

关于知识图的关系机器学习综述

从多关系链接预测到自动化知识图构建

马克西米利安·凯文·墨菲, 沃尔克·Tresp, Evgeniy Gabrilovich

抽象关系机器学习研究关系数据或图形结构数据的统计分析方法。在本文中, 我们回顾了这些统计模型如何在大型知识图上“训练”, 然后用于预测关于世界的新事实(相当于预测图中的新边)。具体来说, 我们讨论两种不同的统计关系模型, 这两种模型都可以扩展到海量数据集。首先是基于张量分解的方法和相关的潜变量模型。第二个是基于在图中挖掘可观察的模式。我们还展示了如何结合这些潜在的和可观察的模式, 以降低计算成本获得改进的建模能力。最后, 我们讨论如何将这样的图形统计模型与基于文本的信息提取方法相结合, 从Web自动构建知识图形。我们特别讨论了Google的知识库项目。

索引术语 - 统计关系学习, 知识图, 知识抽取, 潜在特征模型, 基于图的模型

I. 介绍

我深信, 学习问题的关键是认识到关系, 并能够使用它们。

克里斯托弗·斯特拉奇在1954年给艾伦·图灵的一封信中

MACHINE学习通常与数据矩阵一起工作, 其中每一行表示一个对象属性(可能是数字或分类)的特征向量, 以及主要任务要学习从该特征向量到某种形式的输出预测的映射, 或者执行无监督学习(如聚类或因子分析)的特征向量。在统计关系学习(SRL)中, 对象的表示可以包含与其他对象的关系。因此, 数据是以图形的形式, 由节点(实体)和标记的边缘(实体之间的关系)组成。SRL的主要目标包括预测缺失边缘, 预测节点属性, 以及根据连通性模式对节点进行聚类。这些任务在诸如社交网络和生物路径分析等许多环境中都有出现。有关SRL的更多信息, 请参阅[1, 2, 3]。

在本文中, 我们回顾了来自SRL社区的各种技术, 并解释如何应用它们

Maximilian Nickel是计算和统计学习实验室, 麻省理工学院的Poggio实验室和Istituto Italiano di Tecnologia。Volker Tresp与西门子股份公司, 企业技术部和路德维格公司合作马克西米利安大学慕尼黑。

Kevin Murphy和Evgeniy Gabrilovich在2014年10月19日收到Google Inc. 手稿; 2015年1月16日修订。

到大规模知识图(KGs), 即以实体之间关系的形式存储事实信息的图形结构知识库(KB)。最近, 创造了大量的知识图表, 包括YAGO [4], DBpedia [5], NELL [6], Freebase [7]和Google知识图[8]。正如我们在部分讨论II, 这些图形包含数百万个节点和数十亿的边界。这使我们专注于可扩展的SRL技术, 这需要花费图表大小的线性时间。

除了统计关系学习的典型应用之外, 我们还将讨论SRL如何帮助信息提取方法自动“增长”KG。具体来说, 我们将展示如何使用SRL来训练基于现有KG的“先验”模型, 然后将其预测与使用机器读取方法从网络自动提取的“噪音”事实相结合(参见例如[9, 10])。这是Google的知识库项目中采用的方法, 正如我们在本节中所解释的VIII。

本文的其余部分的结构如下。在部分II 我们介绍知识图和他们的一些属性。部分III 讨论了SRL及其如何应用于知识图。有两种主要的SRL技术类型: 利用潜变量捕获节点/边之间的相关性的那些类, 以及基于图的可观察属性直接使用统计模型捕获相关性的技术。我们在第二节讨论这两个家庭IV 和科V, 分别。部分VI 描述了结合这两种方法的方法, 以获得两全其美。在部分VII 我们讨论使用马尔科夫随机场的关系学习。在部分VIII 我们介绍了如何将SRL用于自动化知识库建设项目。在部分IX 我们讨论所提出的方法的扩展, 以及Section X 提出我们的结论。

II. 知识图

在本节中, 我们将讨论知识图表: 它们如何表示, 如何创建以及如何使用它们。

A. 知识表示

用于幼儿园的关系知识表示在逻辑和人工智能方面有着悠久的历史[11]。最近, 它已经被用在语义网络中以机器可读形式表示信息, 以便使智能代理能够在“数据网络”上运行[12]。虽然语义网的原始视觉仍然有待充分实现, 但部分

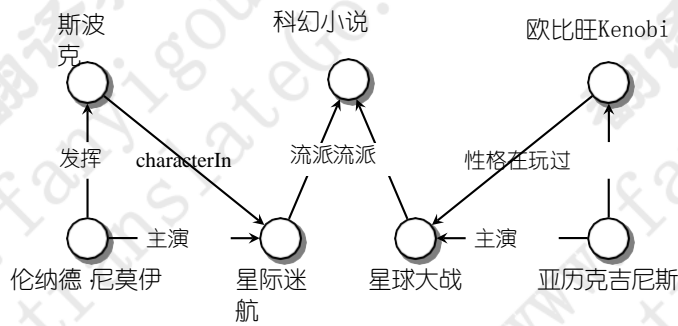


图1. 小例子知识图。节点代表实体，边缘标签代表关系的类型，边代表现有的关系。

知识	表一	项目
创作方法	基于模式的项目	
策划 共同 汽车。半结构化 汽车。非结构化	Cyc/OpenCyc [19], 共发现 [20], UMLS [21] 维基数据[22], 游离碱 [7] YAGO [4, 23], DBpedia中 [5], 游离碱 [7] 知识库[24], NELL [6], PATTY [25], DeepDive /小学[26], Prospera (普罗斯佩拉) [27]	
创作方法	无模式项目	
汽车。非结构化 主角	混响[28], OLLIE [29], 棱柱 [30]	

已经实现。特别是关联数据的概念[13, 14]已经获得了关注，因为它利用W3C资源描述框架（RDF）以关系形式促进了Web上的数据的发布和链接，[15, 16]。

在本文中，我们将松散地遵循RDF标准，以二进制关系的形式表示事实，特别是（主体，谓词，对象）（SP0）三元组，其中主体和对象是实体，谓词是关系的类型。

（我们讨论如何在Section中表示高阶关系IX-A）特定SP0三元组的存在表明了一个现有的事实，即相应的实体处于相应类型的关系中。例如，信息

Leonard Nimoy是科幻电影“星际旅行”（Star Trek）中饰演Spock角色的演员

可以通过以下SP0三元组表示：

学科	谓词	目的
(伦纳德·尼莫伊, 演员)		职业,
(LeonardNimoy, StarTrek)		主演,
(LeonardNimoy, Spock)		出场,
(星际迷航, characterIn, 迷航)		Spock) (Spock,
(星际迷航, 类型, 科幻小说)		characterIn, 星际

我们可以将所有SP0三元组合在一起形成一个多图，其中节点表示实体（所有主题和对象），有向边表示关系。边的方向指示实体是否作为主体或对象出现，即从主体到对象的边缘。不同的关系类型通过不同类型的边缘（也称为边缘标签）来表示。这种结构称为知识图（KG），有时也称为异构信息网络[17]。见图1举一个例子。

除了一系列事实之外，知识图通常提供类型层次结构（Leonard Nimoy是一个演员，它是一个活生生的人）和类型限制（例如，一个人只能嫁给另一个人，而不是一个东西）。

B. 开放与封闭世界的假设

虽然现有的三元组总是编码真实的关系（事实），但是对不存在的三元组的解释有不同的范式：

- 在封闭世界假设（CWA）下，不存在的三元组表示错误的关系。例如，在图中的事实1 伦纳德没有出演

对“星球大战”的尼莫伊的解释是，尼莫伊肯定没有出演这部电影。

- 在开放世界假设（OWA）下，一个不存在的三元组被解释为未知的，即相关关系可以是真或假。继续上面的例子，缺少的边缘不被解释为Nimoy没有在星球大战中出演。这种更谨慎的做法是合理的，因为幼稚园是非常不完整的。例如，有时只列出电影中的主角，而不是完整的角色。此外，即使是你可能认为通常已知的出生地点，在Freebase中包含的所有人中，71%都没有[18]。

RDF和语义Web使得开放世界的假设成为可能。在部分III-D我们还讨论了常用于训练关系模型的局部封闭世界假设（LCWA）。

C. 知识库建设

完整性、准确性和数据质量是决定知识库有用性的重要参数，并受到KB（KG是一种特殊形式）的构建方式的影响。我们可以将KB的构建方法分为四大类：

- 在策划方法中，由封闭的专家组手动创建三元组。
- 在协作方法中，三分之一是由一群开放的志愿者手动创建的。
- 在自动化半结构化方法中，三元组通过手工制定的规则，学习规则或正则表达式从半结构化文本（例如维基百科中的信息框）自动提取。
- 在自动化非结构化方法中，通过机器学习和自然语言处理（NLP）技术自动从非结构化文本中提取三元组（参见例如，[9]为审查）。

策划的知识库的建立通常导致高度准确的结果，但是由于其依赖于人类专家，这种技术不能很好地扩展。用于构建维基百科和Freebase的协作式知识库建设虽然规模较大，但仍然存在一定的局限性。例如，尽管这是模式的强制性属性，但Freebase中包含的所有人的71%都缺少出生地属性[18]。另外，最近的一项研究[31]发现维基百科的增长速度正在放缓。所以，

表二
一些基于模式的知识库的大小

知识图	数量		
	实体	关系类型	事实
游离碱	40 M	35,000	637 M
维基数据	13 M	1,643	50 M
DBpedia中 ¹	4.6 M	1,367	68 M
YAGO2	10 M	72	120 M
Google知识图表	570 M	35,000	18,000 M

自动知识库建设 (AKBC) 方法已经引起了越来越多的关注。

ABKC方法可以分为两种主要方法。第一种方法利用半结构化的数据,如维基百科信息框,这导致了高精度的大量知识图形;示例项目包括YAGO [4, 23]和DBpedia [5]。但是,半结构化文本仍然只覆盖Web上存储的一小部分信息。因此,第二种方法试图“阅读网络”,从网页中的自然语言中提取事实。示例项目包括NELL

[6]和知识库[24]。在部分 VIII,我们展示了如何通过将这些自动提取的事实与从现有的高质量幼儿园中提取的知识相结合来降低“噪音”的水平。

幼稚园,更普遍的知识分子,也可以根据他们是否使用固定或开放的实体和关系词典来分类。特别是,我们区分两种主要类型的KB:

- 在基于模式的方法中,实体和关系通过全局唯一的标识符来表示,并且所有可能的关系都是以固定的词汇表预定义的。例如,Freebase可能代表了奥巴马出生在夏威夷,使用三重 (/ m / 02mjmr, /人/人/天生, / m / 03gh4),其中/ m / 02mjmr是Barack Obama的唯一机器ID。
- 在无模式方法中,实体和关系使用开放信息抽取 (OpenIE) 技术来识别[32]并通过规范化但不歧义的字符串 (也称为表面名称) 表示。例如,OpenIE系统可能包含 (“奥巴马”, “出生”, “夏威夷”), (“奥巴马”, “出生地”, “檀香山”) 等三元组。请注意,从这个表示中清楚地看出,第一个三元组是指第二个三元组的同一个人,还是“生于”是指与“出生地”相同的东西。这是OpenIE系统的主要缺点。

表I列出了按其创建方法和数据模式分类的当前知识库建设项目。在本文中,我们只关注基于模式的知识库。表II显示了这些KB及其大小的选择。

D. 知识图的使用

知识图提供了可由机器解释的语义结构化信息 - 一种被认为是构建更智能的重要组成部分的属性

机器[33]。这样的知识图有各种商业和科学应用。一个典型的例子就是将谷歌的知识图谱 (Google Graph Graph) 整合到谷歌的搜索引擎中[8]。Google KG用于识别和消除文本中的实体,丰富搜索结果和语义结构的摘要,并提供相关实体的链接。(微软有一个类似的知识产权,称为Satori,与其Bing搜索引擎整合在一起 [34].)

利用来自知识图的语义信息来增强搜索结果可以被看作是将基于文本的搜索引擎转换成语义意识的问答服务的重要步骤。展示知识图谱价值的另一个突出的例子是IBM的问答系统Watson,它能够在Jeopardy游戏中击败人类专家。其中,该系统使用YAGO, DBpedia和Freebase作为其信息来源[35].

知识图也用于几个专业领域。例如, Bio2RDF [36], Neurocommons [37]和LinkedLifeData [38]是集成多种生物医学信息来源的知识图表。这些已被用于生命科学中的问题回答和决策支持。

E. 知识图谱建设与管理的主要任务

在本节中,我们将回顾一些典型的KG任务。

链接预测涉及预测图中 (类型) 边 (即,三元组) 的存在 (或正确性的概率)。这是很重要的,因为现有的知识图通常缺少许多事实,而且它们所包含的一些边是不正确的[39]。已经表明,考虑到实体关系的关系模型可以明显优于非关系ML方法 (例如参见[40, 41])。正如我们在Section中详细描述的那样 VIII链接预测可以通过从已知事实中估计三元组的可信性来支持自动化知识库的构建。例如,假设一个信息提取方法返回一个事实,即声称奥巴马出生在肯尼亚,并假设 (出于说明的目的) 奥巴马的真实出生地点尚未存储在知识图中。一个SRL模型可以使用有关奥巴马 (如他的职业是美国总统) 的相关事实来推断这个新事实不太可能是真实的,不应该包括在内。

实体分解 (也称为记录链接[42],对象识别[43],实例匹配[44]和重复数据删除[45]) 是识别关系数据中的哪些对象引用相同的底层实体的问题。在关系设置中,关于哪些对象被假定为相同的决定可以通过图传播,使得针对域中的所有对象集体做出匹配决定,而不是针对每个对象对独立做出决定 (例如,参见[46, 47, 48])。在基于模式的自动化知识库构建中,可以使用实体解析来将提取的表面名称与存储在知识图中的实体进行匹配。

基于链接的集群将基于特征的集群扩展到关系学习设置,并将关系数据中的实体分组

¹2014版,英文内容

基于它们的相似性。然而，在基于链接的聚类中，实体不仅按其特征的相似性分组，而且由其链接的相似性分组。在实体解析中，实体的相似性可以通过知识图传播，这样关系建模可以为这个任务增加重要的信息。在社交网络分析中，基于链接的聚类也称为社区检测[49]。

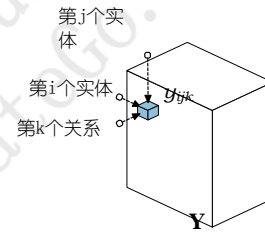


图2. 二元关系数据的张量表示。

III. 统计学的关系学习

知识图

统计关系学习关系到为关系数据创建统计模型。在下面的章节中，我们将讨论如何将统计关系学习应用于知识图。我们将假设知识图中的所有实体和（类型）关系都是已知的。（我们在Section中讨论这个假设的扩展IX-C）。然而，三元组被认为是不完整和嘈杂的；实体和关系类型可能包含重复项。

符号：继续之前，让我们定义我们的数学符号。（变量名将在后面的相应部分介绍。）我们用小写字母来表示标量，比如 a ；列向量（大小为 N ）用粗体小写字母，如 \mathbf{a} ；矩阵（大小为 $N_1 N_2$ ）用大写字母（如 A ）表示；和张量（大小为 $N_1 N_2 N_3$ ）用粗体大写字母和下划线（如 \underline{A} ）表示。张量 \underline{A} 的第 k 个“切片”由 A （它是 $\hat{\sim}_k$ 大小 $N_1 N_2$ ）和 p_i, j, k 个元素 a_{ijk} （这是一个标量）。我们用 $\mathbf{r}_a; \mathbf{b}_s$ 来表示矢量的垂直叠加

\mathbf{a} 和 \mathbf{b} ，即 $\mathbf{r}_a; \mathbf{b}_s$ 。我们可以通过堆叠 \underline{A} 的所有列来将大小为 $N_1 N_2$ 的矩阵 A 转换成大小为 $N_1 N_2$ 的向量 \mathbf{a} ，表示为“ $\text{vec } \underline{A}$ ”。内部（标量）产品两个载体（均为大小 N ）由 $\mathbf{a}^T \mathbf{b}$ 定义。张量（克罗内克）两个向量乘积（大小为 N_1 和 N_2 ）是大小为 $N_1 N_2$ 的矢量，其条目为 $\mathbf{a}^T \mathbf{b}$ 。

矩阵乘法表示为一个向量的L2范数由 $\|\mathbf{a}\|_2$ 表示。照常。按照 \underline{A} 和Frobenius表示

矩阵的范数由 $\|\underline{A}\|_F$ 表示。我们用 $\mathbf{1}$ 表示全1的矢量，用 \mathbf{I} 表示单位矩阵。

A. 概率知识图

我们现在介绍一些数学背景，所以我们可以更正式地定义知识图的统计模型。让 $\mathcal{E} = \{e_1, \dots, e_N\}$ 是所有实体和 $\mathcal{R} = \{r_1, \dots, r_{N_r}\}$ 的集合。是知识库中所有关系类型的集合，边缘图。我们为每个可能的三重 $\langle e_i, r_k, e_j \rangle$ 建模。在这组实体和关系上作为二进制随机变量 $y_{ijk} \in \{0, 1\}$ 表示它的存在。 $\mathcal{E} \times \mathcal{R} \times \mathcal{E}$ 中所有可能的三元组都可以自然地分组为一个三阶张量（三路阵列） $\underline{Y} \in \{0, 1\}^{N \times N_r \times N}$ ，其条目已设置

$y_{ijk} = 1$ ，如果三重 $\langle e_i, r_k, e_j \rangle$ 存在。

我们将这个结构称为邻接张量（参见图2）。 \underline{Y} 的每个可能的实现都可以被解释为一个可能的世界。为了导出整个知识图的模型，我们感兴趣从一个子集估计联合分布 $P(\underline{Y})$ 观察到的三元组。在这样做的时候，我们估计了可能世界的概率分布，这使我们能够根据整个知识图的状态来预测三元组的概率。虽然邻接张量中的 y_{ijk} 表示存在三元组，但是解释是依赖于开放世界，封闭世界还是局部闭合的世界假设。有关详细信息，请参见部分III-D。

请注意，对于大型知识图， \underline{Y} 的大小可能是巨大的。例如，Freebase目前由超过4千万个实体和35万个关系组成，可能的三元组的数量是 $|\mathcal{E} \times \mathcal{R} \times \mathcal{E}|$ 超过 10^{19} 元素。当然，类型限制大大减少了这个数字。

即使在语法上有效的三元组中，只有很小的一部分可能是真的。例如，有超过45万的演员和超过250,000部电影存储在Freebase。但是每个演员只在少数电影中出演。因此，知识图谱的一个重要问题就是如何在有效利用关系稀疏的同时处理大量可能的关系。理想情况下，大规模知识图的关系模型应该是与数据大小最多呈线性关系，即与实体数量 N_e 成线性关系，关系数量 N_r 成线性关系，

并且观察到的三元组 $|\mathcal{D}|$ 的数量是线性的。 N_d 。

B. SRL模型的类型

正如我们所讨论的，关系数据中某些三元组的存在或不存在与某些其他三元组的存在或缺失相关（即预测）。换句话说，随机变量 y_{ijk} 是相互关联的。我们将讨论三种主要的方法来模拟这些相关性：

- M1) 假设所有 y_{ijk} 都是条件独立的，给定与主体、客体和关系类型以及附加参数（潜在特征模型）相关的潜在特征，
- M2) 假设所有 y_{ijk} 都是条件独立的，给定观察图特征和附加参数（图特征模型）

- M3) 假设所有 y_{ijk} 有当地的相互作用（马尔科夫0，否则。

随机场)

下面我们主要关注 M1 和 M2 及其组合; M3 将成为科的话题 VII.

模型类M1和M2通过分数函数 f_{xijk} 预测三元 x_{ijk} 的存在； Θ 代表了模型在给定参数 Θ 的情况下存在三元组的可信度。M1和M2的条件独立性假设允许概率模型写成如下：

$$P(p_{ijk} | D, \theta) = \prod_{ijk} \text{Ber}(p_{ijk} | \sigma(f_{xijk}, \Theta)) \quad (1)$$

其中 $\sigma(p) = \frac{1}{1 + \exp(-p)}$ 是sigmoid (逻辑) 函数,

和

$$P(y_{ijk} | p_{ijk}) = \begin{cases} 1 & \text{如果 } p_{ijk} > 0.5 \\ 0 & \text{如果 } p_{ijk} \leq 0.5 \end{cases} \quad (2)$$

是伯努利分布。

我们将参考方程式的模型 (1) 作为概率模型。除了概率模型之外，我们还将讨论在其他标准下优化 f 的模型，例如最大化现有和不存在三元组之间的余量的模型。我们将把这些模型称为基于分数的模型。如果需要，我们可以通过Platt缩放来推导基于分数模型的概率[50]。

有许多不同的方法来定义 f ，我们将在Section IV 和 V. 在我们继续讨论这些建模问题之前，我们讨论如何一般性地估计模型参数。

C. 惩罚性最大似然训练

假设我们有一组 N 观测三元组，并让第 n 个三元组表示为 x^n 。每个观察到的三元组为真（表示为 $y^n=1$ ）或为假（表示为 $y^n=0$ ）。让这个标记的数据集被表示为 $x^n, y^n, n=1, \dots, N$ 。显然地，估计参数 θ 的自然方式是计算最大后验概率（MAP）估计值：

$$\max_{\theta} \prod_{n=1}^N \text{Ber}(y^n | \sigma(f(x^n, \theta))) \prod_{n=1}^N \exp(-\lambda \|\theta\|_2^2) \quad (3)$$

λ 控制先前的强度。（如果先验是一致的，这相当于最大似然训练。）我们可以等价地说这是一个正则化的损失最小化问题：

$$\min_{\theta} \sum_{n=1}^N L_p(\sigma(f(x^n, \theta)) - y^n) + \lambda \|\theta\|_2^2 \quad (4)$$

其中 $L_p(p, q) = -\log \text{Ber}(p | q)$ 是对数损失函数。另一种可能的损失函数是平方损失 $L_p(p, q) = (p - q)^2$ 。我们在下面讨论其他一些损失函数。

D. 负面例子来自哪里？

一个重要的问题是标签 y^n 来自哪里。问题是大多数知识图只包含积极的训练的例子，因为通常，他们不编码

错误的事实。因此对于所有的 $x^n, y^n=1$ 。为了强调这一点，我们将使用符号 x^n 来表示观察到的正（真）三元组： $D = \{x^n | y^n=1\}$ 。对所有正面

一种解决方法是假设所有（类型一致的）不在 D 中的三元组都是假的。我们将表示这个否定集 x^n

$D = \{x^n | x^n \notin D\}$ 不幸的是，可能是非常大量的，因为虚假事实的数量远大于真实事实的数量。产生负面例子的一个更好的方法是“干扰”真正的三元组。特别是，让我们来定义

$$D' = \{x^n | x^n \in D \text{ 且 } x^n \neq x^n\} \cup \{x^n | x^n \notin D \text{ 且 } x^n \neq x^n\}$$

为了理解这种方法和前一种方法（我们假定所有有效的未知三元组都是假的）之间的区别，让我们考虑图中的例子1. 该

(LeonardNimoy, starredIn, StarWars), (AlecGuinness, starredIn, StarTrek)等，而且还有类型一致的但是“不相关”的负面三元组，例如(BarackObama, starredIn, StarTrek)等等（我们假设（为了这个例子）有一个类型的Person，但不是一个类型的Actor）。第二个方法（基于扰动）不会产生消极的三元组，如(BarackObama, starredIn, StarTrek)，因为BarackObama不参与任何出演的事件。这减小了大小，并鼓励它把重点放在“合理的”消极方面。（一个更好的方法，用在Section VIII, 是从Web上运行的文本提取方法生成候选三元组。由于提取错误，这些三元组中的许多将是错误的，但是它们定义了一组好的“合理的”负数。）

假设KG缺失的三元组是假的，称为封闭世界假设。正如我们之前所说，这并不总是有效的，因为该图表是不完整的。更准确的方法是做一个局部封闭的世界假设（LCWA）[51, 24]。使用LCWA更接近开放世界的假设，因为它只假定一个KG在当地完成。更准确地说，如果我们已经观察到特定主谓关系 ei, rk 的任何三元组，那么我们将假设任何不存在的三重 ei, rk 确实是假的。（这个假设对于函数关系是有效的，比如出生的，而不是集合关系，比如starredIn）。然而，如果我们还没有观察到任何三元组对于 ei, rk ，我们将假设所有的三元组 pei, rk, q 不详。

E. 成对丢失训练

鉴于负面的训练的例子并不总是

真的是消极的，另一种可能性训练的方法是尝试做出概率（或者一般来说，一些得分函数）对于真三元组比对假三元组要大。也就是说，我们可以定义以下目标函数：

$$\min_{\theta} \sum_{n=1}^N L_p(f(x^n, \theta) - y^n) + \lambda \|\theta\|_2^2 \quad (5)$$

其中 L_p, f 是一个基于边际的排序损失函数如

的数据进行培训是非常棘手的，因为这个模型可能会过度泛化。

$$Lpf, \text{fq} \text{ "maxpl`f'f, } 0q_0 \quad (6)$$

这种方法有几个优点。首先，它并不假定消极的例子必然是消极的，只是它们比积极的消极“更负面”。其次，它允许fpq函数是任何函数，而不仅仅是概率（但是

我们假设更大的 f 值意味着三元组更可能是正确的)。

这种目标函数很容易通过随机梯度下降 (SGD) [52]: 在每次迭代中, 我们只是采样一个正面和一个负面的例子。SGD也可以很好地适应大型数据集。但是, 可能需要很长时间才能收敛。另一方面, 我们将在下面看到, 一些模型与平方损失目标相结合, 可以使用交替最小二乘法 (ALS) 进行优化, 这通常要快得多。

F. 知识图的统计特性

知识图通常遵循一定的确定性

规则, 如类型约束和传递性 (例如, 如果Leonard Nimoy出生在波士顿, 而Boston位于美国, 那么我们可以推断出Leonard Nimoy出生在美国)。然而, 幼稚园通常也有各种“较软”的统计模式或规律, 这些并不普遍, 但却具有有用的预测能力。

这种统计模式的一个例子就是同质的, 也就是说, 实体与具有相似特征的其他实体相关的趋势。这已经在各种社交网络中被广泛观察到[53, 54]。在我们的例子中, 一个同质模式的例子可能是Leonard Nimoy更有可能出演在美国制作的电影。对于多关系数据 (具有多种链接的图形), homophily也被称为自相关 [55]。

另一个统计模式被称为块结构。这是指实体可以分成不同的组 (块) 的属性, 这样一个组的所有成员都有与其他团体成员的类似关系 [56, 57, 58]。

例如, 一个块结构的例子可能是科幻小说演员更可能出演科幻小说

电影, 其中的科幻小说演员由伦纳德·尼莫伊 (Leonard Nimoy) 和亚历克·吉尼斯 (Alec Guinness) 等实体以及“星际迷航” (Star Trek) 和“星球大战” (Star Wars) 等实体的科幻电影组成。

图还可以表现出全局和长期的统计依赖性, 即, 可以跨越三元组链并涉及不同类型关系的依赖关系。例如, 伦纳德·尼莫伊 (Leonard Nimoy) (美国) 的公民身份在统计上取决于他出生的城市 (波士顿), 而这种依赖关系涉及多个实体 (伦纳德·尼莫伊, 美国波士顿) 和关系 (出生于, 位于, 公民)。关系型学习的一个显着特征就是能够利用这种模式来创建更丰富, 更准确的关系型域模型。

IV. 潜在的特征模型

在本节中, 我们假设变量 y_{ijk} 是条件独立的, 给定了一组全局的潜在特征和参数, 如等式(1)所示。我们讨论各种可能的形式为得分函数 $f(x; \Theta)$ 下面。什么模型

$$f(x; \Theta)$$

有共同之处在于它们通过实体的潜在特征来解释三元组。例如, 对于亚历克·吉尼斯获得奥斯卡奖的事实的一个可能的解释是他是

一个好演员。这个解释使用实体的潜在特征 (作为一个好演员) 来解释可观察的事实 (吉尼斯获得奥斯卡奖)。我们称这些功能是“潜在的”, 因为它们不是直接在数据中观察到的。所有潜在特征模型的一个任务是从数据中自动推断这些特征。

在下文中, 我们将用向量 $e_i \in \mathbb{R}^H$ 来表示实体 e_i 的潜在特征表示, 其中 H 表示模型中潜在特征的数量。例如, 我们可以模仿亚历克吉尼斯是一个很好的演员, 而奥斯卡奖是通过载体享有声望的奖项

$$e_{\text{Guinness}} = \begin{bmatrix} 0.9 \\ 0.2 \end{bmatrix}, \quad e_{\text{学术奖}} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$$

其中组件 e_{i1} 对应于潜伏特征Good Actor, 并且 e_{i2} 对应于声望奖。(请注意, 与本例不同, 以下模型推断的潜在特征通常很难解释。)

关系潜在特征模型的关键直觉是实体之间的关系可以从它们的潜在特征的相互作用中推导出来。然而, 有很多可能的方法来模拟这些相互作用, 并有许多方法从中推导出一个关系的存在。下面我们讨论几种可能性。见表III对符号进行总结。

A. RESCAL: 双线性模型

RESCAL [59, 60, 61]是通过潜在特征的成对相互作用解释三元组的关系式潜在特征模型。特别是, 我们模拟三重 x_{ijk} 的得分为

$$f_{ijk}^{\text{RESCAL}} = \sum_p W_{pab} e_{ia} e_{jb} \quad (7)$$

其中 $W \in \mathbb{R}^{K \times K}$ 是权重矩阵, 其条目 w_{abk} 指定潜在特征 a 和 b 在第 k 个关系中相互作用的量。我们称之为双线性模型, 因为它使用乘法项捕获两个实体向量之间的相互作用。例如, 我们可以通过一个权重矩阵来模拟好的演员可能获得有声誉的奖项的模式

$$W_{\text{receivedAward}} = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.1 \end{bmatrix}$$

一般来说, 我们可以通过 W_k 条目的大小对块结构模式进行建模, 而我们可以通过对角条目的大小对同构模式进行建模。这些模式中的反相关可以通过 W_k 中的否定条目有效建模。

因此, 在等式(7) 我们通过实体 e_i 和 e_j 的潜在特征之间的所有成对相互作用的加权和来计算三重 x_{ijk} 的得分。参数

模型是 $\Theta = \{e_i, e_j, W_k\}$, 在训练期间我们共同学习潜在特征 e_i 和 e_j 以及权重 W_k 。

表示

潜在特征针对特定关系类型进行交互。

在下文中, 我们将讨论从知识图学习模型的进一步重要性质。

表三
符号概述。

符号	含义	关系数据
N_e	实体数量	
N_r	关系数	
N_d	培训例子的数量	
e_i	数据集中的第 <i>i</i> 个实体 (例如LeonardNimoy)	
r_k	数据集中的第 <i>k</i> 个关系 (例如, bornIn)	
D^+	一套观察到的积极的三元组	
D^-	一组观察到的负三元组	

符号	含义	尺寸
\mathbf{Y}	(部分观察) 所有三元组的标签	$N_e \times N_e \times N_r$
\mathbf{E}	得分为所有可能的三倍	$N_e \times N_e \times N_r$
\mathbf{Y}_k	关系 <i>r_k</i> 的Y片	$N_e \times N_e$
\mathbf{E}_k	关系 <i>r_k</i> 的E片	$N_e \times N_e$

图形和潜在特征模型

符号	含义	尺寸
ϕ_{ijk}	三元组 <i>e_i</i> , <i>r_k</i> , <i>e_j</i> 的特征向量表示	$p \times q$
ω	权重向量得出关系 <i>k</i> 的分数的	
Θ	设置模型的所有参数	
σ_p^q	Sigmoid (后勤) 功能	

符号	含义	尺寸
h_e	实体的潜在特征数量	
h_r	关系的潜在特征的数量	
e_i	潜在的功能repr. 实体 <i>e_i</i>	h_e
r_k	潜在的功能repr. 关系 <i>r_k</i>	h_r
H_a	哈层的大小	
H_b	hb层的大小	
H_c	hc层的大小	
\mathbf{E}	实体嵌入矩阵	$N_e \times h_e$
	关系 <i>k</i> 的双线性权重矩阵	$h_e \times h_r \times h_c$
	实体对的线性特征映射	$p_2 \times h_e \times q_2$
\mathbf{C}	三元组的线性特征映射	$p_2 \times h_e \times q_2 \times h_c$

通过共享表示进行关系学习: 在等式(7), 不管实体具有相同的潜在表示, 它们是否作为关系中的主体或客体出现。此外, 他们有相同的代表

不同的关系类型。例如, 第*i*个实体出现在三重*x_ijk*中, 作为类型为*k*的关系的主体, 而在三重*x_piq*中则作为类型*q*的关系的对象出现。然而, 预测*f_{ijk}*和*f_{piq}*都使用第*i*个实体的相同潜在表示*e_i*。由于所有参数是共同学习的, 这些共享表示允许通过实体的潜在表示和关系的权重在三元组之间传播信息。这允许模型捕获数据中的全局依赖关系。

语义嵌入: RESCAL中的共享实体表示还捕获关系域中实体的相似性, 也就是说, 如果实体通过类似的关系连接到相似的实体, 则它们是相似的[61]。例如, 如果*e_i*和*e_p*的表示相似, 则预测*f_{ijk}*和*f_{pjk}*将具有相似的值。作为回报, 具有许多类似观察关系的实体将具有相似的潜在表示。这个属性可以用于实体解析, 也可以对关系数据进行大规模的层次聚类[59,60]。而且, 因为关系

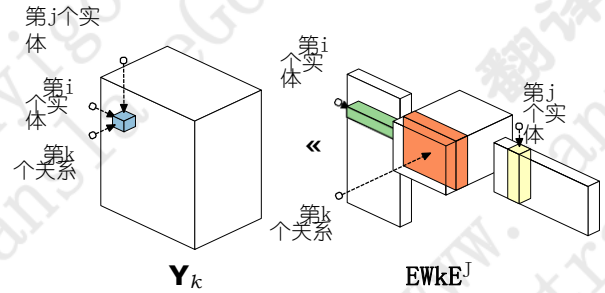


图3. RESCAL作为邻接张量Y的张量因式分解。

相似度通过向量的相似性来表示, 潜在表示*e_i*可以作为代理来给出非关系机器学习算法, 如*k*-均值或内核方法访问实体的关系相似度。

与张量分解的连接: RESCAL类似

到推荐系统中使用的方法[62]和传统的张量分解方法[63]。在矩阵符号中, 等式(7)可以写成如*F_k* *E_k* *E^T*那样紧凑, 其中*F_k* *Rⁿⁿⁿ*是保存第*k*个关系的所有分数的矩阵, *E_k*的第*i*行保持*e_i*的潜在表示。见图3为例证。在下面, 我们将使用这个张量表示来推导一个非常有效的参数估计算法。

拟合模型: 如果我们要计算完整的概率模型, 可以通过使用基于梯度的方法(如随机梯度下降)使对数损失最小化来估计RESCAL的参数[64]。

但是, 如果我们不需要完全的概率模型, 我们可以使用另一种方法来非常有效地估计参数。首先, 让我们做一个封闭的世界假设, 即我们假设*y_{ijk}*的每个值都是已知的, 并且是0或者

1. 其次, 我们将使用平方损失来代替对数损失。因此我们有以下优化问题:

$$\sum_{i,j,k} \{ \mathbf{e}_i^T \mathbf{W}_k \mathbf{e}_j - y_{ijk} \}^2 + \lambda \sum_{i,j,k} \{ \mathbf{e}_i^T \mathbf{W}_k \mathbf{e}_j \}^2. \quad (8)$$

$\lambda \geq 0$, $\lambda \geq 0$ 控制正则化的程度。方程的主要优点(8)它可以通过交替最小二乘法(ALS)进行优化, 其中潜在因子*E*和*W_k*可以通过一系列非常有效的交替闭式更新来计算[59,60]。我们称之为algorithm RESCAL-ALS。

分析表明, RESCAL-ALS中*E*和*W_k*的单一更新与实体*N_e*的数量呈线性关系, 与关系数量*N_r*成线性关系, 与观察到的三元数量成线性关系, *Y*中的零个条目[60]。在实践中, 迭代更新的一小部分(比如30到50)通常足以使算法达到对参数的稳定估计。

给定*E*的当前估计, 可以并行计算每个*W_k*的更新以提高具有大量关系的知识图的伸缩性。此外, 通过利用RESCAL的特殊张量结构, 我们可以为RESCAL-ALS导出更好的更新, 这些更新用于计算单个更新的运行时复杂度为 $O(p^2 h_e q)$ 的参数估计值(与运行时复杂度

H⁵的更新) [61, 65]。因此, 对于可以通过中等数量的潜在特征来解释的关系域, RESCAL-ALS具有高度可扩展性并且计算速度非常快。

解耦预测: 在方程中 (7), 在 H^2 时间内通过简单的矩阵-向量积来计算单一关系的概率。因此, 一旦参数已经估计, 预测任意分数的计算复杂性仅取决于潜在特征的数量, 并且与图的大小无关。然而, 在参数估计期间, 由于共享的潜在表示, 模型可以捕获全局依赖性。

关系学习结果: RESCAL已被显示

在一些关系学习任务上取得最新的成果。例如, [59]显示RESCAL

与马尔可夫逻辑网络（具有结构学习）相比，在许多小基准数据集上提供了可比的或更好的关系预测结果[66]，无限（隐藏）关系模型[67, 68]和贝叶斯聚类张量分解[69]。此外，RESCAL已被用于整个知识图的链接预测，如YAGO和DBpedia [60, 70]。除了链路预测之外，RESCAL还成功应用于实体分辨率和链路聚类类SRL任务。例如，RESCAL在出版物数据库中预测了哪些作者，出版物或出版场所可能是相同的最新结果[59, 61]。此外，由RESCAL计算的实体的语义嵌入已经被利用来通过嵌入空间中的实体的分层聚类来为未分类的数据创建分类[71]。

B. 其他张量分解模型

已经探索了各种其他的张量因子分解方法，用于从知识图和多关系数据中学习。 [72, 73]使用CP分解邻接张量张量分解来分析Web的链接结构页面和语义Web数据。 [74]应用

成对相互作用张量因式分解[75]来预测三元组

在知识图中。 [76]在推荐设置中将分解机器应用于大型单一关系数据集。 [77]

提出了一个具有很多不同关系的知识图的张量分解模型。

也可以使用离散的潜在因素。 [78]提出了布尔张量因子分解来消除提取的事实

与OpenIE方法, 并将其应用于大型数据集[79]. In

与先前讨论的分解对比，布尔张量分解是离散模型，其中邻接张量被分解为基于布尔代数的二元因子。

从知识图学习的另一种方法是基于矩阵分解，其中在因式分解之前，邻接矩阵 $Y \in \mathbb{R}^{N_e \times N_v}$ 通过将行与主 - 客体对相关联而重构成矩阵 $Y \in \mathbb{R}^{N_e \times N_r}$

pe_i , q_j 和关系 rk 的列 (参见 [80, 81]) 或矩阵 R^{MeNe} 通过将行与主体 pe_i 和列与关系/对象 rk, ej 关联 (参见 [82, 83])。不幸的是, 与张量分解相比, 这两个配方都会丢失信息。例如, 如果每个主客体对都是通过不同的潜在表示建模的,

C. 多层感知器

我们可以将RESCAL解释为创建三元组的复合表示并从这个表示中推导出它们的存在。特别是，我们可以将RESCAL重写为

$$f_{ijk}^{\text{RESCAL}} = w_j \phi_{ij}^{\text{RESCAL}} \quad (9)$$

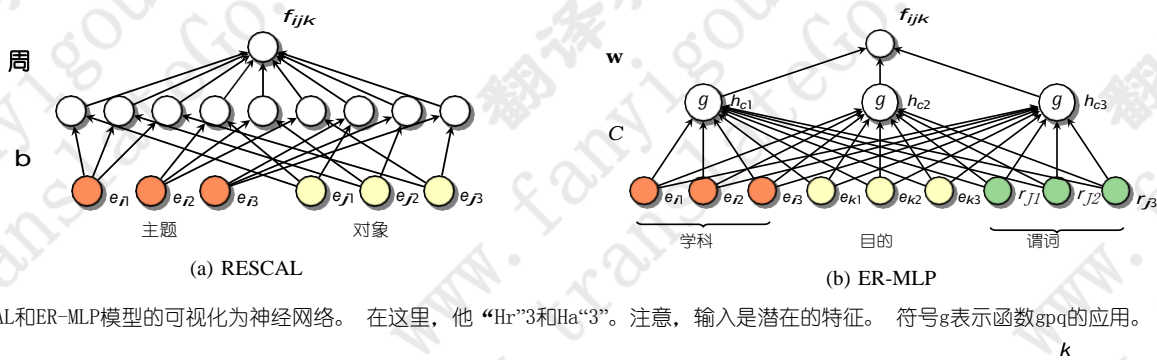


图4. RESCAL和ER-MLP模型的可视化神经网络。在这里，他“Hr”3和Ha“3”。注意，输入是潜在的特征。符号g表示函数gpq的应用。

表IV
KV-MLP在FREEBASE上的语义嵌入

关系	最近的邻居			
孩子	父母 (0.4)	伴侣 (0.5)	出生地 (0.8)	
生日	儿童 (1.24)	性别 (1.25)	父母 (1.29)	
EDU-结束 ²	工作开始 (1.41)	edu-start (1.61)		
	作业结束 (1.74)			

很多参数。另一种方法是使用Hr维矢量rk嵌入关系本身。我们可以定义

$$f_{ijk}^{\text{ER-MLP}} : \text{“w}^J \text{gphc} \text{q} \text{”} \quad (14)$$

$$\text{hc} : \text{“C}^J \phi^{\text{ER-MLP}}_{ijk} \text{”} \quad (15)$$

$$\phi_{ijk}^{\text{ER-MLP}} : \text{“Rei; 没有; RKS.} \text{”} \quad (16)$$

我们称这个模型为ER-MLP，因为它应用了MLP嵌入实体和关系。请注意，ER-MLP为所有关系使用全局权重向量。这个模型在KV项目中使用（参见章节VIII），因为它具有比E-MLP更少的参数（见表I）V；原因是C与关系k无关。

已经在[87] MLP可以学习将“语义相似”的单词靠近嵌入空间，即使他们没有经过专门的训练。在[24]，他们展示

对于关系使用的语义嵌入类似的结果ER-MLP。特别是表IV显示最近的邻居已经选定的关系的潜在表征

用Freebase上的60维模型计算。括号内的数字表示平方欧几里德距离。可以看出，ER-MLP将语义相关关系放在彼此附近。例如，与子女关系最亲密的关系是父母，配偶和出生地。

D. 其他神经网络模型

我们也可以将传统的MLP与双线性模型结合起来，从而产生什么[89]称为“神经张量网络”（NTN）。更确切地说，我们可以将NTN模型定义如下：

$$f_{ijk}^{\text{NTN}} : \text{“W}^J \text{gprha; hbsq} \text{”} \quad (17)$$

$$\text{hb} : \text{“} h_K^1 \cdot \cdot \cdot k^1, H^{\text{血红蛋白}} \text{”} \quad (18)$$

$$h_k : \text{“e}^J \text{B} \phi_{k,j} \text{”} \quad (19)$$

²关系教育启动，教育结束，工作启动，工作结束分别代表参加教育机构和举办特定工作的开始和结束日期

这里 B_k 是张量，其中第B片的尺寸和 H_e ，并且有 H_b 片。这个模型结合了RESCAL的双线性表示和E-MLP的加法模型。NTN模型比MLP或RESCAL（双线性）模型有更多的参数。事实上，[91]和[24]都表明它倾向于过度拟合，至少在在这些论文中使用的（相对较小的）数据集。

除了诸如RESCAL，MLP和NTN之类的乘法模型之外，已经提出了具有潜在特征之间的加性相互作用的模型。特别是潜伏的距离

模型（也称为社交网络中的潜在空间模型分析）从中得出关系的概率

实体潜在在表示之间的距离：实体

如果它们的潜在表示根据某种距离度量是接近的，则可能处于关系中。对于单一关系数据，[92]首先在社会网络的背景下提出这种方法，通过建模关系 x_{ij} 的概率通过得分函数 $f_{ei, ej}^d$ ， $f_{ei, ej}^d$ 其中d是指任意的距离度量，例如欧几里德距离。结构化嵌入（SE）模型[88]通过将三重 x_{ijk} 的得分建模为：将这个想法扩展到多关系数据：

$$f_{ijk}^{\text{SE}} : \text{“A}^s \text{by}_k^s \text{A}_i^s \text{e}^s \text{”} \quad (20)$$

其中 A_k^s 是 A^s 的 k 个转换， A^s 是 A 的 s 个转换。在等式(20)中， A^s 是 A 的 s 个转换， A_k^s 是 A^s 的 k 个转换。

专门地为第k个关系建模关系。这种转换是通过使用排名损失来学习的，使得现有关系中的成对实体比不存在关系中的实体彼此更接近。

为了减少SE模型上的参数数量，TransE模型[90]通过关系特定的偏移来转换潜在的特征表示，而不是通过矩阵乘法对它们进行转换。特别是，三重 x_{ijk} 的得分被定义为：

$$f_{ijk}^{\text{TransE}} : \text{“D} \phi_{ei}^r k, ej \text{q} \text{”} \quad (21)$$

这个模型受到了[87]，谁表示

词之间的某些关系可以通过它们在嵌入空间中的向量差来计算。正如在[90]，

在 ei, ej 的单位范数约束下，使用平方欧几里德距离，我们可以重写方程(21)如下：

$$f_{ijk}^{\text{TransE}} : \text{“p} 2r^J \text{p} ei^r ej^r \text{q}^r 2e^J \text{e}_u^r \text{rk} \text{”}^2 \text{q} \quad (22)$$

表五
内容特征模型概述。 ha, hb和hc是隐藏网络的隐藏层；看到文字的细节。

方法	f	隐藏的功能地图	隐藏的交互阿克	参数数量
结构化嵌入[88]	' } 公顷} 1	$rA_k^{sA}; o_s$	-	$2Nl \text{ HeHa } ^\circ \text{NeHe}$
E-MLP [89]	$w_{k,j}^{sphaq}$	$rA_k^{sA}; A_k^{o_s}$	-	$Nr \text{ 他们是 } 2 \text{ 头}$
ER-MLP [24]	$w_{k,j}^{gphcq}$	-	C	$H \text{ Ne } Ne \text{ NrHe}$
TransE [90]	' p2ha' 2hb` } rk} 2q 2	$rRk \text{ 生存}; 'rk s$	-	$Nr \text{ He } Ne \text{ Ne}$
双线 (RESCAL) [60]	$w_{k,j}^{hb}$	-	-	$Nr \text{ He } ^\circ \text{NeHe}$
NTN [89]	$w_{k,j}^{gprha}; hbsq$	$rA_k^{sA}; A_k^{o_s}$	-	$N \text{ 从来没有}$

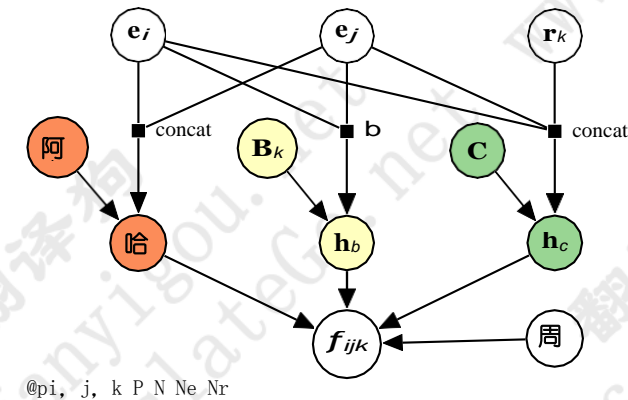


图5. 本文讨论的神经网络的图形模型。

此外，如果我们假设 A_k^{rrk} ； H_b^{1} 和 $B_k^{“}$ ，那么我们可以重写这个模型如下：

$$F_{IJK}^{TransE} = \text{“} \text{ } p2ha' 2hb` \} rk\} 2q_0. \quad (23)$$

我们讨论的所有模型都显示在图中5. 见表V 作总结。 [91]对这些模型进行实验比较；他们表明，RESCAL模型，特别是它的一个对角版本，在两个不同的链接预测任务上表现最好。（他们在那篇文章中没有考虑ER-MLP，但是这在 [24]，大概是由于过度配合。在更大的数据集上进行更详细的实验比较是留待未来的工作的。）

V. 图形特征模型

在本节中，我们假设边缘的存在可以通过从图中观察到的边缘提取特征来预测。例如，由于社会惯例，一个人的父母经常结婚，所以我们可以预测三重（John, marriedTo, Mary）的存在

路径约翰 John 安妮 Anne 玛丽 Mary ，代表一个共同的孩子。与潜在特征模型不同的是，这种推理通过可观察变量来解释三元组，即直接从知识图中观察到的三元组中解释三元组。现在我们来讨论一下这种模型。

A. 单一关系数据的相似性度量

可观察图形特征模型被广泛用于仅包含单一关系的图形中的链接预测，例如，

如社交网络分析（人与人之间的友谊），生物学（蛋白质相互作用）和Web挖掘（网站之间的超链接）。这些方法背后的直觉是类似的实体可能是相关的（同源的），并且实体的相似性可以从节点的邻域或从节点之间的路径的存在中导出。为此，已经提出了各种指标来衡量实体的相似性，这些指标可以分为局部，全局和准局部方法[93]。

本地相似指数，如共同邻居，亚当 - 阿达尔[94]或优惠依恋[95]从它们的普通邻居数量或邻居的绝对数量中推导出实体的相似性。局部相似

对于单一的关系和规模，指数的计算速度很快以及计算依赖的大型知识图

只在涉及的实体的直接邻居。但是，它们可能过于本地化，无法捕获关系数据中的重要模式，也无法模拟远程或全局依赖关系。

全球相似性指数如Katz指数[96]和 *Leicht-Holme-Newman*指数 [97] 从实体之间所有路径的集合中推导实体的相似性，而诸如命中时间，通勤时间和PageRank [98]从图上的随机行走中导出实体的相似度。全球相似性指数往往比本地指数提供更好的预测，但也是计算上更昂贵的 [93, 54]。

准局部相似性指数，如Local Katz指数[54]或本地随机游走[99]尝试通过从路径和随机游走的有限长度中推导出实体的相似性来平衡预测精度和计算复杂度。

在下文中，我们将讨论一种扩展这种准局部相似度指标的方法，以便从关系型大关系知识图中学习单关系网络。

B. 路径排序算法

路径排序算法 (PRA) [100, 101]扩展了使用有界长度的随机游走来预测多关系知识图中的链接的想法。特别地，设 $\pi_{Li, j, k, t}$ 表示形式 $e_1 e_2$ 的长度为 L 的路径 $e_1 \dots e_j$ ，其中 t 表示边缘类型 t_1, t_2, \dots 的序列 t_1, \dots, t_L 。我们还要求存在一个直接弧 $e_i e_j$ ，表示从 e_i 到 e_j 存在一个类型 k 的关系。令 $\Pi_{Li, j, k}$ 表示长度为 L 的所有这些路径的集合，其范围遍及路径类型 t 。（我们可以发现这样的

表六
PRA在免费基础上学习的路径示例预测哪个大学的学生

关系路径	F1	Prec	Rec	w_{ka}
(被起草, 学校)	0.03	1.0	0.01	2.62
(兄弟姐妹, 兄弟姐妹, 教育机构) 0.05 0.55 0.02 1.88				
(配偶, 配偶, 教育机构) 0.06 0.41 0.02 1.87				
(家长, 教育机构)	0.04	0.29	0.02	1.37
(儿童, 教育, 机构)	0.05	0.21	0.02	1.85
(placeOfBirth, peopleBornHere, 教育) 0.13 0.1 0.38 6.4				
(类型, 实例, 教育, 机构)	0.05	0.04	0.34	1.74
(职业, peopleWithProf., edu., inst.) 0.04 0.03 0.33 2.19				

通过枚举类型为ei的实体的所有（类型一致的）路径到类型为ej的实体。如果有太多的关系可以做到这一点，我们可以进行随机抽样。）

我们可以通过假定在每个步骤中随机统一地遵循一个出口来计算遵循这种路径的概率。令 π_{ijk} 为特定路径的概率；这可以通过采样过程递归地计算，类似于PageRank（参见[101]的细节）。PRA中的关键思想是将这些路径概率用作预测丢失边缘概率的特征。更确切地说，定义特征向量

$$\phi_{ijk}^{PRA} = \pi_{ijk} \quad (24)$$

然后，我们可以使用逻辑回归来预测边缘概率：

$$f_{ijk}^{PRA} = w_k^J \phi_{ijk}^{PRA} \quad (25)$$

可解释性：PRA的一个有用特性是其模型易于解释。特别是，关系路径可以被看作是加权规则的主体 - 更确切地说是霍恩（Horn）从句 - 权重指定规则的主体是如何预测的。比如表 VI 显示了PRA在KV项目中学习到的一些关系路径及其权重VIII) 预测一个人参加哪个大学，即预测形式的三倍 (p, college, c)。表中的第一个关系路径 VI 可以解释如下：如果起草这个人的运动队来自同一所大学，那么这个人可能就读了大学。这可以写成Horn子句的形式如下：

$$(p, college, c) \leftarrow (p, draftedBy, t) \wedge (t, school, c)。$$

通过在wk上使用稀疏推理之前，我们可以进行特征选择，这相当于规则学习。

关系学习结果：PRA已被证明优于确定性关系学习方法FOIL [102]用于NELL中的链接预测[101]。在KV的链路预测中也显示出与ER-MLP相当的性能：PRA在ROC曲线下面积得到0.884的结果，而ER-MLP的结果为0.882 [24]。

VI. 合并潜在和图形特征模型

已经通过实验观察到（参见例如[24]）

最先进的关系潜在特征模型（RLFM）
也不是最先进的图形特征模型

从知识图表学习。相反，强项

潜在的和基于图形的模型往往是互补的，因为两个家族都关注关系数据的不同方面：

- 潜在特征模型非常适合通过新引入的潜在变量对全局关系模式进行建模。如果用少量的潜变量来解释三元组，那么它们在计算上是有效的。
- 图形特征模型非常适合建模局部和准局部图形模式。如果三元组可以从实体的邻域或图中的短路径来解释，它们在计算上是有效的。

还有一些比较这两种方法的理论工作[103]。特别是，已经表明，当关系数据由大量强连通分量组成时，张量分解可能是低效的。幸运的是，这种“有问题”的关系通常可以通过基于图形的模型有效地处理。婚姻关系就是一个很好的例子：一个婚姻对应一个单一的强连通的组成部分，因此具有大量婚姻的数据将难以用RLFM进行建模。然而，通过基于图的模型预测婚姻链接是容易的：通过利用关系的对称性，可以简单地从 (Mary, marriedTo, John) 的存在预测三元组 (John, marriedTo, Mary) 的存在。如果 (Mary, marriedTo, John) 边缘是未知的，我们可以使用统计模式，比如共享孩子的存在。

结合潜在的和基于图形的模型的优势因此是增加预测的有希望的方法

图模型的性能。它通常也加快了速度训练。我们现在讨论一下结合这两者的一些方法种类型号。

A. 加性关系效应模型

[103]提出了加性关系效应（ARE），这是一种将RLFM与可观测图模型相结合的方法。

特别是，如果我们将RESCAL和PRA结合起来，就可以得到

$$f_{ijk}^{RESCAL + PRA} = w_k^{p1qJ} \phi_{ij}^{RESCAL} + w_k^{p2qJ} \phi_{ijk}^{PRA} \quad (26)$$

通过用PRA参数交替优化RESCAL参数，可以训练ARE模型。关键的好处是现在RESCAL只需要模拟无法用可观察的图形模式建模的“残差”。这允许该方法使用更低的潜在维度，这显着加快了训练时间。由此产生的组合模型也提高了准确性[103]。

B. 其他组合模型

除了ARE之外，还探索了进一步的模型，以便从关系数据的潜在和可观察模式中共同学习。[80]将一个潜在特征模型和一个加性术语相结合，从多个关系数据的潜在和邻域信息中学习，如下所示：

$$f_{ijk}^{ADD} = w_k^{p1qJ} \phi_i^{SUB} + w_k^{p2qJ} \phi_j^{OBJ} + w_k^{p3qJ} \phi_k^N \quad (27)$$

$$\phi_{ijk}^N = \text{“Ryijk1: k1\%ks”} \quad (28)$$

这里, ϕ_i^{SUB} 是实体 e_i 的潜在表示
 学科 ϕ_j^{OBJ} 是实体的潜在表征
 和 ϕ_j
 目的。此外, [81] 考虑一个附加的术语

$$\hat{f}_{ijk}^{\text{UNI}} = \hat{f}_{ijk}^{\text{ADD}} + w^J \phi_k^{\text{SUB}} + \phi_{ij}^{\text{OBJ}} \quad (29)$$

其中 $\phi_k^{\text{SUB}} + \phi_{ij}^{\text{OBJ}}$ 是一个 (非复合) 潜在特征表示,
 主客体对。 ϕ^N 中的主要思想

方程(28) 和 (29) 是在三元 y_{ijk} 的存在预测同一对实体 (但不同的关系类型) 之间的另一个三元 y_{ijk} 的情况下有效地建模模式。例如, 如果伦纳德·尼莫伊 (Leonard Nimoy) 出生在波士顿, 那他也可能住在波士顿。关系类型 bornIn 和 livedIn 之间的这种依赖关系可以用等式来建模 (28) 通过给 wbornIn 分配一个很大的权重, aliveIn。

注意, 如果我们设置 ϕ_{ijk} , 这种邻域信息可以包含在 ARE 中

是等式中的力 w^{p2q} (26) 有一个零条目 w^{p2q} , 以避免使用 y_{ijk} 预测 y_{ijk} 。

ARE 和 [80] 和 [81] 在精神上与 [104], 其增加具有添加项的评分矩阵的 SVD (即, 矩阵分解) 以包括当地邻居信息。同样, 分解机器 [105] 允许结合潜在的和可观察的模式, 通过低阶因子分解对输入变量之间的高阶相互作用建模 [74]。

结合不同预测系统的一种替代方法是单独拟合它们, 并将其输出用作另一个“融合”系统的输入。这被称为堆叠 [106]。例如, [24] 使用 PRA 和 ER-MLP 的输出作为标量特征, 并通过训练二元分类器学习最终的“融合”层。堆叠的优点是在可组合的各种模型中非常灵活。然而, 它具有个体模型不能合作的缺点, 因此任何个体模型都需要比联合训练的组模型更复杂。例如, 如果我们将 RESCAL 与 PRA 分开安装, 那么我们将需要比联合使用它们更多的潜在特征。

VII. 马尔科夫随机场

在马尔科夫随机场 (MRF) 中, 假定 y_{ijk} 局部相互作用, 即在图中接近的不同 y_{ijk} 之间存在相互作用。交互是通过潜在的功能建模的, 可以用各种方式来定义。常用的方法是使用“马尔科夫逻辑” [107], 这是一种模板语言, 用于通过逻辑公式定义任意大小的依赖关系图 (有时称为“关系型 MRF”) 的潜在函数。

MRF 是建模幼儿园的有用工具, 如 [108]。然而, 他们面临着一些计算问题: 规则学习的难度, 预测的难度以及估计参数的难度。规则学习问题已经在各种论文中进行了研究 (参见章节 V-B 和 [51, 91]), 但总的问题是棘手的。

预测/推理问题 (计算 MAP 状态估计值) 通常也是 NP 难的。在图形模型文献中已经提出了对这个问题的各种近似, 例如吉布斯 (Gibbs) 采样 (参见例如 [26, 109])。

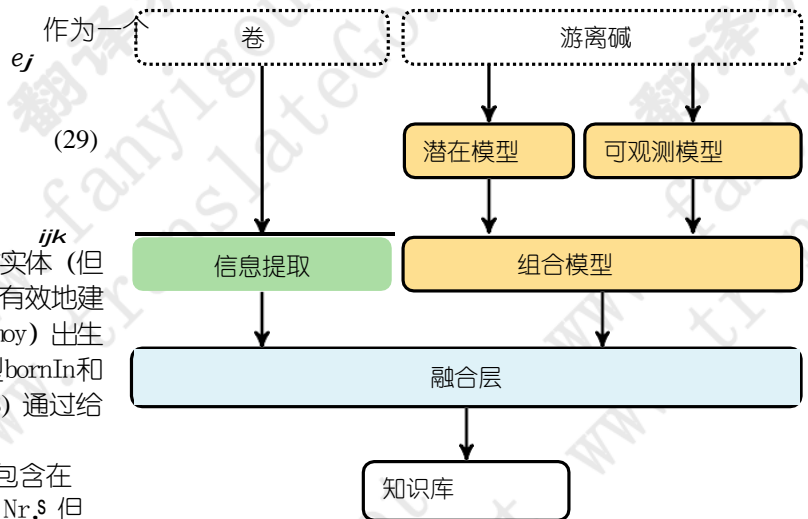


图6. 知识库的体系结构

一个有趣的近期方法, 称为概率软逻辑 (PSL) [110] 是基于不断的放松。最终的系统可以扩展到相当大的知识库, 如图所示 [111]。

参数估计问题 (通常被称为最大似然估计或 MAP 估计) 虽然是凸的, 但通常相当昂贵, 因为它需要将预测称为子程序。已经开发了各种近似, 例如伪似然性 (参见 [112])。然而, 这些方法仍然没有成对损失最小化的灵活性。

总之, 虽然关系型 MRF 是一个有用的工具, 但它不像我们在本文中考虑的其他方法那么容易。

VIII. 知识库: 知识库建设的关系学习

知识库 (KV) [24] 是一个非常大规模自动构建的知识库, 它遵循 Freebase 模式 (KV 使用 4469 个最常见的谓词)。它分三步建造。第一步, 从 Web 源中提取来自不同 Web 源的内容, 例如文本, 表格数据, 页面结构和人类注释 (提取器在 [24])。其次, 在 Freebase 上训练 SRL 模型作为计算 (新) 边缘概率的“先验”。最后, 使用提取分数和先前的 SRL 模型评估自动提取的事实置信度。

KV 系统训练 PRA 模型和 ER-MLP 模型以预测知识图中的链接。如上所述, 这些使用堆叠进行组合。然后基于提取器的置信度, 三元组被找到的 (定义的) 网页的数量等等, 将来自融合链接预测模型的分数与从提取的三元组得到的各种特征组合。参见图 6 为例证。

我们现在给出一个结合事先与提取器 (即融合层) 的好处的定性例子

在图中6)。假设提取管道提取了一个对应于以下关系的三元组：³

(Barry Richter, 威斯康星大学麦迪逊分校)。

这个三元组（通过融合多种提取技术获得）的提取置信度仅为0.14，因为它基于以下两个相当间接的陈述：⁴

在1989年秋天，里希特接受了威斯康星大学的奖学金，在那里他玩了四年，赢得了许多个人的荣誉。

和⁵

极地赛事的原因得益于Andringa, Byce和前UW队友Chris Tancill和Barry Richter等知名教练的影响。

不过，我们从Freebase知道Barry Richter是在威斯康星州麦迪逊出生长大的。这增加了我们以前的信念，他在那里上学，导致最终的融合信仰0.61。

将先前的模型（使用SRL方法学习）与信息提取模型相结合可显着提高性能，将高置信度三元组（标定概率超过90%）从100M（仅基于提取器）增加到271M（基于提取器加之前）。迄今为止，这也许是SRL到KB的最大应用之一。看到[24]了解更多详情。

IX. 扩展和未来的工作

A. 非二元关系

到目前为止，我们完全关注二元关系。在这里我们讨论如何处理其他基数的关系。

一元关系：一元关系是指关于实体属性的陈述，例如一个人的身高。这样的数据自然可以用矩阵来表示，其中行代表实体，列代表属性。[60]提出了联合张量矩阵分解方法，通过共享实体的潜在表示，从二元和一元关系中同时学习。在这种情况下，我们可能还需要修改似然函数，所以对于二值边缘变量是伯努利，对于数字特征是高斯（比如说），对计数数据是Poisson（参见[113]）。

高阶关系：在知识图中，高阶关系通常通过多个二元关系表达。在部分II，(LeonardNimoy, Spock, StarTrek)通过两个二元关系(LeonardNimoy, 扮演Spock)和(Spock, characterIn, StarTrek)表达了三元关系扮演的角色In(LeonardNimoy, Spock, StarTrek)。不过，在不同的星际迷航电影中，有多个演员扮演Spock。为了在不丢失信息的情况下进行建模，我们可以使用辅助节点来识别呼叫关系。例如，建模关系

³为了清晰呈现，我们展示了一个简化的三元组。请参见[24]为实际提取的三元组，包括复合值类型(CVT)。

⁴来源：<http://www.legendsofhockey.net/LegendsOfHockey/jsp/SearchPlayer.jsp?播放器=11377>

⁵来源：http://host.madison.com/sports/high-school/hockey/数字的不断减少换曾经强大-麦迪逊高中曲棍球的程序/article_95843e00-ec34-11df-9da9-001cc4c002e0.html

(LeonardNimoy, Spock, StarTrek-1)，我们可以写

学科	谓词	目的
(LeonardNimoy, 演员, 电影, 1, (MovieRole-1, (MovieRole-1,	演员, 电影, 符号, 1) Spock)	MovieRole-1) StarTrek-1) Spock)

我们使用辅助实体MovieRole-1来唯一标识这个特定的关系。在大多数应用中，辅助实体得到一个标识。如果不是，则称它们为空白节点。在Freebase中，辅助节点被称为复合值类型(CVT)。由于涉及时间和地点的高阶关系比较常见，YAGO2项目将SP0三重格式扩展到(主体，谓词，对象，时间，位置)(SP0TL)格式，以便明确地建立关于关系的时间和空间信息，而不进行变换他们二元关系[23]。

一个相关的问题是，一个事实的真值可能会随着时间而改变。例如，谷歌现任首席执行官拉里·佩奇(Larry Page)，但从2001年到2011年是埃里克·施密特(Eric Schmidt)。这两个事实是正确的，但只在指定的时间间隔。出于这个原因，Freebase允许使用CVT(复合值类型)构造(通过辅助节点表示n元关系)来为开始和结束日期注释一些事实。将来，计划扩展KV系统来模拟这些时间事实。然而，这是不平凡的，因为从文本中推断事实的持续时间并不总是容易的，因为它不一定与相应源的时间戳有关(参见图1)。(114)。

作为辅助节点使用的替代方案，一组n阶关系可以由一个单一的n阶张量表示。RESCAL可以很容易地推广到更高阶的关系，可以通过高阶张量分解或神经网络模型来解决，相应的实体表示数目作为输入[113]。

B. 硬约束：类型，功能限制等等

对知识图中允许的三元组施加严格限制可能是有用的。强大的本体语言，如Web本体语言(OWL)[115]已经被开发出来，其中可以制定复杂的约束条件。然而，本体论的推理在计算上要求很高，并且在现实世界的数据中经常违反严格的约束[116, 117]。幸运的是，在证据矛盾的情况下，机器学习方法是可靠的。

确定性依赖关系：关系中的三元组，例如subClassOf和isLocatedIn遵循清晰的确定性依赖关系，如传递性。例如，如果Leonard Nimoy出生在波士顿，我们可以得出结论，他出生在马萨诸塞州，他出生在美国，出生在北美等地。考虑这种本体论约束的一种方法是预先计算所有可以从约束中派生出真正的三元组，并在学习之前将它们添加到知识图中。根据本体约束的三元组的预计算也被称为物化。但是，在大型知识图表上，完全实现可能在计算上要求很高。

类型约束：通常关系只适用于正确类型的实体。例如，marriageTo的域和范围仅限于个人的实体。明确建模类型约束需要复杂的手动工作。另一种方法是通过简单地考虑关系中观察到的主体和对象的类型来学习近似类型约束。标准RESICAL模型已经扩展了[70]和[65]来有效地处理关系的类型约束。结果，一个好的RESICAL模型所需的等级可以大大降低。此外，[81]考虑学习关系中的参数槽的潜在表示，从数据中学习正确的类型。

功能性限制和互斥性：虽然在章节讨论的方法IV和V可以对三元组之间的长期和全局依赖进行建模，但它们并没有明确强制实施可能的值之间相互排斥的功能约束。例如，一个人出生在一个城市，等等。如果观察到这些值中的一个，则可观察的图模型可以防止其他值被断言，但是如果所有的值都是未知的，那么所得到的互斥约束可以是很难以处理计算。

C. 推广到新的实体和关系

除了遗漏的事实之外，还有很多在网上被提及的实体，但是目前在像Freebase和YAGO这样的知识图表中缺少实体。如果将新的实体或谓词添加到KG，则由于运行时考虑，可能需要避免重新训练模型。给定当前模型和一组新观察到的关系，新实体的潜在表示可以在张量因子分解模型和神经网络中近似计算，通过找到解释相对于当前模型的新观察到的关系的表示。类似地，已经表明，在给定已经导出的实体的潜在表示的情况下，可以有效地计算RESICAL模型中的关系特定的权重 W_k [118]。

D. 查询概率知识图

RESICAL和KV可被视为概率数据库（参见例如[119, 120]）。在知识库中，只查询三元组的概率。一些应用程序可能需要更复杂的查询，如：谁出生在罗马，喜欢爱因斯坦的孩子。众所周知，涉及连接（存在量化的变量）的查询在概率数据库（[119]）。In [118]显示了如何在RESICAL框架内有效地处理一些涉及连接的查询。

X. 结束语

我们已经提供了对用于非常大的知识图的最先进的统计关系学习（SRL）方法的评论。我们还展示了统计关系学习如何与机器阅读和信息提取方法结合使用来自动构建知识库。因此，我们有

展示了如何创建一个真正的庞大的机器可解释的事实“语义记忆”，这已经赋予了许多重要的应用。然而，虽然这些幼儿园的规模令人印象深刻，但它们仍然没有代表人类所拥有的知识。值得注意的是，“常识”事实（例如水湿，潮湿的事物可能滑等事实）以及“程序性”或如何掌握的知识（如如何驾驶汽车，如何发送电子邮件等）。这种知识的表示，学习和推理仍然是AI和机器学习的下一个前沿。

致谢

马克西米利安镍承认由大脑，头脑和机器中心（CBMM）的支持，由NSF STC奖CCF-1231216资助。Volker Tresp承认德国联邦经济与能源部技术计划“智能数据”（授予01MT14001）的支持。

参考

- [1] L. Getoor和B. Taskar编辑，统计关系学习简介。MIT出版社，2007年。
- [2] S. Dzeroski和N. Lavrac，关系数据挖掘。Springer Science&Business Media，2001。
- [3] L. De Raedt，逻辑与关系学习。施普林格，2008年。
- [4] FM Suchanek, G. Kasneci和G. Weikum, “Yago: 语义知识的核心”，在万维网第十六届国际会议论文集。纽约，纽约州，美国：ACM，2007年，第697-706页。
- [5] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak和Z. Ives在“语义Web”中的“DBpedia: 一个开放数据网络的核心”。斯普林格柏林海德堡，2007年，第一卷。4825，第722-735页。
- [6] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, ERH Jr和TM Mitchell, “Toward an Architecture for Never-Ending Language Learning”，第二十四届人工智能会议论文集（AAAI 2010）。AAAI出版社，2010年，第1306-1313页。
- [7] K. Bollacker, C. Evans, P. Paritosh, T. Sturge和J. Taylor, 在2008年ACM SIGMOD数据管理国际会议论文集“Freebase: 一个协作创建的图形数据库，用于构建人类知识”。ACM，2008，第1247-1250页。
- [8] A. Singhal, “介绍知识图谱：事物，而不是字符串”，2012年5月。[在线]。可用：<http://googleblog.blogspot.com/2012/05/> 引入知识，图的事情，not.html
- [9] G. Weikum和M. Theobald在“第二十九届ACM SIGMOD-SIGACT-SIGART数据库系统原理研讨会论文集”中的“从信息到知识：从网络资源收集实体和关系”。ACM，2010年，第65-76页。
- [10] J. Fan, R. Hoffman, AA Kalyanpur, S. Riedel, F. Suchanek和PP Talukdar, “2012年AKBC-WEKEX: NAACL-HLT知识提取研讨会”，2012年。[在线]。可用：<https://开头akbcwekex2012.wordpress.com/>

- [11] R. Davis, H. Shrobe和P. Szolovits, “什么是知识表示?”《AI杂志》, 第一卷。 14, 没有。 1, 1993年第17-33页。
- [12] T. Berners-Lee, J. Hendler和O. Lassila, “The语义网”, 2001. [在线]。 可用: [HTTP://WWW.scientificamerican.com/article/the-semantic-web/](http://WWW.scientificamerican.com/article/the-semantic-web/)
- [13] T. Berners-Lee, “关联数据 - 设计问题”, 2006年7月。 [线上]。 可用: <http://www.w3.org/DesignIssues/LinkedData.html>
- [14] C. Bizer, T. Heath和T. Berners-Lee, “Linked data-the故事到目前为止”, 《国际期刊语义网和信息系统》, 第一卷。 5, 没有。 3, pp.1-22, 2009。
- [15] G. Klyne和JJ Carroll, “资源描述”框架 (RDF): 概念和抽象语法, “2月。 2004. [在线]。 可用: [http://www.w3.org/TR/2004/REC-RDF的概念 - 20040210 /](http://www.w3.org/TR/2004/REC-RDF的概念 - 20040210/)
- [16] R. Cyganiak, D. 木, 和 M. Lanthaler, “RDF 1.1概念和抽象语法”, 2月 2014. [在线]。 可用: [http://www.w3.org/TR/2014/REC-rdf11的概念 - 20140225 /](http://www.w3.org/TR/2014/REC-rdf11的概念 - 20140225/)
- [17] Y. Sun和J. Han, “挖掘异构信息网络: 原则和方法”, “综合数据挖掘和知识发现讲座, 第一卷。 3, 没有。 2, pp.1-159, 2012。
- [18] R. West, E. Gabrilovich, K. Murphy, S. Sun, R. Gupta, 和D. Lin, “通过搜索的知识库完成”基于问题的回答“, 在第23届会议记录《国际万维网大会, 2014年, 515-526。
- [19] DB Lenat, “CYC: 在中国的大规模投资”知识基础设施“, Commun. ACM, vol. 38, 没有。 11, 第33-38页, 1995年11月。
- [20] 遗传算法米勒, “WordNet: 词汇数据库英文”, Commun. ACM, vol. 38, 没有。 11, pp.39-41, 1995年11月
- [21] O. Bodenreider, “统一的医学语言系统 (UMLS): 整合生物医学学术语”, “核酸研究, 第一卷。 32, 没有。 数据库问题, 第D267-270, 2004年1月
- [22] D. Vrandečić和M. Krötzsch, “维基数据: 免费协作知识库”的通讯 ACM, vol. 57, 没有。 10, pp.78-85, 2014。
- [23] J. Hoffart, FM Suchanek, K. Berberich, 和 G. Weikum, “YAGO2: 空间和时间从维基百科增强了知识基础”, “人工”情报, 卷。 194, 第28-61页, 2013年。
- [24] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmman, S. Sun, and W. 张, “知识库: 一个网络规模的方法到概率知识融合“, 在Proceedings 1135-1145。
- [26] F. Niu, C. Zhang, C.R和J. Shavlik, “小学: 通过机器进行大规模的知识库建设学习和统计推断“, 国际期刊在语义Web和信息系统 (IJSWIS) 上 第一卷。 8, 没有。 3, 第42-73页, 2012。
- [27] N. Nakashole, M. Theobald和G. Weikum, “Scalable高精度, 高收获的知识回忆“, 第四届ACM国际会议论文集网络搜索和数据挖掘会议。 ACM, 2011年, 第227-236页。
- [28] A. Fader, S. Soderland和O. Etzioni, “识别关系 对于 打开 信息 萃取, ” in 美国经验方法会议论文集自然语言处理。 Stroudsburg, 宾夕法尼亚州, 美国: 计算语言学协会, 2011年, 1535-1545。
- [29] M. Schmitz, R. Bart, S. Soderland, O. Etzioni 等人, “用于信息提取的开放式语言学习”, 在 2012年联合会议的实证经验自然语言处理和Compu-自然语言学习。 协会 计算语言学, 2012年, 第523-534页。
- [30] J. Fan, D. Ferrucci, D. Gondek和A. Kalyanpur, “棱镜: 从大规模引导知识词汇化关系资源“ NAACL HLT 2010第一届国际研讨会 阅读学习的形式主义与方法论。 计算语言学协会, 2010年, 122-127。
- [31] B. Suh, G. Convertino, EH Chi和P. Pirolli, “The奇点不近: 维基百科发展缓慢 “ 在第五届国际研讨会论文集维基和开放协作。 美国纽约州纽约市: ACM, 2009, 第8: 1-8: 10页。
- [32] O. Etzioni, A. Fader, J. Christensen, S. Soderland, and M. Mausam, “开放信息抽取: 第二代“, 在第二十二届会议记录国际人工智能联合会议 - 第一卷 巴塞罗那, 加泰罗尼亚, 西班牙 AAAI出版社, 2011年, 第3-10页。
- [33] DB Lenat和EA Feigenbaum, “在门槛上的知识”, “人工智能, 第一卷。 47, 没有。 1, pp.185-250, 1991。
- [34] R. 钱, “理解 你的 世界 同 兵, 兵 搜索 博客, ”三月 2013. [上-线]。 可用: <http://blogs.bing.com/search/2013/03/21/理解, 您的世界, 用冰/>
- [35] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, AA Kalyanpur, A. Lally, JW

第20届ACM SIGKDD国际会议
知识发现和数据挖掘。

纽约, NY, USA: ACM, 2014, pp. 601-610。

- [25] N. Nakashole, G. Weikum和F. Suchanek,
“PATY:

关系模式的语义分类

类型”, 在2012年联合会议论文集

自然语言处理中的经验方法与策略

计算自然语言学习, 2012年,

Murdock,

E. Nyberg, J. Prager等人, “建筑沃森: 安
DeepQA项目概述“, AI杂志, 第一卷。 31,
没有。 3, pp. 59-79, 2010。

- [36] F. Belleau, M.-A. Nolin, N. Tourigny, P.
Rigault,

和J. Morissette, “Bio2rdf: 迈向混搭
构建生物信息学知识体系“, ”

生物医学信息学, 第一卷。 41, 没有。 5, pp. 7
706-716,
2008。

- [37] A. Ruttenberg, JA Rees, M. Samwald和A. MS Marshall, “语义网上的生命科学: Neurocommons及其以后”, 生物信息学简报, 第一卷。 10, 没有。 2, 第193-204页, 2009年3月。
- [38] V. Muttchev, D. Pychev, T. Primov和G. Georgiev, “拓展联系生活数据中的途径和交互知识”, Proc. Natl. Acad. Sci. USA, 国际语义网挑战赛, 2009。
- [39] G. Angeli和C. Manning在“第十七届计算自然语言学习会议论文集”中的“哲学家是真人: 推测不可见事实的真相”。 索非亚, 保加利亚: 计算语言学协会, 2013年8月, 第133-142页。
- [40] B. Taskar, M.-F. Wong, P. Abeel和D. Koller, “Link Prediction in Relational Data”, 神经信息处理系统进展 (Advances in Neural Information Processing Systems), S. Thrun, L. Saul和B. Schölkopf, 编辑, 16. 麻省剑桥: 麻省理工学院出版社, 2004年。
- [41] L. Getoor和CP Diehl, “链接挖掘: 一项调查”, ACM SIGKDD探索通讯, 第一卷。 7, 没有。 2, pp. 3-12, 2005。
- [42] HB Newcombe, JM Kennedy, SJ Axford和A. P. James, “重要记录的自动链接计算机可用于从常规记录文件中提取家庭的”后续“统计数据”, Science, 130, 没有。 3381, 第954-959页, 1959年10月。
- [43] S. Tejada, CA Knoblock和S. Minton, “信息集成的学习对象识别规则”, Information Systems, 26, 没有。 8, pp. 607-633, 2001。
- [44] E. Rahm和PA Bernstein, “对自动模式匹配方法的调查”, VLDB Journal, vol. 10, 没有。 4, pp. 334-350, 2001。
- [45] A. Culotta和A. McCallum, “关系数据中多重记录类型的联合重复数据消除”, 第14届ACM信息和知识管理国际会议论文集。 ACM, 2005年, 第257-258页。
- [46] P. Singla和P. Domingos, “实体决议与马尔可夫逻辑”, “数据挖掘, 2006年ICDM'06。 第六国际会议, 2006年12月, 第572-582页。
- [47] I. Bhattacharya和L. Getoor, “集体实体”关系数据中的分辨率”, ACM Trans. Knowl. Discov. 数据, 第一卷。 1, 没有。 1, 2007年3月
- [48] SE Whang和H. Garcia-Molina, “联合实体决议”, 2012年IEEE第28届国际会议在数据工程上。 华盛顿特区, 美国: IEEE 计算机学会, 2012年, 第294-305页。
- [49] S. Fortunato, “图中的社区检测”, 物理报告, 第一卷。 486, 没有。 3, 第75-174页, 2010。
- [50] JC Platt, “SV机器的概率”, Advances 在大型保证金分类器。 麻省理工学院出版社, 1999年, 第61-74。
- [51] LAGalárraga, C. Teflioudi, K. Hose和F. 随机梯度下降“, 在COMPSTAT'2010会议记录。 Physica-Verlag HD, 2010, 第177-186页。
- [53] MEJ纽曼, “科学合作网络的结构”, 国家科学院学报, 第一卷。 98, 没有。 2, 第404-409页, 2001年1月, arXiv: cond-mat / 0007214。
- [54] D. Liben-Nowell和J. Kleinberg, “社会网络的链接预测问题”, 美国信息科学与技术学会杂志, 58, 没有。 7, pp. 1019-1031, 2007。
- [55] D. Jensen和J. Neville, “在关系学习中的连接和自相关原因特征选择偏差”, 第十九届国际机器学习会议论文集。 美国加利福尼亚州旧金山: 摩根考夫曼出版公司, 2002年, 第259-266页。
- [56] PW Holland, KB Laskey和S. Leinhardt, “随机块模型: 第一步”, 社交网络, 第一卷。 5, 没有。 2, pp. 109-137, 1983。
- [57] CJ Anderson, S. Wasserman 和 K. Faust, “Building stochastic blockmodels”, Social Networks, vol. 14, 没有。 1-2, pp. 137-161, 1992, BlockModels的特刊。
- [58] P. Hoff, “对称关系数据中的同构和随机等价性建模”, Advances in Neural Information Processing Systems 20. Curran Associates, Inc., 2008, 第657-664页。
- [59] M. Nickel, V. Tresp和H.-P. Kriegel, “关于多关系数据集体学习的三维模型”, 第28届国际机器学习会议论文集, 2011年, 第809-816页。
- [60] -, “Factorizing YAGO: scalable machine learning for linked data”, Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 271-280。
- [61] M. Nickel, “关系学习的张量分解”, 博士论文, 路德维希马克西米利安大学慕尼黑, 2013年8月。
- [62] Y. Koren, R. Bell和C. Volinsky, “矩阵分解推荐系统技术“, IEEE Computer, 第一卷。 42, 没有。 8, pp. 30-37, 2009。
- [63] TG Kolda和BW Bader, “Tensor Decompositions 和“应用”, SIAM Review, 51, 没有。 3, pp. 455-500, 2009。
- [64] M. Nickel和V. Tresp, “Logistic张量分解多关系数据“, 在”结构化学习: 从结构化和非结构化推断图投入 (SLG 2013)。 2013 ICML研讨会。
- [65] K.-W. Chang, W.-t. Yih, B. Yang和C. Meek, “打字 知识库关系的张量分解 提取“, 在2014年会议记录 自然语言处理中的经验方法研究/ The. ACL - 计算语言学协会, 10月 2014。
- [66] S. Kok和P. Domingos, “统计谓词”

Suchanek,

“AMIE: 不完整的关联规则挖掘

在本体知识基础中的证据“, in

第二十二届国际会议论文集

万维网, 2013年, 第413-422页。

[52] L. Bottou, “大规模机器学习与

发明“ (*Proceedings of the 24th International*)

机器学习会议。

纽约, 纽约,

美国: ACM, 2007年, 第433-440页。

[67] Z. Xu, V. Tresp, K. Yu和H.-P. 克里格尔, “无限

隐藏的关系模型“, 在22日的论文集

- 人工智能不确定性国际会议。 AUA1出版社, 2006年, 第544-551页。
- [68] C. Kemp, JB Tenenbaum, TL Griffiths, T. Yamada 和 N. Ueda, “Learning systems of concepts with a infinite relational model,” Proceedings of the Twenty-First National Conference on Artificial Intelligence, vol. 3, 2006, p. 5。
- [69] I. Sutskever, JBTenenbaum和RR Salakhutdinov, “使用贝叶斯聚类张量因子分解建模关系数据”, Advances in Neural Information Processing Systems 22, 2009, pp. 1821-1828。
- [70] D. Krompaß, M. Nickel和V. Tresp, 2014年 “2014年数据科学与高级分析国际会议论文集” (2014年 DSAA' 2014) 会议记录 “类型受限多关系数据的大规模因子分解”。
- [71] M. Nickel和V. Tresp在 “学习语义学” 中的 “通过基于分层链接的聚类学习来自多关系数据的分类法”。 2011年西班牙格拉纳达NIPS' 11研讨会。
- [72] T. Kolda和B. Bader, “高阶网络链接分析的TOPHITS模型”, 链接分析, 反恐与安全2006年, 2006年。
- [73] T. Franz, A. Schultz, S. Sizov 和 S. Staab, “Triplerank: Ranking semantic web data by tensor decomposition”, The Semantic Web-ISWC 2009, pp. 213-228, 2009。
- [74] L. Drumond, S. Rendle, and L. Schmidt-Thieme, “Predicting RDF Triples in Incomplete Knowledge Bases with Tensor Factorization,” Proceedings of the 27th Annual ACM Symposium on Applied Computing. 意大利Riva del Garda: ACM, 2012年, 第326-331页。
- [75] S. Rendle和L. Schmidt-Thieme, “用于个性化标签推荐的成对相互作用张量因子分解”, 在 Proceedings of the third ACM International Conference on Web Search and Data Mining中。 ACM, 2010, 第81-90页。
- [76] S. Rendle, “缩放分解机器到关系数据”, 在第39届国际超大型数据库会议论文集中。 特伦托, 意大利: VLDB捐赠, 2013年, 第337-348页。
- [77] R. Jenatton, NL Roux, A. Bordes和GR Obozinski, “用于高度多关系数据的潜在因子模型”, Advances in Neural Information Processing Systems 25. Curran Associates, Inc., 2012, 第3167-3175页。
- [78] P. Miettinen, “布尔量子因式分解”, 2011年 IEEE第11届数据挖掘国际会议, 2011年12月, 第447-456页。
- [79] D. 鄂尔多斯和P. Miettinen, “用布尔张量Tucker分解发现事实”, 第22届ACM国际信息与知识管理会议论文集。 纽约, 纽约, 美国: ACM, 2013年, 第1569-1572页。
- [80] X. Jiang, V. Tresp, Y. Huang和M. Nickel, “在多关系图中使用加性模型的链接预测”, ISWC上的语义技术国际研讨会会议推荐系统和大数据会议M. TD Georis, TD Noia,

- P. Lops, T. Lukasiewicz和G. Semeraro编辑, 919. CEUR Workshop Workshop Proceedings, 2012, pp. 1-12.
- [81] 计算语言学协会北美分会的联合人类语言技术会议 / 年会上 S. Riedel, L. Yao, BMMarlin 和 A. McCallum 的 “Relation Extraction with Matrix Factorization and Universal Schemas” (NAACL '13), 2013年6月。
- [82] V. Tresp, Y. Huang, M. Bundschuh 和 A. Rettinger, “物化和查询学到的知识”, Proc. IRMLeS, 第一卷。 2009年, 2009年。
- [83] Y. Huang, V. Tresp, M. Nickel, A. Rettinger 和 H.-P. Kriegel, “语义图中统计学习的可伸缩方法”, Semantic Web journal SWj, 2013。
- [84] P. 斯摩棱斯基, “张量产品变量绑定和在连接系统中的符号结构的表示”, 人工智能, 第一卷。 46, 没有。 1, pp. 159-216, 1990。
- [85] GS Halford, WH Wilson和S. Phillips, “关系复杂性所定义的处理能力: 对比较, 发展和认知心理学的启示”, 行为与脑科学, 第一卷。 21, 没有。 06, pp. 803-831, 1998。
- [86] T. Plate, “用于组成结构的分布式表示方案的通用框架”, 连接主义系统, 用于知识表示和演绎, pp. 15-34, 1997。
- [87] T. Mikolov, K. Chen, G. Crarado, 和 J. Dean, “Efficient Estimation of Word Representations in Vector Space”, Proceedings of Workshop at ICLR, 2013。
- [88] A. Bordes, J. Weston, R. Collobert 和 Y. Bengio, “知识库的学习结构化嵌入”, 第二十五届AAAI人工智能会议论文集, 美国旧金山, 2011年。
- [89] R. Socher, D. Chen, CD Manning和A. Ng, “用神经张量网络进行知识库完成的推理”, Advances in Neural Information Processing Systems 26. Curran Associates, Inc., 2013, 第926-934页。
- [90] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston 和 O. Yakhnenko, “Translating Embeddings for Modeling Multi-Relational Data”, Advances in Neural Information Processing Systems 26. Curran Associates, Inc., 2013, 第2787-2795页。
- [91] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, “在知识基础中嵌入实体与学习和推理的关系”, CoRR, abs / 1412.6575, 2014。
- [92] PD Hoff, AE Raftery和MS Handcock, “用于社会网络分析的潜在空间方法”, 美国统计协会杂志, 97, 没有。 460, 第1090-1098页, 2002。
- [93] L. Lü和T. Zhou, “复杂网络中的链路预测: 一项调查”, Physica A: 统计力学及其应用, vol. 390, 没有。 6, 第1150-1170页, 2011年3月。
- [94] LA Adamic和E. Adar, “网络上的朋友和邻居”, 社会网络, 第一卷。 25, 没有。 3, 第211-230页,

- 2003.
- [95] A.-L. Barabási和R. Albert, “随机网络中的缩放的出现”, *Science*, 286, 没有。 5439, pp. 509-512, 1999.
- [96] L. Katz, “从社会计量分析得出的新的地位指数”, *Psychometrika*, 18, 没有。 1, pp. 39-43, 1953.
- [97] EA Leicht, P. Holme和ME Newman, “网络中的顶点相似性”, *Physical Review E*, 73, 没有。 2, p. 026120, 2006.
- [98] S. Brin和L. Page, “一个大型的超文本网络搜索引擎的解剖”, *计算机网络和ISDN系统*, 30, 没有。 1, 第107-117页, 1998.
- [99] W. Liu和L. Lü, “基于局部随机游走的链路预测”, *EPL (Europhysics Letters)*, vol. 89, 没有。 5, p. 58007, 2010.
- [100] N. Lao和WW Cohen, “使用路径约束随机游走组合的关系检索”
机器学习, 第一卷。 81, 没有。 1, pp. 53-67, 2010.
- [101] N. Lao, T. Mitchell和WW Cohen, “Random
在大范围的知识中进行推理和学习
基地”, 在会议论文集的经验
自然语言处理中的方法。 协会
计算语言学, 2011年, 第529-539页。
- [102] JR Quinlan, “学习关系的逻辑定义 -
tions”, *Machine Learning*, vol. 5, pp. 239-266, 1990.
- [103] M. Nickel, X. Jiang和V. Tresp, “减少
包含关系分解模型的排序
可观察模式”, “神经信息学进展”
处理系统27. Curran Associates, Inc., 2014,
第1179-1187页。
- [104] Y. Koren, “分解与邻里:
一个多方面的协作过滤模型, “
第十四届ACM SIGKDD国际会议论文集
知识发现和数据挖掘会议。
纽约, 纽约, 美国: ACM, 2008, 第426-434页。
- [105] S. Rendle, “使用libFM的分解机器”
ACM智能系统和技术交易
(TIIST), 第一卷。 3, 没有。 3, p. 57,
2012.
- [106] DH Wolpert, “堆积泛化”, 神经网络,
作品, 第一卷。 5, 没有。 2, pp. 241-259, 1992.
- [107] M. Richardson和P. Domingos, “Markov logic net-
作品, “机器学习, 第一卷。 62, 没有。 1, 第
107-136页,
2006.
- [108] S. Jiang, D. Lowd和D. Dou, “Learning to
Refine an
使用马尔可夫自动提取知识库
逻辑”, 2013年IEEE第13届国际会议
数据挖掘, 第912-917页, 2012.
- [109] C. Zhang和C. Ré, “迈向高吞吐量的Gibbs
规模抽样: 跨存储经理的研究”, in
2013 ACM SIGMOD国际会议论文集
数据管理会议。 ACM, 2013年,
397-408.
- 边缘图形识别”, 在Semantic Web-ISWC 2013中。
Springer, 2013, 第542-557页。
- [112] J. Neville 和 D. Jensen, “Relational
dependency networks”, *The Journal of Machine
Learning Research*, vol. 8, 第637-652页,
2007年5月。
- [113] D. Krompaß, X. 江 M. 镍, 和
V. Tresp, “Probabilistic Latent-Factor
Database Models”, 第一届关于知识发现的关联数
据研讨会论文集, 与欧洲机器学习会议和数据库知
识发现的原则和实践共同举办 (ECML PKDD 2014),
2014年。
- [114] H. Ji, T. Cassidy, Q. Li和S. Tamang, “解决时
态知识群体的表示和分类挑战”, *知识与信息系统*,
第1-36页, 2013年8月。
- [115] DL McGuinness, F. Van Harmelen等人, “OWL网络
本体语言概述”, W3C
推荐, 第一卷。 10, 没有。 10, p. 2004年,
2004年。
- [116] A. Hogan, A. Harth, A. Passant, S. Decker,
and
A. Polleres, “编织迂回网”, 3日
国际网络关联数据研讨会
(LDOW2010) 与第19届国际合作
万维网会议。 罗利, 北卡罗莱纳州,
美国: CEUR研讨会会议记录, 2010.
- [117] H. Halpin, P. Hayes, J. McCusker, D.
McGuinness, and
H. 汤普森, “当猫头鹰: sameAs是不一样的:
关联数据中的身份分析, “语义
Web-ISWC 2010, pp. 305-320, 2010.
- [118] D. Krompaß, M. Nickel和V. Tresp, “询问
分解概率三重数据库”
Semantic Web-ISWC 2014. Springer, 2014, pp.
114-
129.
- [119] D. Suciu, D. Olteanu, C. Re和C. Koch, 概率论
数据库。 Morgan&Claypool, 2011年。
- [120] DZ Wang, E. Michelakis, M. Garofalakis和JM
海勒斯坦, “贝叶斯商店: 管理大, 不确定
具有概率图形模型的数据储存库”
VLDB捐赠诉讼, 第一卷。 1, 没有。 1, pp.
340-351, 2008.

- [110] A. Kimmig, SH Bach, M. Broecheler, B. Huang, and L. Getoor, “概率软论”
逻辑“, 在NIPS概率程序研讨会上,
明: 基础和应用, 2012。
- [111] J. Pujara, H. Miao, L. Getoor和W. Cohen,
“Knowl-