

用于简单链接预测的复杂嵌入

西奥·特鲁永^{1,2}
约翰内斯·韦尔布卢³
塞巴斯蒂安·里德尔³
艾瑞克·高西尔²
纪尧姆·布沙尔³

THEO.TROUILLON@XRCE.XEROX.COM
J.WELBL@CS.UCL.AC.UK
S.RIEDEL@CS.UCL.AC.UK
ERIC.GAUSSIER@IMAG.FR
G.BOUCHARD@CS.UCL.AC.UK

¹ 施乐欧洲研究中心, 6 chemin de Maupertuis, 38240 Meylan, FRANCE
² 格勒诺布尔阿尔卑斯大学, 621 avenue Centrale, 38400 圣马丁德赫雷斯, 法国
³ 伦敦大学学院, Gower St, 伦敦 WC1E 6BT, 英国

抽象的

在统计关系学习中,链接预测问题是自动理解的关键
大型知识库的结构。与以前的研究一样,我们提出解决这个问题
通过潜在因子分解。然而,在这里我们
利用复值嵌入。
复杂嵌入的组合可以处理
各种各样的二元关系,其中包括
对称和反对称关系。与最先进的模型(如神经
张量网络和全息嵌入,
我们基于复杂嵌入的方法是
可以说更简单,因为它只使用 Hermitian
点积,实向量之间的标准点积的复数对应项。我们的方法可扩展
到大型数据集,因为它仍然
在空间和时间内都是线性的,同时始终
在标准链接预测基准上优于其他方法。¹

1. 简介

Web 规模的知识库 (KB) 提供了结构化的
世界知识的表示,包括以下项目
DBPedia (Auer 等人, 2007)、Freebase (Bollacker 等人,
2008)或 Google Knowledge Vault (Dong 等, 2014)。
它们可以实现多种应用,例如推荐系统、问答系统或自动个人

这些知识库的不完整性刺激了

¹Code 目前正在接受审批,即将推出
可以在: <https://github.com/ttrouill/complex>

第 33 届国际机器视觉大会论文集
Learning,美国纽约州纽约,2016 年。JMLR:W&CP 卷
48.版权归作者所有,2016 年。

研究预测缺失条目,这项任务被称为
链接预测是统计关系学习 (SRL、Getoor 和 Taskar, 2007)中的主要问
题之一。
知识库将数据表达为带有标记边的有向图
节点 (实体)之间的 (关系)。记录的关系中的自然冗余通常使得能够填补
知识库中缺失的条目。例如,关系 CountryOfBirth 没有记录
所有实体,但如果关系
CityOfBirth 是已知的。链接预测的目标
是自动发现这些规律的过程。然而,许多关系是不确定的:两个事实的组合
IsBornIn(John,Athens) 和
IsLocatedIn(Athens,Greece) 并不总是
暗示事实 HasNationality(John,Greece)。
因此,需要处理涉及这些的其他事实
以概率的方式描述关系或实体。
为了做到这一点,一种越来越流行的方法是声明
将链接预测任务视为 3D 二元张量补全问题,其中每个切片是知识图谱中一个
关系类型的邻接矩阵。基于

低秩分解或嵌入已经流行起来
与 Netflix 挑战赛 (Koren 等人, 2009)进行了部分
观察到的矩阵或张量分解为乘积
嵌入矩阵的秩要小得多,结果
在每个实体的固定维度向量表示中
和数据库中的关系。对于给定的事实 $r(s,o)$,其中
主语 s 通过关系 r 链接到宾语 o ,得分
然后可以恢复为
 s, r 和 o 的嵌入向量 (Nickel 等, 2016a)。
知识库中的二元关系表现出各种类型的模式:层次结构和组合,如 FatherOf、
OlderThan 或 IsPartOf部分/全部,
严格/非严格顺序 以及等价关系
就像 IsSimilarTo。正如Bordes 等人所述。
(2013a),关系模型应该 (a)能够学习

这些属性的所有组合,即反身性/非反身性、对称性/反对称性和传递性,

(b)在时间和内存上都是线性的,以便于扩展达到当今知识库的大小,并跟上其生长。

嵌入的点积具有很好的可扩展性,并且可以自然地处理关系的对称性和 (非)反身性;使用适当的损失函数甚至可以实现传递性 (Bouchard 等人, 2015 年)。然而,到目前为止,处理反对称关系几乎总是意味着

参数数量激增 (Nickel 等, 2011; Socher 等人, 2013 年) (见表1),使得模型容易过度拟合。找到表达能力和参数空间大小之间的最佳比例是嵌入的关键

模型。

在本文中,我们认为嵌入可以是一种非常有效的组合函数,只要使用正确的表示。相反使用包含实数的嵌入,我们讨论并展示复杂嵌入的能力。当使用复杂向量 (即带有 C 语言条目的向量)时,该点积通常称为 Hermitian (或 sesquilinear)点积,因为它涉及

两个向量中的一个。因此,点积不再对称,关于反对称的事实关系可以根据所涉及实体的顺序获得不同的分数。因此,复杂向量可以

有效地捕捉反对称关系,同时保留点积的效率优势,即线性空间和时间复杂度。

本文的其余部分安排如下。我们首先证明在方阵的情况,其中实体之间只有一个关系。然后公式被扩展为

一组三阶张量的堆叠方阵表示多个关系。然后我们描述实验在大规模公共基准知识库中,我们通过经验表明,这种表示不仅可以简化

和更快的算法,而且还提供了系统的准确性比目前最先进的替代方案有所改进。

为了与仅使用实数的现有方法进行清晰的比较,我们还提出了一个

我们的模型的等效重构仅涉及真实嵌入。这应该可以帮助实践者实现我们的方法,而无需在软件实现中使用复数。

2. 低秩实部关系
正态矩阵

在本节中,我们讨论使用复数嵌入进行低秩矩阵分解,并说明这一点

通过考虑简化的链接预测任务,仅单一关系类型。

理解复杂空间中的因式分解可以得到更好地从理论上理解矩阵类实际上可以通过嵌入的点积来近似。这些就是所谓的正态矩阵,

左嵌入和右嵌入共享相同的幺正基。

2.1. 建模关系

假设 E 是一组满足 $|E| = n$ 的实体。两个实体表示为二进制值 $Y_{so} \in \{-1, 1\}$, 其中 $s \in E$ 是关系的主语, $o \in E$ 是其对象。其概率由逻辑逆向链接给出功能:

$$P(Y_{so} = 1) = \sigma(X_{so}) \tag{1}$$

其中 $X \in \mathbb{R}^{n \times n}$ 是分数的潜在矩阵, Y 是部分观察到的符号矩阵。

我们的目标是找到 X 的通用结构,从而现实世界中常见关系的灵活近似标准矩阵分解将 X 近似为矩阵乘积 UV 其中 U 和 V 是两个函数独立的 $n \times K$ 矩阵, K 是矩阵的秩。在这个公式中,假设作为主语出现的实体与作为副语出现的实体不同。

对象。这意味着同一个实体将具有两个不同的嵌入向量,具体取决于它是否出现在现在作为关系的主体或客体。这广泛地研究的模型类型与奇异值密切相关分解 (SVD)并且非常适合矩阵 X 是矩形的。然而,在许多链接预测中问题,同一个实体可以同时作为主体和对象。那么学习实体,这需要共享左右因素,正如几位作者所提出的解决链接预测问题 (Nickel et al., 2011; Bor-des et al., 2013b; Yang et al., 2015) 。

为了对主体和客体使用相同的嵌入,研究人员将点积的概念推广到评分函数,也称为组合函数,以特定方式组合嵌入。我们

简要回顾表1 中评分函数的几个示例,以及本文提出的扩展。

对左右因子使用相同的嵌入会导致进行特征值分解:

$$X = EWE^T - I \tag{2}$$

它通常用于近似实对称矩阵例如协方差矩阵、核函数和距离或相似矩阵。在这些情况下,所有特征值和特征向量存在于实空间中,并且 E 是正交的:

用于简单链接预测的复杂嵌入

模型	评分功能	关系参数	Otime		欧斯派斯
RESICAL (Nickel 等人, 2011)	$e_s^T w_o$ 韦奥	$w_r \in \mathbb{R}^{K^2}$		(K^2)	(K^2)
TransE (Bordes 等人, 2013b)	$\ (e_s + w_r) - e_o\ _p$	$w_r \in \mathbb{R}^K$		好的)	好的)
NTN (Socher 等人, 2013 年)	$\frac{1}{r} (e_s^T w_r)^{[1..D]} + br$	$w_r \in \mathbb{R}^{K^2 D}, br \in \mathbb{R}^K$		$O(K^2 D)$	$O(K^2 D)$
DistMult (Yang 等人, 2015)	$e_s^T e_o$	$w_r \in \mathbb{R}^K$		好的)	好的)
(Nickel 等人, 2016b)	$\frac{1}{r} (F^T [e_s] F [e_o]))$	$K w_r \in \mathbb{R}^K$		$O(K \log K)$	$O(K)$
复杂的	$\text{Re}(\langle w_r, e_s, e_o \rangle)$	$w_r \in \mathbb{C}^K$		好的)	好的)

表 1. 给定事实 $r(s, o)$ 的最先进的潜在因子模型的评分函数,以及它们的关系参数、时间和空间 (内存)复杂度。主语 s 和宾语 o 的嵌入 e_s 和 e_o 在 \mathbb{R}^K 中 (ComplEx)其中 $e_s, e_o \in \mathbb{C}^{K \times D}$ 是 NTN 模型的附加潜在维度。 F 和 F^{-1} 分别表示傅里叶变换及其逆,是两个向量之间逐元素的乘积。

然而,在这项工作中,我们明确感兴趣的是矩阵的问题。因此,它们表示也可以是反对称的。在这种情况下,特征值分解在实空间中是不可能的;只存在复杂空间中的分解。其中嵌入 $x \in \mathbb{C}^K$ 由实向量分量 $\text{Re}(x)$ 和虚向量分量组成 $\text{Im}(x)$ 。对于复数,点积也称为埃尔米特积或半线性形式定义为:

$$u, v := \overline{u}^T v$$

(3)

$$X = \text{Re}(EWE^T)。$$

(5)

其中 u 和 v 是复向量,即 $u = \text{Re}(u) + i\text{Im}(u)$ 其中 $\text{Re}(u) \in \mathbb{R}^K, \text{Im}(u) \in \mathbb{R}^K$ 响应向量的实部和虚部。 i 表示 -1 的平方根。我们在这里看到一个关键的运算是取第一个向量 $u^T = \text{Re}(u) - i\text{Im}(u)$ 。一个简单的方法来证明厄米积用于合成复数向量的原因是:在诱导向量中提供了有效的拓扑范数空间。例如, $x^T x = 0$ 意味着 $x = 0$,而这对于双线性形式 x, x^T 不是这种情况。对 x 因为有很多复向量,其中 $x^T x = 0$ 。即使具有复特征向量 $E \in \mathbb{C}^{n \times n}$,逆在方程(2)的特征分解中, E 可得出计算问题。幸运的是,数学家定义了一个适当的矩阵类,可以防止我们反转特征向量矩阵:我们考虑空间

正规矩阵,即复数 $n \times n$ 矩阵 X ,例如 $XX^T = X^T X$ 。正规矩阵的谱定理指出,矩阵 X 是正规矩阵当且仅当它满足

酉对角化:

$$X = EWE^T$$

(4)

其中 $W \in \mathbb{C}^{n \times n}$ 是特征值的对角矩阵 (模量减小)且 $E \in \mathbb{C}^{n \times n}$ 是特征向量的酉矩阵,其中 E^T 表示其复共轭。

纯实正规矩阵集包括所有对称和反对称符号矩阵 (用于建模

层次关系 (例如 IsOlder)以及所有正交矩阵 (包括置换矩阵),以及许多其他矩阵也可用于表示二元关系,例如表示二分图的分配矩阵。然而,并非所有矩阵都表示为

EWE^T 是纯实数,而公式1要求分数 X 是纯实数。所以我们只保留实数部分分解:

$$X = \text{Re}(EWE^T)。$$

(5)

事实上,在实空间上执行这种投影可以精确分解任何实数方阵 X

而不仅仅是正常的,正如Trouillon 等人所展示的那样。(2016)。

与奇异值分解相比,特征值分解有两个主要区别:

- 特征值不一定是正数或者实数;
 - 因式分解(5)很有用,因为 E 的行可以用作关系矩阵的行和列对应的实体的向量表示
- X . 事实上,对于给定的实体,它的主题嵌入向量是其对象嵌入向量的复共轭。

2.2 低秩分解

在链接预测问题中,关系矩阵未知。目标是从嘈杂的观察中完全恢复它。为了使模型可学习,即推广到未观察到的链接,需要进行一些规律性假设

需要。由于我们处理二元关系,我们假设它们的符号秩较低。符号矩阵的符号秩是具有相同符号秩的实矩阵的最小秩

符号模式为 Y :

$$\text{秩}_\pm(Y) = \min_{A \in \mathbb{R}^{m \times n}} \{ \text{等级}(A) \mid \text{符号}(A) = Y \}。$$

(6)

这在理论上是合理的,因为符号等级是符号矩阵的自然复杂性度量 (Linial 等人, 2007),并且与可学习性相关 (Alon 等人, 2015),并且因式分解模型的广泛成功也从经验上证实了这一点 (Nickel 等人, 2016a)。

如果观测矩阵 Y 是低符号秩的,那么我们的模型可以将其分解为最多为 Y 符号秩两倍的秩,也就是说,对于任何 $Y \in \{-1, 1\}^{n \times n}$,总存在一个矩阵 $X = \text{Re}(EWE^T)$,其中 EWE^T 的秩在符号模式 $\text{sign}(X) = Y$ 最多为 Y 符号秩的两倍(Trouillon et al., 2016)。

虽然两倍听起来很糟糕,但这实际上是一个很好的上限。事实上,符号秩通常比 Y 的秩低得多。例如, $n \times n$ 单位矩阵 I 的秩为 n ,但秩 $\pm(1) = 3$ (Alon 等人, 2015 年)。通过对列 $2j$ 和 $2j + 1$ 进行排列, I 矩阵对应于关系 marriageTo,这是一种已知难以分解的关系 (Nickel 等人, 2014 年)。然而,对于任何 n ,我们的模型都可以用秩 6 来表示它。

通过在 EWE^T 上施加低秩 $K \times n$,只有 $\text{diag}(W)$ 的前 K 个值非零,所以我们可以直接得到 $n \times K$ 和 $W \in \mathbb{C}^{K \times K}$,实体 s 和 o 之间的个体关系得分 $E \in \mathbb{C}^{X_{so}}$ 可以通过它们的嵌入 $e_s, e_o \in \mathbb{C}^K$ 的以下乘积来预测:

$$X_{so} = \text{Re}(e_s^H e_o)$$

我们 (7)

我们将上述讨论总结为三点:

1. 我们的分解涵盖了所有可能的二进制重新关系。

2. 通过构造,它准确地描述对称和反对称关系。

3. 可学习关系可以通过简单的低秩分解有效地近似,使用复数来表示潜在因子。

3. 应用于二元多关系数据

上一节重点介绍了对单一类型的关系进行建模;现在我们将此模型扩展到多种类型的关系。我们通过为每个关系 r 分配一个嵌入 w_r ,并在所有关系之间共享实体嵌入来实现这一点。

令 R 和 E 为知识库中存在的关系和实体的集合。我们想要恢复所有关系 $r \in R$ 的得分矩阵 X_r 。给定两个实体 s 和 $o \in E$,事实 $r(s,o)$ 为真的概率的对数奇数为:

$$P(Y_{rso} = 1) = \sigma(\phi(r, s, o; \theta))$$

(8)

其中 ϕ 是评分函数,通常基于观察到的关系的因式分解, θ 表示相应模型的参数。虽然 X 整体未知,但我们假设观察到一组

真假事实 $\{Y_{rso} | r(s,o) \in \Omega \in \{-1, 1\}^{\Omega}\}$ 对应于不同关系的部分可观察到,对应的邻接矩阵,其中 $\Omega \subset R \otimes E \otimes E$ 是可观察到的三元组的集合。目标是找到对于一组目标未观察到的三元组 $r(s, o) \in \Omega$,条目 Y_{rso} 为真或假的概率。

根据用于预测张量 X 的条目的评分函数 $\phi(s, r, o; \theta)$,我们获得不同的模型。评分函数的示例如表1所示。

我们的模型评分函数是:

$$\phi(r, s, o; \theta) = \text{Re}(\langle w_r, e_s, e_o^H \rangle)$$

(9)

$$= \text{Re} \left(\sum_{k=1}^K w_{rk} e_{sk} e_{ok}^H \right)$$

(10)

$$\begin{aligned} &= \text{Re}(w_r), \text{Re}(e_s), \text{Re}(e_o) \\ &\quad + \text{Re}(w_r), \text{Im}(e_s), \text{Im}(e_o) \\ &\quad + \text{Im}(w_r), \text{Re}(e_s), \text{Im}(e_o) \\ &\quad - \text{Im}(w_r), \text{Im}(e_s), \text{Re}(e_o) \end{aligned}$$

(11)

其中 $w_r \in \mathbb{C}^K$ 是一个复数向量。这些方程提供了该模型的两个有趣的观点:

- 改变表示形式:等式(10)对应于具有实数嵌入的 DistMult,但由于其中一个嵌入 2 的复共轭,可以处理不对称问题。

· 改变评分函数:等式 (11)仅涉及与嵌入和关系的实部和虚部相对应的实向量。

可以很容易地检查出,当 w_r 为纯虚数 (即其实部为零)时,该函数是反对称的,而当 w_r 为实数时,该函数是对称的。有趣的是,通过分离关系嵌入 w_r 的实部和虚部,我们得到关系矩阵 X_r 的分解,即对称矩阵 $\text{Re}(E \text{diag}(\text{Re}(w_r))E^T)$ 和反对称矩阵 $\text{Im}(E \text{diag}(-\text{Im}(w_r))E^T)$ 之和。关系嵌入自然地充当每个潜在维度的权重: $\text{Re}(w_r)$ 在 e_o, e_s 的对称实部上, $\text{Im}(w_r)$ 在 e_o, e_s 的反对称虚部上。实际上,有 $e_o, e_s = e_s, e_o$,这意味着 $\text{Re}(e_o, e_s)$ 是对称的,而 $\text{Im}(e_o, e_s)$ 是反对称的。这使我们能够准确地描述两者

2请注意,在公式(10)中,我们使用了标准分量式多线性点积 $\langle a, b, c \rangle := \text{akbkc}$ 。这不是 Hermitian 扩展,因为它在线性代数文献中没有正确定义。

实体对之间的对称和反对称关系,同时仍然使用实体的联合表示,无论它们是作为关系的主体还是客体出现。

从几何角度来看,每个关系嵌入 w_r 都是实体嵌入 E 定义的基础的各向异性缩放,然后投影到实子空间。

4.实验

为了评估我们的提案,我们在合成数据集和真实数据集上进行了实验。合成数据集基于对称或反对称的关系,而真实数据集包含不同标准知识库中发现的不同类型的关系。我们将我们的模型称为 ComplEx,即复杂嵌入。

4.1. 合成任务

为了评估我们的提案准确建模对称和反对称的能力,我们随机生成了一个包含两个关系和 30 个实体的知识库。一个关系完全对称,而另一个关系完全反对称。

该数据集对应于 $2 \times 30 \times 30$ 张量。图2显示了该随机生成的张量的一部分,具有对称切片和反对称切片,分解为训练集、验证集和测试集。对角线未观察到,因为它与本实验无关。

训练集包含 1392 个观察到的三元组,而验证集和测试集各包含 174 个三元组。图1显示了不同分解模型 (等级最高为 50) 的最佳交叉验证平均精度 (精度-召回率曲线下面积)。使用随机梯度下降法和小批量训练模型,并使用 AdaGrad 调整学习率 (Duchi 等人, 2011 年),通过最小化逻辑正则化对参数 Θ 的负对数似然

考虑 L 的模型: 2

$$\min_{r(s,o) \in \Omega} \log(1 + \exp(-Yr\phi(s, r, o; \theta))) + \lambda ||\theta||_2^2. \tag{12}$$

在我们的模型中, Θ 对应于嵌入 $e_s, w_r, e_o \in \mathbb{C}^K$ 。我们在附录A中描述了完整的算法。

λ 在 $\{0.1, 0.03, 0.01, 0.003, 0.001, 0.0003, 0.0001, 0.0\}$ 中得到验证。正如预期的那样,DistMult (Yang et al., 2015) 无法对反对称性进行建模,只能正确预测对称关系。尽管 TransE (Bordes et al., 2013b) 不是对称模型,但它在实践中表现不佳,特别是在反对称关系上。RESCAL (Nickel et al., 2011) 具有大量参数,随着秩的增加,很快就会过拟合。典型多元 (CP) 分解 (Hitchcock, 1927) 失败

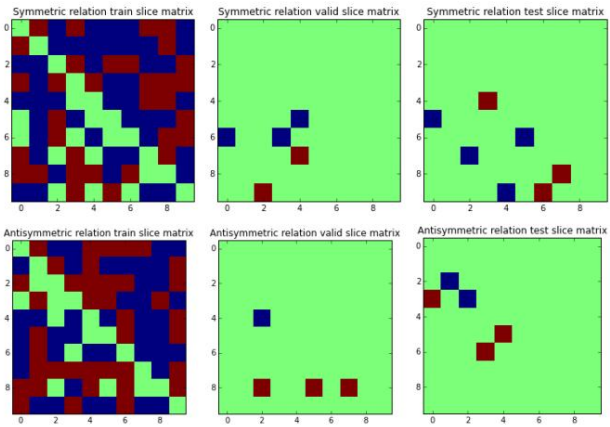


图 2. 生成实验的训练集、验证集和测试集的部分内容,其中包含一个对称关系和一个反对称关系。红色像素表示正三元组,蓝色像素表示负三元组,绿色像素表示缺失三元组。顶部:前 10 个实体的对称切片 (关系)图。底部:前 10 个实体的反对称切片图。

在两种关系上都如此,因为它必须通过实体嵌入来推动对称和反对称模式。令人惊讶的是,只有我们的模型在如此简单的数据上取得成功。

4.2. 数据集:FB15K 和 WN18

数据集 E	$ R $ 训练/验证/测试中的三元组数
WN18	40,943 18 141,442 / 5,000 / 5,000
FB15K	14,951 1,345 483,142 / 50,000 / 59,071

表 3.FB15K 和 WN18 数据集每次分割中的实体、关系和观察到的三元组的数量。

接下来,我们评估模型在 FB15K 和 WN18 数据集上的性能。FB15K 是 Free-base,一个精选的一般事实知识库,而 WN18 是 Wordnet 的一个子集,Wordnet 是一个以单词之间的词汇关系为特色的数据库。我们使用Bordes 等人(2013b) 提供的原始训练、验证和测试集分割。表3总结了这两个数据集的元数据。

两个数据集都只包含正三元组。与Bor-des 等人(2013b)一样,我们使用局部封闭世界假设生成负样本。也就是说,对于三元组,我们随机更改主语或宾语,以形成负样本。在运行时对每一批训练正样本执行此负样本抽样。

为了进行评估,我们测量每个测试三元组在所有可能的主语和宾语替换中的排名质量: $r(s,o)$ 和 $r(o,s)$, $\forall s \forall o \in E$ 。平均互惠排名 (MRR) 和 m 处的命中率是这些数据集的标准评估指标,有两种形式:原始和过滤后(Bordes et al., 2013b)。过滤后的指标

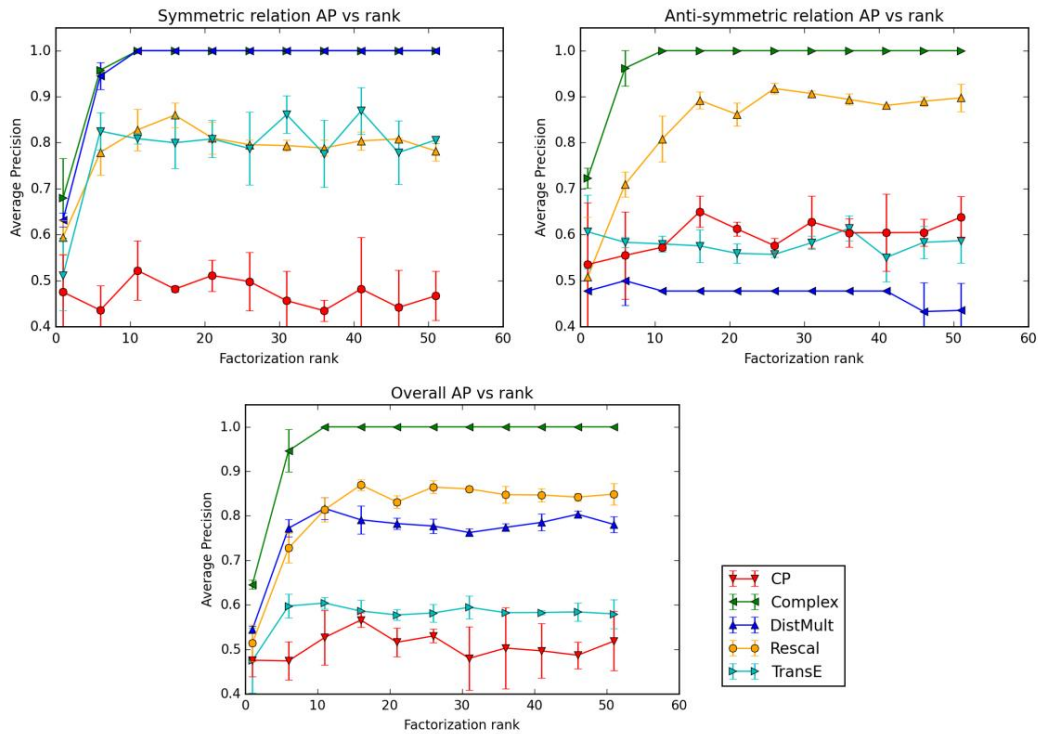


图 1. 在对称和反对称组合实验中,不同最先进模型的每个分解等级 (范围从 1 到 50)的平均精度 (AP)。左上:仅适用于对称关系的 AP。右上:仅适用于反对称关系的 AP。底部:整体 AP。

是在从排名中删除训练、验证或测试集中出现的所有其他正观察三元组后计算的,而原始指标不会删除这些三元组。

由于使用了排名度量,先前的研究通常倾向于使用成对排名损失来完成该任务 (Bordes 等人, 2013b; Nickel 等人, 2016b)。我们选择使用逻辑模型的负对数似然,因为它是符号排名的连续替代,并且已被证明可以学习几个重要关系的紧凑表示,尤其是传递关系 (Bouchard 等人, 2015)。在初步工作中,我们尝试了两种损失,事实上,对数似然比排名损失产生了更好的结果 (TransE 除外),尤其是在 FB15K 上。

我们在表2中报告了评估模型的过滤和原始 MRR,以及过滤后的命中数 1.3 和 10。此外,我们选择 TransE、DistMult 和 HoE 作为基线,因为据我们所知,它们是这些数据上表现最好的模型 (Nickel et al., 2016b; Yang et al., 2015)。我们还与 CP 模型进行了比较,以经验性地强调学习实唯一嵌入的重要性。为了实验公平起见,我们使用 theano (Bergstra et al., 2010)在与 ComplEx 模型相同的框架内重新实现了这些方法。然而,由于

由于时间限制和有效实施 HoE 的复杂性,我们记录了Nickel 等人(2016b) 报告的 HoE 原始结果。

4.3. 结果

WN18 描述了概念之间的词汇和语义层次结构,包含许多反对称关系,例如上位词、下位词或“属于”。事实上,ComplEx 和 HoE 在这方面的表现优于 DistMult 和 TransE 模型,它们的过滤 MRR 得分分别为 0.941 和 0.938。表4显示了所考虑的模型和 WN18 的每个关系的过滤测试集 MRR,证实了我们的模型在反对称关系上的优势,而在其他方面则没有损失。附录B中提供的关系嵌入的 2D 投影在视觉上证实了结果。

在 FB15K 上,差距更加明显,ComplEx 模型的表现远胜于 HoE,其过滤后的 MRR 为 0.692,命中率为 1 的命中率为 59.9%,而 HoE 为 0.524,命中率为 40.2%。我们将此归因于我们模型的简单性和不同的损失函数。与 DistMult (0.654)相比,MRR 的差距相对较小,这支持了这一点;我们的模型实际上可以解释为 DistMult 的复数版本。在这两个数据集上,TransE

用于简单链接预测的复杂嵌入

模型过滤器原始 CP 0.075 0.058 TransE	WN18			FB15K		
	月平均收入		命中数	月平均收入		点击次数
	1	10		1	10	
0.454 0.335 DistMult 0.822 0.532 HoIE*	0.049	3 0.080	0.125	0.152 0.380 0.221	0.219	3 0.532
0.938 0.616 ComplEx 0.941 0.587	0.089	0.823	0.934	0.654 0.242 0.524	0.231	0.376 0.641
	0.728	0.914	0.936	0.232 0.692 0.242	0.546	0.472 0.824
	0.93	0.945	0.949		0.402	0.733 0.739
	0.936	0.945	0.947		0.599	0.613 0.759 0.840

表 2. 在 FB15K 和 WN18 数据集上测试的模型的过滤和原始平均倒数排名 (MRR)。Hits@m 指标为已过滤。*结果来自(Nickel et al., 2016b)针对 HoIE 模型的报告。

关系名称	复杂分布式多业务传输			FB15K 上的过滤 MRR (最高提升 +0.08) 1 到 10 个负面评价) ,但 WN18 上的负面评价并不多。 两个数据集的正则化都很重要 (最高为+0.05 在λ = 0 和最优值之间过滤 MRR) 。我们发现 FB15K 的初始学习率非常重要, 而在 WN18 上则没有那么多。我们认为这也可能 解释我们的模型提供的改进的巨大差距 与之前发布的结果相比,此数据集 – DistMult 结果也比以前更好 报告 (Yang 等人, 2015 年) 以及使用 对数似然目标。似乎一般来说,AdaGrad 对初始学习率相对不敏感,也许 导致对其调整步伐的能力过于自信 尺寸在线,从而减少工作量 选择初始步长。 使用验证集上的早期停止来停止训练,过滤后的 MRR 每 50 个时期计算一次 最多 1000 个时期。
上位词 下位	0.953	0.791	0.446	
词成员 部分	0.946	0.710	0.361	
词成员 完整词实例 上位	0.921	0.704	0.418	
词实例 下位词有部分 部	0.946	0.740	0.465	
分成员 域主题0.930同义	0.965	0.943	0.961	
词集 域主题 成员 域用法	0.945	0.940	0.745	
0.917同义	0.933	0.753	0.426	
词集 域用	0.940	0.867	0.455	
法1.000域成员0.865同义词集 域	0.924	0.914	0.861	
生相关形式0.946类似于1.000 区域	0.919	0.919	0.917	
动词组0.936另见 0.603 –		0.917	0.875	
– – – – –		1.000	1.000	
– – – – –		0.635	0.865	
– – – – –		0.888	0.986	
– – – – –		0.940	0.384	
– – – – –		1.000	0.244	
– – – – –		0.897	0.323	
– – – – –		0.607	0.279	

表 4. 模型的滤波平均倒数秩 (MRR)
在 Wordnet 数据集 (WN18) 的每个关系上进行测试。

和 CP 在很大程度上被抛在后面。这说明了
第一种情况下简单点积的重要性,以及第二种情况下学习唯一实体嵌入的重要性。
由于数量较少,CP 在 WN18 上表现不佳
关系,这放大了这种主体差异。

报告结果给出了最佳的
在验证集上评估的超参数
对于每个模型,经过网格搜索后,得出以下值:K∈{10, 20, 50, 100, 150, 200},

$\lambda \in \{0.1, 0.03, 0.01, 0.003, 0.001, 0.0003, 0.0\}$,
 $\alpha_0 \in \{1.0, 0.5, 0.2, 0.1, 0.05, 0.02, 0.01\}$, $\eta \in \{1, 2, 5, 10\}$
其中 λ 为 L² 正则化参数, α0 初始
学习率 (然后在运行时使用 AdaGrad 进行调整), 以及
η 每次正样本训练产生的负样本数量
三倍。我们也尝试过改变批次大小,但这没有
影响,我们以每 epoch 100 个批次进行处理。最好的
排名一般为 150 或 200,两种情况下的分数均为
对于所有模型来说,总是非常接近。负数
每份阳性样本的样本量也对

4.4 负样本的影响
我们进一步研究了每个正训练样本产生的负样本数量的影响。在之前的实验
中,由于计算限制,
每个训练样本的阴性样本数量 η 经过验证
在可能的数字{1, 2, 5, 10}中。我们想在这里探索增加这些数字是否会导致
更好的结果。为此,我们专注于 FB15K,
从之前的实验中获得最佳验证的λ、K、 α0。然后我们让η在{1、2、5、10、20、50、
100、200}范围内变化。
图3显示了生成的
每积极训练一次消极训练对表现的影响是三倍
我们的模型在 FB15K 上。显然产生更多的负面
改善了结果,过滤后的 MRR 为 0.737,其中 100
负三元组 (和 64.8% 的 Hits@1) ,然后再次减少到 200 个负三元组。该模
型也收敛
使用更少的 epoch,这部分补偿了每个 epoch 的额外训练时间,最多 50 个负
样本。然后
随着负样本数量的增加而线性增长,使得 50 在准确度和训练时间之间取得良
好的平衡。

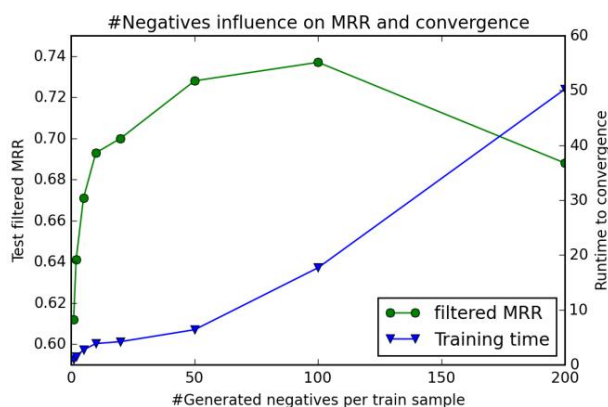


图 3. 每个样本产生的负三元组数量的影响
过滤测试 MRR 上的正训练示例以及 ComplEx 模型在 FB15K 上的收敛训练时间

$K = 200, \lambda = 0.01$ 和 $\alpha = 0.5$ 。时间相对于
每个正样本生成一个负样本的训练时间
训练样本（时间尺度上 = 1）。

5. 相关工作

在线性代数谱理论的早期,复数不用于矩阵分解和

数学家们主要关注双线性形式 (Bel-trami, 1873)。如今在线性代数课程中教授的复数域中的特征分解是在 40 年前提出的。

几年后 (Autonne, 1915)。同样,大多数现有的张量分解方法都是基于实数域中的分解,例如典型

多元 (CP) 分解 (Hitchcock, 1927)。这些

方法在许多应用中非常有效,

不同类型的实体的张量的不同模式。

但在链接预测问题中,关系的反对称性很快被视为一个问题,并研究了张量的不对称扩展,主要是通过考虑独立嵌入 (Sutskever, 2009) 或在 RESCAL 中将关系视为矩阵而不是向量

模型 (Nickel 等, 2011)。直接扩展基于
针对三元组数据的单、双和三元组潜在因子
作为低秩关系矩阵 (Jenatton 等, 2012)。

成对交互模型也被考虑用于提高预测性能。例如,通用

模式方法 (Riedel 等, 2013) 将张量的二维展开 (实体对与关系的矩阵) 分解

而 Welbl 等人 (2016) 也将这一方法扩展到其他对。

在神经张量网络 (NTN) 模型中, Socher 等人。

(2013) 结合线性变换和多个双线性形式的主体和客体嵌入,共同提供

将它们放入非线性神经层。它的非线性和

嵌入之间相互作用的多种方式使其在表达能力上比模型更具优势

使用更简单的评分函数,如 DistMult 或 RESCAL。

缺点是,它的参数数量非常多,
使得 NTN 模型更难训练,并且更容易过度拟合。

原始的多线性 DistMult 模型在

每种关系都有主语和宾语 (Yang et al., 2015), 以及
取得了良好的表现,大概是因为它的简单性。Bordes 等人 (2013b)
的 TransE 模型也将实体和关系嵌入同一空间,并施加了

几何结构偏差进入模型: 主体实体向量应该接近客体实体向量

由关系向量翻译。

处理反对称性的最新新方法是

全息嵌入 (HolE) 模型 (Nickel 等人, 2016b)。在 HolE 中,循环相关用于组合实体嵌入,测量实体之间的协方差

不同维度的嵌入。这通常

这表明除了经典张量积之外的其他复合函数可能会有所帮助,因为它们允许更丰富的

嵌入的相互作用。然而,

HolE 中的合成函数源于不对称性

循环相关的运算为 $O(n \log(n))$, 而

我们的是从复数内积继承而来,复杂度为 $O(n)$ 。

6. 结论

我们描述了一种简单的矩阵和张量分解方法,该方法使用具有

复数值并保留数学定义

点积。正规矩阵类是自然的

适合二元关系,使用实部可以有效地近似任何可学习的关系。结果

标准基准测试表明,无需进行更多修改

需要改进最先进的技术。

这项工作可以扩展几个方向。一个明显的方法是将我们的方法与

已知的张量分解扩展,以进一步提高预测性能。例如,使用

成对嵌入和复数

在许多涉及非组合性的情况下可能会带来更好的结果。另一个方向是

开发更智能的负采样程序,

生成关于

阳性样本,即样本取自该样本。

会减少达到

良好的性能,从而加快训练时间。

另外,如果我们每次都使用复杂的嵌入
模型包括点积,例如在深度神经网络中,
它会带来类似的系统性改进吗?

致谢

这项工作部分由保罗艾伦基金会通过艾伦杰出研究员资助,以及

部分得益于 Google 重点研究奖。

参考

Alon,Noga,Moran,Shay 和 Yehudayoff,Amir。
符号秩与 vc 维度。arXiv 预印本
arXiv:1503.07648,2015年。

奥尔·斯伦、比泽尔、克里斯蒂安·科比拉罗夫、格奥尔基·莱曼、
Jens 和 Ives,Zachary。Dbpedia:网络的核心
开放数据。在第六届国际语义网大会上,
韩国釜山,第 11-15 页。施普林格,2007。

Autonne, L. 关于下层矩阵和关于它们
单位矩阵。安·大学。里昂,Nouvelle Srie I,38:
1-77,1915年。

贝尔特拉米、尤金尼奥。Sulle funzioni 双线性。吉奥纳莱迪
Matematiche ad Uso degli Studenti Delle Universita, 11
(2):98-106, 1873年。

Bergstra,James,Breuleux,Olivier,Bastien,Fred´eric,Lamblin、
Pascal,Pascanu,Razvan,Desjardins,Guil-laume,Turian、
Joseph,Warde-Farley,David 和 Ben-gio,Yoshua。Theano:CPU
和 GPU 数学表达式
编译器。在 Python 科学计算
计算会议 (SciPy) ,2010 年 6 月。口头报告。

博拉克、库尔特、埃文斯、科林、帕里托什、普拉文、斯特奇、
Tim 和 Taylor,Jamie。Freebase:一个协作创建的用于构建人类知识的
图形数据库。
在 2008 年 ACM SIGMOD 的 SIGMOD 08 论文集
国际数据管理会议,第页
1247-1250,2008年。

博德斯、安托万、乌苏尼尔、尼古拉斯、加西亚-杜兰、阿尔贝托、
Weston, Jason 和 Yakhnenko, Oksana。非反身和
层次关系作为翻译。载于 CoRR,2013a。

博德斯、安托万、乌苏尼尔、尼古拉斯、加西亚-杜兰、阿尔贝托、
Weston,Jason,Yakhnenko,Oksana。翻译
用于建模多关系数据的嵌入。《神经信息处理系统进展》,第 154-165
页。
2787-2795,2013b。

Bouchard,Guillaume,Singh,Sameer 和 Trouillon,Theo。
关于低秩向量空间的近似推理能力。在 AAAI 春季知识表示和推理研讨会
(KRR) 中:集成符号
和神经方法,2015 年。

董鑫、加布里洛维奇、叶夫根尼、海茨、杰里米、霍恩、
威尔科、劳、倪、墨菲、凯文、斯特罗曼、托马斯、

孙少华和张伟。知识库:
网络规模的概率知识融合方法。
在第 20 届 ACM SIGKDD 国际会议论文集上
知识发现和数据挖掘会议,
KDD 14,第 601-610 页,2014 年。

Duchi, John, Hazan, Elad 和 Singer, Yoram。自适应
在线学习和随机的次梯度方法
优化。机器学习研究杂志,12:2121-2159,2011。

Getoor, Lise 和 Taskar, Ben. 统计关系学习简介 (自适应计算和
机器学习) 。麻省理工学院出版社,2007 年。ISBN
0262072882。

Hitchcock, FL 张量或多元向量的表达式为
乘积之和。J. Math. Phys,6(1):164-189, 1927。

杰纳纳、鲁道夫、博德斯、安托万、勒鲁、尼古拉斯、
和 Obozinski, Guillaume。潜在因子模型
高度多关系数据。《神经信息处理系统进展》第 25 卷,第 3167-3175 页,
2012 年。

Koren, Yehuda, Bell, Robert 和 Volinsky, Chris。推荐系统的矩阵分解技
术。
计算机,42(8):30-37,2009。

Linial,Nati,Mendelson,Shahar,Schechtman,Gideon 和
Shraibman, Adi。符号矩阵的复杂性度量。
组合,27(4):439-463,2007。

Nickel, Maximilian,Tresp, Volker 和 Kriegel, Hans-Peter。《三向集体
学习模型》
多关系数据。第 28 届国际会议
关于机器学习,第 809816 页,2011 年。

Nickel, Maximilian,Jiang, Xueyan 和 Tresp, Volker。通过纳入可观察模
式来降低关系分解模型中的秩。《神经信息处理系统进展》,第
1179-1187 页,2014 年。

尼克尔、马克西米利安、墨菲、凯文、特雷斯普、沃尔克和
Gabrilovich, Evgeniy. 关系机器的回顾
知识图谱的学习。IEEE 论文集,
104(1):11-33, 2016a。

Nickel,Maximilian,Rosasco,Lorenzo 和 Poggio,
Tomaso A. 知识的全息嵌入
图表。在第三十届 AAAI 会议论文集
关于人工智能,第 1955-1961 页,2016b。

Riedel,Sebastian,Yao,Limin,McCallum,Andrew 和
Marlin,Benjamin M. 使用矩阵进行关系提取
因式分解和通用模式。在人类语言技术:北美会议

计算语言学协会分会,会议录,第 74-84 页,2013 年。

Socher,Richard,Chen,Danqi,Manning,Christopher D,

以及 Andrew Ng。利用神经张量网络进行推理
用于知识库补全。在神经科学进展
信息处理系统,第 926-934 页,2013 年。

Sutskever, Ilya。使用贝叶斯建模关系数据
聚类张量分解。在神经网络进展中

信息处理系统,第 22 卷,第 1-8 页,
2009年。

Trouillon, Theo,Dance,Christopher R.,Gaussier,Eric 和 Bouchard,Guillaume。

通过酉对角化分解实数方阵。arXiv:1605.07103,

2016 年。

Welbl, Johannes,Bouchard, Guillaume 和 Riedel, Se-bastian。用于测试知识库
补全中的二元词嵌入的因式分解机框架。

arXiv:1604.05878,2016年。

杨碧山、易文涛、何晓东、高建峰、

和邓莉。在知识库中嵌入实体和关系以进行学习和推理。国际

学习表征会议,2015 年。