Machine Translated by Google

翻译嵌入以进行建模多关系数据

Antoine Bordes、Nicolas Usunier、Alberto Garcia-Duran´ 康比涅理工大学 – CNRS
Heudiasyc UMR 7253 Compiegne, France {bordesan, nusunier,
agarciad}@utc.fr

Jason Weston、Oksana Yakhnenko Google 111 8th
avenue 纽
约,纽约州,美国
{jweston, oksana}
@google.com

抽象的

我们考虑在低维向量空间中嵌入多关系数据的实体和关系的问题。我们的目标是提出一个易于训练、包含较少参数且可以扩展到非常大的数据库的规范模型。因此,我们提出了 TransE,这是一种通过将关系解释为对实体的低维嵌入进行操作的翻译来建模关系的方法。尽管它很简单,但这个假设被证明是有效的,因为大量实验表明,TransE 在两个知识库上的链接预测中明显优于最先进的方法。此外,它可以在具有 1M 个实体、25k 个关系和超过 17M 个训练样本的大规模数据集上成功训练。

1 简介

多关系数据是指有向图,其节点对应于形式为 (头、标签、尾) 的实体和边（表示为 (h, , t)）,每个边都表示实体头和尾之间存在名称标签关系。多关系数据模型在许多领域都发挥着关键作用。例如,社交网络分析（其中实体是成员,边（关系)是友谊 / 社交关系链接）;推荐系统（其中实体是用户和产品,关系是购买、评级、评论或搜索产品）;或知识库 (KB),例如 Freebase1 、Google Knowledge Graph2或 GeneOntology3 ,其中 KB 的每个实体代表世界的抽象概念或具体实体,关系是表示涉及其中两个事实的谓词。我们的工作重点是从 KB（本文中的 Wordnet [9] 和 Freebase [1])建模多关系数据,目标是提供一种有效的工具来完成它们,通过自动添加新事实,而无需额外的知识。

多关系数据建模一般来说,建模过程归结为提取实体之间的局部或全局连接模式,并利用这些模式概括特定实体与所有其他实体之间观察到的关系,从而进行预测。单一关系的局部性概念可能纯粹是结构性的,例如我朋友的朋友是我朋友的

1freebase.com

2google.com/insidesearch/features/search/knowledge.html 3geneontology.org

Machine Translated by Google

社交网络,但也可能依赖于实体,例如喜欢《星球大战 IV》的人也喜欢《星球大战 V》,但他们可能喜欢也可能不喜欢《泰坦尼克号》。与单关系数据相比,在对数据进行一些描述性分析后可以做出临时但简单的建模假设,关系数据的困难在于局部性概念可能同时涉及不同类型的关系和实体,因此对多关系数据进行建模需要更通用的方法,这些方法可以同时考虑所有异构关系来选择适当的模式。

继协同过滤中的用户/项目聚类或矩阵分解技术成功表示单关系数据中实体连接模式之间的非平凡相似性之后,大多数现有的多关系数据方法都是在从潜在属性进行关系学习的框架内设计的,如[6]所指出的;也就是说,通过学习和操作成分(实体和关系)的潜在表示(或嵌入)。

从这些方法在多关系领域的自然扩展开始,例如随机块模型的非参数贝叶斯扩展 [7, 10, 17] 和基于张量分解 [5] 或集体矩阵分解 [13, 11, 12] 的模型,许多最新方法都专注于提高贝叶斯聚类框架 [15] 或基于能量的框架中的模型的表达能力和通用性,以学习低维空间中的实体嵌入 [3, 15, 2, 14]。这些模型的表达能力更强是以模型复杂性的大幅增加为代价的,这导致建模假设难以解释,计算成本也更高。此外,由于这种高容量模型的适当正则化很难设计,因此这些方法可能出现过拟合,或者由于需要解决具有许多局部最小值的非凸优化问题而导致欠拟合,因此需要解决这些局部最小值才能训练它们。事实上,[2] 表明,在具有相对较多不同关系的多个多关系数据集上,更简单的模型(线性模型而非双线性模型)几乎可以达到与最具表现力的模型一样好的性能。这表明,即使在复杂且异构的多关系领域,简单而适当的建模假设也可以在准确性和可扩展性之间取得更好的平衡。

关系作为嵌入空间中的平移在本文中,我们引入了 TransE,这是一种基于能量的模型,用于学习实体的低维嵌入。在 TransE 中,关系表示为嵌入空间中的平移:如果 (h, , t) 成立,则尾部实体 t 的嵌入应该接近于头实体 h 的嵌入加上某个依赖于关系 的向量。我们的方法依赖于一组精简的参数,因为它只为每个实体和每个关系学习一个低维向量。

我们基于翻译的参数化背后的主要动机是,层次关系在知识库中极为常见,而翻译是表示它们的自然转换。

事实上,考虑到树的自然表示(即 2 维节点的嵌入),兄弟节点彼此接近,给定高度的节点排列在 x 轴上,父子关系对应于 y 轴上的平移。由于零平移向量对应于实体之间的等价关系,因此模型也可以表示兄弟关系。因此,我们选择使用每个关系的参数预算(一个低维向量)来表示我们认为是知识库中的关键关系。

另一个次要动机来自 [8] 的最新研究,其中作者从自由文本中学习词向量,并且不同类型的实体之间的一些 1 对 1 关系(例如国家和城市之间的“首都”)被模型(巧合地而非自愿地)表示为嵌入空间中的翻译。这表明可能存在嵌入空间,其中不同类型的实体之间的 1 对 1 关系也可以通过翻译来表示。我们的模型旨在强制执行这种嵌入空间结构。

第 4 节中的实验表明,尽管这个新模型很简单,其架构主要用于建模层次结构,但它最终在大多数类型的关系上都很强大,并且在现实世界知识库的链接预测中可以明显优于最先进的方法。此外,它的轻量级参数化使其能够在包含 1M 个实体、25k 个关系和超过 17M 个训练样本的 Freebase 大规模分割上成功训练。

在本文的其余部分,我们将在第 2 部分描述我们的模型,并在第 3 部分讨论其与相关方法的联系。我们将在第 4 部分详细介绍对 Wordnet 和 Freebase 的广泛实验研究,并将 TransE 与文献中的许多方法进行比较。最后,我们将在第 5 部分概述未来的一些工作方向。

Machine Translated by Google

算法1学习TransE输入训练集S = {(h, , t)},实体和相关集E和L,边际γ,

嵌入dim.k。6 6 1:对每个实体e ∈ E初始化← uniform(− √ , √ kk for each ∈ L)

) 对每个 ∈ L

2:

3:

4:循环5:

对于每个实体 e ∈ E,e ← e / e

6: Sbatch ←sample(S, b) // 抽取一个大小为 b 的小批量

7: Tbatch ← ∅ // 初始化三元组对集合8: for (h, , t) ∈ Sbatch do (h , t)

←sample(S) // 抽样一个损坏的三元组

9: (h,,t) ,

10: Tbatch ← Tbatch ∪ (h, , t),(h , , t)

11:结束12:更新嵌

入 wrt ∇γ + d(h + , t) − d(h + , 吨) +

(h，,t) ，(h，,t) ∈Tbatch

13:结束循环

2 基于翻译的模型

给定一个由两个实体 h,t ∈ E （实体集)和一个关系 ∈ L （关系集)组成的三元组 (h, · · · , t) 的训练集 S,我们的模型学习实体和关系的向量嵌入。嵌入在 R 中取值（k 是模型超参数），并用相同的字母（粗体)表示。我们模型背后的基本思想是,由 -标记边引起的函数关系对应于嵌入的平移,即,当 (h, · · , t) 成立时（t 应该是 h + 的最近邻居），我们希望 h + ≈ t,否则 h + 应该远离 t。按照基于能量的框架,三元组的能量等于 d(h + · · , t),其中某些相异度 d 我们将其取为L1或L2 范数。

为了学习这样的嵌入,我们最小化了训练集上的基于边际的排名标准：

$$L = \sum_{(h,\ ,t)\in S\ (h,\ ,t)\in S\ (h,\ ,t)} \gamma + d(h + \ \ , t) - d(h + \ \ , 吨)_{+} \quad (1)$$

其中 [x]+表示 x 的正部分,γ > 0 是边缘超参数,S = (h , , t)|h ∈ E ∪ (h, , t)|t ∈ E . (h,,t)

(2)

根据公式 2 构建的损坏三元组集由训练三元组组成,其中头部或尾部被随机实体替换（但不能同时替换）。损失函数 (1) 倾向于训练三元组的能量值低于损坏三元组的能量值,因此是预期标准的自然实现。请注意,对于给定实体,当该实体作为三元组的头部或尾部出现时,其嵌入向量是相同的。

优化是通过随机梯度下降（小批量模式)在可能的 h 和 t 上进行的,附加约束是实体嵌入的 L2 范数为 1（没有对标签嵌入给出正则化或范数约束）。此约束对于我们的模型很重要,就像对于以前基于嵌入的方法 [3、6、2] 一样,因为它可以防止训练过程通过人为增加实体嵌入范数来轻易最小化 L。

算法 1 中描述了详细的优化过程。首先按照 [4] 中提出的随机程序初始化实体和关系的所有嵌入。在算法的每次主迭代中,首先对实体的嵌入向量进行归一化。然后,从训练集中抽取一小组三元组,并将其作为小批量的训练三元组。

对于每个这样的三元组,我们都会抽取一个损坏的三元组。然后通过采用具有恒定学习率的梯度步骤来更新参数。根据算法在验证集上的表现停止算法。

3 相关工作第 1 节描述了大

量关于嵌入知识库的工作。我们在此详细介绍了我们的模型与 [3]（结构化嵌入或 SE)和 [14] 的模型之间的联系。

Machine Translated by Google

表 1:参数数量及其值
对于FB15k（以百万计）。ne和nr是实体和关系的数量;k 是嵌入维度。

方法	注意： FB15K上的参数	
非结构化 [2]	O（负） 0.75	
重新校准 [11]	O(nek + nrk ^{2})	87.80
东南 [3]	O(nek + 2nrk ^{2})	7.47
SME（线性） [2]	O(nek + nrk + 4k ^{2}	0.82
SME（双线性） [2]	O(nek + nrk + 2k ^{3}))	1.06
线性调频 [6]	O(nek + nrk + 10k ^{2})	0.84
转运	O(nek + nrk)	0.81

表 2:所用数据集的统计数据
本文从两篇论文中提取
知识库、Wordnet 和 Freebase。

数据集	西诺 FB15K 40,943 18	FB1M
实体关系训	14,951	1×106
练。例如	1,345 23,382	
	141,442 483,142 17.5×106	
有效 EX。	5,000 50,000 50,000	
测试示例。	5,000 59,071 177,404	

SE [3] 将实体嵌入到 R 中 ^{钾}，并将关系分解为两个矩阵L1 ∈ R $^{千×千}$ 且L2 ∈ R $^{千×千}$
这样， d(L1h, L2t) 对于损坏的三元组 (h, , t) 很大（否则很小）。基本思想
当两个实体属于同一个三元组时,它们的嵌入应该彼此接近
在某个取决于关系的子空间中。使用两个不同的投影矩阵
头和尾是为了解释关系可能出现的不对称。当
相异函数的形式为 d(x, y) = g(x − y),其中 g 为:R ^{钾} → R（例如 g 是
范数），那么嵌入大小为 k + 1 的 SE 比我们的模型更具表现力
大小为 k 的嵌入,因为维度 k + 1 的线性算子可以重现仿射变换
在 k 维子空间中（通过将所有嵌入的第 k+1 维约束为等于
1)SE,以L2为单位矩阵， L1取其平移,则等价于
到 TransE。尽管我们的模型表达能力较低,但我们仍然取得了比
在我们的实验中,SE 是最重要的。我们认为这是因为 (1) 我们的模型是一种更直接的方式来表示
关系的真实属性,以及（2)嵌入模型的优化很困难。对于
SE,更强的表现力似乎更多地意味着欠拟合,而不是更好的性能。
训练错误（见第 4.3 节)倾向于证实这一点。

另一种相关方法是神经张量模型 [14]。该模型的一个特例是
学习分数 s(h, , t)（损坏的三元组分数较低），形式如下：

$$s(h, , t) = h \; ^{电视}\text{长按} \; ^{电视}_{1}\text{氢键} \; ^{电视}2 \qquad (3)$$

其中 L ∈ R $^{千×千}$, L1 ∈ R ^{钾} 且L2 ∈ R ^{钾}, 所有这些都取决于 。

如果我们将 TransE 视为以平方欧氏距离作为相异函数,则有：

$$d(h + , t) = h \; ^{2}_{2} + \; ^{2}_{2} + \text{吨} \; ^{2}_{2} - 2\text{小时}^{电视}\text{吨} + \; ^{电视} (t - h) 。$$

考虑到我们的规范约束（h2 = 1)和排序标准（1），其中 $^{2}_{2}$ 在比较损坏的三元组时不起任何作用,因此我们的模型涉及对三元组进行评分
与 h ^{电视}t + ^{电视}(t − h),因此对应于 [14] 模型（公式 (3)），其中 L 是
单位矩阵,且 = 我们无法用这个模型进行实验（因为它已经
同时发表于我们的论文中），但 TransE 的参数又少得多:这可能
简化训练并防止欠拟合,并可能弥补较低的表达力。

尽管如此,TransE 的简单表述可以看作是对一系列双向
交互（例如通过开发L2版本)存在缺点。对于建模数据，
如果 h 和 t 之间的三向依赖关系至关重要,我们的模型可能会失败。例如,在小规模 Kinships 数据集 [7] 上,TransE 在交叉验证中没有达到预期效果（测量
准确率-召回率曲线下面积)与最先进的 [11, 6] 相媲美,因为
在这种情况下,这种三元相互作用至关重要（参见[2]中的讨论）。尽管如此,我们的实验
第 4 节说明,为了处理像 Freebase 这样的通用大规模知识库,首先应该
正确地模拟最常见的连接模式,就像TransE一样。

4 实验

我们的方法 TransE 是基于从 Wordnet 和 Freebase 中提取的数据进行评估的（它们的统计数据是
如表 2 所示），与文献中几种最近的方法相比,这些方法被证明能够实现
在各种基准测试中取得当前最佳表现,并能扩展到相对较大的数据集。

Machine Translated by Google

4.1 数据集

Wordnet此知识库旨在生成直观易用的词典和同义词库,并支持自动文本分析。其实体（称为同义词集)对应于词义,关系定义它们之间的词汇关系。我们考虑了 [2] 中使用的数据版本,我们在下文中将其表示为 WN。三元组的示例为（得分 NN 1、上位词、评估 NN 1)或（得分 NN 2、有部分、音乐符号 NN 1）。4

Freebase Freebase 是一个庞大且不断增长的一般事实知识库;目前有大约 12 亿个三元组和超过 8000 万个实体。我们使用 Freebase 创建了两个数据集。首先,为了制作一个小型数据集进行实验,我们选择了 Wikilinks 数据库 5 中也存在且在 Freebase 中至少有 100 次提及的实体子集（包括实体和关系）。我们还删除了诸如“！/people/person/nationality”之类的关系,它只是将头和尾与关系“/people/person/nationality”进行了反转。这产生了 592,213 个三元组,其中 14,951 个实体和 1,345 个关系被随机拆分,如表 2 所示。

在本节的其余部分中,此数据集表示为 FB15k。我们还希望拥有大规模数据,以便大规模测试 TransE。因此,我们从 Freebase 创建了另一个数据集,选择了出现频率最高的 100 万个实体。这导致了一个包含约 25000 个关系和超过 1700 万个训练三元组的分割,我们将其称为 FB1M。

4.2 实验装置

评估协议为了进行评估,我们使用与 [3] 中相同的排名程序。对于每个测试三元组,删除头部并依次替换为字典中的每个实体。模型首先计算这些损坏三元组的差异（或能量）,然后按升序排序;最后存储正确实体的排名。重复整个过程,同时删除尾部而不是头部。我们报告这些预测排名的平均值和 hits@10,即排名在前 10 名中的正确实体的比例。

这些指标具有指示性,但当某些损坏的三元组最终成为有效三元组（例如来自训练集)时,这些指标可能会存在缺陷。在这种情况下,这些三元组可能排在测试三元组之上,但这不应算作错误,因为两个三元组都是真实的。为了避免这种误导行为,我们建议从损坏三元组列表中删除训练、验证或测试集中出现的所有三元组（感兴趣的测试三元组除外）。这可确保所有损坏的三元组都不属于数据集。在下文中,我们根据这两种设置报告平均排名和命中率@10:原始（可能有缺陷)称为原始,而较新的称为过滤（或过滤）。我们只提供 FB1M 实验的原始结果。

基线第一种方法是 Unstructured,这是 TransE 的一个版本,它将数据视为单关系并将所有翻译设置为 0（它已在 [2] 中用作基线）。我们还将其与 RESCAL（[11,12] 中提出的集体矩阵分解模型)以及基于能量的模型 SE [3]、SME(linear)/SME(bilinear) [2] 和 LFM [6] 进行了比较。RESCAL 通过交替最小二乘法进行训练,而其他方法则与 TransE 一样通过随机梯度下降进行训练。表 1 将基线与我们模型的理论参数数量进行了比较,并给出了 FB15k 上的数量级。虽然 SME(linear)、SME(bilinear)、LFM 和 TransE 在低维嵌入方面具有与 Unstructured 大致相同的参数数量,但其他算法 SE 和 RESCAL 为每个关系学习至少一个 $k \times k$ 矩阵,因此需要快速学习许多参数。 RESCAL 在 FB15k 上需要大约 87 倍的参数,因为它需要比其他模型大得多的嵌入空间才能实现良好的性能。

出于参数数量或训练时长方面的可扩展性原因,我们没有使用 RESCAL、SME（双线性)和 LFM 对 FB1M 进行实验。

我们使用作者提供的代码训练了所有基线方法。对于 RESCAL,出于可扩展性原因,我们必须将正则化参数设置为 0,如 [11] 中所示,并在 {50, 250, 500, 1000, 2000} 中选择潜在维度 k,这会导致验证集上的平均预测排名最低（使用原始设置）。对于非结构化、SE、SME（线性)和 SME（双线性）,我们

4WN 由意义组成,其实体由单词、词性标记和表示其指的是哪种意义的数字的连接来表示,即分数 NN 1 编码名词“分数”的第一个含义。

5code.google.com/p/wiki-links

Machine Translated by Google

表 3:链接预测结果。测试不同方法的性能。

数据集指标平	西氮		FB15K		FB1M	
均排名命中率@10 (%)平均排名命中率@10 (%)平均排名命中率@10 (%)						
评估设置 原始过滤器 原始 原始非结构化 [2] 315 304 35.3 15,139 RESCAL [11] 1,180 683 37.2 SE [3] 1,011 985 68.5 SME(LINEAR) [2] 545 533 65.1			过滤原煤过滤原煤28.2 1,074 979 4.5	过		生的
SME(BILINEAR) [2] 526 509 54.7 LFM [6] 469 456 71.4 TransE 263 251 75.4	52.8	828 683 28.4	80.5	273 162 28.8	74.1	2.9
	274 154 30.7	61.3	284 158 31.3	81.6	283	滤。 - -
	164 26.0	89.2	243 125 34.9		6.3	22,044 17.5
					44.1	- -
					39.8	- -
					40.8	- -
				41.3	33.1	47.1 14,615 34.0

在{0.001, 0.01, 0.1}中选择学习率,在{20, 50}中选择k,并选择最佳模型
通过使用验证集上的平均等级提前停止（总共最多 1,000 个时期
在训练数据上）。对于 LFM,我们还使用平均验证等级来选择模型,并
在 {25, 50, 75} 中选择潜在维度,在 {50, 100, 200, 500} 中选择因子数量
学习率在{0.01, 0.1, 0.5}之间。

实施对于 TransE 的实验,我们选择了随机
{0.001, 0.01, 0.1} 之间的梯度下降、{1, 2, 10} 之间的边距 γ 以及潜在维度
在每个数据集的验证集上,k 在 {20, 50} 之间。相异度度量 d 设置为
根据验证性能将L1或L2距离设置为最佳配置：
Wordnet 上,k = 20、λ = 0.01、γ = 2 和d = L1；
FB15k;FB1M 上的 k = 50、λ = 0.01、γ = 1 和 d = L2 。对于所有数据集,训练时间都是有限的
在训练集上最多训练 1,000 个 epoch。使用早期停止法选择最佳模型
验证集上的平均预测排名（原始设置）。开源实现
TransE 可从项目网页6获取。

4.3 链接预测

总体结果表 3 显示了所有比较方法在所有数据集上的结果。正如预期的那样,过滤设置提供了较低的平均排名和较高的命中率@10,我们认为这是
更清晰地评估链接预测方法的性能。然而,一般来说
原始和过滤之间的趋势相同。

我们的方法 TransE 在所有指标上都优于所有同类方法,并且通常具有很大的优势,并且
达到了一些令人满意的绝对性能分数,例如 WN 上 89% 的命中率@10（超过
超过 4 万个实体)和 FB1M（超过 100 万个实体)的比例分别为 34% 和 4%。TransE 与
最佳亚军的方法很重要。

我们认为,TransE 的良好表现得益于模型的合理设计
根据数据,也因为它相对简单。这意味着它可以有效地优化
具有随机梯度。我们在第 3 节中表明,SE 比我们的提议更具表现力。
然而,它的复杂性可能会使其很难学习,从而导致性能下降。在 FB15k 上,
在训练集的 50k 个三元组中,SE 的平均排名为 165,命中率为 35.5%。
而TransE达到127和42.7%,表明TransE确实不太容易出现欠拟合
这可以解释其更好的性能。SME（双线性)和 LFM 存在同样的问题
培训问题:我们从来没有对它们进行足够好的培训,以便它们能够充分发挥其
能力。LFM 的糟糕结果也可能由我们的评估设置来解释,基于
排名实体,而 LFM 最初提出来是为了预测关系。RESCAL 可以
在 FB15k 上实现了相当好的命中率@10,但平均排名较差,尤其是在 WN 上,即使
我们使用了较大的潜在维度（Wordnet 上为 2,000）。

翻译项的影响是巨大的。当比较 TransE 和非结构化（即没有翻译的 TransE)的性能时,非结构化的平均排名似乎相当好
（WN 上的最佳亚军）,但 hits@10 非常差。非结构化只是将所有同时出现的实体聚类,与所涉及的关系无关,因此只能进行猜测
哪些实体是相关的。在 FB1M 上,TransE 和非结构化的平均排名几乎
类似,但 TransE 在前 10 名中的预测数量是其 10 倍。

6可从http://goo.gl/0PpKQe获取。

Machine Translated by Google

表 4:按关系类别划分的详细结果。我们比较了 FB15k 上的 Hits@10（单位:%）
我们的模型、TransE 和基线的过滤评估设置。（M. 代表MANY）。

任务相	预测头部				预测尾巴			
关类别非结构化 [2]	1-对-1 1-对-M。	M.-对-1	M.-对-M。	1-对-1	1-对-M。	M.-对-1	M.-对-M。	
	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
东南 [3]	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME（线性） [2]	35.1	53.7	19.0	40.3	32.7	14.9	61.6	43.3
SME（双线性） [2]	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
转运	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0

表 5:使用 TransE 对 FB15k 测试集进行的示例预测。粗体表示测试三元组的
真尾和斜体训练集中存在的其他真尾。

输入（头部和标签）	预测的尾部
JK罗琳受到的影响	GK 切斯特顿、JRR 托尔金、CS 刘易斯、劳埃德亚历山大、 特里·普拉切特、罗尔德·达尔、豪尔赫·路易斯·博尔赫斯、斯蒂芬·金、伊恩·弗莱明
安东尼·拉帕格利亚 (Anthony LaPaglia) 曾演出	马缨丹、萨姆的夏天、快乐的大脚、欢乐之家、 《不忠》、《守护者传奇》、《裸体午餐》、《X战警》、《同名人》
卡姆登县毗邻	伯灵顿县、大西洋县、格洛斯特县、联合县、 新泽西州埃塞克斯县、帕塞克县、海洋县、巴克斯县
《四十岁的处男》获提名	MTV 电影奖最佳喜剧表演奖， BFCA 评论家选择奖最佳喜剧奖， MTV 电影奖最佳荧幕搭档， MTV电影奖最佳突破表演奖， MTV 电影奖最佳电影、 MTV 电影奖最佳吻奖、 DF Zanuck 年度戏剧电影制片人奖， 美国演员工会奖最佳影片男主角
哥斯达黎加足球队有位置	前锋、后卫、中场、守门员、 投手、内野手、外野手、中锋、防守队员
Lil Wayne 出生于	新奥尔良、亚特兰大、奥斯汀、圣路易斯、 多伦多、纽约市、惠灵顿、达拉斯、波多黎各
WALL-E 属于这一类型	动画、电脑动画、喜剧电影、 冒险片、科幻片、奇幻片、定格动画、讽刺片、剧情片

详细结果表 4 根据关系的几个类别和几种方法的预测参数对 FB15k 上的结果（以 hits@10 为单位)进行分类。
我们将其分类为
根据头参数和尾参数的基数,将关系分为四类：
1 对 1、1对多、多对 1、多对多。如果头
最多可以出现一个反面,如果一个正面可以出现多个反面,则为1 -TO-MANY ,如果
多个头可以出现在同一个尾部上,或者如果多个头可以出现在同一个尾部上,则为多对多
多尾关系。我们通过计算每个关系在 FB15k 数据集中出现的平均正面数 h（相对于反面数 t)将关系分为这四类，
给定一对
（,t)（对于一对（h，））。如果这个平均数低于1.5,那么这个论点就被标记为
为 1,否则为MANY 。例如,一个关系平均有 1.2 个正面和反面,并且
每头 3.2 尾的事件被归类为一对多。我们得到 FB15k 的1-TO-1事件率为 26.2%
关系,一对多关系占22.7% ,多对一关系占28.3% ,多对多关系占22.8% 。

表 4 中的详细结果使我们能够精确评估和理解
方法。首先,正如人们所预料的那样,预测“侧面”实体似乎更容易。
1” 的三元组（即,在1 对多中预测头部,在多对一中预测尾部）,即当多个
实体指向它。这些是适定的情况。SME（双线性)被证明是非常准确的
在这种情况下,因为它们是具有最多训练示例的。非结构化执行
1-TO-1关系上表现良好:这表明这种关系的论据必须具有共同点
非结构化模型能够通过将实体聚类到一起来发现隐藏的类型
在嵌入空间中。但这种策略不适用于任何其他类别的关系。添加
翻译术语（例如,将非结构化升级为 TransE)带来了在
嵌入空间,通过遵循关系从一个实体群集到另一个实体群集。这尤其
这些精心设计的案例非常引人注目。

图5给出了TransE在FB15k测试集上的链接预测结果示例
（预测尾部）。这说明了我们模型的能力。给定一个头部和一个标签,顶部

Machine Translated by Google

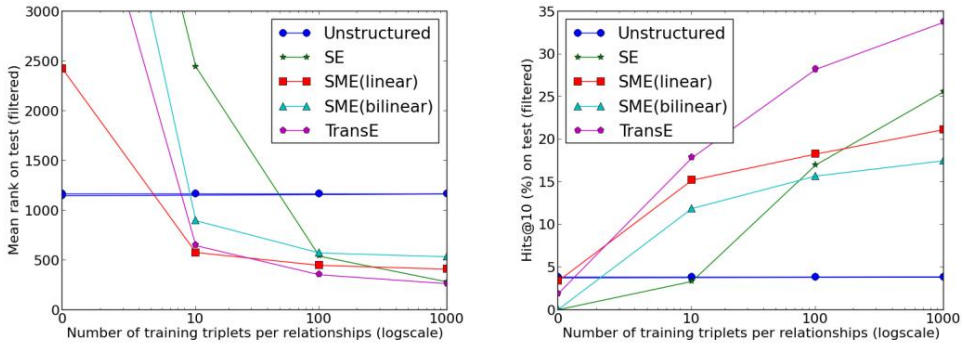

图 1:通过几个示例学习新关系。在 FB15k 数据上进行的比较实验,以平均排名（左)和 hits@10（右)进行评估。文中提供了更多详细信息。

图中描绘了预测的尾部（和真实的尾部）。示例来自 FB15k 测试集。即使好的答案并不总是排在首位,预测也反映了常识。

4.4 学习用少量样本预测新关系使用 FB15k,我们想通过检查方法学习新关系的

速度来测试它们对新事实的推广效果。为此,我们随机选择了 40 个关系,并将数据分成两组:一组（名为 FB15k-40rel）包含这 40 个关系的所有三元组,另一组（FB15k-rest)包含其余三元组。我们确保两个集合都包含所有实体。然后,FB15k-rest 被分成一个包含 353,788 个三元组的训练集和一个包含 53,266 个三元组的验证集,FB15k-40rel 被分成一个包含 40,000 个三元组的训练集（每个关系 1,000 个三元组)和一个包含 45,159 个三元组的测试集。利用这些数据集,我们进行了以下实验：(1)使用 FB15k-rest 训练和验证集训练和选择模型，(2)随后在训练集 FB15k-40rel 上对模型进行训练,但只学习与新的 40 个关系相关的参数，(3)在 FB15k-40rel 测试集（仅包含阶段 (1) 中未见过的关系）上对模型进行链接预测评估。我们在阶段 (2) 中使用每种关系的 0、10、100 和 1000 个示例重复了此过程。

图 1 显示了非结构化、SE、SME（线性）、SME（双线性)和 TransE 的结果。
当没有提供未知关系的示例时,非结构化模型的性能最佳,因为它不使用此信息进行预测。但是,当然,在提供标记示例时,这种性能不会提高。TransE 是学习速度最快的方法:仅使用 10 个新关系示例,hits@10 就已经达到 18%,并且它会随着提供的样本数量而单调提高。我们相信,TransE 模型的简单性使其能够很好地概括,而无需修改任何已经训练过的嵌入。

5 结论和未来工作

我们提出了一种学习知识库嵌入的新方法,重点关注模型的最小参数化,主要表示层次关系。我们表明,与两个不同知识库上的竞争方法相比,该方法效果非常好,并且也是一个高度可扩展的模型,我们将其应用于非常大规模的 Freebase 数据块。虽然我们仍不清楚我们的方法是否能够充分建模所有关系类型,但通过将评估细分为类别（一对一、一对多……），它似乎在所有设置中都比其他方法表现良好。

未来的工作可以进一步分析这个模型,并专注于在更多任务中利用它,特别是受 [8] 启发的学习单词表示等应用。将知识库与文本相结合（如 [2])是我们的方法可能有用的另一个重要方向。因此,我们最近成功地将 TransE 插入到从文本中提取关系的框架中 [16]。

致谢这项工作是在
Labex MS2T (ANR-11-IDEX-0004-02) 框架下开展的,并由法国国家研究机构 (EVEREST-12-JS02-005-01) 资助。我们感谢 X。
Glorot 提供代码基础设施,T. Strohmann 和 K. Murphy 提供有益的讨论。

Machine Translated by Google

参考

[1] K. Bollacker、C. Evans、P. Paritosh、T. Sturge 和 J. Taylor。Freebase:用于构建人类知识的协作创建的图形数据库。2008 年 ACM SIGMOD 数据管理国际会议论文集,2008 年。

[2] A. Bordes、X. Glorot、J. Weston 和 Y. Bengio。用于多关系数据学习的语义匹配能量函数。机器学习,2013 年。

[3] A. Bordes、J. Weston、R. Collobert 和 Y. Bengio。学习知识库的结构化嵌入。第 25 届人工智能年会 (AAAI) 论文集,2011 年。

[4] X. Glorot 和 Y. Bengio。《理解训练深度前馈神经网络的难度》。《国际人工智能与统计会议论文集》（AISTATS），2010 年。

[5] RA Harshman 和 ME Lundy。Parafac:平行因子分析。计算统计与数据分析,18(1):39–72,1994 年 8 月。

[6] R. Jenatton、N. Le Roux、A. Bordes、G. Obozinski 等人。高度多关系数据的潜在因子模型。《神经信息处理系统进展》（NIPS 25），2012 年。

[7] C. Kemp、JB Tenenbaum、TL Griffiths、T. Yamada 和 N. Ueda。具有无限关系模型的概念学习系统。第 21 届人工智能年会 (AAAI) 论文集,2006 年。

[8] T. Mikolov、I. Sutskever、K. Chen、G. Corrado 和 J. Dean。单词和短语的分布式表示及其组合性。《神经信息处理系统进展》（NIPS 26），2013 年。

[9] G. Miller. WordNet:英语词汇数据库。ACM通讯,38(11):39-41,1995。

[10] K. Miller, T. Griffiths 和 M. Jordan. 用于链接预测的非参数潜在特征模型。在神经信息处理系统进展（NIPS 22)中,2009 年。

[11] M. Nickel、V. Tresp 和 H.-P. Kriegel。多关系数据集体学习的三向模型。第 28 届国际机器学习会议 (ICML) 论文集,2011 年。

[12] M. Nickel、V. Tresp 和 H.-P. Kriegel。《YAGO 因式分解:可扩展的关联数据机器学习》。《第 21 届万维网 (WWW) 国际会议论文集》,2012 年。

[13] AP Singh 和 GJ Gordon。通过集体矩阵分解实现关系学习。第 14 届 ACM SIGKDD 知识发现和数据挖掘 (KDD) 会议论文集,2008 年。

[14] R. Socher、D. Chen、CD Manning 和 AY Ng。使用神经张量网络和语义词向量从知识库中学习新事实。《神经信息处理系统进展》（NIPS 26），2013 年。

[15] I. Sutskever、R. Salakhutdinov 和 J. Tenenbaum。使用贝叶斯聚类张量分解对关系数据进行建模。《神经信息处理系统进展》（NIPS 22），2009 年。

[16] J. Weston、A. Bordes、O. Yakhnenko 和 N. Usunier。将语言和知识库与嵌入模型连接起来以进行关系提取。《自然语言处理实证方法会议论文集》（EMNLP），2013 年。

[17] J. Zhu. 用于链接预测的最大边际非参数潜在特征模型。第 29 届国际机器学习会议 (ICML) 论文集,2012 年。