

融合多源信息的知识表示学习研究

(申请清华大学工学硕士学位论文)

培养单位：计算机科学与技术系

学 科：计算机科学与技术

研 生：谢若冰

指导教师：孙茂松教授

二〇一七年五月

Multi-source Information for Knowledge Representation Learning

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Master of Science

in

Computer Science and Technology

by

Xie Ruobing

Thesis Supervisor : Professor Sun Maosong

May, 2017

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

（保密的论文在解密后应遵守此规定）

作者签名： _____

导师签名： _____

日 期： _____

日 期： _____

摘 要

知识图谱是结构化储存人类知识的数据库，在信息检索、问答系统和智能对话等人工智能相关任务中广为运用。而基于分布式表示的知识表示学习，构建了从知识图谱到知识驱动的应用之间的关键桥梁。

本文旨在充分利用多源信息辅助构建更好的知识表示，从而提高知识表示学习相关任务的性能。这些多源信息包括了实体描述的文本信息，实体层次类型的结构化信息，以及实体图像的视觉信息等多模态的信息，能够作为知识图谱三元组结构化信息的有效补充，具有重要的研究意义。

经典的知识表示学习模型往往仅关注知识图谱自身的三元组结构信息，而忽略了蕴含在多源信息中的丰富知识。因此，我们针对实体描述、实体层次类型和实体图像等多源信息，基于平移模型的思想分别设计了不同的模型，尝试将多源信息引入知识表示中。在融合实体描述时，我们对每个实体建立两种表示，并使用神经网络模型构建实体基于描述的表示；在融合实体层次类型时，我们通过基于类型层次结构的映射矩阵，构建实体在不同类型下的不同表示，同时使用了实体类型限制进一步提升知识表示性能；在融合实体图像时，我们通过图像表示模块和图像映射模块构建图像的特征表示，并使用注意力机制进行实体图像的多实例学习，构建实体基于图像的表示。

我们在知识表示学习的经典任务，如知识图谱补全、三元组分类、实体类型分类等任务上，将模型与经典的基线模型进行对比评测。实验结果充分说明，多源信息中蕴含着的丰富信息能够辅助建立更好的知识表示，也说明了我们融合多源信息的知识表示学习模型能够充分利用以上多源信息，在实际知识表示学习任务中取得了更好的实验效果。

关键词：知识图谱；知识表示学习；实体描述；实体层次类型；实体图像

Abstract

Knowledge graphs store structural information of human knowledge, which have been widely used in various tasks of artificial intelligence including information retrieval, question answering and dialogue systems. Knowledge representation learning (KRL), which is based on distributed representations, becomes the essential bridge between knowledge graphs and knowledge-driven applications.

This work aims to fully utilize multi-source information for constructing better knowledge representations and improving knowledge tasks. The multi-source information includes entity descriptions (textual information), entity hierarchical types (structural information) and entity images (visual information), which is an effective supplementary information for the structural information in triples.

Conventional KRL models only focus on the structural information in triples, ignoring the rich multi-source information. To use the multi-source information in entity descriptions, hierarchical types and images, we propose different KRL models based on translation assumption. For entity descriptions, we build two kinds of representations for each entity, and use neural networks to build the description-based representations. For entity hierarchical types, we assume that entities should have different representations in different types with hierarchical type-specific projection matrices, and we also propose type constraints for better performances. For entity images, we extract visual features with image representation module and image projection module, and then introduce attention for multi-instance learning of different images.

We evaluate our models on tasks like knowledge graph completion, triple classification and entity classification compared with classical baselines. Evaluation results demonstrate that multi-source information contains significant information for better knowledge representations. It also indicates that our models can well encode these kinds of multi-source information into knowledge representations, and achieve better performances in real-world KRL tasks.

Key words: knowledge graph; knowledge representation learning; entity description; entity hierarchical type; entity image

目 录

第1章 引言	1
1.1 研究背景	1
1.2 研究内容	3
1.3 相关工作	4
1.3.1 知识表示学习符号体系	4
1.3.2 知识表示学习经典模型	4
1.3.3 平移模型及其拓展模型	6
1.4 章节结构	8
第2章 融合实体描述信息的知识表示学习	10
2.1 引言	10
2.2 相关工作	12
2.3 算法模型	12
2.3.1 符号体系和重要概念	13
2.3.2 模型整体架构和学习框架	13
2.3.3 连续词袋模型	15
2.3.4 卷积神经网络模型	15
2.3.5 模型优化与具体实现	18
2.4 实验设计与结果分析	19
2.4.1 实验数据集	19
2.4.2 实验与模型参数设置	20
2.4.3 知识图谱补全结果与分析	21
2.4.4 实体类型分类结果与分析	23
2.4.5 零样本知识表示学习	25
2.5 本章小结	28
第3章 融合实体层次类型信息的知识表示学习	29
3.1 引言	29
3.2 相关工作	31
3.3 算法模型	32
3.3.1 符号体系与重要概念	32
3.3.2 模型整体架构与学习框架	32

3.3.3	层次类型编码器	33
3.3.4	基于实体类型的限制	36
3.3.5	模型优化与具体实现	38
3.4	实验设计与结果分析	38
3.4.1	实验数据集	39
3.4.2	实验与模型参数设置	39
3.4.3	知识图谱补全结果与分析	40
3.4.4	三元组分类结果与分析	45
3.5	本章小结	46
第4章	融合实体图像信息的知识表示学习	47
4.1	引言	47
4.2	相关工作	48
4.3	算法模型	49
4.3.1	符号体系与重要概念	49
4.3.2	模型整体架构与学习框架	50
4.3.3	实体图像编码器	51
4.3.4	多实例学习中的注意力机制	53
4.3.5	模型优化与具体实现	54
4.4	实验设计与结果分析	54
4.4.1	实验数据集	55
4.4.2	实验与模型参数设置	55
4.4.3	知识图谱补全结果与分析	56
4.4.4	三元组分类结果与分析	58
4.4.5	结果样例分析	59
4.5	本章小结	62
第5章	总结与展望	63
5.1	工作总结	63
5.2	工作展望	64
参考文献	66
致 谢	70
声 明	71
个人简历、在学期间发表的学术论文与研究成果	72

主要符号对照表

<i>KG</i>	知识图谱 (Knowledge graph)
<i>KRL</i>	知识表示学习 (Knowledge representation learning)
<i>TransE</i>	平移模型
<i>DKRL</i>	融合实体描述的知识表示学习模型 (Description-embodied knowledge representation learning)
<i>TKRL</i>	融合实体类型的知识表示学习模型 (Type-embodied knowledge representation learning)
<i>IKRL</i>	融合实体图像的知识表示学习模型 (Image-embodied knowledge representation learning)
<i>CBOW</i>	连续词袋模型 (Continuous bag-of-words)
<i>CNN</i>	卷积神经网络 (Convolutional neural network)
<i>RHE</i>	递归层次编码器 (Recursive hierarchy encoder)
<i>WHE</i>	加权层次编码器 (Weighted hierarchy encoder)
<i>STC</i>	软类型限制 (Soft type constraint)
<i>TCE</i>	测试中的类型限制 (Type constraint in evaluation)
<i>SBR</i>	基于结构的表示 (Structure-based representation)
<i>DBR</i>	基于描述的表示 (Description-based representation)
<i>IBR</i>	基于图像的表示 (Image-based representation)

第1章 引言

1.1 研究背景

知识库（knowledge base, KB）是将人类知识结构化形成的知识系统，其中包含了基本事实、通用规则和其它有关信息，常用于人工智能领域的知识抽取、存储与推理等任务。知识库是人工智能研究和智能信息服务中一项基础核心技术，能够赋予智能体深度理解、精准查询与逻辑推理等能力，被广泛运用于搜索引擎、问答系统、智能对话系统以及个性化推荐等知识驱动的任务。

为了高效地储存与使用知识，人们结合专家标注与机器自动标注等方法，构建了大量开放领域和专有领域的知识库，如Wikidata^[1]，Freebase^[2]，DBpedia^[3]，YAGO^[4]以及WordNet^[5]等经典的知识库。以Wikidata为例，截止至2017年4月5日，Wikidata中已经包含了超过2500万实体。与此同时，国内外各大互联网公司也推出了众多基于知识库的产品，如谷歌知识图谱、百度知心和微软Bing Satori等，各种知识驱动的智能应用背后更是少不了知识库的支持。图1.1给出了一些常用的大规模知识库。



图 1.1 一些常用的大规模知识图谱^[2-5]

知识库将世界上的具象事物与抽象概念等表示为实体（entity），将实体之间的联系表示为关系（relation）。目前，经典的知识库常常使用三元组（triple fact）的形式对实体与实体之间的关系进行储存，类似于万维网联盟发布采用的资源描述框架（resource description framework, RDF）。以“北京是中国的首都”这一知识为例，知识库将此条信息储存为（北京，是...首都，中国），其中北京称为首实体（head entity），中国称为尾实体（tail entity），是...首都称为关系（relation）。通过众多三元组的链接，知识库中储存的知识形成了一张巨大的知识图谱（Knowledge graph, KG），其中实体被看作知识图谱中的节点，而关系则被看作连接节点的边。

在信息爆炸的时代，知识库中已有海量知识储存，同时每天都会产生新的知识，已有知识也面临着补充与更新。如何更好地对知识进行储存与表示，更有效率地利用知识，成为我们如今亟待研究的课题。近年来，基于知识图谱的知识表示学习主要面临着两个难题：（1）计算效率低下：基于图结构的知识表示虽然简洁直观，但是在利用知识图谱进行检索与多步推理时，常常需要设计专门的图算法以完成任务。这些图算法往往计算复杂度较高，在目前的大规模知识图谱上难以快速运行，且难以拓展至其它情况。（2）数据稀疏性：大规模知识图谱中的实体与关系往往也存在着长尾分布，有很多实体只存在着极少数的关系与之相连。对这些稀疏的实体和关系的预测与推理难以得到较高的准确率。

为了解决计算效率低与数据稀疏两个问题，知识表示学习（Knowledge representation learning, KRL）技术被广泛研究与运用。知识表示学习基于分布式表示^[6]（distributed representation）的思想，将实体（或关系）的语义信息映射到低维稠密实值的向量空间中，使得语义相似的两个对象之间距离也相近。而传统对知识的表示通常使用独热表示（one-hot representation），即将知识表示成为一个长向量，只有该知识对应的特定维非零，而其它所有维度都为零。这种表示方式构建简单，在自然语言处理的各个任务中被广泛使用。

与独热表示相比，分布式表示有着以下几点优点：（1）分布式表示学习到的是低维向量。这使得实体与关系之间的语义联系能够在低维空间中得以高速计算，显著提高计算效率。（2）独热表示基于所有对象相互独立的假设，所有向量之间两两正交，丢失了大量对象之间的相似及关联信息。而分布式表示则能通过稠密低维向量之间的相似度计算表达对象之间的关系，较好地缓解了数据稀疏带来的问题。（3）分布式表示能够将多源异质信息映射到同一语义空间中，建立多源跨模态的信息交互，且分布式表示也能更便捷地融入深度学习的模型框架中。近年来，知识表示学习在知识图谱补全、知识获取、融合与推理等任务上被广泛运用，显著地提高了知识驱动型应用的性能。

图1.2给出了一个知识驱动工程的简要流程图。底层的知识构建旨在从多源异质的信息中抽取知识，并将其结构化储存，通常包括关系分类、信息抽取等任务；顶层的知识应用旨在利用已构建好的知识库或知识表示，帮助更好地完成信息检索、问答系统和智能对话系统等知识驱动的任务。而中间层的知识表示，则成为连接底层知识构建与顶层知识应用的不可或缺的桥梁。知识表示学习具有提升计算效率、缓解数据稀疏的影响、便于多源异质信息的交互融合等优点，正受到研究者的广泛关注。

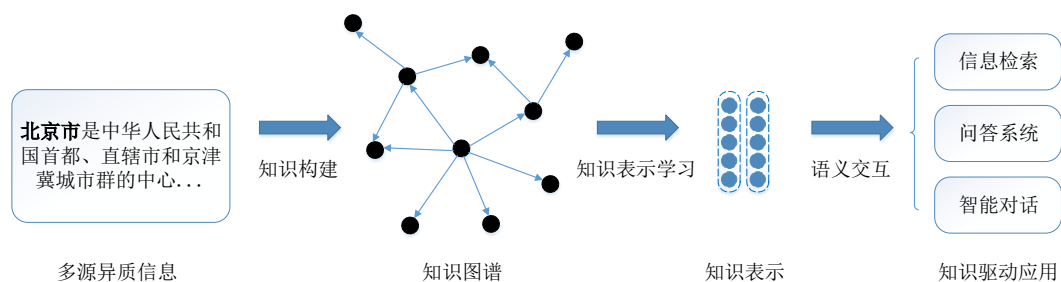


图 1.2 知识驱动工程的简要流程图

1.2 研究内容

世界是复杂且多元的，知识驱动的任务需要与多源信息，如文本、图像、结构化类型等进行交互。多源信息能够帮助构建更加准确的知识表示，而知识表示学习也能使得知识空间与其它多源信息空间的交互变得更为自然、便捷。本文关注融合多源信息的知识表示学习，其中重点解决以下三个任务：（1）融合实体描述信息的知识表示学习；（2）融合实体层次类型信息的知识表示学习；（3）融合实体图像信息的知识表示学习。

第一个任务是融合实体描述信息的知识表示学习。知识库中的（头实体，关系，尾实体）三元组储存的主要是实体之间的结构型知识。目前大多数知识表示学习的模型仅仅关注这些基于三元组的结构知识，而忽略了丰富的基于实体描述的文本知识。实体描述是一段对实体较详尽的文字性描述，通常包含几句话或者一个段落。这些实体描述信息既可以来自于知识库构建时的人工填写，也可以从互联网文本或者百科中自动抽取获得，来源十分丰富。目前，绝大多数知识库都是以基于文字的形式储存获得的知识，并且知识库的构建也往往源于基于文本的知识抽取。可以说，实体描述蕴含着丰富且重要的知识信息，同时也是天然能与知识库进行交互的多源信息之一。在此任务中，我们探究了融合实体描述信息的知识表示学习，提出了几种基于神经网络的模型构建知识表示。模型在知识图谱补充、实体类型分类等任务上都取得了很好的效果，并且能够基于实体描述为知识库中不存在的新实体构建知识表示。

第二个任务是融合实体层次类型信息的知识表示学习。现实世界中的实体往往存在层次结构，而实体层次类型是人工定义的、符合人类认知的实体分类体系，例如莎士比亚这个实体处于人——作家——英国作家的层次体系下。这种层次结构信息符合人类对事物的认识习惯，能够辅助人对实体及实体之间的联系产生更加深入的理解。一些广泛使用的大规模知识图谱，以及作为知识源的各种百科类数据库中，常常维护着自己的实体层次类型信息。实体层次类型信息能够有助于建立更好的知识表示，但是现有的绝大多数知识表示学习模型都未能充分使用层

次结构的丰富信息。在此任务中，我们融合实体层次类型的信息，提出类型特化的实体映射和实体类型限制两种方法，构建更好的知识表示。模型效果在知识图谱补全、三元组分类等任务上得到了验证，同时在具有长尾分布特性的样例上也得到了不错的效果。

第三个任务是融合实体图像信息的知识表示学习。实体图像指的是描述特定实体的图片，能够提供描述实体的直观信息，如实体的外形、颜色等，也能提供实体的行为、相关实体等侧面的信息。图像信息具有较大的信息量，也是人类认知的重要方式之一。研究表明，人类每日接收的信息中，绝大多数来自于视觉。实体的图像信息能够帮助我们从多角度更加全面深入地认识实体，学习更好的知识表示。另一方面，这种跨模态的知识表示也能作为桥梁，将知识空间的信息引入图像相关领域的任务。在此任务中，我们探寻知识空间和图像空间的联合，建立跨模态的知识表示。模型在知识图谱补全、三元组分类等任务上取得了显著的提升效果，证明了跨模型知识表示的可行性。

1.3 相关工作

知识表示学习旨在设计模型与算法，将实体或关系投影至低维连续向量空间中，实现实体或关系的分布式语义表示。知识表示学习具有计算效率高、能缓解数据稀疏性、便于实现多源异质信息融合等优点，在知识图谱补全、自动问答、关系抽取等任务上均被广泛使用。近年来，研究者们提出了很多知识表示学习的方法，如结构向量模型、语义匹配能量模型、隐变量模型、矩阵分解模型、神经张量模型、平移模型等。本章将简单介绍以上知识表示学习模型。

1.3.1 知识表示学习符号体系

我们首先简要地介绍在知识表示学习中常用的符号表示。我们使用 (h, r, t) 表示一个三元组，其中 h 代表头实体， t 代表尾实体，而 r 代表关系。我们使用 E 表示实体集合， R 表示关系集合， T 表示三元组集合。根据定义，我们有 $h, t \in E$ ， $r \in R$ 以及 $(h, r, t) \in T$ 。粗体的符号表示其对应的分布式表示。

1.3.2 知识表示学习经典模型

结构向量模型（Structured embedding, SE）^[7]是知识表示学习较早的尝试。在结构向量模型中，所有实体被投影至同一个低维的向量空间中。针对每一个关系 r ，结构向量模型为头实体与尾实体分别设计了一个关系特定的映射矩阵 $\mathbf{M}_{r,1}$ 和 $\mathbf{M}_{r,2}$ ，这些映射矩阵将会在训练过程中进行自动更新。模型对每个三元

组 (h, r, t) 的损失函数定义如下:

$$E(h, r, t) = \|\mathbf{M}_{r,1}\mathbf{h} - \mathbf{M}_{r,2}\mathbf{t}\|. \quad (1-1)$$

可以看出, 结构向量模型将头实体与尾实体通过关系特化的映射矩阵投影至同一向量空间中, 并希望映射后向量距离越小时三元组存在的可能性越高。

语义匹配能量模型 (Semantic matching energy, SME) [8]与结构向量模型不同, 使用低维向量表示实体及关系。在此之上, 模型使用矩阵映射、点乘等操作, 对实体与关系的联系进行评估。具体地, 语义匹配能量模型设计了线性形式与双线性形式两种对三元组的评价函数:

$$E(h, r, t) = (\mathbf{M}_1\mathbf{h} + \mathbf{M}_2\mathbf{r} + \mathbf{b}_1)^\top (\mathbf{M}_3\mathbf{t} + \mathbf{M}_4\mathbf{r} + \mathbf{b}_2). \quad (1-2)$$

$$E(h, r, t) = ((\mathbf{M}_1\mathbf{h} \otimes \mathbf{M}_2\mathbf{r}) + \mathbf{b}_1)^\top ((\mathbf{M}_3\mathbf{t} \otimes \mathbf{M}_4\mathbf{r}) + \mathbf{b}_2). \quad (1-3)$$

其中 \otimes 表示按位相乘, $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$ 表示待学习的映射矩阵, $\mathbf{b}_1, \mathbf{b}_2$ 表示偏置向量。此外, 也有基于张量对语义匹配能量模型进行改进的模型[9]。

隐变量模型 (Latent factor model, LFM) [10,11]将实体表示成为低维向量, 将关系表示为双线性变换矩阵 \mathbf{M}_r , 在知识表示学习效果与计算复杂度上都有显著改善。模型的评分函数如下:

$$E(h, r, t) = \mathbf{h}^\top \mathbf{M}_r \mathbf{t}. \quad (1-4)$$

此外, 也有方法将关系矩阵限定为对角阵对隐变量模型进行改进的模型, 在降低模型复杂度的情况下获得了更好的模型效果[12]。

矩阵分解模型基于矩阵分解的方式进行知识表示学习, 其中以RESCAL模型[13,14]和HOLE模型[15]为代表。RESCAL模型使用一个三阶张量 $X \in \mathbb{R}^{d \times d \times k}$ 表示三元组, 其中 d 是实体的数量, k 是关系的数量。如果 $X_{htr} = 1$, 则表示三元组 (h, r, t) 存在。设 $X = \{X_1, \dots, X_k\}$, 则有矩阵分解:

$$\mathbf{X}_i \approx \mathbf{A}\mathbf{R}_i\mathbf{A}^\top. \quad (1-5)$$

其中 \mathbf{A} 表示实体向量形成的矩阵，而 \mathbf{R}_i 表示第 i 个关系矩阵。**HOLE**模型则使用了循环相关运算改进**RESCAL**模型，进一步提升了计算效率与知识表示效果。

神经张量模型（Neural tensor network, NTN）^[16]基于单层神经网络模型做出改进，使用双线性张量代替原模型中的线性变换层。神经张量模型对一个三元组 (h, r, t) 的评分函数如下：

$$E(h, r, t) = \mathbf{u}_r^\top \tanh(\mathbf{h}^\top \mathbf{M}_r \mathbf{t} + \mathbf{M}_{r,1} \mathbf{h} + \mathbf{M}_{r,2} \mathbf{t} + \mathbf{b}_r). \quad (1-6)$$

其中 \mathbf{M}_r 是三阶张量， $\mathbf{M}_{r,1}$ 和 $\mathbf{M}_{r,2}$ 是投影矩阵。神经张量模型更精确地刻画了实体与关系之间的联系，获得了更好的知识表示效果，但是复杂的模型也使得计算复杂度提高，同时对知识图谱的稠密程度有更高的要求。

1.3.3 平移模型及其拓展模型

平移模型（TransE）^[17]是Bordes等研究者在2013年提出的知识表示学习算法。平移模型将实体和关系映射至同一个低维向量空间，将实体与实体之间的关系表示为实体向量之间的平移操作。由于只考虑了向量之间的平移操作，平移模型的计算复杂度大大降低，并且学到的知识表示在知识图谱补全等任务上的效果也得到了显著的提升，在知识表示任务中被广为运用。

平移操作的思想在其他任务中早有运用。Mikolov等研究者于2013年提出了词表示学习模型Word2vec^[18-20]，并发现Word2vec学习到的词向量之间有着有趣的语义平移现象，例如：

$$\mathbf{v}(\text{king}) - \mathbf{v}(\text{man}) \simeq \mathbf{v}(\text{queen}) - \mathbf{v}(\text{woman}). \quad (1-7)$$

其中， $\mathbf{v}(x)$ 表示 x 的词向量。这种语义平移现象表明了词和词之间的隐含语义关系被成功地编码进了词向量中。

受到词空间语义平移现象的启发，平移模型将这种隐含语义关系显式地用关系进行表示。具体地，对于给定的三元组 (h, r, t) ，平移模型将关系向量 \mathbf{r} 看做是从头实体 \mathbf{h} 到尾实体 \mathbf{t} 的平移向量，如图1.3所示。基于以上平移假设，平移模型希望一个三元组内的实体与关系向量之间存在 $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$ 的关系。形式化地，模型对三元组 (h, r, t) 定义了如下的能量函数：

$$E(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|. \quad (1-8)$$

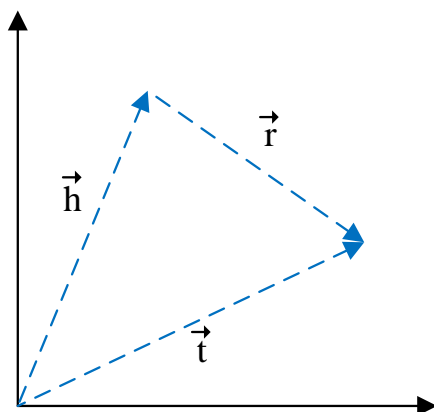


图 1.3 平移模型图示

在实际训练中，平移模型使用最大间隔方法，定义了如下评分函数进行优化：

$$L = \sum_{(h,r,t) \in T} \sum_{(h',r,t') \in T'} \max(\gamma + E(h,r,t) - E(h',r,t'), 0), \quad (1-9)$$

其中， T 和 T' 分别是正例三元组与负例三元组的集合， γ 是正负例三元组得分的间隔距离。平移模型通过最大化正负例三元组之间的得分差来优化知识表示。

尽管平移模型兼顾了效率与结果，被广泛运用于知识表示学习任务中，但平移模型也仍然存在着以下改进与优化的空间：

第一，平移模型简捷高效，能够在大规模知识图谱上获得很好的效果。但是过于理想化的平移假设，在对知识图谱中的复杂关系进行建模时往往存在问题。根据Bordes的定义，知识图谱中存在一对多、多对一甚至多对多的复杂关系^[17]。以是...职业关系为例，我们有（莎士比亚，是...职业，作家）和（鲁迅，是...职业，作家）两条知识。但是莎士比亚和鲁迅仅应在谈论作家这一关系下有相似表示，而在其它属性，如作品风格、年代、国籍等关系上有较大区别。为了解决对复杂关系建模的问题，TransH模型^[21]、TransR模型^[22]、TransD模型^[23]等基于平移模型的改进方法相继被提出，使用向量空间投影与矩阵映射等方式，改进平移模型对复杂关系的建模效果。

第二，平移模型仅仅关注知识图谱三元组的局部信息，而忽略了知识图谱网络的全局结构与关系之间的推理逻辑。知识图谱中的多步关系路径蕴含着丰富的信息，能够帮助我们进行知识推理。例如，如果我们知道（故宫，位于，北京）与（北京，是...首都，中国）两条知识，我们很容易能够推理出（故宫，位于，中国）这条知识。针对这个问题，PTransE模型^[24]在知识表示学习中引入了多步关系路径的信息，KALE模型^[25]也引入了逻辑规则，提高了平移模型的效果。

第三，平移模型只关注了知识图谱自身的网络结构信息，忽略了众多多源异

质的丰富信息，如文本、实体类型与图像等信息。这些多源信息既能辅助知识表示学习模型更好地学习已有的结构化知识，也能提供在知识图谱结构信息之外的多模态信息。目前研究者对如何将多源信息融入知识表示学习的探索还仅仅处于初期阶段，而这也是本文的主要研究课题。

1.4 章节结构

本论文的整体章节结构安排如下：

第一章：引言

引言部分主要从整体阐述了论文的研究背景、相关概念定义与选题意义，介绍了融合多源信息的知识表示学习的研究内容，同时简要介绍了知识表示学习的相关工作与目前存在的问题。

第二章：融合实体描述信息的知识表示学习

在此章节，我们将详细介绍融合实体描述信息的知识表示学习研究工作，包括研究背景，实体描述的定义与作用，主要问题与挑战，利用文本信息辅助知识表示学习的相关工作，本文提出的实验模型与算法原理，以及实验任务和实验结果的展示与分析等。我们设计了不同的神经网络模型，从实体描述中构建实体基于描述的知识表示，在知识图谱补全、实体类型分类等任务上进行评测，同时探索模型在构建新实体知识表示任务上的效果。

第三章：融合实体层次类型信息的知识表示学习

在此章节，我们将详细介绍融合实体层次类型信息的知识表示学习研究工作，包括研究背景，实体层次类型的定义与作用，主要问题与挑战，利用实体类型信息辅助知识表示学习的相关工作，本文提出的实验模型与算法原理，以及实验任务和实验结果与分析等。我们探索了实体层次类型辅助映射矩阵构建以及作为类型限制信息时的可能使用方式，并在知识图谱补齐、三元组分类等任务上进行评测与实验结果分析，也观察了模型在长尾分布数据上的效果。

第四章：融合实体图像信息的知识表示学习

在此章节，我们将详细介绍融合实体图像信息的知识表示学习研究工作，包括研究背景与意义，实体图像的相关定义，主要问题与挑战，结合图像与文字等跨模态信息的相关工作，本文提出的实验模型与算法原理，以及实验任务和实验结果展示与分析等。我们设计神经网络模型，构建知识基于图像信息的跨模态表示，在知识图谱补全、三元组分类等任务上进行评测，同时在样例分析中直观感受了基于图像的知识表示效果。

第五章：总结与展望

在此章节，我们对全文融合文本、层次类型与图像信息的知识表示学习工作进行总结，分析知识表示学习仍存在的挑战，并展望未来工作。

第2章 融合实体描述信息的知识表示学习

2.1 引言

融合实体描述信息的知识表示学习任务，主要关注如何将实体的描述信息与知识图谱自身的结构信息更好地结合起来，从而提升知识表示的性能。文字是知识传承与发展的重要载体之一，人类的交流与社会活动常常离不开文本信息的帮助。目前，绝大多数大规模知识库都是基于文字的形式进行储存，而这些知识库的构建也往往源自从广袤文本数据中的筛选与获取。可以说，以实体描述为代表的文本信息蕴含着丰富且重要的知识，是天然且直观地能与知识空间进行交互的多源信息之一。

实体描述是对实体相关信息较详尽的文字性描述，通常由几句话或者一个段落组成，可以看做是对实体自身信息的浓缩精炼的文字性介绍。实体描述的来源十分丰富，既可以来源于知识库构建时的人工填写，也可以融合百科类数据库的既有信息，亦能从丰富互联网文本中进行抽取。图2.1给出了在维基百科中关于威廉·莎士比亚的一段实体描述节选。从图中我们可以看到，威廉·莎士比亚的实体描述中蕴含关于实体各个方面的丰富信息，如国籍、职业、作品、出身地等，可以很好地作为知识图谱中已有结构化信息的辅助与补充，提供更加深入的细节描述，甚至能挖掘知识图谱中可能遗漏的新知识。

威廉·莎士比亚

威廉·莎士比亚...是英国文学史上最杰出的戏剧家，也是西方文艺史上最杰出的作家之一，全世界最卓越的文学家之一。他流传下来的作品包括38部戏剧、154首十四行诗、两首长叙事诗和其他诗歌。...莎士比亚在雅芳河畔斯特拉特福出生长大，18岁时与安妮·哈瑟维结婚，两人共生育了三个孩子...莎士比亚崇尚高尚情操，常常描写牺牲与复仇，包括《奥赛罗》、《哈姆雷特》、《李尔王》和《麦克白》，被认为属于英语最佳范例。

图 2.1 实体描述节选示例

实体描述中蕴含着丰富的实体信息，这些信息能够作为知识图谱中置信度较高的结构化信息的辅助，帮助模型构建更准确的知识表示。例如，给定三元组（威廉·莎士比亚，写作，哈姆雷特）和相关实体描述，如图2.2所示，在威廉·莎士比亚与哈姆雷特的实体描述中，都有文字暗示了三元组表达的信息。如果能够将此部分信息智能地抽取并融入知识表示学习中，无疑能使知识表示的性能得到提升。但目前大多数知识表示学习模型都仅仅关注知识图谱内部基于三元组的结构

化信息，而忽略了实体描述对于知识表示学习的潜在作用。我们本章的工作重点即集中在如何充分利用实体描述的文本信息，提升知识表示学习的性能。



图 2.2 实体描述辅助三元组结构化知识的示例

融合实体描述信息的知识表示学习主要有着以下难点与挑战：首先，如何从实体描述中自动抽取高质量的实体相关知识。根据实体自身信息量的差异，实体描述往往长短不一，包含的有效文本信息与噪声形式也不尽相同。而根据实体描述构建方式的不同，实体描述的质量往往也参差不齐。另外，由于实体描述基于文本，自然语言的复杂性与灵活性也极大地增加了模型对实体描述文本理解的难度。融合实体描述信息需要考虑以上差异化的文本输入，设计算法自动地从实体描述中获取高质量的知识信息，同时减少文本中无用信息的干扰。其次，当模型抽取到高质量的实体相关知识后，如何将这些基于实体描述的文本信息与基于知识图谱三元组的结构化信息融合。基于实体描述信息学得的表示属于文本空间，而基于三元组结构信息学得的表示属于知识空间。我们需要提出假设并设计算法，实现文本空间与知识空间的对齐与联合训练。

为了解决以上难点与挑战，我们设计了一种融合实体描述的知识表示学习模型（Description-embodied knowledge representation learning, DKRL）^[26]。DKRL模型沿用平移模型的假设与学习框架，同时考虑知识图谱的结构化信息与实体描述的文本信息。具体地，DKRL模型为每个实体设置了两种知识表示：基于结构的表示和基于描述的表示。其中，基于描述的表示使用两种神经网络模型，直接从实体描述中进行构建。我们在知识图谱补全与实体类型分类等任务上对两种实体知识表示进行了评测，并进一步探索了DKRL模型在新实体上的表示效果。概括起来，本章工作的主要创新点有以下几点：

- 与经典的知识表示学习模型相比，我们创新性地知识表示学习任务中考虑了实体描述的文本信息，直接基于实体描述构建知识表示。
- 我们设计了一个全新的融合实体描述的知识表示学习模型，为实体设置基于结构和基于描述的两种知识表示，并使用神经网络模型构建实体基于描述的知识表示。

- 我们的模型学习到的两种知识表示在知识图谱补全和实体类型分类等任务上表现出较好的效果，并证明了模型对新实体的表示能力。

2.2 相关工作

实体描述信息是实体本身精炼的文字性介绍，本质上是一类文本信息。由于文本和知识的天然联系，传统的知识工程相关任务，如知识构建、知识推理等任务中，就有一些与文本结合的相关工作^[27-29]。而在知识表示学习任务中，也有一些模型在学习引入了文本信息。

神经张量模型^[16]是较早使用文本信息的知识表示学习模型之一。它使用组成实体名字的每个词的词向量平均作为此实体的向量。例如对于孟加拉虎 (*Bengal tiger*) 这个实体，神经张量模型使用词向量孟加拉 (*Bengal*) 和虎 (*tiger*) 的平均向量对实体进行表示。这种融合文本和知识的表示学习方法直观且简单，能够挖掘实体名字中有相同单词的相似实体之间可能存在的潜在关系，提高了知识表示的效果。

Wang等研究者提出了联合知识图谱和文本的知识表示学习模型^[30]。该模型基于平移模型^[17]与Skip-gram模型^[18]的思想，对知识空间与文本空间分别进行建模，学习实体向量、关系向量与词向量。同时，该模型设计了对齐模块，使用维基百科锚文本信息或者实体名字信息，实现了知识空间与文本空间的对齐。

Zhong等研究者基于以上联合知识图谱和文本的知识表示学习模型做出了改进^[31]。在对齐模块中，该模型引入了实体描述文本辅助对齐，认为实体向量应与实体描述中的所有词向量尽可能接近。Zhang等研究者也尝试使用实体名字与实体描述的词向量均值表示实体向量^[32]。

与我们提出的DKRL模型相比，以上和文本结合的知识表示学习模型仅孤立地考虑词级别的文本信息，忽略了篇章级别重要的语序语义信息。为了解决这些问题，我们使用神经网络模型整体考虑实体描述的语义信息，直接基于实体描述构建实体的知识表示，同时能够对新实体构建合理的知识表示。

2.3 算法模型

融合实体描述信息的知识表示学习旨在将实体的描述信息与知识图谱的结构信息有机结合起来，从而学习到更好的知识表示。我们提出了一种融合实体描述的知识表示学习模型，模型流程图如图2.3所示。

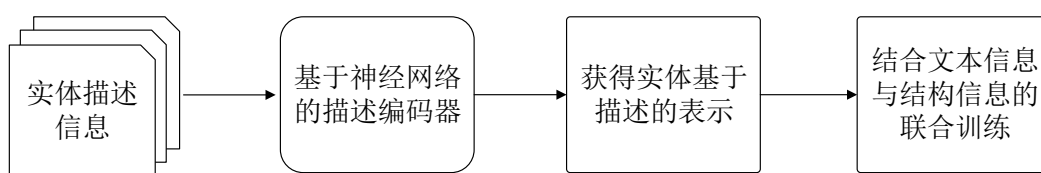


图 2.3 融合实体描述的知识表示学习模型流程图

2.3.1 符号体系和重要概念

我们首先介绍本章所用的符号体系。在知识图谱方面，我们使用 (h, r, t) 表示一个三元组，其中 h 代表头实体， t 代表尾实体，而 r 代表关系。我们使用 E 表示实体集合， R 表示关系集合， T 表示三元组集合，并有 $h, t \in E$ ， $r \in R$ 以及 $(h, r, t) \in T$ 。在实体描述方面，我们使用 $d = (x_0, x_1, \dots, x_n)$ 表示实体描述，其中 x_i 表示实体描述中第 i 个词，而 n 表示实体描述的长度。

我们为每个实体构建两种表示向量。第一种是**基于结构的表示**，使用 \mathbf{h}_S 和 \mathbf{t}_S 分别表示头实体和尾实体基于结构的表示。这种表示向量完全由知识图谱的结构信息决定，与之前平移模型学习到的分布式向量相似。第二种是**基于描述的表示**，使用 \mathbf{h}_D 和 \mathbf{t}_D 分别表示头实体和尾实体基于描述的表示。这种表示向量受到两种信息的联合影响，直接从实体描述中构建所得。

2.3.2 模型整体架构和学习框架

图2.4给出了DKRL模型整体架构的简单图示。如图所示，DKRL模型以实体描述为输入，通过实体描述编码器得到实体基于描述的表示。在顶层，基于描述的表示与基于结构的表示被统一在平移模型的框架下进行学习。学习到的信息通过反向传播更新实体描述编码器的参数和所有向量。

DKRL模型沿用了平移模型的假设，认为三元组内的实体与关系向量之间应该具有 $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$ 的关系。形式化地，我们为DKRL模型的定义了如下能量函数：

$$E(h, r, t) = \alpha_1 \|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_S\| + \alpha_2 \|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_D\| + \alpha_3 \|\mathbf{h}_D + \mathbf{r} - \mathbf{t}_S\| + \alpha_4 \|\mathbf{h}_D + \mathbf{r} - \mathbf{t}_D\|. \quad (2-1)$$

其中 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是控制各项权值的超参数。在能量函数中， $\|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_S\|$ 部分与平移模型的能量函数类似，而 $\|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_D\|$ ， $\|\mathbf{h}_D + \mathbf{r} - \mathbf{t}_S\|$ 和 $\|\mathbf{h}_D + \mathbf{r} - \mathbf{t}_D\|$ 三项则加入了基于描述的实体向量影响。这样，通过混合项的软限制，DKRL模型将实体的两种向量映射到了同一语义空间中，并共享相同的关系向量。

DKRL使用以上改进的能量函数，在训练中也参考了平移模型的学习框架。

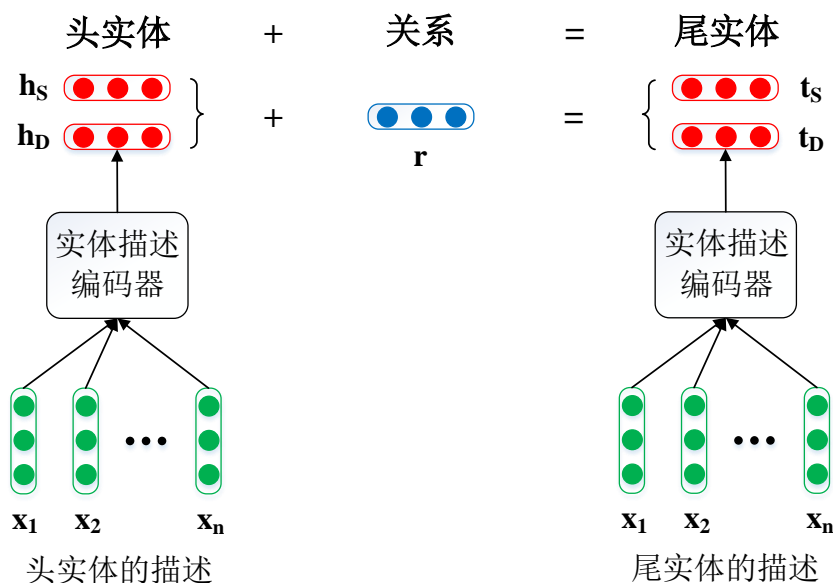


图 2.4 融合实体描述的知识表示学习模型示意图

我们使用了最大间隔方法，定义了如下评分函数对模型进行优化：

$$L = \sum_{(h,r,t) \in T} \sum_{(h',r',t') \in T'} \max(\gamma + E(h,r,t) - E(h',r',t'), 0). \quad (2-2)$$

其中， (h',r',t') 是负例三元组， γ 是表示正负例三元组得分的间隔距离的超参数。与平移模型不同，此处三元组得分的能量函数 $E(h,r,t)$ 包括了两种实体向量表示的四种组合项。最大间隔方法希望正例三元组的能量函数得分比负例三元组的能量函数得分至少小 γ 。 T' 是负例三元组的集合，由于知识图谱中没有显式的负例三元组，对于给定的正例三元组 (h,r,t) ，我们设置负例三元组的集合如下：

$$T' = \{(h',r,t) | h' \in E\} \cup \{(h,r,t') | t' \in E\} \cup \{(h,r',t) | r' \in R\}, \quad (h,r,t) \in T. \quad (2-3)$$

这表示所有负例均由正例三元组中任一实体（或关系）随机替换成另外的实体（或者关系）产生。需要注意的是，我们还在负例生成时加上了以下限制条件：

$$\forall (h',r',t') \in T', \quad (h',r',t') \notin T. \quad (2-4)$$

以避免随机替换后生成的负例三元组中存在真正的正例。

为了从实体描述中尽可能完整地抽取到重要的文本信息，同时减少文本信息中可能存在的噪音，我们设计了两种实体描述编码器，分别是连续词袋模型和卷

积神经网络模型。我们会在下面两节详细介绍这两种模型的框架及设计思路。

2.3.3 连续词袋模型

连续词袋模型（Continuous bag-of-words encoder, CBOW）是简单且有效的对文本描述建模的方法。CBOW模型基于以下假设：相似的实体之间应该具有相似的文本描述，也应该具有相似的关键词。实体描述中的关键词能够很好地概括描述中的文本信息，且一定程度上能够过滤掉文本信息中的噪声。实体之间的联系与潜在关系能够从实体描述中的关键词中得到体现。

具体地，连续词袋模型以实体描述为输入，首先抽取实体描述中排序前 k 的关键词。我们可以使用一些经典的文本特征，如词频、TF-IDF值等，对描述中的词进行排序。然后，我们忽略词序信息，直接将这些关键词的词向量进行加和，从而得到实体基于描述表示。形式化地，我们用 $X' = \{x_1, x_2, \dots, x_k\}$ 表示实体 e 的前 k 个关键词集合，则有：

$$\mathbf{e}_D = \mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_k. \quad (2-5)$$

其中 \mathbf{e}_D 表示实体在连续词袋模型模型下基于描述表示。图2.5给出了连续词袋模型的整体示意图。

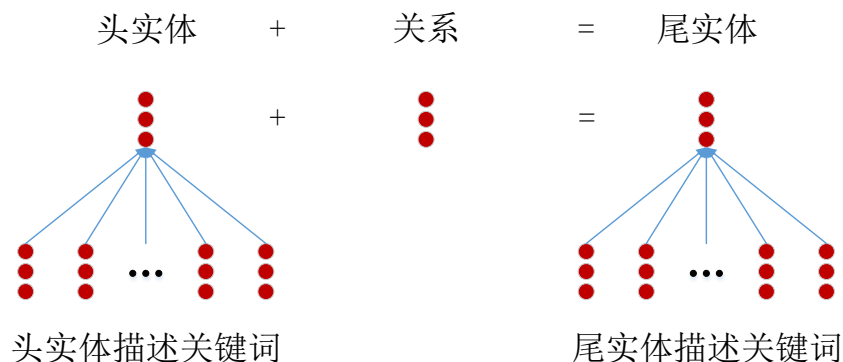


图 2.5 连续词袋模型示意图

2.3.4 卷积神经网络模型

连续词袋模型简单有效，在自然语言处理相关任务上使用广泛。但是连续词袋模型使用关键词构建的实体表示仅关注了词级别的语义信息，忽略了重要的语序信息，且知识表示的效果严重受限于抽取关键词的质量。因此，我们使用卷积神经网络模型构建实体基于描述表示。

卷积神经网络（Convolutional neural network, CNN）模型是一种经典高效的神经网络模型，最初在图像分类^[33]、人脸识别^[34]等图像领域的任务上被广泛使用。近年来，CNN模型在自然语言处理的各种任务上大放异彩，例如句子分类^[35]、关系抽取^[29,36]、词性标注和命名实体识别^[37]等。CNN模型能够通过卷积操作和池化操作，充分考虑词本身与词序的语义信息，在自动抽取变长输入中的有效信息的同时，智能地过滤掉输入中的噪声。

图2.6给出了卷积神经网络的模型示意图。从图中可以看出，为了兼顾实验结果与效率，我们使用了两层卷积层与对应的池化层及非线性层对实体描述进行建模。接下来，我们将从预处理和词表示、卷积层、池化层三个方面详述卷积神经网络模型。

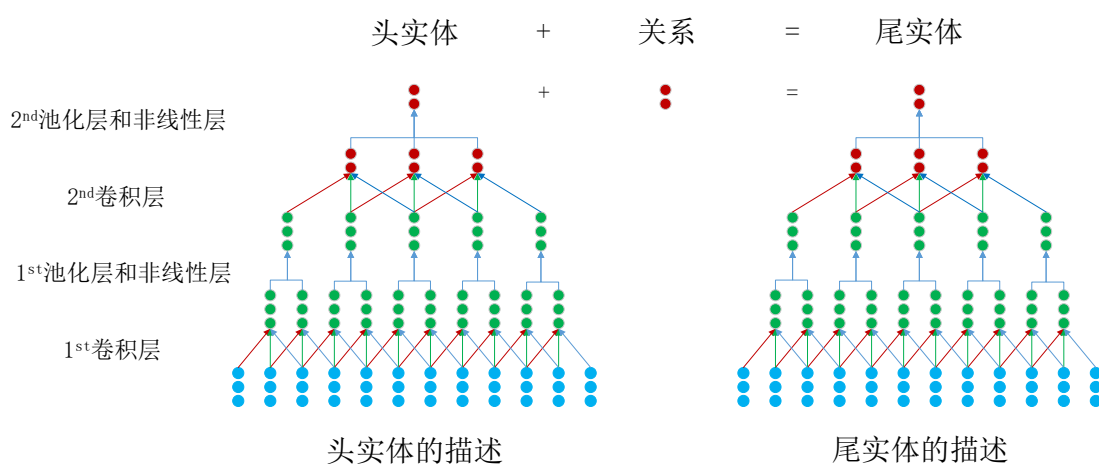


图 2.6 卷积神经网络模型示意图

预处理和词表示关注如何将实体描述处理成适合卷积神经网络的输入。在预处理步骤，我们首先根据停用词词表去除了实体描述中的停用词，并统一了英文的大小写字符。由于部分实体名字由多个词组成，分开成多个词难以抓住整个实体的信息，我们还将实体描述中的所有实体名称看做一个“词”。例如孟加拉虎（*Bengal tiger*）这个实体名字，在描述中被看做一个整体，而不是孤立的两个词孟加拉（*Bengal*）和虎（*tiger*）。我们使用了Word2vec模型^[18]在大规模语料（如维基百科文本）上训练得到词向量，作为卷积神经网络模型的输入。形式化地，卷积神经网络模型的输入使用 $X = \{x_1, \dots, x_n\}$ 表示，其中 n 表示预处理之后实体描述的长度。

卷积层（convolution）是卷积神经网络的核心操作。在一个卷积神经网络中通常会使用多层卷积层叠加，以此增加模型的表达能力。由于我们在融合实体描述的知识表示学习任务中，处理对象是文本序列，所以我们使用的是一组长度为 k 的

一维卷积核。形式化地，我们使用 $\mathbf{X}^{(l)}$ 表示第 l 层输入序列的矩阵， $\mathbf{Z}^{(l)}$ 表示第 l 层的输出矩阵。其中，实体经过预处理后的描述序列 $\mathbf{X} = \{x_1, \dots, x_n\}$ 即为第一层卷积层的输入。

卷积层主要有两步操作。首先，卷积层将会进行窗口操作。对于第 l 层的输入序列 $\mathbf{X}^{(l)} = \{x_1^{(l)}, \dots, x_n^{(l)}\}$ ，我们有以下窗口操作：

$$\mathbf{x}'_i{}^{(l)} = \mathbf{x}_{i:i+k-1}{}^{(l)} = [\mathbf{x}_i^{(l)\top}, \mathbf{x}_{i+1}^{(l)\top}, \dots, \mathbf{x}_{i+k-1}^{(l)\top}]^\top. \quad (2-6)$$

其中 $\mathbf{x}'_i{}^{(l)}$ 是窗口操作后得到矩阵的第 i 个向量， k 是窗口大小，与输入维数联合决定了此层卷积核的大小。可以看出，窗口操作即将输入序列的第 i 个向量到第 $i+k-1$ 个向量共 k 个向量依序头尾相接，连接成一个长向量，作为下一步输入矩阵的第 i 个向量。由于输入序列可能是变长的，所以我们采用补零的方法（zero padding）在序列后补上全零向量。

在窗口操作后，卷积层将会联合卷积核得到输出结果。设 $\mathbf{Z}^{(l)} = \{z_1^{(l)}, \dots, z_n^{(l)}\}$ ，我们有：

$$\mathbf{z}_i^{(l)} = \sigma(\mathbf{W}^{(l)}\mathbf{x}'_i{}^{(l)} + \mathbf{b}_i^{(l)}). \quad (2-7)$$

其中 $\mathbf{z}_i^{(l)}$ 表示第 l 层的第 i 个输出向量， $\mathbf{W}^{(l)}$ 表示第 l 层的卷积核， $\mathbf{b}^{(l)}$ 表示第 l 层的偏置矩阵。 σ 表示非线性函数，通常可以使用tanh函数或者ReLU函数。需要注意的是，在前向和反向传播的时候，输入序列的补零向量对结果都不产生影响。

池化层（pooling）通常接在卷积操作之后，用于减小特征空间，同时过滤噪声。针对我们对文本描述建模的特殊任务，我们对卷积神经网络模型中不同层的池化使用了不同的池化策略。在第一个池化层，我们使用的是 m -最大池化策略（ m -max pooling）。具体地，设第一层通过卷积层和非线性层后的输出矩阵 $\mathbf{Z}^{(1)} \in \mathbb{R}^{n_1 \times n_2}$ ，其中 n_1 为输入描述序列的长度。我们将此输出矩阵按行均等分为窗口大小为 m 的向量组，不能整除时在序列尾部补上全零向量，每个向量组大小为 $m \times n_2$ 。在每个向量组中，针对每一维，我们在 m 个元素中选择最大的一个组成大小为 $1 \times n_2$ 输出向量。形式化地，我们有：

$$\mathbf{x}_i^{(2)} = \max(\mathbf{z}_{n \cdot i}^{(1)}, \dots, \mathbf{z}_{n \cdot (i+1) - 1}^{(1)}). \quad (2-8)$$

其中max操作表示按位取最大值。通过 m -最大池化策略，我们选择了局部区域中

最强烈的特征信号，同时将描述的表达缩小成原来的 $\frac{1}{m}$ ，达到了减少特征空间并过滤噪声的效果。

在第二个池化层，我们使用平均池化策略（mean pooling）代替 m -最大池化策略。由于我们需要获得实体基于描述的表达，所以此池化层应该输出一个向量。具体地，我们有：

$$\mathbf{x}^{(3)} = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{z}_i^{(2)}. \quad (2-9)$$

其中 $\mathbf{x}^{(3)}$ 即表示卷积神经网络模型得到的实体基于描述的表达，而 n_2 表示第二个池化层的输入矩阵长度。以上公式说明，平均池化层将输入矩阵 $\mathbf{Z}^{(2)}$ 按行取均值构成了输出向量，综合考虑了所有输入向量的信息。

如上所述，我们针对文本描述编码的特定任务，在两个池化层设置了不同的池化策略。这是因为文本描述通常包括多个句子，我们认为卷积层对于每个句子内部的局部信息抽取特征时，可能会有一定的冗余和噪声，所以在第一个池化层进行 m -最大池化操作，仅使用局部特征最强烈的值作为整个局部的代表。而在第二个池化层，我们认为描述中的每个句子都应该对最后的知识表示产生影响，所以使用了平均池化的策略综合考虑所有的局部信息。这种不同的池化策略设置，使得卷积神经网络模型能够在尽可能抽取高质量的文本信息的同时，减少模型复杂度与需要训练的参数量，最终建立更好的实体基于描述的知识表示。

2.3.5 模型优化与具体实现

如前所述，我们的模型按照图2.4的整体框架，基于平移模型的假设进行联合学习。在实际训练中，DKRL模型需要学习以下的参数组： $\theta = (\mathbf{X}, \mathbf{E}, \mathbf{R}, \mathbf{P})$ 。其中 \mathbf{X} 表示词向量， \mathbf{E} 表示实体基于结构的表示向量， \mathbf{R} 表示关系向量， \mathbf{P} 表示实体描述编码器的相关参数，主要包括卷积神经网络的卷积核和偏置项等参数。模型使用Mini-batch随机梯度下降（stochastic gradient descent, SGD）优化算法进行训练，使用链式法则对神经网络参数以及对应向量进行更新。

DKRL模型有一些实现细节：在初始化时，卷积神经网络模型各层的卷积核随机初始化，而使用的词向量 \mathbf{X} 则是在大规模语料库（如维基百科）上使用Word2vec工具预训练得到。知识向量 \mathbf{E} 和 \mathbf{R} 可以随机初始化，或者使用平移模型预训练后的向量结果。根据模型的具体需要，基于结构的表示可以在预训练后固定或者进行微调。为了提高模型计算效率，我们实现了一个基于多线程的卷积神经网络模型。

2.4 实验设计与结果分析

在本章节，为了测试融合实体描述的知识表示学习模型的效果，我们进行了知识图谱补全和实体类型分类等实验，并验证了在零样本环境下DKRL模型对于新实体的知识表示构建效果。实验结果证明了实体描述中蕴含的丰富文本信息能够辅助知识表示的构建，同时也证明了我们的模型能够充分抽取实体描述中的有效信息，甚至能够较好地构建训练集中未出现的新实体的知识表示。

2.4.1 实验数据集

在实验中，我们使用一个在知识表示学习中广泛使用的经典数据集——FB15K^[17]——作为我们的实验所用的知识图谱数据。FB15K抽取自一个经典的大规模知识图谱Freebase^[2]。我们直接使用Freebase中维护的实体描述信息作为我们模型输入的实体描述，这些实体描述均抽取自实体对应的维基百科页面中的第一个段落。在经过大小写转换、去除停用词及实体名字处理等预处理后，这些实体描述的平均长度为69个词，而最长的实体描述有343个词。为了保证实体描述有较丰富的意义，我们从FB15K中去除了在预处理后实体描述短于3个词的共47段实体描述及其对应实体。最终，我们在实验中使用的FB15K数据集中一共含有14,904个实体，1,341个关系。我们直接使用Bordes对FB15K的划分^[17]，最终数据集中训练集有472,860个三元组，验证集有48,991个三元组，测试集有57,803个三元组。FB15K的数据统计详见表2.1。

表 2.1 FB15K的统计数据

数据集	#关系	#实体	#训练集	#验证集	#测试集
FB15K	1,341	14,904	472,860	48,991	57,803

为了测试DKRL模型对新实体的知识表示构建效果，我们在FB15K的基础上进行扩展，构建了FB20K的数据集。我们沿用了FB15K中过滤后的1,341个关系作为FB20K的关系集，使用FB15K中过滤后的14,904个实体作为种子集。然后，我们在Freebase全网中抽取所有和种子集中的任意实体有关系相连的实体作为扩展候选。我们从以上的扩展候选实体中随机选择新实体，并且将所有和新实体相关的三元组（包括新实体之间的三元组）抽取出来，作为FB20K中的测试集。需要注意的是，FB20K中的所有关系仅指在关系集中的关系。综上所述，FB20K中的关系、训练集与测试集与FB15K相同，实体数增加到19,923个，测试集增加到88,293个。为了进一步对测试集进行区分，我们根据三元组的头实体/尾实体是

否出现在训练集中，将测试集中的三元组分为四组： $e - e$ 组表示头实体和尾实体都在训练集中， $d - e$ 组表示仅尾实体出现在训练集中， $e - d$ 组表示仅头实体出现在训练集中，而 $d - d$ 组表示头实体和尾实体都未出现在训练集中。第一组即为FB15K中的原测试集，而另外三组则是FB20K新增加的三元组。FB20K的数据统计详见表2.2。

表 2.2 FB20K的统计数据

数据集	#关系	#实体	#训练集	#验证集	# $e - e$	# $d - e$	# $e - d$	# $d - d$
FB20K	1,341	19,923	472,860	48,991	57,803	18,753	11,586	151

2.4.2 实验与模型参数设置

我们使用了两种实体描述编码器对实体描述进行建模，使用连续词袋的模型记录为DKRL(CBOW)，而使用卷积神经网络的模型记录为DKRL(CNN)。为了充分展示融合实体描述的知识表示学习模型的效果，我们实现了平移模型(TransE)作为基线模型，与基于连续词袋和卷积神经网络的模型在多个任务上进行比较。需要注意的是，为了更公平地与DKRL模型进行对比，我们通过加入关系负例等方法改进了平移模型的训练，获得了比原论文中报告更好的结果。另外，我们还报告了使用基于结构的表示与基于描述的表示进行联合预测的结果，使用DKRL(CNN)+TransE进行表示。为保证公平，所有模型的知识表示向量维数都保证一致。

针对DKRL模型，我们探索了不同参数对实验结果的影响。我们基于随机梯度下降进行模型训练，使用了固定的学习率 λ ，在 $\lambda = \{0.0005, 0.001, 0.002\}$ 的范围进行尝试，训练时的最大间隔参数 γ 在 $\{0.5, 1.0, 1.5, 2.0\}$ 中进行尝试。我们将实体向量与关系向量设置成了相同维数 n_k ，在 $n_k = \{50, 80, 100\}$ 的范围中进行尝试。我们还探索了整体能量函数中两种实体表示的四种组合模块的比例 α_i 。最终，我们在实验中选择了如下最优参数： $\lambda = 0.001$ ， $\gamma = 1.0$ ， $n_k = 100$ ， $\alpha_1 = 1.0$ ， $\alpha_2 = 0.4$ ， $\alpha_3 = 0.4$ ， $\alpha_4 = 0.1$ 。

对于连续词袋模型，我们尝试了不同的选择关键词的策略。最后，我们使用TF-IDF值作为各关键词的评分进行排序，并选择了前20个关键词构建实体表示（如果描述长度少于20，则选择描述中全部的词作为关键词）。对于卷积神经网络模型，我们也对卷积层与池化层的不同设置进行了组合与尝试。最终的模型在第一层池化层使用4-最大池化策略，在第二层池化层使用平均池化；第一层卷积层的输入向量维数（即词向量维数）为 $n_w = 100$ ，第二层卷积层的输入向量维数（即

第一层卷积核的个数)为 $n_f = 100$ ，窗口大小为2。卷积神经网络的所有非线性层均使用tanh函数。

2.4.3 知识图谱补全结果与分析

知识图谱补全是知识表示学习的经典评测任务^[8,17,22]，可以被看做是知识图谱上的链接预测任务。具体地，知识图谱补全任务要求模型在给定三元组 (h, r, t) 的某一实体或关系丢失的情况下，利用另外两个对象正确预测出丢失的实体或关系。根据具体丢失的对象，知识图谱补全可以细分为实体预测与关系预测两个任务。知识图谱补全任务旨在验证所学知识表示的准确性，同时也与现实中间答系统的实际任务相呼应。如需补全三元组 $(?, \text{写作}, \text{哈姆雷特})$ ，即相当于回答“谁写了《哈姆雷特》这本书？”这个问题。

2.4.3.1 实验设计

我们使用DKRL模型和基线模型在训练集上学习到的知识向量，对测试集样例进行知识图谱补全。对于测试集中的每个三元组，我们分别隐藏其头实体、关系和尾实体来进行实体预测和关系预测。由于DKRL模型和基线模型都基于平移假设，所以我们直接使用评分函数 $S(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$ 进行补全。例如给定三元组的 h 和 r ，我们首先使用 $\mathbf{h} + \mathbf{r}$ 得到一个向量，然后在所有实体集合 E 中按照和 $\mathbf{h} + \mathbf{r}$ 向量的距离进行排序，得到预测结果。为了更好地展示两种实体表示的效果，在评测中，DKRL(CBOW)模型和DKRL(CNN)模型均只使用实体基于描述的进行预测，而DKRL(CNN)+TransE模型使用实体基于描述和基于结构的两种知识表示联合进行预测。

在评测中，我们参考了Bordes论文中提出的评测方法，使用了Mean Rank和Hits@N两个指标进行评测^[17]。其中，Mean Rank表示所有测例中正确答案的平均排序，关注知识表示的整体效果；而Hits@N表示排名前N的答案中有正确答案的测例比例，关注在实际任务中模型补全的实用性。特别地，我们在关系预测任务中使用Hits@1，在实体预测任务中使用Hits@10对模型进行评测。

需要注意的是，由于知识图谱中存在大量一对多、多对一甚至多对多的关系，仅仅依靠排序直接进行统计可能会存在一些问题。例如，针对 $(\text{莎士比亚}, \text{写作}, ?)$ 这个三元组补全，尾实体有多个正确答案，如哈姆雷特、罗密欧与朱丽叶、仲夏夜之梦等。那么在评测时，即使最好情况下这三个实体排名前三，这些样例的Mean Rank值也只有2，这将与我们的评测初衷相违背。针对这个问题，我们仿照Bordes的设置，构造出“Raw”和“Filter”两个指标^[17]。其中“Raw”表

示不进行额外的处理，而“Filter”表示在评测实体时，不考虑其它所有正确的实体。如果某正确实体排名第4，而排序前3的实体中有2个也是知识图谱中已有的正确答案，那么在“Raw”的设定下，这个实体的Mean Rank 值为4，而在“Filter”的设定下，这个实体的Mean Rank 值为2。

2.4.3.2 实验结果与分析

我们在实体预测与关系预测两个子任务上都进行了评测，实体预测的结果详见表2.3，关系预测的结果详见表2.4。从结果中，我们可以得到以下一些结论：

表 2.3 FB15K上实体预测的评测结果

评测指标	Mean Rank		Hits@10(%)	
	Raw	Filter	Raw	Filter
TransE	210	119	48.5	66.1
DKRL(CBOW)	236	151	38.3	51.8
DKRL(CNN)	200	113	44.3	57.6
DKRL(CNN)+TransE	181	91	49.6	67.4

(1) 在实体预测与关系预测两个任务上，DKRL(CNN)+TransE模型在所有评测指标上都取得了最好的结果。这个结果说明了实体描述中确实蕴含着丰富的文本信息，而这些文本信息能够帮助构建更好的知识表示。同时，这也说明了卷积神经网络编码器能够智能地选择实体描述中有信息量的特征，自动过滤可能的噪声。具体地，在Mean Rank评测指标上的提升，说明DKRL模型学到的知识表示整体效果较好；而在Hits@N评测指标上的提升，则说明DKRL模型对于知识图谱的补全任务的推荐质量更高。

(2) 我们对比两种实体描述编码器的模型结果，发现DKRL(CNN)模型的结果在实体预测的Mean Rank和Hits@10指标，以及关系预测的Hits@1指标上，都显著地优于DKRL(CBOW)模型。在关系预测的Mean Rank指标上，二者的结果也是相近的。这是因为与连续词袋模型相比，卷积神经网络不仅抓住了词级别的语义信息，同时也通过卷积操作获得了局部的词序信息。另外，最大池化操作也使得编码器能够更关注局部特征的强信号，避免长篇描述中的噪声干扰。综上所述，虽然DKRL(CBOW)模型更加简单快速，但DKRL(CNN)模型能够在知识图谱补全任务上获得更好的效果。

(3) 进一步地，我们还通过样例分析证实了有时仅通过结构信息是难以抓住实体相关的知识细节的。例如，如果知识图谱中没有直接提及，我们是很难仅通

过有限的三元组信息回答一个球员到底是前锋还是守门员这种细节问题。而如果通过实体描述，我们就可以在文字信息中挖掘可能相关的知识细节（例如球员在比赛中的表现），从而给出正确的回答。

(4) 融合实体描述的知识表示在某些评测指标上，相比基线模型的提升似乎不太明显。这是因为对于一些三元组的补全任务，仅依靠结构信息就能达到不错的效果。另外，由于实体描述中存在较长的文本，仅使用卷积神经网络还是很难智能地抓住实体相关的每个细节信息。使用更好的实体描述编码器，或者使用更好的实体基于描述的表示方法，可能会是模型未来的提升点。除此之外，DKRL模型也在新实体的零样本知识表示构建任务上展示出了它的能力，而这是传统基于结构信息的知识表示学习模型难以做到的。

表 2.4 FB15K上关系预测的评测结果

评测指标	Mean Rank		Hits@1(%)	
	Raw	Filter	Raw	Filter
TransE	2.91	2.53	69.5	90.2
DKRL(CBOW)	2.85	2.51	65.3	82.7
DKRL(CNN)	2.91	2.55	69.8	89.0
DKRL(CNN)+TransE	2.41	2.03	69.8	90.8

2.4.4 实体类型分类结果与分析

实体类型分类任务旨在预测一个实体的所有类型，可以被看做一个多标签分类的问题，在以往的工作中也常被研究^[38]。目前，大多数大规模百科或知识库都维护有自己的实体类型信息（entity type）。例如在Freebase中，莎士比亚这个实体即有作家、艺术家和获奖提名名人等实体类型。实体类型分类任务能够从实体类型的粒度上证明知识表示的有效性。

2.4.4.1 实验设计

我们首先从Freebase中抽取FB15K训练集中所有实体的全部实体类型，共得到4,054个实体类型。由于这些类型分布具有长尾性，有相当的实体类型仅仅出现了很少的次数，所以我们按照这些实体类型在对应数据集实体中的出现频次从高到低排序，选择了频次前50的实体类型作为我们待分类的实体类型集（我们去除了common/topic这个类型，因为它几乎出现在所有实体的类型列表中）。前50的实体类型覆盖了13,445个实体，我们随机将这些实体分为训练集与测试集，其中训

训练集含有12,113个实体，测试集含有1,332个实体。表2.5展示了排名前5的实体类型及其对应频次。

表 2.5 FB15K中实体类型及其对应频次

排名	实体类型	频次
1	/people/person	4,507
2	/award/award_nominee	4,042
3	/award/award_winner	3,571
4	/film/actor	3,381
5	/location/location	3,184

在训练时，我们使用基线模型和DKRL模型训练得到的知识表示作为实体的特征向量，使用逻辑斯蒂回归算法（Logistic regression）作为分类器，并使用一对其它（one-versus-rest）的训练策略训练多标签分类的逻辑斯蒂回归分类器。在评测中，DKRL(CBOW)模型和DKRL(CNN)模型同样只使用实体基于描述的代表作为特征向量。为了充分显示DKRL模型的优势，我们还实现了一个基于文本的经典词袋模型（BOW）作为补充基线模型。这个模型基于词袋模型，使用实体描述中词的TF-IDF值构建特征向量，然后使用逻辑斯蒂回归算法进行分类。

在评测指标上，我们参考相关工作的设定，使用平均准确率均值（Mean average precision, MAP）作为评测指标^[38]。MAP是多标签分类任务中常用的评测指标之一，在信息检索等领域被广泛使用，其公式如下：

$$MAP = \frac{\sum_{i=1}^{|Q|} AveP(i)}{|Q|}. \quad (2-10)$$

其中 $|Q|$ 是测试集的样例数， $AveP(i)$ 是第 i 个测试样例的平均准确率，我们有：

$$AveP(i) = \frac{\sum_{k=1}^n (P_i(k) \times correct_i(k))}{m_i}. \quad (2-11)$$

其中 n 是候选类型的个数， m_i 是第 i 个测试样例真实拥有的实体类型数。在计算 $AveP(i)$ 时，我们首先将各候选类型的得分从高到低排好序，然后依序判断是否是正确答案。 $P_i(k)$ 是第 i 个测试样例截止至第 k 个候选类型时的准确率，而 $correct_i(k)$ 函数是一个指示函数，当且仅当第 k 个候选类型是正确答案时 $correct_i(k)$ 为1，其余情况下为0。

2.4.4.2 实验结果与分析

表2.6给出了模型在实体类型分类任务上的评测结果，我们分析模型在FB15K数据集上的结果，得到了以下结论：

(1) DKRL(CBOW)模型和DKRL(CNN)模型在MAP评测指标上均能显著超过两个基线模型。这个结果从另一个侧面证明了实体描述中的丰富文本信息能够帮助模型建立更好的知识表示，从而为实体类型分类任务提供更加准确的指导信息。我们认为DKRL模型在实体类型任务上的优势主要来源于联合知识图谱的结构信息与实体描述的文本信息。这是由于与DKRL模型相比，TransE模型使用的知识表示中仅编码了知识图谱三元组的结构信息，而BOW模型也仅编码了实体描述中的文本信息，所以DKRL模型学习到的知识表示在类型分类任务上更加有效。

(2) DKRL(CNN)模型在实体类型分类任务上的表现要优于DKRL(CBOW)模型。这个结果也进一步证明了卷积神经网络模型能够更准确地对实体描述信息进行编码与表示。在实验过程中我们也发现，对于一些描述较长的测试样例，使用DKRL(CNN)模型的准确率通常要高于DKRL(CBOW)。

表 2.6 FB15K上实体类型分类的评测结果

评测指标	MAP
TransE	87.9
BOW	86.3
DKRL(CBOW)	89.3
DKRL(CNN)	90.1

2.4.5 零样本知识表示学习

为了充分展示DKRL模型学到的基于描述的知识表示的效果，我们还探索了模型在零样本环境下的知识表示学习任务表现。零样本学习（Zero-shot learning）是人工智能、特别是深度学习领域的研究热点之一，关注于对新样例的建模。在大数据时代，文字、图像或知识等领域上的模型通常是数据驱动的，而这往往需要在训练时对某类样例进行成百上千的学习。例如在知识表示学习任务上，以往的模型大都要求测试集中的实体曾出现在训练集中，以学得实体的知识表示。然而，在信息爆炸时代，每天都有新实体与新知识的出现，传统知识表示模型无法自动获得新实体的表示，可拓展性差。我们提出的DKRL模型，能够根据实体的描述信息（实体描述信息广泛存在于互联网文本中），构建知识图谱中没有的新实体基于描述的知识表示，是零样本知识表示学习的新思路。

我们使用FB20K数据集模拟零样本知识表示学习的环境，在知识图谱补全与实体类型分类两个任务上进行零样本环境的评测。FB20K与FB15K有着相同的训练集与验证集，而在测试集中，FB20K有着5,019个训练集中未出现的新实体。对这些新实体，我们使用其基于描述的代表作为知识表示。在知识图谱补全任务中，我们主要关注对存在新实体的三元组的补全情况。在实体类型分类任务中，我们主要关注新实体的分类情况。

2.4.5.1 零样本知识图谱补全结果与分析

在零样本知识图谱补全任务中，我们仅关注测试集中存在新实体的三元组的结果，即FB20K数据集中的 $d-e$ 、 $e-d$ 和 $d-d$ 三类三元组。由于传统的平移模型无法对新实体进行表示，所以我们仅测试了DKRL(CBOW)模型和DKRL(CNN)模型的表示效果。在测试时，对于所有新实体，我们均使用其基于描述的代表作为知识表示，而对于测试集中出现在训练集中的实体，我们使用其两种表示分别进行测试。综上，我们在测试中一共使用了4种模型设定，其中CBOW表示测试中所有实体均使用基于描述的代表，而Partial-CBOW表示在训练集中有的实体使用基于结构的代表。同理，CNN和Partial-CNN也分别表示在卷积神经网络模型下使用实体基于描述和基于结构的代表的模型。我们在实体预测与关系预测任务上分别进行了测试，并且展示的不同三元组类的Hits@10（对实体）和Hits@1（对关系）的实验结果。

表2.7展示了模型在零样本环境下实体预测的结果，表2.8展示了模型在零样本环境下关系预测的结果。从这些结果中，我们可以得到以下结论：

(1) 在实体预测与关系预测两个任务上，基于CNN的模型在整体结果与各个子分类上均取得了最佳结果。对于训练集中未出现的新实体，仅依靠其基于描述的代表，即能在关系预测任务上达到将近60%的准确率。这说明DKRL模型学得的基于描述的代表即使在零样本情境下也能获得不错的效果，体现了实体描述文本信息的重要性。需要指出的是， $d-d$ 类三元组中头实体和尾实体均为新实体，是理论上最为困难的测试样例类。然而 $d-d$ 类在零样本实体预测任务中却取得远超其它三元组分类的准确率。这是因为 $d-d$ 类中有很大部分是自环型的关系，即实体自身相连的关系，所以实体预测反而更加容易。

(2) CNN模型在两个任务上的表现均优于CBOW模型。具体地，CNN模型在实体预测任务上比CBOW模型提高约4.2%，在关系预测任务上提高约7.9%。这说明卷积神经网络模型与连续词袋模型相比考虑了词序信息，能够更好地抓住实体描述中关键的文本信息。

(3) Partial-CBOW模型和Partial-CNN模型在两个任务上也得到了不错的效果。这是因为DKRL模型的能量函数是由实体两种表示的四种组合联合形成的。通过 $\|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_D\|$, $\|\mathbf{h}_D + \mathbf{r} - \mathbf{t}_S\|$ 两个混合项的软限制, 实体基于描述和基于结构的表示倾向被学习到同一个语义空间中, 理性状态下能够相互替换。我们在实验过程中尝试过使用更为严格的限制(如强制实体的两种知识表示相等)强化这种语义空间的可替代性, 但是这会使得模型在其它任务上的表现下降。最后我们在模型中仅保留了混合项的软限制, 在确保知识表示质量的前提下, 一定程度维护了实体两种表示的可替代性。

表 2.7 FB20K上零样本实体预测的评测结果

三元组类型	$d - e$	$e - d$	$d - d$	Total
Partial-CBOW	26.5	20.9	67.2	24.6
CBOW	27.1	21.7	66.6	25.3
Partial-CNN	26.8	20.8	69.5	24.8
CNN	31.2	26.1	72.5	29.5

表 2.8 FB20K上零样本关系预测的评测结果

三元组类型	$d - e$	$e - d$	$d - d$	Total
Partial-CBOW	49.0	42.2	0.0	46.2
CBOW	52.2	47.9	0.0	50.3
Partial-CNN	56.6	52.4	4.0	54.8
CNN	60.4	55.5	7.3	58.2

2.4.5.2 零样本实体类型分类结果与分析

在零样本实体类型分类实验中, 我们仿照2.4.4节的实验设计, 将前50类型覆盖的全部13,445个实体(这些实体在FB20K的训练集中出现过)作为训练集, 使用实体基于描述的表示作为实体特征向量, 采用逻辑斯蒂回归训练分类器。我们将FB20K中4,050个被前50类型覆盖的新实体作为测试集, 依旧使用MAP作为评测指标。我们测试了DKRL(CBOW)和DKRL(CNN)模型, 由于平移模型无法对新实体进行表示, 我们沿用了2.4.4节的BOW模型作为基线模型。

实验结果在表2.9中展示。从结果中我们可以发现, DKRL(CNN)模型取得了最佳结果, 比DKRL(CBOW)模型和BOW模型分别提高了9.9%和4.4%。这个结果

再次验证了卷积神经网络构建的基于描述的实体表示较好地将结构信息与文本信息进行了编码。

表 2.9 FB20K上实体类型分类的评测结果

评测指标	MAP
BOW	57.5
DKRL(CBOW)	52.0
DKRL(CNN)	61.9

2.5 本章小结

在本章中，我们提出DKRL模型实现了融合实体描述的知识表示学习任务。我们基于平移模型的思想建立知识表示学习框架，并创新性地为每个实体设置了两种知识表示——基于结构的表示与基于描述的表示。我们使用连续词袋模型与卷积神经网络模型两种实体描述编码器对实体描述进行建模，从实体描述中学习实体基于描述的知识表示。模型在知识图谱补全和实体类型分类等任务上进行了评测，均取得了超过基线模型的效果，同时也在零样本情境下验证了模型对新实体的表示能力。实验结果充分证明了实体描述中蕴含的丰富文本信息能够帮助建立更好的知识表示，同时也说明我们提出的融合实体描述的知识表示学习模型能够准确地对实体描述和三元组结构信息进行联合建模。

第3章 融合实体层次类型信息的知识表示学习

3.1 引言

融合实体层次类型信息的知识表示学习任务，主要关注如何充分利用知识库中实体具有层次结构的类型信息，从而帮助知识表示的构建，提升知识表示相关任务的性能。现实中的实体往往存在着层次结构，有着不同粒度的类型，而人类对于世界的认知往往也是存在层次结构的。层次化的认知系统能够帮助我们自动建立实体之间的联系，并通过实体所属的类型信息先验地判断实体可能存在的属性。实体具有层次结构的类型信息，能够在知识图谱的结构化信息基础上提供体系化的类型知识，从而帮助构建更加精准的知识表示。

实体层次类型信息指的是实体所属类型的信息，这类信息往往具有人工定义的结构化体系。一些广泛使用的大规模知识图谱，如Freebase、DBPedia等，以及一些常见的百科类数据库，如维基百科等，都维护有自己的一套实体层次结构类型信息。实体类型提供了实体结构化的先验知识，这些层次化的先验知识由人工精心构建，能够在知识图谱三元组的结构信息上提供准确的辅助信息，加深模型对实体与三元组的理解。

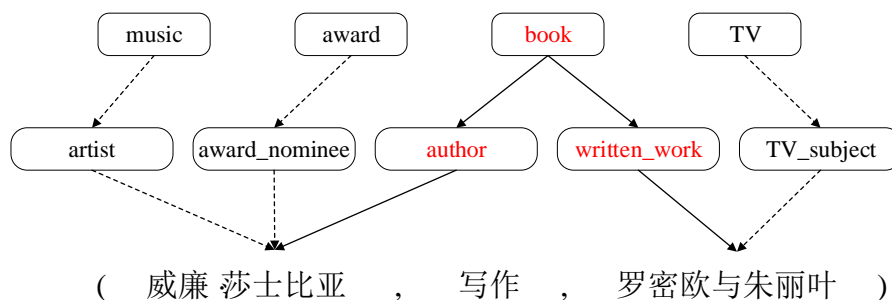


图 3.1 实体层次类型信息示例

图3.1给出了Freebase中样例实体的层次类型结构示意图。图中给出了三元组（威廉·莎士比亚，写作，罗密欧与朱丽叶）及其头实体和尾实体所属的类型。实体类型由自顶向下至实体的一条路径进行表示，如*music/artist*表示一个实体类型，其中*music*和*artist*表示实体不同粒度的子类型。在此分类体系下，一个实体可能属于多个类型，而这些类型具有层次结构，例如威廉·莎士比亚属于*music/artist*，*award/award_nominee*和*book/author*等类型，而罗密欧与朱丽叶则属于*book/written_work*与*TV/TV_subject*等类型。

平移模型基于 $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$ 的平移假设，将实体与关系映射到同一低维连续向量空间中构建知识表示。这种模型虽然简单高效，在大规模知识图谱表示任务上取得了较好的效果，但是在处理复杂关系，如一对多、多对一和多对多关系时往往会出现问题。这是由于有些实体拥有多种类型，需要在不同情境下突出实体不同方面的属性（如威廉·莎士比亚拥有的不同身份），而平移模型仅对每个实体学习了一个实体表示，一定程度上限制了模型对复杂实体的表达能力。

为了解决这个问题，我们尝试通过实体类型信息的帮助，辅助建立实体在不同情境下的多个表示。我们认为在特定关系下的实体应该有更需要突出的实体类型，而实体在突出不同类型时应该有不同的知识表示。以图3.1为例，我们发现在写作这个关系下，威廉·莎士比亚更应突出`book/author`这个类型，而罗密欧与朱丽叶更应突出`book/written_work`这个类型，这些相对重要的实体类型在图中以实线进行连接。所以，对于图中的三元组，应该由威廉·莎士比亚在`book/author`上的知识表示与罗密欧与朱丽叶在`book/written_work`上的知识表示进行交互。

融合实体层次类型信息的知识表示学习主要有以下难点与挑战：首先，如何在不同情境下根据实体的不同表现类型对实体进行不同的表示。大多数实体都拥有着复数的类型，如何判断一个实体在特定三元组中突出了哪些类型特征，是一个挑战性的问题。其次，实体类型往往具有层次结构，这些层次结构信息能够帮助我们建立实体、类型之间的联系。例如通过人类——作家——英国作家这条层次类型，我们可以得知莎士比亚与简·奥斯汀同属英国作家，而与爱因斯坦只是同属人类，其知识表示应该更与前者接近。然而，由于各个知识库维护的类型层次结构在层次深度、分类粒度和分类标准等方面不尽相同，如何设计一个普适且高效的算法，充分利用实体类型的层次结构信息，也是一个值得研究的课题。

为了解决以上难点与挑战，我们设计了一种融合实体类型的知识表示学习模型（Type-embodied knowledge representation learning, TKRL）^[39]。TKRL模型基于平移模型的假设，并在考虑知识图谱结构信息的同时引入实体层次类别信息。具体地，TKRL模型认为每个实体在不同关系下应该突出不同的实体类型，并通过类型特化的映射矩阵，使得实体拥有不同的基于类型的实体表示。为了充分利用实体类型的层次结构信息，我们设计了两种层次类型编码器帮助建立实体类型特化的映射矩阵。另外，我们还在经典的最大间隔学习框架中加入了软类型限制，进一步提高了知识表示的性能。我们在知识图谱补全和三元组分类等任务上对模型进行了评测，在多种实验设定下进行详细的对比分析，并探索了模型在长尾分布数据上的效果。概括起来，本章工作的主要创新点有以下几点：

- 我们创新性地知识表示学习任务中使用实体的类型信息构建类型特化的

映射矩阵，同时在训练过程中提出了软类型限制，使实体在不同类型下拥有不同的知识表示。

- 我们设计了两种层次类型编码器，充分利用了实体类型的层次结构信息，提高了知识表示的效果。
- 我们的模型在知识图谱补全和三元组分类等任务上，相比基线模型均获得了显著的效果提升。

3.2 相关工作

实体层次类型信息通过人工构建的分类体系，对实体进行细致准确的划分与归类。层次结构的信息由于其高质量与结构化的优势，被广泛用于词语上下位关系挖掘^[40]、实体表示^[41]等任务。在知识表示领域，融合实体类型信息的知识表示学习工作还比较初步，其中比较有代表性的是krompaß等研究者的工作^[42]。这篇工作将实体类型信息引入了多种知识表示学习的模型中，提升了知识表示的效果。具体地，在RESCAL模型中，krompaß将实体特征矩阵根据其对应的类型进行了限制；而在平移模型中，krompaß对最大间隔策略中的三元组负例选取，根据实体类型进行了硬限制。与我们提出的TKRL模型相比，上述工作并没有实现实体在不同类型下的多种表示，同时也没有考虑到实体类型具有的层次结构信息。另外，选取负例时使用实体类型进行硬限制，往往使得实体评测结果严重依赖于实体类型的质量与完整度。我们在实验中会对这种情况进行详细分析，并给出可能的改进方法。

我们使用了类型特化的映射矩阵帮助构建实体在不同类型下的知识表示。为了解决平移模型对复杂关系的建模问题，也有一些基于平移假设的模型为实体构建了多种表示。TransH模型^[21]在关系向量之外，为每个关系设置了一个超平面。头实体和尾实体先被投影至关系对应的超平面上，然后再进行相关的平移操作。TransR模型^[22]更进一步，为实体与关系设置了不同的语义空间。在学习实体之间的平移关系时，需要将头实体与尾实体从实体空间映射到特定的关系空间中。TransR模型还将关系细分为多个子关系，提出了CTransR模型，进一步提升了知识表示的效果。TransD模型^[23]基于TransR的思想使用映射矩阵建立实体在不同关系下的不同表示，并且使得映射矩阵与实体和关系均相关。与这些工作相比，我们使用了实体层次类型作为外部信息指导映射矩阵的生成，充分利用了层次类型拥有的实体之间的隐含信息，使得实体在不同类型下的知识表示更为合理。

3.3 算法模型

融合实体层次类型信息的知识表示学习旨在充分利用实体具有层次结构的类型信息，帮助知识表示的构建。我们使用了类型特化的实体映射与训练中的软类型限制两种策略，将实体类型信息编码进知识表示。模型流程图如图3.2所示：

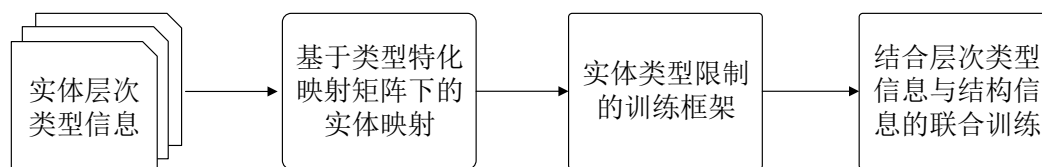


图 3.2 融合实体层次类型的知识表示学习模型流程图

3.3.1 符号体系与重要概念

我们首先介绍本章所用的符号体系。在知识图谱方面，我们使用 (h, r, t) 表示三元组，其中 h 代表头实体， t 代表尾实体， r 代表关系。我们使用 E 表示实体集合， R 表示关系集合， T 表示三元组集合，且 $h, t \in E$ ， $r \in R$ 以及 $(h, r, t) \in T$ 。

在实体类型方面，我们使用 C 表示实体类型集合。如图3.1所示，实体类型存在层次结构。我们认为一个实体层次类型 c 是一条自顶向下的由实体子类型组成的路径，使用 $c = \{c^{(k)} \rightarrow c^{(k-1)} \rightarrow \dots \rightarrow c^{(1)}\}$ 进行表示，其中 $c^{(i)}$ 表示实体层次类型 c 自底向上的第 i 个子类型，即 $c^{(1)}$ 表示实体最细粒度的子类型，而 $c^{(k)}$ 表示最粗粒度的子类型，而 k 表示实体类型的层次数。在Freebase的分类体系下，每个子类型 $c^{(i)}$ 最多只有一个父亲子类型 $c^{(i+1)}$ 。

我们定义了一种关系特化的类型信息，用以确定实体在不同关系下应该表现出的实体层次类型。例如，此信息会告诉我们在写作这个关系下，头实体应该表现出 $book/author$ 这个类型，而尾实体应该表现出 $book/written_work$ 这个类型。关系特化的类型信息可以直接从知识图谱中获取，也可以通过统计等方法进行整理。

3.3.2 模型整体架构与学习框架

融合实体类型的知识表示学习模型基于实体层次类型信息构建类型特化的映射矩阵，为实体在不同类型下建立不同的知识表示。TKRL模型基于平移模型 $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$ 的假设做出改进，设计了如下能量函数：

$$E(h, r, t) = \|\mathbf{M}_{r,h,1}\mathbf{h} + \mathbf{r} - \mathbf{M}_{r,t,2}\mathbf{t}\|. \quad (3-1)$$

其中 $\mathbf{M}_{r,h,1}$ 表示实体 h 的类型特化映射矩阵，由 h 在关系 r 上作为头实体时所突出的类型决定；而 $\mathbf{M}_{r,t,2}$ 则相应地表示尾实体 t 的类型特化映射矩阵。我们提出了两种层次类型编码器，用以构建实体的类型特化映射矩阵，并在关系特化的类型信息帮助下进行匹配和训练。

在训练过程中，我们使用最大间隔策略。这种方法并不是直接优化能量函数的绝对值，而是尝试使得正例的能量函数得分与负例得分之间产生差距。形式化地，我们设计了如下基于最大间隔的优化函数：

$$L = \sum_{(h,r,t) \in T} \sum_{(h',r',t') \in T'} \max(\gamma + E(h,r,t) - E(h',r',t'), 0). \quad (3-2)$$

其中 $(h',r',t') \in T'$ 表示负例三元组， T' 是负例三元组的集合， γ 是最大间隔方法控制正负例得分间隔的超参数。需要注意的是，此处三元组的能量函数 $E(h,r,t)$ 与原始平移模型相比，增加了基于类型特化的矩阵映射操作。对于给定的正例三元组 (h,r,t) ，我们从以下的三元组负例集合中选择负例：

$$T' = \{(h',r,t) | h' \in E\} \cup \{(h,r,t') | t' \in E\} \cup \{(h,r',t) | r' \in R\}, \quad (h,r,t) \in T. \quad (3-3)$$

即将正例三元组中任一实体（或关系）随机替换成另外的实体（或关系），以此产生负例三元组。我们每次从 T' 中选择一个负例三元组，与正例联合进行成对训练。需要注意的是，我们还在负例生成时加上了以下限制条件：

$$\forall (h',r',t') \in T', \quad (h',r',t') \notin T. \quad (3-4)$$

保证随机替换后生成的负例三元组中不存在真正的正例。我们还基于实体类型限制，对最大间隔训练时的负例选择进行了进一步限制，以提升知识表示的精确度。这一点将在后面的章节进行详细阐述。

为了充分利用实体类型具有的层次结构信息，我们设计了两种层次类型编码器，基于子类型矩阵连乘与加权和等形式，构建实体的类型特化映射矩阵。

3.3.3 层次类型编码器

层次类型编码器旨在充分利用实体类型的层次结构信息，并在关系特化的类型信息帮助下构建实体在不同类型下的不同知识表示。在本节，我们首先介绍实体在特定关系下类型映射矩阵的一般形式，然后分别介绍递归层次编码器和加权层次编码器两种对层次结构的不同编码策略。

3.3.3.1 类型映射矩阵的一般形式

我们对每一个层次类型 c 设置一个类型映射矩阵 \mathbf{M}_c 。给定三元组 (h, r, t) ，如前所述，我们需要构建头实体和尾实体在关系 r 下的类型映射矩阵 $\mathbf{M}_{r,h,1}$ 和 $\mathbf{M}_{r,t,2}$ 。对于 $\mathbf{M}_{r,h,1}$ ，我们可以根据关系特化的类型信息，获得 h 在关系 r 下的头实体位置所应该突出的类型集合，以 $C_{r,h,1} = \{c_1^{(r,h,1)}, c_2^{(r,h,1)}, \dots, c_n^{(r,h,1)}\}$ 进行表示。同理，对于尾实体，我们用 $C_{r,t,2} = \{c_1^{(r,t,2)}, c_2^{(r,t,2)}, \dots, c_n^{(r,t,2)}\}$ 表示 t 在 r 和尾实体位置上应该突出的类型集合。综上，我们定义头实体在三元组 (h, r, t) 中类型映射矩阵 $\mathbf{M}_{r,h,1}$ 的一般形式如下：

$$\mathbf{M}_{r,h,1} = \alpha_1 \mathbf{M}_{c_1^{(r,h,1)}} + \alpha_2 \mathbf{M}_{c_2^{(r,h,1)}} + \dots + \alpha_n \mathbf{M}_{c_n^{(r,h,1)}}. \quad (3-5)$$

其中 $c_i^{(r,h,1)}$ 表示 h 在 r 上应该突出的第 i 个类型。 α_i 表示 $c_i^{(r,h,1)}$ 类型在这个关系中的重要程度，可以通过类别频次、互信息等统计量进行度量。同理，尾实体类型矩阵 $\mathbf{M}_{r,t,2}$ 也有类似的表示：

$$\mathbf{M}_{r,t,2} = \alpha_1 \mathbf{M}_{c_1^{(r,t,2)}} + \alpha_2 \mathbf{M}_{c_2^{(r,t,2)}} + \dots + \alpha_n \mathbf{M}_{c_n^{(r,t,2)}}. \quad (3-6)$$

由于实体类型具有层次结构，各子类型之间存在着上下级关系、兄弟关系等重要隐藏信息。我们参考Hu等研究者的工作思想^[41]，为每一个子类型也设置一个映射矩阵，使用这些子类型的映射矩阵联合构建层次类型的映射矩阵。我们设计了递归层次编码器和加权层次编码器两种方法，基于层次结构信息构建类型映射矩阵 \mathbf{M}_c 。

3.3.3.2 递归层次编码器

递归层次编码器（Recursive hierarchy encoder, RHE）是一种简单直观的利用子类型映射矩阵构建层次类型映射矩阵的方法。递归层次编码器自顶向下，通过子类型映射矩阵的递归连乘构建层次类型的映射矩阵。形式化地，对于一个具有 m 层子类型的类型 c ，递归层次编码器构造其类型映射矩阵如下：

$$\mathbf{M}_c = \prod_{i=1}^m \mathbf{M}_{c^{(i)}} = \mathbf{M}_{c^{(1)}} \mathbf{M}_{c^{(2)}} \cdots \mathbf{M}_{c^{(m)}}. \quad (3-7)$$

其中， $c^{(m)}$ 表示 c 的层次结构中最粗粒度的子类型， $c^{(1)}$ 表示 c 最细粒度的子类型，而 $\mathbf{M}_{c^{(i)}}$ 表示子类型 $c^{(i)}$ 对应的映射矩阵。

递归层次编码器具有合理的物理意义：以（莎士比亚，写作，哈姆雷特）这个三元组为例，莎士比亚在写作关系的头实体位置应该表现出人类——作家——英国作家这个层次类型的属性。在递归层次编码器中，莎士比亚的实体向量首先通过人类对应的子类型映射矩阵，从实体空间映射到人类这个子类型的空间中，得到莎士比亚在人类这个子类型下的表示。然后，映射后的表示将被依序映射到更细粒度的子类型空间中，得到莎士比亚在作家下的表示，最终得到了莎士比亚在英国作家语义空间中更精确的知识表示。莎士比亚在英国作家上的表示最终将与哈姆雷特在英文小说上的表示一起进行平移操作，更新编码器的参数与对应的知识向量。图3.3左图给出了递归层次编码器的结构示意图。

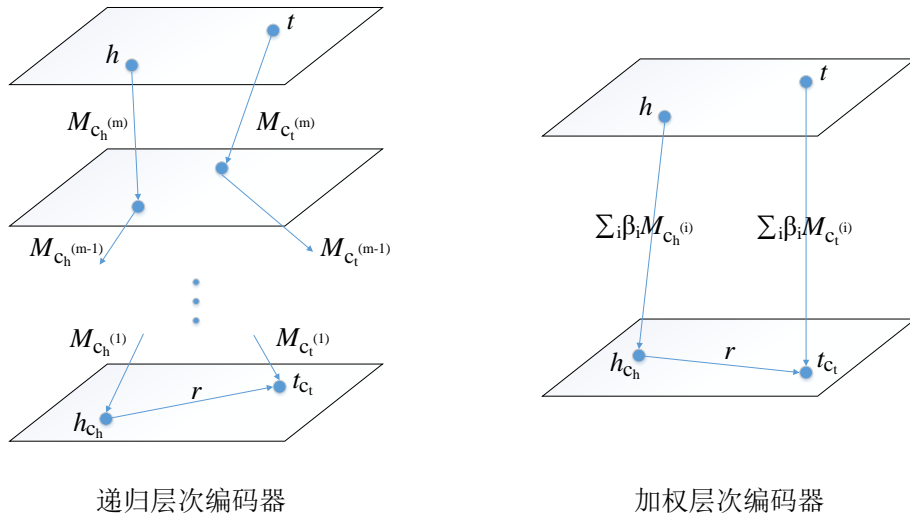


图 3.3 两种层次类型编码器示意图

3.3.3.3 加权层次编码器

递归层次编码器基于矩阵连乘的形式，使实体从粗粒度到细粒度依次映射到对应的类型空间中。但是，在层数较多的复杂分类系统下，连乘运算的计算复杂度较高，而且也未能考虑各层子类型对于实体不同程度的影响。针对这些问题，我们提出了加权层次编码器（Weighted hierarchy encoder, WHE），采用矩阵加权的形式，通过子类型的映射矩阵构建层次类型的映射矩阵，针对实体不同粒度的子类型映射矩阵设定了不同的权值。对于一个具有 m 层子类型的类型 c ，加权层次编码器的具体公式如下：

$$\mathbf{M}_c = \sum_{i=1}^m \beta_i \mathbf{M}_{c^{(i)}} = \beta_1 \mathbf{M}_{c^{(1)}} + \cdots + \beta_m \mathbf{M}_{c^{(m)}}. \quad (3-8)$$

其中 $\mathbf{M}_{c^{(i)}}$ 表示子类型 $c^{(i)}$ 对应的映射矩阵， β_i 表示 $c^{(i)}$ 对应的权值。我们认为实体越细粒度的子类型对实体的表示影响越大，其对应的权值也应该越大，并据此设计了一个等比例下降的加权策略。我们有：

$$\beta_i : \beta_{i+1} = (1 - \eta) : \eta, \quad \sum_{i=1}^m \beta_i = 1. \quad (3-9)$$

其中 $\eta \in (0, 0.5)$ 以保证细粒度子类型的权值更大。

加权层次编码器将实体层次结构的子类型综合起来进行考虑，根据类型的层次结构，为更细粒度的子类型映射矩阵设置更大的权值。另外，通过矩阵加权和而不是矩阵连乘的方法对映射矩阵进行构建，仅使用一步映射便能获得实体在层次类型下的知识表示。实验结果证明，使用加权层次编码器的方法能够在保证知识表示性能的前提下提高计算效率。

3.3.4 基于实体类型的限制

在上一小节，我们介绍了如何利用实体类型的层次结构，构建实体类型特化的知识表示。而除此之外，实体类型信息还能在模型中作为类型限制，帮助在知识表示任务中获得更好的结果。在本节，我们将分别介绍实体类型限制在训练中与测试中的使用。

3.3.4.1 训练中的实体类型限制

在融合实体类型的知识表示学习模型中，我们使用最大间隔方法，通过最大化正例三元组和负例三元组的得分间隔对知识表示进行学习。而由于知识图谱中并没有显式的负例三元组，我们按照公式3-3中的方法构建负例。在构建负例的实体替换中，我们将原三元组 (h, r, t) 替换成 (h', r, t) ，其中 h' 仅需要满足 $h' \in E$ 以及 $(h', r, t) \notin T$ 即可。在这种情况下，拥有相似关系的实体（这些关系往往拥有相似的类型）由于分布式表示学习的假设，在最终的实体语义空间中倾向于聚在一团。由于没有直接针对这些拥有相似类型的实体进行更加精准的学习，这些相似的实体在向量空间中往往难以区分，成为知识图谱补全和三元组分类等任务的主要错误来源之一。

为了解决这个问题，krompaß等研究者曾尝试使用硬类型限制控制负例三元组的生成模式^[42]。具体地，他们认为在生成 (h, r, t) 的负例时，替换的实体 h' 必须满足 $h' \in E_c \subseteq E$ 。其中， E_c 表示一个实体集，在此实体集中的所有实体都必须有类型 c ，而类型 c 是实体 h 在关系 r 的头实体位置所属于的类型。这种硬类型限制方

法强制负例三元组中替换的实体与被替换实体属于同一类型，强化了拥有相同类型的实体之间的差异，在一定程度上能够解决相似实体的知识表示难以区分的问题。但是，硬类型限制方法过于强调相似实体之间的差异，可能扰乱实体向量空间中相似实体的正常聚类现象，甚至可能导致在知识图谱补全和三元组分类等任务上效果的下降。

我们改进了这种硬类型限制方法，在负例选择中提出了软类型限制（Soft type constraint, STC）的方法。这种方法并不是强制所有负例实体满足类型限制，而是以一定比例选择同类型的实体作为替换。在负例选择中替换的实体 e' 属于被替换实体 e 同类型的比例如下式：

$$Pr(e' \in E_c) = \frac{(k+1)|E_c|}{|E| + k|E_c|}, \quad k \in \mathbb{N}. \quad (3-10)$$

其中 c 是实体 e 在此三元组中所属的类型， $|E_c|$ 和 $|E|$ 分别是对应实体集合的大小。 k 是控制同类型替换比例的超参数， k 越小时说明实体类型限制越不严格。这个公式意味着在选择实体的负例时，属于被替换实体同一类型的实体被选中的概率，比不属于同一类型的实体被选中概率多 k 倍。在训练中使用软类型限制时，我们既能够保持实体向量空间中相似实体的正常聚类现象，也能够强调拥有同类型的相似实体之间的差异，使得实体的知识表示更为精确。

我们使用一个实例对软类型限制的意义进行阐述。对于正例三元组（莎士比亚，写作，哈姆雷特），软类型限制会以一定比例提高负例实体属于同一类型的概率，即选择简·奥斯汀作为负例的概率大于选择足球作为负例的概率。在这种情况下，最大间隔法学习到的更多是“简·奥斯汀没有写过哈姆雷特”，而这样的信息比起“足球没有写过哈姆雷特”更有信息量。但是，软类型限制也会产生非同类型的实体负例，也能通过知识图谱的整体学习，使得语义向量空间中莎士比亚和简·奥斯汀的实体表示接近。

3.3.4.2 测试中的实体类型限制

在训练中进行的软类型限制之外，我们也尝试了测试中的类型限制（Type constraint in evaluation, TCE）对知识表示的影响。测试中的类型限制比较直观，即在知识图谱补全任务中，将所有违反了实体类型限制的候选实体从候选集中删除。具体地，如对（莎士比亚，写作，?）进行补全，不进行类型限制的方法将所有实体集中的实体作为候选实体填入尾实体位置，计算能量函数并排序；而类型限制方法直接根据关系特化的类型信息，得到在关系 r 的尾实体位置实体应该表达

的类型，只对拥有这些类型的实体计算能量函数。这样，模型在补全样例时只会选择属于文学作品类型的实体，去除了绝大部分类型不符的错误答案。

测试中的实体类型限制在测试时加入严格的限制条件，大大减少了知识图谱补全时的实体候选数量。然而这种实体类型限制的效果对实体类型的完整度与准确度要求很高。当实体类型完整且准确时，测试中的实体类型限制去除的都是错误答案，测试结果理论上不可能下降；然而，当实体类型信息不完整时，测试中的实体类型限制可能错判而将正确答案去除，这样对知识图谱补全等任务是致命的。我们在实验中详细分析了实体类型限制在各个模型上的测试结果，并进行了具体分析。

3.3.5 模型优化与具体实现

融合实体类型的知识表示学习模型使用类型特化的矩阵映射以及类型限制等方法，在平移模型的框架下学习知识表示。在实际训练中，TKRL模型需要学习以下参数组： $\theta = (\mathbf{E}, \mathbf{R}, \mathbf{M})$ 。其中 \mathbf{E} 代表实体在实体空间的向量表示， \mathbf{R} 代表关系向量， \mathbf{M} 表示所有子类型的映射矩阵。TKRL模型使用Mini-batch随机梯度下降算法进行训练，并根据链式法则反向传播更新所有需要学习的参数。

TKRL模型有以下实现细节：在初始化时， \mathbf{E} 和 \mathbf{R} 可以随机初始化，或者由平移模型进行预训练，所有子类型的映射矩阵均被初始化成单位矩阵。为了提高模型计算效率，我们实现了一个基于多线程的模型版本。

关系特化的类型信息能够确定实体在不同关系下所应该重视的实体类型，对模型效果影响很大。在实验中，我们可以直接使用知识图谱维护的关系特化的类型信息（例如Freebase中的`rdf-schema#domain`和`rdf-schema#range`信息），也可以基于局部封闭世界假设（local closed-world assumption, LCWA）^[43]，根据观测到的三元组信息进行统计。具体地，局部封闭世界假设认为所有出现在同一关系同一位置（头实体或尾实体）的实体具有同一类型，并且这个类型与具体关系和实体所处位置对应。

3.4 实验设计与结果分析

在本章节，为了测试融合实体层次类型的知识表示学习模型的效果，我们进行了知识图谱补全和三元组分类等实验，并探索了模型在具有长尾分布的数据上的表现效果。实验结果证明了实体层次类型信息可以很好地作为知识图谱自身三元组结构信息的补充，辅助构建更好的知识表示，同时也证明了我们的模型能够充分利用实体层次类型信息，构建实体在不同情境下的不同知识表示，在真实分

布的数据集上也能取得较理想的效果。

3.4.1 实验数据集

我们在实验中使用了FB15K数据集^[17]作为知识图谱数据集。FB15K抽取自大规模知识图谱Freebase，共拥有14,951个实体，1,345个关系和592,213个三元组。我们沿用了Bordes论文中对FB15K的划分，将这些三元组划分为训练集、验证集和测试集^[17]。对于实体类型信息，我们在Freebase中通过`type/instance`域收集了FB15K中所有实体的层次类型信息，同时通过`rdf-schema#domain`和`rdf-schema#range`域收集了所有关系对应的关系特化类型信息。进一步地，我们删除了未出现在抽取的关系特化类型信息中的所有实体类型信息，因为这些类型并未对FB15K中的三元组产生直接影响。综上，我们一共得到了571个实体层次类型，并保证FB15K中所有实体都至少属于其中一个类型，而每个实体平均属于8个类型。FB15K的统计数据详见表3.1。

表 3.1 FB15K和FB15K+的统计数据

数据集	#关系	#实体	#类型	#训练集	#验证集	#测试集
FB15K	1,345	14,951	571	483,142	50,000	59,071
FB15K+	1,855	14,951	806	486,446	50,000	62,374

真实世界中的知识图谱往往具有长尾分布的特征，低频的实体与关系占据了知识图谱中大部分比例。而FB15K数据集在构建时设置了实体与关系在Freebase中的最低频次，并去除了那些低频的实体与关系。为了测试TKRL模型在真实世界中对长尾分布数据的效果，我们在FB15K的基础上构造了FB15K+的数据集，用以模拟真实世界的长尾分布。具体地，我们基于FB15K中的实体集，从Freebase中抽取了所有存在于这些实体之间的频次大于1的关系（因为只出现1次的关系不能同时出现在训练集与测试集中，所以我们并未抽取这些关系）。这样，我们在FB15K的基础上增加了510个新关系，并根据这些实体与关系新抽取了6,607个三元组。我们将这些三元组随机划分至FB15K的原始训练集与测试集中，并保证所有关系均在训练集中出现过。对于新增加的关系，我们同样抽取了它们的关系特化类型信息与层次类型信息，最后FB15K+中一共包含806个层次类型。FB15K+的统计数据详见表3.1。

3.4.2 实验与模型参数设置

在实验中，我们使用了两种层次类型编码器对实体类型的层次结构进行建

模，使用递归层次编码器的模型记录为TKRL(RHE)，使用加权层次编码器的模型记录为TKRL(WHE)。我们将训练中使用了软类型限制的模型加上(+STC)的标志，将测试中使用了类型限制的模型加上(+TCE)的标志。为了充分展示融合实体层次类型信息的知识表示学习效果，我们实现了平移模型(TransE)^[17]和基于平移模型改进的TransR模型^[22]，作为实验的基线模型，并加入了关系负例改进了基线模型的训练。为保证公平，基线模型均以其最佳参数进行训练，并保证了知识表示向量维数的一致。另外，我们还报告了其它基线模型的结果，如RESCAL模型^[13,14]、结构向量模型(SE)^[7]、语义匹配能量模型(SME)^[8]、隐变量模型(LFM)^[10,11]等。由于实验设定的一致，我们直接给出了Lin等研究者在论文中报告的结果^[22]。

我们使用随机梯度下降算法对TKRL模型进行训练，训练时的最大间隔参数 γ 在 $\{0.5, 1.0, 1.5, 2.0\}$ 中进行尝试。对于学习率，我们仿照之前的知识表示学习模型尝试了固定值的学习率。另外，我们也设计了一种随迭代轮次上升而自适应下降的学习率策略。对于加权层次编码器，我们尝试了不同的权值下降超参数 η ，在 $\{0.1, 0.15, 0.2\}$ 进行取值。对于训练中的软实体限制超参数 k ，我们尝试了 $k = \{5, 10, 15\}$ 。最终，我们选择了以下最优参数进行结果展示： $\gamma = 1.0, \eta = 0.1, k = 10$ ，模型学习率 λ 使用一种线性下降的策略，在迭代中从0.0025线性下降到0.0001。在初始化中，我们使用平移模型预训练的知识向量作为我们模型中实体与关系向量的初始值。为了方便实现模型之间的比较，所有模型均在知识向量维度 $n = 50$ 的情况下进行训练。

3.4.3 知识图谱补全结果与分析

知识图谱补全是知识表示学习的经典评测任务，要求知识表示模型在测试集三元组中某一实体或关系丢失的情况下，利用三元组剩下的信息正确预测出丢失的实体或关系。知识图谱补全根据需要预测的对象，可以分为实体预测与关系预测两个任务。

3.4.3.1 实验设计

我们使用融合实体层次类型的知识表示学习模型和基线模型在训练集上学到的实体与关系向量，对测试集中的三元组样例进行知识图谱补全。对于测试集中的每个样例，我们分别隐藏其头实体、关系和尾实体，用于进行实体预测和关系预测。在测试时，我们将全部实体(或关系)作为候选集，根据各模型的能量函数为每一个候选三元组进行评分。具体地，对于平移模型，我们使用 $S(h, r, t) =$

$\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$ 的评分函数；对于TransR模型，我们使用 $S(h, r, t) = \|\mathbf{M}_r \mathbf{h} + \mathbf{r} - \mathbf{M}_r \mathbf{t}\|$ 的评分函数；而对于TKRL模型，我们使用 $S(h, r, t) = \|\mathbf{M}_{r,h,1} \mathbf{h} + \mathbf{r} - \mathbf{M}_{r,t,2} \mathbf{t}\|$ 的评分函数。我们根据三元组得分的进行升序排列，完成知识图谱补全。

我们参考Bordes论文中的评测方法，使用Mean Rank和Hits@N作为评测指标^[17]。Mean Rank表示正确实体（或关系）的平均排序，旨在对知识表示的整体效果进行评估；而Hits@N表示排名前N的结果中有正确答案的概率，旨在说明模型在知识图谱补全中的实用性。由于候选实体的数量大大超过了关系数量，我们使用Hits@10评测实体预测任务，而使用Hits@1评测关系预测任务。我们也仿照Bordes的设定使用了“Raw”和“Filter”两个指标对知识图谱补全任务进行更精确的评测。

3.4.3.2 实验结果与分析

我们在实体预测与关系预测两个任务上进行了测试，实体预测的结果详见表3.2，关系预测的结果详见表3.3。分析实体预测与关系预测的结果，我们可以得到以下结论：

表 3.2 FB15K上实体预测的评测结果

评测指标	Mean Rank		Hits@10(%)	
	Raw	Filter	Raw	Filter
RESCAL	828	683	28.4	44.1
SE	273	162	28.8	39.8
SME(linear)	274	154	30.7	40.8
SME(bilinear)	284	158	31.3	41.3
LFM	283	164	26.0	33.1
TransE	238	143	46.4	62.1
TransR	199	77	47.2	67.2
TKRL(RHE)	184	68	49.2	69.4
TKRL(WHE)	186	68	49.2	69.6
TKRL(RHE+STC)	202	89	50.4	73.1
TKRL(WHE+STC)	202	87	50.3	73.4

(1) 在实体预测与关系预测两个任务中，TKRL(RHE)模型与TKRL(WHE)模型在所有评测指标上超过了所有的基线模型。这说明了实体层次类型蕴含着丰富信息，能够帮助模型学到更加准确的知识表示。同时也说明了我们设计的递归层次编码器与加权层次编码器能够很好地对实体类型的层次结构进行建模，将实体

层次类型信息顺利地融入知识表示中。

(2) 我们将TKRL模型与同样基于平移假设的基线模型进行比较。与TransE模型相比，TKRL模型在所有评测指标上的提升都十分显著。这从侧面证实了实体在不同类型下应该有不同知识表示的假设，一定程度上也解决了平移模型框架对于复杂关系建模难的问题。另外，TKRL模型与TransR模型相比也获得了不少提升。TransR模型使用了关系特化的映射矩阵，使实体在不同关系下拥有不同的表示，而我们的模型充分利用了实体层次类型信息以及关系特化的类型信息，使得各个层次类型之间能通过层次结构和共享的子类型产生关联。实体预测上的实验结果也证明，这种基于类型先验知识的映射矩阵融合了更多的信息，从而在实体表示上取得了更好的结果。

表 3.3 FB15K上关系预测的评测结果

评测指标	Mean Rank		Hits@1(%)	
	Raw	Filter	Raw	Filter
TransE	2.79	2.43	68.4	87.2
TransR	2.49	2.09	70.2	91.6
TKRL(RHE)	2.12	1.73	71.1	92.8
TKRL(WHE)	2.22	1.83	70.8	92.5
TKRL(RHE+STC)	2.38	1.97	68.7	90.7
TKRL(WHE+STC)	2.47	2.07	68.3	90.6

(3) TKRL(WHE+STC)模型在实体预测Hits@10, Filter的评价指标上取得了最好的实验结果，在TKRL(WHE)模型的基础上又提升了约3.8%。而训练中的软类型限制方法在递归层次编码器模型上也能带来相似的提升。这是因为在训练时，软类型限制能够增大拥有同类型的实体被选中组成负例三元组的概率。这样的训练方式与之前的方法相比，能够使得拥有相同类型的实体向量之间差异度更高，从而在测试时得以对相似实体进行更精确的区分。然而，使用软类型限制的方法在Mean Rank指标上会稍稍下降。这是因为类型限制方法不可避免地减弱了相似实体在向量空间中的聚类现象，增加了极端错误的可能性，而Mean Rank指标容易受到极端错误结果的影响，最终导致了结果的降低。

(4) 比较基于递归层次编码器和基于加权层次编码器的模型，我们发现，前者在关系预测上表现较好，而后者在实体预测上表现较好，但是两者相差较小。这一部分是因为我们在实验中使用的是基于Freebase的实体类型层次结构，而这些层次结构大都较浅，无法充分体现两种层次编码器的差别。但是在运行速度上，

基于加权层次编码器的模型还是明显快于基于递归层次编码器的模型。

3.4.3.3 测试中的实体类型限制分析

测试中的实体类型限制也能提高模型在知识图谱补全任务上的性能。由于测试中的实体类型限制方法受限于实体类型的完整度与准确度，为了显示模型的鲁棒性，我们仅将此部分结果单独作为辅助实验，并对测试中实体类型限制的效果进行详尽分析。

我们在TKRL(RHE+STC)模型和TKRL(WHE+STC)模型上加入测试中的实体类型限制，得到了TKRL(RHE+STC+TCE)模型和TKRL(WHE+STC+TCE)模型。需要注意的是，与krompaß等研究者的测试设定不同^[42]，我们在测试中仍使用实体集合中全部实体作为候选集（即与传统知识表示学习的测试设定相同）。为了更充分地展示我们模型的优势，我们为基于平移假设的基线模型也增加了训练和测试中的实体类型限制。实验结果详见表3.4。将表3.4的结果与表3.2中对应模型的结果进行比较，我们可以得到以下结论：

表 3.4 测试中的实体类型限制在实体预测上的效果

评测方法	Mean Rank		Hits@10(%)	
	Raw	Filter	Raw	Filter
TransE+TCE	212	116	46.9	63.4
TransR+TCE	182	60	47.9	68.6
TransE+STC+TCE	203	104	49.8	69.9
TransR+STC+TCE	185	63	48.5	71.7
TKRL (RHE+STC+TCE)	169	56	51.4	75.4
TKRL (WHE+STC+TCE)	170	55	51.3	75.6

(1) 包括TKRL模型、平移模型和TransR模型在内的所有模型，在加入测试中的实体类型限制时，实体预测的效果都有提升。其中TKRL(WHE+STC+TCE)模型在Hits@10, Filter评测指标上达到了75.6%，比不使用测试时类型限制的模型进一步提升了2.2%。这是因为测试中的实体类型限制去除了违反类型限制的候选实体，在实体类型较完整与准确时，能够缩小候选范围并较少地引入错误，从而提升实体预测的效果。

(2) 对于基于平移假设的基线模型，在训练中使用软类型限制能够进一步提升测试中类型限制的效果。以TransE模型为例，使用测试中的类型限制能够使TransE模型在Hits@10, Filter评测指标上的结果提高1.3%。而再加上训练时的软

类型限制，TransE+STC+TCE模型能够进一步提高6.5%。这是因为训练中的软类型限制能够使得相似实体间的差异明显化，在联合使用测试中的实体类型限制时能够获得更好的效果。

(3) 对比所有使用了训练/测试中实体类型限制方法的模型，TKRL模型仍然能够取得最佳的结果。这充分说明了根据实体层次类型建立类型特化映射矩阵的重要性与我们模型的鲁棒性。

3.4.3.4 长尾分布数据上的结果与分析

由于真实世界中的知识图谱往往具有长尾分布，所以我们构造了FB15K+数据集，用以测试模型在长尾分布下实体预测与关系预测的效果。与FB15K相比，FB15K+中增加了510个关系，而这些关系往往都具有较低频次。我们根据各关系在FB15K+的三元组中出现的频次对测试集划分成组，展示了模型在实体预测与关系预测中各组测试集的Hits@N, Filter评测指标的结果。为了便于展示，我们仅对TransE模型、TransR模型与TKRL(WHE)模型的结果进行比较。实验结果在表3.5中给出。我们可以得到以下结论：

表 3.5 长尾分布数据集FB15K+上的实体预测与关系预测结果

关系频次	实体预测 Hits@10(%)			关系预测 Hits@1(%)		
	TransE	TransR	TKRL(WHE)	TransE	TransR	TKRL(WHE)
<=10	28.0	32.4	38.1 (+5.7)	13.2	17.0	21.5 (+4.5)
<=100	49.9	54.5	57.9 (+3.4)	45.7	50.5	54.3 (+3.8)
<=1000	66.1	69.1	71.6 (+2.5)	70.9	75.4	77.8 (+2.4)
全部	61.9	67.2	69.2 (+2.0)	80.4	88.8	89.7 (+0.9)

(1) TKRL(WHE)模型在所有频次测试组中的实体和关系测试结果均好于TransE模型和TransR模型，这充分证明了融合实体类型信息的知识表示学习模型在真实的长尾分布数据上相对于基线模型也能具有显著优势，也进一步说明了模型的有效性与鲁棒性。

(2) 通过观察我们发现，TKRL(WHE)模型在低频关系下实体和关系预测的结果相比高频提升更大。表中TKRL(WHE)模型的结果后面标注了其相对TransR模型的提升结果。在关系频次小于等于10的分组下，TKRL(WHE)模型在实体预测和关系预测上分别比TransR模型提高了5.7%和4.5%；而在全部测试集上的结果，TKRL(WHE)模型在实体预测和关系预测上仅比TransR模型提高了2.0%和0.9%。这是因为尽管TKRL模型与TransR模型都使用了映射矩阵的方法构建实体在不同

情境下的表示，但是TKRL的构建基于实体层次类型，而TransR的构建基于关系。在对拥有低频关系的实体进行建模时，TransR模型较容易受到数据稀疏及训练不充分的影响；而TKRL模型能够充分利用实体层次类型之间的联系，对低频关系下的实体也能较好地建立表示。

3.4.4 三元组分类结果与分析

三元组分类任务是知识表示学习中又一经典任务^[16,21,22]。三元组分类要求在给定一个三元组的情况下，判断这个三元组是否是正例，本质上是一个针对三元组的二分类任务。三元组分类能够对实体和关系向量的质量进行评估，同时也与现实问答系统中的实际任务相呼应。如需判断三元组（莎士比亚，写作，哈姆雷特）是否是正例，即相当于回答“莎士比亚是否写了《哈姆雷特》？”这个问题。

3.4.4.1 实验设计

我们使用FB15K的测试集进行三元组分类的评测，然而知识图谱中并没有显式的负例三元组。我们参考Socher等研究者在三元组分类任务上的设定^[16]，基于测试集自动构建负例。具体地，对于测试集中每一个正例三元组 (h, r, t) ，我们随机替换掉其中的头实体或者尾实体，组成负例三元组 (h', r, t) 或者 (h, r, t') 。为了使三元组分类任务更具挑战，更大程度上展示模型的性能差异，我们使用了实体类型限制的方法，强制要求替换的负例实体和被替换的原实体拥有相同的类型。例如，（莎士比亚，写作，哈姆雷特）的负例中头实体应该也是一位作者，因为拥有错误类型的负例三元组很容易就会被知识表示学习模型检测出来，从而失去评测意义。在实验中，对于验证集和测试集中的每一个三元组，我们以均等概率随机替换头实体或尾实体，产生对应的负例，这保证了三元组分类任务中正负例的数量一致。

我们使用TransE模型和TransR模型作为基线模型，与TKRL模型进行比较。在评测时，我们仍然使用在知识图谱补全任务中各模型对应的三元组评分函数 $S(h, r, t)$ 进行判定。具体地，我们会为每一种关系设定一个阈值 δ_r ，当 $S(h, r, t) > \delta_r$ 时，三元组被判断为负例，而当 $S(h, r, t) < \delta_r$ 时，三元组被判断为正例。我们在验证集上进行优化，得到各关系所对应的阈值 δ_r ，用于对测试集进行评测。

3.4.4.2 实验结果与分析

三元组分类的结果在表3.6中给出。从结果中我们可以发现以下结论：

- (1) TKRL模型的准确率超过了所有的基线模型，其中TKRL(WHE+STC)模

型得到了最好的结果。这从另一个角度说明了融合实体类型的知识表示学习模型能够学到更准确的知识表示，从而帮助三元组分类等任务。

(2) 加上训练中的软类型限制方法后，递归层次编码器模型和加权层次编码器模型都能获得进一步的提升。这说明了软类型限制能够学习到相似实体之间的差异，从而在三元组分类任务上获得更加精确的结果。

表 3.6 FB15K上三元组分类的评测结果

模型	准确率(%)
TransE	85.7
TransR	86.4
TKRL(RHE)	86.9
TKRL(WHE)	87.1
TKRL(RHE+STC)	88.4
TKRL(WHE+STC)	88.5

3.5 本章小结

在本章中，我们提出了TKRL模型，融合实体层次类型信息辅助知识表示学习。我们在平移模型的思想做出改进，提出实体在不同类型下应该具有不同知识表示的假设，构建类型特化的映射矩阵，并使用递归层次编码器和加权层次编码器对类型的层次结构进行建模。我们还提出了软类型限制的策略，并在训练与测试中分别使用了类型限制的方法，进一步提高知识表示的精确度。模型在知识图谱补全和三元组分类等任务上进行了评测，均取得了超过基线模型的效果，同时也在具有长尾分布的数据集上验证了模型对低频关系的表示效果。实验结果充分证明了实体层次类型中蕴含了丰富的信息，能够帮助建立更精确的知识表示，同时也说明我们提出的融合实体层次类型信息的知识表示学习模型能够有效地联合编码并学习到实体类型的知识。

第4章 融合实体图像信息的知识表示学习

4.1 引言

融合实体图像信息的知识表示学习任务，主要关注将实体相关的图像信息与知识图谱的结构信息结合起来，构建知识的跨模态表示，提升知识表示的性能。图像信息是人类能够自然接收并认知的信息之一，以图像形式储存的信息往往更加灵活，且信息密度高。研究结果也表明，人类每日从外界进行的信息获取与交互行为中，很大一部分是通过视觉来完成的。通过图像信息的帮助。我们往往得以从多角度全方位地理解实体在文字或结构化信息之外的知识细节。

实体图像是主体部分描述指定实体的图像，能够提供对应实体自身外形、行为和和其它相关实体的视觉信息。实体图像信息来源十分丰富，一些通用知识库（如维基百科等）往往会有对应实体的图像信息，而专门的图像数据集更是储存着海量的实体相关图像信息。例如基于WordNet^[5]的经典图像数据集ImageNet^[44]，截止至2017年4月已包含超过1400万的实体图像信息。实体图像蕴含着极丰富的实体相关视觉信息，能够帮助我们全方位地理解实体，从而提高知识表示的性能。同时，这种跨模态的知识表示学习也能够对知识空间与图像空间的交互进行深入探索，从而为知识和图像的联合应用提供可能。



图 4.1 三元组中实体图像信息示例

图4.1给出了三元组中实体及其对应的实体图像的示例图，直观地展示了实体图像信息对知识表示的帮助。我们分别为盔甲和头盔展示了三张实体图像，对于三元组（盔甲，拥有...部分，头盔）这个知识，我们可以直接从实体图像中得到确认。实体图像包含海量有效的视觉信息，但是目前绝大多数知识表示学习模型在学习知识表示时都忽略了这些图像信息。我们在本章中将尝试融合实体图像中的丰富信息，建立跨模态的知识表示，提升知识表示相关任务的性能。

融合实体图像信息的知识表示学习主要有着以下难点与挑战：首先，图像信息与知识信息是异质的，表达形式也存在巨大的差异。如何对图像信息进行合理

的建模，使之能够与知识信息进行交互与联合学习，是我们首先需要解决的问题。其次，现在的大规模图像数据库中，一个实体往往有成千上万张实体图像。这些实体图像质量良莠不齐，对实体描述的粒度与角度也不尽相同，甚至可能出现噪声与错误标注。如何自动地从这些图像候选中选择信息量较大、质量较高的实体图像进行联合学习，也是一项巨大的挑战。例如对于高尔夫这个实体，有些图像样例给出了高尔夫球的特写，有些图像样例给出了运动员在打高尔夫球的场景，也有些图像给出了高尔夫球场的全景。这些实体图像提供的实体相关信息角度都不相同，选取更有信息量的图像与知识空间进行交互，显然能够进一步提升联合模型的知识表示效果。

为了解决以上难点与挑战，我们设计了一种融合实体图像的知识表示学习模型（Image-embodied knowledge representation learning, IKRL）^[45]。IKRL模型沿用平移模型的学习框架，尝试融合实体图像信息学习实体跨模态的知识表示。具体地，IKRL模型为每个实体设置了两种知识表示，包括基于结构的表示和基于图像的表示。我们使用基于卷积神经网络的图像表示模块和图像映射模块，从实体图像中构造图像在知识空间的表示。然后，我们使用注意力模型自动选择高质量的图像，组成实体基于图像的表示。为了评测模型的效果，我们在知识图谱补全和三元组分类等任务上进行了评测，同时还探索了基于图像的表示在图像-知识空间的语义平移现象，以及注意力机制自动选择高质量图像的效果。概括起来，本章工作的主要创新点有以下几点：

- 我们创新性地在知识表示学习任务中引入图像信息，并提出一个全新的融合实体图像的知识表示学习模型，建立了跨模态的知识表示。
- 我们在真实数据集上对模型进行了验证，在知识图谱补全和三元组分类等任务上均获得了不错的效果。
- 我们探索了跨模态知识表示的语义平移现象，并基于实例研究了注意力机制对模型效果的影响。

4.2 相关工作

图像是人类信息交互的重要媒介，而图像相关领域一直吸引着广大研究者。在图像分类领域，LeCun等研究者提出基于卷积神经网络的LeNet模型，与传统方法相比，效果得到了大幅提升^[46]。随着ImageNet数据集的建立与深度学习的崛起，图像分类相关工作开始爆炸性地增长，其中具有代表性的就是Krizhevsky等研究者的工作^[33]。他们的模型使用了多个卷积层和全连接层对图像建立表示，同时使用了ReLU激活函数作为非线性层。随后，GoogLeNet^[47]等工作在模型上提出了进

一步改进，提高了图像分类的结果。在融合图像信息进行知识表示学习时，我们将参考图像分类上的已有工作成果，构建实体图像的特征向量。

在图像领域也有很多跨模态的任务，其中具有代表性的是图像与文本融合的任务。图像标题生成任务（Image caption generation）是经典的图像和文本空间结合的任务，旨在对给定图像生成对应标题。Vinyals^[48]和Xu^[49]等研究者针对图像标题生成任务，构建了基于神经网络和注意力机制的模型。图像问答系统任务（Visual question answering）给定图像和与图像相关的问题，希望模型能够在图像中发现问题的答案并生成回答^[50]。除此之外，Shutova等研究者^[51]联合图像信息发现文字中的隐喻现象，Kiros等研究者^[52]尝试解决图像-句子排序任务，也尝试了图像与文本信息的结合。尽管图像与文本融合的任务正在引起越来越多的研究兴趣，但图像与知识结合，特别是融合图像信息的知识表示学习还少有人探究。我们将以这些已有的图像文本联合学习模型为参考，设计新的融合图像信息的知识表示学习模型。

4.3 算法模型

融合实体图像信息的知识表示学习旨在充分利用实体图像的视觉信息，提高知识表示的性能，建立知识的跨模态表示。我们使用了图像编码器建立实体图像 的表示，然后引入注意力机制自动选择高质量的实体图像构建实体基于图像的表示，最后在平移模型的框架下进行联合训练。模型流程图如图4.2所示：

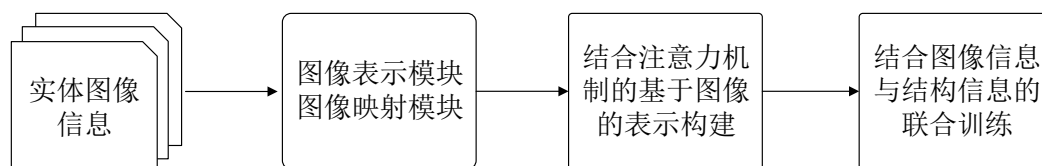


图 4.2 融合实体图像的知识表示学习模型流程图

4.3.1 符号体系与重要概念

我们首先介绍本章所用的符号体系。在知识图谱方面，我们使用 (h, r, t) 表示三元组，其中 h 代表头实体， t 代表尾实体， r 代表关系。我们使用 E 表示实体集合， R 表示关系集合， T 表示三元组集合，有 $h, t \in E$ ， $r \in R$ 以及 $(h, r, t) \in T$ 。在图像方面，每个实体通常对应多个实体图像，第 k 个实体的图像集表示如下： $I_k = \{img_1^{(k)}, img_2^{(k)}, \dots, img_n^{(k)}\}$ ，其中 $img_i^{(k)}$ 表示第 k 个实体的第 i 个图像。

参考融合实体描述的知识表示学习模型，我们为每个实体构建两种表示向量。第一种是基于结构的表示，使用 \mathbf{h}_S 和 \mathbf{t}_S 表示头实体和尾实体基于结构的表示。这种分布式向量基于平移假设，完全由知识图谱三元组的结构信息决定。第二种是基于图像的表示，使用 \mathbf{h}_I 和 \mathbf{t}_I 分别表示头实体和尾实体基于图像的表示。这种表示向量结合三元组的结构信息与实体图像的视觉信息构建得来。

4.3.2 模型整体架构与学习框架

融合实体图像的知识表示模型整体示意图如图4.3。如图所示，对于一个三元组，IKRL模型以其头实体和尾实体的所有实体图像作为输入，经过实体图像编码器得到各个图像在知识空间的表示。由于一个实体往往有多张图像，我们使用注意力机制对各图像表示进行整合，得到了实体基于图像的表达。最后，基于结构和基于图像的表达被统一在平移模型的框架下进行学习，并通过反向传播更新知识向量和实体图像编码器的参数。

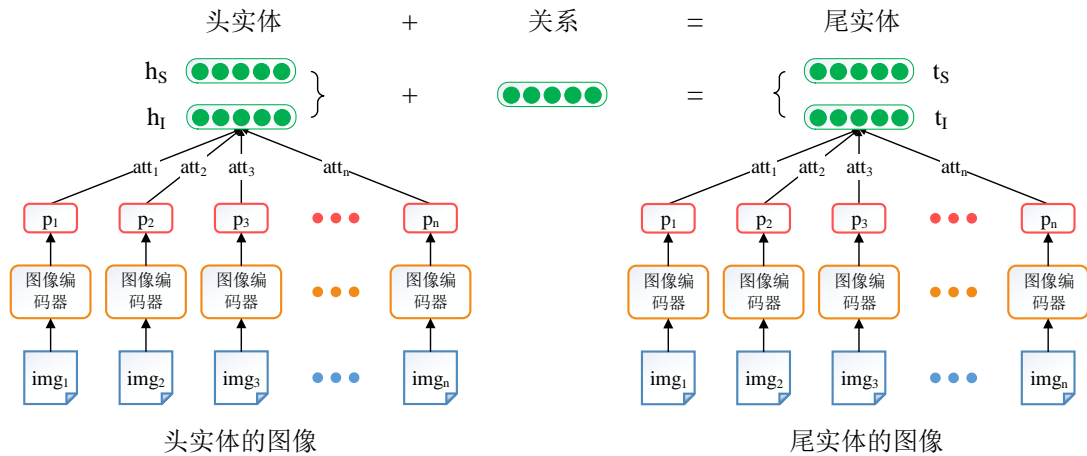


图 4.3 融合实体图像的知识表示学习模型示意图

IKRL模型基于平移模型的框架，即认为三元组内的实体与关系向量之间应该具有 $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$ 的关系。由于在IKRL模型中，每个实体都拥有基于结构和基于图像的两重表示向量，我们将改进后的能量函数形式化如下：

$$E(h, r, t) = \alpha_1 \|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_S\| + \alpha_2 \|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_I\| + \alpha_3 \|\mathbf{h}_I + \mathbf{r} - \mathbf{t}_S\| + \alpha_4 \|\mathbf{h}_I + \mathbf{r} - \mathbf{t}_I\|. \quad (4-1)$$

其中 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是控制各项权值的超参数。能量函数由实体两种表示进行组合，一共分为四个部分： $\|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_S\|$ 部分即为平移模型的能量函数部分， $\|\mathbf{h}_I + \mathbf{r} - \mathbf{t}_I\|$ 部分仅参考基于图像的表示，而 $\|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_I\|$ 和 $\|\mathbf{h}_I + \mathbf{r} - \mathbf{t}_S\|$ 则综合考虑了两种实体表

示。通过以上能量函数混合项的限制，IKRL模型将实体的两种表示学习到了同一语义向量空间中，并共享关系向量。

IKRL模型在训练时依旧使用最大间隔方法，评分函数如下：

$$L = \sum_{(h,r,t) \in T} \sum_{(h',r',t') \in T'} \max(\gamma + E(h,r,t) - E(h',r',t'), 0). \quad (4-2)$$

其中 T' 是负例三元组集合， $(h',r',t') \in T'$ 是负例三元组， γ 是表示正负例三元组得分间隔距离的超参数。由于我们为每个实体设计了两种知识表示，此处三元组能量函数 $E(h,r,t)$ 即如前述，包含了两种实体表示的四种混合组合项。在构建负例三元组集合时，我们也参考前述方法，对所有正例三元组随机替换掉其中任一实体或关系，并保证替换后的负例三元组不存在于正例之中。形式化地，我们构造知识图谱的负例三元组集合如下：

$$T' = \{(h',r,t)|h' \in E\} \cup \{(h,r,t')|t' \in E\} \cup \{(h,r',t)|r' \in R\}, \quad (h,r,t) \in T. \quad (4-3)$$

并保证：

$$\forall (h',r',t') \in T', \quad (h',r',t') \notin T. \quad (4-4)$$

我们设计了一种实体图像编码器抽取实体图像中的视觉信息。由于图像空间与知识空间是异质的，我们首先使用经典的神经网络模型抽取图像特征，然后通过映射矩阵构建图像特征在知识空间的表示。我们在下节将详细介绍实体图像编码器的构造与具体实现。

4.3.3 实体图像编码器

实体图像编码器旨在从实体图像中抽取图像关键特征信息，并形成实体图像在知识空间的表示。具体地，实体图像编码器以单个实体图像作为输入，使用基于神经网络的图像表示模块抽取蕴含实体视觉信息的图像特征。由于知识与图像的异质性，我们在图像表示模块后加入了一个图像映射模块，将图像特征从图像空间映射到知识空间，最后形成实体图像在知识空间中的表示。图4.4给出了实体图像编码器的流程示例图。

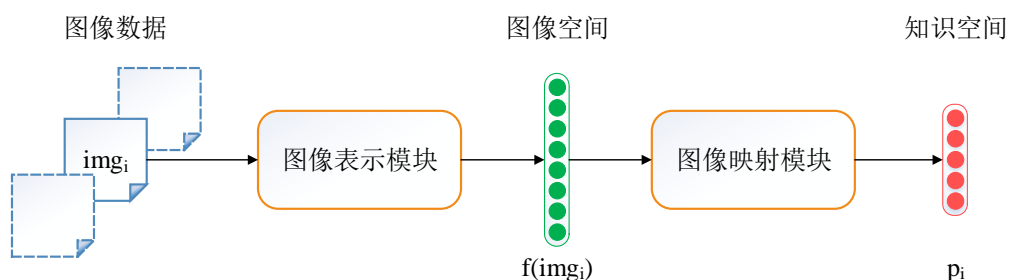
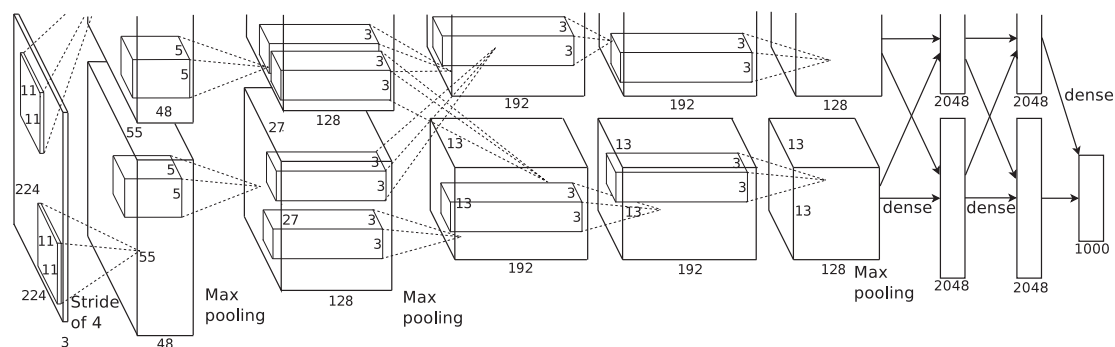


图 4.4 实体图像编码器示例图

4.3.3.1 图像表示模块

图像表示模块负责从图像中抽取实体蕴含的关键视觉信息，建立实体图像的特征向量。现在在图像分类等任务中，基于神经网络的图像特征抽取技术已经十分成熟，其中比较经典的是Krizhevsky等研究者的AlexNet模型^[33]。

我们在实验中直接使用AlexNet模型抽取图像特征。图4.5给出了AlexNet模型的简单示意图，可以看出AlexNet模型由五个卷积层和三个全连接层组成。首先，输入图像经过一定预处理（如镜面复制等），并从图像中心和四个角处抽取224×224大小的像素区域，作为AlexNet模型的输入。然后，输入像素矩阵经过五层卷积层和两层全连接层，得到一个4,096维的向量表示，这个向量表示可以被看做是输入图像经神经网络压缩后的特征表示。模型的非线性层使用ReLU函数，也加入了dropout策略防止模型的过拟合。最后，这个4,096维的向量经过一个softmax层得到1,000维的输出对应分类结果。在实体图像编码器中，我们使用AlexNet模型中第二个全连接层后的4,096维输出向量作为实体图像的特征向量。对于图像输入i，我们使用 $f(img_i)$ 表示实体图像的特征向量。

图 4.5 AlexNet模型示意图^[33]

4.3.3.2 图像映射模块

我们使用了AlexNet模型从实体图像中抽取图像特征，接下来就需要将这部分

图像特征向量从图像空间转移到知识空间。为此，我们构造了一个图像映射模块，使用映射矩阵构建图像在知识空间中的表示。我们有：

$$\mathbf{p}_i = \mathbf{M} \cdot f(\text{img}_i). \quad (4-5)$$

其中 $\mathbf{M} \in \mathbb{R}^{d_i \times d_s}$ 是映射矩阵， d_i 表示图像特征向量的维数，而 d_s 表示知识空间向量的维数。实体图像 img_i 经过图像表示模块得到图像特征向量 $f(\text{img}_i)$ ，然后再经过图像映射模块得到实体图像在知识空间中的表示 \mathbf{p}_i 。

4.3.4 多实例学习中的注意力机制

实体在数据库中往往拥有成千上万的实体图像，这些图像的信息量与质量，描写实体的粒度与角度，以及所处的背景都不尽相同。虽然简单地将这些信息平均化考虑亦能获得一定量的信息，但是相应地也会给实体表示引入大量的噪声。为了解决实体图像多实例学习中的问题，我们引入了注意力机制，使得模型能够自动地选择高质量的实体图像组成更准确的实体基于图像表示。

注意力机制是近年来深度学习领域的突破性进展之一，被广泛运用在图像分类^[53]、文本摘要^[54]、机器翻译^[55]和语音识别^[56]等任务上，显著提升了各任务的效果。在实体图像的多实例学习中，我们联合图像在知识空间的表示和实体基于结构的表示，构建了实例层次上的注意力模型。具体地，对于第 k 个实体 $e^{(k)}$ 的第 i 个图像向量表示 $\mathbf{p}_i^{(k)}$ ，实例层次的注意力公式被定义为：

$$\text{att}(\mathbf{p}_i^{(k)}, \mathbf{e}_S^{(k)}) = \frac{\exp(\mathbf{p}_i^{(k)} \cdot \mathbf{e}_S^{(k)})}{\sum_{j=1}^n \exp(\mathbf{p}_j^{(k)} \cdot \mathbf{e}_S^{(k)})}. \quad (4-6)$$

其中 $\mathbf{e}_S^{(k)}$ 表示 $e^{(k)}$ 基于结构的表示。随后，我们根据各实体图像所得的注意力，在联合构建实体基于图像表示时为各个图像实例赋上权值。实体基于图像表示 $\mathbf{e}_I^{(k)}$ 由以下公式得到：

$$\mathbf{e}_I^{(k)} = \sum_{i=1}^n \frac{\text{att}(\mathbf{p}_i^{(k)}, \mathbf{e}_S^{(k)}) \cdot \mathbf{p}_i^{(k)}}{\sum_{j=1}^n \text{att}(\mathbf{p}_j^{(k)}, \mathbf{e}_S^{(k)})}. \quad (4-7)$$

综合以上两个公式我们可以看出，当实体图像在知识空间的向量与其对应实体的基于结构的表示越相似，注意力的结果就越高，这个实体图像在最终构建实体基于图像表示时影响就会越大。这样自动选择实体图像的策略符合我们模型的假

设与期望。IKRL模型假设实体存在基于结构和基于图像的两种知识表示，并且希望这两种表示能够尽可能一致。这一点在模型能量函数的两种实体表示的混合项 $\|\mathbf{h}_S + \mathbf{r} - \mathbf{t}_D\|$ 和 $\|\mathbf{h}_D + \mathbf{r} - \mathbf{t}_S\|$ 中得到了体现。

为了充分体现注意力机制的效果，我们还实现了另外两种简单的多实例学习策略。第一种是均值方法，使用所有实体图像在知识空间中的向量均值作为此实体基于图像表示，使用AVG进行表示。第二种是最大值方法，是注意力模型的简化版，仅选择注意力值最大的实体图像在知识空间中的向量作为此实体基于图像表示，使用MAX进行表示。我们将会在实践中详细分析这些策略之间的表现差异与可能原因。

4.3.5 模型优化与具体实现

融合实体图像的知识表示学习模型在平移模型的框架下学习知识表示，在训练中需要学习以下参数组： $\theta = (\mathbf{E}, \mathbf{R}, \mathbf{W}, \mathbf{M})$ ，其中 \mathbf{E} 代表实体基于结构的表示向量组， \mathbf{R} 代表关系向量组， \mathbf{W} 代表图像表示模块中神经网络的参数，而 \mathbf{M} 代表图像映射模块中的映射矩阵参数。IKRL模型使用Mini-batch随机梯度下降的方法进行训练与参数更新。

在初始化上， \mathbf{E} 和 \mathbf{R} 可以随机初始化或者由平移模型进行预训练得到，映射矩阵 \mathbf{M} 由随机初始化得到。对于图像表示模块中的AlexNet模型，我们直接使用了经典深度学习工具Caffe^[57]中的实现。Caffe中的AlexNet模型实现在Krizhevsky^[33]原始论文的基础上做了小改进，并在ILSVRC 2012 图像识别数据集上进行了训练。由于预训练后AlexNet的模型参数具有图像识别上的物理意义，所以我们在联合训练时将图像表示模块部分的参数进行固定。最后，为了提高模型计算效率，我们在图像表示模块使用了GPU加速图像特征的提取，并实现了一个基于多线程的联合训练模型版本。

4.4 实验设计与结果分析

在本章节，为了测试融合实体图像的知识表示学习模型，我们在知识图谱补全和三元组分类等知识表示学习经典任务上进行了测试，同时在样例分析中详细说明了实体基于图像的表示的有效性和注意力模型的提升效果。实验结果证明了实体图像信息蕴含的丰富视觉信息能够很好地辅助构建更准确的知识表示，同时也证明了我们的IKRL模型能够充分利用实体图像信息，暗示着跨模态知识表示在实际运用中的可行性。

4.4.1 实验数据集

我们将实体图像信息与知识图谱三元组信息联合进行学习，在实验中需要使用两方面的数据。然而现在并没有构建好的包含丰富实体图像的知识图谱测试集。因此，我们构建了一个包含三元组信息与实体图像信息的WN9-IMG数据集。WN9-IMG数据集基于经典的知识图谱数据集WN18^[9]和图像数据集ImageNet^[44]构建而成，其中WN18数据集从WordNet^[5]中抽取得来。ImageNet是在图像识别领域被广泛使用的标准图像数据集，并且其图像所属的实体与WordNet数据集中实体相互对应。为了保证构建数据集的三元组和图像质量，我们选择从ImageNet中获取实体图像，并从WN18数据集中选择头实体和尾实体均在ImageNet中有实体图像的三元组，构建WN9-IMG跨模态知识图谱数据集。为了平衡实体图像的多样性与模型效率，我们对每个实体在ImageNet中抽取了最多10张实体图像。我们将三元组按照固定比例随机分为训练集、验证集与测试集，并保证测试集和验证集中的所有实体和关系均在训练集中出现过。最终，WN9-IMG中包含了6,555个实体与63,225张图像。数据集的具体统计数据详见表4.1。

表 4.1 WN9-IMG的统计数据

数据集	#关系	#实体	#图像	#训练集	#验证集	#测试集
WN9-IMG	9	6,555	63,225	11,741	1,337	1,319

4.4.2 实验与模型参数设置

融合实体图像的知识表示学习模型基于平移模型的学习框架，也能运用于其它基于平移假设的改进模型。为了对模型效果进行充分细致的分析，我们实现了平移模型（TransE）^[17]以及TransR模型^[22]作为基线模型。为保证公平，基线模型均采用任务中的最佳参数进行训练，并保证所有模型的知识向量维数一致。

融合实体图像的知识表示学习模型采用随机梯度下降算法，基于最大间隔方法进行优化。我们对最大间隔参数 γ 在 $\{1.0, 2.0, 4.0\}$ 中进行了尝试，其中 $\gamma = 4.0$ 时模型具有最佳结果。对于模型学习率的设定，我们尝试了固定学习率 $\lambda = \{0.0002, 0.0005, 0.001\}$ ，还设计了一种随迭代轮次上升自适应下降的学习率策略。最终发现，当模型学习率使用随迭代轮次变化线性下降的策略， λ 从0.001下降到0.0002时，IKRL模型能得到最好结果。我们也探索了能量函数中两种实体表示的四种组合不同的权值比例 α_i ，最后使用了 $\alpha_1 = 1.0$ ， $\alpha_2 = 0.5$ ， $\alpha_3 = 0.5$ ， $\alpha_4 = 1.0$ 的比例。如前所述，每个实体最多拥有10个实体图像，每个实体图像的

图像特征维数为 $d_i = 4,096$ ，所有模型的知识表示维数统一为 $d_s = 50$ 。

4.4.3 知识图谱补全结果与分析

知识图谱补全是知识表示学习领域的经典评测任务，要求模型在给定三元组中某一实体或关系丢失的情况下，通过剩下的实体和关系以及模型学习到的知识表示，正确预测出丢失的实体或关系。知识图谱补全根据预测对象的不同，常被分为实体预测与关系预测两个任务。

4.4.3.1 实验设计

我们使用融合实体图像的知识表示模型和基线模型在训练集上学到的实体与关系向量，对测试集中的三元组进行知识图谱补全。由于WN9-IMG测试集关系较少，所以我们在实验中主要关注实体预测的结果。对于测试集中的每个测试样例，我们隐藏其头实体和尾实体来进行实体预测。预测时，我们直接根据各模型基于平移假设的能量函数，为每一个候选三元组进行评分与排序，最终得到预测结果。由于IKRL模型为每个实体构建了基于结构和基于图像的两种实体表示向量，在测试时，我们根据使用实体表示的不同报告了三种不同设定下的结果：IKRL(SBR)模型代表测试时仅使用基于结构的表示进行预测，IKRL(IBR)模型代表测试时仅使用基于图像的表示进行预测，而IKRL(UNION)模型则代表测试时使用简单的加权策略，联合使用两种实体表示进行预测。

我们参考Bordes论文中的评测方法，使用Mean Rank和Hits@10作为评测指标^[17]。其中，Mean Rank表示所有测试样例中正确结果的平均排序，旨在对知识表示的整体效果进行评估；而Hits@10表示排名前10的结果中有正确答案的概率，旨在说明模型在知识图谱补全中的实用性。同时，我们也仿照Bordes的设定，使用了“Raw”和“Filter”两个指标对知识图谱补全任务进行更精确的评测。

4.4.3.2 实体预测结果与分析

实体预测的结果详见表4.2，从结果中我们可以得到以下结论：

(1) 所有的IKRL模型在全部评测指标上实体预测结果都超过了基线模型，其中IKRL(UNION)模型得到了最好的结果。这说明了实体图像中的丰富视觉信息能够帮助我们更深入地理解实体，同时也说明我们的模型能够成功地将这些信息编码进实体表示中，提升实体预测的性能。

(2) IKRL(SBR)模型和IKRL(IBR)模型在所有评测指标上都超过了基线模型。IKRL(IBR)模型基于实体图像信息直接构建实体的表示向量，融合了实体视觉信

息与实体在知识图谱中的结构信息，因此表示效果与基线模型相比得到了提升。而IKRL(SBR)模型虽然使用的是基于结构的实体表示，但是这些表示向量通过能量函数中两种表示的混合项训练，也间接学习到了一部分图像信息，从而使其表示能力得到提高。

(3) 所有的IKRL模型在Mean Rank评测指标上都得到了显著的提升。这是因为Mean Rank评测指标关注实体表示在向量空间中的整体效果，对于错误样例比Hits@10指标更加敏感。传统基于平移假设的模型仅考虑了知识图谱的结构信息，当所预测的信息丢失时很容易错得比较离谱。而我们的模型通过融合实体图像信息，能够间接地发现知识图谱中没有直接提起的实体潜在关系，从而使得知识表示的整体效果得到提高。

(4) 需要进一步说明的是，IKRL模型是基于平移模型的框架进行训练的，而TransR模型则通过构建关系特化的映射矩阵改进了平移模型。在这种情况下，IKRL模型在实体预测上的效果仍优于TransR模型，更说明了模型的优势与鲁棒性。而融合实体图像的思想也能较容易地运用于基于平移模型的改进模型上。

表 4.2 WN9-IMG上实体预测的评测结果

评测指标	Mean Rank		Hits@10(%)	
	Raw	Filter	Raw	Filter
TransE	143	137	79.9	91.2
TransR	147	140	80.1	91.7
IKRL(SBR)	41	34	81.1	92.9
IKRL(IBR)	29	22	80.2	93.3
IKRL(UNION)	28	21	80.9	93.8

4.4.3.3 注意力机制的影响与分析

为了更深入地展示注意力机制对模型效果的影响，我们还具体分析了在构建基于图像表示时三种不同策略的影响。我们使用IKRL(ATT)表示使用注意力机制加权图像表示的策略（即表4.2中给出的模型），使用IKRL(MAX)表示仅考虑注意力最大的图像表示的策略，使用IKRL(AVG)表示使用所有图像表示均值的策略。为了更全面地进行比较，我们展示了在以上三种策略下基于结构的表示和基于图像表示在实体预测任务上的评测结果。

表4.3展示了三种构建实体基于图像表示的不同策略的实验结果，从结果中我们可以得到以下结论：

表 4.3 WN9-IMG上各组合策略的实体预测结果

表示类型	基于图像的表达				基于结构的表示			
	Mean Rank		Hits@10(%)		Mean Rank		Hits@10(%)	
评测方法	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
IKRL(MAX)	59	52	79.8	92.1	62	55	81.0	92.3
IKRL(AVG)	29	22	79.3	92.9	43	36	80.7	92.8
IKRL(ATT)	29	22	80.2	93.3	41	34	81.1	92.9

(1) IKRL(ATT)模型的两种表示在所有评测指标上都得到了最好的结果。这是由于注意力机制能够智能地从复数实体图像中选择高质量的图像样例，构建实体基于图像的表达，在充分利用实体图像多样性的同时，保证了实体表示尽可能少地受到低质量图像的噪声影响。

(2) 尽管使用了简单的IKRL(MAX)和IKRL(AVG)策略，融合实体图像信息的知识表示模型仍然能超过表4.2中基线模型的结果。这进一步说明了实体图像信息对于知识表示构建的重要性，以及IKRL模型编码图像信息的鲁棒性。

(3) IKRL(AVG)策略比IKRL(MAX)策略的整体结果更好，在Mean Rank指标上尤其如此。这是因为IKRL(AVG)策略综合考虑了所有的实体图像信息，虽然不可避免地会引入一些噪声，但是仍比仅考虑质量最高图像的IKRL(MAX)策略具有优势。这个结果一部分也是由于我们的实体图像是从ImageNet抽取得来，使得图像整体质量较高的缘故。

(4) 通过比较我们可以发现，IKRL(ATT)策略和IKRL(AVG)策略相比似乎优势并不特别明显。这是由于我们构建评测数据集时，为了保证实体图像的质量，选择从ImageNet中直接抽取实体图像的缘故。WN9-IMG数据集中的图像质量整体较高，一定程度上导致了基于注意力机制的策略优势不明显，因为几乎所有图像都能提供实体相关的正确信息，彼此应得的注意力相当。为了充分展示注意力机制对模型的正面影响，我们在结果样例分析中进一步探索，给出了注意力机制从复数实体图像中成功分辨出高信息量图像的实例。

4.4.4 三元组分类结果与分析

三元组分类是检测知识表示质量的另一经典任务，要求模型能够根据所学得的知识表示判断给定三元组是正例或是负例。三元组分类任务也可以被看做是一个针对三元组的二分类任务。

4.4.4.1 实验设计

我们在WN9-IMG数据集上进行三元组分类的评测。由于知识图谱中不存在显式的负例三元组，我们需要自己构建负例三元组。我们沿用Socher等研究者的实验设定^[6]，随机替换测试集和验证集中正例三元组 (h, r, t) 的头实体或者尾实体，生成负例三元组 (h', r, t) 或者 (h, r, t') 。我们还保证负例三元组中的替换实体 h' 和 t' 符合对应关系的类别限制，保证负例三元组从实体类别的粗粒度上有成立的可能，以此增加三元组分类任务的挑战性。对于验证集和测试集中的每一个三元组，我们按照以上策略生成其对应的负例三元组，并保证正例和负例三元组样例数相等，使得正负例标签平衡。

在测试时，我们依旧为每一种关系设置一个关系特化的阈值 η_r 。以IKRL模型为例，当其三元组的评分函数 $\|\mathbf{h} + \mathbf{r} - \mathbf{t}\| > \eta_r$ 时，三元组被判断为负例；而当 $\|\mathbf{h} + \mathbf{r} - \mathbf{t}\| < \eta_r$ 时，三元组被判断为正例。其它模型也按照自己的评分函数进行分类判断。关系特化的阈值 η_r 通过在验证集上的分类效果获得。为了使比较充分，我们测试了IKRL(MAX)、IKRL(AVG)、IKRL(ATT)三种策略下的实验结果，在测试中仅使用了基于图像的知识表示。

4.4.4.2 实验结果与分析

表4.4中给出了IKRL模型和基线模型在三元组分类上的评测结果，从结果中我们可以发现：

(1) 所有IKRL模型都获得了比基线模型更好的效果，这从另一个角度重新证实了实体图像信息的重要性，也说明IKRL模型能够联合编码实体结构信息与视觉信息，具有较高的鲁棒性。

(2) 在三种组合策略中，IKRL(ATT)策略获得了最好的结果。这说明了注意力模型能够自动选取信息量更高的实体图像组成基于图像的实体表示，在保证图像质量的前提下充分利用实体图像的多样性，获得实体相关更完整的视觉信息。而IKRL(AVG)策略也获得了比IKRL(MAX)策略更好的效果，说明了综合考虑多个实体图像所带来的模型效果提升。

4.4.5 结果样例分析

为了充分展示并深入分析融合实体图像的知识表示学习模型效果，我们在量化评测的基础上还进行了细致的样例分析。在本节，我们将进行两个样例分析实验：(1) 图像-知识空间的语义规则探索；(2) 注意力机制的影响。为了展示的方便和明晰，我们在本节给出的图像样例对部分图像进行了截取操作，保留了图像

表 4.4 WN9-IMG上三元组分类的评测结果

模型	准确率(%)
TransE	95.0
TransR	95.3
IKRL (MAX)	96.3
IKRL (AVG)	96.6
IKRL (ATT)	96.9

的主要实体相关内容。

4.4.5.1 图像-知识空间的语义平移规则

Mikolov等研究者提出广为使用的词表示学习工具包Word2Vec时^[18-20]，曾发现了模型学到的词向量之间有趣的语义平移规则，例如：

$$\mathbf{v}(\text{king}) - \mathbf{v}(\text{man}) \simeq \mathbf{v}(\text{queen}) - \mathbf{v}(\text{woman}) \quad (4-8)$$

其中 $\mathbf{v}(x)$ 代表 x 的词向量表示。从上式中我们发现，词向量 $\mathbf{v}(\text{king})$ 和 $\mathbf{v}(\text{queen})$ 词义相近，而 $\mathbf{v}(\text{man})$ 和 $\mathbf{v}(\text{woman})$ 词义相近。在两组相近词向量之间的语义平移现象，似乎说明了词向量之间具有某种隐藏的联系。Kiros等研究者随后也在图像-语义空间上发现了类似的语义平移现象^[52]。这种语义平移现象蕴含了分布式表示的特点，能够从另一个侧面说明模型学习到的向量表示的有效性。

受到以上工作的启发，我们尝试探索IKRL模型学到的跨模态知识表示是否也具有这种语义平移现象。具体地，我们使用实体基于图像表示进行了语义平移规则的探索。实验结果表明，在跨模态的图像-知识表示空间里，语义平移规则也是普遍存在的。图4.6展示了在图像-知识空间中的语义平移规则实例。

以图中的实例为例，设 $\text{IBR}(x)$ 表示实体 x 的基于图像表示，我们首先计算 $\text{IBR}(\text{柜子}) - \text{IBR}(\text{抽屉}) + \text{IBR}(\text{琴键})$ ，得到一个向量，然后以整个实体集为候选，根据实体与此向量的距离进行排序，发现最相近的实体向量是 $\text{IBR}(\text{钢琴})$ 。与之前的工作不同，由于我们在跨模态图像-知识空间进行的测试，实体之间的向量差 $\text{IBR}(\text{柜子}) - \text{IBR}(\text{抽屉})$ 、 $\text{IBR}(\text{钢琴}) - \text{IBR}(\text{琴键})$ 与真实的关系属于成功对应，将语义平移规则中隐藏的联系显式地表现出来，展示了IKRL模型的有效性。



图 4.6 图像-知识空间的语义平移规则示例

4.4.5.2 注意力机制的影响

为了深入探究注意力机制对模型的影响，我们分析了具有高注意力值和低注意力值的实体图像的特点。实验结果显示，整体来说，具有高注意力值的实体图像确实蕴含着更丰富、更准确的实体信息。图4.7给出了三组实体以及它们具有最高和最低注意力值的实体图像。



图 4.7 注意力机制对选择实体图像的影响示例

对于第一组实体手提电脑，高注意力值的实体图像对应着真实的手提电脑，而低注意力值的实体图像却是一个横着摆放的手机，实际上给出了错误的实体。对于第二组实体高尔夫，高注意力值的实体图像显示一个人在打高尔夫的场面，而低注意力值的实体图像显示的是高尔夫球场的远景，对实体本身的信息量和准确度均不如高注意力值的实体图像。对于第三组实体水壶，高注意力值的实体图像明确地给出了水壶的图片，而低注意力值的实体图像过于聚焦在了水壶的花洒部分，反而难以对水壶的整体进行全局的认知。这些实例证明了注意力机制确实

能够智能地选择质量更高的实体图像，从而使得基于图像表示更加准确，甚至暗示了我们的模型能够帮助图像数据库自动推荐高质量实体图像的可能。

4.5 本章小结

在本章中，我们提出了融合实体图像信息和知识图谱结构化信息的IKRL模型，构建知识的跨模态表示。特别地，模型为每个实体设置了基于结构的表示和基于图像的表示，在平移模型的学习框架进行联合训练。我们使用基于深度神经网络的图像表示模块抽取实体图像的特征，然后将图像特征映射到知识空间中，并针对实体图像的多样例学习提出基于注意力的模型，最终构建实体基于图像表示。模型在知识图谱补全和三元组分类等任务上进行了测试，实验结果验证了实体图像蕴含的丰富视觉信息的重要性，也说明我们的模型能够较好地从中实例实体图像中智能准确地抽取实体的相关知识以构建实体表示。我们还结合实例分析，探索了图像-知识空间的语义平移现象和注意力机制对模型的影响。

第5章 总结与展望

5.1 工作总结

知识图谱是将人类已有知识高度结构化形成的知识系统，凝结了人类千百年积累的知识与智慧。知识图谱常被用于信息检索、问答系统和智能对话等知识驱动的人工智能应用，辅助知识抽取、存储与推理，具有重要的实用价值与研究意义。随着互联网时代信息的爆炸性增长，如何对知识图谱中存储的知识进行更好地编码与表示，构建知识图谱到知识驱动的应用之间的桥梁，成为当下热门的研究课题。为了解决计算效率与数据稀疏等问题，知识表示学习应运而生，它基于分布式表示的思想，将实体和关系的语义信息在低维向量空间中进行表示，显著提高了知识表示的灵活性和性能。现在，知识的分布式表示已被广泛运用于关系抽取^[58]、语言模型^[59]和问答系统^[60]等知识驱动的任务中。

我们身处于复杂多元的世界，每时每刻都需要与各种多源信息，如文本、图像和结构化信息等多模态信息进行交互。这些多源信息既是我们不可或缺的知识来源，亦是我们反馈自身信息的对象与媒介。然而，传统的知识表示学习模型往往重点关注知识图谱自身的高度结构化信息，忽略了丰富的跨模态多源信息。在本文中，我们主要关注融合多源信息的知识表示学习，其中重点关注以下三个任务：（1）融合实体描述信息的知识表示学习；（2）融合实体层次类型信息的知识表示学习；（3）融合实体图像信息的知识表示学习。

在第二章，我们重点关注融合实体描述信息的知识表示学习。实体描述是对实体自身凝练的文本描述，其中蕴含着实体各方面的丰富细节信息。这些文本信息能够作为知识图谱结构化信息的补充，帮助构建更准确的知识表示。然而，联合考虑实体描述面临着如何自动抽取描述中高质量的文本信息，以及如何融合结构信息与文本信息进行联合学习等挑战，已有引入文本信息的知识表示模型也仅仅孤立地考虑词级别的文本信息，而忽略了篇章级别语序语义信息的影响。为了解决这些问题，我们基于平移假设提出了融合实体描述的知识表示学习模型，为每个实体设置了基于结构与基于描述的两种表示，并使用神经网络模型对实体描述进行建模。实验结果表明，我们的模型能够充分利用实体描述中的文本信息，提升知识图谱补全和实体类型分类等任务的效果，在对新实体的知识表示上也有较好的表现。

在第三章，我们重点关注融合实体层次类型信息的知识表示学习。实体层次类型指的是实体所属不同粒度的类型信息，这些类型往往储存在层次化的结构中。

实体类型信息能够帮助人类构建层次化的认知体系，提供实体结构化的先验知识，也能暗示实体在不同情境下更应表现出的类型。然而，传统知识表示模型较少考虑实体类型信息，也未能充分利用实体类型的层次结构信息。为了解决这些问题，我们设计了融合实体层次类型的知识表示学习模型，提出实体在不同关系下应该突出不同实体类型，并具有类型特化的实体表示。我们使用两种层次类型编码器对实体类型的层次结构进行建模，构建映射矩阵，同时在训练与测试中进行了实体类型限制，进一步提高知识表示的性能。实验结果表明，我们的模型能够充分利用实体层次结构的类型信息，提升知识图谱补全和三元组分类等任务的效果，并在长尾分布的数据集上也能得到超过基线模型的表现。

在第四章，我们重点关注融合实体图像信息的知识表示学习。实体图像能够提供实体外形、行为和其它相关实体的视觉信息，可以帮助模型全方位地理解实体。融合图像信息与知识图谱结构信息构建跨模态的知识表示，既能通过图像信息提高知识表示的性能，也能将知识引入图像领域，为图像与知识的联合应用提供基础。由于图像信息与知识信息在储存与表示上存在较大差异，如何联合两个异质空间进行学习成为此任务的最大的挑战。另外，海量的图像信息质量良莠不齐，选取的图像质量对知识表示结果也有较大影响。针对这些问题，我们提出了融合实体图像的知识表示学习模型，构建实体基于描述和基于图像的两种知识表示。具体地，我们使用基于神经网络模型的图像表示模块和图像映射模块构建图像特征在知识空间的表示，然后引入注意力机制自动选择高质量的实体图像构建实体基于图像的表达。实验结果证实了实体图像中的视觉信息能够帮助模型构建更准确的知识表示，在知识图谱补全和三元组分类等任务上都表现出更好的效果。经过实例分析，我们也证实了模型在图像-知识联合空间上的语义平移现象，以及注意力机制对模型效果的正面影响。

5.2 工作展望

在本文中，我们针对融合多源信息的知识表示学习任务进行了研究，在知识图谱的结构化信息之外，将实体描述的文本信息、实体层次类型的结构信息、以及实体图像的视觉信息引入知识表示学习中。尽管联合模型显著提升了在知识表示任务上的效果，但是本文的工作仍有可以改进的空间。未来，我们希望能够从以下几个方面进行改进：

(1) 在融合实体描述信息的知识表示学习中，我们可以尝试扩充文本信息的来源，综合考虑整个互联网语料信息，而不是仅仅局限于实体描述。同时，我们可以使用其它模型，如长短期记忆网络^[61,62]和注意力机制等方法，提升实体描述

编码器的效果。最后，我们还可以尝试将融合实体描述的知识表示学习模型框架扩展至其它知识表示学习的模型中，验证模型的鲁棒性与可拓展性。

(2) 在融合实体层次类型信息知识表示学习中，我们可以尝试改进现有的层次类型编码器，进一步增加模型的表达能力，使模型能够更细致地对实体类型的层次结构进行建模。同时，我们也可以让模型在其它更深层次的实体类型分类体系下进行测试，以观察模型的鲁棒性，并根据结果提出可能的改进。

(3) 在融合实体图像信息知识表示学习中，我们可以改进已有的图像表示模块和图像映射模块，并在联合训练时对其参数进行微调，以提升模型抽取实体视觉信息的能力。我们还可以在模型中增加知识空间对图像空间的反馈机制，研究如何将知识引入图像领域。在未来，我们还可以深入挖掘图像内部各实体之间的语义关系，更加充分地利用图像信息建立知识的跨模态表示。

(4) 在未来，我们将联合知识、文本、实体类型、图像等多源信息，在统一的语义向量空间中建立知识的跨模态表示，在综合使用各种多源信息加强知识表示性能的同时，也将知识引入文本、图像等多模态的领域，使跨模态的知识表示能够更好地服务实际的知识驱动应用。

参考文献

- [1] Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 2014, 57(10):78–85.
- [2] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge. *Proceedings of KDD*, 2008. 1247–1250.
- [3] Auer S, Bizer C, Kobilarov G, et al. Dbpedia: A nucleus for a web of open data. *The semantic web*, 2007. 722–735.
- [4] Hoffart J, Suchanek F M, Berberich K, et al. Yago2: A spatially and temporally enhanced knowledge base from wikipedia. *Artificial Intelligence*, 2013, 194:28–61.
- [5] Miller G A. Wordnet: a lexical database for english. *Communications of the ACM*, 1995..
- [6] Hinton G E. Learning distributed representations of concepts. *Proceedings of the eighth annual conference of the cognitive science society*, volume 1. Amherst, MA, 1986. 12.
- [7] Bordes A, Weston J, Collobert R, et al. Learning structured embeddings of knowledge bases. *Proceedings of AAAI*, 2011. 301–306.
- [8] Bordes A, Glorot X, Weston J, et al. Joint learning of words and meaning representations for open-text semantic parsing. *Proceedings of AISTATS*, 2012. 127–135.
- [9] Bordes A, Glorot X, Weston J, et al. A semantic matching energy function for learning with multi-relational data. *Machine Learning*, 2014, 94(2):233–259.
- [10] Sutskever I, Tenenbaum J B, Salakhutdinov R R. Modelling relational data using bayesian clustered tensor factorization. *Proceedings of NIPS*, 2009. 1821–1828.
- [11] Jenatton R, Roux N L, Bordes A, et al. A latent factor model for highly multi-relational data. *Proceedings of NIPS*, 2012. 3167–3175.
- [12] Yang B, Yih W t, He X, et al. Embedding entities and relations for learning and inference in knowledge bases. *Proceedings of ICLR*, 2015.
- [13] Nickel M, Tresp V, Kriegel H P. A three-way model for collective learning on multi-relational data. *Proceedings of ICML*, 2011. 809–816.
- [14] Nickel M, Tresp V, Kriegel H P. Factorizing yago: scalable machine learning for linked data. *Proceedings of WWW*, 2012. 271–280.
- [15] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. *Proceedings of AAAI*, 2016.
- [16] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion. *Proceedings of NIPS*, 2013. 926–934.
- [17] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data. *Proceedings of NIPS*, 2013. 2787–2795.
- [18] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. *Proceedings of NIPS*, 2013. 3111–3119.

-
- [19] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. Proceedings of ICLR, 2013.
- [20] Mikolov T, Yih W t, Zweig G. Linguistic regularities in continuous space word representations. Proceedings of HLT-NAACL, volume 13, 2013. 746–751.
- [21] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes. Proceedings of AAAI, 2014. 1112–1119.
- [22] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion. Proceedings of AAAI, 2015.
- [23] Ji G, He S, Xu L, et al. Knowledge graph embedding via dynamic mapping matrix. Proceedings of ACL, 2015. 687–696.
- [24] Lin Y, Liu Z, Luan H, et al. Modeling relation paths for representation learning of knowledge bases. Proceedings of EMNLP, 2015. 705–714.
- [25] Guo S, Wang Q, Wang L, et al. Jointly embedding knowledge graphs and logical rules. Proceedings of EMNLP, 2016. 1488–1498.
- [26] Xie R, Liu Z, Jia J, et al. Representation learning of knowledge graphs with entity descriptions. Proceedings of AAAI, 2016.
- [27] Lehmann J, Isele R, Jakob M, et al. Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web, 2015, 6(2):167–195.
- [28] Pellissier Tanon T, Vrandečić D, Schaffert S, et al. From freebase to wikidata: The great migration. Proceedings of WWW, 2016. 1419–1428.
- [29] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network. Proceedings of COLING, 2014. 2335–2344.
- [30] Wang Z, Zhang J, Feng J, et al. Knowledge graph and text jointly embedding. Proceedings of EMNLP, volume 14, 2014. 1591–1601.
- [31] Zhong H, Zhang J, Wang Z, et al. Aligning knowledge and text embeddings by entity descriptions. Proceedings of EMNLP, 2015. 267–272.
- [32] Zhang D, Yuan B, Wang D, et al. Joint semantic relevance learning with text data and graph knowledge. ACL-IJCNLP, 2015. 32–40.
- [33] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Proceedings of NIPS, 2012. 1097–1105.
- [34] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: A convolutional neural-network approach. IEEE transactions on neural networks, 1997, 8(1):98–113.
- [35] Kim Y. Convolutional neural networks for sentence classification. Proceedings of EMNLP, 2014.
- [36] Santos C, Xiang B, Zhou B. Classifying relations by ranking with convolutional neural networks. Proceedings of ACL, 2015. 626–634.
- [37] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch. JMLR, 2011, 12:2493–2537.
- [38] Neelakantan A, Chang M W. Inferring missing entity type instances for knowledge base completion: New dataset and methods. Proceedings of NAACL, 2015.

- [39] Xie R, Liu Z, Sun M. Representation learning of knowledge graphs with hierarchical types. Proceedings of IJCAI, 2016.
- [40] Fu R, Guo J, Qin B, et al. Learning semantic hierarchies via word embeddings. Proceedings of ACL, 2014. 1199–1209.
- [41] Hu Z, Huang P, Deng Y, et al. Entity hierarchy embedding. Proceedings of ACL, 2015. 1292–1300.
- [42] Krompaß D, Baier S, Tresp V. Type-constrained representation learning in knowledge graphs. Proceedings of ISWC. 2015: 640–655.
- [43] Krompaß D, Nickel M, Tresp V. Large-scale factorization of type-constrained multi-relational data. Proceedings of DSAA. IEEE, 2014. 18–24.
- [44] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database. Proceedings of CVPR. IEEE, 2009. 248–255.
- [45] Xie R, Liu Z, Luan H, et al. Image-embodied knowledge representation learning. arXiv preprint arXiv:1609.07028, 2016..
- [46] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11):2278–2324.
- [47] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Proceedings of CVPR, 2015. 1–9.
- [48] Vinyals O, Toshev A, Bengio S, et al. Show and tell: A neural image caption generator. Proceedings of CVPR, 2015. 3156–3164.
- [49] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention. Proceedings of ICML, volume 14, 2015. 77–81.
- [50] Antol S, Agrawal A, Lu J, et al. Vqa: Visual question answering. Proceedings of ICCV, 2015. 2425–2433.
- [51] Shutova E, Kiela D, Maillard J. Black holes and white rabbits: Metaphor identification with visual features. Proceedings of NAACL, 2016. 160–170.
- [52] Kiros R, Salakhutdinov R, Zemel R S. Unifying visual-semantic embeddings with multimodal neural language models. Proceedings of NIPS, 2014.
- [53] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention. Proceedings of NIPS, 2014. 2204–2212.
- [54] Rush A M, Chopra S, Weston J. A neural attention model for abstractive sentence summarization. Proceedings of EMNLP, 2015.
- [55] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. Proceedings of ICLR, 2015.
- [56] Chorowski J, Bahdanau D, Cho K, et al. End-to-end continuous speech recognition using attention-based recurrent nn: first results. Proceedings of NIPS, 2014.
- [57] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding. Proceedings of ACM MM, 2014. 675–678.
- [58] Han X, Liu Z, Sun M. Joint representation learning of text and knowledge for knowledge graph completion. arXiv preprint arXiv:1611.04125, 2016..

- [59] Ahn S, Choi H, Pärnamaa T, et al. A neural knowledge language model. arXiv preprint arXiv:1608.00318, 2016..
- [60] Yin J, Jiang X, Lu Z, et al. Neural generative question answering. Proceedings of IJCAI, 2016.
- [61] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation, 1997, 9(8):1735–1780.
- [62] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks. Proceedings of NIPS, 2014. 3104–3112.

致 谢

衷心感谢我的导师孙茂松教授对我的精心指导。他为我创造了良好的学术氛围、优越的科研环境和宝贵的学术交流机会，他的言传身教与谆谆教诲也将使我受益终生。

感谢我的副导师刘知远助理教授一直以来对我的悉心教导。他高屋建瓴的学术视野、严谨认真的治学态度与勤奋踏实的工作作风，都给我留下了深刻的印象。他在学习生活中给予我大量的关怀与帮助，在为人处事上也使我裨益良多，这些都将成为我终生受用。

感谢刘洋副教授在科研上对我的指导与帮助，他对学术研究的真知灼见与追求卓越的学术精神令我受益匪浅。

感谢清华大学计算机科学与技术系的老师们多年来的指导，感谢清华大学自然语言处理组的同学们在研究生活中给予我的交流、探讨和帮助，感谢我的朋友们给我的关心和陪伴。

最后感谢我的家人，你们是最坚实的堡垒和后盾，是我最温暖的心灵港湾，也是我前进的最大动力。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

1992年8月16日出生于湖南省长沙市。

2010年9月考入清华大学计算机科学与技术系计算机科学与技术专业，2014年7月本科毕业并获得工学学士学位。

2014年9月免试进入清华大学计算机科学与技术系攻读硕士学位至今。

发表的学术论文

- [1] Ruobing Xie, Zhiyuan Liu, Maosong Sun. Representation Learning of Knowledge Graphs with Hierarchical Types. The 25th International Joint Conference on Artificial Intelligence, 2016, 2965-2971. (CCF A类, EI 收录, 检索号: 20165103147091)
- [2] Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, Maosong Sun. Representation Learning of Knowledge Graphs with Entity Descriptions. The 30th AAAI Conference on Artificial Intelligence, 2016, 2659-2665. (CCF A 类, EI收录, 检索号: 20165203195410)
- [3] Ruobing Xie, Zhiyuan Liu, Huanbo Luan, Maosong Sun. Image-embodied Knowledge Representation Learning. In Press. (已被The 26th International Joint Conference on Artificial Intelligence录用. CCF A类, EI源刊)
- [4] Ruobing Xie, Xingchi Yuan, Zhiyuan Liu, Maosong Sun. Lexical Sememe Prediction via Word Embeddings and Matrix Factorization. In Press. (已被The 26th International Joint Conference on Artificial Intelligence 录用. CCF A类, EI源刊)
- [5] Hao Zhu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. Iterative Entity Alignment via Joint Knowledge Embeddings. In Press. (已被The 26th International Joint Conference on Artificial Intelligence录用. CCF A类, EI源刊)
- [6] Yilin Niu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. Improved Word Representation Learning with Sememes. In press. (已被The 55th Annual Meeting of the Association for Computational Linguistics录用. CCF A类, EI 源刊)
- [7] Stefan Heinrich, Cornelius Weber, Stefan Wermter, Ruobing Xie, Yankai Lin, Zhiyuan Liu. Crossmodal language grounding, learning, and teaching. Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches, in the 30th Conference on Neural Information Processing Systems, 2016. (EI收录, 检索号: 20170303259854)

- [8] 刘知远, 孙茂松, 林衍凯, 谢若冰. 知识表示学习研究进展. 计算机研究与发展, 2016, 53(2):247-261. (EI收录, 检索号: 20161102091785)