高度多关系数据的潜在因子模型

Rodolphe Jenatton
CMAP, UMR CNRS 7641,
法国帕莱索高等理工学院
jenatton@cmap.polytechnique.fr

Nicolas Le Roux
INRIA - SIERRA项目组,
巴黎高等师范学校，法国巴黎
nicolas@le-roux.name

安托万Bordes
Heudiasyc, UMR CNRS 7253,
法国科技大学（Université de Technologie de Compie`gne）
antoine.bordes@utc.fr

Guillaume Obozinski
INRIA - 法国巴黎高等师范学校的SIERRA项目组
guillaume.obozinski@ens.fr

抽象

诸如社交网络，电影偏好或知识库之类的许多数据是多关系的，因为它们描述实体之间的多个关系。 虽然有大量的工作集中在建模这些数据上，但共同建模这些多种类型的关系依然具有挑战性。 而且，当这些类型的数量增长时，现有的方法往往会崩溃。 在本文中，我们提出了一个建模大型多关系数据集的方法，可能有数千个关系。 我们的模型是基于双线性结构，它捕捉数据的各种交互顺序，并在不同关系之间共享稀疏的潜在因素。 我们用标准张量因子分解数据集来说明我们的方法的性能，在这些数据集中，我们达到或超越了最先进的结果。 最后，NLP应用程序演示了我们的可伸缩性以及我们的模型学习高效的和语义上有意义的动词表示的能力。

1 介绍

统计关系学习（SRL）[7]旨在模拟由实体之间的关系组成的数据。 社交网络，来自推荐系统的偏好数据，用于语义网络或生物信息学的关系数据库，说明了这种建模具有潜在影响的应用的多样性。

关系数据通常涉及实体或属性之间的不同类型的关系。 这些实体在社交网络或推荐系统的情况下可以是用户，在词汇知识库的情况下是词语，或者在生物信息学本体的情况下是基因和蛋白质等等。 对于二元关系，数据自然表示为所谓的多关系图，其由与实体相关联的节点以及与不同类型的关系对应的节点之间的不同类型的边缘组成。 等价地，数据由表单（主题，关系，对象）的三元组集合组成，列出了我们将要分别称为主题和对象的二元关系的第一和第二项的实际关系。 关系数据通常累积很多困难。 首先是大量的关系类型，其中一些关系类型比其他关系类型更具代表性，可能仅涉及实体的子集； 其次，数据通常是嘈杂和不完整的（缺少或不正确的关系，冗余实体）； 最后大多数数据集都是大规模的，具有高达数百万的实体和数十亿个用于现实世界知识库的链接。

除了关系数据库之外，SRL也可以用来模拟自然语言的语义。 表达语言意义的标准方法是识别文本或言语中的实体和关系并对其进行组织。 这可以从单词或句子级别（例如，在解析或语义角色标记）到文本集合（例如在知识提取中）以各种尺度进行。

SRL系统是一个有用的工具，因为它们可以通过构建摘要[22]，感觉分类词典[11]，本体[20]等从所收集的数据中自动提取高级信息。SRL的进展可能会导致自然语言理解的进步。

在本文中，我们引入关系数据模型并将其应用于多关系图和自然语言。 在给所有其他人分配有效关系和低概率的高概率时，该模型提取数据中各种实体和关系的有意义的表示。 不同于其他分解方法（例如[15]），我们的模型是概率的，其优点是明确地说明数据中的不确定性。 此外，由于关系类型的稀疏分布表示，我们的模型可以处理的数据具有比文献（自然语言数据的一个关键方面）所考虑的数量更多的关系类型。 我们凭经验证明，这种方法在链路预测的各种基准测试（SRL方法的标准测试床）上联系或者击败了最先进的算法。

2 相关工作

关系学习的一个分支，由网络中的协作过滤和链接预测等应用程序所驱动，通过这些实体的固有潜在属性模拟实体之间的关系模型[1]工作于我们称之为关系学习的潜在属性（RLA）主要集中在建模单一关系类型的问题上，而不是试图同时建立一个本身类似的关系集合。正如关系型学习提出的几种形式主义[7]所反映的那样，后一种多关系学习问题需要对大规模关系型数据库进行有效建模。 关系可能相似或相关的事实表明，对于每个关系来说，独立学习的模型的叠加将是非常低效的，特别是因为观察到的每个关系的关系是非常稀疏的。

RLA经常转化为学习实体的嵌入，其与代数对应于矩阵分解问题（通常是观察到的关系的矩阵）。 学习多元关系的一个自然延伸在于将矩阵叠加分解并应用经典的张量分解方法，如CANDECOMP / PARAFAC [25,8]。 这种方法在不同的术语和不同的关系之间固有地引入了参数的一些共享，已经被成功地应用[8]，并且启发了一些概率公式[4]。

同时学习几个关系的另一个自然延伸可以是共享共同的
通过RESCAL中提出的集体矩阵分解嵌入或跨越关系的实体
[15]等相关工作[18,23]。

可以与一个实体关联的最简单的潜在属性形式是一个潜在的类：最终的模型是经典的随机块模型[26,17]。 已经提出了几种基于聚类的方法用于多关系学习：[9]考虑了随机块模型的非参数贝叶斯扩展，允许自动推断潜在簇的数量； [14,28]对此进行了细化，以允许实体拥有混合集群成员资格； [10]介绍了马尔可夫逻辑网络中的聚类； [24]使用嵌入在集体矩阵分解公式中的实体的非参数贝叶斯聚类。 为了分享关系之间的参数，[9,24,14,28]和[10]建立模型，不仅聚集实体，还聚集关系。

为了减少参数的数量，文献[2]的语义匹配能量模型（SME）将关系嵌入到与实体相同的空间的向量中，并通过能量将关系向量并且每个矢量编码这两个项。

在可扩展性方面，RESCAL [15]已经被证明可以在多个关系数据集上实现最先进的性能，最近已经被应用于知识库YAGO [16]，从而显示出其能够很好地扩展数据的能力实体数量，尽管建模关系的数量保持适中（少于100）。 至于中小企业[2]，通过向量对关系进行建模，可以扩大到数千个关系。 由于推理的代价，可伸缩性也可能是非参数贝叶斯模型的一个问题（例如[9,24]）。

[1] *这被称为[10]的统计谓词发明。*

3 关系数据建模

我们考虑由三元组构成的关系数据，这些三元组编码两个实体之间关系的存在，我们称之为主体和客体。 具体而言，我们考虑一个ns子集，$\{O_k\}_{k \in 1;\text{没有}}$以及没有与某些nr关系相关的对象$\{R_j\}_{j \in 1;\ nr}$。 三元组编码的关系$R_j$持有主体我和我们将要写的对象k之间$(S_i, O_k) = 1$。因此，我们将把三元组也称为一个关系。 我们将更详细地讨论的典型例子是在自然语言处理中，其中三元组(i, j, k)对应于主体和直接对象通过传递的关联动词。 目标是要学习一个可靠的预测未知的三胞胎的关系模型。 例如，可能有兴趣仅根据主体和客体(S_i, O_k)找到可能的关系R_j。

4 模型描述

在这项工作中，我们将学习关系的问题形成为矩阵分解问题。 根据先前几种方法[15,24]的基本原理，我们考虑一个模型，其中实体嵌入在R^p中，并且关系在这些实体上被编码为双线性算子。 更准确地说，我们假设ns主题（不包括对象）由向量表示$s^i \in R^p$，存储为矩阵的列$[s^1, \ldots, s^{NS}] \in R^{p \times NS}$（或者作为$O\ [o^1, \ldots, o^{\text{没有}}] \in R^{p \times \text{无}}$的列）。 每个p维表示$s^i, o^k$都必须学习。 关系用矩阵集合$(R_j)\ 1 \leqslant j \leqslant n$表示，其中$R_j \in R^{p \times p}$，它们一起形成三维张量。

我们考虑事件$(S_i, O_k) = 1$的概率模型。 假设s^i和o^k是固定的，我们的模型是从对数模型$P[R_j(S_i, O_k) = 1] \sigma \eta^{(j)}$导出的，$\sigma(t) 1 / 1 + e^{-t}$。 $\eta^{(j)}$的一个自然形式是张量积^{i}linear o^k的线性函数，我们可以写出$\eta^{(j)} = s^i, R_j\ o^k\ R^p$中的产物。 如果我们现在认为对于所有的$(i, j, k)$同时学习$s^i, R_j$和$o^k$，这个模型一起学习矩阵$R_j$和实体的最佳嵌入$s^i, o^k$，以便基于$s^i o^k$的通常逻辑回归很好地预测观察到的关系的概率。这是[24]中考虑的初始模型，如果将最小平方损失替换为逻辑损失，它与[16]中考虑的模型相匹配。 我们将通过两种方式来完善这个模型：首先重新定义$\eta^{(j)}$作为一个函数$\eta^{(j)}_{我} E(s^i, R_j, o^k)$考虑了$s^i$，$o^k$和$R_j$，其次，通过潜在的“关系”因素参数化关系$R_j$，减少模型的参数总数。

4.1 一个多阶对数比率模型

一种考虑与三元组(S_i, R_j, O_k)相应的特定关系发生概率的方法是：(a)从个体实体S_i的边际倾向，O_k进入关系和边际倾向关系R_j (b)从(S_i, R_j)，(R_j, O_k)的对应于倾向于发生在关系(c)的右侧项的左边的实体的双向交互中，与实体对的双向交互(S_i, O_k)总体上往往有更多的关系在一起，(d)(S_i, R_j, O_k)之间的三方依赖关系。

在NLP中，我们常常把它们分别称为单字，双字和三字项，这个术语在本文的其余部分将会重复使用。 因此，我们设计(s^i, R_j, o^k)来解释这些各种秩序的相互作用，只保留涉及R_j的术语[2]。

特别地，引入新的参数$y, y^I, z, z^I \in R^p$，我们定义$\eta^{(j)} = E(s^i, R_j, o$

$$E(s^i, R_j, o^k) \quad (y, R_j\ y^I) + (s^i, R_j\ z) + (z^I, R_j\ k\ i, R_j\ o^k),$$

(1)其中$(y, R_j I)$，$(s^i, R_j z) + (z^I, R_j k)$和$(s^i, R_j k)$单，双和三项术语。 这个参数化在一般情况下是冗余的，因为$E(s^i, R_j, o^k)$的形式是$(s^i + z), R_j(o^k + z^I)) + b_j$；但是在正则化模型的背景下却是有用的（见第5节）。

[2]这是因为我们主要关注对关系项的建模，并且没有必要引入所有项来完全参数化模型。

4.2 通过潜在因素共享关系参数

在学习大量关系时，许多关系的观察次数可能很少，导致过度拟合的风险。 Sutskever等人 [24]用一个非参数贝叶斯模型来解决这个问题，从而诱导关系和实体的聚类。 中小企业[2]提出将关系作为R^p的向量嵌入到实体中，以解决数百种关系类型的问题。

为了减少参数的总数，类似的动机不是像RESCAL [16]那样使用矩阵Rj的一般参数化，而是要求所有的Rj在
常用的一组矩阵{Θr}1≤r≤d表示一些规范关系：

$$Rj = \sum_{r=1}^{d} \alpha^{j} \Theta r, \quad \text{对于一些稀疏的} \alpha^{j} \in R^{d} \text{和} \Theta r = urv^{T}, \quad \text{对于} ur, vr \in R^{p}。$$

(2)

(a) 分解的稀疏性和 (b) dnr导致跨关系共享参数的事实。 此外，将θr约束为外积urv^T也加速了依赖于线性代数的所有计算。

5 规范化的制定和优化

表示（或对应）正面（或负面）标记关系的一组指数，我们试图最大化的可能性是

$$L = \prod_{(I,J,K)\in P} P[Rj(Si, Ok) = 1] \cdot \prod_{(i^I, j^I, k^I)\in N} P[RjI(SiI, OkI) = 0]。$$

对数似然是

$$\ell = \sum_{(I,J,K)\in P} \log(1 + \exp(-\eta^{(j)})) + \sum_{(i^I, j^I, k^I)\in N} \log(1 + \exp(\eta^{(j)}))$$

$= E(s^i, Rj, o^k)$。 为了使 (1) 和 (2) 中出现的术语得到正确的规范，我们执行

$\eta^{(j)}$

即在一个特定的约束集上负对数似然的最小化

$$\min_{s,o,\{\alpha^j\},\, y^I, z, z^I} -\log(L), \quad \text{同} \quad \begin{array}{l} z = z^I,\ 0 = S, \\ s^j, o^k, y, y^I, \text{和} vr \\ z, ur \end{array} \quad \text{在球w;} \quad \begin{array}{l} \text{Fl}\alpha^j\text{11} \le \lambda, \ \Theta r = ur \bullet v_{rT的}, \\ \text{lw12} \le 1. \end{array}$$

我们选择在初步实验的基础上约束Ω1范数，这表明它导致了更好的结果，正则化在Ω2-范数内。 正则化参数λ0控制 (2) 中关系表示的稀疏性。 等式约束引发了一个共享表示
在初步实验中显示出改善模型的主题和目标。 考虑到模型是有条件的 (s^i, o^k)，只有一个尺度参数，即α^j_r
在产品α^j (s^i, θro^k) 中是必要的，这激发了所有的欧几里得单位球约束。

5.1 算法的方法

鉴于我们感兴趣的大问题（例如，$|P| \approx 10^6$），并且由于我们可以有效地将其投影到约束集上（这两个投影都可以在线性时间内执行），我们的优化问题很好地适用于随机预计
梯度算法[3]。

为了加快优化，我们使用了几个实用的技巧。 首先，我们考虑一个随机梯度下降方案，包含100个三联体的小批量。 其次，我们使用形式为a / (1 + k) 的步骤，其中k是迭代次数和aa标量（在所有参数中是公共的）在验证集上的对数网格上进行优化[3]

此外，我们不能将NLP应用程序（请参见第8节）视为标准张量分解问题。 事实上，在这种情况下，我们只能访问积极标记的三联体。 继[2]之后，

(i, j′, k), j′≠ j的每个(i, j, k)三元组生成元素{ } / ∈ P。 在实践中，对于每个正三元组，我们抽取一些包含与我们的正三元组相同主题和客体的不同动词的人为负三元组。 这让我们改变了

[3]该代码可以从开源许可证获得http://goo.gl/TGYuh。

把问题归结为一个多元的问题，其目标是正确地把“正面”动词分类，与“负面”动词相对立。

这个问题的标准方法是使用多项式逻辑函数。 然而，这样的功能对否定动词的选择非常敏感，把所有的动词作为否定动词使用都是非常昂贵的。 另一种更加可靠的方法是使用上面定义的似然函数，我们试图将积极动词分类为有效关系，而将否定关系分类为无效关系。此外，多项式逻辑函数的这种近似是渐近无偏的。

最后，我们观察到，减小否定动词的影响以避免正面动词的影响是有利的。

6 与其他型号的关系

我们的模型与其他几个模型密切相关。 首先，如果d很大，则将Rj的参数解耦，并检索RESCAL模型（直至丢失函数的改变）。

其次，我们的模型还涉及经典的张量分解模型，如PARAFAC，ap-
通过形式的低阶张量H^在最小二乘意义上接近张量[Rk (Si, Oj)] i, j,

$\sum_{r=1}^{d} \alpha$ (αr, β_r, γ_r) ∈$R^{nr \times ns \times no}$。 所有Rj参数化为线性组合，

$r \otimes \beta_r$

实际上等同于一个矩阵的国家等同于将张量R = {Rj}$_{j \in 1; nr}$限制为
是低秩张量R =
也可以写成 $\sum_{r=1}^{d} \alpha$r⊗ur⊗vr。 结果，所有三元组的张量[4]
$\alpha_r \otimes \beta_r \otimes \gamma_r$, β = S_r^Tur和 γ = O_r^Tvr。 这表明我们的
模型是张量因子分解的一种特殊形式，当p足够大时，它减少到PARAFAC（直到损失函数的变化）。

最后，[2]中考虑的方法似乎与我们的先验有很大不同，特别是因为关系是作为R^p的向量嵌入的实体而不是矩阵
$R^{p \times p}$在我们的情况。 正如我们所展示的，这种选择对建立复杂的关系模式是不利的在第7节。另外，没有参数化模型[2]能够同时处理bigram和trigram的相互作用。

7 应用于多关系基准

我们在本节中报告我们的模型在标准张量分解数据集上的性能，我们首先简要描述。

7.1 数据集

亲缘关系。 澳大利亚部落因其亲属制度的复杂关系结构而闻名于人类学家之列。 这个由[6]创建的数据集聚焦于澳大利亚中部的Alyawarra部落。 要求104个部落成员提供彼此之间的亲属关系。 这导致了104个实体和26个关系类型的图表，每个图表描绘不同的亲属关系术语，例如Adiadya或Umbaidya。 有关更多详细信息，请参阅[6]或[9]。

UMLS。 这个数据集包含来自[12]收集的统一医学语言系统语义工作的数据。 这包含一个包含135个实体和49个关系类型的图。 这些实体是“疾病或综合症”，“诊断程序”或“哺乳动物”等高级概念。 这些关系表示描述“情感”或“原因”等概念之间的因果关系的动词。

国家。 这个数据集将14个国家（巴西，中国，埃及等）用56种二元关系类型来表示，如“经济援助”，“条约”或“相对外交”，以及描述每个国家的111个特征另有111个实体通过另外的“有特征”关系与该国进行互动[5]。 详见[21]。

[4]其他术语可以用类似的方法分解。-

[5]由此产生的新关系仅用于培训，在考试时不考虑。

数据集		我们的方法	RESCAL [16]	MRC [10]	中小企业[2]
亲缘关系	*PR曲线下面积* *数似然*	**0.946** ± 0.005 **-0.029** ± 0.001	**0.95** N/A	0.84 -0.045 ± 0.002	0.907 ± 0.008 N/A
UMLS	*氩 在PR曲线下* *数似然*	**0.990** ± 0.003 **-0.002** ± 0.0003	0.98 N/A	0.98 -0.004 ± 0.001	0.983 ± 0.003 N/A
国家	*氩在PR曲线下* *数似然*	**0.909** ± 0.009 **-0.202** ± 0.008	0.84 N/A	0.75 -0.311 ± 0.022	**0.883** ± 0.02 N/A

表1：我们的方法获得的性能比较，RESCAL [16]，MRC [10]和SME [2]三个标准数据集。 结果通过10倍交叉验证来计算。

7.2 结果

这三个数据集是相对小规模的，只包含几个关系（几十个）。 由于我们的模型主要是为了处理大量的关系而设计的（见4.2节），所以这个设置对评估我们的方法的潜力是最有利的。如表1中所报道的，我们的方法在精确性 - 回忆曲线（AUC）和对数似然性（LL）下的面积方面与先前的技术相比具有更好的或同样好的性能。 表1中显示的结果通过10次交叉验证[6]，数据集的10次随机分割（交叉验证为90%，测试为10%）进行平均计算。 我们选择将我们的模型与RESCAL [16]，MRC [10]和SME [2]进行比较，因为就我们所知，他们在AUC和LL方面取得了最佳公布结果。

有趣的是，（1）中的trigram项对于获得良好的亲属关系（除去trigram项，我们在AUC中得到0.16和在LL中得到0.14）是必不可少的，因此表明需要在复杂关系数据中建模3-way交互。 而且，正如所期望的，由于关系数量较少，通过交叉验证所选择的 λ 值非常大（$\lambda = nr\ d$），因此在（2）中不会导致稀疏性。 这个数据集的结果也表现出与SME [2]一样的矩阵关系建模的好处，而不是像向量一样。

Zhu [28]最近报道了国家和亲属关于接受者操作特征曲线下面积的评估结果，而不是精确查全率曲线下的面积，如表1所示。使用这个另外的度量，我们的模型得到0.953国家和0.992的亲属关系，因此胜过朱的方法，分别达到0.926和0.962。

8 学习动词的语义表征

通过提供一种模式化语言关系结构的方法，SRL可以很好地用于学习自然语言语义。 因此，本部分提出将我们的方法应用于维基百科的文本数据，以学习词语的表示，重点放在动词上。

8.1 实验设置

数据。 我们分两个阶段收集这些数据。 首先，使用SENNA软件[7][5]对200万维基百科文章进行词性标注，组块化，词形化[8]和语义角色标注。 然后过滤这些数据，以便只选择句法结构（主语，动词，直接宾语）的句子，而三重词的每一项都是WordNet词典中的一个单词[13]。 受试者和直接对象最终都是单个名词，其字典大小为30,605。 这个数据集中的关系总数（即动词数）是4,547：这比以前发布的多关系基准要大得多。 我们保留了100万个这样的关系来建立一个训练集，5万个验证集和25万个测试。 所有的三联体都是独一无二的，我们确保所有出现在验证或测试集中的单词都出现在训练集中[9]

[6] 在nr中搜索 λ，d和p的值 $\times d \cdot \{0.05, 0.1, 0.3\}, \{1, 100, 200\}$ 500和10，25，50。

[7] 可从ronan.collobert.com/senna/获取。

[8] 使用NLTK（nltk.org）进行词形化并将单词转换为其基本形式。

[9] 该数据集根据开源许可证提供http://goo.gl/TGYuh。

	同义词不是 考虑			最好的同义词考虑		
	中位数/平均等级	p@5	p@20	中位数/平均等级	p@5	p@20
我们的方法	50 / **195.0**	**0.78**	**0.95**	19 / **96.7**	**0.89**	**0.98**
中小企业[2]	56 / 199.6	0.77	**0.95**	19 / 99.2	**0.89**	**0.98**
两字	**48** / 517.4	0.72	0.83	**17** / 157.7	0.87	0.95

表2：通过我们的方法，SME [2]和一个bigram模型在NLP数据集上获得的性能。 有关表格统计的细节在文中给出。

实践训练设置。 在训练阶段，我们优化了验证集的各种参数，即大小\p
∈ 25, 50, 100表示潜在分解（2）的尺寸d 50, 100, 200，∈
{ }
正则化参数 λ 的值作为nr的分数1, 0.5, 0.1, 0.05, 0.01 d，0.1, 0.05, 0.01的步长和负
{ } ×
三元组的加权。 此外，为了加快训练速度，我们逐步增加了取样否定动词的数量（参见第
{ }
5.1节），从25增加到50，这有助于提炼训练效果。

8.2 结果

动词预测。 我们首先考虑基于250,000个实例的测试集对我们的方法进行直接评估，通过测量我们如何预测给定一对（主体，直接对象）的相关和有意义的动词。 为此，对于每个测试关系，我们使用给定的一对（主题，直接对象）的概率估计对所有动词进行排名。 表2显示我们的结果有两种度量，即（1）正确动词的等级和（2）正确动词排在最高z%的测试例子的比例。 后者的标准被称为p @ z。 为了评估一些语言语义是否被表达式捕获，我们也考虑一个不太保守的方法，我们不是只关注正确的动词，而是测量从WordNet获得的同义词集合的最小等级。 我们的方法与中小企业[2]的结果进行了比较，该结果显示可以很好地处理具有大关系的数据，并且用一个二元模型来估计对（主体，动词）和（动词，直接宾语）。

第一个观察结果是，通过基于双向交互的简单模型可以很好地解决动词预测的任务，正如由bigram模型获得的好的中值秩所示。 这一点由三元项对模型性能的轻微影响所证实。在这个数据上，我们体验到在我们的能量函数中使用bigram相互作用是实现良好预测的关键。 然而，在我们的方法和仅使用二元模型的模型之间的平均等级下降仍然表明，许多例子需要更丰富的模型才能正确处理。 相比之下，我们倾向于持续匹配或改善中小企业的表现。 重要的是，模型选择导致选择 λ = 0.1 nr d，其中表示（2）的系数 α
在某种意义上说它们是稀疏的，因为它们被很少的大数值所占据（例如， α 的最大值的前2%
× 范数II的约25%）。
约占整个1

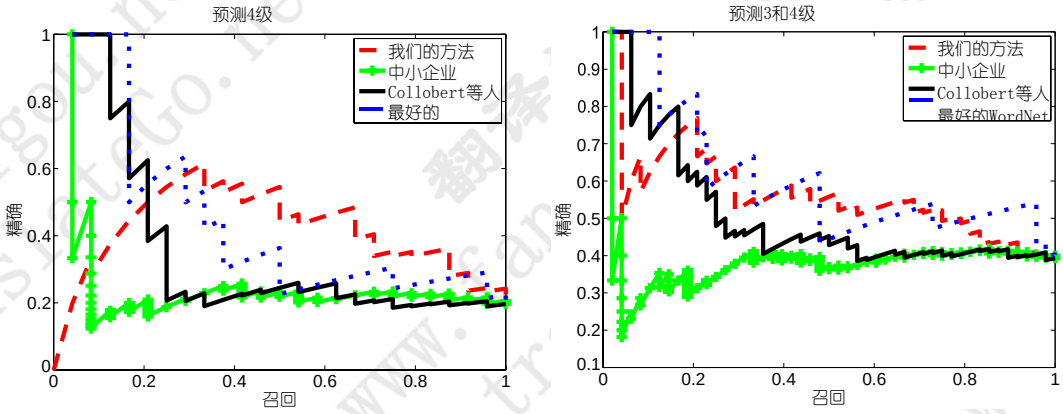

图1：词汇相似度分类任务的精度回忆曲线 根据动词之间不同的相似性度量来计算曲线，即我们的方法SME [2]，Collobert et al。 [5]和最好的（三个）WordNet相似性度量[13]。关于任务的细节可以在文中找到。

	我们的方法	中小企业[2]	Collobert et 人[5]。	最好	WordNet [19]
AUC（4级）	**0.40**	0.21	0.31		**0.40**
AUC（3级和4级）	0.54	0.36	0.48		**0.59**

表3：在词汇相似度分类任务中获得的性能[27]，我们比较了我们的方法，SME [2]，Collobert等人的词嵌入[5]和最好的（3分）WordNet相似性度量[19]在精度 - 召回曲线下使用面积。 细节在文中给出。

词汇相似度分类。 我们的方法学习动词的潜在表示，并通过共享参数强加一些结构，如4.2节所示。 这应该导致类似动词的类似表示。 我们考虑[27]中描述的词汇相似性分类的任务来评估这个假设。 他们的数据集由130对动词标记的动词，分数为0,1,2,3,4。 分数越高意味着组成这对的动词之间的语义相似性越强。 例如，（分割，分割）标记为4，而（延迟，分割）分数为0。

基于我们学习的动词表示Rj之间的成对欧几里德距离[10]，我们试图通过使用Ri和Rj之间的最小距离的假设来预测类别4（以及“合并”类别3,4） （i，j）应该被标记为4.我们将[2]在相同训练数据上学习的表示与[5]的字嵌入（在自然语言处理中被认为是有效的特征）进行比较，并采用WordNet Similarity [19]提供的三种相似性度量。 最近，我们只显示最好的一个，名为“路径”，它是通过计算沿着WordNet的“is-a”层级中的感官之间的最短路径的节点的数目而建立的。

我们报告了我们在图1中显示的精确回忆曲线的结果以及表3中曲线下面的相应面积（AUC）。尽管我们倾向于错过前几个对，但是我们比较好[2]和[5]以及我们的AUC接近于WordNet相似度所建立的参考。 我们的方法能够为动词编码有意义的语义嵌入，即使它已经在有噪声的自动收集的数据上进行了训练，尽管参数空间距离应该满足任何条件并不是我们的主要目标。 通过培训更清洁的三元组，可以提高性能，如[11]收集的。

9 结论

能够处理大量关联关系的设计方法似乎有必要对任何现实世界问题的语义底层关系的丰富性进行建模。 我们通过使用自然适合于多关系数据的关系的共享表示来解决这个问题，其中实体具有在关系类型之间共享的唯一表示，并且我们建议关系本身在潜在“关系”因素上分解。 这种新的方法在标准的关系学习问题和NLP任务上联系或者击败了最先进的模型。关于潜在因素的关系的分解允许显着减少参数的数量，并且由于计算和统计的原因而受到激励。 具体而言，我们的方法在关系数量和数据样本方面都是可扩展的。

有人可能会问我们提出的各个术语的相对重要性。 有趣的是，虽然三项词的存在对于张量分解问题是至关重要的，但它在NLP实验中起到了很小的作用，在这个实验中，大多数的信息都包含在二元和单数词汇中。

最后，我们认为，通过对潜在因素的分析来探索关系的相似性，可以为不同关系类型之间共享的结构提供一些见解。

致谢

这项工作是由Pascal2欧洲卓越网络部分资助。 NLR和RJ由欧洲研究委员会（SIERRA-ERC-239993和SIPA-ERC-256919）提供支持。

[10]其他距离当然可以考虑，为了简单起见，我们选择欧几里德度量。

参考

[1] F.Bach, R.Jenatton, J.Mairal和G.Obozinski。 优化与稀疏诱导惩罚。 机器学习的基础和趋势，4（1）：1-106，2011。

[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，2012。出现。

[3] L. Bottou和Y. LeCun。 大规模在线学习。 神经信息处理系统的进展（Advances in Neural Information Processing Systems），第16卷，第217-224页，2004。

[4] W. Chu和Z. Ghahramani。 不完整多维数组的概率模型。 Journal of Machine Learning Research - Proceedings Track，5：89-96，2009。

[5] R.Collobert, J.Weston, L.Bottou, M.Karlen, K.Kavukcuoglu和P.Kuksa。 自然语言处理（几乎）从头开始。 JMLR，12：2493-2537，2011。

[6] *W. Denham。 Alyawarra非言语行为模式的检测。 博士论文，1973。*

[7] *L. Getoor和B. Taskar。 统计关系学习（自适应计算和机器学习）简介。 麻省理工学院出版社，2007。*

[8] RA Harshman和ME Lundy。 Parafac：平行因子分析。 COMPUT。 统计。 数据分析，18（1）：39-72，1994年8月。

[9] C. Kemp, JB Tenenbaum, TL Griffiths, T. Yamada和N. Ueda。 用无限关系模型学习概念系统。 在Proc。 AAAI，第381-388页，2006年。

[10] S. Kok和P. Domingos。 统计谓词发明。 在第24届国际机器学习会议论文集，第433-440页，2007年。

[11] A. Korhonen, Y. Krymolowski和T. Briscoe。 自然语言处理应用程序的大型子类别词典。 在LREC会议录2006年。

[12] 在McCray。 生物医学领域的上层本体。 比较和功能基因组学，4：80-88,2003。

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

[14] K. Miller, T. Griffiths和M. Jordan。 用于链接预测的非参数潜在特征模型。 在Advances in Neural Information Processing Systems 22，第1276-1284页中。 2009年。

[15] M. Nickel, V. Tresp和H.-P. 克里格尔。 多关系数据集体学习的三维模型。 在第28届国际会议论文集 在马赫上。 学习，第809-816页，2011年。

[16] M. Nickel, V. Tresp和H.-P. 克里格尔。 分解YAGO：可链接数据的可伸缩机器学习。 在 *PROC。 第二十一届国际博览会 在WWW上，第271-280页，2012。*

[17] K. Nowicki和TAB Snijders。 随机块结构的估计和预测。 Journal of the American Statistical Association，96（455）：1077-1087,2001。

[18] A. Paccanaro和G. Hinton。 利用线性关系嵌入学习概念的分布式表示。 IEEE Trans。 在Knowl。 和Data Eng。，13：232-244,2001。

[19] T.Pedersen, S.Patwardhan和J.Michelizzi。 Wordnet ::相似度：衡量概念的相关性。 在2004年HLT-NAACL的演示文件中，第38-41页，2004。

[20] H. Poon和P. Domingos。 从文本的无监督本体论归纳。 在“计算语言学协会第48届年会会刊”，第296-305页，2010年。

[21] RJ Rummel。 国家的维度项目：国家的属性和国家的行为。 在ICPSR数据文件，第1950-1965页。 1999年。

[22] D. Shen, J.-T. Sun, H. Li, Q. Yang和Z. Chen。 使用条件随机字段的文档摘要。 在Proc。 第二十届国际联合会议。 在Artif。 英特尔，第2862-2867页，2007。

[23] AP Singh和GJ Gordon。 通过集体矩阵分解进行关系学习。 在Proc。 SIGKDD'08，第650-658页，2008年。

[24] I. Sutskever, R. Salakhutdinov和J. Tenenbaum。 使用贝叶斯聚类张量分解对关系数据建模。 在Adv。 在神经。 天道酬勤。 PROC。 SYST。 2009年2月22日

[25] LR Tucker。 关于三模式因素分析的一些数学笔记。 Psychometrika，31：279-311，1966。

[26] 王亚军和王永康。 有向图的随机块模型。 美国统计协会杂志，82（397），1987。

[27] D. Yang和DMW Powers。 关于wordnet分类的动词相似性。 Proceedings of GWC-06，第121-128页，2006。

[28] J. Zhu。 用于链路预测的最大余量非参数潜在特征模型。 在2012年第29届国际机器学习会议论文集中。