

控制与决策

Control and Decision

基于联合知识表示学习的多模态实体对齐

王会勇, 论兵, 张晓明, 孙晓领

引用本文:

王会勇, 论兵, 张晓明, 等. 基于联合知识表示学习的多模态实体对齐[J]. *控制与决策*, 2020, 35(12): 2855–2864.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.0331>

您可能感兴趣的其他文章

Articles you may be interested in

阴影条件下基于迁移强化学习的光伏系统最大功率跟踪

Transfer reinforcement learning based maximum power point tracker of PV systems under partial shading condition

控制与决策. 2020, 35(12): 2939–2949 <https://doi.org/10.13195/j.kzyjc.2019.0412>

基于社交网络的双知识表达分类方法

Double knowledge representations based classification method from perspective of social networks

控制与决策. 2020, 35(11): 2653–2664 <https://doi.org/10.13195/j.kzyjc.2019.0141>

基于不变网络模型和故障注入的分布式信息系统故障溯源方法

Fault source location algorithm for distributed information system based on invariant network and fault injection

控制与决策. 2020, 35(11): 2723–2732 <https://doi.org/10.13195/j.kzyjc.2019.0214>

基于无标签、不均衡、初值不确定数据的设备健康评估方法

Equipment health risk assessment based on unlabeled, unbalanced data under uncertain initial condition

控制与决策. 2020, 35(11): 2687–2695 <https://doi.org/10.13195/j.kzyjc.2018.1493>

一种基于稀疏系数匹配学习的图像去雾算法

An image dehazing method based on learning framework with sparse coefficient matching

控制与决策. 2020, 35(11): 2797–2802 <https://doi.org/10.13195/j.kzyjc.2018.1764>

基于联合知识表示学习的多模态实体对齐

王会勇, 论兵, 张晓明[†], 孙晓领

(河北科技大学 信息科学与工程学院, 石家庄 050018)

摘要: 基于知识表示学习的实体对齐方法是将多个知识图谱嵌入到低维语义空间, 通过计算实体向量之间的相似度实现对齐. 现有方法往往关注文本信息而忽视图像信息, 导致图像中实体特征信息未得到有效利用. 对此, 提出一种基于联合知识表示学习的多模态实体对齐方法 (ITMEA). 该方法联合多模态 (图像、文本) 数据, 采用 TransE 与 TransD 相结合的知识表示学习模型, 使多模态数据能够嵌入到统一低维语义空间. 在低维语义空间中迭代地学习已对齐多模态实体之间的关系, 从而实现多模态数据的实体对齐. 实验结果表明, ITMEA 在 WN18-IMG 数据集中能够较好地实现多模态实体对齐.

关键词: 多模态数据; 知识表示学习; 知识图谱; 多模态实体对齐; 翻译模型; 种子集合

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.0331

开放科学 (资源服务) 标识码 (OSID):



引用格式: 王会勇, 论兵, 张晓明, 等. 基于联合知识表示学习的多模态实体对齐 [J]. 控制与决策, 2020, 35 (12): 2855-2864.

Multi-modal entity alignment based on joint knowledge representation learning

WANG Hui-yong, LUN Bing, ZHANG Xiao-ming[†], SUN Xiao-ling

(College of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China)

Abstract: Knowledge representation learning (KRL) based entity alignment represents multi-knowledge graphs (KGs) as low-dimensional embeddings and realizes entity alignment by measuring the similarities between entity vectors. Existing approaches, however, frequently primarily focus on the text information but ignore the image information, and lead entity feature information of the image to being underutilized. To address this problem, we propose an approach for image-text multi-modal entity alignment (ITMEA) via joint knowledge representation learning. The approach joints multi-modal (image and text) data, and embeds multi-modal data into a uniform low-dimensional semantic space by a knowledge representation learning model which combines the translating embedding model (TransE) with the dynamic mapping matrix embedding model (TransD). In low-dimensional semantic space, the link-mapping relations among the aligned multi-modal entities can be learned iteratively from the seed set. The learned link-mapping relations can be applied to unaligned entities. In this way, the multi-modal entity alignment can be implemented. Experiment results on WN18-IMG datasets show that the ITMEA can achieve the multi-modal entity alignment, and it is useful for the construction and completion of the open knowledge graph.

Keywords: multi-modal data; knowledge representation learning (KRL); knowledge graphs (KGs); multi-modal entity alignment; translating model; seed set

0 引言

知识图谱 (knowledge graphs, KGs) 存储着大量结构化知识, 这些结构化知识已成功地应用于知识推理^[1]、问答^[2]和推荐^[3]等领域. 目前, 出现了一些典型的大规模知识图谱, 如 YAGO^[4]、DBpedia^[5]和 Freebase^[6]. 在这些知识图谱中, 结构化知识可表示

为三元组 (h, r, t) , 表示其头实体 h 与尾实体 t 之间存在关系 r . 但随着知识图谱中数据量的不断增加, 长尾效应、推理困难和计算量巨大等问题也随之产生^[7]. 为了解决这些问题, Bordes 等^[8]提出了知识表示学习模型 (knowledge representation learning, KRL). KRL 的关键思想是将知识图谱中结构化三元组表示

收稿日期: 2019-03-21; 修回日期: 2019-09-05.

基金项目: 国家自然科学基金项目 (51271033); 河北省自然科学基金项目 (F2018208116); 河北省研究生创新项目 (CXZXZSS2018092).

[†]通讯作者. E-mail: zxmheburst@163.com.

为低维向量(称为嵌入),同时保留原有三元组之间的语义信息.由于知识表示学习方法在低维语义空间中能够高效计算实体和关系的语义联系,使大规模知识图谱得到更有效的利用,受到学术界和工业界的高度关注.

现有知识表示学习方法更关注于单个知识图谱的嵌入,但随着知识图谱应用环境更加多样化,单个知识图谱中的知识可能很难满足各种需求.因此,一些研究工作将不同来源的知识图谱(如YAGO、DBpedia)嵌入到低维空间,通过计算实体之间的相似性实现实体对齐,并取得了较好的效果.然而,大部分基于知识表示学习的实体对齐方法仅使用单一模态的数据(如文本),而忽视了其他模态的数据(如图像),导致其他模态数据中的实体特征信息未被有效利用.为了弥补实体对齐数据模态较为单一的不足,本文将多模态实体对齐作为主要研究内容.多模态实体对齐的研究能够帮助实现多模态知识图谱的构建和补全,并且有利于基于知识图谱应用技术的提升^[9-10].

多模态实体之间不可避免地存在着异构性,因此,如何消除多模态数据之间的异构性成为一项关键挑战.现有的多源知识图谱实体对齐任务大部分使用多源实体来实现^[11-14].由于能够对齐的多源实体之间具有相同或相似的语义信息,将其嵌入到连续的低维向量语义空间中,实体对齐任务变成了一种向量平移问题.通过计算两个实体之间的相似度便能实现多源知识图谱实体对齐,同时还能保留原有的语义信息.因此,本文遵循基于知识表示学习的多源实体对齐方法,使用多模态实体对(图像-文本实体对),在统一低维语义空间中计算多模态实体之间的相似性以实现多模态实体对齐.但如果训练集中能够实现对齐的多模态实体全部由人工进行标注,则是不可能完成的任务且造成巨大的浪费.为了解决这一问题,受Bootstrapping算法^[15]的启发,本文设置一个小型的由人工进行标注的多模态对齐种子集合.

本文的主要贡献如下:

1) 提出一种联合知识表示学习模型(joint knowledge representation learning, JKRL),使得多模态三元组知识能够嵌入到统一的低维语义空间.该模型结合TransE^[8]和TransD^[16]两种知识表示学习模型中实体距离计算方法,并为这两种距离计算方法设置相应的权重,利用各自模型中的优点实现多模态知识的最优嵌入.

2) 提出一种迭代实体对齐规则(iterative entity alignment, IEA).通过迭代地学习已对齐多模态实体

间的链接映射关系,并将这种关系应用到未对齐实体上以发现更多可对齐的多模态实体,这些新发现的对齐多模态实体可以用来帮助下一轮次的对齐学习.

1 相关工作

1.1 实体对齐

实体对齐又称为实体匹配或实体解析,用于判断两个相同或不同知识库中的实体是否指代同一个对象.传统实体对齐方法通常需要大量人工对数据进行标注或精心设计对齐特征.Yago3方法^[17]通过人工精心设计了特征,使Wikipedia中的实体与已有YAGO中的实体实现对齐,并取得了较高的对齐准确率.基于网络语义标签的多源知识库实体对齐模型^[18]利用多种实体标签,实现中文实体对齐并取得了较高的准确率和召回率.这类传统实体对齐方法可以实现较高的实体对齐效果,但使用人工进行数据标注会造成时间的大量浪费和高昂的劳动力成本,同时,设计出的实体特征往往扩展性较差,不具备普遍适用性.

近些年,也有一些方法专注于在不同知识图谱中使用异源、异构知识进行实体对齐.通过概念标注实现多源实体对齐的方法^[19]利用概念注释来丰富知识图谱中的内部链接,迭代预测新的多源知识对.使用这种迭代对齐的方法虽然可以达到较高的准确率,但常常需要消耗大量的训练时间.随着知识表示学习方法的提出,MTransE方法^[20]直接使用已知的知识三元组实现英语-法语的多语言知识对齐.IPTransE方法^[21]使用共享参数实现知识嵌入,并使用一种软规则的方法实现实体对齐.HolisticEM方法^[22]只使用知识图谱中的结构化信息实现三元组知识的向量化表示,并迭代地实现多源实体对齐.JAPE方法^[23]在统一的语义空间中学习实体的嵌入以及不同知识图谱之间的关系,该方法利用实体属性来细化实体嵌入,但当不同知识图谱之间的相关实体属性不同时,实体嵌入的效果会大大下降.

与以往实体对齐模型相比,本文为了克服训练集中多模态对齐实体数量的不足,提出一个多模态对齐种子实体集合.通过学习集合内已对齐多模态实体之间的链接映射关系,标记可能的对齐多模态实体并迭代地将其添加到种子集合中以帮助多模态数据的嵌入.

1.2 知识表示学习

近些年,以深度学习为代表的知识表示学习方法迅速发展并取得了巨大进步.目前,知识表示学习方法从形式上来讲主要分为基于结构的知识表示

学习和基于语义的知识表示学习(如: neural tensor networks (NTN)^[24]、semantic space projection(SSP)^[25]等). 本文使用结构化三元组知识进行知识表示学习, 因此, 将详细介绍基于结构化的知识表示学习方法.

1.2.1 基于翻译模型的方法

TransE是基于翻译模型的代表, 该方法将实体和关系嵌入到连续低维空间, 并将关系视为头实体到尾实体的一种翻译. TransE在处理一对一的问题上表现出了显著的优越性, 但在处理复杂关系(如: 1-N、N-1、N-N)时存在适用性不足的问题. 为了解决TransE处理复杂关系不理想的问题, TransH方法^[26]在嵌入空间中设置了一个超平面, 所有关系均在超平面上, 并在超平面上进行向量计算, 即头实体和尾实体投影到超平面上并进行平移. TransE和TransH模型都将实体和关系嵌入到相同语义空间中, 但实体和关系是不同属性的对象, 不能简单地嵌入到相同的语义空间中. 为此, TransR/CTransR方法^[27]为每种关系设置了不同的语义空间. CTransR是对TransR的进一步改进, 把关系细分为多个子关系并为每个子关系学习相应的向量表示. TransR/CTransR方法区分了实体和关系, 在处理复杂关系上得到了较大的提升, 但由于需要进行矩阵乘法操作, 导致计算量较大、耗时较长.

1.2.2 基于卷积模型的方法

目前除了翻译模型外, 也出现了一些其他的知识表示学习模型, 其中基于卷积模型的方法在部分问题的处理上优于翻译模型. ConvE方法^[28]将嵌入得到的向量重构成 n 维矩阵, 通过卷积神经网络的方法提取特征, 根据提取的特征实现知识的表示. ConvKB方法^[29]通过卷积神经网络的方法捕获实体和关系在嵌入空间中的全局关系和过渡特征, 更好地实现知识图谱的补全. 但由于卷积神经网络需要进行卷积计算, 训练成本远远高于翻译模型, 且进行卷积操作需要的训练数据量也较为庞大.

考虑到训练数据量及训练成本问题, 本文使用基于翻译模型的方法实现多模态结构化知识的嵌入. 与以往基于翻译模型的知识表示学习方法不同, 本文提出的联合知识表示学习模型, 能够同时将文本知识和图像知识嵌入到相同的语义空间中, 消除二者之间的异构性, 在同一个低维语义空间中利用相似度计算来实现对齐, 提高了计算效率.

2 问题描述

本文研究的重点问题是如何利用知识表示学习模型来消除多模态实体间的异构性, 并基于迭代对齐规则实现多模态实体对齐. 值得提出的是: 以往多源

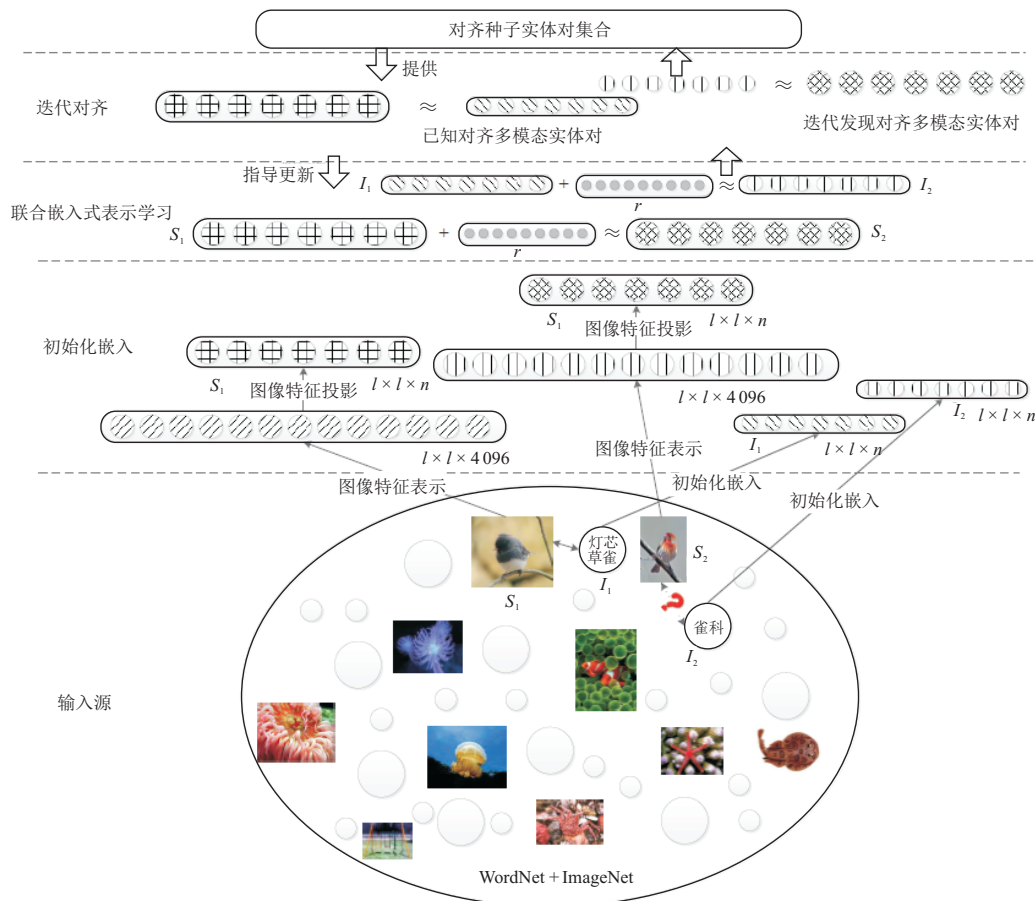


图1 总体框架

知识图谱实体对齐是在相同模态(如文本)下实现多个知识图谱的融合;而本文任务是挖掘不同模态、不同结构知识间实体的链接映射关系,从而实现多模态实体的对齐. 现有基于实体图像的知识表示学习方法大多使用单实体图像,即图像中只有一种实体,这样便于实体识别;同时,对实体的领域也进行了限制,大部分使用易于视觉描述的实体^[30-32]. 本文将使用动物、植物等领域实体实现多模态实体对齐.

本文提出的基于联合知识表示学习的多模态实体对齐(ITMEA)方法的总体框架如图1所示,以图像数据集ImageNet^[33]和文本数据集WordNet^[34]作为输入源,并将这两个数据集进行联合. 其中,假设在新构建的多模态数据集中,已知图像实体 S_1 与文本实体 I_1 (“灯芯草雀”)为已对齐的多模态实体,但图像实体 S_2 与文本实体 I_2 (“雀科”)能否实现对齐尚不确定. 因此,对联合数据集中的图像实体使用图像特征提取方法得到相应的图像实体初始向量,文本实体则通过初始化嵌入得到相应的文本实体初始化向量. 对得到的图像实体和文本实体的初始化向量按照结构化三元组的形式再次嵌入到统一低维语义空间,使图像实体向量和文本实体向量具有相应的语义关系,消除了图像与文本之间的异构性. 通过迭代学习图像实体 S_1 与文本实体 I_1 之间的链接映射关系,发现图像实体 S_2 与文本实体 I_2 也能够建立这样的链接映射关系,同时将新发现的多模态对齐实体添加到种子集合中,并指导下一轮次的多模态知识嵌入.

3 基于知识表示学习的多模态实体对齐方法

首先,对本节中使用到的符号进行介绍. 其中: S 代表图像实体集合; I 代表文本实体集合; R 表示所有关系的集合; E 表示所有实体的集合,即 $E \in S, I$. 本文将嵌入维度表示为 n ,将初始化实体转换矩阵表示为 M .

3.1 基于VGGNet的图像特征提取

图像中包含着重要的语义信息,对实体外形和特征等信息的传递比文本知识更加直观、形象. 因此,在本文方法中使用图像知识作为其中一个输入源.

为了更好地实现图像特征提取,本文对每个图像实体进行图像特征表示,使图像实体初始化嵌入成4096维的向量,实现图像特征向量化表示. 但此时图像特征的维度较高,无法直接与文本实体建立一定的链接关系,因此需要进行图像特征投影,把4096维的图像特征向量映射为 n 维的实体嵌入向量. 其中,图像特征表示采用图像实体识别方法VGGNet^[35],图像

特征投影则由包含 n 个神经元的单层神经网络来完成.

3.1.1 基于VGGNet的图像特征表示

图像特征表示的目的是对图像实体特征进行提取. 本文使用基于卷积神经网络的图像实体识别模型VGGNet,该模型由5个卷积层、3个全连接层和1个softmax层组成. 在图像预处理阶段,本文受Shutova等^[36]和Xie等^[37]的启发,把图像的中心和边缘重构成 224×224 的尺寸,将VGGNet方法的第2个全连接层(fc7)输出的向量作为图像实体特征向量.

3.1.2 基于神经网络的图像特征投影

图像特征投影的目的是把通过图像特征表示得到的图像特征向量投影到实体空间,便于进行图像-文本实体对齐. 本文通过神经网络的方法在图像特征空间与实体特征空间之间搭建一个桥梁,即将图像特征的4096维转换为实体空间的 n 维,其定义为

$$P(d_s, d_i, E_i) = E_i \cdot R^{(d_s, d_i)}. \quad (1)$$

其中: $R^{(d_s, d_i)}$ 是一个转换矩阵, d_s 表示图像特征维度, d_i 表示实体特征维度, E_i 表示第 i 个实体得到的图像特征表示.

3.2 联合知识表示学习

以往知识表示学习方法是在单一模态(如:文本、图像)或单一数据源(如:WordNet、DBpedia等)中进行嵌入化表示. 但由于图像知识与文本知识之间存在异构性,如果分别对二者进行单独嵌入,则会使实体对齐变得极为困难^[37]. 因此,本文提出一种多模态联合知识表示学习模型,旨在将图像知识和文本知识能够嵌入到统一的低维语义空间,消除两者之间的异构性. 如图2所示,该联合知识表示学习模型结合

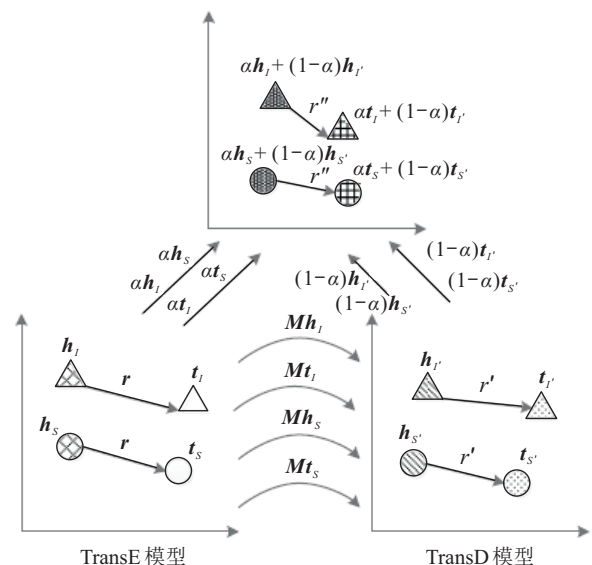


图2 联合知识表示学习模型

TransE 和 TransD 中实体距离计算方法, 使图像知识和文本知识能够同时嵌入到统一的低维语义空间。

TransE 模型使用简单的向量平移(翻译)思想实现知识三元组的表示, 但这种思想处理简单关系更为实用. 与 TransE 相比, TransD 模型设置了一个动态映射矩阵, 将知识三元组按照关系映射到不同低维空间, 对复杂问题的处理更为有效. 本文认为图像与文本的对齐关系是一种简单关系, 而图像实体之间与文本实体之间更多的是复杂关系. 多模态知识仅使用 TransE 模型或 TransD 模型并不能很好地实现实体和关系的准确嵌入, 因此, 在本文提出的 JKRL 方法中结合使用 TransE 模型和 TransD 模型, 并设置了相应的权重, 从而更好地实现多模态知识的嵌入. 具体来说, 本模型为距离计算方法设置权重 α , α 的取值范围为 $[0.00, 1.00]$, 采用网格搜索算法得到最优权重, 设置网格搜索的步长为 k . JKRL 的距离计算方法定义为

$$f_r(h, t) = \alpha(\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp) + (1 - \alpha)(\mathbf{h} + \mathbf{r} - \mathbf{t}) \|^2_2 \quad (2)$$

其中: $\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp$ 为 TransD 的距离计算方法, \mathbf{h}_\perp 和 \mathbf{t}_\perp 分别为 $\mathbf{h}_\perp = \mathbf{M}_{r_h}\mathbf{h}$ 和 $\mathbf{h}_\perp = \mathbf{M}_{r_t}\mathbf{t}$, \mathbf{M}_{r_h} 和 \mathbf{M}_{r_t} 则分别表示头实体和尾实体所对应的空间转换矩阵. TransD 模型为每个关系设置了一个不同的映射矩阵, 使得实体通过映射矩阵嵌入相应关系空间, 这

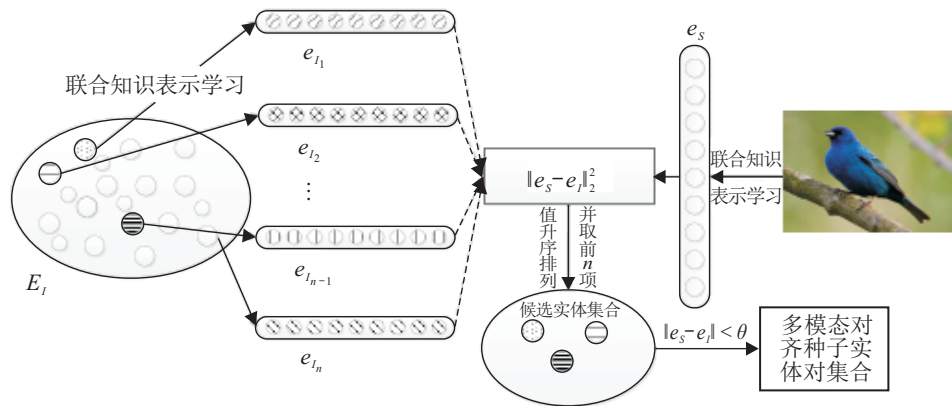


图 3 对齐规则

对于未对齐的图像实体 e_s , 会遍历整个文本实体集合, 计算每个文本实体向量与该图像实体向量之间的距离, 对计算结果进行升序排列, 并取前 m 个实体组成候选对齐实体集合 \hat{E}_I , 并设定一个距离阈值 θ 作为超参数. 如果 $d(e_s, \hat{e}_I) < \theta, \hat{e}_I \in \hat{E}_I$, 则认为 \hat{e}_I 与 e_s 能够实现对齐, 否则认为两个实体之间不能够实现对齐. 逐个将新对齐的图像-文本实体 (e_s, \hat{e}_I) 添加到对齐种子集合 L 中, 将不能实现对齐的图像-文

样, 不同属性的同一实体将会表示为不同向量. $\mathbf{h} + \mathbf{r} - \mathbf{t}$ 为 TransE 的距离计算方法. 损失函数定义为

$$L = \sum_{\mathbf{E}=\{\mathbf{E}_S, \mathbf{E}_I\}} l, \quad (3)$$

$$l = \sum_{(h, r, t) \in T} \sum_{(h', r, t') \in T'} [\gamma + f_r(\mathbf{h}, \mathbf{t}) - f_r(\mathbf{h}', \mathbf{t}')]_+. \quad (4)$$

其中: L 表示损失函数, 损失函数趋近于零时, 模型效果最优; E 表示图像实体 E_I 和文本实体 E_S 的集合; γ 是正例与负例之间的最小间隔, 正例表示知识图谱中已有的结构化三元组, 负例表示训练过程中随机生成的错误三元组, 其形式化定义为

$$T' = \{(h', r, t) | h' \in E\} \cup \{(h, r, t') | t' \in E\}; \quad (5)$$

$[x]_+$ 表示当得到的值为正值时取原值, 否则为零, 这样能够使得正例与负例区分开.

3.3 迭代实体对齐

实体对齐的目的是构造图像实体特征向量与文本特征向量间的链接映射关系. 本文在联合嵌入结果的基础上, 根据多模态实体在低维语义空间中的距离, 实现图像-文本多模态实体对齐. 本文提出的 IEA 方法的距离计算公式为

$$d(e_s, e_I) = \|e_s - e_I\|_2^2, \forall e_s \in E_S, e_I \in E_I. \quad (6)$$

本文的对齐规则如图 3 所示.

本实体对进行舍弃, 且每迭代完一次后, 根据对齐结果更新多模态对齐种子集合, 此时, 多模态对齐种子集合中所有实体对均为已对齐实体对.

3.4 算法描述

ITMEA 方法使用结构化多模态三元组 $\{(h_s, r, t_s), (h_s, r, t_I)\}$ 和 $\{(h_I, r, t_s), (h_I, r, t_I)\}$ 作为数据输入. 在数据初始化过程中, 将正确三元组的头实体或尾实体进行替换以形成负例, 并使正例与负例的数量

相等. 首先,利用距离计算公式 $\|\alpha(\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp) + (1 - \alpha)(\mathbf{h} + \mathbf{r} - \mathbf{t})\|_2^2$,计算头、尾实体间的距离,得到嵌入向量;然后,根据 $\|e_S - e_I\|_2^2 < \theta$ 找到更多的对齐实体对;最后,更新损失函数,将新找到的对齐实体对添加进对齐种子实体对集合 L 中,并进行下一轮的迭代训练. 算法描述如下:

step 1: 训练集 $S = \{(h_S, r, t_S), (h_S, r, t_I)\} \cup \{(h_I, r, t_S), (h_I, r, t_I)\}$ //文本实体集合 E_I ,图像实体集合 E_S ,关系集合 R ,最小间隔 γ ,嵌入维度 k ;

step 2: 对齐种子集合 $L = (e_S, e_I) // e_S \in E_S, e_I \in E_I$;

step 3: $r \leftarrow$ 每种关系 $r \in R$ 初始化嵌入;

step 4: $e \leftarrow$ 每个实体 $e \in \{E_I, E_S\}$ 初始化嵌入;

step 5: $M \leftarrow re^\top + I^{m \times n}$;

step 6: $e_\perp \leftarrow Me + I^{m \times n}$;

step 7: $S_{batch} \leftarrow \text{Sample}(S, b)$ //小批量取样;

step 8: $T_{batch} \leftarrow \emptyset$ //初始化三元组集合;

step 9: for $(h, r, t) \in S_{batch}$ do

$(h', r, t') \leftarrow \text{Sample}(S'(h, r, t))$ //打破

正确三元组,生成负例

$T_{batch} \leftarrow T_{batch} \cup \{(h, r, t), (h', r, t')\}$

$d(\mathbf{h} + \mathbf{r} - \mathbf{t}) = \|\alpha(\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp)\|_2^2 + \|(1 - \alpha)(\mathbf{h} + \mathbf{r} - \mathbf{t})\|_2^2$;

step 10: for $e_S \in E_S$ do

for $e_I \in E_I$ do //为每个图像实体遍历文本实体集合并计算距离

$d(e_S, e_I) = \|e_S - e_I\|_2^2$;

step 11: Align $\leftarrow (\min_n d(e_S, e_I)) < \theta$ //对计算结果进行排序;

step 12: 嵌入向量

w.r.t $\sum_0^{T_b} \text{SGD}[\gamma + d(\mathbf{h} + \mathbf{r} - \mathbf{t}) - d(\mathbf{h}' + \mathbf{r} - \mathbf{t}')]_+$

//使用SGD随机梯度下降方法优化参数;

step 13: 实体对齐种子集合 $L \leftarrow L \cup \text{Align}$;

step 14: h_S, h_I, t_S, t_I, r //此时得到的图像实体表示向量和文本实体表示向量都带有相应的语义信息和对齐信息.

4 实验分析

4.1 数据集

本文构建一个包含图像和文本的多模态数据集 WN18-IMG. WN18-IMG 中的文本数据来自于文本知识库 WordNet, 图像数据来自于图像知识库 ImageNet, 并将 WN18-IMG 随机分为训练集、测试集

和验证集. 为了保证能够实现多模态实体对齐, 应使 WN18-IMG 中所有的图像实体都有对应的文本实体. 由于在构建 WN18-IMG 数据集时没有设置明显的负例, 本文将遵循以往知识表示学习模型^[8,16,25,37]的负例生成方法, 把三元组中的头实体和尾实体随机替换形成负例, 且使正例与负例数量相等. 同时, 为了能够验证本文所提出的方法具有普遍适用性, 还使用 WN9-IMG^[37]进行了对比实验. 数据集统计信息如表 1 所示.

表 1 数据集统计信息

数据集	关系	实体	训练集	验证集	测试集
WN18-IMG	18	5 684	13 634	466	766
WN9-IMG	9	6 555	11 741	1 337	1 319

4.2 实验设置

本文使用 VGGNet 实验中得到的图像识别最优参数, 这样可节省大量数据预处理时间. 此外, 本文采用随机梯度下降法 (SGD) 对上述方法的值进行最优化处理. 对于本文提出的 JKRL 模型中正例、负例之间最小间隔 γ 设置为 $\{0.5, 1.0, 1.5, 2.0\}$, 经验学习率设置为 0.001 ^[8,21], 迭代次数设定为 $\{500, 1500, 2500, 3500, 4500\}$, θ 设置为 $\{0.5, 1.0, 2.0, 3.0, 4.0\}$, 对于网格搜索的步长 k 设置为 $\{0.05, 0.10, 0.15, 0.20\}$. 在实验中用于对比实验的 TransE 和 TransD 模型的参数依据原论文进行设置. 在本文中, 图像特征向量 d_i 为 4096, 关系和实体嵌入维数 n 设置为 100. 通过实验得到本模型的最优参数为 $\gamma = 1.0, k = 0.5, \theta = 1.0$.

本文的实验是在配置了 Intel Xeon E5-1620 3.6 GHz 的 CPU, Nvidia GeForce GTX 1070 Ti 的 GPU 和 32 GB 内存的个人工作站上进行的, 操作系统为 Ubuntu.

4.3 多模态实体对齐

4.3.1 评价标准

多模态实体对齐任务的目的是判断不同模态数据中相同或相似的实体能否建立一定的对齐关系. 以往实体对齐任务使用正确实体得分函数的平均排名 (mean rank of correct entity, MR) 和预测结果前 10 项 (Hit@10)、前 1 项 (Hit@1) 实体的准确率作为评价标准, 因此, 本文也使用 Hit@10、Hit@1 和 MR 作为多模态实体对齐的评价标准. 为了进行对比实验, 本文使用 MTransE 效果较好的两个模型 (基于翻译的模型和基于线性变化的模型). 同时, 在传统表示学习模型 TransE 和 TransD 对多模态数据进行嵌入

的基础上,使用迭代对齐规则(IEA)实现多模态实体对齐的实验来进一步验证 ITMEA 方法的有效性. 本文将 MTransE 的两个模型分别命名为 MTransE(T) 和 MTransE(L),将使用迭代对齐规则的 TransE 和 TransD 模型命名为 TransE+IEA 和 TransD+IEA.

4.3.2 实验结果

表 2 中给出了在多模态数据集 WN18-IMG 和 WN9-IMG 上 Hit@10、Hit@1 和 MR 的结果. 从表 2 中可以看出:

1) TransE+IEA、TransD+IEA 和 ITMEA 的结果优于 MTransE(T) 和 MTransE(L) 方法. 由此可以说明,通过迭代学习少量已对齐的多模态之间的关系便能够对实体对齐的结果有所提升,且能够大大减少标注数据的工作量.

2) 在使用迭代对齐规则的方法中,ITMEA 的结果与 TransE+IEA 和 TransD+IEA 的结果相比,在一定程度上有所提升. 这也验证了本文之前介绍的传统表示学习模型在对多模态数据进行嵌入表示时,在一定程度上会损失掉一些视觉信息从而影响对齐结果. 同时,结合多种表示学习模型的优点能够实现更加精准的实体对齐.

表 2 多模态实体对齐结果

方法	WN18-IMG			WN9-IMG		
	Hit@10	Hit@1	MR	Hit@10	Hit@1	MR
MTransE(T)	61.0	38.5	237.7	79.2	46.1	419.2
MTransE(L)	35.4	13.6	547.7	62.4	30.8	647.6
TransE+IEA	62.3	35.4	30.0	84.7	58.9	138.9
TransD+IEA	72.4	47.7	27.1	84.3	50.2	159.9
ITMEA	79.7	52.8	38.4	86.7	61.9	139.3

图 4 展示了在 WN9-IMG 数据集上,不同模型迭代相同次数的耗时情况. 从图 4 中可以看出 MTransE(T) 耗时最短,而本文提出的 ITMEA 方法耗时最长. MTransE(T) 模型直接进行不同模态实体间的相互计算,且使用 TransE 进行训练,其时间复杂度为 $O(Nt)$. 同样,进行不同模态实体间相互计算的

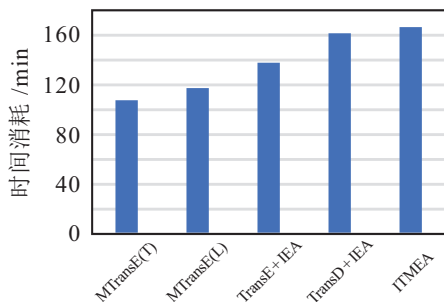


图 4 不同算法时间消耗

MTransE(L) 模型由于需要进行线性变换,其时间消耗比 MTransE(T) 长. 然而,本文提出的 ITMEA 方法由于结合了 TransE 和 TransD 两种距离计算方法且迭代对齐不同模态实体,其时间复杂度高于对比模型,导致耗时较长.

4.4 三元组分类

4.4.1 评价标准

三元组分类任务的目的是判断给定的三元组 (h, r, t) 是否正确,其本质是一个二分类问题. 本文为每种关系设置一个专用的阈值 δ_r ,使用验证集对阈值 δ_r 进行优化. 在测试阶段,使用 3.2 节中提出的损失函数对测试集中每个三元组进行计算,当得分小于阈值时判断为分类正确,否则认为错误. 同时,为了验证添加图像信息是否有助于三元组分类的效果,使用未添加图像信息的 WordNet 的子集 WN11 来进行对比.

4.4.2 实验结果

表 3 中给出了 TransE、TransD 和 JKRL 知识表示学习模型,在 WN11、WN18-IMG 和 WN9-IMG 三个数据集上进行三元组分类的实验结果. 纵向对比表 3 中的结果可以看出:在 WN18-IMG 和 WN9-IMG 多模态数据集上,JKRL 模型均优于 TransE 和 TransD 模型;但在纯文本数据集上,JKRL 模型优于 TransE 模型但低于 TransD 模型. 这是由于 JKRL 模型为每种模态的数据分别学习了相应权重,但对单模态数据进行嵌入时损失了部分信息,从而造成三元组分类结果降低. 横向对比三元组分类结果可以发现,任意一种知识表示学习模型在多模态数据集 WN18-IMG 和 WN9-IMG 上得到的结果均优于单模态数据集 WN11. 由此表明,向文本中添加图像信息能够提升三元组分类任务的准确率.

表 3 三元组分类结果

模型	WN11	WN18-IMG	WN9-IMG
TransE	75.9	89.4	90.8
TransD	86.4	90.6	92.4
JKRL	81.6	91.7	93.8

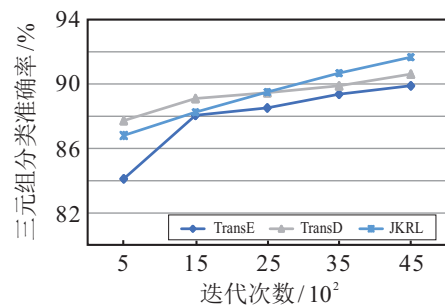


图 5 三元组分类准确率

图5展示了在WN18-IMG多模态数据集中,三元组分类结果随迭代次数的变化情况.可以看出,随着迭代次数的增加,三元组分类的准确率不断提升,但迭代3500次之后,增长的趋势有所减缓.

4.5 链接预测

4.5.1 评价标准

链接预测的目的是当结构化三元组中缺失头实体、关系或尾实体中的任意一个时,均能够成功预测缺失的实体或关系.链接预测主要关注的是实体预测,这个任务被广泛应用于知识表示学习模型的评价中.本文使用两种评价标准作为评测指标:1)正确实体得分函数的平均排名(mean rank of correct entity, MR);2)前10个实体的正确率(Hit@10).此外,本文还遵循TransE模型中使用的“raw”和“filter”两个评估指标.其中:“raw”表示原始未经处理的数据,而“filter”表示删掉负例生成环节中产生的正确三元组数据(也称“污染”三元组).与其相比,在本文实验中还使用了文献[38]提出的无类型约束(TransE(N)、TransD(N)、JKRL(N))和有类型约束(TransE、TransD、JKRL)两个指标作为评价标准.其中,有类型约束指的是为每种关系固定好头实体和尾实体的类型.

4.5.2 实验结果

图6展示了在WN18-IMG多模态数据集中,正确实体得分函数的平均排名(MR)随迭代次数的变化趋势.可以看出,随着迭代次数的增加,正确实体得分函数的平均排名(MR)逐渐降低并趋于稳定.且TransE、TransD和JKRL模型的值优于TransE(N)、TransD(N)、JKRL(N)的值.

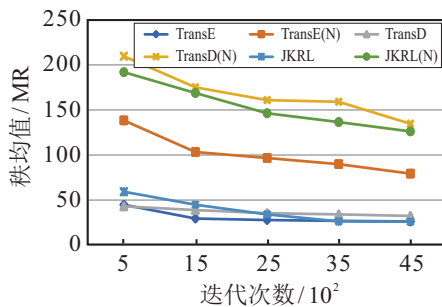


图6 MR随迭代次数的变化

在WN18-IMG多模态数据集中,Hit@10准确率随迭代次数的变化情况如图7所示.从图7中可以看出,随着迭代次数的增加,Hit@10的准确率不断提升,同时TransE、TransD和JKRL模型的准确率均优于TransE(N)、TransD(N)、JKRL(N)的准确率.

从图6和图7可以得出:1)JKRL模型在处理实体链接预测任务上优于TransE、TransD模型,说明对于

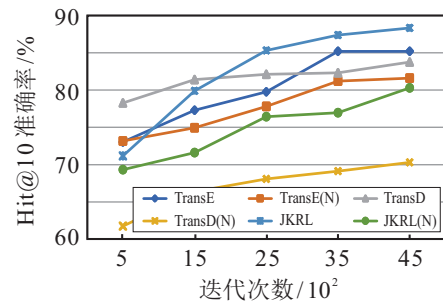


图7 Hit@10随迭代次数的变化

多模态数据的嵌入需要更加有效的知识表示学习模型;2)对实体的类型进行约束能够提升实体链接预测的效果.

表4中进一步展示了MR和Hit@10在多模态数据集WN18-IMG和WN9-IMG上取得的最优值.从表4中可以发现:1)JKRL模型在MR和Hit@10上取得的结果均优于TransE和TransD模型,说明知识表示学习模型JKRL能够帮助多模态数据的嵌入;2)在相同数据集中,有类型约束的实验结果与无类型约束的实验结果相比有较为明显的提升,说明对实体的类型进行约束有助于提高知识表示学习的效果;3)使用任意一种知识表示学习模型对多模态数据进行嵌入,删掉其他正确实体后的结果明显优于未经处理的数据,说明对数据进行预处理有助于实体链接预测结果的提升.

表4 实体链接预测结果

模型	WN18-IMG				WN9-IMG			
	mean rank		Hit@10		mean rank		Hit@10	
	raw	filter	raw	filter	raw	filter	raw	filter
TransE(N)	111.1	47.8	74.8	77.7	230.1	221.6	71.4	83.0
TransE	89.7	26.2	81.6	85.2	142.8	134.4	72.0	83.2
TransD(N)	166.1	61.7	63.9	74.5	260.8	252.9	72.3	84.2
TransD	141.5	34.1	70.3	82.3	175.1	167.3	75.0	85.9
JKRL(N)	147.5	47.2	74.3	79.3	236.0	227.7	74.0	84.7
JKRL	126.1	25.9	80.2	88.3	162.4	124.1	75.2	87.5

5 结论

现有基于知识表示学习的实体对齐方法处理数据模态较为单一,为此,本文提出了一种基于联合知识表示学习的多模态实体对齐方法ITMEA.在理论价值方面,ITMEA使用联合知识表示学习模型JKRL,将图像知识和文本知识嵌入到统一低维语义空间,通过IEA方法迭代学习对齐种子实体集合中已对齐多模态实体间的关系,并将学习到的关系应用到全部实体中以实现多模态实体对齐.实验结果表

明,本文提出的多模态实体对齐方法与现有实体对齐方法相比,对齐准确率在一定程度上有所提升.此外,本文在多模态数据集上使用实体链接预测和三元组分类对所提出的JKRL模型进行评价.实验结果显示,JKRL模型优于基准线.在应用价值方面,本文提出的多模态实体对齐方法有助于开放的多模态知识图谱构建,同时为跨模态检索和视觉问答等领域提供了新的思路和方法.

本文对迭代过程中新找到的对齐实体未进行置信度评估,这可能会将错误的多模态实体对添加到种子集合中,从而影响之后的迭代效果.为了解决这一问题,后续工作将从以下几点入手:1)引入软规则方法对新对齐的实体对进行置信度评估;2)探究更加有效的知识表示学习模型,使多模态知识实现最优嵌入;3)研究更高效的图像特征提取模型并引入注意力机制,从而实现图像实体特征的精确提取.

参考文献(References)

- [1] Yang F, Yang Z L, Cohen W. Differentiable learning of logical rules for knowledge base reasoning[C]. Proceedings of the International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc, 2017: 2319-2328.
- [2] Dubey M, Banerjee D, Chaudhuri D, et al. EARL: Joint entity and relation linking for question answering over knowledge graphs[C]. Proceedings of the 17th International Semantic Web Conference. Monterey: Springer Int Publishing, 2018: 108-126.
- [3] Wang H W, Zhang F Z, Xie X, et al. DKN: Deep knowledge-aware network for new recommendation[C]. Proceedings of the 2018 World Wide Web Conference. Lyon: Springer Int Publishing, 2018: 1835-1844.
- [4] Suchanek F M, Kasneci G, Weikum G. Yago: A core of semantic knowledge[C]. Proceedings of the 2007 International World Wide Web Conference. Banff: Springer Int Publishing, 2007: 697-706.
- [5] Lehmann J, Isele R, Jakob M, et al. DBpedia—A large-scale, multilingual knowledge base extracted from wikipedia[J]. Semantic Web, 2015, 6(2): 167-195.
- [6] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge[C]. Proceedings of the 14th ACM SIGMOD International Conference on Management of Data. Vancouver: ACM, 2008: 1247-1250.
- [7] 刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. 计算机研究与发展, 2016, 53(2): 247-261. (Liu Z Y, Sun M S, Lin Y K, et al. Knowledge representation learning: A review[J]. Journal of Computer Research and Development, 2016, 53(2): 247-261.)
- [8] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]. Proceedings of the International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc, 2013: 2787-2795.
- [9] Antol S, Agrawal A, Lu J, et al. VQA: Visual question answering[C]. Proceedings of the IEEE International Conference on Computer Vision. Chile: IEEE, 2015: 2425-2433.
- [10] Huang F Q, Zhang X M, Xu J, et al. Network embedding by fusing multimodal contents and links[J]. Knowledge-Based Systems, 2019, 171: 44-55.
- [11] Liu W, Liu J, Wu M, et al. Representation learning over multiple knowledge graphs for knowledge graphs alignment[J]. Neurocomputing, 2018, 320: 12-24.
- [12] 赵军, 刘康, 何世柱, 等. 知识图谱[M]. 北京: 高等教育出版社, 2019: 1-254. (Zhao J, Liu K, He S Z, et al. Knowledge graph[M]. Beijing: Higher Education Press, 2019: 1-254.)
- [13] Chen M H, Tian Y, Chang K W, et al. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment[C]. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: Morgan Kaufmann Inc, 2018: 3998-4004.
- [14] Xu K, Wang L W, Yu M, et al. Cross-lingual knowledge graph alignment via graph matching neural network[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: ACL, 2019: 3156-3161.
- [15] Sun Z Q, Hu W, Zhang Q H, et al. Bootstrapping entity alignment with knowledge graph embedding[C]. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: Morgan Kaufmann Inc, 2018: 4396-4402.
- [16] Ji G, He S, Xu L, et al. Knowledge graph embedding via dynamic mapping matrix[C]. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Beijing: ACL, 2015: 687-696.
- [17] Mahdisoltani F, Biega J, Suchanek F M. Yago3: A knowledge base from multilingual wikipeidias[C]. Proceedings of the 7th Biennial Conference on Innovative Data Systems Research. Asilomar: ACM, 2013.
- [18] 王雪鹏, 刘康, 何世柱, 等. 基于网络语义标签的多源知识库实体对齐算法[J]. 计算机学报, 2017, 40(3): 701-711. (Wang X P, Liu K, Heng S Z, et al. Multi-source knowledge bases entity alignment by leveraging semantic tags[J]. Chinese Journal of Computers, 2017, 40(3): 701-711.)
- [19] Wang Z, Li J. Boosting cross-lingual knowledge linking via concept annotation[C]. Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Beijing: Morgan Kaufmann Inc, 2013: 2733-2739.
- [20] Chen M H, Tian Y T, Yang M H, et al. Multi-lingual

- knowledge graph embeddings for cross-lingual knowledge alignment[J]. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: Morgan Kaufmann Inc, 2017: 1511-1517.
- [21] Zhu H, Xie R B, Liu Z Y, et al. Iterative entity alignment via joint knowledge embeddings[C]. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: Morgan Kaufmann Inc, 2017: 4258-4264.
- [22] Pershina M, Yakout M, Chakrabarti K. Holistic entity matching across knowledge graphs[C]. Proceedings of the IEEE International Conference on Cloud and Big Data Computing. Beijing: IEEE, 2015: 1585-1590.
- [23] Sun Z Q, Hu W, Li C K. Cross-lingual entity alignment via joint attribute-preserving embedding[C]. Proceedings of the 16th International Semantic Web Conference. Vienna: Springer, 2017: 628-644.
- [24] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]. Proceedings of the International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc, 2013: 926-934.
- [25] Xiao H, Huang M, Meng L, et al. SSP: Semantic space projection for knowledge graph embedding with text descriptions[C]. Proceedings of the 31st Association for the Advancement of Artificial Intelligence. San Francisco: AAAI, 2017: 3104-3110.
- [26] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]. Proceedings of the 28th Association for the Advancement of Artificial Intelligence. Québec: AAAI, 2014: 1112-1119.
- [27] Lin Y, Liu Z, Zhu X, et al. Learning entity and relation embeddings for knowledge graph completion[C]. Proceedings of the 29th Association for the Advancement of Artificial Intelligence. Hyatt Regency: AAAI, 2015: 2181-2187.
- [28] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2d knowledge graph embeddings[C]. Proceedings of the 32nd Association for the Advancement of Artificial Intelligence. New Orleans: AAAI, 2018: 1811-1818.
- [29] Nguyen D Q, Nguyen T D, Nguyen D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network[C]. Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: ACL, 2018: 327-333.
- [30] Pezeshkpour P, Chen L Y, Singh S. Embedding multimodal relational data for knowledge base completion[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018: 3208-3218.
- [31] Cui P, Liu S W, Zhu W W. General knowledge embedded image representation learning[J]. IEEE Transactions on Multimedia, 2017, 20(1): 198-207.
- [32] Collett G, Zhang T, Moens M F. Imagined visual representations as multimodal embeddings[C]. Proceedings of the 35th Association for the Advancement of Artificial Intelligence. San Francisco: AAAI, 2017: 4378-4384.
- [33] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 248-255.
- [34] Miller G A. WordNet: A lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [35] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of the International Conference on Learning Representations. San Diego: Arxiv-prints, 2015: 1-14.
- [36] Shutova E, Kiela D, Maillard J. Black holes and white rabbits: Metaphor identification with visual features[C]. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: ACL, 2016: 160-170.
- [37] Xie R B, Liu Z Y, Luan H B, et al. Image-embodied knowledge representation learning[C]. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: Morgan Kaufmann Inc, 2017: 3140-3146.
- [38] Han X, Cao S, Lv X, et al. OpenKE: An open toolkit for knowledge embedding[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018: 139-144.

作者简介

王会勇(1980—),男,讲师,博士,从事模式识别与机器学习、语义Web、知识图谱、知识表示学习等研究, E-mail: wanghuiyong815@163.com;

论兵(1994—),男,硕士生,从事语义Web、知识图谱、知识表示学习的研究, E-mail: bing_lun1994@163.com;

张晓明(1975—),男,教授,博士,从事语义计算、知识图谱等研究, E-mail: zxmhebust@163.com;

孙晓颖(1994—),女,硕士生,从事语义Web、知识图谱的研究, E-mail:15028104457@163.com.

(责任编辑:李君玲)