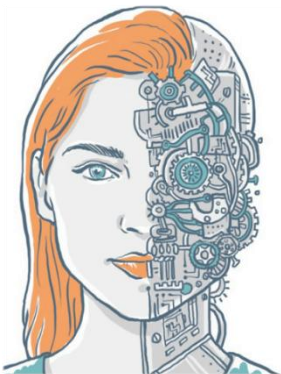


知识表示学习

(Knowledge Representation Learning)

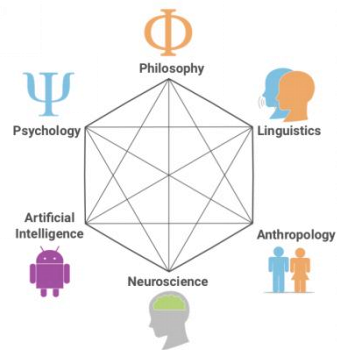


Mind or Machines
Cognitive Science Changing
Artificial Intelligence

汪 鹏

pwang@seu.edu.cn

东南大学 KGCODE实验室



东南大学 计算机科学与工程学院/人工智能学院

- 一、原理及概念
- 二、表示学习方法
 - 翻译模型
 - 语义匹配模型
 - 融合多源信息的模型
- 三、模型评测
- 四、挑战

什么是知识表示学习

- 表示学习：将研究对象的语义信息表示为低维稠密的向量。
研究对象：文字（词汇、短语、句子、文章）、图片、语音等。
。
- 知识表示学习：将知识图谱中的实体和关系表示为低维稠密的向量。

知识图谱中的“知识”

- 知识图谱包括实体和关系
 - 节点代表实体
 - 连边代表关系
- 知识通常用三元组表示
 - $(head, relation, tail)$

例子：（北京，是…的首都，中国）

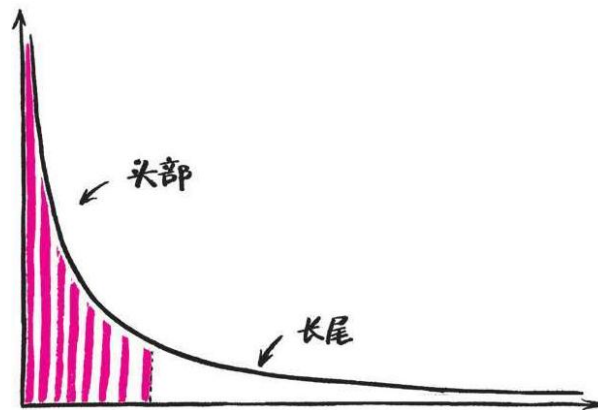
基于网络的表示存在的问题

- 计算效率问题

基于图结构的知识表示虽然简洁直观，但是需要专门的图算法（复杂度高、可扩展性差）。

- 数据稀疏问题

长尾分布，长尾上的实体和关系的语义难以捕获。



独热表示存在的问题

独热表示 (one-hot representation)

假设所有研究对象都是独立的，将研究对象表示为向量，只有某一维非零，其余维度上的值均为0。显然不符合实际情况，导致丢失大量信息。

例子：苹果 $(0, 1, 0, 0, 0, 0, 0)$ ，香蕉 $(0, 0, 0, 1, 0, 0, 0)$

知识表示学习的意义

- 低维向量提高计算效率
- 稠密向量缓解数据稀疏
- 多源的异质信息表示形式统一，便于迁移和融合

提纲

- 一、原理及概念
- 二、表示学习方法
 - 翻译模型
 - 语义匹配模型
 - 融合多源信息的模型
- 三、模型评测
- 四、挑战

翻译模型 — TransE

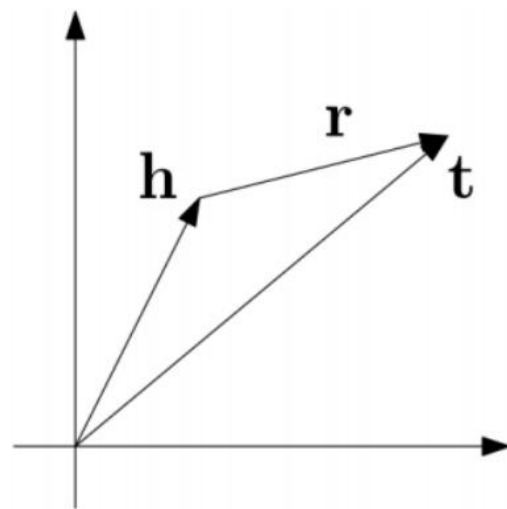
将每条知识 (head, relation, tail) 中的relation看作从head到tail的翻译操作。得分函数：

$$f(h, r, t) = \| h + r - t \|_{l1/l2}$$

例子：

(北京，是…的首都，中国)

$$L = \sum_{\xi \in T} \sum_{\xi' \in T'} [\gamma + f(\xi) - f(\xi')]_+$$



简化版的TransE — UM

仅利用头实体和尾实体的共现信息，而忽略了它们之间的关系。得分函数：

$$f(h, r, t) = - \| h - t \|_{l_1/l_2}$$

Bordes A, Glorot X, Weston J, et al. Joint learning of words and meaning representations for open-text semantic parsing. Artificial Intelligence and Statistics. 2012: 127-135.

与UM相比，SE添加了关系信息：将关系建模为分别针对头实体和尾实体的矩阵。得分函数：

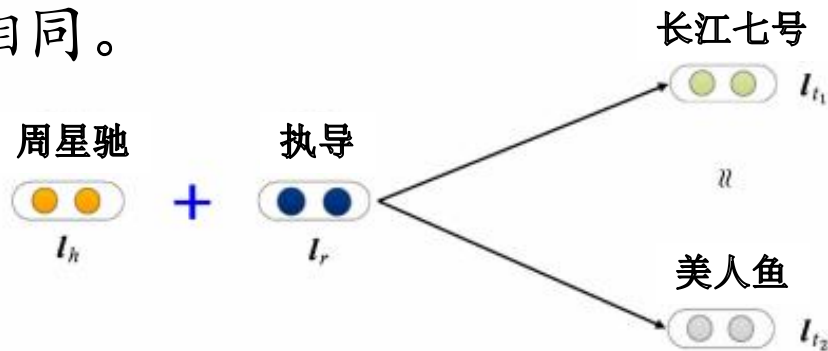
$$f(h, r, t) = - \| M_r^h h - M_r^t t \|_{l_1/l_2}$$

Bordes A, Weston J, Collobert R, et al. Learning structured embeddings of knowledge bases. AAAI. 2011.

TransE的不足

TransE存在的问题：不能很好的处理复杂关系，如：1-N，N-1，N-N关系。

例如：TransE会使“奥巴马”和“布什”的向量变得相同。



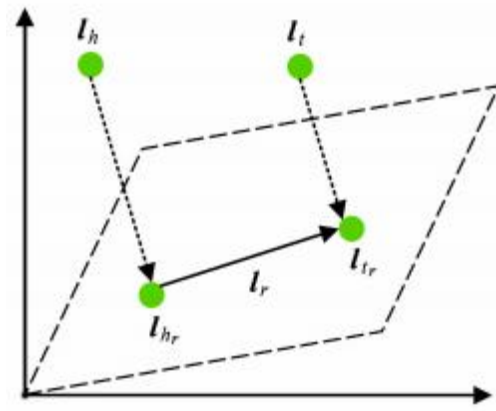
Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data. NIPS2013: 2787-2795.

翻译模型—TransH

为了解决TransE无法很好处理复杂关系的问题，TransH将每种关系建模为一个超平面，将三元组中的头实体和尾实体分别映射到该超平面中。得分函数：

$$h_{\perp} = h - w_r^T h w_r, \quad t_{\perp} = t - w_r^T t w_r$$

$$f(h, r, t) = -\|h_{\perp} + r - t_{\perp}\|_{l1/l2}$$

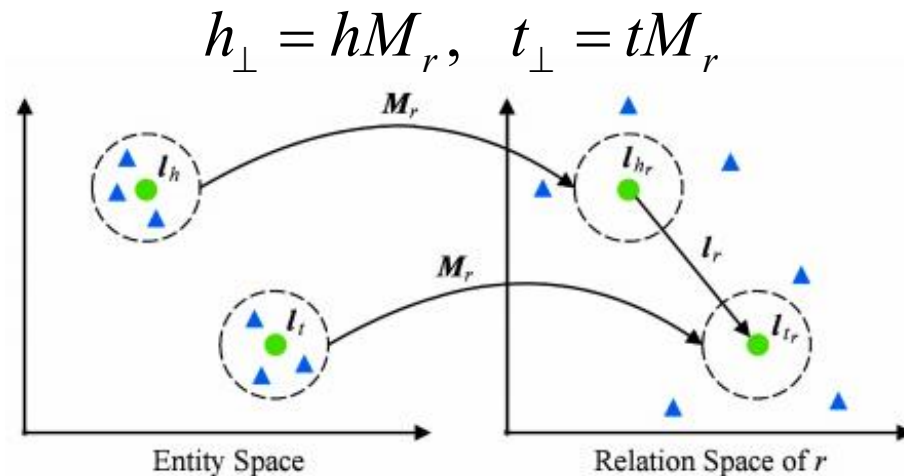


Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes. AAAI2014.

翻译模型—TransR/CTransR

- TransE和TransH都将实体和关系嵌入到同一个向量空间中。
- TransR认为不同的关系应该拥有不同的语义空间。其为每个关系引入一个映射矩阵，将头尾实体映射到关系的子空间中。
- CTransR将属于同一种关系的头尾实体对分成多个类，针对每个类学习不同的关系矩阵。

翻译模型—TransR/CTransR



原来在实体空间中与头尾实体（圆圈）相似的实体（三角形），在关系 r 的子空间中被区分开了。

Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion. AAAI2015.

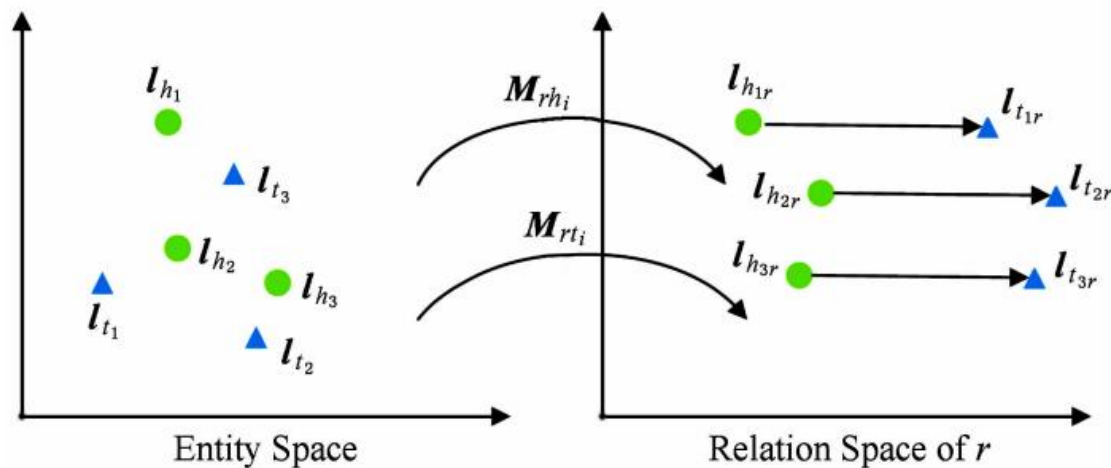
翻译模型—TransD

- 在TransH和TransR/ CTransR中，不同的种类的实体共享相同的映射向量或矩阵，但一个关系的头尾实体的种类和属性往往有较大差别。
- TransR引入了空间映射，导致模型参数量急剧增加。
- TransD将每个对象（实体、关系）嵌入为两个向量：语义向量、映射向量。

翻译模型—TransD

$$M_r^h = r_p h_p^T + I^{k*d}, \quad M_r^t = r_p t_p^T + I^{k*d}$$

$$h_{\perp} = M_r^h h, \quad t_{\perp} = M_r^t t$$



Ji G, He S, Xu L, et al. Knowledge graph embedding via dynamic mapping matrix. ACL2015: 687-696.

翻译模型—TranSparse

为了克服关系的异质性（有的关系连接的头尾实体对较多，有的较少）和不平衡性（同一关系的头实体和尾实体的数量不对称），TranSparse（share/separate）用自适应的稀疏矩阵代替一般的映射矩阵，稀疏度由关系连接的头尾实体对的数量决定。

$$h_{\perp} = M_r(\theta_r)h, \quad t_{\perp} = M_r(\theta_r)t \quad \theta_r = 1 - (1 - \theta_{min})N_r/N_{r^*}$$

$$h_{\perp} = M_r^h(\theta_r^h)h, \quad t_{\perp} = M_r^t(\theta_r^t)t \quad \theta_r^l = 1 - (1 - \theta_{min})N_r^l/N_{r^*}^l \quad (l = h, t)$$

Ji G, Liu K, He S, et al. Knowledge graph completion with adaptive sparse transfer matrix. AAAI. 2016.

翻译模型—TransM

TransM放宽了前面几种模型所使用的基本条件，即： $h + r \approx t$ ，方法是在前面得分函数的基础上加了权重：

$$w_r = \frac{1}{\log(h_r p t_r + t_r p h_r)}$$

$$f_r(h, t) = w_r ||\mathbf{h} + \mathbf{r} - \mathbf{t}||_{L_1/L_2}$$

Fan M, Zhou Q, Chang E, et al. Transition-based knowledge graph embedding with relational mapping properties. Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing. 2014.

翻译模型—ManifoldE

ManifoldE将约束 $h + r \approx t$ 放宽为一种基于流形的约束。
得分函数是：

$$f(h, r, t) = \|M(h, r, t) - D_r\|_{l1/l2}^2$$

$$M(h, r, t) = \|h + r - t\|_{l2}$$

$M()$ 是一个流形，以 $(h, r, *)$ 为例，所有合适的尾实体都分布在高维流形上。例如 $M()$ 是一个高维球体，那么所有的 $*$ 都在以 $h + r$ 为球心，以 D_r 为半径的球面上

Xiao H, Huang M, Zhu X. From one point to a manifold: Knowledge graph embedding for precise link prediction. arXiv preprint arXiv:1512.04792, 2015.

翻译模型—TransF

TransF将约束 $h + r \approx t$ 放宽为：只要求 $h + r$ （或 $t - r$ ）的方向与 t （或 h ）一致。得分函数同时衡量了 $h + r$ 和 t ， $t - r$ 和 h 的方向：

$$f(h, r, t) = (h + r)^T t + (t - r)^T h$$

Feng J, Huang M, Wang M, et al. Knowledge graph embedding by flexible translation. Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning. 2016.

翻译模型—TransA

- TransE及其扩展模型均存在两个问题：①得分函数只采用L1或L2距离，灵活度不够；②损失函数过于简单，实体和关系向量的每一维等同考虑。
- TransA将损失函数中的距离度量改为马氏距离，并为每一维学习不同的权重， W_r 是非负对称矩阵。

$$f(h, r, t) = (h + r - t)^T W_r (h + r - t)$$

Xiao H, Huang M, Hao Y, et al. TransA: An adaptive approach for knowledge graph embedding. arXiv preprint arXiv:1509.05490, 2015.

翻译模型—KG2E

KG2E使用高斯分布来表示实体和关系，将其表示为从多变量高斯分布中抽取的随机向量。高斯分布的均值表示实体或关系在语义空间中的中心位置，协方差则表示该实体或关系对其他元素的影响与的不确定度。

- 头尾实体间的关系的概率分布：

$$P_e \sim N(h - t, \Sigma_h + \Sigma_t)$$

- 实际关系的概率分布：

$$P_r \sim N(r, \Sigma_r)$$

- 损失函数：KL散度/期望似然（计算上述两个概率的相似度）

He S, Liu K, Ji G, et al. Learning to represent knowledge graphs with gaussian embedding. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, 2015: 623-632.

翻译模型—TransG

TransG也使用高斯分布来表示实体和关系。但其认为一个关系连接不同的头尾实体时拥有不同的语义。所以关系的分布应该是高斯分布的混合。

$$h \sim N(\mu_h, \sigma_h^2 I), t \sim N(\mu_t, \sigma_t^2 I)$$

$$r = \sum_i \pi_r^i \mu_r^i, \mu_r^i \sim N(\mu_h - \mu_t, (\sigma_h^2 + \sigma_t^2) I)$$

得分函数：

$$f(h, r, t) = \sum_i \pi_r^i \exp\left(\frac{-\|\mu_h + \mu_r^i - \mu_t\|_{l_2}}{\sigma_h^2 + \sigma_t^2}\right)$$

Xiao H, Huang M, Zhu X. TransG: A generative model for knowledge graph embedding. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016, 1: 2316-2325.

提纲

- 一、原理及概念
- 二、表示学习方法
 - 翻译模型
 - 语义匹配模型
 - 融合多源信息的模型
- 三、模型评测
- 四、挑战

语义匹配模型—LFM

LFM利用基于关系的双线性变换，刻画实体和关系之间的二阶联系。协同性较好，计算复杂度低。得分函数：

$$f(h, r, t) = h^T M_r t$$

Jenatton R, Roux N L, Bordes A, et al. A latent factor model for highly multi-relational data. NIPS. 2012: 3167-3175.

语义匹配模型—RESCAL

RESCAL是矩阵分解模型的典型代表。在该模型中，知识图谱中的所有实体对和所有关系构成一个张量 X ，如果三元组 (h, r, t) 存在，则 $X_{hrt} = 1$ ，否则为0。

张量分解旨在将每个三元组 (h, r, t) 分解为实体和关系表示 $X_{hrt} = h^T M_r t$ ，使得 X_{hrt} 尽量接近于1。

语义匹配模型—RESCAL

RESCAL的基本思想和LFM类似。不同之处在于，RESCAL会优化张量中的所有位置，包括值为0的位置；而LFM只优化知识图谱中存在的三元组。

Nickel M, Tresp V, Kriegel H P. A Three-Way Model for Collective Learning on Multi-Relational Data. ICML. 2011, 11: 809-816.

语义匹配模型—DistMult

DistMult将LFM中关系的表示矩阵限制为对角阵，这种简化极大降低了模型复杂度，模型效果反而得到显著提升。

$$f(h, r, t) = h^T \text{diag}(M_r) t$$

Yang B, Yih W, He X, et al. Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint arXiv:1412.6575, 2014.

语义匹配模型—ComplEx

为了更好的对非对称关系建模，ComplEx将DistMult中实体和关系的表示扩展到复数向量空间中，即向量的每一维是复数，而不再是实数。得分函数如下， $Re(x)$ 是取 x 的实部：

$$f(h, r, t) = \text{Re}(h^T \text{diag}(M_r) t)$$

Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction. International Conference on Machine Learning. 2016: 2071-2080.

语义匹配模型—ANALOGY

ANALOGY着重对知识图谱中的类比关系进行建模，例如：太阳系中的太阳之于行星正如原子系统中的原子核之于核外电子。

- 得分函数同LFM:

$$f(h, r, t) = h^T M_r t$$

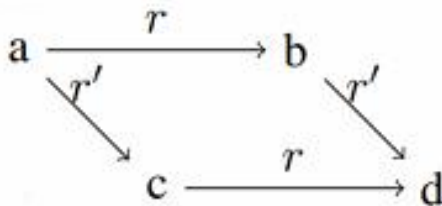
Liu H, Wu Y, Yang Y. Analogical inference for multi-relational embeddings. Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 2168-2178.

语义匹配模型—ANALOGY

- 约束条件

①正规性： $M_r M_r^T = M_r^T M_r$ ，如：对称矩阵可以对对称关系建模（好友关系），反对称矩阵对非对称关系建模（父子关系）。

②交换性： $M_r M_{r'} = M_{r'} M_r$ ，可以对如下情况建模：



语义匹配模型—HoIE

HoIE定义了一种循环相关操作，更好地捕获实体间的语义关联。

- 循环相关操作定义：

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

- 得分函数：

$$f(h, r, t) = r^T (h \otimes t) = \sum_{k=0}^{d-1} \left([r]_k \sum_{i=0}^{d-1} [h]_i * [t]_{(i+k) \bmod d} \right)$$

Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. AAAI. 2016.

语义匹配模型—SLM

SLM利用标准非线性单层神经网络来连接实体。
得分函数：

$$f(h, r, t) = u_r^T \tanh(M_r^h h + M_r^t t)$$

Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion. NIPS. 2013: 926-934.

语义匹配模型—NTN

NTN用张量神经网络捕获头尾实体间的语义关联，得分函数：

$$f(h, r, t) = u_r^T \tanh \left(h^T M_r t + V_r \begin{pmatrix} h \\ t \end{pmatrix} + b_r \right)$$

u_r 是与关系相关的线性层， M_r 是所有关系共享的三阶张量， V_r 是与关系相关的投影矩阵。

Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion. NIPS. 2013: 926-934.

语义匹配模型—MLP

MLP使用标准的多层神经网络捕获头尾实体间的语义关联，得分函数：

$$f(h, r, t) = w_r^T \tanh(M_r \cdot (h, r, t))$$

M_r 是与关系相关的第一层神经网络的权重， w_r 是与关系相关的第二层神经网络的权重，二者为所有关系共享。

Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 601-610.

语义匹配模型—NAM

NAM使用 L 层网络捕获头尾实体间的语义关联：

输入： $a^{(l)} = W^{(l)} z^{(l-1)} + b^{(l)}, l = 1, \dots, L$

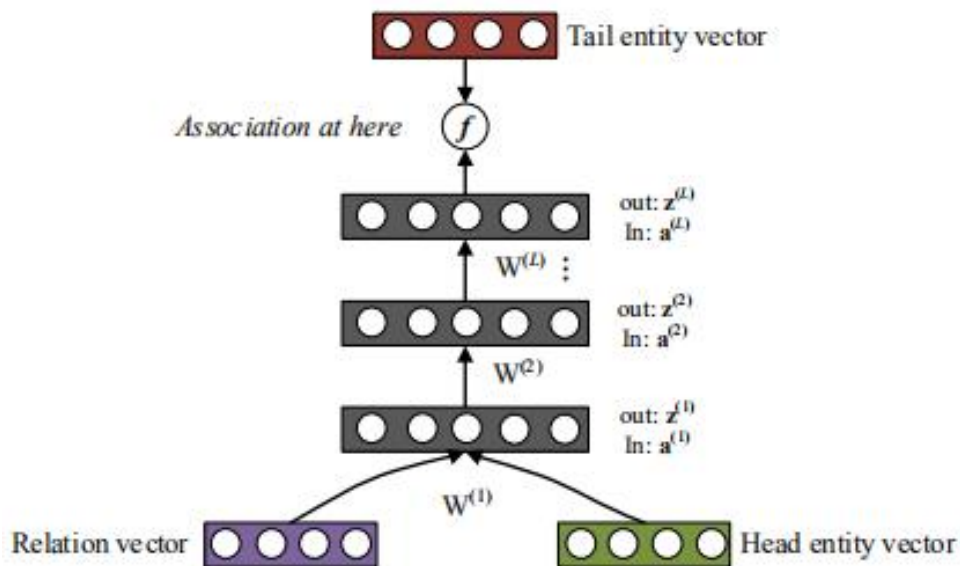
输出： $z^{(l)} = \text{ReLU}(a^{(l)}), l = 1, \dots, L$

$(h; r)$ 作为整个神经网络的输入，即 a^1

得分函数： $f(h, r, t) = t^T z^{(L)}$

Liu Q, Jiang H, Evdokimov A, et al. Probabilistic reasoning via deep learning: Neural association models. arXiv preprint arXiv:1603.07704, 2016.

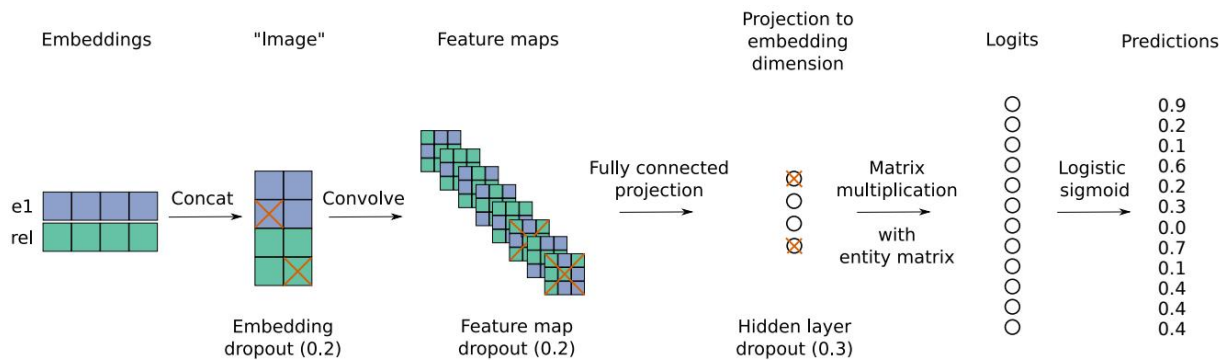
语义匹配模型—NAM



DNN for NAM

语义匹配模型—ConvE

ConvE利用卷积神经网络捕获实体间的语义关联：

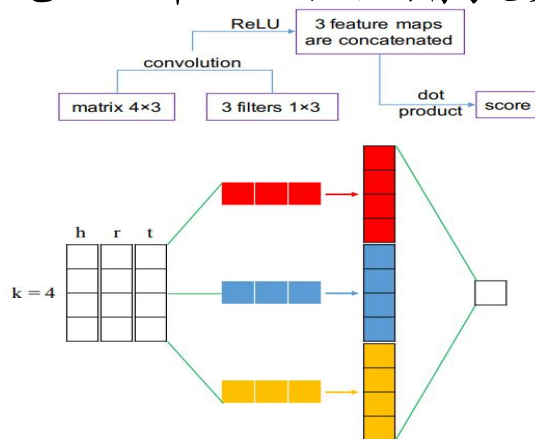


$$f(h, r, t) = g(\text{vec}(g([\bar{e}_s; \bar{e}_r] * \omega))W)e_o$$

Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2d knowledge graph embeddings. AAAI. 2018.

语义匹配模型—ConvKB

ConvKB使用卷积神经网络捕获实体间的全局关系：

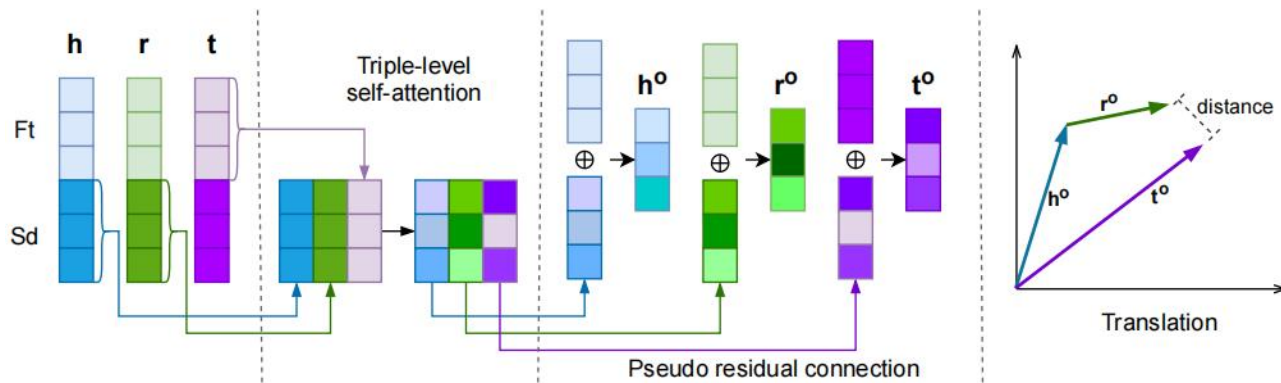


$$f(h, r, t) = \text{concat}(g([v_h, v_r, v_t] * \omega) \cdot w$$

Nguyen D Q, Nguyen T D, Nguyen D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network. NAACL. 2018.

语义匹配模型—AprilE

AprilE利用三元组级别的注意力机制和伪残差连接来进行表示学习：



Liu Y Z, Wang P, Li Y T, et al. AprilE: attention with pseudo residual connection for knowledge graph embedding. COLING. 2020.

提纲

- 一、原理及概念
- 二、表示学习方法
 - 翻译模型
 - 语义匹配模型
 - 融合多源信息的模型
- 三、模型评测
- 四、挑战

实体类型—SSE, TKRL

- SSE假设属于同一类型的实体在嵌入空间中应该距离较近，并使用流形学习算法对这种假设进行建模。
 - TKRL利用层级类型信息建立实体的映射矩阵。
-
- Guo S, Wang Q, Wang B, et al. Semantically smooth knowledge graph embedding. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015, 1: 84-94.
 - Xie R, Liu Z, Sun M. Representation Learning of Knowledge Graphs with Hierarchical Types. IJCAI. 2016: 2965-2971.

关系路径

PTransE在TransE的基础上加入了实体对间的路径信息，得分函数为：

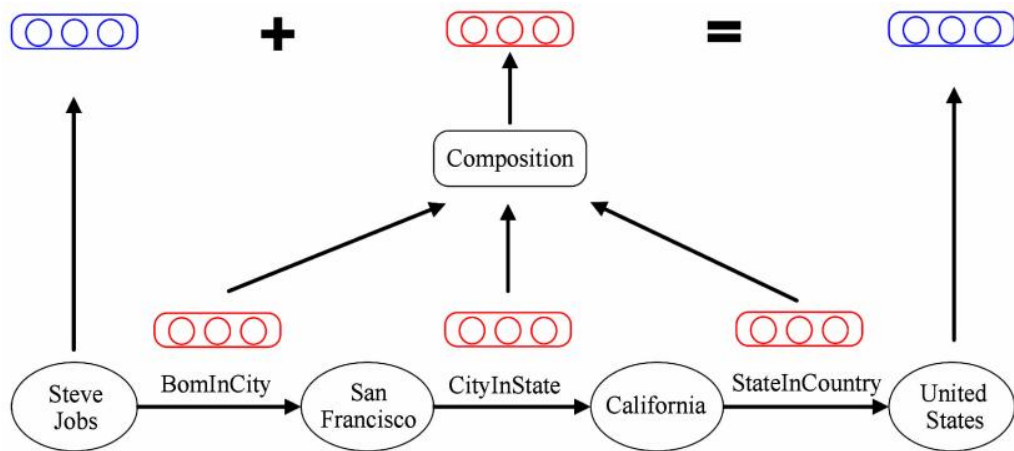
$$f(h, r, t) = E(h, r, t) + E(h, p, t)$$

等式右边的两项分别为基于单一关系和基于关系路径的三元组的得分。

Lin Y, Liu Z, Luan H, et al. Modeling relation paths for representation learning of knowledge bases. arXiv preprint arXiv:1506.00379, 2015.

关系路径—PTransE

为了对路径上所有关系的向量进行语义组合产生关系路径向量，即 p ，PTransE尝试了3种代表性的语义组合操作，分别是向量加法、向量哈达玛积、循环神经网络。实验表明，**向量加法**效果最好。



文本描述—NTN

NTN:

先用辅助新闻语料库学习词汇的向量表示，然后用实体名字描述中的词汇的表示向量的平均值来初始化实体的表示。例如，用smart和phone的向量的平均值初始化实体smartphone。

Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion. NIPS. 2013: 926-934.

文本描述—DKRL

DKRL: 每个实体有两个表示, 分别是基于三元组结构的表示 e_s 和基于文本描述的表示 e_d 。

得分函数:

$$\begin{aligned} f(h, r, t) = & \| h_s + r - t_s \| + \| h_d + r - t_d \| \\ & + \| h_s + r - t_d \| + \| h_d + r - t_s \| \end{aligned}$$

基于文本描述的向量: 连续词袋模型和卷积神经网络。

Xie R, Liu Z, Jia J, et al. Representation learning of knowledge graphs with entity descriptions. AAAI. 2016.

逻辑规则—ILP

ILP模型将推理看作整数线性规划问题，目标优化函数包括两部分：具体的优化函数、一系列基于规则的约束。

$$\begin{aligned}
 & \max_{\{x_{ij}^{(k)}, \epsilon_{ij}^{(k)}\}} \sum_k \sum_i \sum_j w_{ij}^{(k)} x_{ij}^{(k)} - \sum_{t^+ \in \mathcal{O}} \epsilon_{ij}^{(k)}, \\
 & \text{s.t.} \quad \text{R1. } x_{ij}^{(k)} + \epsilon_{ij}^{(k)} = 1, \forall t^+ \in \mathcal{O}, \\
 & \quad \text{R2. } x_{ij}^{(k)} = 0, \forall k, \forall i \notin \mathcal{H}_k, \forall j \notin \mathcal{T}_k, \\
 & \quad \text{R3. } \sum_i x_{ij}^{(k)} \leq 1, \forall k \in \mathcal{R}_{1-M}, \forall j, \\
 & \quad \text{R3. } \sum_j x_{ij}^{(k)} \leq 1, \forall k \in \mathcal{R}_{M-1}, \forall i, \\
 & \quad \text{R3. } \sum_i x_{ij}^{(k)} \leq 1, \sum_j x_{ij}^{(k)} \leq 1, \forall k \in \mathcal{R}_{1-1}, \forall i, \forall j, \\
 & \quad \text{R4. } x_{ij}^{(k_1)} \leq x_{ij}^{(k_2)}, \forall r_{k_1} \mapsto r_{k_2}, \forall i, \forall j, \\
 & \text{where } x_{ij}^{(k)} \in \{0, 1\}, \forall k, i, j; \epsilon_{ij}^{(k)} \in \{0, 1\}, \forall t^+ \in \mathcal{O}.
 \end{aligned}$$

Wang Q, Wang B, Guo L. Knowledge base completion using embeddings and rules. IJCAI. 2015.

逻辑规则—KALE

KALE将三元组事实表示为原子公式，将规则表示为基于三元组事实的复合公式，然后用一个损失函数同时优化两者。例如：

$$I(h, r, t) = 1 - \frac{1}{3\sqrt{d}} \|h + r - t\|$$

$$I(f_1 \Rightarrow f_2) = I(f_1) \cdot I(f_2) - I(f_1) + 1$$

$$I(f_1 \wedge f_2 \Rightarrow f_3) = I(f_1) \cdot I(f_2) \cdot I(f_3) - I(f_1) \cdot I(f_2) + 1$$

其中， $I(h, r, t) \in [0, 1]$ 代表三元组为真的概率。

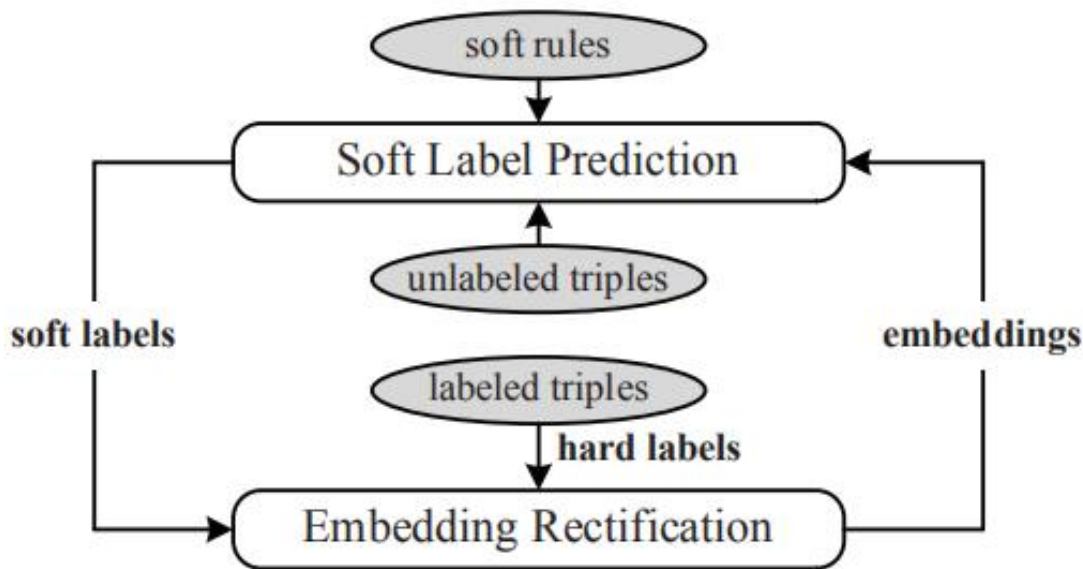
Guo S, Wang Q, Wang L, et al. Jointly embedding knowledge graphs and logical rules. EMNLP. 2016: 192-202.

逻辑规则—RUGE

RUGE重复迭代下述步骤完成对向量表示的学习：

- ①根据从知识图谱中自动抽取出的规则，为未标记的三元组生成置信度；
- ②使用已标记的三元组和带有置信度的未标记三元组共同进行表示学习。

逻辑规则—RUGE



RUGE模型框架

Guo S, Wang Q, Wang L, et al. Knowledge graph embedding with iterative guidance from soft rules. AAAI. 2018.

实体属性—Nickel et al.

大部分KGE模型将实体属性也看做关系处理，但实际中，部分属性值是不可能出现在关系中的。例如，(Obama, Gender, Male)。这会导致模型的一部分是浪费的，例如：属性值会显著增加RESCAL中的张量维度，从而导致张量中的大量值是无效的。因此，Nickel等人为实体属性单独建模处理，改进了RESCAL。

Nickel M, Tresp V, Kriegel H P. Factorizing yago: scalable machine learning for linked data. WWW. 2012: 271-280.

时序信息—Jiang et al.

例如：(霍金, 出生于, 英国牛津), (霍金, 逝世于, 英国剑桥), “出生于”和“逝世于”有严格的时间先后顺序。

Jiang等人在三元组的基础上，添加时间维度，构成四元组(h, r, t, T)，构建时序转换矩阵M，捕捉关系 r_1 和 r_2 的先后关联。

Jiang T, Liu T, Ge T, et al. Encoding temporal information for time-aware link prediction. EMNLP. 2016: 2350-2354.

图结构—GAKE

GAKE将知识图谱看做有向图，定义了实体的三种上下文，帮助捕获三元组的语义。

- 邻居上下文：实体的所有出边及出边所连接的尾实体。
- 边上下文：实体的所有出边。
- 路径上下文：包含实体的关系路径。

Feng J, Huang M, Yang Y. GAKE: graph aware knowledge embedding. COLING. 2016: 641-651.

邻居信息—TransN

TransN利用实体在知识图谱中的邻居节点，将每个实体或关系都嵌入为两个向量：

- 语义向量：用于表示实体或关系本身的语义
- 上下文向量：用于表示其他实体或关系的上下文。

由语义向量和上下文向量的线性组合得到的向量作为实体的向量，参与到基于翻译的三元组学习中。上下文向量的合成用到了注意力机制。

Wang C C, Cheng P J. Translating Representations of Knowledge Graphs with Neighbors. SIGIR. 2018: 917-920.

实体与概念—TransC

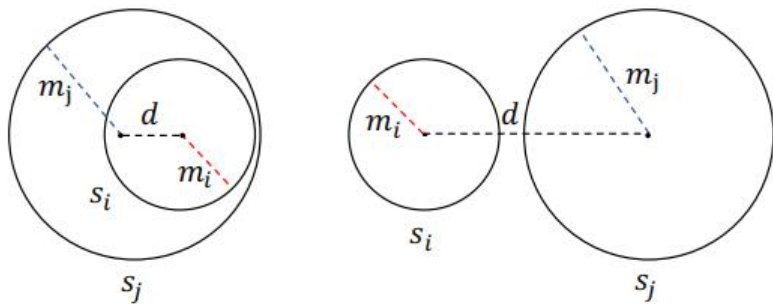
TransC将知识图谱中的实例和概念区别对待：
实例嵌入为向量，概念嵌入为球体。用点和球
、球和球之间的相对位置关系对instanceOf和
subClassOf两种关系建模，普通关系采用
TransE模型。

$$f_e(i, c) = \|i - p\|_2 - m$$

$$d = \|p_i - p_j\|_2$$

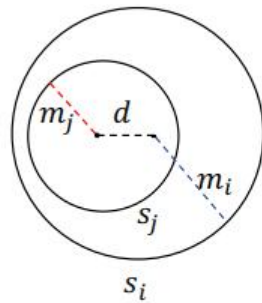
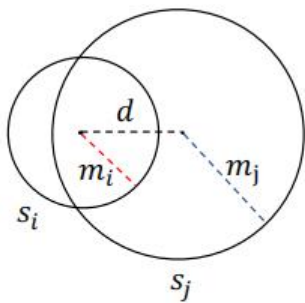
Lv X, Hou L, Li J, et al. Differentiating Concepts and Instances for Knowledge Graph Embedding. EMNLP. 2018: 1971-1979.

实体与概念—TransC



(a) $d < |m_i - m_j| \wedge m_i < m_j$ (GOAL)

(b) $d \geq |m_i + m_j|$

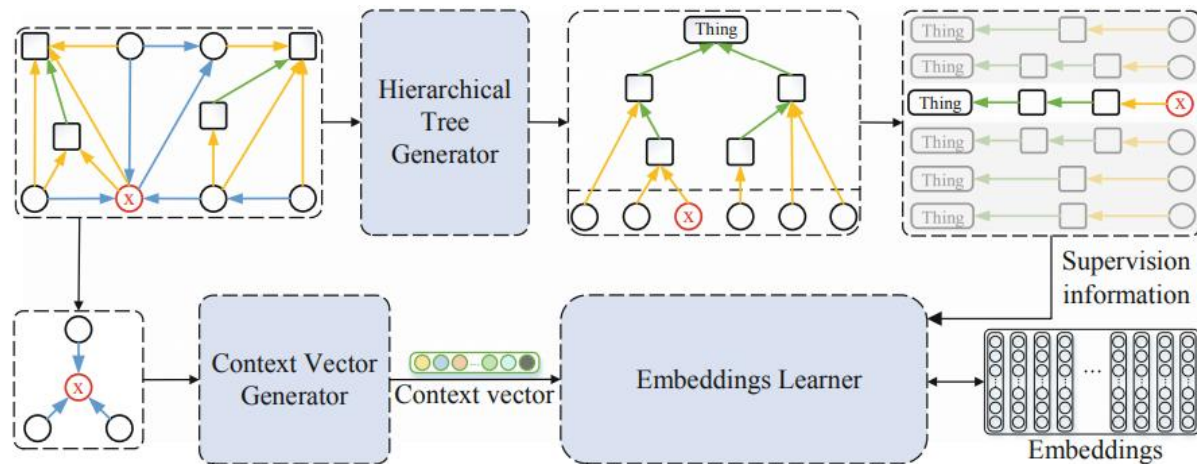


(c) $|m_i - m_j| \leq d < |m_i + m_j|$

(d) $d < |m_i - m_j| \wedge m_i \geq m_j$

实体与概念—JECI

JECI将知识图谱中的实例和概念联合进行嵌入表示。

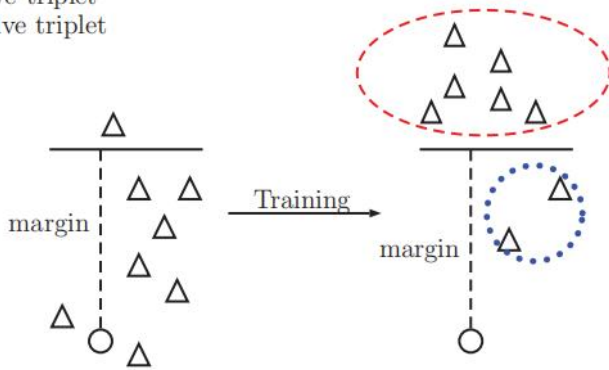


Zhou J, Wang P, Pan Z, et al. JECI: a joint knowledge graph embedding model for concepts and instances. JIST. 2019.

GAN-based Framework

GAN-based Framework利用生成对抗网络中的生成器进行高质量的负采样，用判别器进行表示学习。因而该框架可以应用于不同的传统模型。

○ Positive triplet
△ Negative triplet



生成器实际采用的
是一种强化学习模
型：Policy Gradient

。

Wang P, Li S, Pan R. Incorporating GAN for negative sampling in knowledge representation learning. AAAI. 2018.

开放域知识图谱补全

- 封闭域知识图谱补全

给定一个需要补全的知识图谱 $G = (E, R, T)$, 找到一组缺失的三元组 $T' = (h, r, t) | h \in E, r \in R, t \in E, (h, r, t) \notin T$ 来完成对 G 的补全。

- 开放域知识图谱补全

给定一个需要补全的知识图谱 $G = (E, R, T)$, 找到一组缺失的三元组 $T' = (h, r, t) | (h, r, t) \notin T, h \in E^i, r \in R, t \in E^i$ 来完成对 G 的补全, 其中 E^i 是 E 的超集。

开放域知识图谱补全

ConMask 是一种以文本为中心的方法，利用注意力机制对实体和关系的名称和描述进行建模，得到基于文本的头实体、关系和尾实体嵌入表示，并通过对嵌入表示的成对匹配来对三元组进行评分。

OWE 将图嵌入和文本嵌入分开进行训练，然后学习两个嵌入空间的映射。

Shi B, Wenginger t. Open-World knowledge graph completion. AAAI. 2018.

Shah H, Villmow J, Ulges A, et al. An open-world extension to knowledge graph completion models. AAAI. 2019

BERT meet KG

ERNIE 从知识库中提出的高信息量的实体信息，通过特殊的语义融合模块，来增强文本中对应的表示。

KG-BERT 将知识图谱中的三元组看作是文本序列，然后使用BERT进行训练，充分利用大规模预料的丰富信息增强实体和关系的表示。

KEPLER 使用预训练语言模型对实体的描述信息进行编码，联合优化知识嵌入表示和语言模型。

Zhang Z, Han Z, Liu Z. ERNIE: enhanced language representation with informative entities. ACL. 2019.

Yao L, Mao C. KG-BERT: BERT for knowledge graph completion. arXiv. 2019.

Wang X, Gao T, Zhu Z, et al. KEPLER: a unified model for knowledge embedding and pre-trained language representation. TACL. 2020

- 一、原理及概念
- 二、表示学习方法
 - 翻译模型
 - 语义匹配模型
 - 融合多源信息的模型
- 三、模型评测
- 四、挑战

常用数据集—WordNet

WordNet是最著名的词典知识库，拥有极高准确率的本体知识，主要用于词义消歧。其主要定义了名词、动词、形容词和副词之间的语义关系。例如名词之间的上下位关系（如：“猫科动物”是“猫”的上位词），动词之间的蕴含关系（如：“打鼾”蕴含着“睡眠”）等。

常用数据集—Freebase

Freebase将WordNet与Wikipedia二者的知识结合，即利用WordNet的本体知识补充Wikipedia中实体的上位词知识，从而获取大规模高质量、高覆盖的知识库。其数据基于RDF三元组模型，底层采用图数据库进行存储。Freebase的分类系统包含Topic（条目）、Type（类型）、Property（属性）和Schema（模式）。

常用数据集—YAGO

YAGO是一个综合型知识库，整合了Wikipedia、WordNet 以及GeoNames（全球地理数据库）等数据源，特别是将维基百科中的分类体系与WordNet的分类体系进行了融合，构建了一个复杂的类别层次结构体系。

常用数据集

datasets	#relation	#entity	#triple		
			train	valid	test
WN11	11	38696	112581	2609	10544
WN18	18	40943	141442	5000	5000
FB13	13	75043	316232	5908	23733
FB15K	1345	14951	483142	50000	59071
FB1M	23382	10^6	17.5×10^6	50000	177404
FB5M	1192	5385322	19193556	5000	59071

详细介绍: <https://blog.csdn.net/jingOlivia/article/details/85142789>

评测任务—链接预测

知识图谱中的链接预测指预测三元组中丢失的头实体或尾实体。预测结果是所有候选实例的得分排名，而不是直接给出最匹配的实例。

举例：(南京，是...的城市，中国)

评测标准通常使用以下两种：

- **MRR**：所有原始三元组排名的倒数的平均值；
- **Hits@N**：原始三元组的排名中不大于 N 的比例。

评测任务—三元组分类

知识图谱中的三元组分类本质上是一个二分类问题：判断给定的三元组是否是知识图谱中真实存在的。

评测标准通常使用评测标准使用准确率（Accuracy）、精确率（Precision）、召回率（Recall）和 F1 值。