

本科毕业论文（设计）

**论文题目（中文）**  基于知识感知的关系抽取方法研究

**论文题目（英文）** Research on Relation Extraction Based on

Knowledge-Aware Methods

学生姓名 张若寒

指导教师 赵志立

学 院 信息科学与工程学院

专 业 数据科学与大数据技术

年 级 2019级

兰州大学教务处

**诚信责任书**

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人、集体已经发表或未发表的论文。

本声明的法律责任由本人承担。

论文作者签名： 日 期：

**关于毕业论文（设计）使用授权的声明**

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文（设计）的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文（设计）。本人离校后发表、使用毕业论文（设计）或与该毕业论文（设计）直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文（设计）研究内容：

□可以公开

□不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名： 导师签名：

日 期： 日 期：

**基于知识感知的关系抽取方法研究**

中文摘要

关系抽取任务作为信息抽取的重要一环，也是构建知识图谱的关键支撑，具有重要的研究意义。关系抽取旨在从文本中抽取实体及实体间的关系，并以结构化的形式存储。近年来，预训练模型在自然语言处理领域展现了出色的性能，提示学习的提出使预训练模型的性能得到了更好的发挥。然而对于关系抽取任务来说，使用提示学习存在着难以构建模板、无法利用关系标签中的丰富知识的问题。因此，本文针对关系抽取任务，研究如何将提示学习应用到基于预训练模型的关系抽取模型中，并提出基于知识感知的协同优化方法，以达到更好的性能和泛化能力。具体研究内容有：

（1）设计了一个关系抽取模型K-Prompt，整体框架为：首先对输入文本进行分词得到输入序列，然后采用预训练模型RoBERTa对输入序列和模板一起编码，得到向量表示，接着使用关系解码模型对向量进行解码，最终输出关系类型。模型的构建结合了提示学习方法，设计了简单的模板，并且采用了基于知识感知的协同优化方法，以充分利用关系标签内的知识和关系三元组的结构化知识。

（2）在SemEval、DialogRE等关系抽取数据集上进行实验，并与基线模型对比。本文从模型对比和数据集对比两个方面对实验结果进行分析。实验表明，与其他关系抽取基线模型相比，本文模型在整体上的性能最优，证明了在模型中融入提示学习和知识感知的有效性。在数据集方面，模型在传统关系抽取数据集上表现较好，在对话数据集上表现较差。另外，本文还在ACE2005数据集上进一步测试了K-Prompt模型的性能，结果表明其在全监督设置和低资源设置下表现都较为良好。

**关键词：**关系抽取；预训练模型；提示学习；知识感知

RESEARCH ON RELATION EXTRACTION BASED ON KNOWLEDGE-AWARE METHODS

Abstract

As an important part of Information Extraction, Relation Extraction is also a key support for building knowledge graphs, which has important research significance. Relation Extraction aims to extract entities and relationships between entities from text and store them in a structured form. In recent years, the pre-trained models have shown excellent performance in the field of Natural Language Processing, and the introduction of prompt learning has made the performance of the pre-trained models better. However, for relation extraction tasks, using prompt learning has the problem of being difficult to construct templates and unable to utilize the rich knowledge in relation labels. Therefore, for relation extraction tasks, this paper studies how to apply prompt learning to relation extraction models based on pre-trained models, and proposes a knowledge-aware collaborative optimization method to achieve better performance and generalization ability. The specific research contents include:

(1) A relation extraction model K-Prompt is designed. The overall framework is: first, the input text is tokenized to obtain the input sequence, then the pre-trained model RoBERTa is used to encode the input sequence and the template together to obtain a vector representation, then use the relation decoding model to decode the vector and finally output the relation type. The construction of the model combines prompt learning methods, simple templates are designed, and a knowledge-aware based collaborative optimization method is adopted to make full use of the knowledge within relation labels and the structured knowledge of relation triples.

(2) Conduct experiments on relation extraction datasets such as SemEval and DialogRE, and compare with baseline models. This paper analyzes the experimental results from two aspects: model comparison and dataset comparison. Experiments show that compared with other relation extraction baseline models, our model has the best overall performance, proving the effectiveness of incorporating prompt learning and knowledge into the model. In terms of datasets, the model performs better on traditional relation extraction datasets and poorer on the dialogue dataset. In addition, this paper further tests the performance of the K-Prompt model on the ACE2005 dataset, and the results show that it performs well in both fully supervised and low-resource settings.

**Keywords:** relation extraction; pre-trained model; prompt learning; knowledge -aware

目 录

[中文摘要 I](#_Toc134718056)

[Abstract II](#_Toc134718057)

[第一章 绪 论 1](#_Toc134718058)

[1.1 选题背景及意义 1](#_Toc134718059)

[1.2 研究现状 2](#_Toc134718060)

[1.2.1 基于模式匹配的方法 2](#_Toc134718061)

[1.2.2 基于机器学习的方法 3](#_Toc134718062)

[1.2.3 基于深度学习的方法 4](#_Toc134718063)

[1.3 本文主要研究内容及组织结构 5](#_Toc134718064)

[1.3.1 研究内容 5](#_Toc134718065)

[1.3.2 组织结构 6](#_Toc134718066)

[第二章 研究路线及相关理论 7](#_Toc134718067)

[2.1 研究路线 7](#_Toc134718068)

[2.2 预训练模型 8](#_Toc134718069)

[2.3 提示学习 9](#_Toc134718070)

[2.4 本章小结 10](#_Toc134718071)

[第三章 基于知识感知的关系抽取模型 11](#_Toc134718072)

[3.1 概述 11](#_Toc134718073)

[3.2 模型构建 12](#_Toc134718074)

[3.3 处理流程 13](#_Toc134718075)

[3.4 本章小结 15](#_Toc134718076)

[第四章 实验与分析 16](#_Toc134718077)

[4.1 实验设置 16](#_Toc134718078)

[4.1.1 实验环境设置 16](#_Toc134718079)

[4.1.2 实验数据集 16](#_Toc134718080)

[4.1.3 实验参数设置 17](#_Toc134718081)

[4.2 评估标准 18](#_Toc134718082)

[4.3 实验实施与结果分析 18](#_Toc134718083)

[4.3.1 对比实验 18](#_Toc134718084)

[4.3.2 ACE2005实验 20](#_Toc134718085)

[4.4 本章小结 21](#_Toc134718086)

[第五章 总结与展望 22](#_Toc134718087)

[5.1 工作总结 22](#_Toc134718088)

[5.2 工作展望 22](#_Toc134718089)

[参考文献 23](#_Toc134718090)

[致 谢 25](#_Toc134718091)

**图 目 录**

[图 1.1 知识图谱网状结构 1](#_Toc135135356)

[图 1.2 本文组织结构 6](#_Toc135135357)

[图 2.1 本文研究路线 7](#_Toc135135358)

[图 4.1 ACE2005实验结果 20](#_Toc135135359)

**表 目 录**

[表 4.1 数据集统计信息 16](#_Toc131334432)

[表 4.2 超参数设置 17](#_Toc131334433)

[表 4.3 混淆矩阵 18](#_Toc131334434)

[表 4.4 评价指标计算公式及含义 18](#_Toc131334435)

[表 4.5 对比实验结果 19](#_Toc131334436)

1. 绪 论

1.1 选题背景及意义

长久以来，搜索引擎在满足互联网用户的搜索需求方面发挥了重要作用。传统的搜索引擎通过捕捉用户输入的关键词，查询与其最相关的文档，按照相关程度进行排名，作为结果呈现给用户，然而这些文档有时不符合用户需求。因此，一些搜索引擎试图通过建立一个包含实体及其关系的图，来准确理解用户的搜索需求，知识图谱（Knowledge Graph）应运而生。

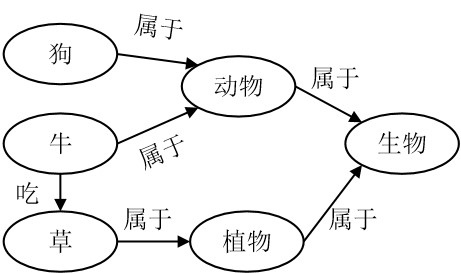


图 1.1 知识图谱网状结构

2012年，Google首次提出知识图谱的概念，它是用来理解真实世界中的实体及实体间关系的一张“图”，使用户能够搜索Google了解的事物并立即获取与查询相关的信息。如图1.1所示，知识图谱是一个网状结构，由“实体-关系-实体”三元组组成，图中的节点为实体，节点之间的边表示实体间的关系。知识图谱的发展为信息搜索、智能问答、决策支持等智能应用提供了技术支撑。

信息抽取（Information Extraction）作为构建知识图谱的第一步，是支撑知识图谱构建的关键。信息抽取任务包含实体抽取（Entity Extraction）、关系抽取（Relation Extraction）和事件抽取（Event Extraction）。关系抽取任务描述为从文本中抽取出实体间的关系，并以关系三元组（主体，关系，客体）的形式输出。例如，给定一段文本“刘翔，1983年7月13日出生于上海”，采用关系抽取相关技术从中抽取出的关系三元组为（刘翔，出生地，上海）。关系抽取作为信息抽取的子任务之一，旨在将无结构的信息以结构化的形式存储，对处理海量数据、提高检索效率、促进智能应用的发展提供了支持，具有重要的理论意义和广阔的应用前景。

早期关系抽取大多采用基于模式匹配或基于机器学习的算法，基于神经网络的深度学习方法进一步提升了关系抽取模型的性能。随着预训练模型（Pre-trained Models, PTMs）的兴起，使用预训练模型如BERT（Bidirectional Encoder Representations from Transformers, BERT）[1]作为关系抽取的骨干架构逐渐成为研究主流，各种基于“预训练+微调”范式的关系抽取模型在相关的关系抽取数据集上取得了很好的效果。“微调”是指针对不同的下游任务来改变预训练模型的参数。然而，预训练模型在关系抽取任务上还存在以下两点不足：

（1）微调过程虽然降低了数据量的需求，但仍离不开高质量的人工标注，导致任务模型难以泛化到不同的下游任务中，且随着预训练模型越来越大，微调的成本也越来越高；

（2）预训练模型在训练和微调过程存在的差异，会降低任务模型的性能。

GPT-3[2]模型被提出后，提示学习（Prompt Learning）引起了人们的关注与研究热潮，并迅速在广泛的自然语言处理（Natural Language Processing, NLP）任务中取得了优异的性能。因此，最近一些研究人员使用提示学习来解决以上问题，该方法的核心思想是不调整预训练模型的参数，而让下游任务向预训练模型靠近。具体来说，通过设计模板将关系抽取任务转变为与预训练相同的任务，将预训练模型直接作为预测器，而不需要在预训练模型的基础上增加额外的分类器，这样就弥补了预训练和微调阶段的差距，有助于通用模型的构建。

虽然提示学习在许多自然语言处理任务上取得了很好的结果，如文本分类中的情感分类任务，但对于关系抽取任务来说，使用提示学习的方法还面临着一些难以解决的挑战。例如使用传统提示学习方法在模板生成方面较为困难，手工生成模板要求掌握相关领域的专业知识，耗费人力资源；自动生成模板需要在模板空间进行搜索，耗费计算成本。另外，对于关系抽取任务来说，关系标签中蕴含的丰富知识有利于关系的预测，应该融入模型的构建中。

因此，本文在提示学习的基础上做出改进，基线模型采用预训练模型结合提示学习，并将关系标签中的先验知识和关系三元组中隐含的结构化知识融入模型的构建中。该方法克服了以上方法的缺陷，具有较强的理论意义。

1.2 研究现状

关系抽取任务自1998年在MUC会议上被提出后，一直备受人们的关注，并被持续地研究。关系抽取任务的主要难点在于关系重叠，即一个实体和另一个实体间可能存在多种关系，或者一个实体和其他多个不同实体间存在多种关系，这种关系重叠问题会影响关系抽取的性能。随着关系抽取的方法逐渐改进完善，从早期的基于模式匹配的方法，再到基于机器学习、深度学习的方法，关系抽取的性能也在不断提高。本节主要对关系抽取方法进行概述。

1.2.1 基于模式匹配的方法

早期关系抽取任务的研究大多采用基于模式匹配的方法。所谓模式是指关系抽取过程中使用的实体关系的词典或规则，需要提前运用语言学知识人工构造，然后将语句片段与模式进行匹配来完成关系抽取。Mary等人[3]提出了一个从信息抽取任务中学习模式的方法RAPIER，使用样本文档和填充模板对来诱导模式的生成。Miller等人[4]提出了一个具有头部规则的词汇化、概率上下文无关的解析器LPCFG-HR，可以有效地进行信息抽取。由于中文在语法、语义等方面与英文差距较大，一般的模式匹配方法在中文数据集上表现的性能不佳，邓擘等人[5]使用了中文的词汇语义匹配方法来进行中文数据集的关系抽取，效果优于一般的模式匹配方法。

基于模式匹配的方法使用预定义的词典或规则，这些词典或规则需要领域专家来构造，因此人工标注成本高、难度大且移植性差，为此，研究人员转而尝试使用机器学习方法建模实体间关系。

1.2.2 基于机器学习的方法

基于机器学习的方法本质上是构建分类模型，包含两个阶段：首先在语料库上进行特征学习，构建关系抽取模型；然后用该关系抽取模型对测试文本进行预测。以数据标注情况作为分类标准，基于机器学习的方法可分为三类：有监督方法、半监督方法和无监督方法。

有监督方法使用完全标注的语料进行训练，将关系抽取任务转化成分类问题。目前流行的有监督关系抽取方法有基于特征向量的方法和基于核函数的方法。针对前者，Kambhatla[6]提出使用结合文本特征的最大熵模型进行关系抽取的方法，文本特征包括词汇、句法和语义特征。Sun等人[7]提出一种考虑长距离依赖的新的特征表示方法，在向量空间模型中加入了特征的位置信息，并使用朴素贝叶斯和投票感知机混合算法进行关系分类。针对后者，Zelenko等人[8]使用定义在文本的浅层解析表示上的核函数，利用支持向量机和投票感知机来进行关系抽取。Zhou等人[9]提出了基于卷积树核函数的关系抽取方法，构建语义关系树来整合丰富的语法和语义信息，并用卷积树核来捕获树结构中的结构化信息。有监督方法依赖于大量标注的语料库，耗费人力，扩展性差。

半监督方法仅使用少量标注数据，在大量未标注的数据上进行训练。半监督方法主要包括自举（Bootstrapping）方法、协同训练（co-training）方法和标注传播（label propagation）方法。Brin[10]采用自举方法，将少量实例作为初始种子模板，从大量来自网络的训练数据中学习新的实例和抽取模板，构建了DIPRE系统，用于关系抽取。Agichtein等人[11]提出了Snowball抽取系统，在Brin方法的基础上对新抽取的关系实例进行可信度评分，并完善了关系描述方法。Balcan等人[12]提出协同训练的半监督学习算法，使用两个分类器进行分类，相互学习以提高关系抽取性能；Zhu等人[13]提出基于图的标注传播算法，图中节点和边分别表示实体和关系，利用已标记的节点的信息标注未标记的节点。

无监督方法用来解决大规模语料中关系类型未知的问题，采用基于聚类的方法将可能具有同一种关系的实体对聚成一类，选择合适的标签表示。Hasegawa等人[14]首次提出无监督的关系抽取方法，核心思想是利用上下文判断哪些实体间具有相似的语义关系；为了避免关系重叠问题，Rozenfold等人[15]对其做了进一步完善，剔除了具有多种关系的实体对或在同一语料库中聚类实体对。

1.2.3 基于深度学习的方法

传统的机器学习方法一般通过人工构造的准则来选取有效的特征，由于特征学习和模型学习是两个分开的阶段，导致学习到的特征可能无法提升模型的性能。随着深度学习和神经网络模型的提出，采用基于深度学习的方法不断提高了关系抽取的性能。基于深度学习的方法主要分为有监督方法和远程监督方法，本节分别对这两种方法进行介绍。

⑴ 有监督方法

不同于传统机器学习中的有监督方法，深度学习中的有监督方法可以自动从训练语料中学习特征，避免了人工特征选择带来的人力消耗和错误传播问题。按模型结构分类，有监督的深度学习关系抽取模型可分为两大类：流水线（Pipeline）模型和联合（Joint）模型，下面将分别介绍这两种模型。

A、流水线模型

流水线模型是指实体识别和关系分类作为两个独立的子任务进行，即先抽取出文本中的全部实体，再判断每两个实体对之间可能存在的关系。流水线模型下的关系抽取任务，本质上是在命名实体识别（Named Entity Recognition）工作已经完成的基础上进行关系分类。

基于流水线模型的关系抽取方法常采用神经网络架构，如卷积神经网络（Convolutional Neural Networks, CNNs）、循环神经网络（Recurrent Neural Networks, RNNs）、长短期记忆网络（Long Short-Term Memory, LSTM）等等。Socher等人[16]最早使用具有内部记忆的RNN模型进行关系抽取，Hashimoto等人[17]提出基于句法依存分析树的RNN模型；Zeng等人[18]首次提出将CNN用于关系抽取任务，后续研究人员将注意力机制和句法树等融入CNN以提升性能；Xu等人[19]提出使用融合句法树的最短路径以及句法类型等特征的LSTM模型进行关系抽取。

流水线模型简单、易于实现，并且由于实体识别和关系分类是两个互不干扰的过程，因而灵活性较高。但流水线模型存在着明显的缺点：首先是误差传播问题，实体识别模型产生的错误会对关系分类模型的性能造成影响；其次，由于实体识别阶段会将所有实体识别出来，在关系分类阶段对实体两两配对再进行关系抽取，因此彼此之间没有关系的实体对造成了大量冗余信息，提升了计算复杂度，降低了准确率；另外，两个子任务之间没有交互，这两个任务间的内在联系被忽视，这些被忽视的信息可能会影响关系抽取的效果。

B 联合模型

为解决流水线模型的明显缺陷，研究人员提出了联合抽取模型。联合抽取模型是指将实体识别和关系分类进行联合建模，通过改变模型结构直接输出文本中包含的关系三元组。联合模型的主流方法可以分为基于参数共享的方法和基于序列标注的方法。

在基于参数共享的方法下，实体抽取和关系抽取不是同步的，这两个过程通过共享参数来实现信息交互，不断更新以找到全局关系抽取任务的最优参数。2016年Miwa等人[20]首次提出使用联合模型的思路来将实体识别和关系抽取任务融合，通过共享编码层的LSTM来更新全局参数。

虽然基于参数共享的联合模型解决了流水线模型的误差传播问题和信息交互问题，但仍存在着实体对冗余的缺点，因此Zheng等人[21]提出可同时提取实体和关系的基于序列标注的方法。该方法通过标注实体中词的位置信息、实体间关系类型信息以及实体角色信息，以端到端的方式直接得到关系三元组。

实验证明，在大多数关系抽取任务中，联合模型的性能优于流水线模型，这是因为联合模型不仅解决了误差传播问题，还使两个子任务之间进行信息交互以辅助任务学习。另外，通过一个模型解决问题可以减小训练与预测时的差距。

⑵ 远程监督方法

Mintz等人[22]首次使用远程监督方法进行关系抽取。远程监督方法的基本假设是：如果一对实体在远程知识库存在某种关系，那么在数据集里所有包含这对实体的句子中，这对实体之间都会具有这种关系。基于远程监督的方法，可以利用远程知识库给无标签数据自动打标签，无需人工标注，节省了大量人力。但远程监督方法会带来两个问题：误差传播和噪声。对标注后的数据先进行特征提取，再训练关系抽取模型，会带来特征误差的传播；使用远程知识库标注数据会使大量实体对间的关系被标注错误，成为噪声。

深度学习中的分段卷积神经网络（Piecewise Convolutional Neural Network, PCNN）可以有效解决误差传播问题，PCNN以实体所在位置作为划分点，将句子分成三段分别进行池化，可以获得更多实体上下文信息。为了缓解噪声问题，Zeng等人[18]将PCNN与多示例学习结合进行关系抽取，Lin等人[23]在其基础上融入注意力机制，更加充分利用包内的信息，Ji等人[24]在PCNN与注意力机制结合的基础上，进一步加入了实体表示信息，使得分类的准确率更高。

随着深度神经网络的进一步发展，预训练模型（Devlin等人[1]）被提出，并凭借优异的性能刷新了多项NLP任务上的记录。许多研究人员将预训练模型应用在关系抽取任务上，提出了许多效果优良的模型，如MTB[25]、SPANBERT[26]、R-BERT[27]等等，使用预训练模型进行关系抽取逐渐成为主流范式。

1.3 本文主要研究内容及组织结构

1.3.1 研究内容

通过分析关系抽取任务的研究背景和研究现状可知，基于预训练模型的方法是目前的主流范式，提示学习方法对于提升预训练模型的性能有很大帮助。本文将研究如何将提示学习应用到基于预训练模型的关系抽取模型中，以获得更好的性能和泛化能力，并加入内部知识来改进模型。具体内容有：

⑴ 构建一个基于预训练模型的关系抽取模型，同时引入提示学习方法以增强模型的性能；

⑵ 针对关系抽取任务特点设计提示学习模板，将实体与关系的先验知识融入模型；

⑶ 进行实验验证，证明方法的有效性。

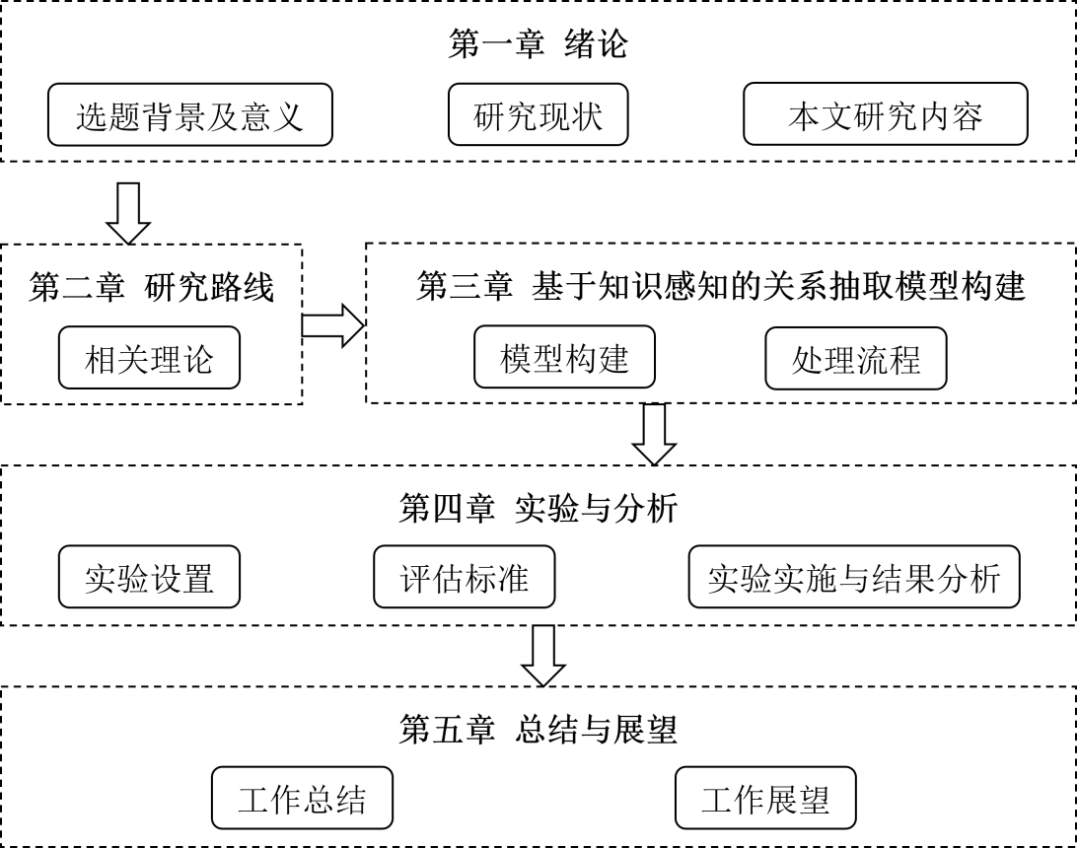


图 1.2 本文组织结构

1.3.2 组织结构

本文的组织结构如图1.2所示：第一章为绪论，介绍了本文选题背景及意义，梳理了关系抽取领域的研究现状；第二章介绍了本文的研究路线及相关理论基础；第三章详述如何构建基于知识感知并引入提示学习的关系抽取模型；第四章介绍了实验设置，展示了实验结果，并进行结果分析；第五章总结本文的工作，并对未来工作进行规划。

1. 研究路线及相关理论

对于预训练模型来说，知识感知（Knowledge-Aware）是为了增强预训练模型理解更高层次文本的能力，而将特定的外部知识引入模型的方法。特定的外部知识是指与特定任务或领域相关的知识，对于NLP任务，其外部知识包含语言、语义等多种知识，通常情况下，预训练模型通过知识图谱来引入丰富的外部知识。预训练模型从大规模未标注语料中学习到的是通用知识，由于知识感知方法引入了特定的外部知识，因而能进一步提升预训练模型的能力。

本文研究基于知识感知的关系抽取方法，该方法建立在预训练模型之上，结合提示学习方法，并将知识感知融入其中。本章首先介绍本文的研究路线，然后介绍相关的理论基础，包括预训练模型和提示学习。

2.1 研究路线

本文针对关系抽取任务设计了对应的研究路线，采用结合提示学习的预训练模型作为关系抽取的基础模型，模型的大致框架如图2.1所示。根据关系抽取框架，对输入文本进行关系抽取的关键步骤如下：

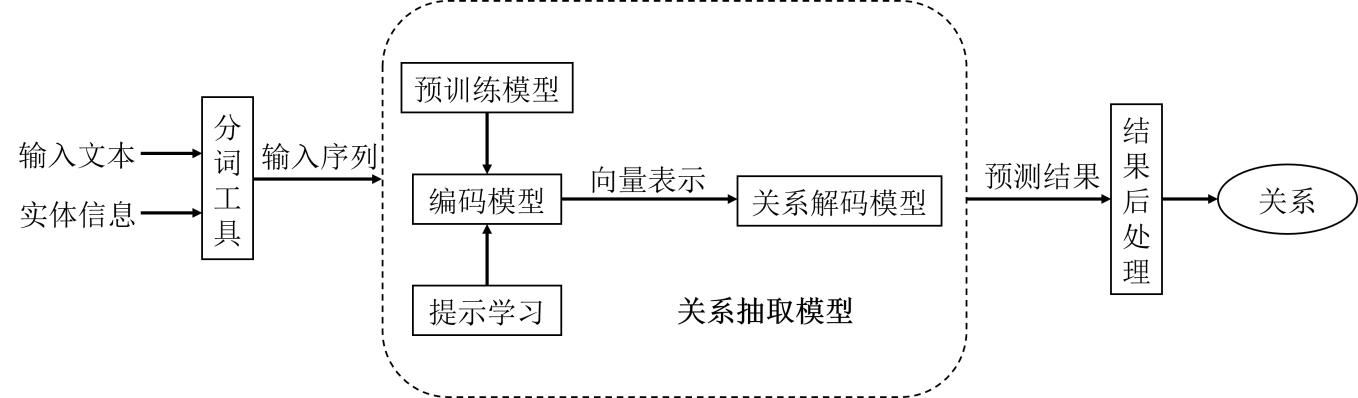


图 2.1 本文研究路线

⑴ 首先对输入文本中的实体进行标注，因为本文采用的是流水线模型，即在已知实体的情况下抽取关系，并进行关系分类。得到实体信息后，利用分词工具对标注好实体信息的文本进行处理，将连续的文本转化成模型能够识别的输入词序列，作为关系抽取模型的输入；

⑵ 关系抽取模型主要包含编码模型和关系解码模型。编码模型使用预训练模型作为编码器，并引入提示学习模块。提示学习模块针对关系抽取任务构造合适的模板和标签词集，然后将对文本分词得到的词序列和构造的模板合并，作为预训练模型的输入。之后预训练模型对输入进行编码，使其成为固定大小的、包含文本信息的向量表示，并送入关系解码模型。关系解码模型将文本的向量表示通过合适的神经网络，解码为输出的预测结果；

⑶ 关系抽取模型的预测结果不能直接作为最终结果，需要进行后处理，转化为需要的格式，如以关系三元组的形式输出，并过滤掉明显错误的结果。通过结果后处理就得到最终的关系。

2.2 预训练模型

预训练模型属于迁移学习的一种，旨在将从特定任务学习到的知识迁移至不同的下游任务中，从而提升模型在下游任务中的性能表现。预训练模型最早出现在计算机视觉领域，视觉预训练模型也被称为主干网络，一般先在大规模的图像标注数据集上进行训练，得到通用的特征提取器，再将其用于不同的下游任务，以提取更加良好的图像特征表示。但在自然语言处理领域，对于一些语法、语义的相关任务，人工标注成本很高，难以构建大规模标注数据集用于训练。

自监督预训练任务的引入解决了这一问题：自然语言处理领域的无标注数据资源非常丰富，大规模未标注语料库较容易构建，因此可以先从未标注语料中学习良好的语言表示，再将这些表示用于其他任务。基于自监督预训练的方式，许多大型预训练（语言）模型被提出，比如BERT、GPT等等。BERT[1]是2018年由谷歌提出的一个经典语言预训练模型，刷新了当时11项NLP任务上的性能记录。BERT旨在训练一个通用的语言表征模型（Language Representation Model），其核心思想是基于Transformer的双向编码器预训练。为了更好的学习大规模语料库中的知识，BERT设计了两个预训练任务，分别是掩码语言模型（Masked Language Model, MLM）和“下一句预测”（Next Sentence Prediction, NSP）。MLM的目标是使表征能够融合上下文语境，而NSP任务可以训练模型理解两个句子之间关系的能力。BERT模型一经提出，迅速席卷了自然语言处理领域，被广泛用于各种文本相关下游任务中，目前的许多研究工作均采用BERT作为编码模型。

大量工作表明，在大规模语料库上训练的预训练模型可以学习到通用的语言表示，因此对自然语言处理任务十分有帮助。早期NLP任务中，文本特征表示通常为离散表示，如one-hot向量表示，这导致文本特征维度爆炸，且任意两个词的向量表示均正交，无法衡量相似度。而预训练模型可以学习词的分布式表示（又叫“词嵌入”），即使用一个低维实值向量来表示该词，并使得两个词之间的相似度可以用对应的词嵌入来计算。根据学习到的词嵌入的不同，可将预训练模型的发展大致分为两个阶段：

第一代预训练模型的目标是学习上下文无关的词嵌入，将词映射为静态词向量，即词向量的表示不会随着上下文的变化而改变，代表模型有Word2Vec、Glove。但这种上下文无关的词向量表示存在显著的缺陷——无法解决自然语言中常见的一词多义问题，因此阻碍了预训练模型的发展；第二代预训练模型的目标为学习上下文相关的词嵌入，根据上下文的语境不同，相同的词会被映射为不同的词向量，代表模型有ELMo、BERT。

基于预训练模型的研究思路通常是“预训练+微调”，这也被称为NLP领域的“第三范式”（“第一范式”和“第二范式”分别为完全监督机器学习、完全监督深度学习）。在“预训练+微调”范式下，处理一个任务主要分为两步：首先从大规模语料中训练出一个模型（即预训练模型）；然后根据具体的下游任务对预训练模型进行微调，不断更新模型参数以使模型在该任务上达到最优的性能。预训练模型并非针对特定任务训练，而是从大量数据中学习通用的知识，保存在预训练模型中，因此具有良好的可扩展性。

预训练模型开启了自然语言处理的新篇章，随着预训练模型的发展，其在模型结构、研究任务、应用领域等方面不断扩展，训练方法也在不断创新，极大地促进了自然语言处理的研究与应用。

本文使用BERT的改进版RoBERTa作为编码模型，它相比BERT采用了更大的预训练数据、更长的训练时间以及进行了一些其他的修改。

2.3 提示学习

自从BERT、GPT等预训练模型被提出后，采用预训练模型完成NLP任务迅速成为主流。然而，基于“预训练+微调”范式处理NLP任务存在一些问题。首先，预训练模型需要针对不同的下游任务单独进行微调，且微调任务与预训练任务间存在差异；其次，微调的成本随着预训练模型的增大而提高，为了一个特定的任务去微调预训练模型会造成资源的浪费；最后，“预训练+微调”范式在零样本或少样本的设置下学习能力较差。

后来，研究发现只需将自然语言提示（prompt）和输入文本一起输入到预训练模型中，就可以在不改变原本模型参数的情况下，在少样本或零样本的环境中完成任何NLP任务。这表明了预训练模型本身具有零样本、少样本学习能力，只是基于微调的方法不能很好地利用预训练模型中的知识，于是研究人员转而研究提示学习方法。提示学习致力于使下游任务向预训练的目标靠近，通过将下游任务重构为与预训练任务相同的形式，从而构建一个可应用于不同任务的统一模型。提示学习对NLP领域的发展产生了深远的影响，因此“预训练+提示+预测”的范式被称为NLP领域的“第四范式”。

在提示学习方法中，“提示”的作用是辅助预训练模型判断下游任务。提示包括一个模板和一个标签词集，以及标签词和真实类别标签间的映射。下面以情感分类任务为例阐述提示学习的处理过程，情感分类任务是一个二分类问题，目标为判断文本表达的情感是正向（positive）还是负向（negative），给定输入文本x=“I like this movie”，需要输出“positive”或“negative”。使用提示学习方法来解决这一问题，首先为其构造一个模板“[x]it is a [MASK] movie”，该模板的作用是将训练数据以自然语言的形式表述，并添加[MASK]以将其改造成与预训练任务相同的形式（例如BERT的预训练就是采用这种完形填空形式）；然后将其输入到预训练模型中，得到对应的向量表示；之后，将向量表示进行解码，预测出[MASK]位置上的标签词是“interesting”还是“boring”（这里的“interesting”和“boring”构成了标签词集）；最后将标签词映射到真实标签，完成该句的分类。

提示分为硬提示（离散提示）和软提示（连续提示）。硬提示是人工设计或自动构建的基于离散令牌（token）的提示，是人能够看懂的提示。人工构建提示要求算法工程师具备丰富的下游任务经验，而机器自动搜索到的提示具有较弱的可读性（如不符合人类的写作规范）和可解释型。对于硬提示来说，提示设计的微小差别对模型的影响非常大，因此需要仔细设计模板和标签词集。于是人们把研究目光转向软提示，软提示是机器自己学习到的提示，即将提示的构建也作为模型的一个任务来学习。具体来说是在提示模板中插入一段任务特定的tokens，这些tokens可以让机器进行微调，这种方法也叫做提示调优（Prompt tuning）。

提示学习的思想简单易懂，却能取得十分优秀的效果，尤其在零样本和少样本的情境下表现出色。提示学习蕴含的假设是预训练模型中包含的知识已经足够丰富，具备了独立解决NLP任务的能力，使用提示学习方法重构下游任务以适应预训练模型，能够激发模型中的知识，引导预训练模型完成特定的任务。由于第一代预训练模型学习的是上下文无关的词向量，这种假设可能并不成立，而BERT等上下文相关的预训练模型的出现，使得提示学习得以发挥作用。随着预训练模型的发展，提示学习的概念也在逐步延申，研究人员进一步提出了示例学习（Demonstration Learning）和上下文学习（In-Context Learning），GPT-3模型即是使用上下文学习的成功典例。

综上所述，将提示学习与预训练模型结合的思路有助于关系抽取任务，因此本文将以RoBERTa作为编码模型、结合提示学习的方法应用于关系抽取，以期望达到更好的抽取性能。同时，由于关系抽取与情感分类任务不同，是一个多分类问题，且关系标签是由多个词汇组合而成的，因此在模板及标签词的构建上需要加以调整。本文将在传统提示学习方法的基础上做出相应改进。

2.4 本章小结

本文研究基于知识感知的关系抽取方法，研究建立在预训练模型及提示学习等相关理论的基础上。本章先简要介绍了知识感知方法的概念，引出使用结合提示学习的预训练模型进行关系抽取的研究路线，然后介绍了涉及的相关理论，包括预训练模型、提示学习，这些理论支撑着模型的构建，为后续章节奠定了基础。

1. 基于知识感知的关系抽取模型

本章在现有的关系抽取方法的基础上，设计了一个基于知识感知的关系抽取模型。其中基础模型采用预训练模型，主要用于编码向量表示，并融入了知识感知的提示学习方法。本章内容首先是概述，介绍模型提出的原理依据；然后是模型的框架及构建方法；最后是使用本文模型进行关系抽取的处理流程。

3.1 概述

在过去的20多年里，关系抽取工作吸引了许多研究人员，在关系抽取发展的各个阶段，提出了许多富有成效的模型，如早期基于模式匹配的方法，基于CNN/RNN的方法和基于图的方法。其中，深度学习的发展极大地推进了关系抽取任务的研究进程，使得关系抽取的性能得到了很大的提升。但基于深度学习的方法进行关系抽取需要为其单独构建模型，且随着参数规模的增大，对训练数据量的要求也越来越大，而有标注训练数据的稀缺会对模型的性能产生影响。

自从自然语言处理领域的预训练模型被提出后，关系抽取任务的研究又产生了新的突破。大量研究表明，基于BERT的模型性能明显优于基于CNN或基于图的模型。另外，由于预训练学习到的通用知识可能在特定任务上表现不佳，一些研究将外部知识融入预训练模型以增强模型性能，提出了一些知识增强型（Knowledge-Enhanced）的预训练模型。例如MTB[25]是一种基于BERT模型的空白匹配方法，可以从文本中学习关系模式；SPANBERT[26]使用知识来增强学习目标；KNOWBERT[28]提出将知识融入输入特征中；LUKE[29]利用知识来改进模型架构。上述知识增强型预训练模型被应用在关系抽取任务上，取得了不错的效果。

然而，预训练模型存在明显的问题，即在微调阶段需要为每个下游任务微调一个模型。为了弥补预训练和微调间的差距，研究人员提出提示学习的方法，将下游任务表述为预训练任务的形式，使用“预训练+提示+预测”的范式处理NLP任务。该方法在一些文本分类任务，如情感分类上表现出色，但由于关系抽取任务是多分类任务，应用提示学习方法进行关系抽取的研究较少。Han等人[30]提出了基于规则的多类文本分类的提示调优方法（Prompt Tuning with Rules, PTR），该方法应用逻辑规则构造具有多个子提示的提示，以编码每个类别的先验知识。

通过分析以上研究现状和各种方法的优缺点，参考Chen等人[31]的研究工作，本文提出一种具有协同优化的知识感知提示调优方法，用于关系抽取任务。该方法的核心思想是将关系标签中包含的的先验知识注入模型，并使用提示学习方法引导预训练模型，以提升模型性能。具体来说，首先依据关系标签中的先验知识，在提示模板的主体和客体周围分配实体类型词，来表示主客体的实体类型；然后将原始输入及带有屏蔽位（[MASK]）的模板输入到预训练模型中，预测[MASK]位置的关系类型词；再使用适当的损失函数协同优化实体类型词和关系类型词的表示。

3.2 模型构建

关系抽取任务目标为从文本中抽取出正确的关系三元组，首先给出关系抽取的符号表示：给定一个输入句子，其中和分别对应主体和客体，关系抽取任务需要抽取出主体和客体之间的关系，其本质上是一个多分类问题。

由于关系抽取任务是一个多分类问题，如果直接使用预训练模型，则预训练和微调阶段的训练目标不同。具体来说，预训练阶段是完形填空任务，而微调时是分类任务，这个差距会影响预训练模型中的知识向分类任务转移，从而降低模型性能。因此，本文在预训练模型的基础上引入提示调优方法，通过构造提示模板将关系抽取任务也转化为完形填空任务，弥补预训练和微调间的差距。

然而，关系抽取任务不同于其他文本分类任务。一方面，为关系抽取任务确定合适的模板是较为困难的，需要专业领域的知识，并且自动生成高质量的模板需要额外的生成和验证的计算成本；另一方面，因为关系标签的长度不一致，所以关系标签词搜索的计算复杂度很高，而且在词汇表中获取一个合适的标签词来表示特定的关系标签并非易事，例如关系标签“country\_of\_residence”很难用词汇表中的某个词来表示。此外，关系标签中存在丰富的语义知识，关系三元组中的结构化知识也不容忽视。

因此，本文将模板简化为仅包含主体、客体及屏蔽位[MASK]的句子。例如，输入到预训练模型中的句子为“[CLS][*E1*]Obama[*/E1*] is the first black president of [*E2*]the United States[*/E2*].[SEP]”，其中主体和客体分别用[*E1*]和[*E2*]标注，[CLS]和[SEP]为BERT的特殊标记，依据本文方法为其构造的模板为“Obama[MASK]the United States”。另外，在结合提示调优的预训练模型的基础上，本文进一步融入了基于知识感知的协同优化方法，与其他知识感知的方法不同，本文关注内部知识而非引入外部知识。主要体现在以下两方面：

第一，关系标签与实体类型之间会相互约束。关系标签中包含丰富的语义知识，假设给定一个关系标签“country\_of\_birth”，可以大致推断出它的主体类型属于person，客体类型属于country；反过来，如果能够判断出主客体的类型分别属于person和country，就会对抽取的关系进行约束，减少将其分类为不合适的关系标签，如“date\_of\_birth”的概率。因此，对应于提示调优方法中的模板和标签词，本文设计了可学习的实体类型词和关系类型词，分别表示实体类型和关系标签，将关系标签内部的知识融入到提示的构建中。在训练过程中，实体类型和关系之间的相互约束，使得实体类型词和关系类型词的表示不断优化，越来越接近于真实的实体类型和关系标签，这就是协同优化。

第二，关系三元组（主体，关系，客体）之间隐含着结构化知识，通过计算主体、关系和客体间的距离，可以判断抽取出的关系三元组是否正确。具体来说，在向量的表示空间上，一个三元组包含的主体、客体、关系三者的向量表示存在以下关系：主体的向量表示与关系的向量表示之和，应该等于客体的向量表示，可以此作为评价抽取的关系三元组质量的标准。这种约束关系可以看作一种隐式的结构化知识，本文通过引入该隐含的结构约束，可以更好地协同优化实体类型词和关系类型词的表示。

综上所述，本文在提示调优的基础上，进一步利用关系标签内的先验知识及关系三元组的结构化知识，提出了具有协同优化的知识感知提示调优方法。该方法主要包含三个模块：实体类型词模块、关系类型词模块、结构化约束模块。相比传统的提示调优方法，该方法主要做出以下改进：

（1）在模板中引入了可学习的实体类型词和关系类型词。不同于传统提示调优方法使用的提示词，本文使用的实体类型词和关系类型词不是真正存在于词汇表中的词，而是向量表示，因此是可以让模型学习、优化的。其中，实体类型词是指表示实体类型的嵌入向量，而关系类型词是指语义接近关系标签的嵌入向量，能更好的融入关系标签中的先验知识。

（2）协同优化实体类型词和关系类型词的表示。协同优化时既考虑实体类型词和关系类型词与上下文的交互，也考虑到关系三元组间隐含的结构化损失，并通过适当的损失函数来优化实体类型词和关系类型词的表示，使得它们表达的语义越来越接近于真实实体类型和关系标签。

3.3 处理流程

根据上述构建的模型框架，当输入一个句子时，将经过以下流程进行处理，以得到最终预测的关系。

⑴ 实体类型词和关系类型词的初始化

**A、实体类型词**

实体类型词即表示实体类型的嵌入向量，通过在实体周围分配实体类型词，可以将实体知识注入到模型中，以约束关系预测的范围。具体做法为：首先依据关系标签中的先验知识，得到潜在实体类型的候选集和，通过频率统计量来估计实体类型的先验分布和，即每种实体类型在候选集中所占比例作为先验分布。然后根据实体类型的先验分布在实体类型候选集上进行聚合嵌入，得到实体类型词的初始化向量。

(1)

(2)

其中和表示围绕主体和客体的实体类型词的嵌入向量，和表示去重后的实体类型集合，是预训练模型的词嵌入层。

初始化完成后，实体类型词将会被分配到模板中主体和客体实体的两边，以3.2节中的输入为例，输入句子为“[CLS][*E1*]Obama[*/E1*] is the first black president of [*E2*]the United States[*/E2*].[SEP]”，则最终的模板为“[*sub*]Obama[*sub*][MASK][*obj*]the United States[*obj*]”。其中，[*sub*]和[*obj*]分别为表示当前句子中主体和客体类型的实体类型词，是可学习的嵌入向量。分配给每对主客体的实体类型词的初始化嵌入向量相同，后续将根据上下文进行优化调整。

**B 关系类型词**

以往关于提示学习的研究通常将词汇表中的一个标签词与任务标签进行一对一映射，使得搜索过程的计算复杂度较大，且无法利用关系标签中丰富的语义知识。因此，本文假设在预训练模型的词汇空间中有一个关系类型词，可以表示该关系标签的隐含语义。也就是说，关系类型词是语义接近关系标签的嵌入向量，而非词汇表中真实存在的词汇，这样就使得关系类型词中注入了关系知识。

关系类型词初始化的具体做法为：首先将关系标签拆分成单词集（关系标签通常为单词聚合而成的，如founded\_by是由两个单词连接组成的），然后根据单词的概率分布来初始化该关系标签对应的关系类型词。

(3)

其中，为对应于关系标签的关系类型词的表示，是关系标签拆分后的单词集，是单词集上的先验分布，是预训练模型的词嵌入层。

初始化后的关系类型词将被拓展为模板的答案空间，相当于标签词集，后续将根据上下文与实体类型词一同进行优化。

⑵ 协同优化

对于关系抽取任务来说，实体类型和关系标签之间存在不容忽视的密切交互和联系，且构建的实体类型词和关系类型词需要与上下文相关联，因此本文提出一种具有结构化约束的协同优化方法，来优化实体类型词和关系类型词的表示。

**A、上下文校准**

虽然实体类型词和关系类型词是根据知识进行初始化的，但它们在潜在变量空间可能不是最优的，需要与周围的上下文联系。因此，本文使用上下文来校准它们的表示。具体来说，给定屏蔽位[MASK]上预测的概率分布，将y与之间的交叉熵作为损失函数来优化实体类型词和关系类型词的表示：

(4)

其中表示训练数据集的样本数量。

**B、隐式结构约束**

此外，为了将关系三元组的结构知识注入到提示的构建中，本文使用隐式结构化约束来对实体类型词和关系类型词进行优化。一个关系可以用关系三元组(s, r, o)来表示，其中s和o分别表示主体和客体的类型，r是预定义关系类型词集中的关系标签，隐式结构约束的定义如下：

(5)

(6)

其中，s和o为预训练模型输出的实体类型词的嵌入向量，分别表示主体和客体，r是输出的关系类型词的嵌入向量。是采样的负样本，对于负采样，我们将正确的关系类型词分配在[MASK]位置，对主体实体或客体实体进行随机抽样，替换为与该关系类型不相关的实体，构造错误的三元组，也就是说对当前关系来说实体具有不可能的类型。是间隔，表示sigmoid函数，l是得分函数。

⑶ 训练过程及目标函数

为了更好的训练本文模型，将训练过程分为两阶段。第一阶段的目标是协同优化实体类型词集、和关系类型词集的表示，以得到最优的提示（模板和标签词集）。该阶段使用较大的学习率，优化以下目标函数：

(7)

其中是超参数。

在第二阶段，基于优化后的实体类型词和关系类型词，使用作为目标函数，以较小的学习率对带提示的预训练模型进行微调，优化整体参数。

3.4 本章小结

本章首先对关系抽取任务的研究进行概述，分析了关系抽取领域研究方法的优缺点，讨论了基于知识感知的关系抽取模型提出的依据和可行性；然后阐述了本文提出的具有协同优化的知识感知提示调优方法的基本原理和框架；接着仔细描述了基于本文方法的关系抽取处理流程，对实验的进行提供了理论支撑。

1. 实验与分析

本章在上一章构建的模型的基础上，使用SemEval、DialogRE等数据集进行了实验，研究了该模型的效果，并与其他的基线模型对比。结果表明，本文模型整体优于其他模型，在关系抽取任务中体现出良好的性能。

4.1 实验设置

4.1.1 实验环境设置

实验环境使用深度学习PyTorch框架，版本为1.8.1，配合Python 3.8和Cuda 11.1，并使用4个RTX3090 GPU进行训练。

4.1.2 实验数据集

实验使用的数据集有：SemEval 2010 Task 8（SemEval）、DialogRE、TACRED、TACREV、Re-TACRED、ACE2005，表4.1中列出了这些数据集的统计信息。数据集的简要介绍如下：

表 4.1 数据集统计信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据集 | 训练集 | 验证集 | 测试集 | 关系类型 |
| SemEval | 6,507 | 1,493 | 2,717 | 19 |
| DialogRE | 5,963 | 1,928 | 1,868 | 36 |
| TACRED | 68,124 | 22,631 | 15,509 | 42 |
| TACREV | 68,124 | 22,631 | 15,509 | 42 |
| Re-TACRED | 58,465 | 19,584 | 13,418 | 40 |
| ACE2005 | 2453 | 611 | 685 | 7 |

SemEval：传统的关系分类数据集，共有10717条注释样本，以及包括“no\_relation”在内的19种关系类型。

DialogRE：第一个人工注释的对话级关系抽取数据集，包含来自美国情景喜剧的1788个完整对话和36种关系类型。该数据集的特殊之处在于，每个实体对之间包含多种关系，是一个多标签分类问题。

TACRED：大规模句子级关系抽取数据集，包含从年度TACKBP4挑战中提取的超过106K个句子，定义了包含“no\_relation”在内的42种关系类型。

TACREV：基于原始TACRED构建的数据集，纠正了原始验证集和测试集中的错误，训练集保持不变。

Re-TACRED：对原始训练集、验证集、测试集进行重构后的TACRED数据集，并且将关系类型修改为40种。

ACE2005：数据集包括中文、英文和阿拉伯语三部分数据，本文使用的英文数据包含599篇文档，共7大类、25小类关系。

4.1.3 实验参数设置

对于普通基于“预训练+微调”方法的模型，以及基于“预训练+提示学习”方法的模型，为了更好的对比，均采用RoBERTa-large作为预训练模型。特别地，对于对话数据集DialogRE，使用RoBERTa-base作为预训练模型来与其他方法进行对比。

表 4.2 超参数设置

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 超参数 | SemEval | DialogRE | TACRED | TACREV | Re-TACRED | ACE2005 |
| 优化器  （Optimizer） | AdamW | | | | | |
| 学习率  （lr） | 3e-5 | 5e-5 | 4e-5 | 3e-5 | 3e-5 | 3e-5 |
| 时期  （max\_epochs） | 10 | 20 | 4 | 5 | 5 | 10 |
| 批量大小  （batch\_size） | 16 | 8 | 16 | 16 | 16 | 16 |
| 工作进程  （num\_workers） | 8 | | | | | |
| 训练权重  （t\_lambda） | 0.001 | | | | | |
| 间隔  （） | 1 | | | | | |
| 权重衰减  （weight decay） | 0.01 | | | | | |
| 最大序列长度  （max\_seq\_length） | 256 | 512 | 256 | 256 | 256 | 256 |

超参数（Hyper Parameter）的设置对模型的性能有重要的影响，例如学习率的大小会影响模型的收敛速度。本实验使用网格搜索（Grid Search）方法进行超参数的优化，每个数据集训练时设置的超参数如表4.2所示。

4.2 评估标准

在分类任务中，通常使用精确率（Precision）、召回率（Recall）和F1值（F1-score）作为评价指标来评估模型的性能，下面给出这三项指标的计算方法。

首先，根据分类结果得到混淆矩阵（Confusion Matrix），如表4.3。混淆矩阵中TP和TN表示预测值和真实值相同，即预测正确；FP和FN表示预测值和真实值不同，即预测错误。表4.4中给出了三项评价指标的计算公式及含义。

表 4.3 混淆矩阵

|  |  |  |
| --- | --- | --- |
| 真实值  预测值 | Positive | Negative |
| Positive | TP | FP |
| Negative | FN | TN |

表 4.4 评价指标计算公式及含义

|  |  |  |
| --- | --- | --- |
| 评价指标 | 计算公式 | 含义 |
| 精确率 |  | 在所有模型预测是正例的结果中，真正正例所占的比例，越高越好。 |
| 召回率 |  | 在所有真正是正例的结果中，模型预测为正例所占的比例，越高越好。 |
| F1值 |  | 综合了精确率与召回率的结果，越高越好。 |

本实验使用F1值作为评价指标，它是精确率和召回率的调和平均，能够更好地评估模型整体性能。

4.3 实验实施与结果分析

4.3.1 对比实验

为了验证本文提出的模型的性能，设计了对比实验，分为两个角度——不同模型在同一数据集下的效果对比和相同模型在不同数据集上的效果对比。实验用到的数据集有SemEval、DialogRE、TACRED、TACREV、Re-TACRED。

⑴ 不同模型实验对比

对于采用“预训练+微调”范式的模型，除了选取普通的预训练模型作为基线模型之外，考虑到实体信息对关系抽取模型很有帮助，本文选取了一些使用知识图谱作为外部知识进行增强的预训练模型进行对比，包括MTB[25]、SPANBERT[26]、KNOWBERT[28]、LUKE[29]，它们分别使用外部知识来增强预训练策略、学习目标、输入特征和模型架构。

除此之外，对于更具挑战性的数据集DialogRE，还选取了两个在其上取得最先进性能的模型GDPNet[32]和DUAL[33]来进行对比。

对于采用提示学习方法的模型，本文选取了PTR[30]作为对比的基线模型，它通过应用逻辑规则构造具有多个子提示的提示，来编码每一类别的先验知识。本文方法用K-Prompt表示，实验结果如表4.5所示。

表 4.5 对比实验结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 方法 | 引入外部知识 | SemEval | DialogRE | TACRED | TACREV | Re-TACRED |
| 预训练+微调 | | | | | | |
| Fine-Tuning | 否 | 87.6 | 57.3 | 68.7 | 76.0 | 84.9 |
| MTB | 是 | 89.5 | - | 70.1 | - | - |
| SPANBERT | 是 | - | - | 70.8 | 78.0 | 85.3 |
| KNOWBERT | 是 | 89.1 | - | 71.5 | 79.3 | 89.1 |
| LUKE | 是 | - | - | 72.7 | 80.6 | - |
| GDPNet | 否 | - | 64.9 | 71.5 | 79.3 | - |
| DUAL | 否 | - | **67.3** | - | - | - |
| 预训练+提示学习 | | | | | | |
| PTR | 否 | 89.9 | 63.2 | **72.4** | 81.4 | 90.9 |
| K-Prompt | 否 | **90.6** | 65.6 | 69.2 | **81.8** | **91.1** |

首先，从表4.5中可以看到，知识增强的预训练模型MTB、SPANBERT、KNOWBERT、LUKE在所有的数据集上的结果全面优于普通的“预训练+微调”模型，这表明在预训练模型中融入特定于任务的知识是可行的，使得模型更加专业化，从而提升模型在特定任务上的性能。

其次，本文方法K-Prompt相比知识增强型预训练模型取得了更好的性能，这表明了应用提示学习的必要性。即使知识增强型预训练模型中已经融入了任务特定的知识，由于预训练和微调任务形式的不同，在微调阶段这些知识也难以应用到下游任务中。因此引入提示学习是一个很好的改进方法，本文提出的模型结合了简单有效的提示调优方法，从而取得了更好的结果。

与本文方法类似，PTR也使用了提示学习，并融入了关系标签的先验知识，不同的是，PTR将各种条件函数的子提示组合成一个完整的任务特定提示。实验表明，与PTR进行对比，本文方法依旧表现出更好的性能。分析发现本文方法优于PTR的原因有：第一，PTR采用包含多个词的关系标签形式，需要手动制定规则，花费人力，而K-Prompt设计的关系类型词只包含一个词，避免了劳动密集型工作；第二，在多标签分类数据集DialogRE上，本文方法相比PTR提高了2.4%，这可能是由于PTR使用预测多个[MASK]的规则会混淆多标签预测。

⑵ 不同数据集实验对比

本次实验共使用5个数据集，每个模型在这5个数据集上的表现都存在差异。如表4.5所示，在传统关系抽取数据集SemEval上，所有模型都取得了较好的效果。模型在TACRED数据集上的性能较差，可能是因为该数据集中存在错误，但在修改后的TACREV和Re-TACRED数据集上表现良好，尤其是在Re-TACRED上表现最佳。由于DialogRE是一个对话数据集，实体对间存在多种关系，是关系抽取任务中的难点，因此在该数据集上所有模型的性能最低。

4.3.2 ACE2005实验

通过上述对比实验，可以看出本文提出的模型效果较好。除了上述实验，本文还将构造的模型K-Prompt应用在ACE2005英文数据集上进行关系抽取，并在全监督设置和少样本设置下得到了实验结果，实验结果如图4.1所示。图中，K-shot中的K表示少样本设置下每个类别中的实例数目，full-shot既全监督设置。

图表, 折线图

描述已自动生成

图 4.1 ACE2005实验结果

从实验结果可以看出，在全监督设置下，本文模型达到了91.8的F1分数，体现出本文模型在ACE2005数据集上进行关系抽取任务的优异性能。在少样本设置下，K=8时的效果较差，只取得了30.0的F1分数，随着K的增加，关系抽取的效果也逐渐变好，当K=32时，F1分数达到了65.8。整体来说，本文提出的模型在全监督设置和少样本设置下的关系抽取性能均较为良好。

4.4 本章小结

基于具有协同优化的知识感知提示调优方法，本文构造了关系抽取模型K-Prompt，为了验证该模型在关系抽取任务上的实际性能，本章设计了相应的实验，从两个角度进行了研究：不同模型在同一数据集上的性能对比，同一模型在不同数据集上的性能对比。实验表明，本文的模型整体上优于其他关系抽取模型，通过与普通预训练模型、知识增强型预训练模型和“预训练+提示学习”模型进行对比，分别证明了知识感知、提示学习和K-Prompt方法的有效性。从数据集来看，模型在常规的关系抽取数据集SemEval、TACREV和Re-TACRED上表现出良好的性能，而在对话数据集DialogRE上表现较差。另外，本文还使用提出的K-Prompt模型在ACE2005数据集上进行关系抽取，并取得了优异的性能。

1. 总结与展望

5.1 工作总结

关系抽取任务具有重要的研究意义，本文针对关系抽取领域的研究现状，分析了现有方法的优缺点，提出了一个基于知识感知的、具有协同优化的提示调优关系抽取方法，并通过实验证明了该方法在关系抽取任务中的优异性能。具体来说，本文的研究工作主要有以下几点：

⑴ 将提示调优方法与预训练模型结合，消除预训练与微调间的差距，更好的利用预训练模型中分布的知识，以提升模型性能；

⑵ 为了将关系标签内的先验知识及关系三元组的结构化知识融入模型中，设计了可学习的实体类型词和关系类型词，并通过上下文校准和结构化约束来协同优化它们的表示。

5.2 工作展望

本文对关系抽取方法进行研究，提出了关系抽取模型K-Prompt，并设计了相关实验，取得了一些成果。基于目前的研究，发现未来的探索方向有以下方面：

⑴ 扩展到半监督设置。本文实验环境为全监督设置，所有的数据都带有关系标签，然而实际情况下可能不满足这一条件，会由于各种限制导致缺少数据标签。因此，研究如何将本文方法扩展到半监督设置，以利用大量未标注数据，将具有重要的意义。

⑵ 提示的优化。结合提示学习是本文模型取得优异性能的关键因素，研究如何优化提示的构建是未来的一个探索方向，可能有助于模型性能的提升。

参考文献

1. Jacob Devlin, Ming-Wei Chang, Kenton Lee, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]. NAACL-HLT 2019: 4171-4186.
2. Tom B. Brown, Benjamin Mann, Nick Ryder, et al. Language Models are Few-Shot Learners[J]. NeurIPS 2020, 33: 1877-1901.

1. [Mary Elaine Califf](https://dblp.org/pid/37/6181.html), [Raymond J. Mooney](https://dblp.org/pid/m/RaymondJMooney.html). Relational Learning of Pattern-Match Rules for Information Extraction[C]. [AAAI/IAAI 1999](file:///\\文字文稿1#CaliffM99): 328-334.

1. [Scott Miller](https://dblp.org/pid/18/1985.html), [Heidi Fox](https://dblp.org/pid/93/4131.html), [Lance A. Ramshaw](https://dblp.org/pid/02/767.html), et al. A Novel Use of Statistical Parsing to Extract Information from Text[C]. [ANLP 2000](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#MillerFRW00): 226-233.
2. 邓擘, 樊孝忠, 杨立公. 用语义模式提取实体关系的方法[J]. 计算机工程, 2007, (10): 212-214.
3. Nanda Kambhatla. Combining Lexical, Syntactic, and Semantic Features with Maximum Entropy Models for Information Extraction[C]. [ACL (Poster and Demonstration) 2004](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#Kambhatla04): 22-26.
4. Xia Sun, Lehong Dong. Feature-based approach to chinese term relation extraction[C]. International Conference on Signal Processing Systems 2009: 410-414.

1. [Dmitry Zelenko](https://dblp.org/pid/75/6783.html), [Chinatsu Aone](https://dblp.org/pid/76/5955.html), [Anthony Richardella](https://dblp.org/pid/24/2476.html). Kernel Methods for Relation Extraction[C]. [EMNLP 2002](file:///\\文字文稿1#ZelenkoAR02): 71-78.

1. [Guodong Zhou](https://dblp.org/pid/42/6620.html),[Longhua Qian](https://dblp.org/pid/38/7161.html),[Jianxi Fan](https://dblp.org/pid/93/2399.html). Tree kernel-based semantic relation extraction with rich syntactic and semantic information[J].[Information Sciences 2010 180(8)](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#ZhouQF10): 1313-1325.

1. [Sergey Brin](https://dblp.org/pid/b/SergeyBrin.html).Extracting Patterns and Relations from the World Wide Web[C].[WebDB 1998](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#Brin98): 172-183.

1. [Eugene Agichtein](https://dblp.org/pid/26/1185.html),[Luis Gravano](https://dblp.org/pid/g/LuisGravano.html). Snowball: extracting relations from large plain-text collections[C].[ACM DL 2000](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#AgichteinG00): 85-94.

1. [Maria-Florina Balcan](https://dblp.org/pid/b/MariaFlorinaBalcan.html), [Avrim Blum](https://dblp.org/pid/b/AvrimBlum.html), [Ke Yang](https://dblp.org/pid/80/4136.html). Co-Training and Expansion: Towards Bridging Theory and Practice[J]. [NIPS 2004](file:///\\文字文稿1#BalcanBY04): 89-96.

1. [Xiaojin Zhu](https://dblp.org/pid/z/XiaojinZhu.html), [Zoubin Ghahramani](https://dblp.org/pid/g/ZoubinGhahramani.html), [John D. Lafferty](https://dblp.org/pid/46/6823.html). Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions[C]. [ICML 2003](file:///\\文字文稿1#ZhuGL03): 912-919.

1. [Takaaki Hasegawa](https://dblp.org/pid/65/3503.html), [Satoshi Sekine](https://dblp.org/pid/10/163.html), [Ralph Grishman](https://dblp.org/pid/32/3175.html). Discovering Relations among Named Entities from Large Corpora[C]. [ACL 2004](file:///\\文字文稿1#HasegawaSG04): 415-422.

1. [Binyamin Rosenfeld](https://dblp.org/pid/07/5233.html), [Ronen Feldman](https://dblp.org/pid/f/RonenFeldman.html). High-Performance Unsupervised Relation Extraction from Large Corpora[C]. [ICDM 2006](file:///\\文字文稿1#RozenfeldF06): 1032-1037.

1. [Richard Socher](https://dblp.org/pid/79/128.html), [Brody Huval](https://dblp.org/pid/117/4058.html), [Christopher D. Manning](https://dblp.org/pid/m/ChristopherDManning.html), et al. Semantic Compositionality through Recursive Matrix-Vector Spaces[C]. [EMNLP-CoNLL 2012](https://dblp.org/db/conf/emnlp/emnlp2012.html#SocherHMN12): 1201-1211.

1. [Kazuma Hashimoto](https://dblp.org/pid/76/2653.html), [Makoto Miwa](https://dblp.org/pid/29/456.html), [Yoshimasa Tsuruoka](https://dblp.org/pid/18/3787.html),et al. Simple Customization of Recursive Neural Networks for Semantic Relation Classification[C]. [EMNLP 2013](file:///\\文字文稿1#HashimotoMTC13): 1372-1376.

1. [Daojian Zeng](https://dblp.org/pid/133/1954.html), [Kang Liu](https://dblp.org/pid/42/4903.html), [Siwei Lai](https://dblp.org/pid/121/4224.html), et al. Relation Classification via Convolutional Deep Neural Network[C]. [COLING 2014](file:///\\文字文稿1#ZengLLZZ14): 2335-2344.

1. [Yan Xu](https://dblp.org/pid/03/4702-13.html), [Lili Mou](https://dblp.org/pid/127/0779.html), [Ge Li](https://dblp.org/pid/24/712-1.html), et al. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths[C]. [EMNLP 2015](file:///\\文字文稿1#XuMLCPJ15): 1785-1794.

1. [Makoto Miwa](https://dblp.org/pid/29/456.html), [Mohit Bansal](https://dblp.org/pid/32/5243.html). End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures[C]. [ACL (1) 2016](file:///\\文字文稿1#MiwaB16).

1. [Suncong Zheng](https://dblp.org/pid/133/2598.html), [Feng Wang](https://dblp.org/pid/90/4225-23.html), [Hongyun Bao](https://dblp.org/pid/34/10083.html), et al. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme[C]. [ACL (1) 2017](file:///\\文字文稿1#ZhengWBHZX17): 1227-1236.

1. [Mike Mintz](https://dblp.org/pid/50/9013.html), [Steven Bills](https://dblp.org/pid/65/9013.html), [Rion Snow](https://dblp.org/pid/42/1066.html), et al. Distant supervision for relation extraction without labeled data[C]. [ACL/IJCNLP 2009](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#MintzBSJ09): 1003-1011.

1. [Yankai Lin](https://dblp.org/pid/161/0001.html), [Shiqi Shen](https://dblp.org/pid/169/3386.html), [Zhiyuan Liu](https://dblp.org/pid/53/3245-1.html), et al. Neural Relation Extraction with Selective Attention over Instances[C]. [ACL (1) 2016](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#LinSLLS16).

1. [Guoliang Ji](https://dblp.org/pid/151/1438.html), [Kang Liu](https://dblp.org/pid/42/4903.html), [Shizhu He](https://dblp.org/pid/136/8650.html), et al. Distant Supervision for Relation Extraction with Sentence-Level Attention and Entity Descriptions[C]. [AAAI 2017](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#Ji0H017): 3060-3066.

1. [Livio Baldini Soares](https://dblp.org/pid/178/3562.html), [Nicholas FitzGerald](https://dblp.org/pid/85/9686.html), [Jeffrey Ling](https://dblp.org/pid/157/8170.html), et al. Matching the Blanks: Distributional Similarity for Relation Learning[C]. [ACL (1) 2019](file:///\\文字文稿1#SoaresFLK19): 2895-2905.

1. [Mandar Joshi](https://dblp.org/pid/85/1261.html), [Danqi Chen](https://dblp.org/pid/87/7949.html), [Yinhan Liu](https://dblp.org/pid/238/0128.html), et al. SpanBERT: Improving Pre-training by Representing and Predicting Spans[J]. [Trans. Assoc. Comput. Linguistics 8](file:///\\文字文稿1#JoshiCLWZL20) 2020: 64-77.

1. [Shanchan Wu](https://dblp.org/pid/16/2030.html), [Yifan He](https://dblp.org/pid/70/6707.html).Enriching Pre-trained Language Model with Entity Information for Relation Classification[C]. [CIKM 2019](file:///C:\Users\z1513\Desktop\新建%20DOCX%20文档.docx#WuH19a):2361-2364.

1. [Matthew E. Peters](https://dblp.org/pid/48/9898.html), [Mark Neumann](https://dblp.org/pid/37/6418.html), [Robert L. Logan IV](https://dblp.org/pid/210/2652.html),et al. Knowledge Enhanced Contextual Word Representations[C]. [EMNLP/IJCNLP (1) 2019](file:///\\文字文稿1#PetersNLSJSS19): 43-54.

1. [Ikuya Yamada](https://dblp.org/pid/150/1853.html), [Akari Asai](https://dblp.org/pid/213/8066.html), [Hiroyuki Shindo](https://dblp.org/pid/97/8317.html), et al. LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention[C]. [EMNLP (1) 2020](file:///\\文字文稿1#YamadaASTM20): 6442-6454.

1. [Xu Han](https://dblp.org/pid/19/3011-7.html), [Weilin Zhao](https://dblp.org/pid/197/5702.html), [Ning Ding](https://dblp.org/pid/04/4910.html), et al. PTR: Prompt Tuning with Rules for Text Classification[J]. [CoRR abs/2105.11259](file:///\\文字文稿1#abs-2105-11259) (2021).
2. Xiang Chen, Ningyu Zhang, Xin Xie, et al. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction[C]. Proceedings of the ACM Web Conference 2022: 2778-2788.

1. [Fuzhao Xue](https://dblp.org/pid/248/1245.html), [Aixin Sun](https://dblp.org/pid/78/5155.html), [Hao Zhang](https://dblp.org/pid/55/2270-48.html), et al. GDPNet: Refining Latent Multi-View Graph for Relation Extraction[C]. [AAAI 2021](https://dblp.org/db/conf/aaai/aaai2021.html#XueSZC21): 14194-14202.

1. [Xuefeng Bai](https://dblp.org/pid/18/2759-1.html), [Yulong Chen](https://dblp.org/pid/157/4604-1.html), [Linfeng Song](https://dblp.org/pid/136/3610.html), et al. Semantic Representation for Dialogue Modeling[C]. [ACL/IJCNLP (1) 2021](https://dblp.org/db/conf/acl/acl2021-1.html#BaiCS020): 4430-4445.

致 谢

回顾我的本科毕业论文的撰写历程，有太多的人对我给予了帮助、关心和支持，使我能够成功完成我的研究工作。

首先，我要感激我的导师赵志立副教授和胡艳丽副教授。他们在整个毕业设计的研究阶段，给我提供了高质量的指导和鼓励，并且不断地为我提供观点和建议，让我更加深入地理解我研究的课题，尽力保证我完成了自己的研究。

其次，我向一些师兄、师姐表示感谢。在我的研究中，他们给予了我很大的帮助和鼓舞。他们分享了他们的经验和知识，并提供了一些实用的方法和技巧，以促进我的研究更好地进行，帮助我度过了研究期间的艰难岁月。

此外，我还要感谢我的家人和朋友。他们在生活和学习中给了我无私的鼓励和支持。没有他们，在繁琐的实验、紧张的论文答辩等过程中，我将少了很多关键的精神力量和帮助。

最后，我要向我的祖国和人民表示感激。是我们伟大的祖国给予了我这样的发展机会，也是中国优秀的教育培养我成为一名有用之才。本次研究所用到的所有材料和方法，都来源于国内外不同领域的前沿技术。

总而言之，没有这些人士的支持、期待和鼓励，我的本科学业将毫无意义。愿您们在此得到表彰和荣誉，同时也接受我真挚的感激之情！

毕业论文（设计）成绩表

|  |
| --- |
| 导师评语  **建议成绩 指导教师（签字）** |
| **答辩委员会意见**  **答辩委员会负责人（签字）** |
| **成绩 学院（盖章）**  **年 月 日** |