本科生毕业论文（设计）

毕业论文（设计）题目：

面向异构信息网络对齐的神经网络模型研究

部　　院　系：　信息科学与技术学院

专　　　业：　计算机科学与技术

学　　　号：　201511210102

学 生 姓 名：　吕青松

指 导 教 师：　王志春

指导教师职称：　副教授

指导教师单位：　北京师范大学信息科学与技术学院

2019 年 4 月 21 日

北京师范大学本科生毕业论文（设计）诚信承诺书

本人郑重声明： 所呈交的毕业论文（设计），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

本人签名： 年 月 日

北京师范大学本科生毕业论文（设计）使用授权书

本人完全了解北京师范大学有关收集、保留和使用毕业论文（设计）的规定，即：本科生毕业论文（设计）工作的知识产权单位属北京师范大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许毕业论文（设计）被查阅和借阅；学校可以公布毕业论文（设计）的全部或部分内容，可以采用影印、缩印或扫描等复制手段保存、汇编毕业论文（设计）。保密的毕业论文（设计）在解密后遵守此规定。

本论文（是、否）保密论文。

保密论文在 年 月解密后适用本授权书。

本人签名： 年 月 日

导师签字： 年 月 日

面向异构信息网络对齐的神经网络模型研究

摘　要

异构信息网络融合旨在将多个异构信息网络进行合并，对大规模网络的构建具有重要作用。其中，实体对齐是异构信息网络融合过程中最关键的一步。实体对齐在人工智能很多领域都有应用，例如，跨语言知识图谱实体对齐能够辅助跨语言信息检索、机器翻译；跨社交网络用户对齐可以用于用户兴趣挖掘。本文提出了一种基于图卷积神经网络的异构信息网络对齐模型，给定一个预先匹配好的实体对集合，我们的方法用图卷积神经网络把两个异构信息网络中的实体嵌入到同一个向量空间中，通过计算实体之间的距离发现新的匹配实体。图卷积神经网络的嵌入表示方法可以利用异构信息网络中的结构信息和属性信息进行实体的嵌入表示，两种表示最后结合起来从而得到准确的面向匹配的嵌入表示。通过在现实世界的异构信息网络上进行测试，我们的模型相比其他基于嵌入表示的模型表现更优。

关键词：异构信息网络 实体对齐 跨语言知识图谱 图神经网络

Research on Heterogeneous Information Network Alignment Oriented Neural Network Models

ABSTRACT

Heterogeneous information network (HIN) fusing aims at merging multiple HINs, which plays an important part in building large-scale networks. Entity alignment (EA) is the most critical step during HIN fusing. EA is useful for many AI applications. For example, cross-lingual knowledge graph entity alignment is able to assist cross-lingual information retrieval and machine translation, and social network user alignment can be used for user interest mining. In this paper, we propose a novel approach for HIN entity alignment via graph convolutional networks (GCNs). Given a set of pre-aligned entities, our approach trains GCNs to embed entities into a unified vector space. Entity alignments are discovered based on the distances between entities in the embedding space. Embeddings based on GCNs can be learned from both the structural and attribute information in HINs, and the results of structure embedding and attribute embedding are combined to get accurate alignment embeddings. In the experiments on aligning real multilingual HINs, our approach gets the best performance compared with other embedding-based alignment approaches.

KEY WORDS： heterogeneous information network, entity alignment, cross-lingual knowledge graph, graph neural network

目 录

1 引言

1.1 研究背景

信息网络由事物（节点）以及事物之间的关系（边）构成，很多现实生活中的问题都可以归结为信息网络分析问题，比如，社交网络研究人和人际关系，参考文献网络研究论文和引用关系，知识图谱研究实体和实体之间的关系。传统的信息网络分析模型通常假设信息网络中的事物属于同一类别，事物之间的关系也只有一种，这种信息网络被称为同构信息网络。基于这种假设的模型只能处理诸如朋友关系网络[1]、论文合著者网络[2]、互联网文档网络[3]等同构信息网络分析问题。然而，现实世界中的系统往往具有异构性和复杂性，事物的性质不尽相同，事物之间的关系也复杂多样，因此同构信息网络模型在应用中具有很大的局限性。2009 年，Y. Sun 等人[4]提出 RankClus 算法，对论文作者和国际会议组成的二类信息网络进行分析，达到了比单纯使用合著者网络更好的权威度预测效果，这也是异构信息网络的概念首次提出。2011 年，Y. Sun 等人[5]又提出元路径的概念，并发现在异构信息网络中，用不同的元路径可以从不同的层面衡量节点的相似度，而这一点在同构信息网络中是无法做到的。此后，异构信息网络便成为学术界研究的热点问题，至今，更加深入的研究仍在如火如荼地进行。

异构信息网络最大的特点就是异构性，这种异构性一方面表现在构成网络的节点和边的类型可以有多种，另一方面表现在用于构建网络的信息源可以有多个。异构性虽然能够让异构信息网络有更强的语义表达能力，但是也会给异构信息网络问题的研究带来巨大的挑战，其中，如何融合从不同信息源获得的网络就是挑战之一。在异构信息网络融合的过程中，实体对齐是最关键的一步。目前，许多领域都对异构信息网络实体对齐问题进行了相关的研究工作。在社交网络领域，该问题被定义为锚链接预测问题[6]，通过一部分已知的用户锚链接获取新的锚链接，将两个社交网络中原本没有匹配的相同用户进行连接，从而更全面地理解用户兴趣、提供更优质的推荐服务；在知识图谱领域，主要研究跨语言知识图谱的实体匹配问题[7][8]，通过将不同语言的知识进行对齐，不仅可以缓解不同语言的知识不平衡的现象、实现跨语言的知识共享，还能辅助信息检索和机器翻译等应用获得更好的体验。

1.2 研究现状

传统的对齐模型使用人工定义的特征表示网络中的节点，例如，X. Kong 等人[6]提出的 MNA 算法使用扩展公共邻居、扩展 Jaccard 系数等作为结构特征，使用社交网络中用户

发表文章的时间、地点和内容作为标签特征，对社交网络进行锚链接预测；Z. Wang 等人[7]提出的概率图模型使用邻居的相似度、节点的类别、作者兴趣作为主要特征对维基百科和百度百科进行跨语言对齐。但是，各个领域的网络往往具有不同的特点，在一种网络中的特征在另一种网络不一定可用，即使可用也不能保证有同样的表现，即使是元路径的方法[19]，也需要预先人工设定一个或多个元路径来衡量节点相似度。因此，通过表示学习的方法自动地从信息网络中获取特征也一直是学者们的关注点。

网络表示学习最初使用独热向量表示每个节点，通过一阶邻居向量或高阶邻居向量衡量节点间的相似度。但是这种向量具有高维、稀疏、离散的特点，使得其很难应用于大规模网络，并且效果欠佳。虽然稀疏表示可以通过奇异值分解转换为低维稠密表示，但是这种方法由于复杂度太高仍然难以用于大规模网络。2013 年在自然语言处理领域，T. Mikolov 等人[9][10]提出词嵌入模型，将词表示为低维、稠密、连续的向量，被成为词的嵌入表示或分布式表示。这种思路极大地启发了社交网络、知识图谱等领域对网络表示学习的研究，许多将网络中的实体和关系嵌入到低维向量空间的模型开始提出。2013 年，A. Bordes 等人[11]提出 TransE 模型，利用（头实体，关系，尾实体）的三元组的线性运算和梯度下降，把知识库中的实体和关系嵌入到低维向量空间；2014 年，B. Perozzi 等人[12]提出 DeepWalk 模型，首先利用随机游走的方式，将社交网络转换为节点序列，然后利用 Skip-gram 等词嵌入方法把节点嵌入到低维向量空间。此后，许多借助 TransE 或 DeepWalk 等模型的网络对齐模型也相继被提出[14][15][16][17][20][21][25]。

然而，目前大多数基于嵌入表示的对齐模型存在以下两个问题：

- 首先，基于 TransE 和 DeepWalk 等模型的对齐模型需要在嵌入损失函数和对齐损失函数之间权衡，这种权衡需要在模型中加入一个或多个超参数，导致模型调参复杂度增加。
- 其次，大多数嵌入模型只能考虑网络中的结构信息，而现实世界中的异构信息网络还存在大量的附加在节点上的属性信息，这部分信息在实体对齐问题中往往能够起到重要作用。Z. Sun 等人[16]为解决这个问题提出了 JAPE 模型，借助 Skip-gram 的思想利用属性在实体中的共现关系对属性进行嵌入，但是该模型也只能利用属性的类型信息，而无法利用属性值的信息。

近几年，图神经网络处理异构信息网络的方法[22][23][31]开始受到越来越多的关注。图神经网络一方面可以同时处理网络的结构信息以及节点的属性信息，另一方面可以端到端地提取深度特征而无需人工干预。因此不论相比传统的人工特征模型还是近年来的图结构嵌入模型，都具有更强的网络信息处理能力。

1.3 本文的主要研究内容

为了解决当前大多数网络嵌入模型存在的属性信息无法利用、嵌入损失和对齐损失难以权衡的问题，我们提出了一种基于图卷积神经网络的异构信息网络对齐模型。具体来说，我们以实体和关系种类最多、结构最复杂的异构信息网络——知识图谱——的跨语言实体对齐为例，通过实验说明了以下三点：

- 图卷积神经网络可以直接面向异构信息网络的对齐问题建立损失函数，而无需通过超参数权衡嵌入和对齐的损失函数。
- 图卷积神经网络可以利用属性信息辅助对齐模型进一步提升效果。
- 基于图卷积神经网络的模型可以达到比现有嵌入表示学习的模型[1]更好的对齐效果。

2 相关工作

传统特征工程的方法在对齐异构信息网络时，对不同的网络需要用不同的模型，例如：在社交网络对齐时，需要借助用户的附加信息[24][13]，如用户名、性别、语言风格、兴趣等等；在跨语言知识图谱对齐时，需要借助机器翻译[7]。而在近几年提出的基于嵌入表示学习的框架下，主要关注点由节点的附加信息变成了图的网络结构信息，这些模型相比传统的方法具有更强的通用性。

2.1 社交网络对齐

2.1.1 IONE

IONE（Input-Output Network Embedding）[25]借助 Skip-gram 的思想把节点嵌入低维向量空间。总体来看，目标函数有两部分——嵌入损失函数和对齐损失函数，两部分损失函数都是通过节点向量预测周边节点向量来定义（锚链接也可以看作一种强邻居）。通过优化该损失函数，既保持了每个网络各自的结构信息，又让两个网络中相同的用户进行对齐，达到了多社交网络嵌入的效果。此外 IONE 还通过把一个节点的向量拆分成三部分——节点向量、输入向量、输出向量，根据边方向的不同让节点向量分别影响输入向量和输出向量，从而考虑了边的方向信息。

[1] 本文提出的模型发表于 EMNLP 2018，现有模型指此前提出的模型，相关工作也未对同期工作和之后的工作进行梳理

2.1.2 PALE

PALE（Predicting Anchor Links via Embedding）[20]把对齐过程看作两个步骤，第一步是两个社交网络分别做嵌入表示，第二步是通过锚链接学习两个向量空间的映射。在论文中，作者使用的嵌入方法是链接预测，让有边的点对向量点积更大，这种嵌入方法后来被 DeepWalk 代替从而形成了 PALE 的变体 PALE-DeepWalk。第二步的映射作者尝试了线性映射和多层感知机。

2.2 跨语言知识图谱对齐

2.2.1 JE

JE[14]基于 TransE 进行跨语言知识图谱嵌入，通过在 TransE 的损失函数中加入对齐的部分，把两个知识图谱嵌入同一个向量空间，这种方法与社交网络中的 IONE 思路近似。

2.2.2 MTransE

MTransE[15]的思路与社交网络中的 PALE 相近，第一步用 TransE 分别学习两个知识图谱的向量表示，第二步通过一些对齐的三元组学习两个向量空间的映射，在论文中作者尝试了 5 种映射方法。

2.2.3 JAPE

JAPE[16]的嵌入过程分为两部分：结构信息嵌入和属性信息嵌入。对于属性信息嵌入，作者借助 Skip-gram 的思想对属性进行嵌入表示，而实体由周边属性的向量均值得到；对于结构信息嵌入，作者把预先对齐的种子实体对合并，然后用 TransE 进行嵌入作为结构信息损失函数，通过实体的属性向量构造了辅助修正损失函数。JAPE 需要关系和属性的对齐信息才能获得预期的效果。

2.2.4 ITransE

ITransE[17]同样基于 TransE 模型进行对齐，但是采用的是迭代的方法，新发现的对齐实体可以作为训练种子从而获取更多的对齐实体。这种方法容易造成错误的传播，因此作者假设所有的关系都是已经对齐的。

通过相关工作的梳理，可以发现现有的嵌入表示对齐模型都是基于某种网络嵌入方法进行网络表示学习，然后利用某种映射实现两个网络的对齐，而没有直接面向对齐的嵌入

表示学习模型。相比这些方法，我们的模型采用一个完全不同的框架：用图卷积神经网络把实体嵌入向量空间，而直接把对齐作为目标函数，让等价的实体尽量接近。

3 基于图卷积神经网络的对齐模型

3.1 问题定义

我们把异构信息网络定义为五元组，$G = (E, R, A, L, T_R, T_A)$。其中H是一个异构信息网络；E是网络中的实体集合；R是网络中的关系集合；A是网络中的属性集合；L是网络中的文字集合；$T_R \subseteq E \times R \times E$是关系三元组集合，集合中元素的形式为<*实体, 关系, 实体*>；$T_A \subseteq E \times A \times L$是属性三元组集合，集合中元素的形式为<*实体, 属性, 文字（属性值）*>。例如在知识图谱 YAGO[26][27]中，graduatedFrom 是一种关系，(Albert Einstein, graduatedFrom, ETH Zurich)是一个关系三元组；diedOnDate 是一种属性，(Albert Einstein, diedOnDate, 1955)是一个属性三元组。关系三元组和属性三元组都描述了实体相关的重要信息，我们的模型在对齐两个网络的时候会把两种信息都考虑在内。

现在假设G_1和G_2是两个不同的异构信息网络，$S = \{(e_1, e_2)|e_1 \in E_1, e_2 \in E_2\}$是网络中一些预先匹配好的实体对构成的集合。对齐任务就是要在现有的匹配实体的基础上，发现新的匹配实体。

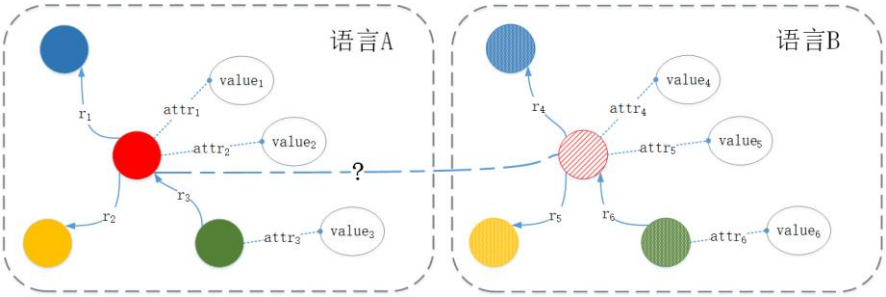


图 1：跨语言知识图谱对齐示意图。同色代表相同实体，阴影表示待发现的对应实体。

3.2 图卷积神经网络面向对齐的嵌入表示

图卷积神经网络（Graph Convolutional Networks，以下简称 GCN）[22][28][29][30]是一种直接处理图数据的神经网络，它可以端到端地学习任意规模和结构的图中节点的表示。

给定节点的特征向量和图的结构信息，GCN 可以为每个节点产生一个嵌入表示的向量，向量中包含了节点的特征信息和它周边的邻居节点的信息。在对齐问题的情景下，我们有以下两个设定：（1） 等价实体的特征向量是相似的；（2）等价实体的邻居通常也含有等价实体。而 GCN 恰好可以把节点的特征信息和邻居信息结合，产生节点的向量表示，因此，我们的模型使用 GCN 把两个网络嵌入到低维向量空间，在这个向量空间中，等价的实体比非等价实体更接近。

一个 GCN 模型由多个堆积的图卷积层组成，输入层是节点的特征向量矩阵，此后每一层都由上一层的特征矩阵通过图卷积操作产生新的特征矩阵，作为下一层的输入。设节点个数为 n，$d^{(i)}$表示第 i 层的特征维度，$H^{(i)} \in R^{n\times d^{(i)}}$表示第 i 层的特征矩阵，那么图卷积层的公式如下：

$$H^{(i+1)} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(i)}W^{(i)})$$

其中，σ是激活函数，$A \in R^{n\times n}$是图的邻接矩阵，$\hat{A} = A + I$，I 是单位矩阵；$\hat{D}$是一个对角矩阵，(i, i)位置上是第 i 个节点的度。$W^{(i)}$是第 i 层的权重矩阵。

3.2.1 结构嵌入和属性嵌入

我们的模型框架如图 2 所示。对于结构嵌入（Structure Embedding，以下简称 SE），输入层的特征向量是单位阵；对于属性嵌入（Attribute Embedding，以下简称 AE），输入层的特征向量是实体的属性特征。属性特征使用的是独热表示，每一列代表一种属性，如果实体有对应属性则为 1，否则为 0。这种表示可以扩展为属性值，例如，数值型属性可以用数值扩展 0/1，文本型属性可以用预训练的词向量扩展 0/1。我们把属性值的扩展作为未来工作。

如果用$H_s^{(i)}$和$W_s^{(i)}$代表第 i 层的结构嵌入和参数矩阵，$H_a^{(i)}$和$W_a^{(i)}$代表第 i 层的属性嵌入，那么模型前向传播的公式如下：

$$[H_s^{(i+1)}; H_a^{(i+1)}] = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}[H_s^{(i)}W_s^{(i)}; H_a^{(i)}W_a^{(i)}])$$

其中，[;]代表矩阵拼接，这里的激活函数σ为$\mathrm{ReLU}(x) = \max(0, x)$。

最后，每个实体的嵌入表示是结构嵌入和属性嵌入的带权拼接：

$$H_{GCN} = [\beta H_s; (1-\beta)H_a]$$

其中，β是一个超参数。

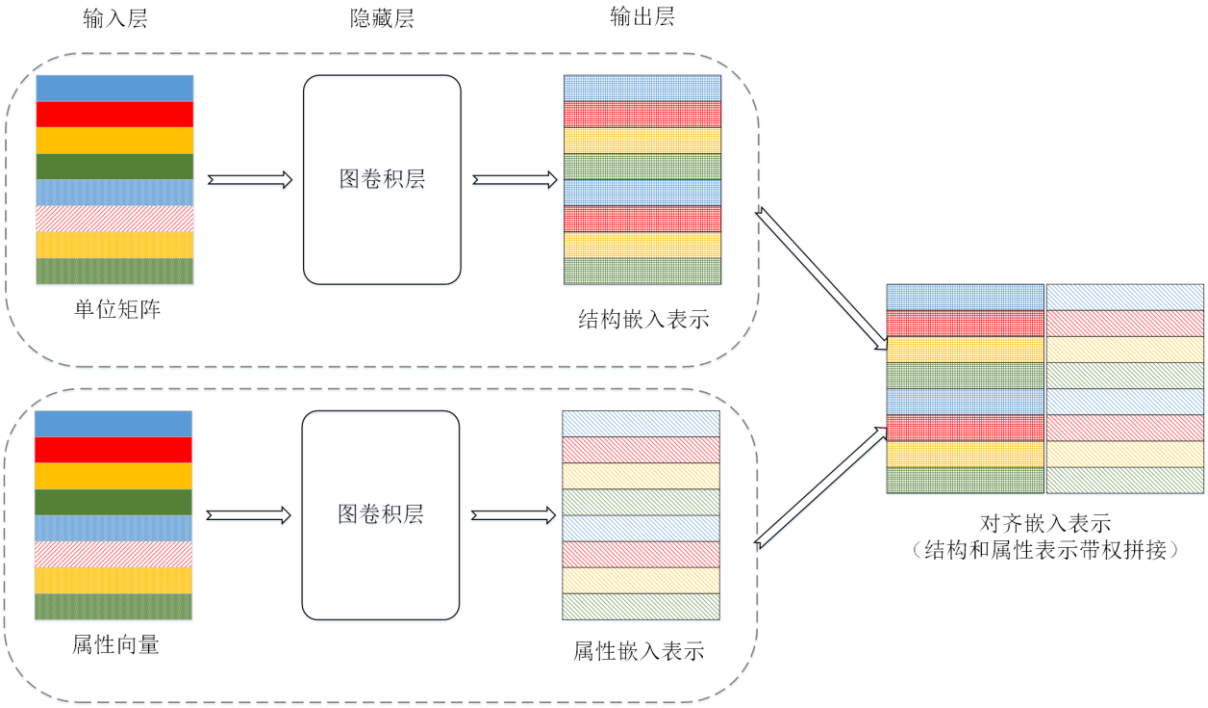


图 2：面向实体对齐的图卷积神经网络模型框架示意图

3.2.2 模型配置

更具体地，我们的模型使用了两层图卷积，设$n = |E_1 \cup E_2|$，$a = |A_1 \cup A_2|$，最终 SE 的目标维度是d_s，AE 的目标维度是d_a，各层的参数维度如表 1 所示。其中$W^{(0)}$用归一化的正态分布初始化，$W^{(1)}$用单位矩阵且不更新，因此模型需要训练的参数只有$W^{(0)}$。

	SE	AE
$H^{(0)}$	$n \times n$	$n \times a$
$W^{(0)}$	$n \times d_s$	$a \times d_a$
$H^{(1)}$	$n \times d_s$	$n \times d_a$
$W^{(1)}$	$d_s \times d_s$	$d_a \times d_a$
$H^{(2)}$	$n \times d_s$	$n \times d_a$

表 1：各层参数维度

3.2.3 邻接矩阵计算

原始的 GCN 模型在计算 A 矩阵时把图看作无向无权图，即没有区分边的方向和边的类型。但是知识图谱是一种有向的、多关系类型的异构信息网络，因此，我们设计了一种特别的方式去计算知识图谱的 A 矩阵。令a_{ij}表示 A 矩阵中 i 节点对 j 节点的影响权重，这个值程度很大程度上取决于连接两个节点的关系类型。例如，hasParent 的影响显然应该大于 hasFriend。所以，我们定义了两个函数，用来计算这种关系类型的影响权重：

$$\mathrm{fun(r)} = \frac{\#\text{关系 r 的头实体个数}}{\#\text{关系 r 的三元组个数}}$$

$$\mathrm{ifunc(r)} = \frac{\#\text{关系 r 的尾实体个数}}{\#\text{关系 r 的三元组个数}}$$

然后，a_{ij}被定义为：

$$a_{ij} = \sigma\Big(\sum_{<e_i,r,e_j>\in G} \mathrm{ifun(r)} + \sum_{<e_j,r,e_i>\in G} \mathrm{fun(r)}\Big)$$

其中，σ是 sigmoid 函数。

3.2.4 模型训练和实体匹配

为了让 GCN 得到的实体向量表示能让等价的实体尽量接近，我们用已有的匹配实体对集合 S 作为训练数据。模型训练的过程是用随机梯度下降来最小化以下损失函数：

$$L = \sum_{(e,v)\in S} \sum_{(e',v')\in S'_{(e,v)}} \big[f\big(h(e),h(v)\big) + \gamma - f(h(e'),h(v'))\big]_+$$

其中，$[x]_+ = \max\{0, x\}$；$S'_{(e,v)}$表示与(e, v)相关联的反例集合，反例从两个网络中随机选择；$\gamma > 0$是一个区分正例和反例的超参数。f 是衡量两个向量之间距离的函数，这里我们选用 L1 距离，h 用于获取某个实体的向量表示。该损失函数对 SE 和 AE 都适用。

实体匹配的过程就是一个 KG 里实体在另一个 KG 里寻找相似度最大的实体的过程。训练的损失函数使用的是 L1 距离，很自然地，相似度指标我们也使用 L1 距离（曼哈顿距离）。

4 实验

4.1 数据集

我们使用由 Z. Sun 等人[16]构建的 DBP15K 作为实验数据集。该数据集是从 DBpedia 中通过一定规则抽取的知识图谱子集，包括汉语、英语、日语、法语四种语言。表 2 描述了该数据集的细节信息。DBP15K 包括三部分：汉英对照图谱、日英对照图谱、法英对照图谱，每个部分包括两种语言的知识图谱、以及 1.5 万对匹配实体。实验中我们把这些已知实体对划分成两部分，一部分用来训练，另一部分用来测试。

数据集		实体个数	关系个数	属性个数	关系三元组个数	属性三元组个数
DBP15K-汉英	汉语	66469	2830	8113	153929	379684
	英语	98125	2317	7173	237674	567755
DBP15K-日英	日语	65744	2043	5882	164373	354619
	英语	95680	2096	6066	233319	497230
DBP15K-法英	法语	66858	1379	4547	192191	528665
	英语	105889	2209	6422	278590	576543

表 2：DBP15K 数据集统计信息

4.2 实验设定

4.2.1 基线模型

在实验中，我们将我们的方法与 JE、MTransE 和 JAPE 进行了比较。我们还构建了一个 JAPE 的变体——JAPE’，该模型在 JAPE 的基础上去掉了预先匹配好的关系和属性信息。

我们没有把迭代的方法作为基线进行对比。一方面，我们认为迭代的方法可以与任意嵌入表示的方法组合，所以把 GCN 加入迭代作为未来工作。另一方面，ITransE 假定所有关系是预先匹配好的，这种假设在 DBP15K 的数据集上并不适用。

4.2.2 测试方法

我们把 DBP15K 中的 1.5 万对匹配实体作为标准答案，对于所有模型，都把这 1.5 万对实体按照 3:7 的比例划分为训练集和测试集。我们用 Hits@k 作为评测指标去测试模型的表现。对于一个测试样本(e_1, e_2)，将e_1与所有候选实体进行相似度计算，e_2在这些相似度中排名记为 r，Hits@k 衡量的是在测试集中，r<=k 的测试样本所占比例。

4.2.3 参数设置

$d_s = 200$，$d_a = 100$，$\gamma = 3$，$\beta = 0.9$。这些参数的设定属于经验性的设置。对于学习率，设学习率α迭代 500 次以后 SE 得到的损失函数值为 loss，我们取让 loss 最小的学习率作为模型的学习率。这样做的好处在于无需验证集。图 3 显示了随着学习率的变化 loss 和 Hits@1 之间的相关性，为了方便查看，表中显示的是-log(loss)和 hits/10。（中英数据集，训练和测试按 3:7 划分）

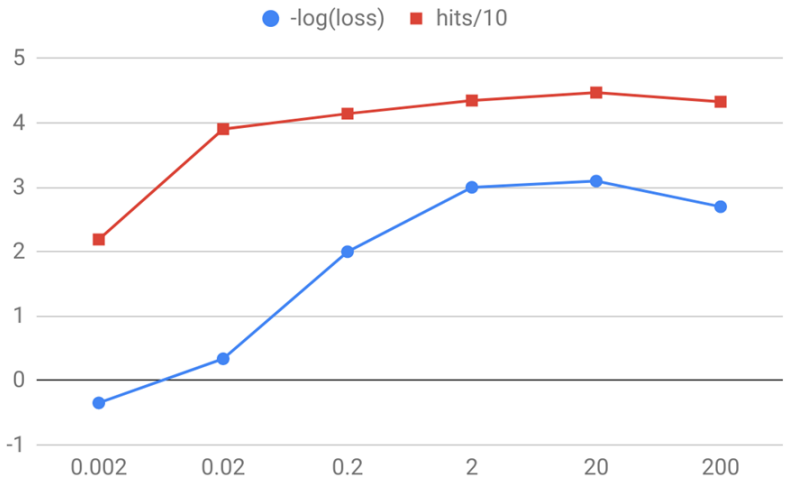


图 3：不同学习率下 loss 和 Hits@1 之间的相关性

4.3 实验结果和分析

4.3.1 实验结果

表 3 展示了所有对比方法在 DBP15K 数据集上的结果。对于 Hits@k 来说，我们汇报了 Hits@1、Hits@10 和 Hits@50。由于我们使用的数据集与 Z. Sun 等人[16]相同，因此 JE、MTransE 和 JAPE 的结果是从[16]中获得。对于 JAPE 和 JAPE’，他们各自都有三种变体：

结构嵌入且无负例三元组（SE w/o neg.），结构嵌入（SE），结构属性联合嵌入（SE+AE）。我们用 GCN（SE）和 GCN（SE+AE）来代表我们模型的两种变体：前者只使用关系三元组做结构嵌入，后者同时使用关系三元组和属性三元组做结构属性联合嵌入。

DBP15K-汉英		汉语→英语			英语→汉语		
		Hits@1	Hits@10	Hits@50	Hits@1	Hits@10	Hits@50
*JE		21.27	42.77	56.74	19.52	39.36	53.25
*MTransE		30.83	61.41	79.12	24.78	52.42	70.45
*JAPE	SE w/o neg.	38.34	68.86	84.07	31.66	59.37	76.33
	SE	39.78	72.35	87.12	32.29	62.79	80.55
	SE+AE	41.18	74.46	**88.90**	40.15	71.05	**86.18**
JAPE'	SE w/o neg.	30.10	62.58	80.28	23.04	52.91	72.17
	SE	30.54	66.41	83.94	23.91	57.02	77.31
	SE+AE	33.32	69.28	86.40	33.02	66.92	85.15
GCN	SE	41.96	74.10	83.78	38.62	69.58	80.70
	SE+AE	**44.75**	**77.31**	87.76	**40.61**	**72.61**	84.10
DBP15K-日英		日语→英语			英语→日语		
		Hits@1	Hits@10	Hits@50	Hits@1	Hits@10	Hits@50
*JE		18.92	39.97	54.24	17.80	38.44	52.48
*MTransE		27.86	57.45	75.94	23.72	49.92	67.93
*JAPE	SE w/o neg.	33.10	63.90	80.80	29.71	56.28	73.84
	SE	34.27	66.39	83.61	31.40	60.80	78.51
	SE+AE	36.25	68.50	85.35	38.37	67.27	82.65
JAPE'	SE w/o neg.	28.90	60.61	80.03	25.34	53.36	71.94
	SE	29.35	63.31	82.76	26.37	57.35	76.87
	SE+AE	31.06	64.11	81.57	32.45	62.21	79.08
GCN	SE	42.73	75.63	84.61	40.59	72.51	82.15
	SE+AE	**45.10**	**78.34**	**88.57**	**42.62**	**75.41**	**86.62**
DBP15K-法英		法语→英语			英语→法语		
		Hits@1	Hits@10	Hits@50	Hits@1	Hits@10	Hits@50

*JE		15.38	38.84	56.50	14.61	37.25	54.01
*MTransE		24.41	55.55	74.41	21.26	50.60	69.93
*JAPE	SE w/o neg.	29.55	62.18	79.36	25.40	56.55	74.96
	SE	29.63	64.55	81.90	26.55	60.30	78.71
	SE+AE	32.39	66.68	83.19	32.97	65.91	82.38
JAPE'	SE w/o neg.	28.23	60.99	78.47	24.68	55.25	74.19
	SE	27.58	62.03	79.98	24.93	58.95	77.79
	SE+AE	30.21	65.81	82.57	31.42	63.86	80.95
GCN	SE	42.26	78.48	88.86	40.86	76.68	88.04
	SE+AE	**43.76**	**80.12**	**90.81**	**42.28**	**79.03**	**90.58**

表 3：跨语言知识图谱实体匹配实验对比结果（标*表示结果引用自论文[16]）

4.3.2 结果分析

首先是 GCN（SE）对比 GCN（SE+AE）。通过实验结果发现，属性信息的加入可以让结果有所提升，这种提升大约在 1%到 10%之间，与 JAPE 的 SE 和 SE+AE 非常相似。这说明我们的模型可以用统一的框架有效结合关系和属性这两种不同的信息。

其次是 GCN（SE+AE）对比基线模型。从实验结果可以看出，我们的模型在所有数据集的几乎所有指标上都比基线模型表现更好。尤其是在法英数据集上，GCN 相比基线模型提升 10%的 Hits@1。虽然在汉英数据集上 Hits@50 比 JAPE 要略低一些，但是值得注意的是，JAPE 使用了关系的对齐信息，如果考虑它的变体 JAPE’，我们的模型是占绝对优势的。

此外，GCN 还有一个很好的特性，即在不同语言的数据集下表现稳定，而没有像基线模型一样在不同语言数据集上结果浮动较大。

4.4 扩展实验

4.4.1 与 TransE 的嵌入表示结合

为了探究我们的模型与 TransE 模型（JAPE’的 SE）的嵌入表示包含的信息是否一致，我们把两者的向量表示进行了加权拼接。拼接的公式为：

$$\mathrm{H_{align}} = [\beta_2 \mathrm{H_{GCN}}; (1-\beta_2)\mathrm{H_{TransE}}]$$

其中，β_2经验性地设置为 0.7。

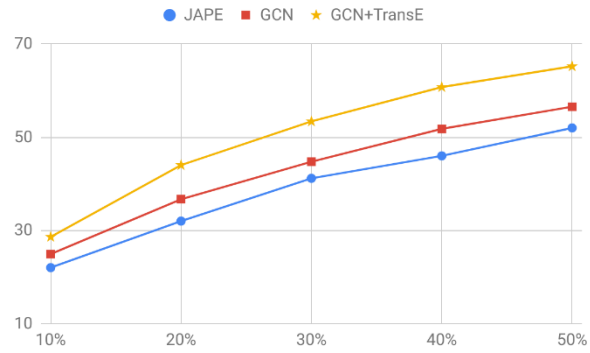
结果如表 4 所示，可以发现，混合的嵌入表示可以获得比两个独立模型更高的命中率，因此说明，两个模型所嵌入的信息不完全一样，并且能够在对齐任务中相互辅助。我们把如何自动训练拼接权重作为未来工作。

	汉语→英语			英语→汉语		
DBP15K-汉英	Hits@1	Hits@10	Hits@50	Hits@1	Hits@10	Hits@50
GCN	44.75	77.31	87.76	40.61	72.61	84.10
GCN+TransE	**53.40**	**82.73**	**92.43**	**47.79**	**77.55**	**88.47**
	日语→英语			英语→日语		
DBP15K-日英	Hits@1	Hits@10	Hits@50	Hits@1	Hits@10	Hits@50
GCN	45.10	78.34	88.57	42.62	75.41	86.62
GCN+TransE	**52.02**	**82.52**	**92.00**	**49.28**	**78.67**	**89.30**
	法语→英语			英语→法语		
DBP15K-法英	Hits@1	Hits@10	Hits@50	Hits@1	Hits@10	Hits@50
GCN	43.76	80.12	90.81	42.28	79.03	90.58
GCN+TransE	**53.06**	**84.74**	**93.64**	**51.07**	**83.91**	**93.49**

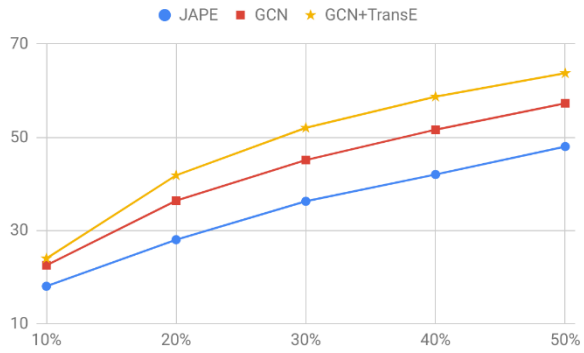
表 4：GCN 与 TransE 结合的实验结果

4.4.2 不同比例划分训练集测试集

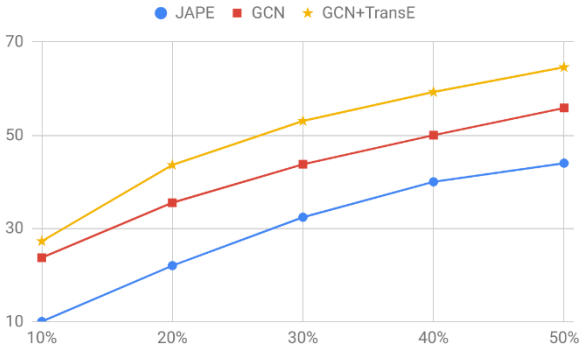
为了探究 GCN 对齐模型如何受训练集大小的影响，我们进一步与 JAPE 在不同比例的训练集下进行了比较。我们测试了 5 种比例的训练集，分别为 10%到 50%，步长为 10%。对于 GCN 与 TransE 结合的方法，我们也进行了实验。实验结果如图 4 所示，可以看出在任何比例的训练集下，GCN 模型比 JAPE 都有更好的表现，GCN 与 TransE 的结合也能让结果有稳定的提升。



汉语-英语



日语-英语



法语-英语

图 4：不同比例的训练集三种模型 Hits@1 对比

4.4.3 社交网络数据集测试

对于社交网络的用户匹配，我们在 IONE[25]提供的数据集上进行了测试。该数据集包括 Foursquare 和 Twitter 两个社交网络，用户数为 5 千余个，锚链接 1609 个。通过在不同比例的训练集和测试集下评测，发现基于 GCN 的对齐模型稳定优于 IONE 的表现。结果如图 5 所示，与 IONE 相同，我们采用 Precision@30（两个方向 Hits@30 的均值）作为评判标准。

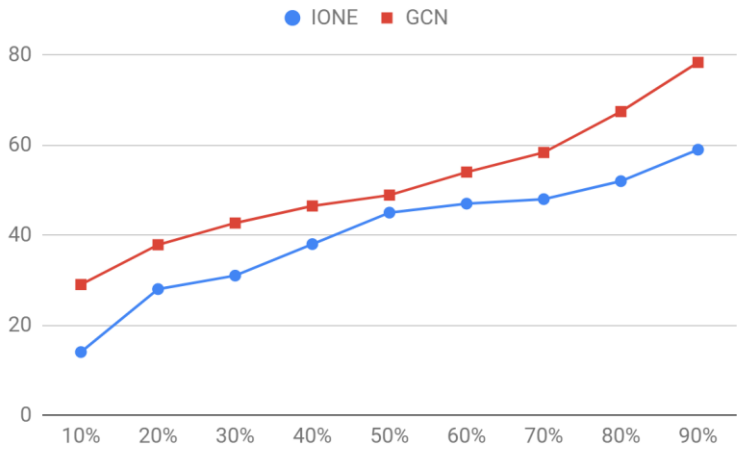


图 5：GCN 与 IONE 在社交网络匹配情景下对比

5 总结和展望

本文提出了一种全新的基于图卷积神经网络的异构信息网络对齐模型，可以同时利用关系和属性信息来得到实体的嵌入表示。我们在真实的多语言知识图谱上评测了我们的方法，结果显示，我们的模型相比基线模型有巨大的优势。

对于未来工作，我们有以下四点展望：

- 结合多种模型时自动学习权重而非人工调整权重；
- 把图卷积模型与迭代的方法结合让效果进一步提升；
- 探索其他图神经网络（如关系型图卷积神经网络[31]、图注意力网络[23]）应用在实体对齐任务上的效果；
- 与传统的机器翻译方法结合，提升实体对齐或机器翻译的效果。

参考文献

[1] Leroy V, Cambazoglu B B, Bonchi F. Cold start link prediction[C]//Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010: 393-402.

[2] Lichtenwalter R N, Lussier J T, Chawla N V. New perspectives and methods in link prediction[C]//Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010: 243-252.

[3] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: Bringing order to the web[R]. Stanford InfoLab, 1999.

[4] Sun Y, Han J, Zhao P, et al. Rankclus: integrating clustering with ranking for heterogeneous information network analysis[C]//Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology. ACM, 2009: 565-576.

[5] Sun Y, Han J, Yan X, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks[J]. Proceedings of the VLDB Endowment, 2011, 4(11): 992-1003.

[6] Kong X, Zhang J, Yu P S. Inferring anchor links across multiple heterogeneous social networks[C]//Proceedings of the 22nd ACM international conference on Information & Knowledge Management. ACM, 2013: 179-188.

[7] Wang Z, Li J, Wang Z, et al. Cross-lingual knowledge linking across wiki knowledge bases[C]//Proceedings of the 21st international conference on World Wide Web. ACM, 2012: 459-468.

[8] Hao Y, Zhang Y, He S, et al. A joint embedding method for entity alignment of knowledge bases[C]//China Conference on Knowledge Graph and Semantic Computing. Springer, Singapore, 2016: 3-14.

[9] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.

[10] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119.

[11] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]//Advances in neural information processing systems. 2013: 2787-2795.

[12] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 701-710.

[13] Zhang J, Philip S Y. Integrated Anchor and Social Link Predictions across Social Networks[C]//IJCAI. 2015: 2125-2132.

[14] Hao Y, Zhang Y, He S, et al. A joint embedding method for entity alignment of knowledge bases[C]//China Conference on Knowledge Graph and Semantic Computing. Springer, Singapore, 2016: 3-14.

[15] Chen M, Tian Y, Yang M, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment[J]. arXiv preprint arXiv:1611.03954, 2016.

[16] Sun Z, Hu W, Li C. Cross-lingual entity alignment via joint attribute-preserving embedding[C]//International Semantic Web Conference. Springer, Cham, 2017: 628-644.

[17] Zhu H, Xie R, Liu Z, et al. Iterative entity alignment via joint knowledge embeddings[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. AAAI Press, 2017: 4258-4264.

[18] Wang Z, Lv Q, Lan X, et al. Cross-lingual Knowledge Graph Alignment via Graph Convolutional Networks[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 349-357.

[19] Zhang J, Yu P S, Zhou Z H. Meta-path based multi-network collective link prediction[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 1286-1295.

[20] Man T, Shen H, Liu S, et al. Predict Anchor Links across Social Networks via an Embedding Approach[C]//IJCAI. 2016, 16: 1823-1829.

[21] Guo L, Sun Z, Cao E, et al. Recurrent Skipping Networks for Entity Alignment[J]. arXiv preprint arXiv:1811.02318, 2018.

[22] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.

[23] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.

[24] Liu J, Zhang F, Song X, et al. What's in a name?: an unsupervised approach to link users across communities[C]//Proceedings of the sixth ACM international conference on Web search and data mining. ACM, 2013: 495-504.

[25] Liu L, Cheung W K, Li X, et al. Aligning Users across Social Networks Using Network Embedding[C]//IJCAI. 2016: 1774-1780.

[26] Suchanek F M, Kasneci G, Weikum G. Yago: A large ontology from wikipedia and wordnet[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2008, 6(3): 203-217.

[27] Rebele T, Suchanek F, Hoffart J, et al. YAGO: A multilingual knowledge base from wikipedia, wordnet, and geonames[C]//International Semantic Web Conference. Springer, Cham, 2016: 177-185.

[28] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[J]. arXiv preprint arXiv:1312.6203, 2013.

[29] Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data[J]. arXiv preprint arXiv:1506.05163, 2015.

[30] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[C]//Advances in neural information processing systems. 2016: 3844-3852.

[31] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//European Semantic Web Conference. Springer, Cham, 2018: 593-607.

致　谢

感谢我的指导老师王志春教授对本人的悉心指导，您潜心治学的态度和紧跟前沿技术的精神让我终生受益。没有您对的信任、鼓励和帮助，就没有这项成果的提出和发表。感谢我的父母、姥姥姥爷、爷爷奶奶，没有你们就没有今日的我。感谢我的新生导师尹乾和郑新教授、ACM 校队的教练冯速教授，是你们培养了我最初的科研素养和代码能力。特别感谢兰晓翰和张雨同学，因为有你们的合作，这项成果才得以呈现，与你们一起奋战过的无数个昼夜让我终生难忘。

本课题承蒙国家自然科学基金和国家重点研发计划的资助，特此致谢。