

## 摘要

知识图谱 (Knowledge Graph, KGs) 是一种关键的结构化知识数据库, 旨在帮助计算机系统更好地理解 and 推理人类知识。知识图谱帮助机器理解知识的语义信息, 并作为核心数据基础广泛应用于人工智能领域, 如智能语义搜索 (Semantic Searching)、智能问答系统 (Question Answering) 等。目前, 知识图谱更是成为解决大模型 (Large Language Models) 幻觉 (Hallucination) 问题的主要技术手段之一, 具有极其重要的研究价值。近年来, 知识图谱表示学习 (Knowledge Graph Representation Learning, KRL) 一直是人工智能领域的研究热点, 它旨在通过学习实体和关系的嵌入向量表示 (embeddings), 能够更好地捕捉语义关联。知识图谱表示学习在促进语义理解、推理能力、跨模态关联、领域专业化以及个性化服务等方面都发挥着重要的作用, 是推动人工智能技术不断发展的关键之一。

本论文主要开展知识图谱表示学习研究。目前, 图注意力网络 (GATs) 已被证明是一种有效的图学习模型。然而, 该类模型仅考虑了实体邻域内利用局部特征, 对于需要深入挖掘全局上下文信息的信息表示学习, 仍需探索更加高效的全局信息集成方法。为了解决这个问题, 本文专注于考虑如何在知识图谱表示学习中将全局信息 (Global Information) 合理融入局部注意力 (Local Attention) 中。基于以上思想, 本文提出了一种新的考虑实体重要性的

图神经网络模型 (Entity Importance-aware Graph Attention Networks, 简称为 EIGAT), 它在局部注意力机制中融入全局实体重要性以学习有效的知识图谱知识表示。具体而言, 本文通过使用一个全局随机游走 (Random Walk) 算法在整个图结构信息上估计的实体重要性 (Entity Importance, EI), 允许在图神经网络模型中很好的利用全局信息。最后, 本文在多种不同类型的知识图谱上进行了广泛的实验, 通过实体关系预测与基准模型进行对比, 充分证明了本文提出的 EIGAT 能够成功将全局信息融入局部注意力中, 以提高知识表示学习的效果。

**关键词:** 知识图谱、表示学习、图神经网络

## Abstract

Knowledge Graphs (KGs) are a crucial structured knowledge database designed to assist computer systems in better understanding and reasoning human knowledge. Knowledge graphs help machines understand the semantic information of knowledge, which are widely applied as core data foundations in the field of artificial intelligence, such as semantic searching, question-answering systems, and more. Currently, knowledge graphs have become one of the main technical means to address the hallucination problem in large language models, holding significant research value. In recent years, Knowledge Graph Representation Learning (KRL) has been a research hotspot in the field of artificial intelligence. It aims to better capture semantic correlations by learning embeddings of entities and relations. Knowledge graph representation learning plays an important role in promoting semantic understanding, reasoning ability, cross-modal correlations, domain specialization, personalized services, and more, making it one of the key drivers of the continuous development of artificial intelligence technology.

This thesis mainly focuses on the research of knowledge graph representation learning. Currently, Graph Attention Networks (GATs) have been proven to be an effective model for graph learning. However, such models only consider utilizing

local features within entity neighborhoods. It requires deep exploration of global context information for knowledge representation learning. There is still a need to explore more efficient methods for integrating global information.

To address this issue, this thesis focuses on considering how to reasonably integrate global information into local attention in knowledge graph representation learning. Based on this idea, a new graph neural network model called Entity Importance-aware Graph Attention Networks (EIGAT) is proposed, which incorporates global entity importance into the local attention mechanism to learn effective knowledge graph representations. Specifically, by using a global random walk algorithm to estimate entity importance (EI) across the entire graph structure, this paper allows for the effective utilization of global information within the graph neural network model. Finally, it conducted extensive experiments on various types of knowledge graphs, comparing EIGAT with baseline models in entity relation prediction, which fully demonstrate that EIGAT successfully integrates global information into local attention to improve the effectiveness of knowledge representation learning.

**Keywords:** Knowledge Graph, Knowledge Representation Learning, Graph Neural Networks

# 目 录

<b>摘要</b>	<b>I</b>
<b>Abstract</b>	<b>III</b>
<b>1. 绪论</b>	<b>1</b>
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	6
1.2.1 基于翻译的表示学习 .....	6
1.2.2 基于张量因子分解的表示学习 .....	8
1.2.3 基于卷积的表示学习 .....	10
1.2.4 基于图神经网络模型的表示学习 .....	11
1.3 研究内容及创新点 .....	11
1.4 论文组织结构 .....	15
<b>2. 相关知识与关键技术</b>	<b>17</b>
2.1 知识图谱 .....	17
2.1.1 通用知识图谱 .....	17

2.1.2	垂直领域知识图谱 .....	19
2.2	图神经网络模型 .....	20
2.2.1	同质图神经网络模型 .....	21
2.2.2	异质图神经网络模型 .....	22
2.3	节点重要性评估 .....	23
2.4	本章小结 .....	24
<b>3.</b>	<b>融合全局信息与局部信息的知识图谱表示学习模型</b> .....	<b>26</b>
3.1	单层神经网络模块 .....	26
3.1.1	符号定义 .....	26
3.1.2	局部注意力评估 .....	27
3.1.3	全局实体重要性评估 .....	28
3.1.4	全局信息与局部信息融合方法 .....	29
3.2	模型整体框架 .....	31
3.2.1	编码模块 (Encoder Model) .....	31
3.2.2	解码模块 (Decoder Model) .....	32
3.3	优化算法 .....	32
3.4	本章小结 .....	33
<b>4.</b>	<b>实验设置与结果分析</b> .....	<b>35</b>
4.1	数据集 .....	35

---

4.2	实验设置 .....	37
4.2.1	基线模型 .....	37
4.2.2	评价指标 .....	40
4.2.3	训练参数设置 .....	41
4.3	关系预测结果分析.....	43
4.4	消融实验分析.....	45
4.5	案例研究 .....	46
4.5.1	实体预测 .....	46
4.5.2	实体重要性评估 .....	47
4.6	本章小结 .....	49
<b>5.</b>	<b>总结和展望</b> .....	<b>50</b>
5.1	工作总结 .....	50
5.2	未来展望 .....	51
	<b>参考文献</b> .....	<b>52</b>
	<b>致谢</b> .....	<b>70</b>
	<b>在读期间科研成果</b> .....	<b>72</b>



# 1. 绪论

本章节首先介绍了课题的研究背景和意义，然后概述了课题的研究现状，接着阐述了本文的主要研究工作和创新点，最后对本文的结构进行了简要介绍。

## 1.1 研究背景与意义

随着互联网信息科学技术的迅猛发展，尤其是在进入 Web 2.0 时代之后，用户拥有了更多的自主生成和传播信息的能力，这导致信息量呈现出爆炸式的增长。然而，这些信息主要以人类可理解的形式存在，机器很难理解其语义。为了克服这一挑战并帮助计算机更好地理解语义信息，早在上世纪 60 和 70 年代，先驱学者 Simon 和 Minsky 等进行了早期的探索。Simon 在进行自然语言理解 (Natural Language Understanding) 的应用研究中提出了语义网络 (Semantic Network) 的概念，试图通过将语言中的概念和关系表示为图形结构，使计算机能够更好地理解和推理。然而，由于当时计算机处理能力的限制和数据稀缺，这些尝试在实际应用中受到了一些限制。1998 年，万维网联盟的蒂姆·伯纳斯-李 (Tim Berners-Lee) 提出了语义网 (Semantic Web) 的构想，标志着对语义信息的研究迈出了新的一步。语义网的目标是通过为网络内容赋予机器可理解的语义，促进信息的自动处理和整合。这一构想推动了

一系列技术和标准的发展，包括 RDF（资源描述框架）和 OWL（Web 本体语言），为机器理解和处理信息提供了更多可能性。

与此同时，随着互联网的广泛普及，大量的资源描述框架数据（Resource Description Framework, RDF）已经在网络上迅速增加，其中包括用户生成内容（UGC）和链接开放数据（LOD）等。这些数据的增长使得互联网变得更加丰富和多样化，为计算机系统提供了更多语义信息的可能性。用户生成内容的涌现，如社交媒体上的评论、博客和论坛，为知识图谱的构建提供了丰富的用户参与和实时数据。另一方面，互联网正在逐渐从主要通过网页超链接构建的文档万维网（Document Web）向描述各种实体及其丰富关系的数据万维网（Data Web）转变。这一转变意味着互联网不再仅仅是存储和传递文档，而是成为了一个巨大的、结构化的知识库。在数据万维网中，信息以数据的形式存在，使得计算机可以更直接地获取、理解和利用其中的知识。这也为知识图谱的发展提供了更加丰富和多维的数据来源。

在这个背景下，谷歌公司于 2012 年 5 月提出了知识图谱（Knowledge Graph, KGs）的概念，并将其用于改善其搜索引擎的性能。知识图谱不仅是一个巨大的知识库，更是一个连接各种实体之间关系的网络。谷歌的知识图谱不断地从互联网上的各种可信数据源中抽取知识，并以图形结构的形式表示。这种表示方式使得搜索引擎能够更好地理解用户的查询意图，提供更准确、全面的搜索结果。知识图谱的引入标志着对于互联网信息处理的一次巨大飞跃，不仅提高了搜索引擎的性能，还促进了语义搜索、智能问答系统等人工智能领域的发展。未来，随着互联网的不断演进和技术的不断创新，知识图谱将在更多领域发挥关键作用，为人们提供更智能、个性化的信息服务。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/675110314102012020>