计算机研究与发展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(11): 2687−2696, 2025

# 注意力增强的语义融合知识图谱表示学习框架

郭文斌[1] 王 鑫[1] 李 钊[1] 陈子睿[1] 赵 军[2] 冯 锋[2] 杜 方[2]
[1](天津大学智能与计算学部 天津 300072)
[2](宁夏大学信息工程学院 银川 750021)
(wenff@tju.edu.cn)

# Attention-Enhanced Semantic Fusion Knowledge Graph Representation Learning Framework

Guo Wenbin[1], Wang Xin[1], Li Zhao[1], Chen Zirui[1], Zhao Jun[2], Feng Feng[2], and Du Fang[2]
[1](*College of Intelligence and Computing*, *Tianjin University*, *Tianjin* 300072)
[2](*School of Information Engineering*, *Ningxia University*, *Yinchuan* 750021)

**Abstract** Knowledge graphs often face the challenge of incompleteness, which can be alleviated by completing missing information through link prediction tasks. However, most knowledge graph completion work overly focuses on extracting embedding features without sufficiently considering the complex semantics contained in the predicted node neighborhood information, global feature information, and directional feature information, making it difficult to accurately predict the missing information. We propose a general representation learning semantic enhancement framework, ASFR, which utilizes an attention mechanism to extract local association information of the knowledge graph and structural features of the knowledge graph, and enhances existing knowledge graph representation learning models by incorporating positional information. By embedding these three types of additional knowledge graph information into the entity vectors of the knowledge graph, the quality of the knowledge graph representation vectors is improved. Comparative experiments are conducted using five different categories of classical methods, and the results indicate that this framework can effectively enhance the predictive capability of models, achieving an improvement of 6.89% on three public datasets.

**Key words** knowledge graph; representation learning; link prediction; knowledge graph embedding; knowledge graph completion

**摘 要** 当前知识图谱通常存在不完整性的挑战,可以通过链接预测任务对缺失信息进行补全来缓解这一问题.然而,大部分知识图谱补全方法过度关注对嵌入特征的提取,没有充分考虑预测节点邻域信息、全局特征信息和方向特征信息中所包含的复杂语义,难以准确预测出缺失的信息.提出一种通用的表示学习语义增强框架ASFR,利用注意力机制提取知识图谱局部关联信息、知识图谱结构特征,结合位置信息对现有的知识图谱表示学习模型进行增强.将3种知识图谱附加信息嵌入到知识图谱的实体向量中,提高知识图谱表示向量的质量.在5种不同类别的经典方法中进行对比实验,结果显示ASFR框架在3个公开数据集上性能的提升幅度为6.89%,能够有效增强模型的预测能力.

**关键词** 知识图谱;表示学习;链接预测;知识图谱嵌入;知识图谱补全

**收稿日期**:2024-08-13;**修回日期**:2025-02-27
**基金项目**:国家科学自然基金面上项目(62472311);宁夏回族自治区重点研发计划项目(2023BEG02067)
This work was supported by the General Program of the National Natural Science Foundation of China (62472311) and the Key Research and Development Program of Ningxia Hui Autonomous Region (2023BEG02067).
**通信作者**:王鑫(wangx@tju.edu.cn)

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202440669 CSTR: 32373.14.issn1000-1239.202440669

知识图谱作为一种高度结构化的信息表示模型,以图形化形式展示实体间丰富的关联,构建了一个包罗万象、层次分明的认知领域框架.然而,由于数据源的限制和人工标注的疏忽,当前的知识图谱存在不完整性问题,使知识图谱的广泛应用成为挑战.当前通过学习数据低维稠密表示来对知识图谱进行补全的方法已经成为解决知识图谱不完整性问题的主流,现存许多工作已经实现了高质量的知识图谱表示[1].

知识图谱表示学习方法将关系视作实体之间的连接桥梁,运用平移、旋转、同相和反射等转换方式来进行建模.一些最新的方法引入了注意力机制,使模型能够更专注于特定任务相关的内容,从而提高三元组内部特征交互的效果.然而,目前的研究很少将实体放入整个图谱中考虑,没有将邻域信息、全局信息和位置信息对实体表示的影响进行深入探索.

知识图谱邻域信息指的是与给定实体直接相关的其他实体和关系所蕴藏的信息,它们可以提供有关给定实体语义上下文的重要线索,发现实体间的隐藏关联和语义联系.全局信息可以提供对整个知识图谱的综合理解和把握,捕捉知识图谱在不同领域展现出特定的规则模式[2],从而确保推理过程中的一致性.自适应位置信息针对异构知识图谱隐藏的真实世界复杂关联和多样性,对不同类型的实体和关系进行显式区分,从而提高模型在异构知识图谱上推理的准确性.

由于当前知识图谱表示学习方法已经能够有效捕捉三元组内部实体和关系之间的交互特征,因此本文设计了利用注意力增强的语义融合知识图谱表示学习框架(attention-enhanced semantic fusion knowledge graph representation learning framework, ASFR)来关注邻域和全局信息的融合.如图1所示,ASFR利用不同的注意力机制分别对知识图谱的邻域特征和全局特征进行捕捉融合,增强实体嵌入的表达能力,同时构建自适应位置嵌入特征,将实体处于图谱不同位置所蕴含的语义信息进行显式表达.并将3种用于语义增强的特征嵌入融入到传统知识图谱表示学习模型的训练中,设计了统一的链接预测训练框架更新知识图谱的表示,提高嵌入向量的质量.本文的主要贡献有3点:

1) 设计了一种知识图谱表示学习框架ASFR,利用注意力机制注重捕捉邻域和全局信息并进行语义融合,从而提升实体嵌入的表达能力.

2) 引入自适应位置嵌入特征,有效地展现实体在不同图谱位置的语义信息,进一步丰富实体嵌入的语义表达.

3) 在5种不同类别的经典方法中进行对比实验,结果表明该框架能够有效增强模型的预测能力,在3个公开数据集上的提升幅度为6.89%.同时消融实验证明ASFR框架各个组成部分的有效性.

## 1 相关工作

本文的准备工作主要包括知识图谱表示学习相关的概念定义.

### 1.1 传统的知识图谱表示学习模型

传统的知识图谱表示学习模型假设存在一种关联性,即如果2个实体之间存在某种关系,则它们在向量空间中的表示应该通过某种运算相近.该类模型将关系看作为一种在高维空间中的平移运算符,通过计算头实体向量经过关系向量平移后与尾实体向量之间的欧氏距离来评估三元组的真实性.其中,TransE模型[3]将连续性假设引入知识图谱表示学习中,提出将关系向量视为从头实体向量到尾实体向量平移操作的思想,之后一系列的工作使用超平面[4]、黎曼球面[5]、关系映射空间、多重映射空间[6]、旋转空间[7]等不同维度空间,或采用反演、反射、旋转、同质[7]等运算方式,增强模型对复杂关系模式的理解,使关系的表示不再过于刚性.但在处理具有多重关系或非线性关系的知识图谱时,该类模型往往难以准确捕捉其深层的语义信息.

随着深度学习方法在自然语言处理和计算机视觉领域的快速发展,卷积神经网络(CNN)技术被引入知识图谱预测任务中.其主要目的在于利用CNN的局部感知和参数共享特性,捕获实体和关系之间的复杂模式,并对缺失的信息进行预测.ConvE将实体向量和关系向量转换为二维矩阵,通过拼接头实体和关系嵌入矩阵来进行特征聚合,使用交叉熵损失函数确定尾实体的正确类别[8].后续工作的优化改进集中在增强实体和关系的交互角度,即ConvKB将三元组语义特征向量组合成三阶矩阵作为输入[9],InteractE改进实体和关系特征组合与提取方法[10],AcrE采用了反转卷积核来替代传统的外部卷积核[11],

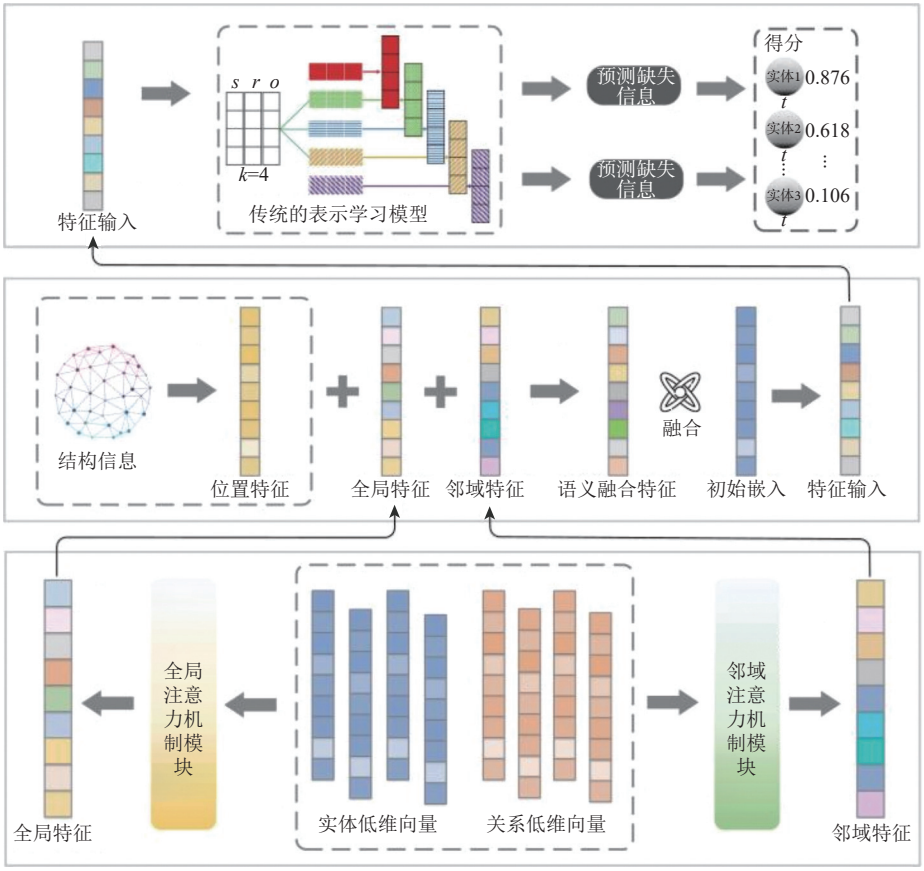



Fig. 1 Structure diagram of ASFR framework

图 1 ASFR 框架结构图

而 ConvR 使用关系嵌入充当卷积核，使卷积操作只围绕实体和关系的嵌入进行[12]. 实验证明，CNN 确实能够捕捉到知识图谱中复杂的关系语义和图谱层次结构，实现对知识图谱的深度特征提取.

注意力机制的引入为深度学习中的表示学习模型提供了一种灵活且高效的建模方式. Sheikh 等人提出了一种关系感知的掩码注意机制，运用关系和邻居信息来计算邻居的重要性，将这种信息传播到实体特征中，再进行嵌入表示的更新[13]；HRAN 在每个关系路径下聚合实体的邻居特征，更新不同关系路径的重要性[14]；KGT5 使用 Transformer 架构构建可扩展且通用的知识图谱表示学习模型[15]；HittER 使用 2 个不同的 Transformer，利用实体邻域信息实现对关系上下文和源实体本身信息的平衡[16]. 该类模型很好地与注意力机制进行了融合，但是并没有融入先前的表示学习工作之中.

当前知识图谱表示学习领域的工作在捕获知识图谱浅层和深层语义方面已经取得一定进展，但在与知识图谱邻域信息和全局信息有效融合方面尚未有令人满意的方法. 结合注意力机制的知识图谱表示学习方法在获取实体邻域信息和全局信息方面表现出良好的潜力，然而，与先前表现优异的方法相比，它们在提取三元组内部特征方面尚有不足之处. 因此本文设计了注意力增强的语义融合知识图谱表示学习框架来增强知识图谱的表示效果.

### 1.2 知识图谱表示学习模型框架

最近，一些研究提出了用一些框架增强知识图谱表示学习模型的性能，这些框架通常集成了先进的技术和方法，以克服传统模型在处理复杂知识图谱时的局限性，并提高模型在各种任务中的表现.

提出了一个规范化增强的已知事实感知框架 CEKFA[17]，该框架通过降低关系短语和三元组的稀疏性来提高链接预测性能. CAKE 从具有实体概念的事实三元组中自动提取常识，生成常识增强，实现高质量的负采样，是一个常识感知的知识嵌入框架[18]. R-GCN 引入关系图卷积网络应对真实知识库中高度多关系的数据[19]. PUDA 提出了一种数据增强策略，用于解决传统知识图谱表示学习模型中的假阴性和数据稀疏性问题[20]. REP 是一种后处理技术框架，它将预训练的 KG 嵌入与图上下文相适应[21].

与现有方法不同，本文旨在对知识图谱邻域信息和全局信息进行提取，同时对异构知识图谱中的位置信息进行显示构建，增强现有知识图谱表示学习模型的推理能力.

## 2 准备工作

介绍 ASFR 框架前，本文进行了一些准备工作，主要给出知识图谱和表示学习的相关概念定义.

**定义 1.** 知识图谱. 给定一组有限的实体集合$E$、关系集合$R$和可观察三元组集合$P$，一个知识图谱$\vartheta$由三元组$(E,R,P)$表示，三元组集合中的每个元素可以表示为$(h,r,t)$的形式，其中$h\in E$，$t\in E$，都为实体集合内的某个实体，$r\in R$是关系集合内的某个关系.

**定义 2.** 知识图谱表示学习. 对于知识图谱$(E,R,P)$，每个实体和关系都被赋予一个唯一的标识符或符号进行离散化表示，知识图谱表示学习方法将实体和关系映射到低维连续向量空间中，用向量$\boldsymbol{e}_h$、$\boldsymbol{e}_t$表示实体向量，向量$\boldsymbol{e}_r$表示关系向量.

**定义 3.** 知识图谱链接预测. 令$P_{\text{rea}}$表示真实三元组集合，其中$P\in P_{\text{rea}}$. 知识图谱链接预测任务旨在预测二元事实中缺失的元素，如通过相关算法预测知识图谱中缺失的实体，可能是头实体$(?,r,t)$，也可能是尾实体$(h,r,?)$，或者预测其他实体之间可能存在的未知关系$(h,?,t)$，?表示元组缺失的元素.

传统的知识图谱表示学习模型利用三元组作为训练数据，通过将三元组映射到高维空间中的向量表示，再对这些向量进行操作，以预测知识图谱中缺失的部分. 许多算法通过将向量表示转化为二维甚至三维的形式，增强特征之间的交互程度，从而实现更深层次的语义提取. 本文研究旨在通过引入新的框架，增强传统表示学习模型的能力. 我们在经典的表示学习模型基础上，增加了 3 种不同的特征表示向量或矩阵，以在高维向量层次上增加语义信息的交互，从而提升模型的表达能力[22-23].

通过在传统模型的基础上引入额外的特征表示向量或矩阵，本文方法能够在高维向量层次上增加语义信息的交互，从而捕捉到更丰富的特征信息. 这种增强的特征交互有助于提升模型在复杂语义关系中的表现. 这种多层次的语义提取有助于模型更好地理解和表示复杂的知识图谱. 由于引入了额外的特征表示向量或矩阵，模型的训练复杂度相应增加，需要更多的计算资源和时间来处理高维向量和多层次语义信息，但是使模型在处理复杂关系和预测缺失部分时更加准确和有效.

## 3 方 法

传统的知识图谱表示学习侧重于通过三元组内部的交互来提取特征，然而，这种方法在融合图谱整体复杂语义信息方面存在不足，同时尚未对知识图谱结构进行深入理解.

我们发现在进行实体和关系的嵌入特征提取与捕捉时，必须充分考虑它们的邻域特征以及不同位置之间的复杂语义关联. 从这一角度出发，我们利用邻域注意力机制来抽取局部结构中的语义信息，将周围图谱信息嵌入到实体向量中；同时采用全局注意力机制提取知识图谱整体结构特征，从而使得框架能够适用于不同领域和规模的图谱；此外，构建了自适应位置嵌入特征，用于标识实体在图谱中的不同位置，以增强模型的推理效果. 下面进行详细介绍.

### 3.1 知识图谱表示学习模型框架

1)邻域特征提取. 给定知识图谱中的三元组$(h,r,t)$，首先构造头部实体$h$的初始嵌入向量$\boldsymbol{e}_h^{\text{o}}\in\mathbb{R}^{d_h}$和关系$r$的初始嵌入向量$\boldsymbol{e}_r^{\text{o}}\in\mathbb{R}^{d_r}$. 传统的知识图谱嵌入模型直接使用初始向量$\boldsymbol{e}_h^{\text{o}}$进行特征交互，忽略了知识图谱整体和局部的结构信息. 如图 2 所示，ASFR 将复杂语义信息直接嵌入到初始向量中，得到蕴含图谱信息的嵌入向量$\boldsymbol{e}_h$. 针对图谱中的每个实体，设计邻域信息提取算法以获取局部语义提取路径，旨在选择出与实体语义高度融合或对实体属性有增强的融合路径，计算方式如式(1)所示.

$$y(\boldsymbol{e}_h)=neighbor(\boldsymbol{e}_h^{\text{o}},I)=[(\boldsymbol{e}_{t_1},\boldsymbol{e}_{r_1}),\cdots,(\boldsymbol{e}_{t_m},\boldsymbol{e}_{r_m})]. \tag{1}$$

其中$neighbor(\cdot)$表示邻域信息提取算法；$\boldsymbol{e}_h$在该算法中为中心实体；$y(\boldsymbol{e}_h)$为中心实体的邻域信息；$I$为局部语义提取路径；$m$为相关路径的数量，针对于不同的实体，相关路径的数量可能不同.

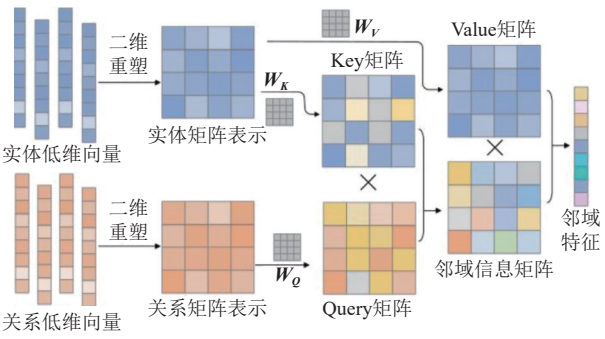



Fig. 2 Neighbourhood information extraction

图 2 邻域信息提取

注意力机制可以帮助模型在处理数据时捕捉到依赖关系，因此设计邻域注意力机制对路径集合中的每条路径进行计算，提取出与中心实体$\boldsymbol{e}_h$强相关的语义特征. 具体过程如式(3)所示.

$$\boldsymbol{L}_i = softmax\left(\frac{\boldsymbol{Q}^{\boldsymbol{L}}\boldsymbol{K}^{\boldsymbol{L}}}{\sqrt{d}}\right)\boldsymbol{V}^{\boldsymbol{L}}, \tag{2}$$

$$\boldsymbol{Q}^{\boldsymbol{L}} = \boldsymbol{e}_{t_i}\times\boldsymbol{W}_{\boldsymbol{Q}}^{\boldsymbol{L}},\ \ \boldsymbol{K}^{\boldsymbol{L}} = \boldsymbol{e}_{r_i}\times\boldsymbol{W}_{\boldsymbol{K}}^{\boldsymbol{L}},\ \ \boldsymbol{V}^{\boldsymbol{L}} = \boldsymbol{e}_{t_i}\times\boldsymbol{W}_{\boldsymbol{V}}^{\boldsymbol{L}}, \tag{3}$$

其中$\boldsymbol{Q}^{\boldsymbol{L}}\in\mathbb{R}^{k\times k}$、$\boldsymbol{K}^{\boldsymbol{L}}\in\mathbb{R}^{k\times k}$和$\boldsymbol{V}^{\boldsymbol{L}}\in\mathbb{R}^{k\times k}$为计算邻域注意力分布的特征矩阵，由路径集合中的向量与映射矩阵$\boldsymbol{W}_{\boldsymbol{Q}}^{\boldsymbol{L}}$，$\boldsymbol{W}_{\boldsymbol{K}}^{\boldsymbol{L}}$和$\boldsymbol{W}_{\boldsymbol{V}}^{\boldsymbol{L}}$计算得到；$\boldsymbol{L}_i$表示邻域路径的语义信息.

$\boldsymbol{L}_i$只包含1条邻域路径中的语义关联，需要将中心实体的所有路径信息进行融合，充分利用邻域中的复杂语义，如式(4)所示.

$$\boldsymbol{L} = \sum_{i=1}^{m}\boldsymbol{L}_i, \tag{4}$$

$\boldsymbol{L}$包含了所有的邻域信息，用于之后增强初始嵌入向量$\boldsymbol{e}_h^{\mathrm{o}}$和$\boldsymbol{e}_t^{\mathrm{o}}$.

2)全局特征提取. 在推理任务中，知识图谱在不同邻域展现出特定的规则模式，对于模型的性能起着至关重要的作用. 然而，当前的很多工作在多数据集的效果不尽如人意. 它们忽视了不同知识图谱所包含的特定信息，导致模型在不同邻域的泛化能力很差.

为了提取整个图谱的模式信息，一种常见的方法是对每个三元组进行线索提取，并利用相应的算法将这些信息进行聚合. 通过这种方式，可以得到整个图谱的模式嵌入，从而增强模型对知识图谱特定领域信息的感知能力，并提高模型在不同数据集上的泛化性能.

注意力机制能够根据知识图谱中的不同实体和关系以及它们之间的关联性，自适应地调整注意力权重，提升模型对于重要邻域信息的关注度. 如图3所示，在ASFR中采用全局注意力机制来捕捉知识图谱的整体结构特征. 利用全局嵌入来表示知识图谱中特定的规则模式，从而增强了实体嵌入的质量.

$$\boldsymbol{G}_i = softmax\left(\frac{\boldsymbol{Q}^{\boldsymbol{G}}\boldsymbol{K}^{\boldsymbol{G}}}{\sqrt{d}}\right)\boldsymbol{V}^{\boldsymbol{G}}, \tag{5}$$

$$\boldsymbol{Q}^{\boldsymbol{G}} = \boldsymbol{e}_{h_i}\times\boldsymbol{W}_{\boldsymbol{Q}}^{\boldsymbol{G}},\ \ \boldsymbol{K}^{\boldsymbol{G}} = \boldsymbol{e}_{r_i}\times\boldsymbol{W}_{\boldsymbol{K}}^{\boldsymbol{G}},\ \ \boldsymbol{V}^{\boldsymbol{G}} = \boldsymbol{e}_{h_i}\times\boldsymbol{W}_{\boldsymbol{V}}^{\boldsymbol{G}}. \tag{6}$$

与局部注意力机制不同，全局注意力机制的$\boldsymbol{Q}^{\boldsymbol{G}}\in\mathbb{R}^{k\times k}$，$\boldsymbol{K}^{\boldsymbol{G}}\in\mathbb{R}^{k\times k}$，$\boldsymbol{V}^{\boldsymbol{G}}\in\mathbb{R}^{k\times k}$矩阵由知识图谱三元组的数据通过特征映射得到，特征映射矩阵为$\boldsymbol{W}_{\boldsymbol{Q}}^{\boldsymbol{G}}$，$\boldsymbol{W}_{\boldsymbol{K}}^{\boldsymbol{G}}$和$\boldsymbol{W}_{\boldsymbol{V}}^{\boldsymbol{G}}$，$\boldsymbol{G}_i$表示从一个三元组中提取全局特征的线索特征. 之后将所有线索特征融合：

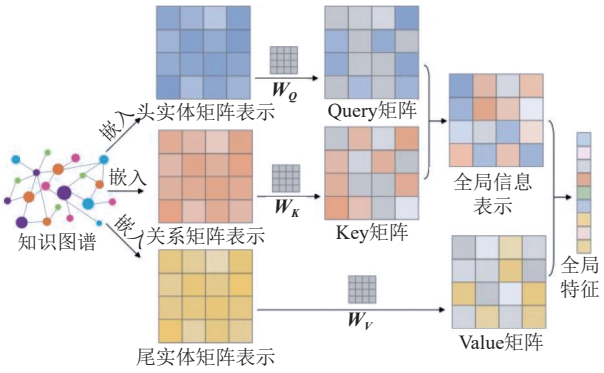



Fig. 3 Global information extraction

图3 全局信息提取

$$\boldsymbol{G} = \sum_{i=1}^{n}\boldsymbol{G}_i, \tag{7}$$

其中$\boldsymbol{G}$为全局特征嵌入表示，包含了知识图谱整体结构特征，可以帮助模型适应不同领域的知识图谱，提高模型的泛化能力.

3)位置特征嵌入. 异构知识图谱中的实体和关系类型可能具有不同的语义和结构特征，涵盖多个存在复杂的关联关系领域或主题. 异构知识图谱更能够反映真实世界中的复杂关联和多样性，同时也使得在异构知识图谱中进行跨领域的知识推理变得更具挑战性.

考虑到异构知识图谱中，不同关系限制了实体类型和语义的表达，ASFR针对不同的关系构建自适应位置嵌入特征. 对与同一关系相关的不同实体构建相同的位置嵌入，并在训练过程中与实体嵌入独立训练：

$$\boldsymbol{e}_{\mathrm{loc}} = \boldsymbol{e}^{\mathrm{o}}\times\boldsymbol{W}_r. \tag{8}$$

其中$\boldsymbol{W}_r$为与关系相关联的转换矩阵，构造位置嵌入特征；$\boldsymbol{e}_{\mathrm{loc}}$为位置嵌入，表示异构知识图谱中不同类型的语义信息.

通过使用位置嵌入特征对初始嵌入向量进行增强，可以将异构信息融入到知识图谱的实体表示中，使模型感知到实体和关系类型的复杂关联和多样性，从而提高模型的推理性能.

### 3.2 特征融合

ASFR利用邻域注意力机制提取局部特征信息，生成局部特征嵌入$\boldsymbol{L}$；利用全局注意力机制捕捉知识图谱的整体结构特征，生成全局特征表示$\boldsymbol{G}$；针对异构知识图谱不同的类别信息，构建位置特征嵌入$\boldsymbol{e}_{\mathrm{loc}}$；将3种信息进行融合：

$$\boldsymbol{e}_{\mathrm{sem}} = \boldsymbol{L} + \tau_1\times\boldsymbol{G} + \tau_2\times\boldsymbol{e}_{\mathrm{loc}}, \tag{9}$$

其中$\tau_1$和$\tau_2$是2个超参数，用于表示整体结构特征和

异构信息特征在信息融合中的比重.

将语义增强的融合特征向量与初始嵌入向量连接,然后输入到传统的知识图谱表示学习模型中进行训练,能够将多种复杂语义信息从嵌入的层次融入模型训练中,增强模型的表达能力.

### 3.3 联合训练

在训练模型参数时,由于嵌入了统一的 ASFR 框架,采用统一的方式对传统的知识图谱表示学习模型进行训练. 不同于“1-1 训练方法”将每个三元组视为独立的训练样本,ASFR 框架引入了“1-$N$ 训练方法”一次性将一个头实体-关系对 $(h, r)$ 与所有可能的尾实体进行匹配,即对于缺失的尾实体,模型会同时计算它与所有候选实体的得分,输出一个维度为实体总数的向量,表示每个实体作为尾实体的可能性. 这种方式可以并行处理大量候选实体,显著加快训练和预测速度,提高整体效率. 具体来说,1-$N$ 训练方法使用实体集的所有候选实体替代缺失的实体,计算 $(h,t,?)$,结果使用一个向量来表示,其维度为数据集中实体的数量,具体算法如算法 1 所示.

**算法 1.** ASFR 框架训练过程.

输入:可观测元组 $\vartheta=(E,R,P)$,迭代次数 $n_{\text{iter}}$,批量大小 $m_{\text{b}}=128$;

输出:每个三元组的分数.

① $\boldsymbol{L}\leftarrow$使用公式计算邻域特征;/* 获取局部语义提取路径*/

② $\boldsymbol{G}\leftarrow$使用公式计算全局特征;

③ $\boldsymbol{E}_{\text{loc}}\leftarrow$构建位置特征嵌入;/*将异构信息融入*/

④ for $t=1,\cdots,n_{\text{iter}}$ do

⑤ 采样一个大小为 $m_{\text{b}}$ 的小批量 $\vartheta_{\text{batch}}$ 属于 $\vartheta$;

⑥ end for

⑦ for each tuple $g\in\vartheta_{\text{batch}}$ do

⑧ $\boldsymbol{E}_{\text{sam}}\leftarrow$使用公式计算语义融合特征;

⑨ $Score\leftarrow$使用公式得到元组的最终得分;

⑩ end for

⑪ 基于公式目标更新可学习的参数.

模型评分函数为:

$$\varsigma(h,r,t)=\sigma(ReLU(\boldsymbol{v}_{\text{out}})\times\boldsymbol{W}_{\text{out}}+b)\boldsymbol{e}_t, \tag{10}$$

其中,$\boldsymbol{v}_{\text{out}}$ 为模型预测尾实体的输出,表示 ASFR 预测特征结果,$\boldsymbol{W}_{\text{out}}$ 为预测特征映射矩阵,$b$ 为偏置项,$\sigma(\cdot)$ 为 sigmoid 函数,将预测特征转化为预测尾部实体嵌入向量. $\boldsymbol{e}_t$ 为真实尾部实体嵌入矩阵,与预测尾部实体嵌入向量相乘得到向量三元组 $(\boldsymbol{e}_h,\boldsymbol{e}_r,\boldsymbol{e}_t)$ 的得分.

知识图谱链接预测任务可以视为对实体的分类任务,即判断 2 个实体之间是否存在某种关系. 相应地,分类任务的类别数量等于实体集中实体的种类数量. 因此,在训练过程中使用在分类任务中应用广泛的交叉熵损失(cross entropy loss)函数. 具体损失函数为:

$$\phi(h,r,t)=-\frac{1}{|\varepsilon|}\sum_{o\in\varepsilon}y_t\log_a(\varsigma)+(1-y_t)\log(1-\varsigma), \tag{11}$$

其中 $\varepsilon$ 是候选实体的个数,$\varsigma$ 表示模型分数 $\varsigma(h,r,t)$. $y_t$ 是一个标签,如果 $(h,r,t)$ 是正确的三元组,则 $y_t=1$,否则 $y_t=1$.

ASFR 使用小批量梯度下降法进行模型参数优化. 在每次训练迭代中,整个数据集被分割为多个较小的三元子集,依次使用这些子集进行训练以提高训练速度. 利用批处理归一化方法来加速运算,并采用 Adam 优化器和标签平滑方法来训练模型.

## 4 实 验

本节介绍实验数据集的选择和评估指标的确定,选择了经典的知识图谱表示学习模型进行实验对比以验证 ASFR 框架的性能,并设计了消融实验验证邻域信息、全局信息和位置信息对框架的贡献.

### 4.1 数据集与评价指标

实验选取了 3 个不同规模的公开数据集,包括 FB15K-237, WN18RR 和 DB100K 数据集. 其中,FB15K-237 和 WN18RR 数据集是链接预测任务最常用的公开数据集,创建过程中移除了验证集和测试集中使用到的训练集中的逆关系,以解决数据泄露问题;而 DB100K 则是一个具有大规模数据量的公开知识图谱数据集,用于测试模型在大规模知识图谱上的泛化能力. 3 个公开数据集的统计信息如表 1 所示.

**Table 1 The Dataset Used in the Experiment**

**表 1 实验中使用的数据集**

| 数据集 | 关系 | 实体 | 训练集 | 验证集 | 测试集 |
|---|---|---|---|---|---|
| FB15K-237 | 237 | 14 541 | 272 115 | 17 535 | 20 466 |
| WN18RR | 11 | 40 943 | 86 835 | 3 034 | 3 134 |
| DB100K | 470 | 99 604 | 597 572 | 50 000 | 50 000 |

链接预测任务采用 2 个评价指标:

1)平均倒数排名(mean reciprocal rank, *MRR*). *MRR* 是测试集中每个正确三元组在模型预测的候选三元组中的倒数排名的平均值,用于强调排名靠前的正确预测.

$$MRR=\frac{1}{\sum\limits_{h\in H_{\text{test}}}|r|}\sum_{h\in H_{\text{test}}}\sum_{p=1}^{|r|}rank_p(h), \tag{12}$$

其中，$H_{\text{test}}$代表测试元组的集合，$rank_p(h)$表示正确元组在替换实体后的候选元组中的排名.

2）命中率*Hits*@*N*. 在前*N*个候选三元组中是否包含了正确的三元组. 常见的值包括*Hits*@1、*Hits*@3和*Hits*@10. 计算公式为：

$$Hits@N=\frac{\sum\limits_{h\in H_{\text{test}}}\sum\limits_{p=1}^{|r|}cond(rank_p(h)\leqslant k)}{\sum\limits_{h\in H_{\text{test}}}|r|}, \tag{13}$$

其中$k\in\{1,3,10\}$，与*Hits*@1、*Hits*@3和*Hits*@10相对应. *Hits*@*N*提供了明确的阈值，可以用来评估模型是否能够在给定数量的候选中成功地找到正确的元组，使得任务需求更加清晰.

### 4.2 基础模型

ASFR框架通过进行邻域和全局的复杂语义融合增强传统知识图谱表示学习模型的性能. 实验选择了不同类别的经典模型作为基础模型，加入ASFR框架进行对比验证，基础模型为TransE, DistMult, RotatE, ConvE和ConvR.

### 4.3 链接预测任务结果

实验基于PyTorch框架进行，训练批大小为128，初始学习率为0.003，标签平滑系数为0.1. 针对每个数据集，我们选择在验证集中具有最高*MRR*的超参数配置，并将其性能在测试集上进行报告. 在FB15K-237, WN18RR和DB100K数据集上提出的框架进行的链接预测实验结果如表2所示. 在FB15K-237和WN18RR数据集上基础模型的实验结果均取自原论文，而由于原论文没有提供DB100K数据集的实验结果，我们在本地环境中使用源代码进行补充.

**Table 2 Link Prediction Results for FB15K-237,WN18RR and DB100K**

**表2 FB15K-237、WN18RR和DB100K的链接预测结果**

| 模型 | | FB15K-237 | | | WN18RR | | | DB100K | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *MRR* | *Hits*@10 | *Hits*@1 | *MRR* | *Hits*@10 | *Hits*@1 | *MRR* | *Hits*@10 | *Hits*@1 |
| TransE | 原模型 | 0.287 | 0.475 | 0.192 | 0.193 | 0.445 | 0.003 | 0.111 | 0.270 | 0.016 |
| | ASFR | 0.294 | 0.491 | 0.206 | 0.210 | 0.459 | 0.058 | 0.134 | 0.284 | 0.059 |
| DistMult | 原模型 | 0.241 | 0.419 | 0.155 | 0.430 | 0.490 | 0.390 | 0.233 | 0.448 | 0.115 |
| | ASFR | 0.261 | 0.442 | 0.170 | 0.446 | 0.504 | 0.407 | 0.247 | 0.462 | 0.123 |
| RotatE | 原模型 | 0.338 | 0.533 | 0.241 | 0.476 | 0.571 | 0.428 | 0.352 | 0.526 | 0.258 |
| | ASFR | 0.351 | **0.548** | 0.253 | 0.495 | **0.586** | 0.440 | 0.370 | 0.537 | 0.272 |
| ConvE | 原模型 | 0.316 | 0.491 | 0.239 | 0.460 | 0.480 | 0.390 | 0.325 | 0.530 | 0.236 |
| | ASFR | 0.332 | 0.504 | 0.253 | 0.475 | 0.492 | 0.401 | 0.338 | 0.541 | 0.253 |
| ConvR | 原模型 | 0.350 | 0.528 | 0.261 | 0.475 | 0.537 | 0.443 | 0.362 | 0.560 | 0.265 |
| | ASFR | **0.367** | 0.544 | **0.279** | **0.497** | 0.560 | **0.463** | **0.385** | **0.586** | **0.288** |

注：黑体数值为当前列的最优结果，下划线数值为当前列的次优结果.

根据实验结果，ASFR在FB15K-237数据集上相较于基准方法，*MRR*提高了4.11%，*Hits*@10提高了2.98个百分点，*Hits*@1提高了5.77个百分点，在所有评估指标的平均改善幅度为4.29%. ASFR在平均关系密度相对较低的WN18RR数据集上也表现出优势，相较于基准方法，*MRR*提高了4.23%，*Hits*@10提高了2.64个百分点，*Hits*@1提高了21.34个百分点. 在所有评估指标的平均改善幅度为9.40%. 值得注意的是，一些浅层模型在平均关系密度较低的数据集中，*Hits*@1指标的表现效果很差，ASFR可以有效提高*Hits*@1的预测效果.

为了验证框架的泛化能力，在DB100K数据集上的实验结果表明，ASFR明显优于基准模型，证明了ASFR在更大规模数据集上具有强大的泛化能力.

经过实验证明，ASFR模型在多个不同的公共数据集上表现出优越的性能. 在所有模型评估指标上，ASFR均优于基准模型的性能. 证明了ASFR框架能够聚合知识图谱局部和全局信息，增强传统知识图谱表示学习的性能.

### 4.4 消融实验

本研究选择了FB15K-237, WN18RR和DB100K三个标准的知识图数据集，并使用与链接预测实验相同的超参数进行了消融实验，实验结果如表3所示. 本研究进行了3组不同的消融实验，分别是没有

邻域信息增强的 ASFR、没有全局信息增强的 ASFR 和没有位置信息增强的 ASFR.

**Table 3 Results of Ablation Experiments**

**表 3 消融实验结果**

| 模型 | FB15K-237 | | | WN18RR | | | DB100K | | |
|---|---|---|---|---|---|---|---|---|---|
| | *MRR* | *Hits*@10 | *Hits*@1 | *MRR* | *Hits*@10 | *Hits*@1 | *MRR* | *Hits*@10 | *Hits*@1 |
| 去掉邻域信息增强 | 0.354 | 0.534 | 0.266 | 0.477 | 0.541 | 0.448 | 0.367 | 0.564 | 0.270 |
| 去掉全局信息增强 | 0.364 | 0.541 | 0.275 | 0.492 | 0.553 | 0.458 | 0.379 | 0.577 | 0.284 |
| 去掉位置信息增强 | 0.359 | 0.538 | 0.270 | 0.484 | 0.549 | 0.453 | 0.372 | 0.569 | 0.278 |
| ASFR（ConvR） | 0.367 | 0.544 | 0.279 | 0.497 | 0.560 | 0.463 | 0.385 | 0.586 | 0.288 |

从实验结果可以看出，ASFR 在 FB15K-237 数据集上所有模型评估指标的性能降低范围为 1.84%~4.66%，在 WN18RR 数据集上为 3.24%~4.02%，在 DB100K 数据集上为 3.75%~6.25%，邻域信息增强是对模型性能影响最大的因素，除此之外位置信息要比全局信息更加重要.

### 4.5 注意力机制性能实验

本实验旨在探讨注意力机制在捕捉邻域特征、全局特征以及不同位置之间复杂语义关联方面的能力. 为此，我们通过设置不同的注意力头数来考察 ASFR 模型的性能变化. 实验选用最通用的公开数据集 FB15K-237. 将全局注意力机制和邻域注意力机制的头数分别设置为 2、4、6、8，旨在通过逐步减少注意力机制的作用，从而降低模型对全局特征和邻域特征的捕捉能力. 该实验使用了最适配的 ConvR 模型作为基础模型，性能评估使用了 *MRR*，*Hits*@1，*Hits*@3，*Hits*@10指标，以全面评估模型的预测能力. 实验结果如表 4 所示.

**Table 4 Results of Attention Mechanism Performance Experiments**

**表 4 注意力机制性能实验结果**

| 注意力头数 | *MRR* | *Hits*@10 | *Hits*@3 | *Hits*@1 |
|---|---|---|---|---|
| 2 | 0.285 | 0.492 | 0.329 | 0.235 |
| 4 | 0.328 | 0.497 | 0.364 | 0.258 |
| 6 | 0.360 | 0.531 | 0.387 | 0.271 |
| 8 | 0.367 | 0.544 | 0.395 | 0.279 |

从表 4 可以看出，随着注意力头数的增加，模型的性能呈现出先上升后稳定的趋势. 注意力头数为 2 时，模型的性能相对较低，因为较少的注意力头数限制了模型捕捉复杂特征的能力. 随着头数增加至 4、6、8，模型性能逐渐提升，表明增加注意力头数有助于模型更细致地捕捉和学习知识图谱中的复杂语义关系. 当注意力头数达到 8 时，性能提升趋于稳定，意味着超过一定数量的注意力头数后，模型性能的提升效果不再显著，或数据集的复杂性已经达到模型能够处理的极限.

通过在 FB15K-237 数据集上调整注意力头数，我们观察到模型性能随着注意力头数的增加而提升，直至达到一个饱和点. 这一现象表明，适量的注意力头数能够显著提高模型对知识图谱深层次语义的理解能力，但超过一定阈值后，性能提升效果趋于平稳，暗示着模型性能可能受到数据集本身复杂性的限制.

## 5 结 论

本文提出了一种注意力增强的语义融合知识图谱表示学习框架 ASFR，旨在增强传统知识图谱表示学习模型对复杂语义的理解能力. 该框架利用 2 种不同的注意力机制分别提取图谱邻域信息和全局信息，并构建自适应位置特征嵌入，以对实体的初始嵌入进行语义增强. ASFR 框架具备融合大多数现存知识图谱表示学习方法的能力，并能够提升模型的推理效果. 广泛的实验验证表明，该框架能够成功捕捉知识图谱不同层次的语义信息，从而实现对传统模型的语义增强，提高模型的预测性能和推理能力. 消融实验证明了框架各组成部分的有效性.

**作者贡献声明：** 郭文斌制定了 ASFR 框架的算法设计和实验方案，实施实验以及收集和分析数据，并撰写论文；王鑫提供了 ASFR 框架设计的指导意见，并在论文的修改和完善过程中发挥了关键作用；李钊、陈子睿和赵军参与了实验设计和结果讨论，为实验的准确性和可靠性提供了宝贵的建议；冯锋和杜方在理论分析和文献综述方面做出了贡献，确保了论文内容的深度和广度.

## 参 考 文 献


[1] Liu Zhiyuan, Sun Maosong, Lin Yankai, et al. Knowledge representation learning: A review[J]. Journal of Computer Research and Development, 2016, 53(2): 247−261 (in Chinese)
(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)

[2] Wang Xin, Zuo Lei, Wang Chaokun, et al. Research on knowledge graph data management: A survey[J]. Journal of Software, 2019, 30(7): 2139−2174 (in Chinese)
(王鑫, 邹磊, 王朝坤, 等. 知识图谱数据管理研究综述[J]. 软件学报, 2019, 30(7): 2139−2174)

[3] Bordes A, Usunier N, Garcia A, et al. Translating embeddings for modeling multi-relational data[J]. Advances in Neural Information Processing Systems, 2013, 26: 2787−2795

[4] Wang Zhen, Zhang Jianwen, Feng Jianlin, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proc of the 28th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2014: 1112−1119

[5] Lin Yankai, Liu Zhiyuan, Sun Maosong, et al. Learning entity and relation embeddings for knowledge graph completion[C]//Proc of the 29th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI , 2015: 2181−2187

[6] Ji Guoliang, He Shizhu, Xu Liheng, et al. Knowledge graph embedding via dynamic mapping matrix[C]//Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2015: 687−696

[7] Nayyeri M, Vahdati S, Aykul C, et al. 5* knowledge graph embeddings with projective transformations[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021, 35(10): 9064−9072

[8] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings[C]//Proc of the 32th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI , 2018, 32(1): 1811−1818

[9] Nguyen D, Nguyen Q, Phung D. A novel embedding model for knowledge base completion based on convolutional neural network[C]//Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). Stroudsburg, PA: ACL, 2018: 327−333

[10] Vashishth S, Sanyal S, Nitin V, et al. Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020, 34(3): 3009−3016

[11] Ren Feiliang, LI Juchen, Zhang Huihui, et al. Knowledge graph embedding with atrous convolution and residual learning[C]//Proc of the 28th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2020: 1532−1543

[12] Jiang Xiaotian, Wang Quan, Wang Bin. Adaptive convolution for multi-relational learning[C]//Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Stroudsburg, PA: ACL, 2019: 978−987

[13] Sheikh N, Qin X, Reinwaid B, et al. Knowledge graph embedding using graph convolutional networks with relation-aware attention[J]. arXiv preprint, arXiv: 2102.07200, 2021

[14] Li Zhifei, Liu Hai, Zhang Zhaoli, et al. Learning knowledge graph embedding with heterogeneous relation attention networks[J]. IEEE Trans on Neural Networks and Learning Systems, 2021, 33(8): 3961−3973

[15] Saxena A, Kochsiek A, Gemulla R. Sequence-to-sequence knowledge graph completion and question answering[C]//Proc of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2022: 2814−2828

[16] Chen Sanxing, Liu Xiaodong, Gao Jianfeng, et al. HittER: Hierarchical transformers for knowledge graph embeddings[C]//Proc of the 2021 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2021: 10395−10407

[17] Wang Yilin, Hu Minghao, Huang Zhen, et al. A canonicalization-enhanced known fact-aware framework for open knowledge graph link prediction[C]//Proc of the 32nd Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2023: 2332−2342

[18] Niu Guanglin, Li Bo, Zhang Yongfei, et al. CAKE: A scalable commonsense-aware framework for multi-view knowledge graph completion[C]//Proc of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2022: 2867−2877

[19] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//Proc of the 15th Int Semantic Web Conf. Berlin: Springer, 2018: 593−607

[20] Tang Zhenwei, Pei Shichao, Zhang Zhao, et al. Positive-unlabeled learning with adversarial data augmentation for knowledge graph completion[C]//Proc of the 31st Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2022: 2248−2254

[21] Wang Huijian, Dai Siming, Su Weiyue, et al. Simple and effective relation-based embedding propagation for knowledge representation learning[C]//Proc of the 31st Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2022: 2123−2130

[22] Wang Meng, Wang Haofen, Li Bohan, et al. Survey on key technologies of new generation knowledge graph[J]. Journal of Computer Research and Development, 2022, 59(9): 1947−1965 (in Chinese)
(王萌, 王昊奋, 李博涵, 等. 新一代知识图谱关键技术综述[J]. 计算机研究与发展, 2022, 59(9): 1947−1965)

[23] Ma Ang, Yu Yanhua, Yang Shengli, et al. Survey of knowledge graph based on reinforcement learning[J]. Journal of Computer Research and Development, 2022, 59(8): 1694−1722 (in Chinese)
(马昂, 于艳华, 杨胜利, 等. 基于强化学习的知识图谱综述[J]. 计算机研究与发展, 2022, 59(8): 1694−1722)



**Guo Wenbin**, born in 2000. PhD candidate. Student member of CCF. His main research interest includes knowledge representation learning.

**郭文斌**, 2000 年生，博士研究生. CCF 学生会员. 主要研究方向为知识表示学习.



**Wang Xin**, born in 1981. PhD, professor. Distinguished member of CCF. His main research interests include knowledge graph data management, graph databases, and big data distributed processing.

**王　鑫**, 1981 年生. 博士，教授. CCF 杰出会员. 主要研究方向为知识图数据管理、图数据库、大数据分布式处理.



**Li Zhao**, born in 1994. PhD candidate. Student member of CCF. His main research interests include knowledge graph, and representation learning and reasoning.

**李　钊**, 1994 年生. 博士研究生. CCF 学生会员. 主要研究方向为知识图谱、表示学习与推理.



**Chen Zirui**, born in 1998. PhD candidate. Student member of CCF. His main research interests include knowledge representation learning and knowledge graph question answering.

**陈子睿**, 1998 年生. 博士研究生. CCF 学生会员. 主要研究方向为知识表示学习、知识图谱问答.



**Zhao Jun**, born in 1971. PhD, professor. Member of CCF. His main research interests include information systems, computational experiments, and intelligent applications of big data.

**赵　军**, 1971 年生. 博士，教授. CCF 会员. 主要研究方向为信息系统、计算实验、大数据智能应用.



**Feng Feng**, born in 1971. PhD, professor. Senior member of CCF. His main research interests include the integration and application of information systems and Internet of things engineering.

**冯　锋**, 1971 年生. 博士，教授. CCF 高级会员. 主要研究方向为信息系统的集成与应用、物联网工程.



**Du Fang**, born in 1974. PhD, professor. Senior member of CCF. Her main research interests include big data analytics, natural language processing, and biomedical multimodal data analysis and mining.

**杜　方**, 1974 年生. 博士，教授. CCF 高级会员. 主要研究方向为大数据分析、自然语言处理、生物医学多模态数据分析与挖掘.