带文字的知识图嵌入综述：哪个模型在文字上链接得更好？

杰内・阿塞法・格塞斯、鲁萨・比斯瓦斯、迈赫威・阿拉姆、哈拉德・萨克

卡尔斯鲁厄 - 莱布尼茨信息基础设施研究所 & 应用信息学和形式研究所

描述系统 (AIFB) ，卡尔斯鲁厄理工学院，德国卡尔斯鲁厄

电子邮件: [genet-asefa.gesese@fiz-karlsru.de、russa.biswas@fiz-karlsruhe.de、mehwish.alam@fiz-karlsru.de、harald.sack@fiz-karlsruhe.de](mailto:genet-asefa.gesese@fiz-karlsru.de、russa.biswas@fiz-karlsruhe.de、mehwish.alam@fiz-karlsru.de、harald.sack@fiz-karlsruhe.de)

**摘要.**知识图（KGs）由有关特定领域的结构化信息以实体和关系的形式组成。除结构化信息外，KGs 还有助于促进链接数据云中表示的不同资源之间的互连性和互操作性。KGs 已用于各种应用程序中，例如实体链接，问题解答，推荐系统等。但是，KGs 应用程序承受着高昂的计算和存储成本。因此，需要一种能够将高维 KGs 映射到低维空间（即，嵌入空间），保留结构以及相关信息的表示形式。本文对 KG 嵌入模型进行了调查，该模型不仅考虑 KG 中实体和关系形式所包含的结构化信息，而且还考虑以文本，数值，图像等文字形式表示的非结构化信息。除了到目前为止提出的用于生成带有文字的 KG 嵌入的方法的理论分析和比较之外，还针对链接预测的一般任务对相同设置下的不同方法进行了经验评估。

**关键字：**知识图、知识图嵌入、具有文字的知识图嵌入、链接预测、调查

# 引言

知识图（KGs）对于存储结构化信息已变得至关重要。突然之间，人们开始关注将 KGs 用于各种应用程序，主要是在人工智能领域。例如，从更一般的意义上讲，KGs 可用于支持决策过程并改善不同的机器学习应用程序，例如问题回答[1]，推荐系统[2]和关系提取[3]。一些最流行的可公开获得的通用 KGs 是 DBpedia[4]，Freebase[5]，Wikidata[6] 和 YAGO[7]。这些通用 KGs 通常由使用数十亿个实体（表示为节点）和关系（作为连接这些节点的边）构造的大量事实组成。

尽管 KGs 可以有效地表示结构化数据,但是仍然存在一些妨碍其有效操作的问题，例如 i）不同的 KGs 通常基于不同的严格符号框架，这使得在其他应用程序中难以利用其数据[8]，并且 ii)事实证明，有效操纵和分析图所需的大量重要图算法都是 NP 完全的[9]。为了解决这些问题并更有效地使用 KG，将其转换为低维向量空间同时保留其基本语义是有益的。为此，迄今为止已经进行了各种尝试来学习 KGs 的向量表示（嵌入）。但是，这些方法中的大多数，包括当前最新的 TransE[10]，都是基于结构的嵌入，不使用任何文字信息，即通常只考虑由通过属性连接的实体组成的三元组。这是一个主要缺点，因为在捕获某个实体的语义时，将不使用原义编码的信息。

文字学习者可以通过两种主要方式将广告范式带入学习 KG 嵌入的过程:

1. 学习新实体的嵌入：新实体是未与 KG 中的任何其他实体链接但具有与其关联的文字值（例如其文字说明）的实体。在大多数现有的基于结构的嵌入模型中，无法学习此类新颖实体的嵌入。但是，这可以通过利用文字中表示的信息来学习嵌入来解决。例如，考虑数据集 FB15K-20[11]（它是 Freebase 的子集），实体 ‘/m/0gjd61t' 是一个新颖的实体，在任何训练三元组中都不会出现，但它有如下形式的描述<主题，关系，对象>。

</m/0gjd61t, http://rdf.freebase.com/ns/common.topic.description, "文森特・富兰克林是一位英语演员，以在喜剧电视节目中的作用而闻名">

为了了解该特定实体（即 /m/0gjd61t）的嵌入，该模型应足以利用该实体的描述。DKRL[11] 是使用其描述为新颖实体提供嵌入的那些方法之一。

1. 在基于结构的嵌入模型中改善实体的表示：在要求实体至少出现在最小数量的关系三元组中的情况下，文字在改善表示学习方面起着至关重要的作用。例如，仅考虑从 DBpedia 中提取的图1所示样本 KG 中提供的信息，就不可能将实体 dbr:David\_Prowse，dbr:Carrie\_Fisher 和 dbr:Peter\_Mayhew 彼此区分开。之所以如此，是因为以下事实：关于该 KG 中这些实体的唯一可用信息是它们都在电影 dbr:Return\_of\_the\_Jedi 中起作用，而这不足以知道哪些实体彼此相似，哪些不相似。因此，如果仅使用该 KG 训练某些 KG 嵌入模型，则无法获得实体dbr:David\_Prowse，dbr:Carrie\_Fisher 和 dbr:Peter\_Mayhew 的良好表示。

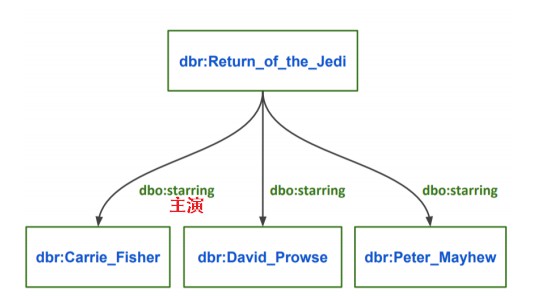


图1.取自 KG DBpedia 的一小部分三元组

但是，如图 2 所示，使用包含这些实体字面值的更多三元组训练模型，将会改善实体的嵌入。例如，查看数据关系 foaf:gender的值，dbr:David\_Prowse 和 dbr:Peter\_Mayhew 都是男性，而 dbr:Carrie\_Fisher 是女性。仅凭此信息，就可以使模型学习更好地表示这些实体，以使实体dbr:David\_Prowse 和 dbr:Peter\_mayhew 彼此之间的相似性比 dbr:Carrie\_fisher 更高。

上面的示例表明，使用文字及其各自的实体将增加更多的语义，以便相似的实体可以在向量空间中彼此接近地表示，而那些不相似的实体则进一步分开。

最近，人们提出了一些利用文本信息生成 KG 嵌入的方法。这些嵌入方法中考虑的文字类型是文本，数字，图像或多模式文字，即多种信息介质的组合。这些方法使用不同的技术，以便将文字合并到 KG 嵌入中。但是，在这些 KG 嵌入模型和对 KG 嵌入进行的调查中，未解决数据类型的文字。数据类型文字（例如日期和时间）的主要挑战在于，它们需要在 KG 嵌入中表示其他语义。

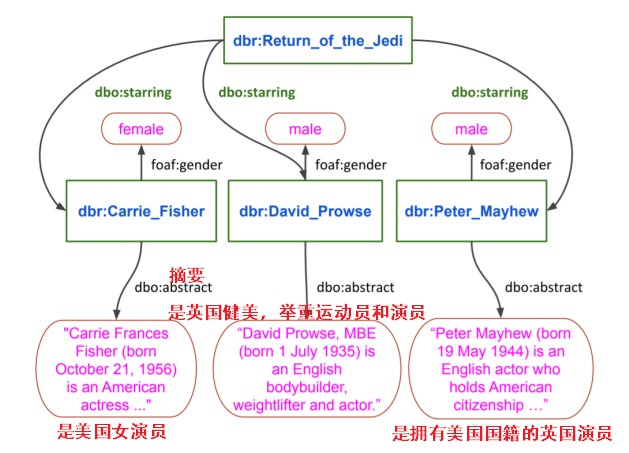


图2.三元组的一小部分，其文字取自 KG DBpedia

本文重点分析了不同的嵌入方法，突出了它们在处理不同挑战方面的优缺点。此外，还对不同 KG 嵌入模型在模型评估中的应用进行了评述，并针对链接预测任务进行了实验。本文的贡献总结如下：

1. 详细分析了现有文献丰富的 KG 嵌入模型及其方法。 此外，根据所使用的文本类型，模型被分为不同的类。
2. 在相同的实验设置下，对链接预测任务上的现有模型进行了面向评估的比较。
3. 指出了 KG 嵌入在文献使用方面存在的研究空白，为进一步的研究开辟了方向。

本文的其余部分安排如下：第二部分简要介绍了相关工作。第三部分给出了问题的表述，包括定义、预备知识、文本类型和研究性问题，第四部分分析了不同的 KG 文本嵌入技术。第五部分回顾了训练和评估嵌入模型的不同任务。第六部分讨论了使用现有的 KG 嵌入模型对链接预测任务进行文字编码的实验。最后，结束语总结了我们对带有文字的KGs 的发现，并在第七部分中介绍了未来的方向。

# 2. 相关工作

本节介绍了用于生成 KG 嵌入的最新算法。它还简要概述了按照这些原则已经发表的调查以及这些研究中缺少的内容。

到目前为止，已经提出了不同的 KG 嵌入技术，可以分为基于翻译的模型、语义匹配模型、包含实体类型的模型、包含关系路径的模型、包含逻辑规则的模型、包含时间信息的模型、包含图结构的模型和包含文本信息的模型。表1中相对于先前定义的类别，简要概述了最流行的方法，包括生成 KG 嵌入的最新方法。这项研究的主要重点是分析表 1 的最后一个类别，即包含以 KG 中的文字表示的信息的模型。

很少有人尝试对 KG 嵌入的技术和应用进行调查[52-54]。该调查[52]基于因子分解，基于随机游动和基于深度学习的网络嵌入方法（例如 DeepWalk，Node2vec 等）进行。[53,54]仅讨论 RESCAL[17] 和 KREAR[55] 作为使用实体属性进行 KG 嵌入的方法，重点讨论了基于结构的嵌入方法，即使用非属性三元组的方法，例如表1中列出的基于翻译的嵌入模型。但是，RESCAL 是一种用于关系学习的矩阵分解方法，该方法将每个对象/数据属性编码为张量的切片，从而导致张量的维数自动增加。如果在生成 KG 嵌入时使用文字，则该方法会遇到效率问题。同样，KREAR 仅考虑具有分类值（即固定数量的值）的那些数据属性，而忽略那些将任何随机文字作为值的数据属性。最近的一项调查[56]总结了迄今为止提出的提炼 KGs 的方法。然而，此调查并没有局限于嵌入技术，也没有考虑大多数利用文本的方法。

上述调查均未包括所有利用文字的现有 KG 嵌入模型，例如被归类为包含表 1 所示文字信息的模型。据我们所知，这是首次尝试分析迄今为止提出的使用文字生成 KG 嵌入的算法。在本文中，讨论了文字类型，嵌入方法以及评估嵌入模型的应用程序/任务。还提供了基于模型使用的文字类型的分类。

表1 KG 嵌入模型及其类别

|  |  |
| --- | --- |
| **类别** | **模型** |
| 平移距离模型 | TransE[10] 及其扩展名：TransH[12] TransR[13]，TransD[14]，TranSparse[15]，TransA[16] 等 |
| 语义匹配模型 | RESCAL[17]及其扩展名：DistMult[18]，HolE[19]，ComplEx[20]等。神经网络的语义匹配：SME[21]，NTN[22]，MLP[23]等 |
| 使用实体类型的模型 | 扩展 RESCAL[24]，SSE[25]，TKRL[26]，类型约束表示学习[27]，包含 KG 完成模型的规则[28]，TRESCAL[29]，实体层次结构嵌入[30] |
| 使用关系路径的模型 | PTransE[31]，在向量空间中遍历 KG[32]，RTRANSE[33]，成分向量空间[34]，使用 RNN 进行推理[35]，上下文相关的 KG 嵌入[36] |
| 使用逻辑规则的模型 | 包含 KG 完成模型的规则[28]，大规模知识库完成[37]，KALE[38]，用于关系提取的逻辑背景知识[39]等。 |
| 使用时间信息的模型 | 时间感知链路预测[40]，事件和 KG 的协同进化[41]，Knowevolve[42] |
| 使用图结构的模型 | GAKE[43]，多关系图中的链接预测[44] |
| 使用文字的模型 | LiteralE[45]，TransEA[46]，KBLRN[47]，MTKGNN[48]，MKBE[49]，KDCoE[50]，DKRL[11]，IKRL[51]等。 |

这个调查[57]是一个已经发表的简短调查的延伸。两个版本之间的主要区别在于: (1) 本调查包含了迄今为止提出的有文字的 KG 嵌入模型的更详细的理论分析，(2) 在相同的实验环境下，以链接预测为例对所讨论的模型进行了实证评价。

# 3．问题表述

本节简要介绍 KGs 和 KG 嵌入的基本原理，然后给出带有文字的 KG 嵌入的正式定义。 这也提出了各种研究问题，为什么进行这项研究是未来发展的踏脚石。

## *3.1. 准备工作*

**知识图.** 知识图（KG）由一组三元组 K⊆E×R×(E∪L) 组成，其中 E 是一组称为实体的资源，L 是一组文字，而 R 是一组关系。实体由表示实际对象或抽象概念的 URI 标识。关系（或属性）是一个二进制谓词，而文字是一个字符串，日期或数字，后跟它的数据类型。对于三元组 <*x*，*r*，*y*>，x 是对象，r 是关系，y 是对象。主体和客体通常分别称为头和尾实体。由文字作为对象的三元组通常称为定语三元组。

**关系（或属性）:**

根据对象的性质，关系可分为两大类：

* **对象关系**将一个实体链接到另一个实体。例如，在三元组 <dbr:Albert\_Einstein,dbo:field,dbr:Physics>, dbr:Albert\_Einstein和dbr:Physics 都是实体，dbo:field 是对象关系。
* **数据类型关系**将实体链接到其值，即文字。例如，在 <dbr:Albert\_Einstein,dbo:birthDate,"1879-03-14"> 中，其中 "1879-03-14" 是文字值，关系 dbo:birthDate 是数据类型关系。

## *3.2. 文字类型*

KG 中的文字对实体或关系未捕获的其他信息进行编码。 KG 中存在不同类型的文字：

* **文本文字：**KG 可以以自由文字的形式储存各种各样的信息，例如名称，标签，标题，描述，注释等。在大多数带有文字的 KG 嵌入模型中，文本信息进一步分类为 **Short 文本**和 **Long 文本**。 诸如名称，标题，标签等关系之类的相当短的文字被视为 Short 文本。 另一方面，对于更长的字符串，例如实体描述，注释等，被视为Long 文本，通常以自然语言提供。
* **数字文字：**以整数、浮点数等形式编码的信息，例如高度，日期，人口等，也提供有关实体的有用信息。在嵌入模型中，数字作为不同的实体是值得考虑的，因为它有自己的语义，不能被字符串距离度量所覆盖。 例如，777 更类似于 788 而不是 77。
* **度量单位：**数字字面值通常表示测量的单位到一个确定的大小。 例如，维基数据属性 wdt: P2048 采用 mm，cm，m，km，inch，foot 和 pixel 的值。 因此，丢弃单位而只考虑数值而不进行规范化会导致语义损失，特别是当单位不可比时，例如长度单位和重量单位。
* **图像文字：**图像还为实体建模提供了潜在的有用信息。 例如，可以通过对描绘此人的图像进行视觉分析来推断其详细信息，例如年龄，性别等。
* **其他类型的文字：**以其他文字形式编码的有用信息，例如外部 URI，可能会链接到一个图像，文本，音频或视频文件。

## *3.3. 研究问题*

从以上讨论可以看出，KG 中表示的信息是多种多样的，对这些实体进行建模是一项艰巨的任务。这项研究进一步针对的挑战如下：

* **研究问题1** -如何将 KGs 中的结构化 (带对象关系的三元组) 和非结构化信息 (属性三元组) 结合到表征学习中？
* **研究问题2** -KGs 中存在的文字类型的异质性如何被捕捉并结合到表征学习中？

# 4. 带文字的知识图嵌入

本节研究 KG 嵌入模型，根据文本类型将文本分为以下不同类别: (i) 文本，(ii) 数字，(iii) 图像，和 (iv) 多模式。一个使用至少两类文本提供互补信息的 KG 嵌入模型被认为是多模式的。在随后的章节中，对前述的每个类别的模型进行描述，分析了它们的相似性和差异，然后讨论了潜在的缺点。

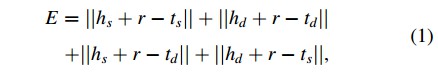
## *4.1. 具有文本文字的模型*

在本节中，我们将讨论四种利用文本文字的 KG 嵌入模型，即带有文字的扩展 RESCAL[24]，DKRL[11]，KDCoE[50] 和 KGloVe[58]。 首先，对这些模型进行了详细描述，然后进行了总结，介绍了模型的比较及其缺点。

**扩展 RESCAL** 旨在通过扩展其算法来更有效地处理文字值并解决张量带来的稀疏性的缺陷，从而改进原始 RESCAL 方法。在原始的 RESCAL 方法中，关系数据被建模为大小为 *n*×*n*×*m* 的三向张量 X，其中 *n* 是实体数，m 是关系数。*Xijk* = 1 表示存在三元组，其中第 i 个实体作为主语，第 k 个关系作为谓语，第 j 个实体为宾语。如果 *Xijk* 设置为 0，则表示该三元组不存在。提出了一种在 X 上进行张量分解的新方法。有关更多详细信息，请参见 [24]。如果属性三元组必须以这种方式建模，那么即使它们不能作为主语出现在三元组中，文本也将被视为实体。包含文字可能会导致运行时间增加，因为必须将较大的张量分解。

与原始算法相反，扩展的 RESCAL 算法在单独的矩阵中处理属性三元组。矩阵分解与非属性三元组的张量分解共同执行。仅包含文本文字的属性三元组以这样的方式在实体属性矩阵 D 中进行编码：行是实体，列是 <数据类型关系，值> 对。对于具有文本数据类型 (如rdfs:label 或 yago:hasPreferredMeaning) 的三元组，可以通过对对象文字中的文本进行标记和截断来创建一个或多个这样的对。然后将矩阵 *D* 分解为 *D* ≈ *AV*，其中 A 和 V 分别表示实体和属性的潜在成分。尽管此方法具有处理多值文字的优点，但它并未考虑文字值的单词顺序。注意，Extended RESCAL 表示 RDF (s) 数据的方式不区分 A-Box 和 T-Box，也就是说，类和实例都被建模为张量中的实体。T-box 相当于软约束，而不是让它们对模型施加硬约束。

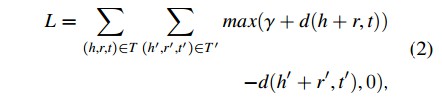
**DKRL**通过利用实体的描述来扩展 TransE[10]。 对于每个实体 *e*，学习两种矢量表示，即基于结构的 *es* 和基于描述的 *ed*。将这两种实体表示同时学习到相同的向量空间中，但不强制统一，只有描述的新实体才能得到表示。为了达到这个目的，给定一个三元组 (*h*，*r*，*t*) ，DKRL 模型的能量函数定义为:



其中 *hs* 和 *ts* 是基于结构的表示，*hd* 和 *td* 是对应实体的基于描述的表示。

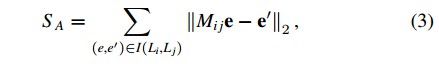
为了学习基于结构的表示方法，本文直接采用 TransE 方法三元组中的关系，如从头实体到尾实体的转换。另一方面，连续词袋（CBOW）和深层卷积神经网络（CNN）模型已用于生成基于描述的头和尾实体。在使用 CBOW 的情况下，根据关键词描述生成短文本，并对相应的词嵌入进行归纳，生成实体嵌入。在 CNN 模型中，在对描述进行预处理之后，会提供来自 Wikipedia 的经过预训练的单词向量作为输入。该 CNN 模型有五层，并且在应用每个卷积层池之后，可以减少 CNN 的参数空间和滤波器噪声。最大池适用于第一个池，平均池适用于最后一个池。使用的激活函数可以是 tanh 或 ReLU。CNN 模型比 CBOW 更好，因为它保留了单词的顺序。

为了训练 DKRL，以下基于边距的得分函数被视为目标函数，并使用标准的反向传播（使用随机梯度下降（SGD））将其最小化



其中 *γ*> 0 是余量超参数，*d* 是相异函数，*T’* 是损坏的三元组。实体的表示可以基于结构或基于描述。

**KDCoE** 致力于通过基于利用实体描述的嵌入方法来创建新的语言间链接（ILL），从而在多语言 KGs 的实体之间创建对齐方式。该模型使用弱对齐的多语言 KG 进行半监督的跨语言学习。它迭代执行多语言 KG 嵌入模型（KGEM）和多语言实体描述嵌入模型（DEM）的联合训练，以便每个模型交替提出新的 ILL。 KGEM 由知识模型和对齐模型两个部分组成，用于根据来自 KGs 的结构化信息（非归因三元组）学习嵌入。给定一组语言 𝓛，每种语言 *L* ∈ 𝓛 使用单独的 k1 维嵌入空间 Rk1L 表示对应的关系 *RL* 和实体 *EL*。为了学习 *RL* 和 *EL* 的嵌入，知识模型采用 TransE，因此将铰链损耗作为其目标函数。另一方面，对准模型采用基于线性变换的技术，该技术在进行跨语言推理时具有最佳性能。该技术采用以下目标函数：



其中 *I*(*Li*, *Lj*) 是语言 *Li* 和 *Lj* 之间的 ILLs， *Mij* 是 *k1×k1* 矩阵，用作从 *Li* 到 *Lj* 的实体向量的线性变换。

令 SK 为知识模型使用的铰链损失函数，然后 KGEM 模型将 *SKG = SK +αSA* 最小化，其中 *α* 为正超参数。对于 DEM 模型，采用一种专注的门控循环单元编码器 (AGRU) 对多语种实体描述进行编码。DEM 应用多语言单词嵌入，以便从单词级别捕获多语言实体描述的语义信息。 迭代地训练两个模型，即 KGEM 和 DEM，以便每个模型交替提出新的 ILL。

**带文字的 KGloVe** 是将实体描述纳入 KGloVe KG 嵌入方法的实验性尝试。 该实验是在 DBpedia 上进行的，将实体的摘要和注释作为其描述。 主要目标是从文本描述中提取命名实体，并为文本中的每个实体提取表示实体的单词，然后将其替换为实体本身，然后将其相邻的单词和实体作为上下文。 该方法的工作原理是独立创建两个共现矩阵，然后在最后合并它们，以便可以执行联合嵌入。 使用与 KGloVe[59] 中相同的技术来生成第一矩阵，即，通过对（加权）图执行个性化 PageRank（PPR），然后进行与 GloVe[60] 方法中相同的优化。

为了创建第二个矩阵，使用 KG 的实体和谓词列表作为输入，对实体描述文本执行 “命名实体识别（NER）” 任务。 NER 步骤采用简单的精确字符串匹配技术，这会导致许多缺点，例如由于具有相同语义的不同关键字而导致实体缺失。与任何实体标签都不匹配的所有英语单词都将添加到实体谓词列表。然后使用实体谓词和单词列表作为输入，将文本的 GloVe 共现应用于修改后的文本（即 DBpedia 摘要和注释）。最后，将两个同时出现的矩阵求和，以创建一个统一的矩阵。所提出的方法已经在分类和回归任务上进行了评估，结果表明，除了 SVM 之外，对于大多数使用的分类器，该方法并未对 KGloVe 带来显著改善。但是，可以通过大量实验使用参数调整来改进该方法。

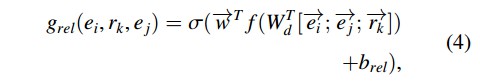
**总结** 这些模型之间的基本区别在于用于利用文本文字中给出的信息并将其与基于结构的表示形式相结合的方法。 与基于文本文字的嵌入模型相比，KDCoE 的一个主要优势是它考虑了多语言 KG 中存在的描述。 此外，DKRL 和 KDCoE 嵌入模型均设计用于在 KG 中仅具有属性三元组的新颖实体上表现良好。 所提出的带有文本文字的方法主要集中于描述，即长自然语言文本。其他类型的文本文字，例如名称，标签，标题等，并未得到广泛考虑。

## *4.2. 具有数字文字的模型*

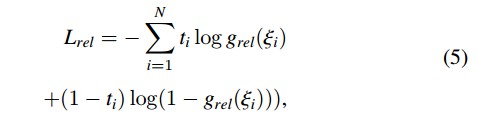
在本节中，将对使用数字文字（即 MT-KGNN[48]，KBLRN[47]，LiteralE[45] 和 TransEA[46]）提出的 KG 嵌入模型进行分析，然后进行总结。

**MT-KGNN** 是一种用于知识图上的关系学习和非离散属性预测的方法，目的是学习实体，对象属性和数据属性的嵌入。它由两个网络组成，即关系网络（RelNet）和属性网络（AttrNet）。 RelNet 是用于三重态预测的二进制（逐点）分类器，而 AttrNet 是用于属性值预测的回归任务。分别给定 n，m 和 l 作为实体，关系和文字嵌入维，该模型作为输入 [*ei* , *rk* , *ej* , *t* ] 传递给 RelNet，并作为 [*ai* , *vi* , *aj* , *vj*] 传递给 AttrNet，其中 *ei* , *ej ∈ Rn*, *rk ∈ Rm*, *t* 是 0 或 1 的分类目标，*ai* , *aj ∈ Rl*，并且 *vi*和 *vj* 是区间 [0，1] 中的归一化连续值。请注意，AttrNet 的输入，即 [*ai*, *vi*, *aj*, *vj*] 取自具有非离散文字值的属性三元组。在给定这些输入为一热编码索引的情况下，将使用嵌入查找层来检索相应的矢量表示。

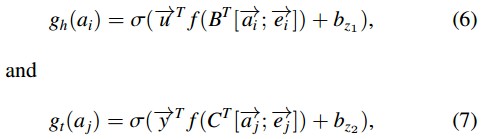
在 RelNet 中，串联的三元组通过非线性变换，然后应用 S 型函数来获得线性变换：



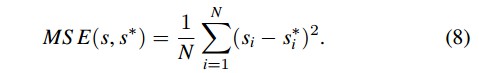
其中 *ω* ∈ *Rh×1* 和 *Wd* ∈ *R3n×h* 是网络的参数。 *σ*，*f* 和 *brel* 分别是 S 型函数，双曲正切函数 tanh 和标量偏差。 通过最小化以下交叉熵损失函数来训练 RelNet：



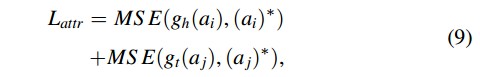
其中 *ξi* 表示批次 *N* 中的三元组 *i*，*ti* 的取值为 0 或 1。在 AttrNet 的情况下，将执行两项回归任务，一项用于头数据属性，另一项用于尾数据属性。 为这两项任务定义了以下评分函数：



其中，*u*，**y** ∈ *Rha×1* 和 *B*，*C* ∈ *R2n×ha* 是 AttrNet 的参数。 *ha* 是隐藏层的大小，*bz1*，*bz2* 是标量偏差。 通过优化均方误差（MSE）损失函数来训练每个 AttrNet：

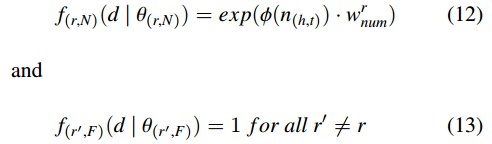
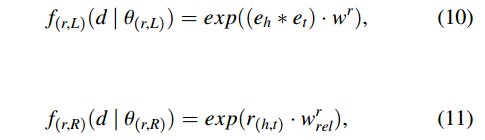


Attrnet 的总体损失是通过将其头部的 MSE 和尾部的 MSE 相加计算得到的，具体如下:

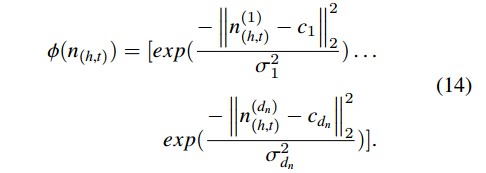


其中(*ai*)∗，(*aj*)∗是地面真相标签。 最后，使用共享的嵌入空间以多任务方式训练这两个网络。

**KBLRN** 将关系特征 (r)、潜在特征 (l) 和数值特征 (n) 结合起来。该模型主要是为了完成 KG 而设计的。它使用概率 PoE (专家产品) 方法将这些特征类型组合起来，并对它们进行端到端的联合训练。 通过采用规则挖掘方法 AMIE + [61] ，将每个关系特征表示为一个逻辑公式，在知识库中计算特征值。 潜在功能通常是使用嵌入方法（例如 DistMult）生成的。 在假设某些关系类型使用数字特征的前提下，头和尾之间的差异可以视为关系本身的特征。 给定一个三元组 *d* =（*h*，*r*，*t*），对于每个（关系类型 r 和特征类型 *F* ∈{*L*，*R*，*N*}）对，基于线性模型和 DistMult 嵌入方法定义各个专家，如下所示：



其中 *wr*，*wrrel*，*wrnum* 是与关系 r 相对应的潜在特征、关系特征和数值特征的参数向量，\* 是元素乘积，・是点乘，*φ* 是按元素方式应用于值 n(*h*,*t*) 的差的径向基函数（RBF），其计算如下：

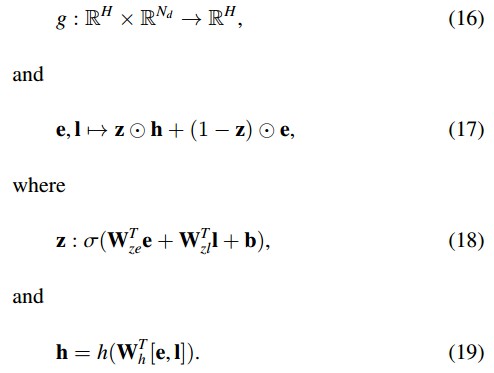


*dn* 在此对应于相关的数值特征。 三元组 *d* =（*h*，*r*，*t*）的 PoE 概率分布定义如下：



其中 *c* 表示所有可能的三元组。 实体嵌入模型的参数由所有专家共享，以便在它们之间创建依赖关系。 在这种方法中，采用负采样和交叉熵损失对 PoE 进行训练，从而使观察到的三元组具有很高的概率。

LiteralE 将文字整合到现有的潜在特征模型中，这些模型是为链接预测而设计的。 在这种方法中，不失一般性的重点在于将数值文本合并到三种最先进的嵌入方法中: DistMult、 ComplEx 和 ConvE。给定一个基本模型（例如 Distmult），LiteralE 修改了 Distmult 使用的计分函数 *f*，用文字强化表示 *eliti* 替换 *f* 中实体 *ei* 的向量表示。利用可学变换函数 *g*，以 *ei* 及其相应的文字向量 *li* 作为输入，将它们映射到一个新的向量，从而生成新的矢量。 函数 *g* 的定义，如下所示，基于 GRU 的概念，以使其灵活，可学习，并能够决定，是否有利于纳入文字信息:

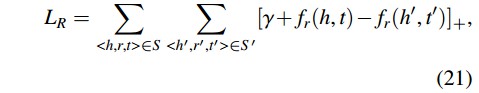


注意，**W***ze* ∈ R*H×H*，**W***zl*∈ R*Nd×H*，**b** ∈ R*H* 和 **W***h* ∈ R*H + Nd×H* 是 g 的参数，*σ* 是 S 型函数，⊙ 表示按元素相乘，而 h 是分量非线性。 得分函数 *f*(e*i*，e*j*，r*k*)已替换为 *f*(*g*(e*i*，l*i*），*g*(e*j*，l*j*)，r*k*)并按照与基本模型相同的过程进行训练。

**TransEA** 具有两个组件模型； 直接采用的基于翻译的结构嵌入模型（即 TransE）和新提出的属性嵌入模型。 在前者中，给定三元组 <h，r，t> 的评分函数定义如下：



其中 ||x||1/2 表示 L1 或 L2 范数。 对于 KG 中的所有关系三元组，结构嵌入的损失函数定义为：



其中 *S'* 表示通过破坏头或尾实体构成的一组三元组，[*x*] + = *max*{0，*x*}，且 *γ* > 0 是余量超参数。

对于属性嵌入，它使用包含数值的所有属性三元组作为输入，并应用线性回归模型来学习实体和属性的嵌入。 给定属性三元组 <*e*，*a*，*v*>，评分函数定义为：



其中 **a** 和 **e** 是属性 *a* 和实体 *e* 的向量，*ba* 是属性 *a* 的偏差。另一方面，给定所有数值为 *T* 的属性三元组，计算属性嵌入的损失函数如下:



TransEA 的主要损失函数（即，*L* = (1-*α*)・*LR* +*α*・*LA*）是通过对具有超参数的组件模型的各个损失函数之和进行定义的，从而为每个模型分配权重。 最后，通过共享实体的嵌入，在训练过程中共同优化了两个模型。

**总结** 尽管支持数字文字，但是所有讨论的嵌入方法都无法解释单位/数据类型文字背后的语义。例如，给定从 DBpedia 中取得的以下两个三元组，

<http://dbpedia.org/resource/Anton\_Baraniak, dbp:weight,"110.0"^^<http://dbpedia.org/datatype/kilogram>,  
<http://dbpedia.org/resource/Katelin\_Snyder, dbp:weight,"110.0"^^<http://dbpedia.org/datatype/pound>

如果忽略公斤和磅类型的语义，则可以认为第一个三元组的文字值 “110.0” 和第二个三元组的文字值 “110.0” 完全相同。 而且，大多数模型没有适当的机制来处理多值文字。

关于模型复杂度，表 2 中列出了每个模型中使用的参数数量，以显示参数方面的复杂性。 注意，模型的复杂度取决于数据集的大小，与其他模型相比，TransEA 具有较低的复杂度。

## *4.3. 具有图像文字的模型*

**IKRL**[51] 通过联合训练基于结构的表示和基于图像的表示来学习 KGs 的嵌入。 通过改编传统的嵌入模型（如 TransE）来学习实体的基于结构的表示形式。对于基于图像的表示，给定一个实体可能具有多个图像实例的事实，将图像编码器应用于为多值图像关系的每个实例生成嵌入。图像编码器由神经表示模块和投影模块组成，以从图像中提取判别特征并将这些表示分别从图像空间投影到实体空间。

表2

在参数数量方面，具有数字文字的模型的复杂性。 Θ 是基本模型中的参数数量，H 是实体嵌入大小，Nd 是数据关系数，Λ 是 MTKGNN 的 Attrnet 网络中隐藏层的大小，Nr 是关系数，M 是属性嵌入大小。

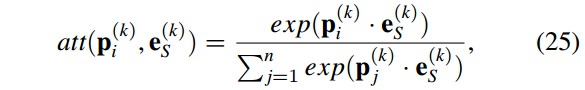
|  |  |
| --- | --- |
| **模型** | **参数** |
| LiteralE with g | Θ + 2*H*2 + 2*Nd*H + H |
| LiteralE with *glin* | Θ + (*Nd*H + H)H |
| MTKGNN | Θ + *Nd*H + 2(2HΛ + Λ) |
| KBLN | Θ + *NrNd* |
| TransEA | Θ + *M* |

对于第 i 个图像，其在实体空间中基于图像的表示 pi 计算为：

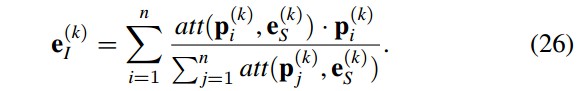


其中 *M* ∈ R*di×ds* 是投影矩阵，其中 *di* 和 *ds* 分别代表图像特征的维数和实体的维数。 *F*(*imgi*) 是图像空间中的第 i 个图像特征表示。

基于注意力的多实例学习用于通过自动计算应给予每个实例的注意力来集成为每个图像实例学习的表示。 第 k 个实体的第 i 个图像表示 pi(k)的注意事项为：

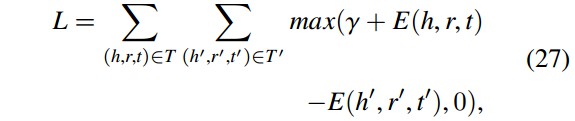


其中 *e(k)S* 表示第 k 个实体的基于结构的表示。 注意力越高，就越类似，基于图像的表示与其对应的基于结构的表示类似，这表示在聚合基于图像的表示时应给予更多的重视。 第 k 个实体的基于聚集门控图像的表示形式定义如下：



给定三元组，则通过组合四个能量函数（即 *E*(*h*，*r*，*t*) = *ESS* + *EII* + *ESI* + *EIS*）来定义整体能量函数。这些能量函数基于两种实体表示形式（即结构 能量函数（即 *ESS* = ||*hS* + *r* - *tS*||）与 TransE 和第二函数（即 *EII* = ||*hI* + *r* - *tI*||）相同。 ）将其相应的基于图像的表示形式用于头部和尾部实体，第三个函数（即 *ESI* = ||*hS* + *r* - *tI*||）基于头部的基于结构的表示形式和基于图像的 尾部实体的表示，而第四个函数（即 *EIS* = ||*hI*+ *r* - *tS*||）正好相反，这些第三和第四个函数确保将基于结构的表示和基于图像的表示都学习到相同的向量空间。

给定能量函数 E（h，r，t），基于边距的评分函数定义如下：



其中 *γ* 是余量超参数，*T’*是通过替换 T 中每个三元组的头部实体，尾部实体或关系而生成的 T 的负样本集。请注意，已从 *T'* 中删除了 T 中已存在的三元组。

**总结** IKRL 通过将基于结构的表示与基于图像的表示相结合，将实体的图像用于 KG 表示学习。 但是，给定三元组 <*h*，*r*，*t*>，以实现对实体 h 和 t 的非常好的表示，要求两个实体都具有与它们关联的图像。 此模型的另一个问题是，仅将图像视为与其关联的那些实体的属性。 例如，如果存在一个包含两个实体 e1 和 e2 的图像，但是该图像仅与 e1 相关联，那么它将被视为 e1 的一个图像实例，而不是 e2 的一个图像实例。 但是，将图像与它们表示的所有实体明确关联在一起，然后再将其用于学习 KG 嵌入，将更为有益。

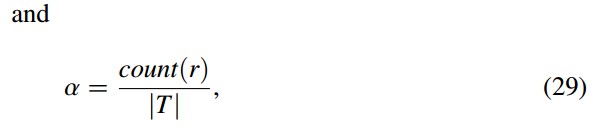
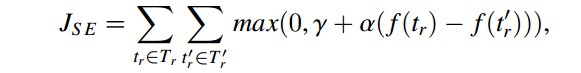
## *4.4. 具有多模态文本的模型*

本节使用至少两种提供补充信息的文字来分析嵌入模型。 首先，讨论具有数字和文本文字的类别，然后讨论具有数字，文本和图像的类别。

### *4.4.1. 具有数字和文本文字的模型*

**带有阻塞功能的 LiteralE**[62] 建议通过将 LiteralE 与 CER 阻塞策略[63]相结合来提高数据链接任务的效率。与 LiteralE 不同，给定三元组 < *h*,*d*,*v* >，除了对象文字值 *v* 之外，它还从头实体 *h* 和数据关系 *d* 的 URI 前缀中获取文字。 CER 阻止基于两次遍历索引方案。在第一遍中，Levenshtein 距离度量用于处理文字对象和 URI 前缀，而在第二遍中，使用 WordNet[64] 进行语义相似度计算以处理对象/数据关系。所有提取的文字都被标记为单词列表，以创建倒排索引。使用与 LiteralE 中相同的训练过程来训练该模型。对于每个给定的三元组 <h,r,t>，采用 LiteralE 的评分函数 f 来计算知识图中所有三元组 < *h*,*r*,*t*’ > 的分数。使用 S 形函数 *p* = *σ*(*f*(.)) 来产生概率。然后，通过相对于三元组的真值向量最小化产生的概率函数向量的二进制交叉熵损失来训练模型。

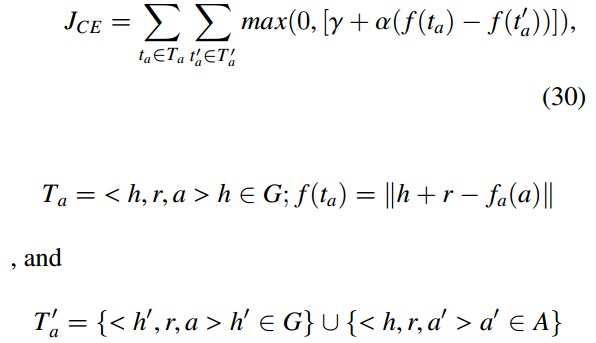
**EAKGAE**[65] 是一种通过学习 KG 的统一嵌入空间来设计 KG 之间的实体对齐的方法。 实体对齐任务具有三个主要模块：谓词对齐，嵌入学习和实体对齐。 谓词对齐模块通过重命名相似的谓词将两个 KG 合并在一起，从而为关系嵌入创建统一的向量空间。 嵌入学习模块使用结构嵌入（通过适应 TransE）和属性字符嵌入来联合学习两个 KG 的实体嵌入。 定制后的 TransE 可以使谓词对齐的三元组得到更多的关注。 这是通过增加权重 *α* 来控制三元组中的嵌入学习而获得的。 因此，为基于结构的嵌入定义了以下目标函数 *JSE*：



其中 *Tr* 和 *T'r* 分别是有效三元组和损坏的三元组，*count*(*r*) 是关系 *r* 的出现次数，|*T*| 是合并的 KG 中三元组的总数。

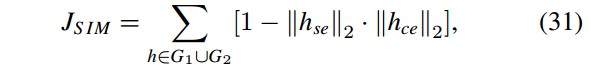
另一方面，属性字符嵌入旨在从出现在与实体相关联的属性中的字符串中学习实体的嵌入。目的是使来自两个 KG 的实体嵌入能够落入相同的向量空间，尽管属性来自不同的 KG。属性字符嵌入的灵感来自 TransE 中的翻译概念。给定一个三元组（*h*，*r*，*a*），数据属性 *r* 被解释为从头部实体 *h* 到文字值 *a* 的转换，即 *h* + *r* = *fa*(*a*) 其中 *fa*(*a*) 是一个合成函数。此函数将属性值编码为单个向量，将相似的属性值映射为相似的表示形式。已经提出了三种不同的组成函数 SUM，LSTM 和基于 N-gram 的函数。 SUM 定义为属性值的所有字符嵌入的总和。在 LSTM 中，最终的隐藏状态被视为属性值的矢量表示。基于 N 元语法的函数根据其实验显示出比其他函数更好的性能，该函数使用属性值的 N 元语法组合的总和。

为属性字符嵌入定义了以下目标函数：



其中，*Ta* 和 *T'a* 是有效属性三元组和损坏的属性三元组，而 A 是给定 KG G 中的属性集。损坏的三元组是通过将头部实体替换为随机实体或将属性替换为 a 随机属性值。 在此，*f*(*ta*) 是基于头实体 h 的嵌入，关系 r 的嵌入以及使用合成函数 *f*(*ta*) 之一获得的属性值的矢量表示而计算出的合理性评分。

通过最小化以下目标函数，利用嵌入 hce 的属性特征将嵌入 hse 的结构转换到同一个向量空间:



其中 ||*x*||2 是向量 x 的 L2 - 范数。 这样，通过基于实体关系的结构嵌入和通过基于属性值的属性嵌入，可以获取来自两个 KG 的实体的相似性。

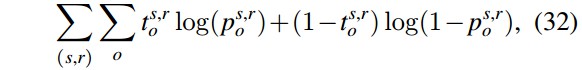
所有这三个函数都被汇总为一个总体目标函数 J（即 *J* = *JSE* + *JCE* + *JSIM*），用于共同学习结构和属性嵌入。 最后，通过定义具有指定阈值的相似性方程式来完成对齐。 此外，已应用传递性规则来丰富 KG 中的三元组，以获得更好的属性嵌入结果。

**总结** 这两种方法(带阻塞的 LiteralE 和 EAKGE)的共同缺点是文本和数值文本以相同的方式处理。 他们的方法也不考虑文字数据类型语义或多值文字。 此外，由于 EAKGAE 使用基于字符的属性嵌入，它无法捕捉音节共现背后的语义。

### *4.4.2. 具有数字，文本和图像文字的模型*

**MKBE**[49] 是一种多模式 KG 嵌入，其中文本，数字和图像文字被一起建模。这种方法的主要目标是利用所有观察到的主题，对象和关系（对象属性和数据属性），以预测是否存在任何事实。它扩展了 DistMult，它通过为不同数据类型添加神经编码器来创建实体和关系的嵌入。给定一个三元组 <*s*，*r*，*o*>，头实体 *s* 和关系 *r* 使用一个热编码通过密集层编码为独立的嵌入矢量。类似地，如果对象 *o* 是分类值，则它将通过具有 relu 激活的密集层表示，该层具有与嵌入空间维数相同的节点数。另一方面，如果对象 *o* 是一个数值，则在标准化输入之后使用前馈层，以便通过将 *o* 投影到更高维度的空间来学习 *o* 的嵌入。如果 *o* 是短文本（例如名称和标题），则使用基于字符的堆叠式双向 GRU 对其进行编码，并且顶层的最终输出将用作 *o* 的表示形式。相反，如果 *o* 是诸如实体描述之类的长文本，则将使用单词嵌入上的 CNN 来获取 *o* 的嵌入。对象 *o* 也可以是图像，在这种情况下，使用 ImageNet[66] 上 VGG 预训练网络的最后一个隐藏层，然后进行紧凑的双线性池化，即可获得 *o* 的嵌入。给定实体，关系和属性的矢量表示形式，使用 DistMult 相同的评分功能确定三元组的正确性概率。

如下定义的二进制交叉熵损失用于训练模型：



对于给定的主题关系对（*s*，*r*），所有实体上的二进制标签向量 *ts,r* 用于指示在训练过程中是否观察到 <*s*，*r*，*o*>。 pos,r 表示使用 S 型函数计算出的任何三元组 <*s*，*r*，*o*> 的模型的真实概率。

此外，使用这些学习的嵌入和不同的神经解码器，引入了一种新颖的多峰插补模型，以从知识库中的信息生成缺失的多峰值，例如数值数据，分类数据，文本和图像。 为了预测缺失的数字和分类数据（例如日期，性别和职业），使用了一个简单的实体嵌入前馈网络。 对于文本，对抗性正则化自动编码器（ARAE）已用于训练生成器，该生成器从连续代码解码文本，并使生成器以实体嵌入而不是随机噪声向量为条件。 类似地，BE-GAN 结构与 pix2pix-GAN 模型的组合用于生成图像，从而将生成器置于实体嵌入上。

**总结** 尽管已尝试将文本文字，数字文字和图像合并到 KG 嵌入中，但是模型（MKBE）未能捕获（数字）文字值的数据类型/单位的语义。 此外，类似于 IKRL，仅当 *I* 最初与所考虑的数据集中的 *e* 相关联时，才将图像 *I* 作为某个实体 *e* 的实例（有关更多详细信息，请参见第 4.3 节）。

# 5. 应用

本节讨论了 KG 嵌入的不同应用，在这些应用中已经对前面描述的方法进行了训练 和/或 评估。

***链接预测***。 一般来说，链路预测可以定义为在复杂网络中识别缺失信息的任务。 特别是在 KGs 的情况下，连结预测模式旨在预测利用现有连结进行训练的实体之间的新关系[67, 68]。除了预测实体之间的关系，链接预测还着重于预测任何一个关系的头或尾实体。然后，它决定一个新的三元组是否有效(KG 中没有观察到这一点)。 在形式上，设 *G* 是一个 KG，具有一组实体 *E* = { *e1*,...，*en* }和一组对象关系 *R* = {*r1*，...，*rm*} ，则链路预测可以用一个映射函数*ψ*来定义: *E* × *E* × *R* → R，该映射函数给每个可能的三元组(*ei*，*ej*，*rk*) ∈ *E × E × R* 分配一个得分。 高分表明那个三元组最有可能是正确的。

链接预测是用于评估 KG 嵌入性能的最常见任务之一。头部预测，尾部预测和关系预测是与链接预测有关的不同种类的子任务。头部预测旨在识别给出了关系和尾部实体的丢失的头部实体，并且类似地用于尾部预测和关系预测。在第 4 节中讨论的大多数模型已针对这些预测任务中的一项或多项进行了评估。头尾预测用于评估模型 LiteralE [45]，TransEA[46]，KBLRN[47]，KDCoE[50]，EAKGAE[65] 和 IKRL[51]。另一方面，DKRL[11] 已在各种链接预测任务上进行了评估：头，尾和关系预测，而 MKBE[49] 已在尾部预测上进行了评估。在 Extended RESCAL[24] 中，已经对 Yago 2[69] 进行了两种链接预测实验，即 i）通过将关系类型固定为 rdf：type 进行尾部预测，ii）所有关系的通用链接预测实验类型。不幸的是，不可能对所有这些模型的评估结果进行比较，因为实验是在不同的数据集上进行的，并且遵循了不同的链接预测程序。考虑到这一点，在本次调查中，针对这些模型的头和尾预测任务进行了实验（请参见第 6 节）。

***三重分类***。 三分类任务的目标与链路预测的目标是一致的。 一个潜在的三元组 (ei，ej，rk) 被分类为 0 (false) 或 1 (true) ，即二进制分类任务。 嵌入模型 MTKGNN[11] 和 IKRL[51] 已针对此任务进行了评估。 但是，由于他们不使用公共评估数据集，因此无法直接比较报告的结果。

***实体分类***。 给定一个 KG G，它具有一组实体 E 和类型 T，并且具有实体 *e* ∈*E* 和类型 *t* ∈ *T*，则实体分类的任务是确定是否未观察到潜在的实体类型对（*e*，*t*）在 G((*e*，*t*) ∉ *G*）中是否存在缺失事实。 此任务是使用多标签分类算法的实体类型预测，将 G 中的实体类型视为给定类。 在 DKRL[11] 中，已使用数据集 FB15K[10] 对提出的模型进行了评估。

***实体对齐***。 给定两个 KGs *G1* 和 *G2*，实体对齐任务的目标是确定那些实体对（*e1*，*e2*），其中 *e1* 是 *G1* 中的实体，*e2* 是 *G2* 中的实体，它们表示相同的现实世界实体，因此 通过这些统一的实体，即实体对，可以实现 *G1* 和 *G2* 的集成。 最近针对实体对齐任务提出了不同的基于嵌入的模型。 在本次调查所包括的模型中，已经提出了 EAKGAE[65] 和 KDCoE[50] 用于实体对齐任务。具体来说，KDCoE[50] 使用跨语言的实体对齐任务，该任务确定不同语言的相似实体。尽管这两个模型都使用相同的评估任务（实体对齐任务），但是由于它们基于不同的数据集，因此无法比较它们的实验结果。

***其他应用***。属性值预测，最近邻分析，数据链接和文档分类是用于评估正在讨论的模型的其他应用场景。 属性值预测是预测 KG 中（离散）属性值的过程。 例如，可以使用通常被视为 KG 完成任务的属性值预测任务来识别体重的缺失值。 在 MTKGNN[48] 中，属性值预测是使用特定于属性的线性回归分类器进行评估的。 通过估算不同的多峰属性值，MKBE[49] 中采用了相同的任务进行模型评估。

最近邻分析是一项任务，用于检测通过嵌入模型学习的潜在空间中某些给定实体的最近邻。 已在 LiteralE[45] 中执行了此任务，以将 DistMult + LiteralE 与基本模型 DistMult 进行比较。 另一方面，数据链接和文档分类任务已分别在 LiteralE 中使用阻塞[62] 和带有文字[58] 的 KGlove 使用（有关更多详细信息，请参阅[62] 和[58]）。 表 9 总结了评估了带文字的 KG 嵌入模型的所有应用程序。

# 6. 链路预测实验

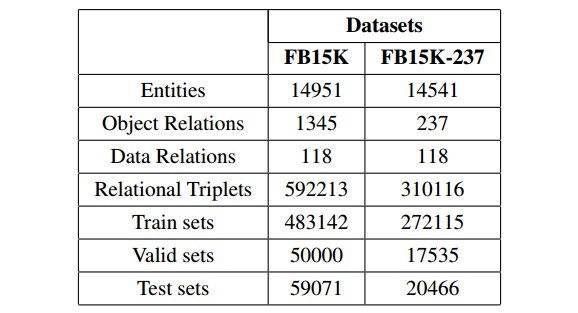
本节对前一节在统一环境条件下讨论的方法进行了实证评价，并基于应用于链接预测任务的方法的性能讨论了结果。 在这项工作中，选择链接预测是因为对大多数带有文字的 KG 嵌入模型进行了训练和评估。 进行这些实验时遇到的主要问题之一是这些模型中某些模型的源代码不是公开可用的，并且不容易复制。 这种方法被排除在实验之外。 在随后的部分中，将介绍数据集以及带有文本，数字，图像和多模式文字的实验。

## *6.1. 数据集*

使用两个最常用的链接预测数据集（即 FB15K[10] 和 FB15K-237[70]）测量了上述模型的性能。 FB15K 是 Freebase[5] 的子集，其中大部分包含描述电影，演员，奖项，体育和运动队的事实的三元组。 它包含随机分组的训练，验证和测试集。 该数据集的问题在于，测试集包含大量的三元组，而这些三元组是通过简单地将训练集中的三元组反转而获得的。 这使得相对于头和尾实体对称的简单嵌入模型能够获得出色的性能。 为了避免这种情况，已通过从 FB15K 中删除逆关系创建了数据集 FB15K-237。 表 3 给出了这些数据集的统计信息。

表3

FB15K 和 FB15K-237 数据集的实体数量，对象关系，数据关系，关系三元组，训练集，有效集和测试集。



## *6.2. 文本文字实验*

如第 4.1 节所述，带有文字的嵌入模型 Extended RESCAL，DKRL，KDCoE 和 KGloVe 使用文本文字。 但是，由于以下问题，除 DKRL 以外的所有这些模型均未考虑进行试验：

* 带有文字的 KGloVe 模型的实现不是公开可用的，并且不容易复制。
* KDCoE 专为跨语言实体对齐任务而设计，这使其很难将其应用于链接预测。
* 在扩展 RESCAL 的情况下，该方法实际上在计算上是昂贵的，因此不被视为合并文字的可行嵌入模型。 此外，本文讨论的带文字的模型均未在实验中考虑扩展 RESCAL。

***扩展数据集：***为了用文本文字进行实验，表 3 中给出的数据集 FB15K 和 FB15K-237 都用一组仅包含文本文字的 15239 属性三元组进行了扩展。 对于文本（实体描述）的预处理，已使用 spacy.io1。 这包括令牌化，命名实体识别以及将数字转换为文本，即 16 已转换为 “sixteen”。 在预处理步骤之后，将删除所有实体以及不包含少于 3 个单词的简短描述的相应三元组。 而且，如作者在论文中提到的那样，删除了包含这些实体的三元组。 此外，对于具有多个文本描述的实体，仅随机选择一个描述。

***实验设置：***用于 DKRL 的超参数如下：学习速率 0.001，嵌入大小 100，损失余量 1，批量大小 100，单位时间 1000。实验在具有 503GiB RAM 和 2.60GHz 速度的 Ubuntu 16.04.5 LTS 系统上进行。

***评估程序和结果：***基于链接预测任务评估模型的性能。 对于测试集中的每个三元组，将针对头部或尾部实体生成一组损坏的三元组。 如果该头实体被 KG 中的任何其他实体替换，则称其三元组相对于其头实体已损坏，并且类似地，针对其尾部实体损坏了三元组。 评估指标 MR（平均排名），MRR（平均倒数排名），hits@1，hits@3 和 hits@10 用于评估模型的性能。 由于在训练，验证或测试集中出现的损坏的三元组可能会低估指标，因此在使用指标计算分数之前，将其过滤掉。 因此，本文报道的实验结果基于上述（过滤）设置。

表 4 中显示了 FB15K 和 FB15K-237 数据集上的链接预测结果。这些结果分别针对头部实体，尾部实体和关系预测进行报告。还提供了通过对头和尾预测的平均值获得的总体结果。最好的分数是用粗体突出显示的分数。要注意的是，与 FB15K 相比，数据集 FB15K-237 取得了更好的结果，因为前者不包含对称关系。此外，结果表明，对于两个数据集，具有伯努利分布的 DKRL 模型（DKRLBern）的性能要优于具有均匀分布的模型（DKRLunif）。如[12] 中定义的用于采样的伯努利分布是概率分布，*tph*/(*tph* + *hpt*)，其中 *tph* 是每个头实体的尾部实体的平均数量，*hpt* 是每个尾部实体的头部实体的平均数量。给定黄金三元组（*h*，*r*，*t*），具有上述概率，则三元组通过替换头部而损坏，而概率 *hpt*/(*tph* + *hpt*)，则三元组通过替换尾部而损坏。 DKRLBern 最适合预测 MRR，Hits@1 和 Hits@3 的头部，关系和尾部，而 DKRLUnif 方法根据这两个数据集的 MR 均能更好地预测。对于 FB15K-237 数据集，DKRLBern 比 DKRLUnif 稍好。

## *6.3. 使用数字文本进行实验*

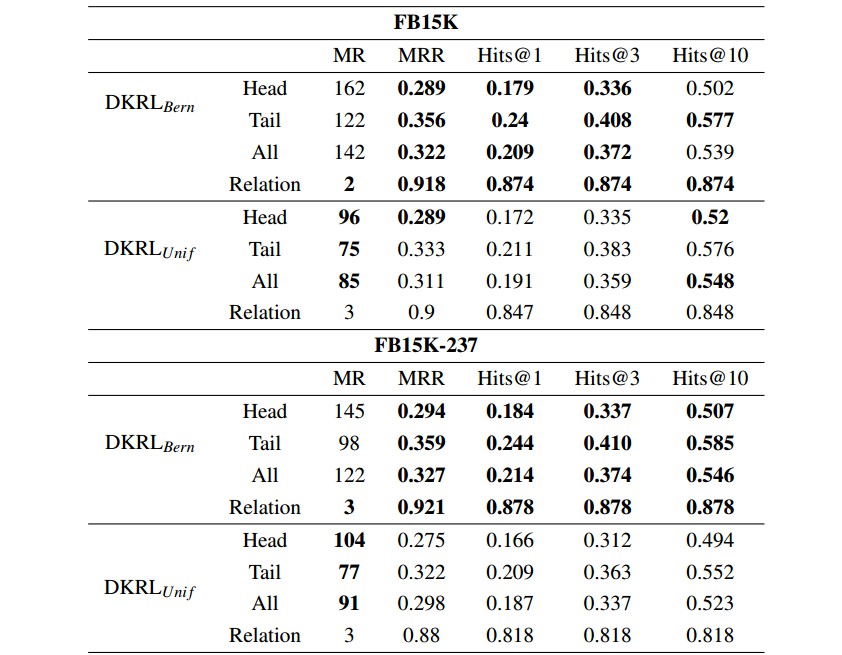
MT-KGNN，KBLRN，LiteralE 和 TransEA 是使用数字文字的 KG 嵌入模型（请参见第 4.2 节）。实验中使用 KBLN（KBLRN 的子模型）代替了主要模型 KBLRN，该子模型排除了图特征方法提供的关系信息。之所以如此，是因为 KBLN 可直接与其他三个模型（即 MT-KGNN，LiteralE 和 TransEA）进行比较，而 KBLRN 则不能。 TransEA 模型的 code2 是 TransEA[46] 的原始实现，其中 MT-KGNN，KBLN 和 LiteralE 模型的源代码 3 来自 LiterlaE[45] 的实现。如本节所述。 4.2，MT-KGNN 的基于结构的嵌入组件基于神经网络，称为 RelNet。然而，在 LiteralE[45] 中实现的版本中，他们已经用 DistMult 代替了 RelNet 作为基线，以便能够直接将类似 MTKGNN 的方法与他们提出的方法进行比较。因此，在本次调查中，使用了类似 MT-KGNN 的模型来代替原始的 MT-KGNN 模型。

此外，根据基准模型和所使用的转换函数，模型 LiteralE 具有不同的品种。 如第 4 节所述，在 LiteralE 中，有两个转换函数：*g*（基于 GRU 的函数）和 *lin*（一个简单的线性函数），并且有三个基线模型：DistMult，ConvE 和 ComplEx。 因此，在本实验中，考虑了 LiteralE 模型的六个变体：DistMult-Literaleg，ComplEx-Literaleg，ConvELiteralelin，DistMult-Literalelin，ComplEx-Literalelin 和 ConvE-Literalelin。 数据集，实验设置和评估结果将在后续章节中讨论。

***扩展数据集：***为了用数字文字进行实验，表 3 中给出的数据集 FB15K 和 FB15K-237 都扩展了一组 23521 属性三元组，它们仅包含数字文字。 这些三元组是基于 TransEA 的属性三元组创建的[46]。 在 TransEA 中，作者提供了一组属性三元组，其中对象值是数字。 但是，不可能直接使用此数据，因为文字值会按照模型的要求在 [0-1] 区间中进行归一化，但是本实验中的其他模型（如 LiteralE）会使用原始的未归一化的文字值来代替。 因此，有必要查询 Freebase 以替换 TransEA 属性三元组数据中每个（主题，数据关系）对的规范化对象文字值。 此外，仅考虑至少出现在三个三元组中的那些数据关系。

表4

在 FB15K 和 FB15K-237 数据集上使用 DKRL 模型进行实验的结果。



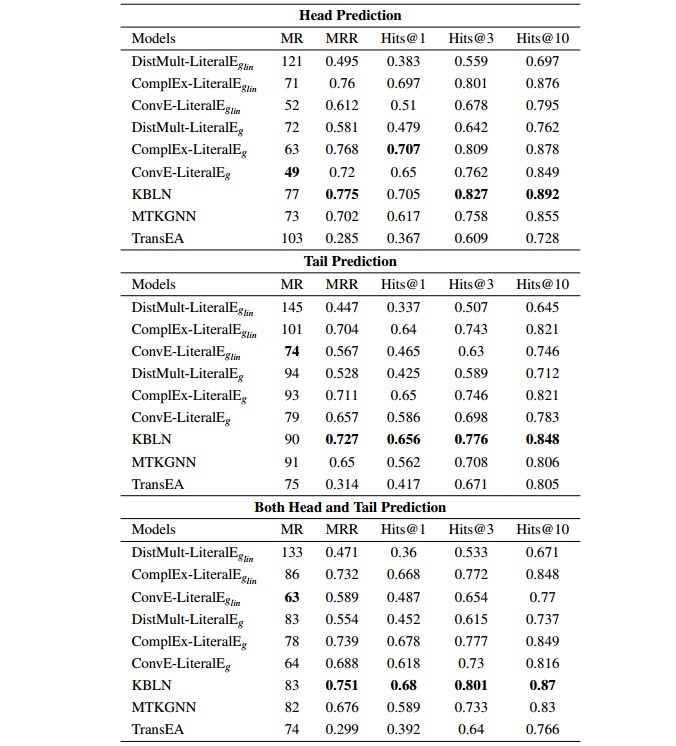
***实验设置：***与 LiteralE 中一样，除了 TransEA 以外，所有两个数据集用于所有模型的超参数是：学习率 0.001，批处理大小 128，嵌入大小 100 或 200，嵌入丢失概率 0.2，标签平滑 0.1 和历元 100。 该实验遵循与 LiteralE [45] 实验相同的 1-N 训练方法（有关更多详细信息，请参考 LiteralE [45]）。 对于 TransEA，对于这两个数据集，参数均为：时代 3000，维度 100，批次 100，保证金 2 和学习率 0.3。 实验是在具有 503GiB RAM 和 2.60GHz 速度的 Ubuntu 16.04.5 LTS 系统上进行的。 TITAN X（Pascal）GPU 已用于 LiteralE，KBLN 和 MTKGNN 模型。

***评估过程和结果:*** 在第6.2节中讨论的相同的评估指标已经被用来评估链接预测任务的数字文字模型的性能。 如表5所示，从总体结果来看，KBLN 模型在除先生之外的所有指标上都比其他模型具有更好的性能。这是因为这个模型能够处理 FB15K 中的逆关系，当实体被用作尾部时应用实体嵌入的共轭复数，当实体被用作头部时应用正常嵌入。

另一方面，参考表 6 所示的 FB15K-237 数据集的总体结果，根据所有指标，模型 DistMult-LiteralEg 优于其他模型。 这意味着将 LiteralE 应用于 FB15K-237 上的 DistMult 会提供比将其应用于其他基准模型更好的性能。 请注意，DistMult-LiteralEg 模型在 FB15K-237 数据集上获得最佳结果的原因是该数据集没有任何对称关系。 关于两个变换函数 *g* 和 *glin*，根据两个数据集上的结果，函数 *g* 比 *glin* 产生了更好的结果。

表5

使用过滤设置的 FB15K 数据集上的链接预测结果。



## *6.4. 图片实验*

注意，不可能将整个 MKBE [49] 与任何其他模型进行比较，因为它是唯一同时使用三种类型的文字的嵌入模型：文本，数字和图像。 因此，将其仅使用图像的子模型 S + I 与嵌入模型 IKRL[51] 进行了比较。 由于这种比较已经由 MKBE[49] 的作者完成，因此表 7 中显示的结果直接取自他们的论文。 他们比较了 DistMult+S+I，ConvE+S+I 和 IKRL 模型，其中 S 代表结构，I 代表图像。 DistMult+S+I 和 ConvE+S+I 都是 MKBE 的子模型，它们仅使用关系三元组和图像。 结果表明，在 YAGO-10 数据集的所有指标上，ConvE+S+I 均优于其他两个模型（有关更多详细信息，请参阅 MKBE[49]）。

## *6.5. 多模式文本的实验*

如第 4 节所述，现有的多模式嵌入分为两种类型：i）具有文本文字，数字文字和图像文字的模型，以及 ii）具有文本和数值文字的模型。但是，由于 MKBE 是第一类中唯一的模型，因此只能将其子模型 *S+I* 与 IKRL 进行比较（请参见 6.4 节）。对于带有文字和数字文字的模型，即带有阻塞功能的 LiteralE 和 EAKGAE，它们也未包含在实验中。 EAKGAE 的问题与 KDCoE 的问题相同，即，它是在实体对齐任务上接受训练的，因为没有 LiteralE 带阻塞的原因是其代码不可公开获得。相反，在作者进行的实验中，还采用了 LiteralE（带有数字的模型）将文字文字包含在内。同样，在我们的实验中，已尝试将文本和数字文字结合使用 LiteralE 方法，即表 8 中的模型 DistMult-LiteralEg-text。然后，将结果与只使用数字字面值 (DistMult-LiteralEg) 和只使用文本字面值的模型 DKRL 进行比较，以探讨利用不同类型字面值表示的信息的好处。

DistMult-LiteralEg 文本是一个模型，该模型通过使用数字和文本文字，将 LiteralE 方法应用于 DistMult。请注意，此处选择 DistMult 作为基准，是因为使用 6.3 节中讨论的模型可以在 FB15K-237 数据集上进行数字实验的最佳结果。表 3 中列出的数据集也与其他文本属性三元组（对实体的描述）一起用于本实验。还比较了 DistMult-LiteralEg 文本与其仅等效数字的 DistMult-LiteralEg 和 DKRLBern。

在数据集 FB15K 和 FB15K-237 上获得的实验结果如表 8 所示。 结果表明，在 FB15K 数据集上采用 DistMult-LiteralEg-text 方法结合文本和数值文本，并不能产生比其他模型 DistMult-LiteralEg 和 DKRLBern 更好的结果。如前所述，该数据集包含一组逆关系，这可能导致三元组的逆具有不同的标签。鉴于 DistMult 无法为这种不对称关系建模的事实，将更多文字与 DistMult 结合使用可能会带来比提高性能更多的噪音。 另一方面，对于 FB15K-237 数据集，根据除 MR 之外的所有度量，与其他两个模型相比，DistMult-LiteralEg-text模型对头部实体预测的效果更好。 对于尾部实体预测，对于相同数据集的所有度量，DKRLBern 效果更好。

表6

过滤设置在 FB15K-237数据集上的链路预测结果。

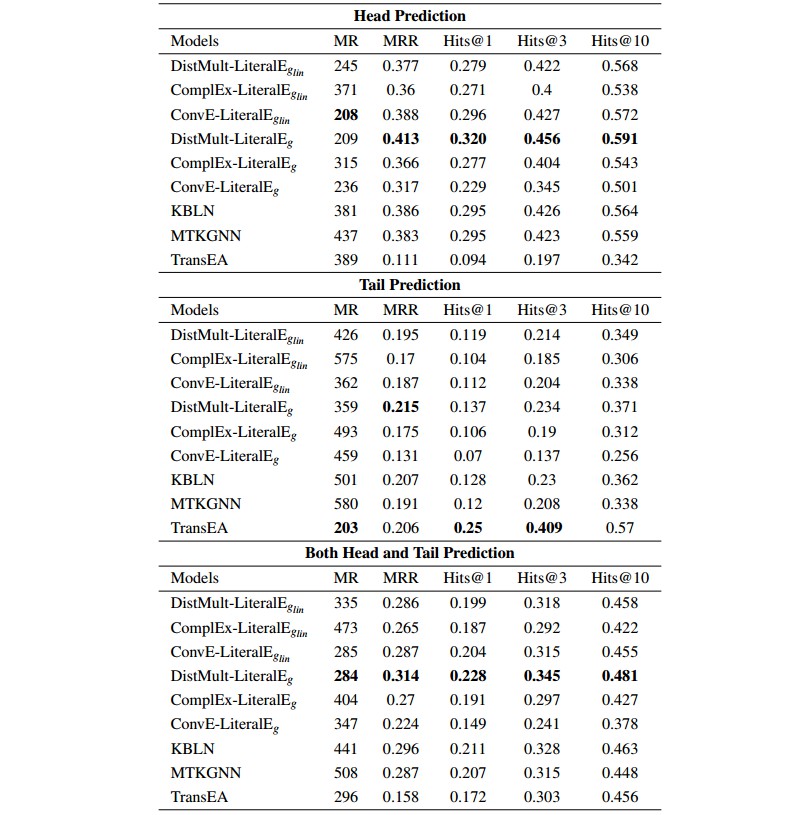
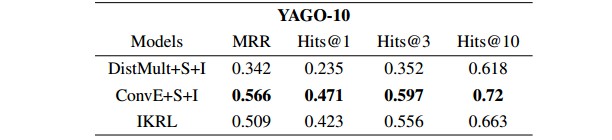


表7

来自 MKBE 的 YAGO-10 上的链接预测任务的 MRR 结果 [49]。



# 7. 讨论及结论

鉴于最近大量关注在各种 KG 中和 KG 外应用中使用 KG（即，KG 中的应用属于输入 KG 的范围，而 KG 外的应用又涵盖了输入 KG 的范围） 在输入 KG 的范围之外的域）中，已经提出了不同的 KG 嵌入技术来实现 KG 的有效使用。 在这些技术中的一些中，已经尝试利用 KG 中存在的文字表示的信息，以更好地嵌入 KG 的元素，即实体和关系。在本文中，对带有文字的 KG 嵌入模型进行了全面的调查。该调查基于所提出的方法以及它们的应用场景和局限性对这些模型进行了详细的分析和分类。此外，针对这些模型的链接预测任务进行了各种实验，以比较模型的性能。

如第 4 节所述或从第 6 节的实验结果可以看出，这些嵌入模型具有不同的缺点，例如：

* 任何模型都未考虑数据类型/单元对文字语义的影响。
* 大多数使用数字文字的嵌入模型，例如 LiteralE，TransEA，MT-KGNN 和 KBLN，仅考虑日期类型文字的年份部分，而忽略月和日值。这阻碍了正确捕获以此类文字表示的信息的能力。例如，给定以下三个日期键入的文字值：

"1999-10-29"^^xsd:date,

"1999-04-14"^^xsd:date,

"1999-10-30"^^xsd:date,

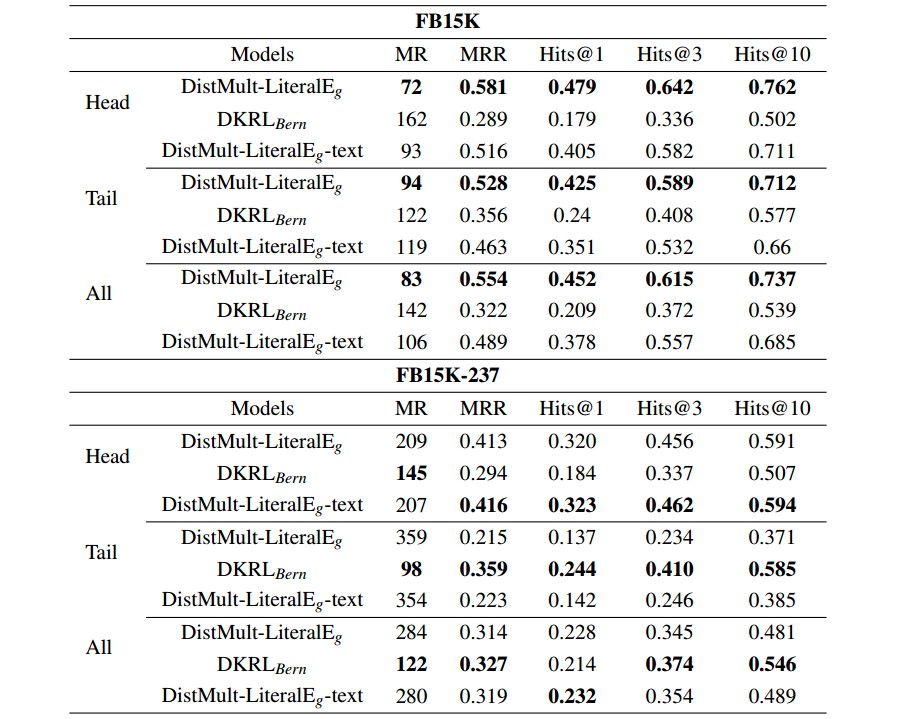
尽管第一个日期值比第二个值更接近第三个值，但仅使用这些值的年份部分的模型将所有这些值视为完全相同。

* 大多数模型也没有适当的机制来处理多值文字。
* 大多数模型的性能取决于用于训练和测试的数据集，这表明这些模型并不可靠。例如，参考表 8，模型 DistMult-LiteralEg-text 的结果表明，由于模型中使用的技术和数据集的性质，将文本和数字文字组合在一起可在 FB15K-237 上产生更好的性能，但在 FB15K 上则不会，（请参阅第 6.5 节）。
* 并非所有模型都能有效地组合不同类型的文字。例如，与 DistMult-LiteralEg（仅数字文字）相比，在数据集 FB15K 上组合文本和数字文字的 DistMult-LiteralEg-text（数字 + 文字）的性能较低。
* 仅提出了几种方法用于多模式 KG 嵌入，并且没有一种方法考虑到带有与诸如音频，视频或 pdf 文件之类的项目相关联的 URI 的文字。

现有模型的上述缺点清楚地表明，需要对如何解决获得不同的固有语义的不同类型的文字进行深入研究。例如，通过这种详细分析得出的一个可能的观点是，需要正确处理数据类型的文字，例如以 *千克* 和 *磅* 为单位的数据关系 *权重* 的值。针对此问题的一种可行解决方案是将这些文字值标准化为标准化的度量，并在表示学习过程中分别处理权重和长度之类的不同度量。

表8

使用过滤集链接 FB15K 和 FB15K-237 数据集上的预测结果。



我们不能期望通过遗漏原始 KG 中存在的可用信息，它的潜在表示形式只是原始 KG 的近似值，将在依赖于其语义信息内容的任务上同样出色地执行。 总体而言，将具有语义正确表示的数据类型文字包括在表示学习过程中，将增加模型的语义内容，从而可能改善质量。

# 参考文献

[1] A.Bordes, S.Chopra和J.Weston,《带子图嵌入的问答》, 载于:《 2014 年自然语言处理经验方法会议(EMNLP)会议录》, 计算语言学协会, 多哈, 卡塔尔, 2014年, 第615–620页.doi:10.3115/v1/D14-1067. https://www.aclweb.org/anthology/D14-1067

[2] F.Zhang, N.J.Yuan, D.Lian, X.Xie和W.-Y.Ma, 面向推荐系统的协同知识库嵌入, 载于: 第22届 ACM SIGKDD知识发现和数据挖掘国际会议论文集, KDD’16, ACM, 美国纽约, 2016年, 第353-362页.ISBN 978-1-4503-4232-2.doi:10.1145/2939672.2939673

[3] J.Weston, A.Bordes, O.Yakhnenko和N.Usunier, 将语言和知识库与关系提取的嵌入模型相连接, 载于: 2013年自然语言处理经验方法会议论文集, 2013年

[4] J.Lehmann, R.Isele, M.Jakob, A.Jentzsch, D.Kontokostas, P.N.Mendes, S.Hellmann, M.Morsey, P.Van Kleef, S.Auer等人, DBpedia–大型多语言知识库, 摘自 Wikipedia, 语义网(2015)

[5] K.Bollacker, C.Evans, P.Paritosh, T.Sturge和J.Taylor, Freebase: 一个协同创建的人类知识结构图数据库, 载于: ACM SIGMOD 国际数据管理会议上, 2008年

[6] D.Vrandeciˇc和M.Krötzsch, 维基数据: 免费的协作知识库(2014)

[7] F.Mahdisoltani, J.Biega和F.M.Suchanek, Yago3:多语言维基百科的知识库, CIDR, 2013年

[8] A.Bordes, J.Weston, R.Collobert和Y.Bengio,《学习结构化的知识嵌入》,载于:《第二十五届AAAI人工智能会议论文集》, AAAI'11, AAAI出版社, 2011年, 第301-306 页.http://dl.acm.org/citation.cfm?id=2900423.2900470

[9] M.R. Garey和D.S. Johnson, 《计算机与难处理》, NP 完全性理论指南, W.H.Freeman＆Co., 美国纽约, 1990年, ISBN 0716710455

[10] A.Bordes, N.Usunier, A.Garcia-Duran, J.Weston和O.Yakhnenko,《翻译多关系数据建模的嵌入》, NIPS, 2013年

[11] R.Xie, Z.Liu, J.Jia, H.Luan和M.Sun, 带有实体描述的知识图的表示学习, 载于:AAAI, 2016.

[12] Z.Wang, J.Zhang, J.Feng和Z.Chen,通过在超平面上进行翻译来嵌入知识图,2014 年.https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8531

[13] Y. Lin, Z. Liu, M.Sun, Y.Liu和X.Zhu, 《用于知识图完成的学习实体和关系嵌入》, 2015年, https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9571

[14] G.Ji, S.He, L.Xu, K.Liu和J.Zhao, 通过动态映射矩阵嵌入知识图, 2015年, 第687–696页.doi:10.3115/v1/P15-1067

[15] G.Ji, K.Liu, S.He和J.Zhao,《具有自适应稀疏转移矩阵的知识图完成》, 2016年.https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11982

[16] Y.Jia, Y.Wang, H.Lin, X.Jin和X.Cheng, 知识图嵌入的局部自适应翻译, 载于:第三十届AAAI人工智能会议论文集, AAAI'16, AAAI出版社, 2016年, 第992-998页.http://dl.acm.org/citation.cfm?id=3015812.3015960

[17] M.Nickel, V.Tresp和H.-P.Kriegel,《基于多关系数据的集体学习的三向模型》,载于:ICML, 2011年

[18] B.Yang, W.Yih, X.He, J.Gao和L.Deng, "在知识基础中嵌入实体和关系以进行学习和推理", 第3届国际学习代表大会, ICLR, 2015年, 圣地亚哥, 美国加利福尼亚州, 2015年5月7日至9 日, 会议记录, 2015 年.http://arxiv.org/abs/1412.6575

[19] M.Nickel, L.Rosasco 和 T.Poggio, 知识图的全息嵌入, 载于:第三十届AAAI人工智能会议论文集, AAAI’16, AAAI 出版社, 2016年, 第1955–1961页.http://dl.acm.org/citation.cfm?id=3016100.3016172

[20] T.Trouillon, J.Welbl, S.Riedel, E.Gaussier和G.Bouchard, "用于简单链接预测的复杂嵌入", 载于:第33届国际机器学习国际会议论文集-第48卷, ICML, 16, JMLR.org, 2016年, 第2071-2080页. http://dl.acm.org/citation.cfm?id=3045390.3045609

[21] A.Bordes, X.Glorot, J.Weston和Y.Bengio, 一种用于多关系数据学习的语义匹配能量函数, 机器学习94(2)(2014), 233-259.doi:10.1007/s10994-013-5363-6

[22] R.Socher, D.Chen, C.D.Manning和A.Ng, 使用神经张量网络推理完成知识库, 载于:神经信息处理系统的进展26, C.J.C.Burges, L.Bottou, M.Welling, Z.Ghahramani和K.Q.Weinberger编辑, Curran Associates, Inc., 2013年, 第926–934页

[23] X.Dong, E.Gabrilovich, G.Heitz, W.Horn, N.Lao, K.Murphy, T.Strohmann, S.Sun和W.Zhang, 知识库:概率论的网络规模方法Fusion, 载于:第20届ACM SIGKDD知识发现和数据挖掘国际会议论文集, KDD'14, ACM, 纽约, 美国, 2014年, 第601-610页.ISBN978-1-4503-2956-9.doi:10.1145/2623330.2623623

[24] M.Nickel, V.Tresp和H.-P.Kriegel, 《分解Yago:链接数据的可伸缩机器学习》, 载于:第21届万维网国际会议论文集, ACM, 2012 年

[25] S.Guo, Q.Wang, B.Wang, L.Wang和L.Guo, SSE:知识图的语义平滑嵌入, IEEE知识与数据工程学报, PP(2016年), 1-1.doi:10.1109/TKDE.2016.2638425

[26] R.Xie, Z.Liu和M.Sun, “具有分层类型的知识图的表示学习”, 载于:IJCAI, 2016年

[27] D.Krompa β, S.Baier和V.Tresp, 知识图的类型约束表示学习, 载于:第14届语义网ISWC国际会议论文集-第9366卷, 施普林格出版社, 柏林, 海德堡, 2015年, 第640-655页.ISBN978-3-319-25006-9

[28] Q.Wang, B.Wang 和 L.Guo, 《使用嵌入和规则完成知识库》, 载于:《第24届国际人工智能会议论文集》, IJCAI’15, AAAI出版社, 2015年, 第1859-1865页.ISBN978-1-57735-7384.http://dl.acm.org/citation.cfm?id=2832415.2832507

[29] K.-W.Chang, W.-t.Yih, B.Yang和C.Meek, “用于关系提取的知识库的类型张量分解”, 载于:2014年自然语言处理中的经验方法会议论文集(EMNLP), 计算语言学协会, 多哈, 卡塔尔, 2014年, 第1568-1579页. doi:10.3115/v1/D14-1165.https://www.aclweb.org/anthology/D14-1165

[30] Z.Hu, P.Huang, Y.Deng, Y.Gao和E.Xing, 实体层次结构嵌入, 载于:计算语言学协会第53届年会和第七届国际自然语言联合会议论文集处理(第1卷:长论文), 计算语言学协会, 北京, 中国, 2015年, 第1292-1300 页.doi:10.3115/v1/P15-1125.https://www.aclweb.org/anthology/P15-1125

[31] Y.Lin, Z.Liu和M.Sun, “知识库表示学习的关系路径建模”, EMNLP, 2015年

[32] K.Guu, J.Miller和P.Liang, 在向量空间中遍历知识图谱, 载于:《2015年自然语言处理经验方法会议论文集》, 计算语言学协会, 里斯本, 葡萄牙, 2015年, 第11页.318–327.doi:10.18653/v1/D15-1038 https://www.aclweb.org/anthology/D15-1038

[33] A.García-Durán, A.Bordes和N.Usunier, 《与翻译的关系》, 载于:《2015年自然语言处理经验方法会议论文集》, 计算语言学协会, 葡萄牙里斯本, 2015年, 第11页.286–290.doi:10.18653/v1/D15-1034 https://www.aclweb.org/anthology/D15-1034

[34] A.Neelakantan, B.Roth和A.Mccallum, 知识库完成的成分向量空间模型1(2015).doi:10.3115 /v1/P15-1016

[35] R.Das, A.Neelakantan, D.Belanger和A.Mccallum, 使用递归神经网络的实体, 关系和文本推理链, 2017年, 第132–141.doi:10.18653/v1/E17-1013

[36] Y.Luo, Q.Wang, B.Wang和L.Guo, 上下文相关的知识图嵌入, 载于:《自然语言处理经验方法的会议纪要》, 计算语言学协会, 里斯本, 葡萄牙, 2015年, 第1656–1661页.doi:10.18653/v1/D15-1191.https://www.aclweb.org/anthology/D15-1191

[37] Z.Wei, J.Zhao, K.Liu, Z.Qi, Z.Sun和G.Tian, 大型知识库完成:通过对选定实例进行接地网络采样推断, 载于:CIKM, 2015年

[38] S.Guo, Q.Wang, L.Wang, B.Wang和L.Guo, 联合嵌入知识图和逻辑规则, 2016年, 第192–202.doi:10.18653/v1/D16-1019

[39] T.Rocktäschel, S.Singh和S.Riedel, 将逻辑背景知识注入嵌入关系中以进行关系提取, 载于:计算语言学协会北美分会2015年会议论文集:人类语言技术, 计算语言学, 科罗拉多州丹佛, 2015年, 第1119-1129 页.doi:10.3115/v1/N15-1118.https://www.aclweb.org/anthology/N15-1118

[40] T.Jiang, T.Liu, T.Ge, L.Sha, S.Li, B.Chang和Z.Sui, 编码时间信息以进行时间感知的链接预测, 载于:2016年经验会议论文集《自然语言处理中的方法》, 计算语言学协会, 德克萨斯州奥斯汀, 2016年, 第 2350–2354.doi:10.18653/v1/D16-1260.https://www.aclweb.org/anthology/D16-1260

[41] C.Esteban, V.Tresp, Y.Yang, S.Baier和D.KrompaÃ§, 预测事件和知识图的共同演化, 载于:2016 年第19届国际信息融合会议(FUSION), 2016年, 第98–105页

[42] R.Trivedi, H.Dai, Y.Wang和L.Song, 《知识进化:动态知识图的深度时间推理》, 载于:《第34届机器学习国际会议论文集》, 第70卷, ICML'17, JMLR.org, 2017年, 第3462-3471页.http://dl.acm.org/citation.cfm?id = 3305890.3306039

[43] J.Feng, M.Huang, Y.Yang和X.Zhu, GAKE:Graph Aware Knowledge Embedding, 载于:COLING 2016会议录中, 第26届国际计算语言学会议:技术论文, COLING2016组织委员会, 日本大阪, 2016年, 第641-651页.https://www.aclweb.org/anthology/C16-1062

[44] X.Jiang, V.Tresp, Y.Huang和M.Nickel, 使用加性模型的多关系图中的链接预测, 载于:2012年语义技术国际会议论文集与推荐系统和大数据量919会合, SeRSy'12, CEUR-WS.org, 德国亚琛, 德国, 2012年, 第1–12页.http://dl.acm.org/citation.cfm?id=2887638.2887639

[45] A.Kristiadi, M.A.Khan, D.Lukovnikov, J.Lehmann和A.Fischer, 《将文字融入知识图嵌入》, 载于:ISWC2019, 2019

[46] Y.Wu和Z.Wang, 知识图嵌入实体的数字属性, 载于:Rep4NLP@ACL, 2018年

[47] A.García-Durán和M.Niepert, KBlrn:具有潜在关系和数值特征的知识库表示的端到端学习, UAI, 2018年

[48] Y.Tay, A.T.Luu, M.C.Phan和S.C.Hui, 用于知识图中非离散属性预测的多任务神经网络, CoRR(2017 年)

[49] P.Pezeshkpour, L.Chen和S.Singh, 《嵌入多峰关系数据以完成知识库》, 载于:《2018年自然语言处理经验方法会议》(EMNLP)会议录, 2018年, 第3208–3218页

[50] M.Chen, Y.Tian, K.-W.Chang, S.Skiena和C.Zaniolo, 知识图和跨语言实体对齐的实体描述的共同训练嵌入, arXiv预印本arXiv:1806.06478(2018)

[51] R.Xie, Z.Liu, T.-S.Chua, H.-B.Luan和M.Sun, 图像体现的知识表示学习, 载于:IJCAI, 2017

[52] P.Goyal和E.Ferrara, “图形嵌入技术, 应用程序和性能:调查,知识”, 基于系统.(2018)

[53] H.Cai, V.W.Zheng和K.C.-C.Chang,《图嵌入的综合调查:问题,技术和应用》,TKDE(2018)

[54] Q.Wang, Z.Mao, B.Wang和L.Guo, 知识图嵌入:方法与应用概述, TKDE(2017)

[55] Y.Lin, Z.Liu和M.Sun, 带有实体, 属性和关系的知识表示学习, 种族(2016)

[56] H.Paulheim, 知识图细化:方法和评估方法概述, 语义网8(3)(2017), 489-508.doi:10.3233/SW-160218

[57] GAGesese, R.Biswas和H.Sack,《对具有文字的知识图嵌入的全面调查:技术与应用》, 来自:与第16届扩展语义网络会议同期举办的知识图深度学习研讨会(DL4KG2019)的论文集 2019(ESWC 2019), 斯洛文尼亚 Portoroz, 2019年6月2日, 2019, 第31–40页.http://ceur-ws.org/Vol-2377/paper\_4.pdf

[58] M.Cochez, M.Garofalo, J.Lenßen和M.A.Pellegrino, 《将文本文字包含在KGloVe中的第一个实验》, arXiv预印本arXiv:1807.11761(2018)

[59] M.Cochez, P.Ristoski, S.P.Ponzetto和H.Paulheim, 全球 RDF 矢量空间嵌入, 载于:国际语义网会议, 施普林格, 2017年

[60] J.Pennington, R.Socher和C.Manning,《手套：用于单词表示的全球载体》, 载于:EMNLP, 2014年

[61] L.Galárraga, C.Teflioudi, K.Hose和F.M.Suchanek, 使用AMIE+, VLDB(2015)在本体知识库中进行快速规则挖掘

[62] G.de Assis Costa和J.M.P.de Oliveira, "致力于探索文字以丰富知识图中的数据链接", 载于:AIKE, 2018年

[63] G.de Assis Costa和J.M.P.de Oliveira，语义网中实体解析的阻塞方案，载于:AINA, 2016

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

[65] B.D.Tresedya, J.Qi和R.ZhangâLU, 使用属性嵌入的知识图之间的实体对齐, 载于AAAI, 2019年

[66] K.Simonyan和A.Zisserman, 用于大规模图像识别的超深度卷积网络, 见:国际学习表示会议(ICLR), 2015年

[67] D.Liben-nowell和J.Kleinberg,《社交网络的链接预测问题》, 美国信息科学技术学会学报 58(2003).doi:10.1002/asi.20591

[68] L.Lü和T.Zhou,《复杂网络中的链接预测:一项调查》,《物理学》A：统计力学及其应用390(6)(2011), 1150–1170年

[69] FMSuchanek, G.Kasneci和G.Weikum, Yago:语义知识的核心, 载于:第16届国际互联网会议论文集, WWW'07, ACM, 美国纽约, 2007年, 第697-706页.ISBN 978-1-59593-654-7.doi:10.1145/1242572.1242667

[70]K.Toutanova和D.Chen, 知识库和文本推理的观察与潜在特征, 载于:连续向量空间模型及其组成的第三届研讨会论文集, 计算语言学协会, 北京, 中国, 2015年, 第57–66页.doi:10.18653/v1/W15-4007.https://www.aclweb.org/anthology/W15-4007

# 附录 A. 应用摘要

表 9：带文字的 KG 嵌入模型的应用场景摘要。

