增加, SparseMult 的效果有明显提升。在比赛中我们最终选择  $d = 3200$ 。对于分块对角矩阵中每个小方阵的维度  $k$ , 我们选择不同大小的值训练多个不同的模型, 并把根据多个模型的平均 MR 的对数进行重新排序, 融合多个模型的结果。最终在测试集上 MRR@10 效果达到 0.189, 位列第二名。

## 4 总结和展望

本文提出了一种新的基于稀疏关系矩阵的知识图谱张量分解模型: SparseMult, 用来解决知识图谱链接预测问题。SparseMult 模型有效的减少了张量分解模型 (e.g., Rescal) 的参数量, 提高了模型的可扩展性。我们参加了 CCKS2021 表型-药物-分子多层次知识图谱的链接预测任务, 我们提出的 SparseMult 模型在该任务中在测试集上 MRR@10 达到 0.189, 取得了第二名的成绩。实验结果表明本文提出的 SparseMult 模型效果优于很多基线模型。

## 参考文献

1. Balazevic, I., Allen, C., Hospedales, T.M.: Tucker: Tensor factorization for knowledge graph completion. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. pp. 5184–5193 (2019)
2. Nickel, M., Tresp, V., Kriegel, H.: A three-way model for collective learning on multi-relational data. In: Getoor, L., Scheffer, T. (eds.) Proceedings of the 28th International Conference on Machine Learning. pp. 809–816. Omnipress (2011)
3. Sun, Z., Deng, Z., Nie, J., Tang, J.: Rotate: Knowledge graph embedding by relational rotation in complex space. In: 7th International Conference on Learning Representations (2019)
4. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: Balcan, M., Weinberger, K.Q. (eds.) Proceedings of the 33rd International Conference on Machine Learning. JMLR Workshop and Conference Proceedings, vol. 48, pp. 2071–2080. JMLR.org (2016)
5. Yang, B., Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: 3rd International Conference on Learning Representations (2015)