

知识表示学习模型的研究

Anonymous

北京航空航天大学计算机学院

摘要 摘要。

关键字： 表示学习，知识图谱，知识表示

1 引言

知识图谱是一种可以结构化地表示真实世界中的实体和实体间关系的网络，广泛应用于搜索引擎等领域。近年来，大量开源知识图谱，如 WordNet [9]、FreeBase [1] 和 Yago [13] 的出现，验证了知识图谱模型在真实世界数据集中的表现。而知识表示学习 [8]，则是将知识图谱中的实体和关系用词向量的形式来表示，使得知识图谱相关研究不必拘束于图的算法。本文将针对近年来具有重要意义的知识表示学习模型进行研究和对比，并介绍知识表示学习的相关应用。

2 背景介绍

2.1 知识图谱

知识图谱是用来表示真实世界中的实体和实体间关系的图结构网络，由节点和边组成，其中节点代表实体，边代表实体之间的关系，通常用三元组 $G = \{E, R, S\}$ 来表示知识图谱，其中 E 为实体的集合， R 为关系的集合， S 是实体-关系-实体的三元组集合，即 $S \subseteq E \times R \times E$ 。此外，一个实体-关系-实体的三元组也可以表示为 (h, r, t) 的格式，其中 h 和 t 分别代表头实体和尾实体， r 表示头实体和尾实体之间的关系。

2.2 知识图谱数据集

著名知识图谱开源数据集有 WordNet、FreeBase 和 Yago 等。其中 WordNet 数据集是由专家构建的词典知识库，FreeBase 数据集则是由志

愿者构建的世界知识库，包含人物、地点、分类等信息，Yago 则是一个从半结构化文本中学习出的知识库。知识表示学习的相关算法，大多基于 WordNet 和 FreeBase 进行验证。此外，由于具体应用的特殊性，往往需要将三元组 S 根据实体之间的对应关系分为 $1-1$ 、 $1-n$ 、 $n-1$ 、 $n-n$ 四种。但在 ConvE 模型 [5] 中发现，这两个数据集的子集 FB15K 和 WN18 存在严重的测试集泄漏问题，即对于某个训练集中的三元组，反转其头实体和尾实体，就可以在测试集中找到相同的数据。故在此后的研究中，普遍使用去除泄漏数据的 FB15K-237 和 WN18-RR 数据集进行训练、验证和预测。

3 知识表示学习

知识图谱通常以图的形式存储，这种图结构不仅不利于对知识图谱的进一步研究，更需要设计特定的算法进行相应的计算。因此通常将图结构转化为向量的形式，便于更加高效的解决知识图谱领域的关系提取、关系预测、实体预测等问题。

通过向量表示实体的方法通常有两种。在传统方法中，采用独热编码来构建向量。对一个有 n 个实体数据集，需要对每个实体使用一个 n 维向量来表示。每个向量中只有 1 维的值为 1，其余值均为 0。这种方法存在明显的问题：随着数据集规模的增加，向量维数过大；又由于向量之间两两正交，不仅无法表示出实体的语义信息，更无法表示实体之间的相似程度。另一种构建词向量的方法则是通过机器学习算法，得到每一个实体的低维稠密向量。对于一个词向量而言，单独拿出一维数值可能毫无意义，但组成词向量后可以表示出相应实体的语义信息，这就是表示学习。

所谓知识表示学习，即对知识库运用表示学习的方法。通过用低维度的稠密向量来表示知识库中的实体或关系，从而进行实体预测、实体分类、语义相似性分析等工作。常用的知识表示学习模型可以分为线性模型、神经网络模型、翻译模型等。

4 知识表示学习模型

4.1 距离模型及其扩展模型

结构表示模型 (Structured Embeddings, SE) [4] 是最经典的线性模型，SE 模型也被称为距离模型。对于知识图谱中的一个头实体-尾实体-关

系三元组 $\{E_h, E_t, R_k\}$ ，SE 模型通过将第 x 个实体 E_x 表征为一个 d 维（约50维）的词向量 l_x ，将第 k 个关系 R_k 构造两个 $d \times d$ 维的投影矩阵 W_k^h 和 W_k^t ，并定义了如下的损失函数来计算实体之间的语义相关性：

$$S_k(E_h, E_t) = \|W_k^h l_h - W_k^t l_t\|_p$$

在 SE 模型中，取范数 $p = 1$ ，此时损失函数为：

$$S_k(E_h, E_t) = \sum_{x=1}^d S_k(E_h, E_t)_x$$

即将头实体和尾实体的词向量分别用一个投影矩阵投影，投影后的距离为投影后得到的向量差按位相加的结果。SE 模型认为，实体投影后的距离越小，头实体和尾实体的越有可能存在该种关系。因此只需不断迭代使得损失函数最小，求出实体向量 l 和投影矩阵 W 两个参数的最优值。应用在实体预测时，只需通过头实体词向量和两个关系矩阵，即可计算出尾实体映射在该关系空间中的词向量。由于损失函数采用 $L1$ 范数，SE 模型又被称为距离模型。

但使用两个投影矩阵来定义一种关系，很难直观的表征出实体之间的关系。在现实世界中，实体 king 与 queen 之间的关系，应该和实体 man 与 woman 的关系有一定联系。换言之，这两个关系向量应该具有某种相似性。但 SE 模型通过两个投影矩阵来表示一种联系后，很难反映出其中的语义联系。此后出现的单层感知机模型（Single Layer Model, SLM）[12]和多层感知机模型（Multi Layer Perceptron, MLP）[6]，试图在 SE 模型的基础上解决这个问题。SLM 模型采用了和 SE 模型一样的两个投影矩阵，其评价函数为：

$$g(E_h, E_t, R_k) = u_k^T \tanh(W_k^h l_h + W_k^t l_t)$$

其中 u_k 为关系 R_k 相关的参数。而 MLP 模型可以看作在 SLM 模型的基础上引入多个隐藏层的结果。SLM 模型和 MLP 模型并非线性模型，而是利用神经网络进行计算，由此带来了挖掘实体和关系之间隐含语义联系的可能性，一定程度上解决了 SE 模型中无法计算实体之间和关系之间的语义联系的问题，但与此同时也大幅提高了参数计算的复杂度，且模型的可解释性较差。

为了进一步解决距离模型中，使用两个投影矩阵无法直观表现关系之间的语义联系的问题，出现了双线性模型。隐变量模型（Latent Factor

Model, LFM) [7]和 DISTMULT 模型 [15]就是双线性模型的代表。LFM 模型可以直观的通过其评价函数来理解:

$$g(E_h, E_t, R_k) = l_h^T W_k l_t^T$$

即头实体向量 l_h 或尾实体向量 l_t 都可以通过一个和关系矩阵的线性变换得到对方。双线性将关系表征为一个二维矩阵,大大简化了距离模型的投影矩阵。而 DISTMULT 模型根据 LFM 模型中和权重乘积的过程,将关系矩阵限制为对角矩阵,即主对角线外的元素值都为0,不仅提升了 LFM 模型的效果,还进一步降低关系矩阵参数的规模。

针对翻译模型中的 TransE 模型 [3]中采取了平移的方法,该论文还提出了优化后的 DISTADD 模型 [15], DISTADD 模型和 DISTMULT 模型在实体分类问题上的准确率见表1。相比之下, DISTMULT 模型的效果更胜一筹。

表 1. DISTMULT 模型和 DISTADD 模型在实体分类问题上的准确性

模型	头实体预测准确率				尾实体预测准确率			
	1-1	1-n	n-1	n-n	1-1	1-n	n-1	n-n
DISTADD	70.0	76.7	21.1	53.9	68.7	17.4	83.2	57.5
DISTMULT	75.5	85.1	42.9	55.2	73.7	46.7	81.0	58.8

除此之外,还有双线性的语义匹配能量模型 (Semantic Matching Energy, SME) [2]等模型,同样在 SE 模型的基础上尝试构建实体和关系之间的潜在语义联系。

4.2 张量分解模型及其扩展模型

张量分解模型 (RESICAL) [11]又被称为矩阵分解模型,它和 LFM 的思想类似,对第 k 个头实体-关系-尾实体的三元组,定义一个三阶张量:

$$\chi_k = AR_k A^T$$

其中 A 为因子矩阵,是所有实体向量拼接成的矩阵; R_k 为核心矩阵,是关系 k 的参数。参数的求解是通过矩阵分解和交替最小二乘法,图1表示出了 RESICAL 模型的矩阵分解过程。

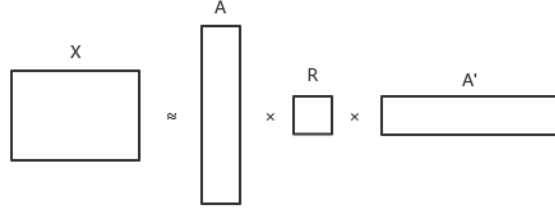


图 1. RESCAL 模型的矩阵分解

在 RESCAL 模型中，定义损失函数为：

$$\min_{A, R_k} = \frac{1}{2} \left(\sum_k \|\chi_k - AR_k A^T\|_F^2 \right) + \frac{1}{2} \lambda (\|A\|_F^2 + \sum_k \|R_k\|_F^2)$$

对于每一个可能的三元组，计算其张量 χ_k 值。在关系预测上，认为张量 χ_k 大于阈值则认为关系存在（即认为 $\chi_k = 1$ ），小于阈值则认为关系不存在（即认为 $\chi_k = 0$ ）。RESCAL 模型也是一个双线性模型，同样也存在参数过大的问题。

张量神经网络模型（Neural Tensor Networks, NTN）[12]则是双线性模型和神经网络模型的结合。NTN 模型是在 SLM 模型的基础上引入实体和关系的三阶双线性张量，如图2所示。

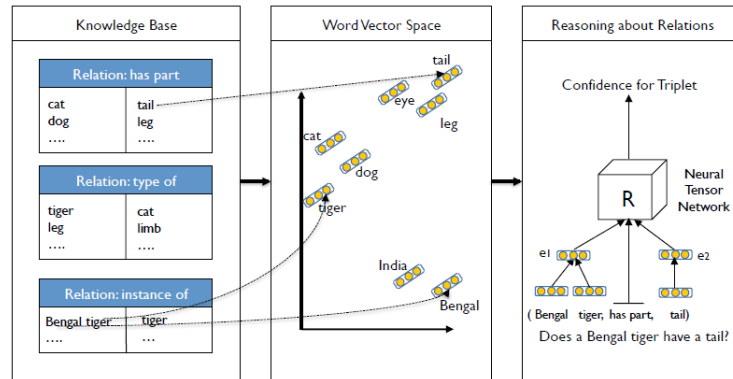


图 2. NTN 模型

在 NTN 模型中，还对向量进行预训练作为初始值，而不是通常采用的随机生成，这样做的好处是尽可能使用实体和关系的语义信息，尤其是在比较稠密的知识图谱中，对模型效果的提高十分明显。NTN 模型相比上述模型，对实体和联系的语义信息体现的更好，但 NTN 模型所需要的参数较多，难以在知识图谱较稀疏时发挥作用。

此外，还有使用循环相关（Circular Correlation）代替 RESCAL 模型张量乘积操作的全息表示模型（Holographic Embeddings, HolE）[10]，该循环相关降低了 RESCAL 模型中的参数维度。

4.3 翻译模型及其扩展模型

在距离模型及其扩展模型中，关系通常被表征为一个或两个矩阵进行点乘映射。TransE 模型 [3]被称为翻译模型或隐距离模型，它将关系表征为一个向量 l_r 而非矩阵，其“翻译”的过程可以表示为：

$$l_t \approx l_h + l_r$$

即将尾实体的词向量看作头实体按照关系进行平移的结果，图3直观的反映

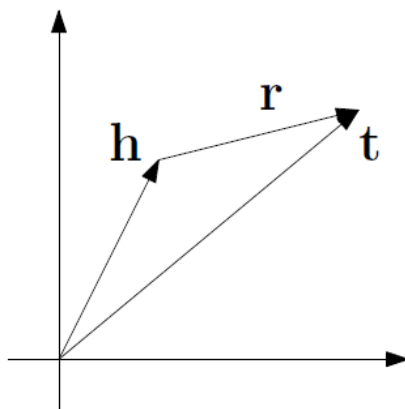


图 3. TransE 模型的翻译过程

出了 TransE 模型翻译的过程。相应的，TransE 模型的评分函数为：

$$f_r(E_h, E_t) = d(l_h + l_r, l_t)$$

通过这个平移的过程，TransE 模型可以很好的表示出实体 king 与 queen 的联系和实体 man 与 woman 的联系语义相关性。但实验 [3]表明，TransE 模型对 $1-n$ 问题的效果并不理想，在实体预测问题上，TransE 模型的准确率见表2。

表 2. TransE 模型在实体预测问题上的准确性

模型	头实体预测准确率				尾实体预测准确率			
	1-1	1-n	n-1	n-n	1-1	1-n	n-1	n-n
SME	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
TransE	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0

对于 $1-n$ 问题，如“苹果公司的创始人有谁”应该有多个尾实体符合。但经过 TransE 模型训练后，由 $l_h + l_r$ 得到的尾实体向量相同，即多个创始人有着相同的词向量，这显然与显示不符。但由于 TransE 模型可以通过向量直接建立语义联系，且参数大幅减少，性能提升较大，逐渐成为了相当经典的模型。后续许多扩展模型都是根据 TransE 翻译模型进行的进一步改进。

针对 TransE 模型中在 $1-n$ 、 $n-1$ 和 $n-n$ 问题上的缺陷，提出了 TransH 模型 [14]。不同于 TransE 模型中，简单的将头实体词向量根据关系向量平移，TransH 模型采用了向量投影的方法，如图4所示。

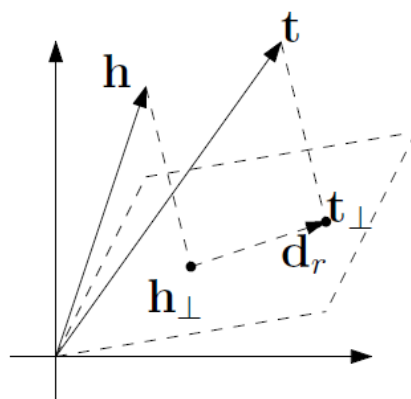


图 4. TransH 模型的翻译过程

在 TransH 模型中, 认为在不同的关系中的实体应该有不同的表示。TransH 模型通过定义投影矩阵 W_r , 将头实体向量 l_h 和尾实体向量 l_t 投影到投影空间, 得到 $l_{h\perp}$ 和 $l_{t\perp}$ 向量。此后再使用 TransE 模型的方法进行翻译:

$$l_{t\perp} \approx l_{h\perp} + l_r$$

而投影的过程, 可以表示为:

$$l_{h\perp} = l_h - W_r^T h W_r$$

$$l_{t\perp} = l_t - W_r^T t W_r$$

通过投影操作, TransH 模型实际只使用了实体词向量的一个分量, 避免出现 TransE 模型在非 1-1 情况下得到相同尾实体的问题。

5 结论

参考文献

1. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: A collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. SIGMOD '08 (2008)
2. Bordes, A., Glorot, X., Weston, J., Bengio, Y.: Joint learning of words and meaning representations for open-text semantic parsing. In: Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012, La Palma, Canary Islands, Spain, April 21-23, 2012. pp. 127-135 (2012)
3. Bordes, A., Usunier, N., García-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States. pp. 2787-2795 (2013)
4. Bordes, A., Weston, J., Collobert, R., Bengio, Y.: Learning structured embeddings of knowledge bases. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011 (2011)
5. Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.: Convolutional 2d knowledge graph embeddings. In: Proceedings of the Thirty-Second AAAI Conference on

Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018. pp. 1811–1818 (2018)

6. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., Zhang, W.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014. pp. 601–610 (2014)
7. Jenatton, R., Roux, N.L., Bordes, A., Obozinski, G.: A latent factor model for highly multi-relational data. In: Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States. pp. 3176–3184 (2012)
8. Lin, Y., Han, X., Xie, R., Liu, Z., Sun, M.: Knowledge representation learning: A quantitative review. Computing Research Repository abs/1812.10901 (2018)
9. Miller, G.A.: Wordnet: A lexical database for english. COMMUNICATIONS OF THE ACM 38(11) (Nov 1995)
10. Nickel, M., Rosasco, L., Poggio, T.A.: Holographic embeddings of knowledge graphs. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA. pp. 1955–1961 (2016)
11. Nickel, M., Tresp, V., Kriegel, H.: A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011. pp. 809–816 (2011)
12. Socher, R., Chen, D., Manning, C.D., Ng, A.Y.: Reasoning with neural tensor networks for knowledge base completion. In: Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States. pp. 926–934 (2013)
13. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Proceedings of the 16th international conference on World Wide Web. pp. 697–706 (2007)
14. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada. pp. 1112–1119 (2014)

15. 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, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015)