中图分类号：TP391　　单位代号：10280
密　　级：公开　　学　　号：22721587

# 上海大学

# 硕士学位论文

# SHANGHAI UNIVERSITY
# MASTER'S DISSERTATION

| 题目 | 基于对比学习与知识迁移的信息抽取方法研究 |
| --- | --- |

作　　者　凌晨帆
学科专业　计算机应用技术
导　　师　张瑞
完成日期　二〇二五年四月

姓　　名：凌晨帆　　学号：22721587

论文题目：基于对比学习与知识迁移的信息抽取方法研究

# 上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学硕士学位论文质量要求。

答辩委员会签名：

主　　席：

委　　员：

导　　师：

答辩日期：2025年 8月 13日

姓　　名：凌晨帆　　学号：22721587

论文题目：基于对比学习与知识迁移的信息抽取方法研究

## 上海大学学位论文原创性声明

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。除了文中特别加以标注和致谢的内容外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他研究者对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：凌晨帆

日期：2025 年 8 月 13 日

## 上海大学学位论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

**(保密的论文在解密后应遵守此规定)**

学位论文作者签名：凌晨帆　　导师签名：张瑞

日期：2025 年 8 月 13 日　　日期：2025 年 8 月 13 日

上海大学工学硕士学位论文

# 基于对比学习与知识迁移的信息抽取方法研究

作　　者: 凌晨帆

导　　师: 张瑞

学科专业: 计算机应用技术

计算机工程与科学学院

上海大学

2025 年 4 月

A Dissertation Submitted to Shanghai University for the
Degree of Master in Engineering

# A Study of Information Extraction Methods Based on Contrastive Learning and Knowledge Transfer

Candidate: Chenfan Ling
Supervisor: Rui Zhang
Major: Computer Application Technology

**School of Computer Engineering and Science**
**Shanghai University**
**April, 2025**

# 摘　要

信息抽取是从海量非结构化文本中自动获取结构化知识的关键技术之一，其核心任务包括命名实体识别与关系抽取。前者旨在定位并分类文本中的特定实体，后者则致力于识别这些实体间存在的语义关系。在信息抽取流水线中，两个子任务串行执行。然而，任务间有限的交互导致关系抽取任务未能充分利用命名实体识别过程中的丰富特征信息，从而制约了性能进一步提升。同时，作为自监督表示学习领域的一项重要进展，对比学习的出现尽管为知识迁移提供了有力的技术支持，但在应用于命名实体识别任务时，现有方法在实体跨度特征表示上存在不足。这不仅影响了实体识别的效果，也难以有效支撑后续知识向关系抽取的迁移。为了解决这些问题，本文具体工作如下：

（1）针对现有命名实体识别方法在长跨度表示上语义信息不足的问题，提出了一种基于偶数卷积的跨度表示增强模块。该模块有效捕捉了实体跨度内部的语义信息，增强了长跨度的特征表示，进而提升了模型对长跨度的识别效果。

（2）针对现有基于对比学习的命名实体识别框架对比维度单一的问题，引入了基于跨度和类别的双重对比学习目标，与仅对于跨度维度进行优化的方法相比，本文方法学习到了对实体类别更具区分性的表示。同时，在类别端采用轻量级可学习嵌入层代替预训练语言模型，在保证性能的同时提升了模型效率。

（3）针对信息抽取方法中子任务间交互不足的局限性，提出了一种知识迁移方法，旨在将基于对比学习的命名实体识别模型参数迁移至关系抽取模型。具体而言，该方法通过复用命名实体识别阶段训练的文本编码器，使得模型能够更有效地编码实体上下文信息。同时，引入基于类别编码器迁移的类别感知融合模块，强化了关系抽取阶段对实体类别的感知能力，这两项迁移机制的协同作用增强了实体关系抽取的性能。

经过实验验证，本文研究不仅显著提升了信息抽取的整体性能，而且深化了信息抽取子任务间的知识交互与共享，为信息抽取领域贡献了新的思路与优化路径。

**关键词**：对比学习；命名实体识别；关系抽取；迁移学习；信息抽取

# ABSTRACT

Information extraction is a key technique for automatically obtaining structured knowledge from large volumes of unstructured text, and its core tasks include named entity recognition and relation extraction. Named entity recognition locates and classifies specific entities within the text, while relation extraction identifies the semantic relationships between those entities. In a typical information extraction pipeline, these two subtasks are executed sequentially, but limited interaction between them hinders the relation extraction task from fully leveraging the rich representations produced by the named entity recognition process, thereby constraining performance improvements. Furthermore, while contrastive learning offers robust support for knowledge transfer as a significant advancement in self-supervised representation learning, its application to the named entity recognition task is hampered by the inadequacy of existing methods in representing entity span features. This inadequacy not only impairs the performance of entity recognition itself but also hinders the effective transfer of knowledge to the downstream relation extraction task. To address these issues, the specific contributions of this paper are as follows:

(1) To address the insufficient semantic information in long-span representations of existing contrastive learning-based named entity recognition methods, a span representation enhancement module based on even-sized convolution is proposed. The module effectively captures the semantic information inside the entity span and enriches the feature representations of the long span, which improves the recognition performance on the long spans.

(2) To address the limitation of a single contrastive dimension in the existing contrastive learning-based named entity recognition framework, a dual contrastive learning objective based on span-level and type-level is introduced. The approach enables the model to learn more discriminative entity type representations compared to methods that optimize for spans alone. Meanwhile, a lightweight learnable embedding layer is used instead of a pre-trained language model at the type end, which improves the model efficiency while maintaining the performance.

(3) To address the limitation of insufficient interaction between subtasks in information extraction methods, a knowledge transfer method is proposed, which aims to transfer the features from the contrastive learning-based named entity recognition model to the relation extraction model. Specifically, the method enables the model to encode entity context information more efficiently by reusing the text encoders trained in the named entity recognition stage. Furthermore, a type-aware fusion module transferred from the type encoder is introduced to strengthen the ability to perceive entity types in the relation extraction phase. The combined effect of these two transfer mechanisms significantly improves relation extraction performance.

Through experimental validation, this study not only markedly enhances the overall performance of information extraction but also strengthens knowledge interaction and sharing among its subtasks, providing novel insights and optimization pathways for the field.

**Keywords**: Contrastive Learning; Named Entity Recognition; Relation Extraction; Transfer Learning; Information Extraction

# 目 录

# 第一章　绪论

## 1.1　课题来源

本课题得到国家重点研发计划（编号:2022YFB3707800）、国家自然科学基金（编号:52273228）、云南省科技攻关计划（编号:202302AB080022）、上海市科技青年人才扬帆计划（编号:23YF1412900）的资助。

## 1.2　课题背景概述

人类社会正迈入信息爆炸的时代，全球数据的产生速度和规模正以指数级增长。尽管这些数据中潜藏着丰富的信息和潜在价值，但绝大部分都是以非结构化或半结构化的文本形式存在。其复杂性、多样性和海量性，为从中有效获取信息带来了巨大的挑战。在此背景下，信息抽取技术作为从非结构化文本获取数据的关键技术，其重要性日益凸显。在信息抽取领域，命名实体识别（Named Entity Recognition，NER）和关系抽取（Relation Extraction，RE）是两个基础且核心的子任务。命名实体识别旨在从文本中识别出具有特定意义的实体边界及其类别；关系抽取致力于识别并分类文本中实体之间存在的语义关系。这两项任务的准确性直接影响着许多下游应用的效果，例如知识图谱构建、信息检索、问答系统等。因此，如何持续提升命名实体识别和关系抽取的性能一直是学术界和工业界共同关注的焦点。

近年来，得益于深度学习理论的突破性进展和大规模预训练语言模型的广泛应用，命名实体识别技术日趋成熟，其性能获得了显著的提升[1-3] 。由于命名实体识别作为关系抽取的上游关键环节，其性能的提高能够给关系抽取模型提供更精准的边界和实体信息，很大程度上减轻了关系抽取因错误传播（命名实体识别阶段的错误结果）而导致的性能瓶颈。与此同时，关系抽取模型也在不断优化，能够更好的利用命名实体识别的结果，从而对实体间的语义关系做出更精确的判断。

尽管关系抽取模型本身在不断演进，但其性能提升近年来呈现出放缓的趋势。这一瓶颈在很大程度上归因于当前主流的流水线式抽取范式所固有的局限性[4-5]。具体而言，该范式将命名实体识别与关系抽取作为两个独立的串联任务，关系抽取模型通常仅依赖于命名实体识别模型的识别结果，其训练过程中所捕捉到的深层语义特

征，在任务传递时被忽略了。关系抽取模型难以探知命名实体识别模型做出判断的具体依据，从而造成了信息的损失。因此如何有效协同关系抽取模型与命名实体识别模型之间的交互，并探索高效的知识迁移机制以期进一步提升关系抽取性能，是信息抽取领域一个具有重要理论意义和实践价值的研究方向。

## 1.3 课题研究的目的与意义

在信息抽取任务中，命名实体识别和关系抽取是两个高度关联的子任务。针对这两个任务之间交互不足的问题，近年来备受关注的对比学习为此提供了新的解决思路。作为一种强大的自监督表示学习范式，对比学习的核心优势在于其能够从未标注或少量标注数据中学习到判别性强、迁移能力高的特征表示[6]。这一特性使其在提升模型对复杂数据的理解、促进深层知识迁移方面尤为突出。因此，探索利用对比学习构建两个任务之间的有效协同与知识迁移机制，是克服传统流水线模式局限性和提升关系抽取性能的一个值得深入探索的方向。

目前已有研究尝试将对比学习引入命名实体识别模型中，并取得了一定进展[1]。然而，该类工作主要聚焦于优化命名实体识别任务本身，尚未探索如何将学到的知识有效地迁移或应用于关系抽取任务中。更重要的是，现有基于对比学习的命名实体识别方法存在实体跨度表示不充分的问题，这不仅影响了命名实体识别的效果，也限制了其结果和知识向关系抽取模型的有效传递。因此，为了将基于对比学习的命名实体识别模型所学习到的知识迁移至关系抽取模型，仍需克服以下关键挑战：

- 现有基于对比学习的命名实体识别模型都是以跨度为单位进行分类的。其跨度表示往往通过简单聚合构建，未能捕捉到跨度内部丰富的语义及上下文信息，影响模型对长跨度实体的识别。
- 在对比学习的设计上，现有方法在正负样本的构造围绕仅考虑了实体跨度维度，未能考虑从实体类别维度构建正负样本，这导致模型学习到的实体表示在类别维度上缺乏足够的区分度。
- 将命名实体识别模型的知识迁移至下游的关系抽取模型也是一个挑战。由于命名实体识别的优化目标是识别实体，而关系抽取的优化目标是推理关系。这种目标不一致性，导致了两者模型在架构设计上的异构性，给跨任务的知识迁移带来了困难。

针对上述挑战，本文着力于研究对比学习和上下文表示技术，增强命名实体识别模型的实体跨度特征表示，从而传递更加精准的实体边界信息。随后，将命名实体识别阶段的知识迁移应用于关系抽取任务，从而提高模型对实体间关系的判别能力。这项工作不仅提升了信息抽取任务的整体性能，还为解决两个子任务间模型交互不足的问题提供了一种可行且有效的解决方案。

## 1.4 国内外研究现状

本论文主要聚焦于信息抽取中命名实体识别和关系抽取这两个核心任务，并探究了通过对比学习将命名实体识别的高质量表示有效迁移应用于关系抽取任务，因此本节将分别介绍命名实体识别、关系抽取和对比学习的国内外研究现状，这些研究为本工作提供了启发和借鉴。

### 1.4.1 命名实体识别

命名实体识别作为信息抽取的核心任务，其发展历程大致可分为传统方法和基于深度学习的方法。

传统的命名实体识别方法主要包括基于规则的方法、无监督机器学习方法以及基于特征的有监督机器学习方法。基于规则的方法主要依靠人工设计领域特定的启发式规则进行识别，Kim 和 Woodland[7]提议使用 Brill 规则推理方法进行语音输入，该系统会根据 Brill 的词性标签器自动生成规则。这类方法在规则覆盖全面的情况下能够获得较高的精度，但存在召回率低、规则构建和维护成本较高以及领域迁移能力不足的问题。以聚类为代表无监督学习方法通常基于词汇资源或统计信息，通过相似的上下文和语料库统计推断命名实体，Zhang 和 Elhadad[8]提出了一种无监督的方法从生物医学文本中提取命名实体。他们的模型没有进行监督，而是依靠术语、语料库统计和浅层句法知识。尽管无需标注数据，这类方法在性能稳定性和识别精度方面通常不及监督方法。基于特征的监督学习方法依靠人工特征工程，将命名实体识别问题转化为序列标注或分类任务并结合经典机器学习算法（如隐马尔可夫模型、决策树、支持向量机和条件随机场等）进行识别[9-16]。这类方法在特征设计完善的情况下表现优异，但特征工程本身需要大量人力，且在面对新领域或语言时，通用性和扩展性均受限。

近年来，得益于深度学习理论与技术的快速发展，基于深度学习的命名实体识别

模型取得了显著进展并占据主导地位，相比于传统方法，基于深度学习的命名实体识别方法能够在大规模数据集上自动学习特征，并在跨领域应用中展现更强的泛化能力。当前研究主要围绕提升嵌套实体识别能力、提高推理效率和增强模型泛化能力展开，形成了多种主流范式。早期的命名实体识别方法主要基于序列标注范式[17-18]，将文本视为词元序列，并通过以 RNN、LSTM 为代表的循环神经网络结合 CRF 等模型进行标注。例如，罗等人[19]提出了融合笔画级 ELMo 和多任务学习的 BiLSTM-CRF 模型，用于中文电子病历命名实体识别。然而，传统的序列标注方法难以有效处理嵌套实体，即一个实体包含另一个实体的情况，因为单一的序列标签无法同时表示多层实体信息[20]。为了解决嵌套实体问题及提升效率，研究者提出了多种新的技术路径。

在生成式命名实体识别方法中，命名实体识别任务被建模为一个序列生成问题，通过自回归或非自回归的方式生成实体文本或实体指针。Yan 等人[21]提出了一种基于 BART 的生成方法，将实体识别任务视为结构化序列生成问题，并利用指针机制预测实体的边界和类别。之后，Tan 等人[22]提出的 Seq2Set 方法进一步优化了生成式命名实体识别方法的解码过程，通过实现多个实体的并行预测，有效规避了自回归生成的低效性，从而大幅提升了模型的推理效率。此外，Shen 等人[3]提出了基于扩散模型的命名实体识别方法，利用扩散去噪过程来逐步优化实体边界预测，以提高模型的稳定性和泛化能力。

另一类重要的命名实体识别研究方向是基于阅读理解（Machine Reading Comprehension，MRC）的方法[23]，该方法通过构造问题来引导模型识别特定类别的实体。这一方法的优势在于可以利用预训练语言模型的强大特征表示能力，提高对长距离依赖的建模能力，并有效解决数据稀缺问题。然而，MRC 方法在多实体类别场景下计算成本较高，并且依赖于手工设计的查询模板，这限制了其在大规模、多类别命名实体识别任务中的应用。为了解决这一问题，Shen 等人[2]提出了并行实例查询网络，通过并行实例查询机制，同时预测多个实体，并采用动态标签分配策略，从而减少计算成本并提高模型的预测效率。

针对嵌套实体识别的挑战，基于跨度的命名实体识别方法提供了一种有效的解决方案，这类方法将命名实体识别任务从对单个词元的标注转变为对跨度的分类，其关键在于如何为每个候选跨度进行特征表示。一种直接的方法是池化跨度内所有词元的表示[4]。之后，Yu 等人[24]借鉴了依赖解析中的 Biaffine 解码器[25]的思想，将跨

度分类转化为对起始和结束词元组合对的分类。这种以其前后标识跨度的框架受到了广泛的关注和应用，并取得了不少的成果[1,26-30]。但是这种方法仍存在一定的局限性，由于其仅依赖边界词元的策略存在中间语义缺失的问题，从而导致对较长跨度的表示能力不足。随后，Yan 等人[31]通过卷积神经网络对跨度分数矩阵进行建模，成功捕捉了矩阵中的空间关系，为跨度补充了上下文语义。Zhu 和 Li[32]提出使用多粒度二维卷积对跨度分数矩阵进行卷积，以捕捉近距离和远距离跨度之间的交互关系。然而，这些直接将卷积神经网络应用于跨度矩阵的方法也存在一些问题。首先，跨度矩阵与图像在结构和性质上存在显著差异，跨度矩阵中跨度之间的关系相比于图像中像素间的关系更具规律性。其次，跨度矩阵的隐藏层维度较高，每增加一层卷积层都会显著增加模型的参数量，从而增加计算复杂度。

### 1.4.2 关系抽取

传统的关系抽取方法主要采用基于规则的技术，也称为手工构建的模式方法。这些方法为预定义的关系集设计一系列抽取模式，通过模式与文本的匹配来实现关系抽取，许多早期的研究均采用了这种方法[33-36]。然而，基于规则的方法存在一定的局限性。由于模式的设计高度依赖于人工经验和领域知识，这类方法通常只能适用于特定领域。当应用于新领域或处理异构语料时，需要重新设计关系模式，这不仅增加了大量人工成本，也导致方法难以灵活扩展。此外，手工构建的规则难以覆盖语言表达的多样性，限制了该方法的泛化能力以及对复杂文本场景的适应性。

基于深度学习的流水线关系抽取方法将任务划分为实体识别与关系分类两个阶段，并在每个阶段构建独立的模型来完成相应任务。实体识别阶段通常采用序列建模方法，而关系分类阶段则广泛应用卷积神经网络或依存句法分析技术。Zeng 等人[37]使用 CNN 对实体对的上下文语义进行编码，用于关系分类。武等人[38]提出了基于高层语义注意力机制和 CNN 的中文实体关系抽取方法。Chen 和 Manning[39]引入依存句法信息，增强了模型的结构建模能力。这类方法具有结构清晰、便于模块化训练的优点，但由于任务之间缺乏有效协同，容易导致误差在阶段之间传播，从而限制整体抽取性能。

为克服流水线式关系抽取方法中任务分离、误差传播等问题，近年来联合抽取方法逐渐成为实体识别与关系抽取任务的重要发展方向。此类方法致力于在统一框架下实现实体与关系的协同建模，通过增强任务间的信息交互，有效缓解前后阶段

之间的依赖瓶颈与性能损耗。CasRel[40]模型采用级联二分类标签方式依次识别头尾实体，具备较强的关系特异性建模能力。TPLinker[41]是基于 token pair 的统一抽取框架，通过表格填充方式直接建模 token 对之间的实体与关系，支持三元组重叠的表示，兼具高效性与灵活性。Yan 等人[42]进一步提出 PFN 模型，引入任务门控机制划分神经元为共享与专属区域，实现实体与关系任务之间的精细化双向交互。Sui 等人[43]将联合抽取建模为集合预测问题，以无序三元组目标集替代顺序标注机制，在提升效率的同时有效应对关系重叠与冗余候选生成问题。中文研究领域也涌现了针对特定场景的联合抽取模型，这些模型尝试将联合抽取方法应用于中文文本的特定领域，以解决实际应用中的挑战。例如，孙等人[44]提出了面向中文电子病历的 PIAN 模型。尽管联合抽取模型在解决流水线式方法的痛点上展现出巨大潜力，然而，此类模型也普遍面临着结构复杂、训练成本较高、可解释性较弱等挑战。

由于联合抽取模型在结构复杂性、训练成本和可解释性等方面存在一定问题，一些研究者重新将目光投向流水线方法。最具代表性的改进之一是 PURE[45]，该方法通过在文本中插入特殊标记以突出实体对，从而在关系抽取阶段提升实体间的语义关注。尽管 PURE 在处理效率上存在限制，其复杂度为 $O(N^2)$，但它有效地桥接了实体识别和关系抽取任务，增强了两者的协同作用。随后，Ye 等人[46]在 PURE 的基础上通过固定一个主语实体并将其他对象附加到文本末尾，降低了计算复杂度至 $O(N)$，并在多个对象的关系抽取上表现出较好的性能。然而，PL-Marker 依然存在实体间缺乏相互可见性的问题，限制了其对复杂关系的建模。为了解决这一问题，Feng 等人[47]提出了一种创新的两阶段训练方法，结合了候选标签标记机制和对象间相互定向注意力机制。该方法通过将关系抽取任务转化为标签选择问题，并增强实体对之间的交互性。这些方法在一定程度上缓解了传统流水线方法中误差传播对关系抽取阶段的影响。然而，尽管上述工作在流水线架构下进行了多项改进，但关系抽取阶段在很大程度上仍只利用了命名实体识别阶段识别出的实体结果，而未能充分利用命名实体识别模型在训练过程中学习到的语言表示。也有工作尝试将命名实体识别阶段训练好的编码器直接传递给关系抽取模型，但是由于任务特性差异等原因而产生了负面影响[45]。因此，如何设计更为有效的方法，以将命名实体识别任务中蕴含的深层知识迁移并整合到关系抽取阶段，从而促进其性能提升，仍然是关系抽取领域一个值得深入探索的关键研究方向。

### 1.4.3 对比学习

近年来，对比学习已成为深度学习领域的主流研究方向之一。作为一种强大的自监督表示学习范式，它使得模型能够学习到高质量、可迁移的特征表示，在跨任务迁移上展现出潜力。对比学习的核心思想在于构建一个表示空间，通过优化目标拉近正样本对的样本表示，同时推远负样本对的样本表示，从而促使模型从未标注或少量标注数据中学习到具有良好判别性的深层特征[48-49]。

对比学习的许多基础概念和框架最初在计算机视觉领域得到了验证和发展。Chen 等人[50]通过强数据增强和大规模批次训练证明了对比学习的有效性。He 等人提出的 Moco 框架[51]利用动量编码器和队列优化了负样本管理，提高了训练效率。Grill 等人[52]则探索了无需负样本的对比学习路径。这些计算机视觉领域的开创性工作不仅证明了对比学习在学习高质量表示方面的能力，也为后续将对比学习思想应用于其他模态奠定了基础。

受计算机视觉领域成功的启发，对比学习被引入自然语言处理并取得了显著进展，尤其在构建高质量的句子嵌入方面表现突出。早期的工作如 Sentence-BERT[53]，通过利用孪生网络结合自然语言推理、语义文本相似度等特定任务的监督数据进行微调，有效地优化了句子嵌入的相似度计算，使其更适用于下游任务。Gao 等人提出的 SimCSE[54] 则在此基础上，进一步推动了句子表示学习的发展。其无监督版本巧妙地利用 Dropout 机制为同一句子构造正样本对，而有监督版本则借助自然语言推理数据中的蕴含和矛盾关系来生成正负样本。SimCSE 不仅显著提升了句子表示的对齐性和均匀性，还有效缓解了句子嵌入的各向异性问题。受 SimCSE 成功的启发，研究者也开始探索在更细粒度的词元级别应用对比学习。TaCL[55] 旨在通过对比学习提升词元表示的区分度和各向同性，从而改善模型在 GLUE[56]、SQuAD[57] 等多个自然语言理解任务上的性能。这些在提升句子和词元表示质量上的成功，以及双编码器架构在计算相似度方面的固有高效性，使得对比学习与双编码器架构的结合自然而然地成为了处理需要高效相似度计算或大规模检索场景的主流方法，尤其是在信息检索、开放域问答等任务中。密集检索模型（Dense Passage Retriever，DPR）[58]正是这一趋势下的典型代表。DPR 采用非对称双编码器架构，并利用对比学习进行训练，极大地提升了端到端密集检索的性能。与此同时，除了具体的模型探索，也有研究深入聚焦于对比学习的关键技术细节，包括不同损失函数的选择、正样本的构建

策略（如数据增强、利用标注数据），以及负采样策略（如批内负采样、难负样本挖掘）。这些研究共同构建了对比学习在自然语言处理领域的应用图景，使其成为提升预训练语言模型表示能力和下游任务迁移性能的重要手段。

## 1.5 论文主要工作

本论文聚焦于增强基于对比学习的命名实体识别模型的特征表示，并将其所学习到的知识有效迁移至关系抽取任务中，以提升信息抽取的整体性能。如图 1.1 所示，针对命名实体识别和关系抽取的问题，本文的主要研究工作与创新性贡献体现在以下三个方面：

（1）针对现有基于对比学习的命名实体识别方法在跨度表示上语义信息不足的问题，提出了一种基于偶数卷积的跨度表示增强模块。该模块利用偶数卷积核不规则感受野的特性，使得跨度更加关注与其有着被蕴含关系的子跨度信息，进而有效捕捉并补充实体跨度内部的中间语义信息，显著增强了跨度特征的表示，提升了模型对于长跨度的识别能力。

（2）针对现有基于对比学习的命名实体识别框架对比维度单一的问题，设计并引入了基于跨度和类别的双重对比学习目标。该优化目标通过更精细化的损失计算方式，显著增强了模型区分目标跨度与其正负类别之间语义相似性的能力，从而学习到类别区分性更强的实体表示，有效提升了模型识别的能力。作为对框架整体优化的补充，本文还将原用于表示实体类别的预训练语言模型替换为轻量的可学习嵌入层，此举在保证模型性能的同时，也有效简化了模型结构并提升了运算效率。

（3）针对现有关系抽取模型未能充分利用命名实体识别阶段信息的问题，本文将命名实体识别阶段的文本编码器和类别编码器迁移至下游的关系抽取模型中，分别用于提升实体表示质量和为关系抽取决策提供额外的实体类别信息。不同于传统信息抽取流水线仅在输出层面进行结果交互，本文方法在模型内部实现了命名实体识别阶段特征语义与关系抽取阶段特征语义的深度共享与融合，提升了关系抽取的性能，为构建高效协同的信息抽取系统提供了有效路径。

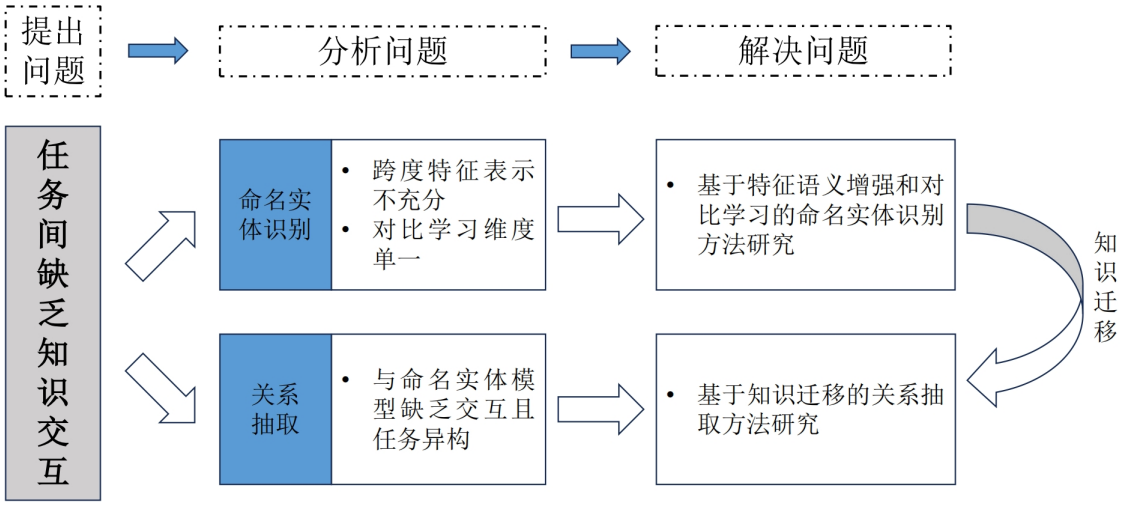



**图 1.1** 本论文的技术路线图

## 1.6 论文组织架构

针对现有基于对比学习的命名实体识别方法中特征表示不充分，以及关系抽取模型对命名实体识别阶段信息利用不足的问题，本论文研究上下文表示学习方法以及知识迁移方法，提出了一个基于特征语义增强和对比学习的命名实体识别模型，并构建了从命名实体识别模型到关系抽取模型的有效知识迁移。

本论文其他章节的内容安排如下：

第二章介绍了相关的理论基础与核心技术。主要内容包括：命名实体识别方法、预训练语言模型、卷积神经网络、注意力机制、双编码器架构以及关系抽取方法。

第三章致力于增强基于对比学习的命名实体识别方法的特征语义。本章首先分析现有方法的局限性，然后详细阐述提出的 SEE 模型。接着，通过在公开命名实体识别数据集上进行详尽实验，验证模型各组件的有效性及超参数影响。随后，通过对比实验将 SEE 模型与基线命名实体识别模型进行性能比较。最后，进行了效率分析，评估模型的实际应用潜力。

第四章研究了如何将第三章的命名实体识别模型中的知识迁移至下游关系抽取任务。为此，本章提出了关系抽取模型 TransRE，并详细阐述了其中的知识迁移机制。接着，在标准关系抽取数据集上进行了一系列实验分析，探究了迁移策略的独立贡献与协同效应，以及模型关键组件和迁移知识的特性。最后，通过对比实验，评估了整合知识迁移的 TransRE 模型相对于基线方法的性能优势。

第五章是对全文工作的总结与未来展望。系统梳理本论文的主要研究内容与创

新贡献，分析所提出方法的优势与存在的局限性，并对未来可能的研究方向和改进思路提出展望。

# 第二章　相关理论和方法概述

本章将系统回顾支撑本论文研究的相关理论与技术，为后续章节提出和阐述具体方法提供坚实的基础。

## 2.1　命名实体识别方法

命名实体识别是一项基础且关键的自然语言处理任务，旨在从非结构化文本中自动抽取出特定语义类别的实体片段（如人物、地点、组织、时间、数字表达式等），并将其划分到预定义的类别中[59]，从而为下游的关系抽取、知识图谱构建、问答系统等任务提供重要的支撑[60]。命名实体识别最初被定义为一个序列标注问题，即为输入文本序列中的每个词元赋予一个表示实体类别的标签，传统上通常假设文本中的实体之间互不交叠，即每个词元仅属于一个实体或不属于任何实体。

然而，随着对实际文本的深入研究发现，现实中的实体结构往往更为复杂，一个实体可能完整地包含另一个实体，这种情况被称为嵌套实体。如图 2.1 所示，在文本片段“南京市长江大桥”中，“长江”实体嵌套于“长江大桥”实体中。相较于传统的非嵌套实体识别，嵌套实体识别因需考虑实体间的层次与交叠，复杂度显著提升。传统序列标注方法难以有效应对此挑战，从而推动研究人员寻求新的技术路径。

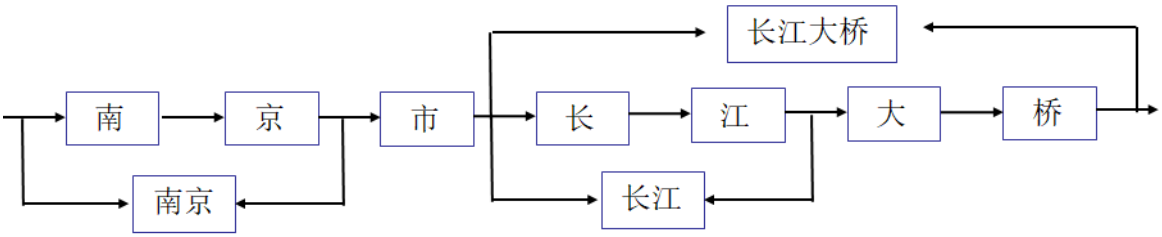



**图 2.1**　嵌套命名实体示例

为了有效处理包括嵌套实体在内的复杂实体结构，学术界涌现了多种技术范式，基于跨度的命名实体识别方法便是其中一种极具代表性且强有力的解决方法。该方法的核心思想是将命名实体识别任务从对单个词元的标注转变为对跨度的分类。具体而言，模型会枚举句子中所有可能构成实体的连续词元序列，并为每个候选跨度生成一个特征表示。以“南京市长江大桥”为例，如图 2.2 所示，这段文本共有七个

字，其所有可能的跨度可以通过一个 7x7 的矩阵进行概念上的可视化。在这个矩阵中，横轴代表跨度的起始字符位置，纵轴代表跨度的结束字符位置。矩阵中位于主对角线及右上方的单元格 $(i, j)$（满足 $i \leq j$）唯一对应着文本中从第 $i$ 个字符开始到第 $j$ 个字符结束的某个特定跨度。在矩阵中对角线上的单元格代表单个字符构成的跨度，而越偏离对角线、向右上角移动的单元格则代表由多个字符组成的、长度更长的跨度。矩阵中对角线以下的单元格（$i > j$）由于结束位置在起始位置之前，在表示连续跨度上没有意义。

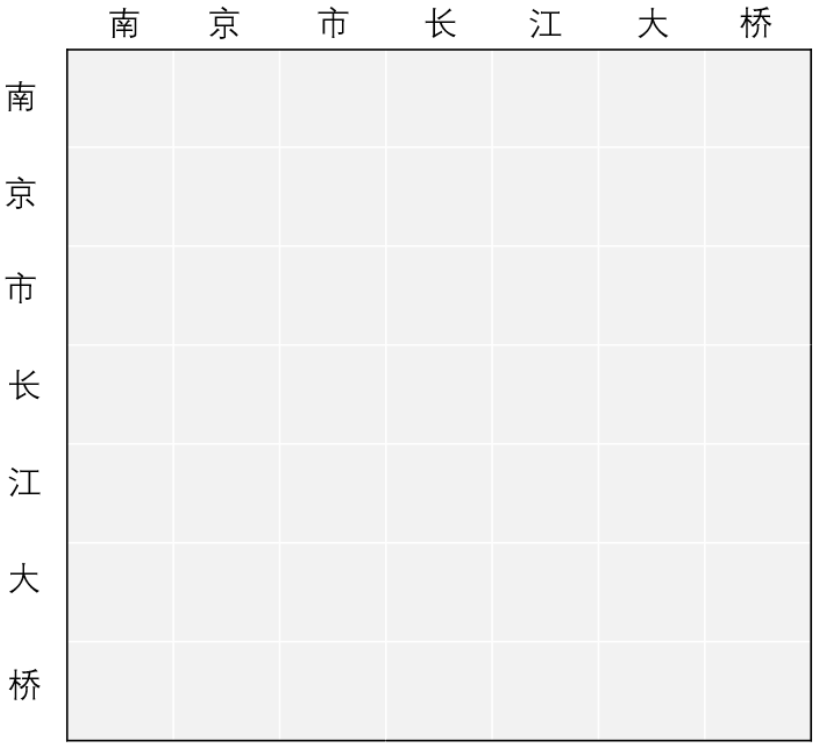



**图 2.2** 跨度特征矩阵示例

在此范式下，命名实体识别任务被分解为两个步骤：首先，为矩阵上三角区域中的每个候选跨度 $(i, j)$，通过融合其内部语义与外部上下文信息，生成一个高维特征表示；然后，基于该特征表示对跨度 $(i, j)$ 进行分类，判定其所属的实体类别或判定为非实体。由于每个候选跨度都经过独立地判断和分类，该范式能够处理相互重叠或嵌套的实体，从而有效克服了传统序列标注方法的局限性。

## 2.2 预训练语言模型

预训练语言模型（Pre-trained Language Models，PLMs）是近年来自然语言处理领域取得突破性进展的关键技术。这类模型通常首先在大规模、无标注的文本语料库上进行无监督或自监督的预训练，旨在让模型学习通用的语言知识，如词汇语义、语法结构。完成预训练后，这些模型可以作为强大的基础特征提取器，通过在特定下游任务的标注数据上进行微调，便能快速适应各种具体的 NLP 应用，如文本分类、命名实体识别、关系抽取、机器翻译等，并显著提升任务性能。预训练语言模型的巨大优势在于其能够从未标注数据中有效学习深层上下文表示，从而大大减少了对特定任务标注数据的依赖。

其中由 Devlin 等人提出的 BERT （Bidirectional Encoder Representations from Transformers）是最具里程碑意义的预训练语言模型之一[61]。如图 3.2 所示，BERT 基于 Transformer[62]的编码器架构，其核心创新在于设计了两个巧妙的预训练任务来捕捉双向的上下文信息：掩码语言建模（Masked Language Model，MLM）——随机遮盖输入序列中 15% 的词元，然后训练模型依据该词元左右两边的上下文来预测被遮盖掉的原始词元；以及下一句预测（Next Sentence Prediction，NSP）——给定两个句子 A 和 B，判断句子 B 是否为原文中紧跟在句子 A 后面的下一句。通过这两个任务的联合训练，BERT 能够学习到深度的、与上下文相关的词元表示。在应用于下游任务时，通常只需在预训练好的 BERT 模型顶部添加一个简单的任务特定层（例如，用于分类的全连接层），然后使用少量任务标注数据对整个模型进行微调，即可取得非常好的效果。BERT 的提出极大地推动了“预训练-微调”这一新范式在 NLP 领域的普及和应用。

基于 BERT 的巨大成功，研究界迅速发展出一系列对其预训练策略、训练数据或模型结构进行改进的变体模型，以期获得更强的性能或针对特定领域的更优表现。本论文的研究工作除了基础的 BERT-base 之外，还涉及了以下几种 BERT 的重要变体：

- **RoBERTa[63]**: 作为 BERT 的直接优化版本，RoBERTa 证明了精心调整预训练过程的重要性。其主要改进包括：(1) 使用了更大规模、更多样化的训练数据 (包括 BooksCorpus[64]，CC-News[65]，OpenWebText[66]，Stories[67]等，总计约 160GB)；(2) 移除了 NSP 预训练任务，实验发现该任务对某些下游任务性能提升有限甚

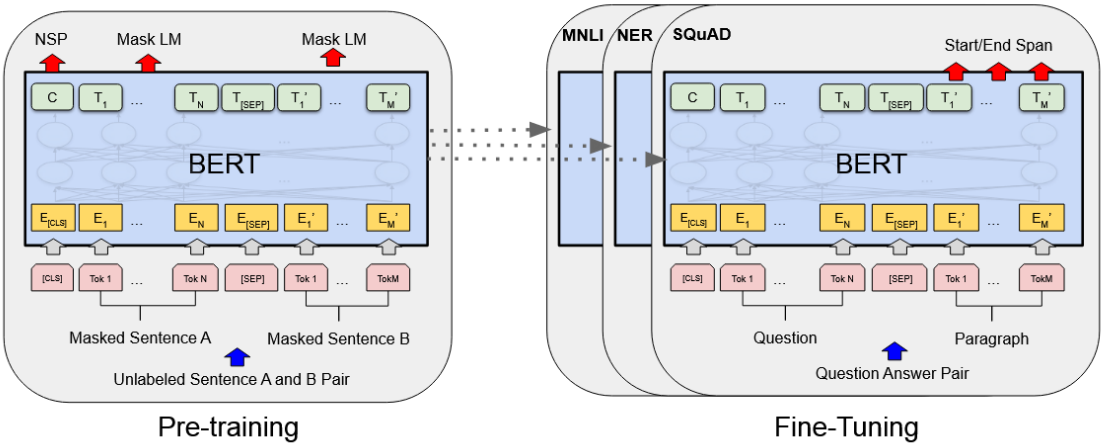



图 2.3　BERT 模型的训练过程[61]

至有害；(3) 采用了动态掩码策略，即每次向模型输入序列时都会重新生成掩码模式，而不是像 BERT 那样在数据预处理时固定掩码；(4) 使用了更大的训练批次和更长的训练步数。这些综合优化使得 RoBERTa 在多项 NLP 基准（如 GLUE[56]，SQuAD[57,68]）上显著超越了 BERT 的性能。

- **PubMedBERT**[69]: 考虑到通用领域预训练模型在特定专业领域的词汇和表达上可能存在局限，研究者提出了面向特定领域的预训练模型。PubMedBERT 就是一个专门为生物医学领域设计的 BERT 模型。与标准 BERT 不同，它完全使用生物医学文献（PubMed 摘要和来自 PMC 的全文）作为预训练语料，并且是从零开始进行训练的，而非在 BERT 基础上继续训练。这种领域专属的预训练使得 PubMedBERT 能够更好地理解生物医学术语、缩写和独特的语言模式，在生物医学相关的 NLP 任务上，通常展现出优越的性能。
- **SciBERT**[70]: 与 PubMedBERT 类似，SciBERT 是另一款面向科学文献领域的预训练模型。它的预训练语料来自于覆盖计算机科学和生物医学领域的科学论文文本。通过在大量科学文本上进行训练，SciBERT 旨在提升模型对科学术语、研究方法、实验结果等复杂科学内容的理解能力。本论文在处理 SciERC 数据集上的关系抽取任务时，便采用了基于 SciBERT 的预训练权重，期望发挥其在科学文本上的优势。

## 2.3 卷积神经网络

卷积神经网络（Convolutional Neural Network，CNN）是一种深度学习模型，广泛用于各种深度学习领域，尤其是在计算机视觉领域[71]。CNN 通过模仿生物视觉系统的工作原理，利用局部连接、权重共享和池化等机制，逐层提取输入数据中的特征。其核心优势在于自动从数据中学习有效的特征表示，而无需人工设计特征提取器。典型的 CNN 架构主要由卷积层、激活函数、池化层和全连接层等核心组件构成。

卷积层（Convolutional Layer）是 CNN 的核心计算单元。它通过一组可学习的卷积核（Kernel，也称滤波器）作用于输入数据。对于图像等二维数据，卷积核是一个小的权重矩阵，它在输入图像上按步长滑动，与其感受野内的局部区域进行卷积运算。该运算旨在检测输入中的局部模式，如图像的边缘、角点和纹理等。

一个关键的机制是权重共享，即同一个卷积核的参数（权重 $W$ 和偏置 $b$）在处理输入数据的不同空间位置时保持不变。这相当于用同一个特征检测器扫描整个输入空间，不仅显著减少了模型需要学习的参数量，提高了计算效率，而且使得模型学习到的特征具有一定的平移不变性。通常，一个卷积层会包含多个卷积核，每个核负责提取一种不同的局部特征，从而产生多个输出特征图或通道。对于一个单通道输入 $I$（例如灰度图像矩阵）和尺寸为 $k_h \times k_w$ 的卷积核 $W$，卷积层在输出特征图 $O$ 的位置 $(i, j)$ 生成的值 $O_{i,j}$ 可以表示为:

$$O_{i,j} = f\left(\sum_{m=0}^{k_h-1}\sum_{n=0}^{k_w-1} W_{m,n} \cdot I_{i+m,j+n} + b\right) \tag{2.1}$$

其中 $I_{i+m,j+n}$ 代表输入在相应位置的值；$W_{m,n}$ 是卷积核的权重；$b$ 是偏置项；$f(\cdot)$ 代表紧随其后的非线性激活函数。

激活函数的引入对于使网络能够学习和表示非线性关系至关重要。常用的激活函数是修正线性单元（Rectified Linear Unit，ReLU），其计算方式如下：

$$\text{ReLU}(z) = \max(0, z) \tag{2.2}$$

ReLU 函数将所有负激活值置零，保留正激活值，有助于缓解深度网络训练中的梯度消失问题，并被广泛应用于许多 CNN 架构中。

近年来，尤其在 Transformer 等模型中，高斯误差线性单元（Gaussian Error Linear Unit，GELU）也被证明是一种非常有效的激活函数[72]。GELU 可以看作是 ReLU 的

一种平滑近似，它根据输入的数值大小（更准确地说，是其在标准正态分布下的累积概率）来对其进行随机正则化的加权，其数学表达式为：

$$\text{GELU}(x) = x \cdot \Phi(x) \tag{2.3}$$

其中 $\Phi(x)$ 是标准高斯分布的累积分布函数（Cumulative Distribution Function，CDF），即 $\Phi(x) = P(X \leq x)$，$X \sim N(0,1)$。GELU 的特点在于其平滑的非线性特性，并且它允许负值传递（乘以一个小于 1 的系数），这被认为有助于模型学习更复杂的函数，并在一些大型语言模型中取得了优于 ReLU 的性能表现。

池化层，也称为下采样层，通常周期性地插入卷积层之间。其主要目的是降低特征图的空间维度（宽度和高度），从而减少网络后续部分的计算量和参数量，增强模型的稳健性，并提供一定程度的平移不变性。

全连接层通常位于 CNN 架构的末端。在经过多层卷积和池化提取到足够抽象和鲁棒的特征之后，这些特征会被展平成一个向量，输入到一个或多个全连接层中。全连接层负责将学习到的高级特征进行整合，并最终映射到任务的输出空间，例如在分类任务中输出属于各个类别的概率。

尽管 CNN 最初是为处理图像这类二维网格数据而设计的，但其核心原理亦可应用于某些自然语言处理任务中产生的二维表示。具体而言，在基于跨度的命名实体识别方法中，一种常见的做法是通过枚举文本中所有可能的词语跨度，并通常基于跨度的起始和结束词元的嵌入表示，来构建一个跨度表示矩阵。这个矩阵实际上也构成了一种二维的结构化数据，其中每个元素对应一个候选的文本跨度及其初始表示。这种二维的跨度表示结构使得直接应用 CNN 的局部感知特性成为可能。

## 2.4 注意力机制

注意力机制是近年来深度学习中广泛应用的一种技术。其灵感来源于人类视觉系统的注意力机制，即人眼在处理视觉信息时，并非同时处理整个视野中的所有信息，而是将注意力集中在与当前任务最相关的部分，从而提高处理效率。在深度学习中，注意力机制允许模型在处理输入数据时动态地聚焦于重要的部分，从而提高模型的性能和表达能力[73]。

具体来说，注意力机制通过计算查询（Query，Q）、键（Key，K）和值（Value，

V）之间的关系来生成加权和。查询是当前要关注的输入，键是与查询相关联的其他输入，值则是与每个键关联的实际信息。在标准的注意力机制中，查询和键的相似度决定了权重分配，即相似度较高的键会获得更大的权重，而相似度较低的键则会被赋予较小的权重。这种加权和的计算使得模型能够动态地“关注”输入序列中最相关的部分：

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.4}$$

在 Transformer 模型[62] 中，其核心计算单元之一是多头自注意力（Multi-Head Self-Attention）机制。这种机制允许模型在处理输入序列时，不仅能够直接捕捉序列内部任意两个位置之间的依赖关系，而且能够从多个不同的表示子空间并行地学习和整合这些依赖关系。自注意力意味着 $Q$、$K$ 和 $V$ 均派生自同一个输入序列 $X$，使得序列能够“关注自身”的不同部分。多头机制则通过将注意力计算分散到 $h$ 个并行的“头”（head）来增强模型的表达能力。多头自注意力的计算过程如下：首先，将来自输入序列 $X$（维度为 $d_{model}$）的 $Q$、$K$ 和 $V$ 表示，通过 $h$ 组不同的、可学习的线性投影矩阵分别投影到 $h$ 个子空间，得到 $Q_i$、$K_i$ 和 $V_i$:

$$Q_i = QW_i^Q, \quad K_i = KW_i^K, \quad V_i = VW_i^V \tag{2.5}$$

其中 $Q, K, V \in \mathbb{R}^{N\times d_{\text{model}}}$ 分别表示输入序列的查询、键和值矩阵；$W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d_{\text{model}}\times d_k}$ 是第 $i$ 个头的可学习投影矩阵，$d_k = d_{\text{model}}/h$。

随后，在每个子空间上并行地应用缩放点积注意力函数（如公式 2.4 定义，但尺度因子分母中的维度为 $d_k$），计算出第 $i$ 个头的注意力输出 $\text{head}_i$:

$$\text{head}_i = \text{softmax}\Big(\frac{Q_i K_i^\top}{\sqrt{d_k}}\Big)V_i \tag{2.6}$$

其中 $\text{head}_i \in \mathbb{R}^{N\times d_k}$ 表示第 $i$ 个头的注意力输出，计算在各自子空间并行完成，这一步使得模型能在 $h$ 个不同的子空间中，独立地计算序列不同部分之间的关联性。

接着，将所有 $h$ 个头的输出 $\text{head}_i$ 在特征维度上进行拼接，并通过一个最终的可学习线性变换进行融合，产生多头自注意力层的最终输出：

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\, W^O \tag{2.7}$$

其中 Concat(·) 是拼接操作；$W^O \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$ 是对拼接结果进行线性映射的可学习投影矩阵。

## 2.5 双编码器架构

双编码器架构是一种常用于自然语言处理和信息检索中的网络结构。它通过使用两个独立的编码器来分别处理输入的两个部分，这两个部分的数据类型可以相同（如文本对）、也可以不同（图像-文本对、实体-类别对等），并分别对其进行编码，得到各自的嵌入表示。这两个编码器可以共享或不共享参数。然后，两个嵌入表示会被映射到同一空间中，通过计算它们之间的相似度，来度量它们的相关性。这种架构通常用于需要计算输入对之间相似度的任务，尤其适用于对比学习、语义匹配和信息检索等场景。

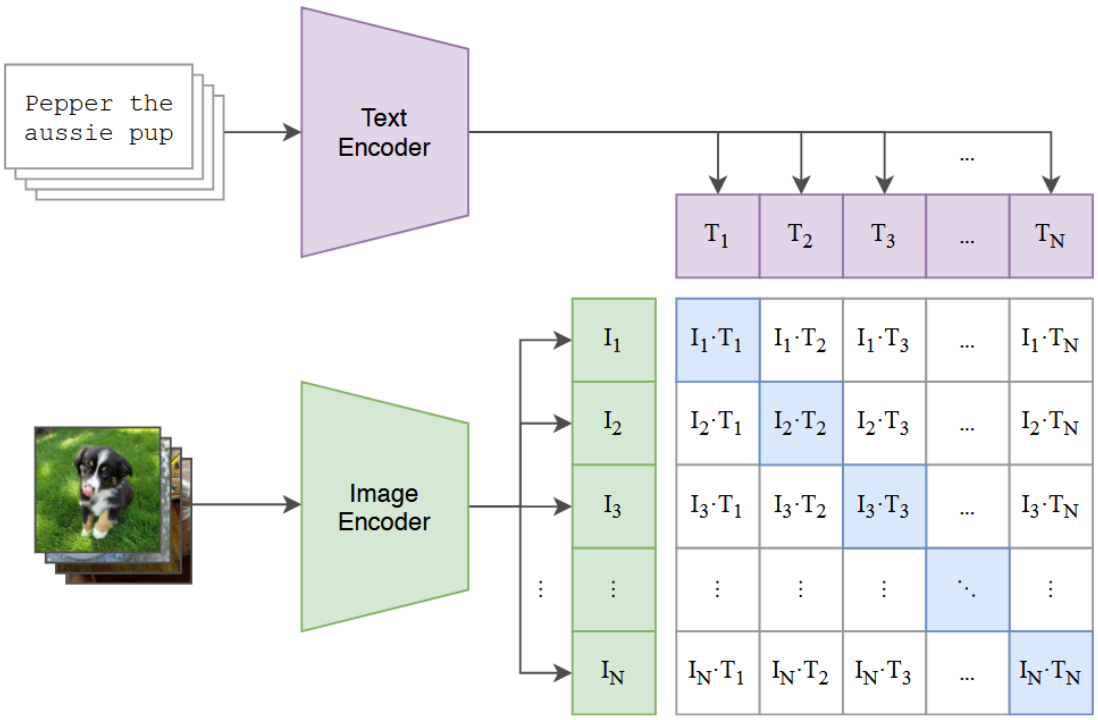



**图 2.4** CLIP 模型架构[74]

CLIP（Contrastive Language-Image Pretraining）[74]是一个极具代表性的模型。如图 2.4 所示，CLIP 通过引入双编码器架构，分别对图像和文本进行编码，对文本使用 Bert 预训练模型，对图像则使用 Vision Transformer（ViT）进行编码，并将它们的

表示映射到共享的语义空间中。通过在海量的文本-图像对上进行训练，CLIP 建立了图像与文本间强大的对齐能力。具体而言，CLIP 通过计算图像和文本对之间的相似度来进行对比学习，从而优化其表示，使得图像和文本的嵌入可以在同一个空间中进行比较和匹配。

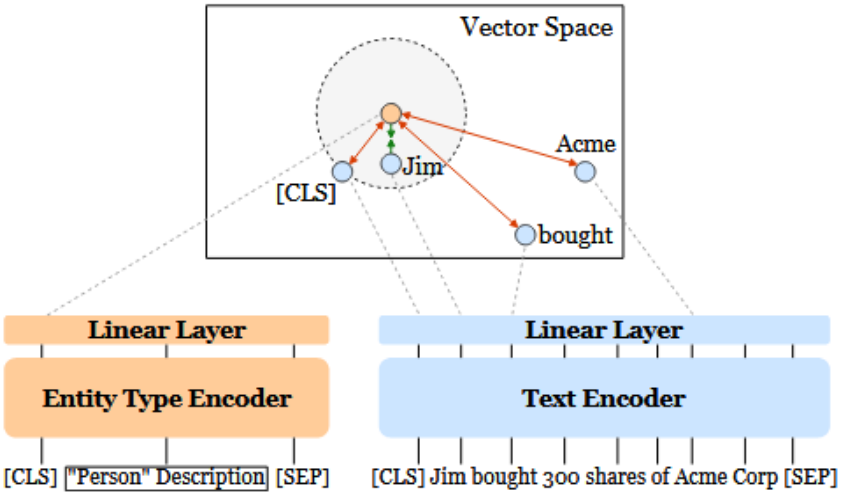



图 2.5　BINDER 模型架构[1]

BINDER（**BI**-encoder for **N**ame**D** **E**ntity **R**ecognition）[1]是一个基于双编码器架构的命名实体识别模型。如图 2.5 所示，其核心思想是将命名实体识别任务视为一个表示学习问题。模型包含两个独立的编码器：一个实体类别编码器和一个文本编码器。实体类别编码器负责将预定义的实体类别及其描述性文本映射到向量空间中，生成实体类别嵌入。同时，文本编码器处理输入文本，为其中的每个词元生成相应的向量表示，随后进一步构建跨度表示。该模型的目标是通过对比学习将文本跨度与其对应的实体类别映射到同一个向量空间中，使得匹配的实体类别和文本跨度表示具有高相似度。更关键的是，该模型还提出了一种新颖的动态阈值方法，它不再使用固定阈值，而是利用全局上下文信息（[CLS]）与实体类别的相似度来动态设定分类边界，从而更有效地将实体片段与非实体片段区分开来。

## 2.6　关系抽取方法

关系抽取是一项旨在从非结构化文本中抽取出已识别实体之间特定语义关系的自然语言处理任务。它通常在命名实体识别完成之后进行，负责发现并连接文本中

提及的实体。关系抽取任务的目标是将文本中两个或多个实体之间的关联信息，归类到预定义的关系类别中。例如，在识别出“苹果公司”和“库比蒂诺”两个实体后，判断它们之间是否存在“总部位于”的关系；在识别出“张三”和“清华大学”后，判断是否存在“就读于”或“任职于”的关系。通过识别和分类实体间的关系，关系抽取能够将文本中分散的实体信息转化为结构化的事实或知识片段（常以三元组形式表示：主体实体 - 关系类别 - 客体实体），从而为下游更高级别的知识获取、知识图谱构建、复杂问答系统、深度文本分析、信息检索等任务提供关键的、连接性的结构化知识支撑。

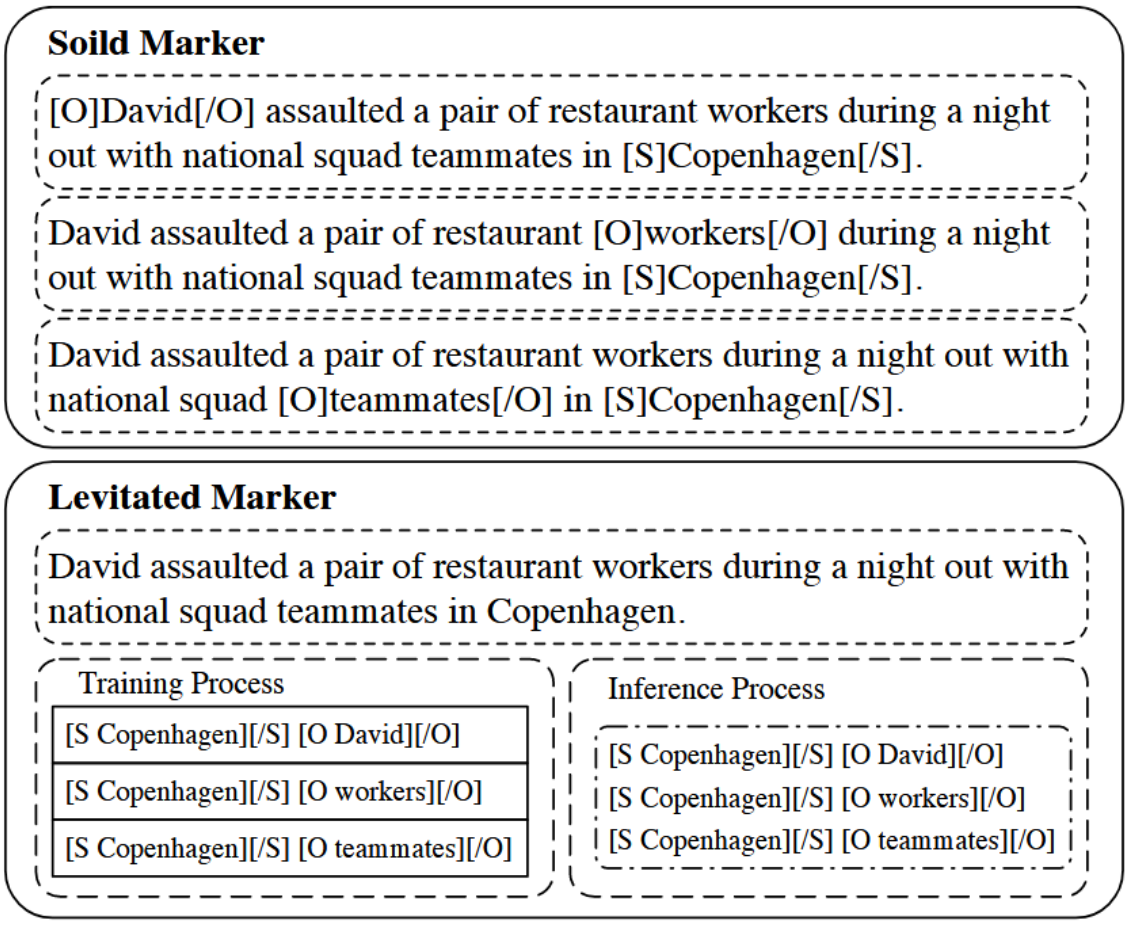



**图** 2.6　两类标记示例

在关系抽取任务中，如何让模型精确聚焦于待判断关系的特定实体对，是一个核心挑战。为应对此挑战，一种主流思路是利用特殊标记来显式地引导模型。基于这一思路发展而来的基于标记（Marker-based）的关系抽取方法得到了广泛的关注与发展[45-46] 。这类方法的核心思想在于，向关系抽取模型输入文本序列之前，会先基

于命名实体识别结果在文本中插入特殊的标记，以这种形式来显式地标明待判断关系的目标实体对的位置。标记的主要作用是引导模型关注特定的实体跨度，并帮助编码器编码出能够有效用于关系分类的实体特征表示。图 2.6 展示了本文的方法所涉及的两种标记：

(1) 实心标记（Solid Markers）：实心标记是一种直接的实体指示技术，其核心思想是在目标实体跨度的前后显式插入特定符号。这种方法直接改变了原始文本序列，从而以显式形式清晰地标示出目标实体的边界。实心标记的主要优势在于能为模型提供强烈而直接的实体边界信号，并且能够有效强化特定实体（如主语）的表示。然而，实心标记也存在局限性。当需要同时处理文本中的多个实体对时，序列中存在的多对实心标记会使得模型难以准确区分和关联特定实体对的起始和结束标记，这限制了其并行处理多实体对的能力。此外，大量实心标记的直接插入还会显著增加输入序列的长度，这可能导致计算成本的上升并影响整体处理效率，因此本文关系抽取模型仅对主语实体采用实心标记。

(2) 悬浮标记（Levitated Markers）：悬浮标记则采用了另一种不同的策略，其核心思想是让标记“悬浮”于文本之上，即在不直接改变文本序列的前提下标识实体。具体而言，悬浮标记与其对应的实体边界词元共享相同的位置嵌入，但拥有独立的词嵌入。它们依赖于定制的注意力机制来发挥作用，一对悬浮标记（起始和结束标记）在注意力计算中被设置为相互可见，并且它们都能关注到文本内容，而不同标记对之间则通常是相互屏蔽的。这种设计使得模型能够在一次编码过程中高效地并行处理多个实体对，并且由于标记不实际增加序列长度，因此特别有利于处理长文本。同时，它也能有效地将实体边界的位置信息传递给模型。与实心标记相比，悬浮标记对实体边界的指示更为间接，但是效率更好。为了平衡效率与性能，本文关系抽取模型对宾语实体采用悬浮标记。

## 2.7 本章小结

本章系统地介绍了支撑本论文研究的核心理论与关键技术。首先，阐述了命名实体识别任务的演进，特别是为解决嵌套实体问题而兴起的、将任务转化为对文本片段分类的基于跨度的方法。接着，介绍了以 BERT 及其领域特定变体为代表的预训练语言模型，它们通过“预训练-微调”范式为下游任务提供强大的上下文感知特

征表示。进一步，本章回顾了 CNN 的核心原理及其在处理跨度表示矩阵等二维结构数据中的应用潜力，并详细说明了作为现代模型基础的注意力机制。此外，还探讨了常用于对比学习和语义匹配的双编码器架构以及在关系抽取中用以明确实体边界的标记方法。这些理论和方法共同构成了后续章节提出新模型的技术基石。

# 第三章　基于特征语义增强和对比学习的命名实体识别方法研究

本文旨在探索如何在命名实体识别阶段通过基于对比学习的双编码器架构，学习高质量的文本和类别信息，并将其有效迁移到关系抽取任务中。现有基于对比学习的命名实体识别方法在特征表示能力上存在一些问题。首先，这类方法对跨度的特征表示通常采用首尾拼接的方式，这往往忽略了跨度内部词元所蕴含的丰富语义信息，造成了跨度内部语义信息的损失。特别是对于较长的实体，这种表示上的信息损失可能更为严重。其次，在对比损失的设计上，当前主要框架仅在跨度层面构建对比样本对，忽略了在类别维度上进行显式对比的可能性。这意味着未能充分利用不同实体类别之间的区分信息来进一步优化表示空间，导致学到的特征在区分细粒度或相似类别实体时鲁棒性低。针对上述问题，本章提出了一个基于对比学习的特征语义增强双编码器模型（contrastive learning-based feature Semantic Enhanced bi-Encoder model，SEE），旨在学习到更加丰富、鲁棒的文本和类别特征表示，以便在下一章中用于知识迁移，最终提升关系抽取任务的性能。

## 3.1　方法概述

本章聚焦于提升基于对比学习的命名实体识别方法的特征表示能力。针对现有方法在跨度内部语义信息表示不足以及对比损失优化维度单一等方面的局限性，本章提出了 SEE 模型，整体流程如图 3.1 所示。首先，模型将预定义的实体类别和句子中的候选跨度映射到向量空间中。其次，针对跨度表示中存在的中间语义不足问题，本章利用基于偶数卷积核的卷积神经网络（Even-sized kernel Convolutional Neural Network，ECNN）来增强跨度表示，该模块通过捕捉子跨度的信息有效补充了跨度内部的语义信息。最后，为了解决现有方法对比学习维度单一的局限性，SEE 模型采用了基于跨度和类别的双重对比学习优化目标，这使得模型能够同时从跨度维度和类别维度进行优化，以充分发挥对比学习在区分性表示学习上的优势。通过上述结构设计，SEE 模型能够学习到更加丰富、鲁棒的文本和类别特征表示，为后续章节的知识迁移奠定基础。详细的模型架构和原理将在后续小节中进行阐述。

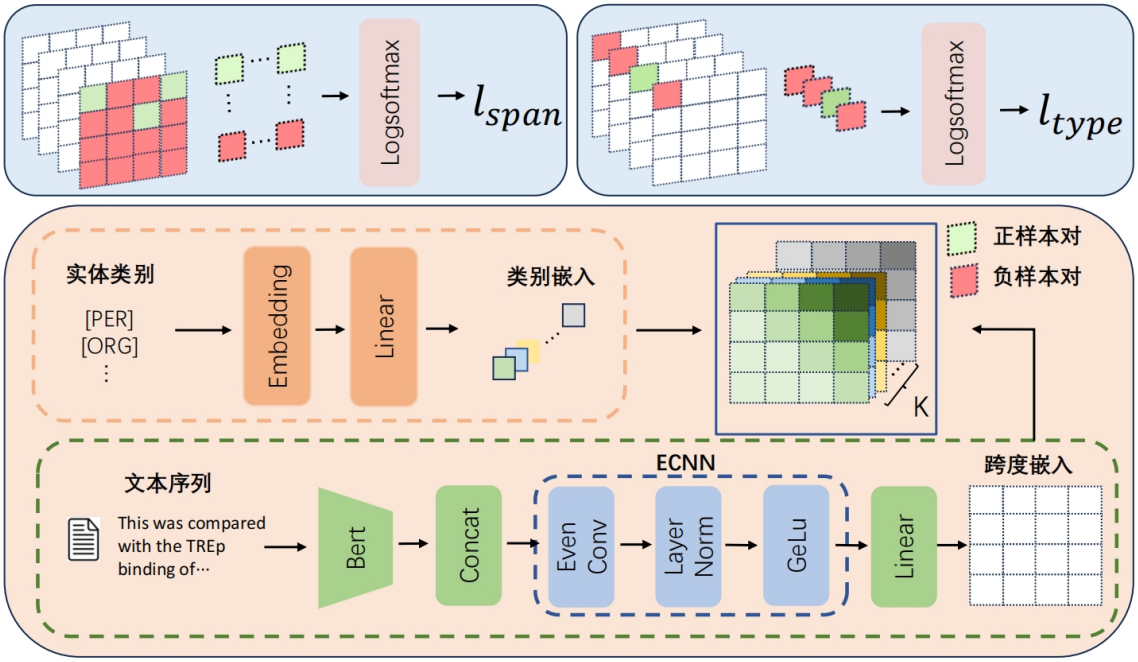



图 3.1　SEE 模型的整体结构

### 3.1.1　类别嵌入表示

SEE 是基于对比学习框架的模型，该模型首先对每个跨度和类别进行特征表示。对于任务预定义的 $m$ 个实体类别 $\epsilon = \{E_1, E_2, \ldots, E_m\}$，SEE 通过嵌入层和额外的线性投影层，将实体类别 $E_i$ 映射到向量空间，从而计算相应的实体类别嵌入:

$$t_i = \text{Linear}_{type}(\text{Embed}_{type}(i)) \tag{3.1}$$

其中 $i$ 是实体类别的索引，$t_i \in \mathbb{R}^d$ 代表第 $i$ 个实体类别的嵌入表示；$\text{Embed}_{type} \in \mathbb{R}^{m \times 3d}$ 代表实体类别嵌入层; $\text{Linear}_{type} \in \mathbb{R}^{3d \times d}$ 代表实体类别端的线性投影层。

### 3.1.2　跨度嵌入表示初始化

对于一个包含 $N$ 个词元的句子 $X = [x_1, x_2, ..., x_N]$，SEE 模型使用上下文编码器 BERT 对其进行编码处理。具体而言，BERT[61] 作为 Transformer[62] 模型，其中的注意力机制能够在编码每个词元的过程中同时捕捉句子中的上下文信息。此外，句子在编码前的分词操作后会产生两个额外的词元 [cls] 和 [sep]，其中 [cls] 蕴含整个句子的语义信息，[sep] 标识句子的结尾。通过将句子输入到 BERT 中，编码器会输出一个序列:

$$h_{[\text{cls}]}, h_1, h_2, \ldots, h_N, h_{[\text{sep}]} = \text{BERT}(x_1, x_2, \ldots, x_N) \tag{3.2}$$

其中 $h_{[\text{cls}]}, h_1, h_2, \ldots\ h_N, h_{[\text{sep}]} \in \mathbb{R}^{3d}$ 代表每个词元编码后的特征表示。

由于 SEE 是基于跨度的模型，因此需要为每个候选跨度构建有效的嵌入表示。对于一个由起始词元坐标 $i$ 和结尾词元坐标 $j$ （满足条件 $j \geq i$）所定义的候选跨度 $span(i,j)$, SEE 模型首先借鉴了 YU 的做法[24]，通过拼接该跨度的起始词元向量、结尾词元向量以及编码了跨度长度信息的嵌入向量，来初始化该跨度的表示：

$$span_{raw}(i,j) = h_i \oplus h_j \oplus \text{Embed}_{length}(\text{j - i}) \tag{3.3}$$

其中 $h_i$ 和 $h_j$ 分别是跨度的开头和结尾词元的特征表示；$\text{Embed}_{length} \in \mathbb{R}^{N \times d}$ 代表跨度长度嵌入层；$span_{\text{raw}} \in \mathbb{R}^{N \times N \times 7d}$ 代表所有候选跨度的初始嵌入矩阵，该矩阵的构建方式，即枚举所有词元的两两组合以覆盖所有潜在跨度。

### 3.1.3 ECNN 跨度增强模块

之前也有一些工作[31-32]使用 CNN 对跨度矩阵中的空间关系进行建模，并取得了一定成效。然而，这些方法未能充分考虑跨度矩阵的独特性，将其视为与图像同质，忽略了两者本质上的归纳偏置差异。具体而言，图像中像素之间的空间关系复杂多变且高度依赖内容，CNN 通过学习这些局部结构模式来提取特征。然而，跨度矩阵中跨度间的关联性更为结构化，如第 2.1 章所述，右上的跨度在文本上蕴含了左下方子跨度的语义信息。传统的跨度表示方法常仅依赖首尾词元，导致跨度越长（对应跨度矩阵越靠右上），中间语义信息损失越多。基于上述分析，本章提出一种基于偶数卷积的跨度增强模块 ECNN，利用其不规则感受野侧重捕捉左下子跨度信息的特性，进而有效补充长跨度表示中所损失的特征信息，增强跨度的语义表示能力。

本章通过卷积核相对于中心的左上、右上、左下和右下的偏移值来描述感受野 $R = [p_{lt}, p_{rt}, p_{lb}, p_{rb}]$，通常卷积神经网络使用的卷积核大小为奇数，故以下述方式表示奇数卷积核卷积神经网络的感受野 $R^O$：

$$R^O = [(-k,-k),(k,-k),(-k,k),(k,k)], k = \lceil \frac{c-1}{2} \rceil \tag{3.4}$$

其中 $k$ 表示从四个边到原点的最大像素数；$\lceil . \rceil$ 是向上舍入函数；$c$ 是卷积核大小。

Wu 等人[75]的研究提到了偶数卷积核的特殊性，由于其在卷积操作时没有明确的中心点，卷积过程不可避免地会产生不对称性。在大多数深度学习框架（如 TensorFlow 和 PyTorch）中，这种不对称性通常被忽略，并通过预定义的偏移量来掩盖，

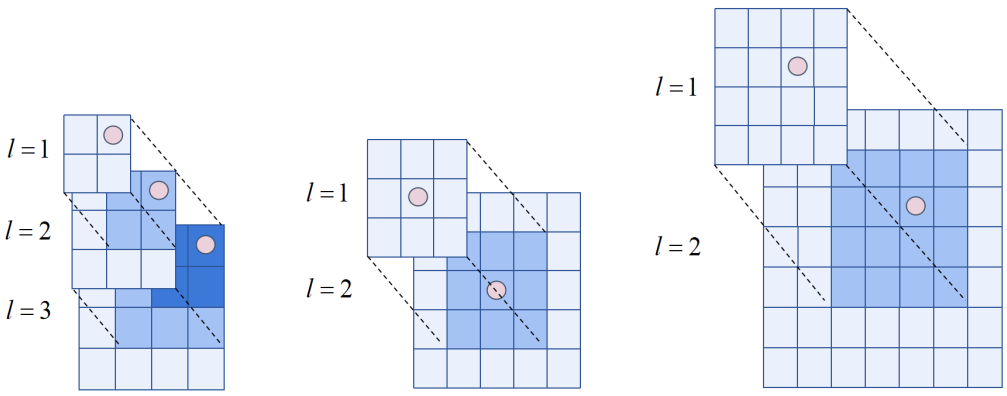
默认选择中心左上角的像素作为卷积中心，感受野 $R^E$ 表示如下:

$$R^E = [(1-k, 1-k), (1+k, 1-k), (1-k, 1+k), (1+k, 1+k)], k = \lceil \frac{c-1}{2} \rceil \quad (3.5)$$

由于本章方法旨在重点关注左下方的子跨度信息，故对偶数卷积核的默认中心进行了调整，选择其偏右上角的像素作为新的卷积中心。当卷积神经网络堆叠了 $L$ 层，奇数卷积核心的 CNN 感受野 $R_L^o$ 和偶数卷积核心的 CNN 感受野 $R_L^E$ 可以分别表示为:

$$R^O_{L=l} = [(-l \cdot k, -l \cdot k), (l \cdot k, -l \cdot k), (-l \cdot k, l \cdot k), (l \cdot k, l \cdot k)], k = \lceil \frac{c-1}{2} \rceil \quad (3.6)$$

$$R^E_{L=l} = [l(-1-k, 1-k), l(k-1\ 1-k), l(-1-k, 1+k), l(k-1, 1+k)], k = \lceil \frac{c-1}{2} \rceil \quad (3.7)$$




（a）$c = 2$　　（b）$c = 3$　　（c）$c = 4$

**图** 3.2　不同卷积核及层数下的感受野示意图。圆心代表卷积核的中心，高层与底层感受野的重叠部分已做颜色加深处理，以方便观察其范围变化。

如图 3.2 所示，随着卷积层的堆叠，ECNN 的感受野呈现出显著的不对称性。尤其在使用 2×2 卷积核时，其感受野几乎完全集中在左下区域。相较之下，传统 CNN 的奇数核感受野则是中心对称的。基于跨度矩阵中任意跨度其子跨度都处于左下角的特性，ECNN 可以利用偶数卷积核的非对称感受野，使其在更新跨度的表示时，能够侧重于聚合其内部子跨度的信息。这种非对称的聚合方式比对称聚合更能有效地

建模跨度的层次化语义结构，从而为长跨度实体补充了被边界表示法忽略的关键内部信息。因此 SEE 模型利用 ECNN 对跨度表示进行增强：

$$span = \text{GELU}(\text{LayerNorm}(span_{raw} + \text{EvenConv}(f(span_{raw})))) \tag{3.8}$$

其中 $f(\cdot)$ 是一个将跨度矩阵左下半元素置零的操作; EvenConv，LayerNorm 和 GeLU 分别是卷积、层归一化和激活函数。

最后，通过一个线性投影层将跨度嵌入映射到和类别嵌入相同的向量空间中：

$$s = \text{Linear}_{span}(span) \tag{3.9}$$

其中 $s \in \mathbb{R}^{N\times N\times d}$ 代表增强后的候选跨度嵌入矩阵；$\text{Linear}_{span} \in \mathbb{R}^{7d\times d}$ 代表文本跨度端的线性投影层。

### 3.1.4 基于对比学习的训练和推理

基于前文讨论的实体类别嵌入与跨度嵌入，SEE 模型的优化过程采用了两种不同的对比学习优化目标：其一是 Zhang 等人[1] 提出的基于跨度维度的对比学习目标；其二是本章提出的基于实体类别维度的对比学习目标。本章采用 InfoNce[51] 作为损失函数。相似度的计算则基于余弦距离，并通过一个可调节的温度系数 $\tau$ 进行缩放，其公式为 $\text{sim} = \cos(\cdot)/\tau$。

**基于跨度维度的优化目标:** 对于跨度特征矩阵 s 中的跨度 $s(i,j)$，基于跨度的损失函数可以定义为如下形式：

$$l_{\text{span}} = -\log \frac{\exp(\text{sim}(s_{i,j}, t_m))}{\sum_{s'\in S_m^- \cup s_{i,j}} \exp(\text{sim}(s', t_m))} \tag{3.10}$$

其中 $s_{i,j}$ 是所有属于类别 $E_k$ 的正样本跨度嵌入向量集合；$S_m^-$ 是所有不属于 $E_m$ 的负样本跨度嵌入向量集合，它们的并集是输入文本中所有跨度嵌入向量集合；$t_m$ 是类别 $E_m$ 的嵌入向量。

在基于 Transformer 的语义表征体系中，[cls] 向量通过自注意力机制聚合了全局上下文信息，其语义空间分布同时受到正样本跨度与负样本跨度的联合约束。因此，在一个经过充分优化的表示空间中，[cls] 向量与某个实体类别的相似度，其理想位置应该介于该类别下的正样本跨度与负样本跨度之间。基于这一内在约束，可设计

出如下新的基于跨度的优化目标：

$$l^{+}_{\text{span}} = \beta l_{\text{span}} - (1-\beta)\log \frac{\exp(\text{sim}(s_{0,0}, t_m))}{\sum_{s' \in S^{-}_{m}} \exp(\text{sim}(s', t_m))} \tag{3.11}$$

其中 $s_{0,0}$ 是 [cls] 的跨度嵌入向量，因为 [cls] 的开头和结尾坐标都是 0；$\beta$ 是可调整的超参数。

**基于实体类别维度的优化目标:** 当前多模态对比学习框架[74]通常采用跨模态相似度对齐策略，其核心是通过异构模态间的余弦相似度矩阵进行联合优化。在命名实体识别任务中，这种范式被调整为计算候选跨度与预定义实体类别嵌入之间的语义相似度。然而，现有方法在优化时仅关注跨度维度，忽略了实体类别维度的潜在优化空间。因此，本文在原有框架的基础上，加入了一个基于实体类别维度的优化目标。这种优化策略能进一步增强模型在区分类别时的能力，基于实体类别的损失函数可以定义为如下形式：

$$l_{\text{type}} = -\log \frac{\exp(\text{sim}(s_{i,j}, t_m))}{\sum_{t' \in t} \exp(\text{sim}(s_{i,j}, t'))} \tag{3.12}$$

其中 $t_m$ 是跨度 $s_{i,j}$ 所属于类别的嵌入表示；$t$ 是除去 $t_m$ 以外的类别嵌入集合。

最后，结合上述两个不同维度的对比学习优化目标，最终的损失函数定义如下所示：

$$L = \alpha l^{+}_{\text{span}} + (1-\alpha) l_{\text{type}} \tag{3.13}$$

其中 $\alpha$ 是可调整的超参数。

由于 SEE 是基于对比学习的模型，其输出的结果是跨度和类别的相似度矩阵。在推理过程中，SEE 枚举所有潜在跨度，并基于每个跨度嵌入和每个实体类别嵌入计算相似度分数。随后，模型会进行筛选，仅保留那些相似度分数高于特定阈值的跨度，并将其预测为正样本。具体而言，即保留满足条件 $\text{sim}(s_{i,j}, t_m) > \text{sim}(s_{0,0}, t_m)$ 的跨度。通过这一系列计算，模型能够直接得到支持实体嵌套的命名实体识别结果。若目标是处理非嵌套任务，则额外需要一个去重的后处理步骤。完整的推理算法如算法 3.1 所示。

**算法 3.1:** SEE 模型推理算法

**输入:** 跨度集合 $S = \{(i, j) \mid i, j = 1 \ \ldots, N, 0 \leq j - i \leq L\}$ 表示所有可能的跨度，其中 $i$ 和 $j$ 分别是跨度的起始和结束位置，$N$ 是序列的最大长度，而 $L$ 是允许的最大跨度长度。集合 $\epsilon = \{E_1, \ldots, E_m\}$ 表示所有实体类别的集合，*flat* 表示推断是否针对 flat NER。

**输出:** 分配实体类别的跨度集合 M。

$M \leftarrow \{\}$ // 定义跨度集合 $M$
**for** $E_m \in \epsilon$ **do** // 迭代类别
  计算阈值分数 $s_{\text{thred}}$
  **for** $(i, j) \in S$ **do** // 迭代跨度
    计算跨度相似度分数 $s_{i,j,m}$
    **if** $s_{i,j,m} > s_{thred}$ **then** // 判断跨度和类别的相似度是否大于阈值
      $M \leftarrow M \cup \{(i, j, E_m)\}$; // 将该分类结果加入跨度集合 $M$
    **end**
  **end**
**end**
**if** *flat is true* **then** // 判断任务是否是 flat NER
  $\hat{M} \leftarrow \{\}$ // 定义跨度集合 $\hat{M}$
  将 $M$ 按以下规则排序：
    1. 按相似度分数 $s_{i,j,m}$ 降序排列
    2. 若分数相同，按起始位置 $i$ 升序排列
    3. 若 $i$ 相同，按结束位置 $j$ 升序排列
  **for** $(i, j, E_m) \in M$ **do** // 迭代分类结果
    **if** *no span in* $\hat{M}$ *overlaps with* $(i, j)$ **then**
      $\hat{M} \leftarrow \hat{M} \cup \{(i, j, E_m)\}$ // 将该分类结果加入跨度集合 $\hat{M}$
    **end**
  **end**
  $M \leftarrow \hat{M}$
**end**
**return** $M$;

## 3.2 实验分析

### 3.2.1 数据集介绍

为了验证 SEE 模型识别命名实体的有效性和泛化性，本章选取了五个命名实体识别公开数据集进行了实验，其中 ACE2005[76]、ACE2004[77]和 Genia[78]是非嵌套数据集，Ontonotes[79]和 CoNLL03[80]是嵌套数据集。

**（1）ACE 数据集**

本章实验中的 ACE2004 和 ACE2005 数据集均源自语言数据联盟（Linguistic Data Consortium，LDC）发布的多语言训练语料库。该语料库的文本主要源自新闻领域，内容涵盖英语、阿拉伯语和中文，并包含了对实体、关系和事件的精细标注。本章的实验与评估仅使用了其中的英文命名实体识别部分。ACE 数据集中包含七类命名实体，分别为 PER、ORG、LOC、FAC、WEA、VEH 和 GPE，各类别具体含义见表 3.1。在数据划分方面，ACE2005 数据集按照 Luan 等人[81]提出的方案划分为训练集、验证集和测试集。ACE2004 数据集的划分则遵循 Lu and Roth[82]及 Muis and Lu[83]的处理方案，将数据按比例 8:1:1 划分为训练集、验证集与测试集。

**表** 3.1　ACE2004 与 ACE2005 数据集实体类别及描述

| 序号 | 实体类别 | 描述 |
|---|---|---|
| 1 | PER | 人类个体或群体 |
| 2 | ORG | 公司、企业、机构、院校和其他人群组成的团体 |
| 3 | VEH | 移动、牵引或运输物体的物理设施 |
| 4 | GPE | 由政治或社会群体定义的地理区域 |
| 5 | FAC | 人造建筑物或场所 |
| 6 | WEA | 用于造成物理伤害的工具 |
| 7 | LOC | 地理位置 |

**（2）GENIA 数据集**

本章实验所用的 GENIA 数据集，源自 GENIA 项目所构建的同名生物医学文本语料库，其版本为 v3.0.2。该语料库主要用于支持分子生物学领域的信息抽取与文本挖掘系统的开发和评估。GENIA v3.0.2 包含共 1999 篇 Medline 文献摘要，这些摘要通过 PubMed 查询 MeSH 术语 “human”、“blood cells” 以及 “transcription factors” 得到。语料库在多个语言和语义层面进行了注释，其中包括词性标注、句法结构、语义角色及命名实体等信息。本章仅关注其中的命名实体识别部分，数据集中共标注了

五类生物医学实体类别，分别为 DNA、RNA、Protein、cell_line 和 cell_type。各实体类别的定义与示例详见表 3.2。在数据划分方面，本章遵循 Finkel and Manning[84] 以及 Lu and Roth[82] 的设置，将 GENIA v3.0.2 语料库按 8:1:1 的比例划分为训练集、验证集和测试集。

**（3）CoNLL03 数据集**

本章实验使用了 CoNLL03 命名实体识别数据集。该数据集是 CoNLL-2003 共享任务的一部分，涵盖英语和德语两种语言，本章仅采用其中的英文部分进行实验。CoNLL-2003 的英文数据源自路透社语料库，其内容为 1996 年 8 月至 1997 年 8 月期间的新闻报道。该数据集共标注了四类命名实体，分别为 PER、ORG、LOC 和 MISC。各实体类别的定义与示例详见表 3.3 。在数据划分方面，本章遵循了原始文件的数据划分格式。

**表** 3.2　GENIA 数据集实体类别及描述

| 序号 | 实体类别 | 描述 |
|---|---|---|
| 1 | DNA | 由多糖-磷酸骨架构成，并带有嘌呤和嘧啶的突出部分，形成通过这些嘌呤和嘧啶之间的氢键结合而成的双螺旋结构。 |
| 2 | RNA | 一种多核苷酸，主要由重复的磷酸和核糖单元构成的链组成，氮碱基附着在其上。 |
| 3 | protein | 在核糖体上合成的线性多肽，可能经过进一步修饰、交联、切割或组装成具有多个亚基的复杂蛋白质。氨基酸的特定序列决定了多肽在蛋白质折叠过程中形成的形状以及蛋白质的功能。 |
| 4 | cell_line | 在适于其生长的特殊培养基中进行体外培养的细胞。培养细胞用于研究发育、形态、代谢、生理和遗传过程等。 |
| 5 | cell_type | 构成生物体的基本结构和功能单元或亚单元。它们由包含各种细胞器的细胞质和细胞膜边界组成。 |

**（4）OntoNotes 数据集**

本章实验所用的 OntoNotes 数据集，是大型多语种语料库 OntoNotes 5.0 中的命名实体识别部分。该语料库涵盖了多种文本类别，包括新闻、电话交谈语音、博客、Usenet 新闻组、广播、脱口秀等，并包含英语、中文和阿拉伯语三种语言。OntoNotes 5.0

**表** 3.3　CoNLL03 数据集实体类别及描述

| 序号 | 实体类别 | 描述 |
|---|---|---|
| 1 | PER | 人类个体或群体 |
| 2 | ORG | 公司、企业、机构、院校和其他人群组成的团。 |
| 3 | LOC | 地理位置。 |
| 4 | MISC | 不属于上述类别的其他命名实体 |

提供了结构信息和浅层语义信息。本研究仅采用其中的英文部分进行实验。该部分数据共包含十八种实体类别，即 PERSON、NORP、FAC、ORG、GPE、LOC、PRODUCT、DATE、TIME、PERCENT、MONEY、QUANTITY、ORDINAL、CARDINAL、EVENT、WORK_OF_ART、LAW 及 LANGUAGE。各实体类别的定义与示例详表 3.4。在数据划分方面，本章严格遵循 Gu 等人[69]的处理方法，对数据集进行了预处理和划分。

**表** 3.4　OntoNotes 5.0 数据集实体类别及描述

| 序号 | 实体类别 | 描述 |
|---|---|---|
| 1 | PERSON | 人类个体或群体 |
| 2 | NORP | 民族、宗教或政治团体 |
| 3 | FAC | 人造建筑物或场所 |
| 4 | ORG | 公司、企业、机构、院校和其他人群组成的团体 |
| 5 | GPE | 由政治或社会群体定义的地理区域 |
| 6 | LOC | 地理位置 |
| 7 | PRODUCT | 产品、商品、车辆 |
| 8 | DATE | 日期 |
| 9 | TIME | 时间 |
| 10 | PERCENT | 百分比 |
| 11 | MONEY | 货币金额 |
| 12 | QUANTITY | 数量、度量单位 |
| 13 | ORDINAL | 序数词（first，second 等） |
| 14 | CARDINAL | 基数词（one，two 等） |
| 15 | EVENT | 自然灾害、战争、体育赛事等事件 |
| 16 | WORK_OF_ART | 书籍、歌曲等艺术作品的标题 |
| 17 | LAW | 成文法律、法规 |
| 18 | LANGUAGE | 语言名称 |

### 3.2.2 评价指标

为全面评估模型在命名实体识别任务中的性能，本章遵循该领域的通用评估方法，采用精确率（Precision，P）、召回率（Recall，R）以及它们的调和平均值 F1 值（F1-Score，F1）作为评价指标。这些指标均基于混淆矩阵计算得出，而混淆矩阵则反映了模型预测结果与真实标注之间的一致性程度。

如表 3.5 所示，混淆矩阵是一个 2×2 的表格，用于总结分类任务的预测结果。以命名实体识别任务为例，通常将识别实体视为正例，不识别为负例。矩阵的每一行代表数据的真实类别，每一列代表模型的预测类别。

**表** 3.5　混淆矩阵

| | 预测为正例 | 预测为负例 |
|---|---|---|
| 标注为正例 | 真正例（TP） | 假负例（FN） |
| 标注为负例 | 假正例（FP） | 真负例（TN） |

精确率衡量的是模型预测为实体的实例中，真正是实体的比例。它反映了模型预测的准确性。较高的精确率意味着模型在识别实体时，很少将非实体错误地判定为实体，降低了“误报”的比例，精确率的计算公式如下：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.14}$$

召回率衡量的是模型能够正确识别出所有真实实体的能力。它表示在所有实际为实体的实例中，模型成功识别出的比例。较高的召回率意味着模型能够覆盖更多的真实实体，减少了“漏报”的情况，召回率的计算公式如下：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.15}$$

F1 值是精确率和召回率的调和平均数，它综合考虑了模型的精确率和召回率，提供了一个更全面的性能评估指标。F1 值更侧重于精确率和召回率中的较低值，只有当两者都较高时，F1 值才会得到一个较高的分数。F1 值的计算公式如下：

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.16}$$

理想的模型应该同时具备高精确率和高召回率。高精确率意味着模型预测的实体边界和类别更加准确，减少了错误的实体识别，而高召回率则意味着模型能够尽

可能多地识别出文本中所有真正的实体，避免遗漏。F1 分数综合了精确率和召回率，能够更平衡地反映模型在分类任务中的整体性能，因此被广泛采用。

### 3.2.3 实验设置

本章实验根据目标数据集的领域特性，针对性地选择了不同的预训练语言模型作为编码器。具体而言，对于 ACE2004、ACE2005、CoNLL03 和 OntoNotes 这四个通用领域数据集，本章分别基于 BERT-base[61] 和 RoBERTa-base[63] 进行了实验，而针对生物医学领域的 Genia 数据集，考虑到其专业性，本章采用了在该领域预训练的 PubMedBERT-base-uncased[69] 模型。模型的关键超参数根据所选编码器进行了适配，以期达到各自场景下的最优性能：

- 当编码器为 BERT-base 或 PubMedBERT-base-uncased 时，对比损失权重 $\alpha$ 设置为 0.5，学习率为 $3\times10^{-5}$，批量大小为 8，最大输入序列长度 $N$ 设置为 128。在此配置下，模型总计训练 20 个 epoch。
- 当编码器为 RoBERTa-base 时，对比损失权重 $\alpha$ 设置为 0.8，学习率为 $3\times10^{-5}$，批量大小为 8，最大输入序列长度 $N$ 扩展至 256。在此配置下，模型总计训练 40 个 epoch。

在本章训练过程中，模型统一使用 AdamW 优化器进行参数更新。对于本章提出的 ECNN 模块，其配置固定为 4 个卷积层，每层均采用 2×2 尺寸的卷积核。本章实验采取了标准的验证策略：每完成 50 个训练步骤，就在对应的验证集上以 F1 分数为指标评估模型性能，并保存性能最佳的模型状态。最终报告的测试集结果均来自于加载该最佳模型状态进行的评估。为了确保实验结果的稳定性和可靠性，本章报告的所有实验结果均为五次独立实验运行的中位数结果。

### 3.2.4 消融实验

为了系统性地评估本论文提出的 SEE 模型中各个核心组件的实际贡献，并验证关键设计选择的有效性，本章进行了一系列详尽的消融实验。这些实验旨在剥离或改变模型的特定部分以量化它们各自对模型最终性能的影响。所有消融实验均在 ACE2005 数据集上进行。

首先，本章考察了 ECNN 模块与基于类别的对比损失这两个核心创新点对模型性能的独立及联合贡献，结果如表 3.6 所示。实验中，ECNN 配置为 4 个 2×2 的卷

积层，编码器有 BERT-base 和 RoBERTa-base。对于 BERT-base 编码器，单独加入基于类别的对比损失后，模型的精确率有明显提升，从 88.5% 增至 89.5%，但召回率略有下降，使得 F1 值小幅提升；而单独加入 ECNN 模块，同样带来了显著的精确率提升，同时召回率也有小幅增加，使得 F1 值提升了 0.6%。对于 RoBERTa-base 编码器，各组件的效果更为突出。单独加入基于类别的对比损失后，精确率提升显著，从 89.6% 增至 90.3%，召回率也有提升，从 90.6% 增至 91.2%，最终 F1 提升了 0.7% 并达到 90.8%。而单独加入 ECNN 模块，精确率的提升幅度更大，增加了 1.3% 提升到 90.9%，尽管召回率略微下降，但依然带来了 0.4% 的 F1 提升达到 90.5%。当同时引入 ECNN 模块和基于类别的对比损失时，模型性能达到了最佳状态。特别是在 RoBERTa-base 上，精确率大幅提升至 91.4%，召回率提升了 0.3% 并达到 91.1%，F1 增长了 1.0% ，最终达到 91.1%。这些结果表明，两个模块都能独立提升性能，尤其对精确率的贡献更为一致和显著，并且它们之间存在积极的协同效应，共同优化了模型的整体识别性能。

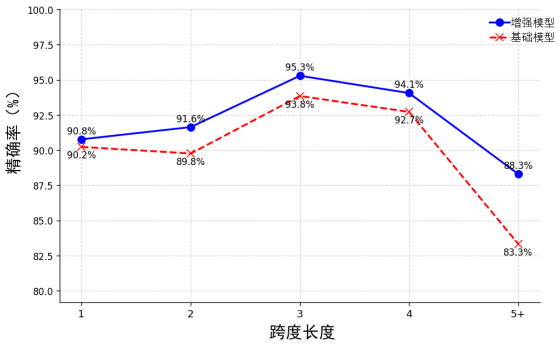
**表** 3.6　ECNN 模块在 ACE2005 上，不同配置的消融实验结果

| ECNN 模块 | 类别损失 | BERT-base | | | RoBERTa-base | | |
|---|---|---|---|---|---|---|---|
| | | P(%) | R(%) | F1(%) | P(%) | R(%) | F1(%) |
| | | 88.5 | 90.0 | 89.3 | 89.6 | 90.6 | 90.1 |
| | ✓ | 89.5 | 89.7 | 89.6 | 90.3 | 91.2 | 90.8 |
| ✓ | | 89.5 | 90.3 | 89.9 | 90.9 | 90.2 | 90.5 |
| ✓ | ✓ | 89.7 | 90.7 | 90.2 | 91.4 | 90.9 | 91.1 |

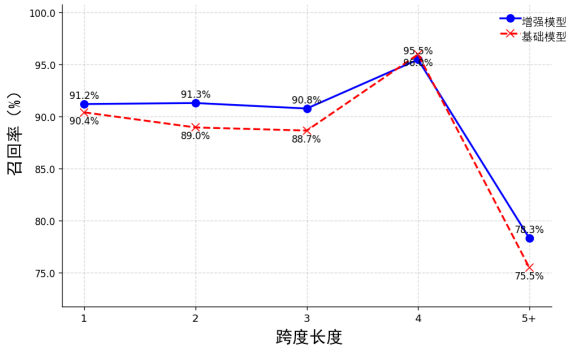
为了更深入地探究 ECNN 模块提升性能的具体机制，特别是其在处理不同长度实体时的作用，本章进一步分析了模型在不同跨度长度下的表现。基于跨度的命名实体识别方法，尤其是那些主要依赖边界词元构建跨度表示的模型，往往在处理较长实体时表现不佳。本研究假设通过 ECNN 模块补充跨度内部的语义信息能够改善这一状况。表 3.7 和图 3.3 展示了在 ACE2005 数据集上，未加入 ECNN 模块的基础模型与加入了 ECNN 模块的增强模型在精确率、召回率和 F1 分数三个指标上随实体跨度长度变化的性能表现以及趋势。其中 $\Delta$ 符号代表增强模型相对于基础模型的性能差异。

**表 3.7** ECNN 模块对模型识别不同长度实体跨度的定量分析

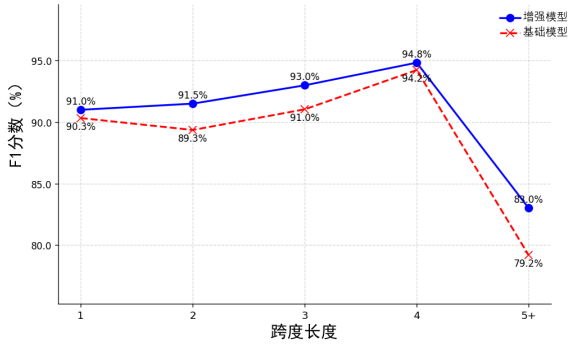
| 跨度长度 | P(%) | | R(%) | | F1(%) | |
|---|---|---|---|---|---|---|
| | 值 | Δ | 值 | Δ | 值 | Δ |
| 1 | 90.8 | +0.5 | 91.2 | +0.8 | 91.0 | +0.7 |
| 2 | 91.6 | +1.9 | 91.3 | +2.4 | 91.5 | +2.1 |
| 3 | 95.3 | +1.4 | 90.8 | +2.1 | 93.0 | +2.0 |
| 4 | 94.1 | +1.3 | 95.5 | −0.5 | 94.8 | +0.6 |
| 5+ | 88.3 | +5.0 | 78.3 | +2.8 | 83.0 | +3.8 |




(a) 精确率对比




(b) 召回率对比




(c) F1 分数对比

**图 3.3** ECNN 模块对模型识别不同长度实体跨度的影响

实验结果可以观察到，实体跨度长度对模型性能有着显著影响。对于长度为 1 的实体，即单个词元构成的跨度，基础模型与增强模型的各项指标均十分接近，这表明在这种简单情况下，模型本身已具有较好的性能，ECNN 模块带来的额外增益有限。然而，当跨度长度开始增长后，引入 ECNN 模块的增强模型相比基础模型展现出明显的性能优势。从图 3.3 所呈现趋势上来看，在精确率方面，增强模型的曲线

始终位于基础模型之上，尤其在长度为 5 及以上跨度的表现差距更为显著，相较基础模型提升高达 5 个百分点，这有力表明 ECNN 模块能够有效捕捉跨度内部的语义信息，从而帮助模型更准确地进行实体分类和边界判断，显著减少误报。在召回率方面，增强模型除了 4 以外所有长度都超过了基础模型的表现，这一结果进一步证明 ECNN 能够增强模型识别那些仅依赖边界信息难以确定的实体的能力，有效降低了漏报率。在综合性能指标 F1 分数的表现中，增强模型的性能曲线同样在所有跨度长度上均持续高于基础模型，充分体现了本文提出的 ECNN 模块在处理不同长度实体时的整体有效性和鲁棒性。综上所述，实验结果有力地支持了本文的初始假设：通过显式地建模和利用跨度内部的语义信息，ECNN 模块能够有效弥补传统方法在长跨度表示上的不足，显著提升模型对于长跨度实体识别的综合性能。

为了证明偶数卷积核的优越性，本章继续深入探究不同卷积核配置对模型性能的影响。由于数据集中长度小于等于 4 的跨度占比最多，同时考虑到计算资源的限制，本章选择了四种不同配置的卷积神经网络进行试验，分别为：2 层 2×2 卷积核 CNN、3 层 2×2 卷积核 CNN、1 层 3×3 卷积核 CNN 以及 1 层 4×4 卷积核 CNN。这四种卷积神经网络的参数量逐级增大。

各种尺寸卷积核的实验结果及参数量如表 3.8 所示。其中，1 个 3×3 卷积核的表现最差，未能对模型整体性能产生提升；相比之下，2 个 2×2 和 1 个 4×4 卷积核的性能相似，而 3 个 2×2 卷积核的组合则取得了最佳的实验效果。这表明采用偶数卷积核的卷积神经网络能够有效增强跨度表示，显著提高模型整体性能。进一步分析模型参数发现，采用 2 个 2×2 卷积核的网络参数量仅为 1 个 3×3 卷积核和 1 个 4×4 卷积核网络的 0.8 倍和 0.5 倍，却实现了相当甚至更优的性能，这意味着 2×2 卷积核的 CNN 结构在保持性能优势的同时，也更具备资源高效性和应用灵活性。综上所述，2×2 卷积核的 CNN 在保持高性能的同时对计算资源要求最低，是 ECNN 模块的最佳卷积核心选择。

此外，本章考察了两种对比损失之间的权重系数 $\alpha$ 对模型性能的影响，结果如图 3.4 所示。结果显示，$\alpha$ 的取值对模型的三个指标均有影响。随着 α 的增大，精确率呈现较明显的上升趋势，并在 $\alpha = 0.9$ 时达到峰值；F1 值也随之整体上升。而召回率则在 $\alpha$ 约为 0.6 时达到顶峰，之后随着 $\alpha$ 增大反而下降。这可能反映了两种损失的侧重不同：基于跨度的损失更侧重于提升判别的精确率，而基于类别的损失则有助于保证识别的召回率。在实验中，模型依据验证集上的 F1 表现选择了最优的 $\alpha$

**表** 3.8　不同卷积核尺寸与层数在 ACE2005 数据集上的性能比较

| 卷积核尺寸 (层数) | P (%) | R (%) | F1 (%) | 参数 (M) |
|---|---|---|---|---|
| 2×2 (2) | 89.3 | 90.3 | 89.8 | 4.71 |
| 3×3 (1) | 89.0 | 89.8 | 89.4 | 5.30 |
| 2×2 (3) | 89.5 | 90.3 | 89.9 | 7.07 |
| 4×4 (1) | 89.4 | 90.3 | 89.8 | 9.43 |

值（$\alpha = 0.8$），以求在精确率和召回率之间取得最佳平衡。

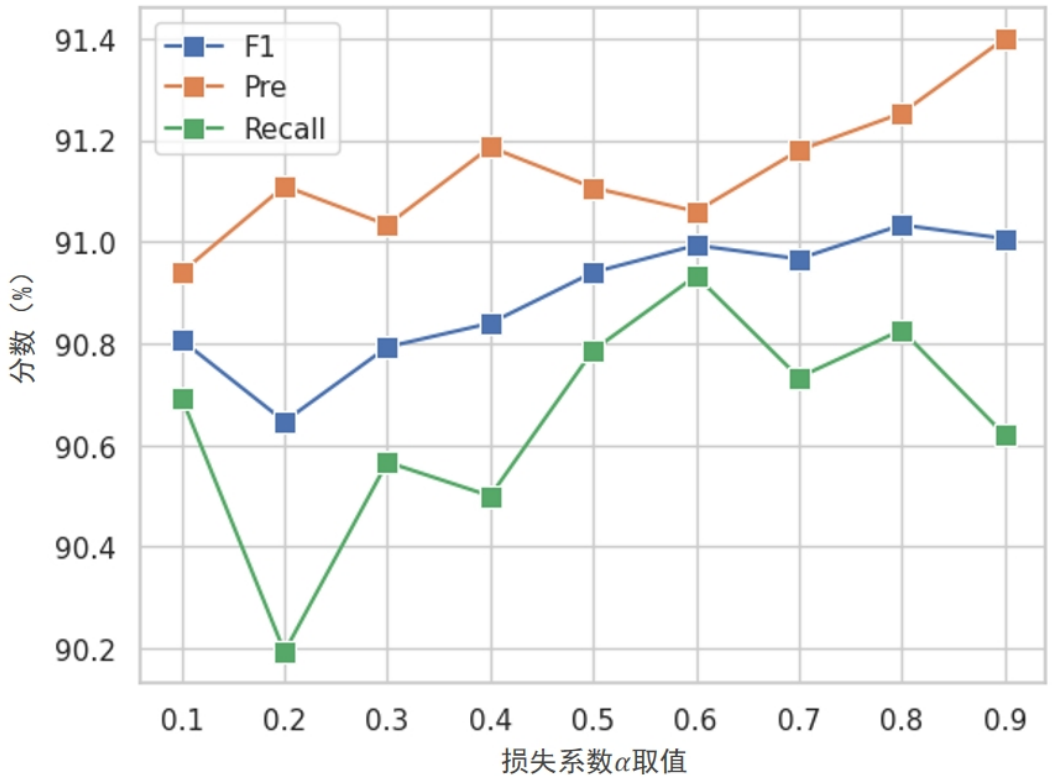



**图** 3.4　不同损失系数 $\alpha$ 下，模型的性能趋势

在现有基于对比学习的命名实体识别框架中[1]，类别嵌入方法通常会先收集数据集中各类别对应的注释指南，并将其作为该类别的文本描述。随后通过预训练模型对这些预定义描述进行编码，并提取 [cls] 对应的向量作为该类别的嵌入表示。然而，在基于对比学习的命名实体识别架构中，类别端可用于编码的数据量仅与类别数量相等，这种数据稀缺性可能限制了预训练模型充分发挥其语义理解能力，也难以使模型从有限的类别描述中有效学习到丰富的语义信息。

基于这一猜想，本章对类别嵌入的表示方式展开了消融研究。表 3.9 中研究对

比了三种不同的类别嵌入策略：使用预训练模型编码注释指南、使用预训练模型编码类别数字编号，以及直接使用可学习嵌入层。实验结果表明，在最终模型性能上，这三种嵌入方式的差异并不显著。甚至采用可学习嵌入层的方法在 F1 值上展现出微弱的领先。考虑到可学习嵌入层在显著减少模型参数量、提升计算效率方面的优势，本研究最终选择以可学习嵌入层来表示类别嵌入。

**表** 3.9　不同类别嵌入表示方法的性能比较

| 方法 | 预训练模型 | P(%) | R(%) | F1(%) | 参数 (M) |
|---|---|---|---|---|---|
| 注释指南 | ✓ | 89.7 | 90.2 | 90.0 | 21.90 |
| 数字 | ✓ | 89.8 | 90.1 | 90.0 | 21.90 |
| 嵌入层 | | 90.4 | 90.0 | 90.2 | 11.01 |

### 3.2.5 对比实验

为全面评估本文所提 SEE 模型在当前技术格局下的性能水平，本章选择了一系列涵盖不同技术范式的代表性基线模型，并与之进行了广泛比较。这些基线方法包括：基于序列标注的模型 Pyramid[85]；基于生成的模型 Sequence-To-Set[22]、BART-NER[21]、De-Bias[86]和 DiffusionNER[3]；基于 LLM 的模型 GPT-NER[87]和 Padellm-NER[88]；以及基于跨度的模型 Biaffine[24]、SLG-NER[26]、BINDER[1]和 SGT[30]。

本章选取了五个广泛使用的基准数据集进行了系统的对比实验。这些数据集覆盖了不同的任务特性，包括三个嵌套命名实体识别数据集 ACE2004、ACE2005、GENIA 以及两个非嵌套命名实体识别数据集 CoNLL03、OntoNotes。具体的实验结果与各基线方法的详细比较分别呈现在表 3.10、表 3.11 及表 3.12 中。在这些结果表格中，本章遵循了通用的性能标注约定：最优结果以**粗体标识**，次优结果以下划线标识。同时，为了清晰展示各方法所依赖的基础模型，表格中也列出了所使用的预训练编码器，其缩写说明如下：Bl 代表 BERT-large；BioB 代表 BioBERT；BioBl 代表 BioBERT-large；BAl 代表 BART-large；T5b 代表 T5-base；Bb 代表 BERT-base；Rb 代表 Roberta-base；Pub 代表 PubMedBERT-base；而 LLM 则指代大语言模型（Large Language Model）。

在通用领域嵌套命名实体识别基准数据集 ACE2004 和 ACE2005 上，SEE 模型展现了卓越的性能。如表 3.10 所示，在 ACE2004 数据集上，基于 RoBERTa-base 编码器的 SEE 模型在数据的 F1 分数达到了 90.6% ，超越了该表中所列的所有基线方

**表** 3.10　不同方法在嵌套命名实体识别数据集上的定量结果对比

| | 编码器 | ACE2004 | | | ACE2005 | | | GENIA | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | P(%) | R(%) | F1(%) | P(%) | R(%) | F1(%) | P(%) | R(%) | F1(%) |
| Pyramid[85] | Bl/BioB | 86.1 | 86.5 | 86.3 | 84.0 | 85.4 | 84.7 | 79.5 | 78.9 | 79.2 |
| Biaffine[24] | Bl/BioB | 87.3 | 86.0 | 86.7 | 85.2 | 85.6 | 85.4 | 81.8 | 79.3 | 80.5 |
| Sequence-To-Set[22] | Bl/BioB | 88.5 | 86.1 | 87.3 | 87.5 | 86.6 | 87.1 | 82.3 | 78.7 | 80.4 |
| BART-NER[21] | BAl | 87.3 | 86.4 | 86.8 | 83.2 | 86.4 | 84.7 | 78.6 | 79.3 | 78.9 |
| De-Bias[86] | T5b | 86.5 | 84.5 | 85.4 | 83.3 | 86.6 | 84.9 | 81.0 | 77.2 | 79.1 |
| SLG-NER[26] | Bb | 86.5 | 84.5 | 85.4 | 83.3 | 86.6 | 84.9 | 81.0 | 77.2 | 79.1 |
| GPT-NER[87] | LLM | 73.3 | 75.1 | 74.2 | 72.8 | 75.5 | 73.6 | 61.9 | 67.0 | 64.4 |
| DiffusionNER[3] | Bl/BioB | 88.1 | 88.6 | 88.4 | 86.1 | 87.7 | 86.9 | 82.1 | 80.9 | 81.5 |
| BINDER[1] | Bl/BioB | 89.7 | 89.7 | 89.7 | 89.6 | 90.5 | 90.0 | 83.4 | 78.3 | 80.8 |
| Padellm-NER[88] | LLM | - | - | - | - | - | 85.0 | - | - | 80.8 |
| SGT[30] | Rb/BioB | 89.1 | 88.9 | 89.0 | 89.1 | 89.3 | 89.2 | **84.3** | **82.5** | **83.4** |
| SEE(ours) | Bb/Pub | 90.2 | 88.7 | 89.4 | 90.4 | 90.0 | 90.2 | 81.8 | 80.5 | 81.2 |
| SEE(ours) | Rb | **91.0** | **90.3** | **90.6** | **91.4** | **90.9** | **91.1** | - | - | - |

法，相较于次优基线 BINDER[1]的 89.7% 提升了 0.9%。同样，在 ACE2005 数据集结果中，基于 RoBERTa-base 的 SEE 的 F1 分数达到了 91.1%，同样取得了最佳结果，领先次优基线高达 0.9%。

在生物医学领域的嵌套数据集 GENIA 上的表现如表 3.10所示。SEE 模型取得了 81.2% 的 F1 分数。虽然这一成绩略低于当前该数据集上的最优模型 SGT[30]的 83.4%，这一差距可能源 SGT 能够联合捕捉重复跨度的信息，因此在 GENIA 这种重复跨度较多的数据集表现更加优异。但 SEE 的性能仍然显著优于众多其他先进基线模型，这表明 SEE 模型具有良好的领域适应性。

在 CoNLL03 和 OntoNotes 5.0 这两个经典的非嵌套命名实体识别数据集上，SEE 模型同样展示了非常出色的性能。根据表 3.11 和表 3.12，使用 RoBERTa-base 编码器的 SEE 模型在 CoNLL03 上达到了 93.5% 的 F1 分数，在 OntoNotes 上达到了 91.3% 的 F1 分数。这两个结果均与其他基线方法的水平相当或更优。考虑到基于对比学习的跨度方法在处理非嵌套命名实体识别任务时需要进行额外的去重后处理操作，这可能会对性能产生一定影响，但是 SEE 模型仍然取得了优秀的效果，进一步凸显了本文方法的准确性和鲁棒性。

综合各数据集的对比结果，SEE 模型相较于其他基于跨度的命名实体识别模型（包括基于对比学习的命名实体识别模型），展现出稳定且显著的性能优势。这种跨数据集的一致性提升，有力地验证了本文所提出的 ECNN 跨度增强模块以及改进的对

**表** 3.11　不同方法在 CONLL03 上的定量结果对比

| 模型 | 编码器 | P(%) | R(%) | F1(%) |
|---|---|---|---|---|
| Biaffine | Bl | **93.7** | 93.3 | **93.5** |
| GPT-NER | LLM | 89.7 | 92.0 | 90.9 |
| BINDER | Rb | 93.0 | 93.5 | 93.3 |
| DiffusionNER | Bl | 92.1 | 93.7 | 92.9 |
| Padellm-NER | LLM | - | - | 92.5 |
| SEE (ours) | Rb | 93.2 | **93.8** | **93.5** |

**表** 3.12　不同方法在 OntoNotes 上的定量结果对比

| 模型 | 编码器 | P(%) | R(%) | F1(%) |
|---|---|---|---|---|
| Biaffine | Bl | **91.1** | 91.5 | **91.3** |
| BART-NER | LLM | 79.8 | 84.5 | 82.2 |
| DiffusionNER | Bl | 90.3 | 91.0 | 90.6 |
| Padellm-NER | LLM | 89.5 | 90.3 | 89.9 |
| SEE (ours) | Rb | 90.7 | **91.9** | **91.3** |

比学习优化目标在提升实体表示质量的有效性与实用价值。此外，尽管部分现有方法依赖更强大的预训练编码器来追求性能，但 SEE 模型仅需使用相对较小的 BERT-base 和 RoBERTa-base 编码器，便能在多个数据集上取得与其相近或更好的结果。例如，在 ACE2005 数据集上，SEE 使用 BERT-base 编码器所取得的效果超越了许多使用更大模型的基线，这充分体现了 SEE 模型在有限计算资源下的实用价值。同时，从表中基于 LLM 的命名实体识别方法的结果可以看出，虽然该类方法展现出巨大潜力，并且在 CoNLL03、OntoNotes 等数据集上取得了可观的 F1 分数，但与现有基于预训练模型的最优方法相比，其性能表现仍略逊一筹，且往往伴随高昂的计算和部署成本。相比之下，SEE 模型以远低于 LLM 的资源消耗，实现了相当甚至更优的性能，这使得 SEE 对于需要平衡性能与效率的实际应用场景，具有更强的吸引力和广泛的适用性。

### 3.2.6　效率分析

在评估模型的实用价值时，运行效率与预测性能同等重要，它们是衡量模型综合能力的关键依据。因此，本节对所提出的 SEE 模型的计算效率进行分析，并着重考察其推理速度。为确保评估的公平性，所有效率测试均在一张 RTX 4090 上进行。

实验尽可能采用开源实现和公开可用的预训练权重，并设置相同的批量大小和最大序列长度。对比结果汇总于表 3.13。

表中呈现了 SEE 模型不同配置（包括 ECNN 层数和类别嵌入表示方式）与几种代表性命名实体识别基线模型的推理速度对比。从中可以看出，相较于将命名实体识别任务建模为复杂阅读理解问题的 MRC-NER[23] 或依赖双仿射解码器的 Biaffine-NER[24] 等模型，本研究所采用的基于对比学习的框架展现出显著的推理速度优势：SEE（depth=4）的推理速度分别达到 MRC-NER 和 Biaffine-NER 的 12.9 倍和 1.2 倍以上。与结构相对精简的 BERT-CRF 或同样基于对比学习的 BINDER 模型[1] 相比，集成 ECNN 模块的 SEE 模型推理速度有所降低，SEE（depth=4）的速度约为 BINDER 的 74%。额外卷积层带来的计算开销在预期范围内，如前文所述，这种轻微的速度牺牲，在嵌套实体识别等复杂任务上带来了显著的性能提升，充分体现了性能与效率间的有效权衡。

对 SEE 模型内部不同配置的效率分析显示，类别嵌入方式对推理速度影响显著。如表 3.13 所示，采用作为默认配置的可学习嵌入层的推理速度是编码注释指南方式的 1.5 倍。结合表 3.9 的消融实验进一步表明，采用可学习嵌入层在不显著影响性能的前提下能大幅减少类别端参数量。这些分析共同证明了文本端预训练模型结合类别端可学习嵌入层是实现高性能与高效率的有效策略。此外，ECNN 模块的层数对推理速度的影响同样重要：随着卷积层深度从 depth=3 增加到 depth=4，推理速度相应下降。因此，在实际应用中，可根据性能需求和硬件限制灵活选择 ECNN 层数，以在精度与速度之间取得最佳平衡。在主要对比实验中，本章最终采用了性能最优的 4 层 ECNN 配置。

**表** 3.13　不同模型的推理效率对比

| 模型 | 处理速度 (Tokens/s) | 加速比 |
|---|---|---|
| MRC-NER[23] | 1686 | 1.0× |
| Biaffine-NER[24] | 17710 | 10.5× |
| BERT-CRF | 25964 | 15.4× |
| BINDER[1] | 29349 | 17.4× |
| SEE[depth=3+ 注释指南] | 17028 | 10.1× |
| SEE[depth=3] | 25627 | 15.2× |
| SEE[depth=4] | 21749 | 12.9× |

## 3.3 本章小结

本章聚焦于提升基于对比学习的命名实体识别的特征表示能力，并为此提出了SEE 模型。该模型通过引入 ECNN 来补充实体跨度内部的语义信息，并设计了基于跨度和类别的双重对比学习优化目标，从而有效解决了现有方法在跨度表示不足和优化维度单一这两方面的局限性。详尽的实验结果表明，SEE 模型在多个命名实体识别数据集上，特别是在嵌套命名实体识别任务和中长跨度实体识别上，取得了优于基线方法的性能，并展现出良好的效率和鲁棒性。本章的贡献成功验证了 SEE 模型在处理各类命名实体识别任务中的有效性和高效性，为其在后续章节进行知识迁移至关系抽取任务奠定了坚实基础。

# 第四章　基于知识迁移的关系抽取方法研究

现有关系抽取方法大多聚焦于如何高效利用命名实体识别结果。然而，这些方法未能充分挖掘和利用命名实体识别模型在训练过程中所学到的、蕴含在模型参数中的深层语义信息。这些未被利用的知识对于理解上下文以及实体类别信息至关重要，是进一步提升关系抽取性能的关键所在。此外，命名实体识别和关系抽取作为两个异构任务，这也给知识迁移带来了困难。在第三章中，本文提出了一种基于对比学习的特征语义增强双编码器模型，该模型有效地学习了高质量的文本和实体类别信息。基于上一章的结果，本章提出一个基于知识迁移的关系抽取模型（knowledge Transfer for Relation Extraction，TransRE），通过设计有效的知识迁移策略，将命名实体识别阶段所学习的深层知识，有效地迁移到关系抽取模型中，从而显著提升关系抽取性能。

## 4.1　方法概述

本章聚焦于解决信息抽取领域中，命名实体识别与关系抽取两个任务阶段间交互不足、知识共享有限的核心问题。为此，本章提出了 TransRE 模型，该模型旨在利用上一章节在同一数据集经过对比学习预训练的双编码器桥接命名实体识别任务和关系抽取任务两个子任务，解决任务异构性带来的知识迁移困难的问题。模型整体流程如图 4.1 所示。首先，采用面向主语的打包策略对输入序列进行预处理，为文本序列中的主语和宾语分别插入不同的标记，以高效处理多实体对关系。随后，利用迁移的预训练文本编码器对打包后的序列进行编码，并进一步获取带有上下文信息的实体跨度表示。接着，实体类别感知融合模块（Entity Type-Aware Fusion module, ETAF）利用迁移的类别编码器所提供的类别嵌入以及宾语的类别预测概率，对宾语实体表示进行特征增强，将实体类别先验知识注入到关系分类的决策过程中，强化了模型对关系和实体类别间强相关性的建模。最后，模型通过综合损失函数进行优化，该函数结合了主关系抽取损失和辅助实体类别预测损失。详细的模型架构和原理将在后续小节中进行阐述。

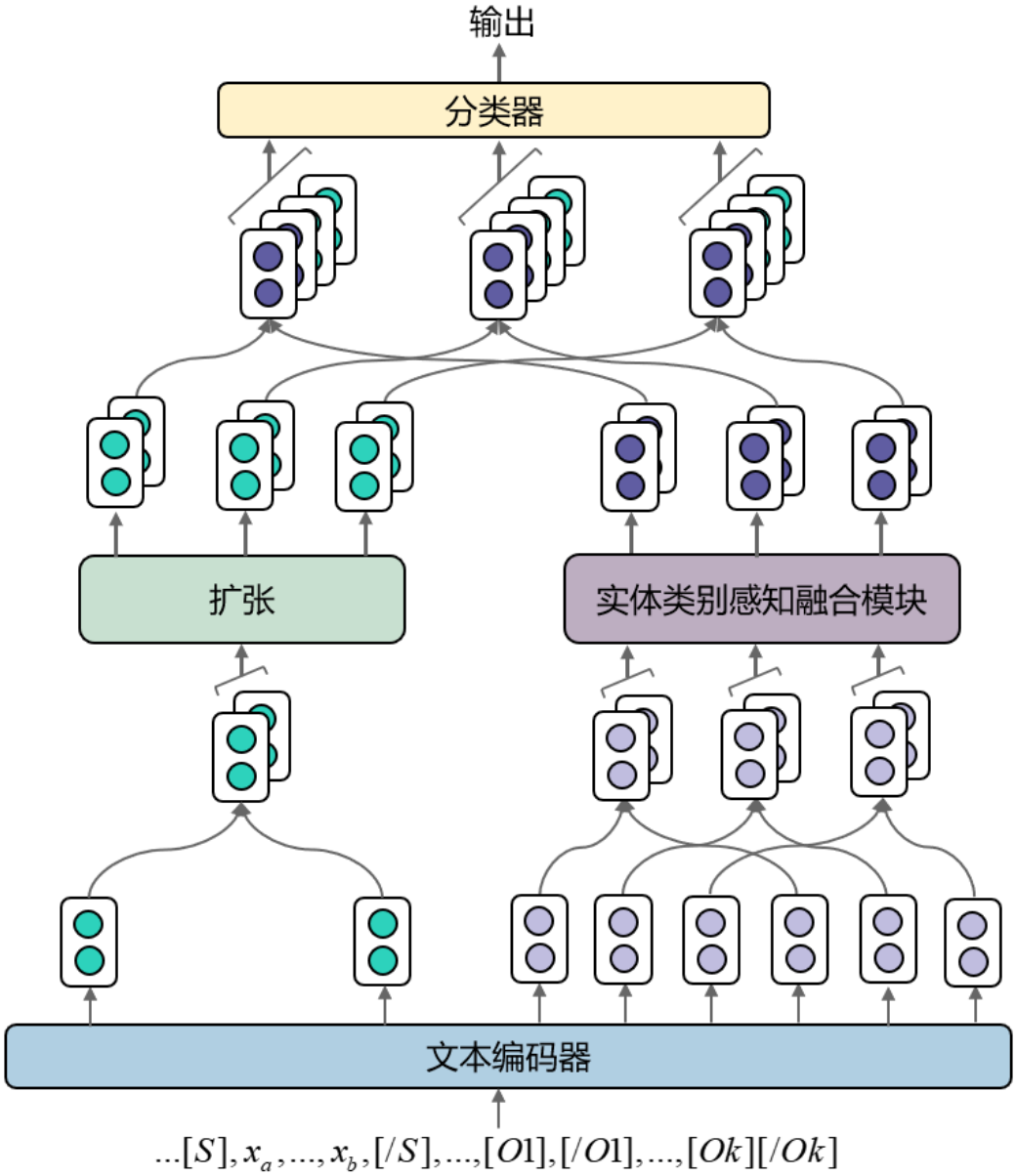



图 4.1　TransRE 整体框架

### 4.1.1　数据预处理

在关系抽取领域，基于标记的方法是一类重要且常用的技术。这类方法首先利用命名实体识别模型识别出的实体信息，然后在其周围插入特殊的标记，来显式地标识出实体的边界与跨度。这种处理方式能够更直接地向关系抽取模型提供实体位置信息，从而帮助模型更有效地关注并捕捉文本中实体间的潜在关系。

PL-Marker[46]是基于标记方法的经典代表之一，其在关系抽取阶段引入了面向主语的打包方法。该方法旨在通过更高效的数据组织方式，在一定程度上提升推理

速度，并进一步优化模型对实体关系的判断精度。对于句中每一个被识别出的实体，先将其视为潜在的主语，然后将该主语与句中其余所有实体构建成一系列候选的“主语-宾语”实体对，并进行统一处理。图 4.2 展示了经过面向主语的打包预处理后的序列示例。

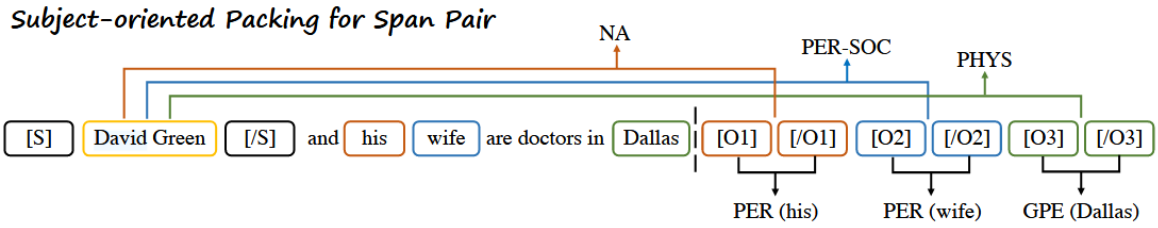



图 4.2　面向主语打包后序列示例[46]

**算法 4.1:** 面向主语的打包方法

**输入:** 文本序列 $X$；预测实体集合 Entities $= \{(start, end, label)\}$

**输出:** 打包后序列集合 $\mathcal{D}_{\text{pack}}$

$\mathcal{D}_{\text{pack}} \leftarrow []$; // 空列表，用于收集所有打包序列

**foreach** $sub \in$ Entities **do**

  $\hat{X} \leftarrow X$; // 重置为原始文本

  $objs \leftarrow$ Entities $\setminus \{sub\}$; // 定义宾语实体集合

  $\hat{X} \leftarrow Insert(\hat{X}, sub, start, \text{“}[S]\text{”})$; // 插入主语左标记

  $\hat{X} \leftarrow Insert(\hat{X}, sub, end + 1, \text{“}[/S]\text{”})$; // 插入主语右标记

  **for** $obj \in objs$ **do**

    $\hat{X} \leftarrow Insert(\hat{X}, len(\hat{X}), \text{“}[O]\text{”})$; // 在末尾追加宾语左标记

  **end**

  **for** $obj \in objs$ **do**

    $\hat{X} \leftarrow Insert(\hat{X}, len(\hat{X}), \text{“}[/O]\text{”})$; // 在末尾追加宾语右标记

  **end**

  $\mathcal{D}_{\text{pack}} \leftarrow \mathcal{D}_{\text{pack}} \cup \{\hat{X}\}$; // 加入结果集合

**end**

**return** $\mathcal{D}_{\text{pack}}$;

具体而言，对于给定的文本序列，首先从预测的实体集合中依次选取一个实体作为主语，并在其起始和结束位置前后插入一对实心标记（[S]，[/S]），以显式界定主语边界。随后，对于除当前主语外的所有其他实体都视为潜在宾语。区别于主语的处理方式，为了保持原有上下文结构不被干扰，宾语标记并未在原序列中进行局部插入，而是在打包后的序列 $\hat{X}$ 的末尾，一次性追加与潜在宾语实体数量等同的若干对悬浮标记（[O]，[/O]）。两种标记的详细说明可参考第 2.6 章，面向主语的打包方法详细流程如算法 4.1 所示。该方法使得模型能够在一次前向传播中并行计算该主语与所有潜在宾语之间的关系，从而显著提升了关系推理的并行度与整体计算效率。

### 4.1.2 基于迁移文本编码器的实体特征提取

给定一个输入序列 $X = [x_1, x_2, \ldots, x_N]$，对于句子中一个由起始位置 a 和结束位置 b 确定的主语实体 $s(a, b)$ 以及其包含 $k$ 个候选宾语集合 $\tau = \{(c1, d1), (c2, d2), \ldots, (ck, dk)\}$，先对其进行面向主语打包的预处理，本章用 $\hat{X}$ 表示预处理后的插入标记的序列 $\hat{X} = [\ldots [\text{S}], x_a, \ldots, x_b, [/\text{S}] \ldots, x_{c1} \cup [\text{O1}], \ldots, x_{d1} \cup [/\text{O1}], \ldots, x_{c2} \cup [\text{O2}], \ldots, x_{d2} \cup [/\text{O2}], \ldots]$ 这里用 $\cup$ 符号表示悬浮标记属于该词元，并与其共享位置编码。

在模型的编码阶段，编码器 BERT 是由第三章节中通过对比表示学习预训练得到的（与公式3.2相同）。鉴于该编码器在命名实体识别阶段已经学习到丰富的、与当前数据集相关的实体跨度特征。它能够更有效地对本章预处理后的序列 $\hat{X}$ 进行编码，为后续的关系分类任务提供高质量的特征表示:

$$H = \text{BERT}(\hat{X}) \tag{4.1}$$

其中 $H \in \mathbb{R}^{(N+2k)\times d}$ 代表编码后的词元序列特征表示，$h_i$ 是序列 $H$ 中第 $i$ 个词元的特征表示。

不同于命名实体识别阶段通过拼接开头和结尾词元来表示跨度，基于标记的关系抽取模型通过实体跨度的所对应一对标记来表示其特征，对于由起始位置 $a$ 和结束位置 $b$ 确定的主语跨度 $s = (a, b)$ 和宾语实体跨度集合 $O$ 中的一个由起始位置 $c$ 和

结束位置 $d$ 确定的宾语跨度 $o=(c,d)$，以如下形式进行表示：

$$s(a,b)=h_{a-1}\oplus h_{b+1} \tag{4.2}$$

$$o(c,d)=h_c^{(s)}\oplus h_d^{(e)} \tag{4.3}$$

其中 $\oplus$ 是拼接操作；$s(a,b),o(c,d)\in\mathbb{R}^{2d}$ 分别代表主语实体跨度和宾语实体跨度的特征表示。由于标记主语实体的是一对实心标记，因此该标记的位置在主语实体的两侧，故该实体跨度的语义由 $h_{a-1}$ 和 $h_{b+1}$ 共同表示；而标记宾语实体的是结尾的悬浮标记，因此该实体跨度的语义由结尾处 $h_c^{(s)}$ 和 $h_d^{(e)}$ 共同表示。

### 4.1.3 实体类别感知融合模块

关系抽取模型的目标是准确预测实体对之间存在的特定关系。然而，为了更好地捕捉实体类别与关系之间的复杂依赖关系（例如，某些关系仅适用于特定类别的实体），关系抽取模型通常会引入一个辅助损失函数，用于预测实体的类别。通过对实体类别进行显式预测并施加监督，能够增强模型对构成关系三元组的实体类别信息的感知能力，从而间接辅助主任务的关系预测。模型基于上一小节获得的宾语实体表示集合 $O$ 计算交叉熵损失：

$$y=\text{softmax}(\text{Linear}_{object}(O)) \tag{4.4}$$

$$\mathcal{L}_e=-\sum_{i=1}^{k}\log P_e(y_i^*|y_i) \tag{4.5}$$

其中 $\text{Linear}_{object}\in\mathbb{R}^{2d\times m}$ 代表将宾语实体映射成实体类别得分的线性层；$y_i^*$，$y_i$ 分别是宾语实体跨度 $o_i$ 的真实标签和预测概率分布；交叉熵损失函数有良好的数学特性，如可导性、凸性等，非常便于使用梯度下降法寻找最优，本章的损失函数都使用交叉熵衡量预测结果与真实标签的差异。

尽管此前的方法已尝试在损失函数层面建立实体类别与实体关系之间的关联，但这种关联主要体现在优化过程中，并未将模型对实体类别的预测结果显式地注入到用于最终关系判断的实体表示中。本章在此基础上，突破了仅在损失层面建立关联的范式，提出了一个实体类别感知融合模块 ETAF（如图 4.3 所示），用于进一步增强关系抽取过程中的类别感知能力。

具体而言，首先引入一个包含 $m$ 种实体类别的嵌入层 $\mathbf{e}^{type}$。为更好地捕捉和利

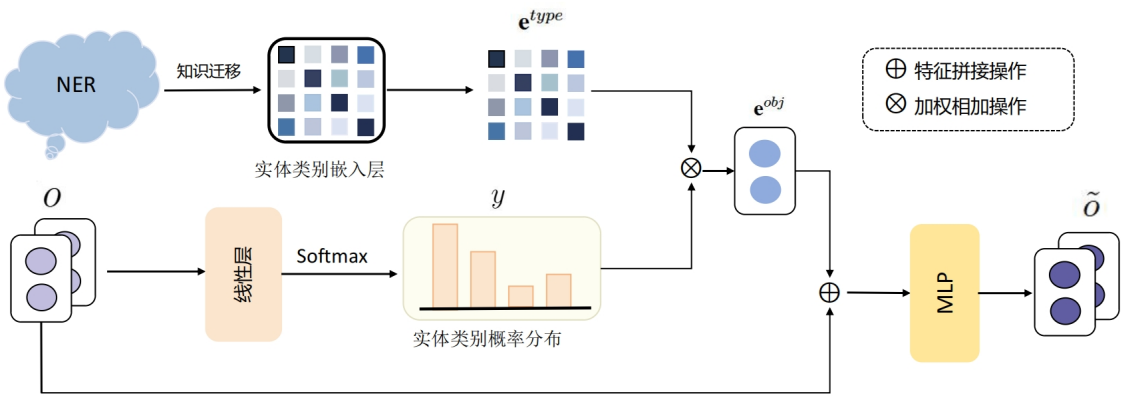



**图** 4.3　ETAF 模块架构图

用实体类别的语义信息，该嵌入层采用上一章节 3.1 中的实体类别编码器进行初始化。通过这种方式，嵌入层得以继承命名实体识别阶段所捕获的语义知识。这使得实体类别嵌入在特征空间中的分布能够更准确地反映其在自然语言中的语义关联。对于一个实体跨度 $o_i$，模型根据其预测的所有类别概率与相对应类别嵌入加权求和以获得该实体类别特征 $\mathbf{e}_i^{obj}$:

$$\mathbf{e}_i^{obj} = \sum_{n=1}^{m} y_i^n \cdot \mathbf{e}_n^{type} \tag{4.6}$$

其中 $\mathbf{e}_n^{type} \in \mathbb{R}^d$ 代表第 $n$ 个实体类别的嵌入向量；$y_i^n$ 是预测的宾语实体 $i$ 属于第 $n$ 个实体类别的概率；$\mathbf{e}_i^{obj} \in \mathbb{R}^d$ 代表第 $i$ 个宾语实体所对应的实体类别特征。

随后，为了将提取到的实体类别信息用于指导关系判断，模型将实体类别特征向量 $\mathbf{e}^{obj}$ 与原始上下文相关的宾语实体表示 $O$ 进行拼接，并通过多层感知机进行融合:

$$\tilde{O} = \text{MLP}(O \oplus \mathbf{e}^{obj}) \tag{4.7}$$

其中 $\tilde{O} \in \mathbb{R}^{k \times 2d}$ 代表增强后的宾语实体特征集合；$\oplus$ 代表向量拼接操作; $\text{MLP} \in \mathbb{R}^{3d \times 2d}$ 代表多层感知机。

融合了两类信息的增强宾语实体表示 $\tilde{O}$ 将作为后续关系分类模块的关键输入。利用这种同时包含上下文和类别线索的增强表示进行关系判断，能够显著提高模型对类别敏感关系的识别能力，实现更精准的关系抽取。

### 4.1.4 损失函数

为训练模型能够准确预测句子中实体对之间的语义关系，本章采用了一个综合损失函数对模型进行优化。该损失函数由两部分构成：主关系抽取损失和辅助实体类别预测损失。实体类别的损失函数已经在公式 4.5 中进行介绍。关系抽取损失是根据模型对实体间关系的预测计算得出的，此预测过程结合了主语实体表示与由 ETAF 模块增强的宾语实体表示：

$$\mathcal{L}_r = -\sum_{s_i \in S, \tilde{o}_j \in \tilde{O}} \log P_r(y^*_{i,j} | s_i, \tilde{o}_j) \tag{4.8}$$

其中 $y^*_{i,j}$ 是第 $i$ 个实体跨度和第 $j$ 个实体跨度之间的真实关系类别标签，并且 $i$ 和 $j$ 不能相同，这样可以确保不考虑同一实体自身之间的关系预测；$S$ 代表句子中得到的原始实体跨度特征集合；$\tilde{O}$ 代表增强过的实体跨度特征集合。

最终模型结合公式 4.5 和 4.8 共同进行优化:

$$\mathcal{L} = \lambda \mathcal{L}_r + (1 - \lambda)\mathcal{L}_e \tag{4.9}$$

其中 $\lambda$ 代表平衡两个损失的权重。

## 4.2 实验分析

### 4.2.1 数据集介绍

为全面评估本章提出的基于知识迁移的关系抽取模型 TransRE 的性能，本章在三个广泛使用的基准数据集上进行了系统的实验。这三个数据集分别为 ACE2004（简称 ACE04-R）、ACE2005（简称 ACE05-R）和 SciERC，其内容覆盖了通用新闻领域和特定的科学文献领域。

**（1）ACE05-R 数据集**

该数据集源自 LDC 发布的多语言训练语料库的英文部分，其详细来源、文本类别、领域介绍及数据划分已在 3.2.1 节中阐述。本章重点关注其关系抽取任务，该任务共包含六种预定义的关系类别，即 PHYS、PART-WHOLE、PER-SOC、ORG-AFF、ART 及 GEN-AFF，各类别具体含义详见表 4.1。

**（2）ACE04-R 数据集**

**表** 4.1　ACE05-R 数据集主要关系类别及描述

| 序号 | 关系类别 | 描述 |
|---|---|---|
| 1 | PHYS | 物理位置关系（如位于） |
| 2 | PART-WHOLE | 部分-整体关系 (实体是另一个实体的组成部分) |
| 3 | PER-SOC | 个人-社会关系 (家庭或职业等人际关系) |
| 4 | ORG-AFF | 组织-隶属关系 (个体与组织的隶属关系) |
| 5 | ART | 智能体-工具关系 (施事关系) |
| 6 | GEN-AFF | 一般-附属关系 (通用附属关系) |

该数据集同样源自 LDC 发布的多语言训练语料库的英文部分，其详细来源与数据划分等信息已在 3.2.1 节中介绍。本章重点关注其关系抽取任务，该任务共包含六种关系类别，即 PER-SOC、OTHER-AFF、ART、GPE-AFF、EMP-ORG 及 PHYS。各类别具体定义详见表 4.2。

**表** 4.2　ACE04-R 数据集关系类别及描述

| 序号 | 关系类别 | 描述 |
|---|---|---|
| 1 | PER-SOC | 个人-社会关系（家庭或职业等人际关系） |
| 2 | OTHER-AFF | 其他隶属关系 |
| 3 | ART | 智能体-工具关系（施事关系） |
| 4 | GPE-AFF | 地缘政治实体-隶属（如公民） |
| 5 | EMP-ORG | 雇佣/成员/组织关系 |
| 6 | PHYS | 物理位置关系（如位于） |

**（3）SciERC 数据集**

SciERC 数据集[81] 是一个面向科学信息抽取任务的语料库。该数据集包含了来自 12 个 AI 相关会议/研讨会的 500 篇科学论文摘要，旨在促进对科学文献中实体、关系和共指进行联合识别的研究。

SciERC 标注了 6 种面向科学领域的实体类别，涵盖了研究中的核心要素，包括 Task、Method、Metric、Material、OtherST 和 Generic，各类别具体定义详见表 4.3。同时，该数据集定义了 7 种用于描述实体间语义联系的关系类别，包括 Used-for、Feature-of、Hyponym-of、Part-of、Evaluate-for、Compare 和 Conjunction。这些关系类别的具体定义详见表 4.4。

本章使用 SciERC 原始的标准数据划分进行训练、验证和测试。选用 SciERC 数据集是为了评估本章提出的模型在特定专业领域（科学文献）的性能，并考察其处理复杂实体关系结构的能力。

**表** 4.3　SciERC 数据集实体类别定义

| 序号 | 实体类别 | 描述 |
|---|---|---|
| 1 | Task | 应用、待解决问题、待构建系统 |
| 2 | Method | 使用的方法、模型、系统、工具、组件 |
| 3 | Metric | 衡量系统/方法质量的指标或度量 |
| 4 | Material | 数据、数据集、资源、语料库 |
| 5 | OtherST | 其他明确定义的科学术语 |
| 6 | Generic | 指代实体但不具体的通用词或代词 |

**表** 4.4　SciERC 数据集关系类别定义

| 序号 | 关系类别 | 描述 (B 指向 A 或对称) |
|---|---|---|
| 1 | Used-for | (B for A) B 被用于 A；B 为 A 建模；A 基于 B |
| 2 | Feature-of | (B of A) B 是 A 的一个特征、属性或方面 |
| 3 | Hyponym-of | (B of A) B 是 A 的一种类别 (下位词) |
| 4 | Part-of | (B of A) B 是 A 的一个组成部分 |
| 5 | Evaluate-for | (B for A) B (Metric) 被用来评估 A (Method/Task) |
| 6 | Compare | (对称) 比较实体 A 和 B |
| 7 | Conjunction | (对称) 实体 A 和 B 并列提及，功能或角色类似 |

### 4.2.2　评价指标

为了从不同维度全面评估本章提出的知识迁移关系抽取模型的性能，本章实验将采用以下三种评价标准，并主要汇报其 F1 分数：

实体评估（Entity Evaluation，Ent）：此标准旨在评估模型准确识别实体信息（包括边界和类别）的能力，这通常是关系抽取的基础。一个实体被视为正确识别，当且仅当模型预测的实体边界与实体类别两项均与标注数据完全一致。

关系评估（Relation Evaluation，Rel）：此标准关注模型抽取关系事实的核心能力。一个关系实例被视为正确识别，当且仅当模型预测的主语实体边界、宾语实体边界以及它们之间的关系类别这三项均与标注数据完全一致。此评估标准不强制要求实体的类别预测正确。

严格关系评估（Strict Relation Evaluation，Rel+）：这是对关系评估更为严格的标准。在关系评估（Rel）的基础上，此标准额外要求模型预测的主语实体类别和宾语实体类别也必须与标注数据完全一致。因此，一个关系实例只有在主/宾语实体的边界、主/宾语实体的类别以及关系类别这五项全部正确时，才被视为正确识别。

对于上述三种评价标准，均计算精确率和召回率，并主要汇报 F1 值。F1 值作为精确率和召回率的调和平均数（其基本计算方式已在 3.2.2 介绍），能够更全面地衡

量模型在各项评价标准下的综合性能。

### 4.2.3 实验设置

本章根据目标数据集的领域特性针对性地选择了不同的预训练语言模型作为文本编码器。具体而言，对于通用领域的 ACE04-R 和 ACE05-R 数据集，实验采用 Bert-base 作为编码器；而针对科学文献领域的 SciERC 数据集，考虑到其专业性，采用了经过该领域语料预训练的 Scibert 模型。本章提出的 TransRE 模型所使用的文本编码器和类别编码器模型均直接来自于上一章的实验结果。考虑到不同编码器的特性差异，实验超参数根据编码器进行了适配调整：

- 当编码器为 Bert-base 时，学习率设置为 $2\times10^{-5}$，批量大小为 8，最大实体对数目为 32，最大输入序列长度为 256。在此配置下，模型共训练 10 个 epoch，每经过 5000 个训练步长保存一次模型。
- 当编码器为 Scibert 时，学习率设置为 $2\times10^{-5}$，批量大小为 8，最大实体对数目为 16，最大输入序列长度为 256。在此配置下，模型共训练 10 个 epoch，每经过 2500 个训练步长保存一次模型。

在本章训练过程中，模型统一使用 AdamW 优化器进行参数更新。本章采用了标准的验证策略：每完成相应的训练步长，就在对应的验证集上以 Rel+ 分数作为指标进行性能评估，并保存表现最佳的模型状态。最终报告的测试集结果均来自于加载该最佳模型状态进行的评估。为确保实验结果的稳定性和可靠性，本章报告的所有结果均为在五个不同的随机种子下进行实验所得结果的平均值。

### 4.2.4 消融实验

为系统性地评估本文所提 TransRE 模型中各核心组件的实际贡献，并验证关键设计选择的有效性，本章开展了一系列详尽的消融实验。这些实验旨在通过剥离或改变模型的特定部分，来量化其各自对最终性能的影响。为保证对比的公平性，所有消融实验均在 ACE05-R 数据集上进行。

表 4.5 展示了以本文第三章提出的 SEE 模型预测结果作为输入时，两种知识迁移机制对关系抽取模型性能影响的消融研究结果。当模型仅引入文本知识迁移（即使用经过命名实体识别阶段预训练的编码器），模型在 Rel 和 Rel+ 上表现出明显的提升，分别增长了 0.4% 和 0.5% 。这清晰地表明从命名实体识别阶段学习到的高质

量上下文表示，能够有效增强下游模型对关系及其涉及实体边界的判断能力。当模型仅引入类别知识迁移（即 ETAF 模块）时，同样对模型产生了积极的影响。Rel 提升了 0.3%，并且 Rel+ 提升了 0.3%，这两项数值提升直接反映了提供的类别感知特征对于关系抽取任务的价值。最重要的是，当同时采用两种知识迁移策略时，模型在所有指标上均达到最佳性能，显著超过了基模型。其中在 Ent 上提升了 0.2%，与输入的 SEE 模型本身的实体识别性能更为接近，表明两种知识的结合不仅优化了关系判断，也帮助模型在关系抽取时更好地维持了对实体信息的准确理解；在 Rel 上也提升了 0.8%，在综合性最强的 Rel+ 指标上提升最为显著，达到了 1.4% 并且均超过了仅有一种知识迁移的效果。由此可见，它们之间存在积极的协同效应，共同提高了模型的整体识别性能，也验证了本章所提知识迁移框架的有效性。

**表** 4.5　知识迁移组件在 ACE05-R 数据集上的定量结果

| 类别迁移 | 文本迁移 | Ent (%) | Rel (%) | Rel+ (%) |
|---|---|---|---|---|
| | | 90.0 | 69.3 | 66.5 |
| | ✓ | 90.0 | 69.7 | 67.0 |
| ✓ | | 90.1 | 69.6 | 66.8 |
| ✓ | ✓ | 90.2 | 70.1 | 67.9 |

为了评估 TransRE 模型在处理不同质量命名实体识别结果时所展现的鲁棒性与泛化能力，进一步开展了数据敏感性分析。具体而言，在 ACE05-R 数据集上，本章分别使用三种不同命名实体识别模型产生的实体预测结果作为输入（输入的命名实体识别结果如表 4.6 所示，根据其性能表现，依次为 Binder NER、PL-Marker NER 及本文第三章提出的 SEE NER）。通过这一设置，本章对比了包含完整知识迁移的 TransRE 模型与未使用知识迁移的基模型之间的性能表现。

实验结果如图 4.4 所示。整体结果表明，在采用不同质量的命名实体识别结果作为输入时，TransRE 在各项评估指标上的性能均一致且稳定地优于基模型 PL-Marker。这表明命名实体识别阶段迁移的知识具有良好的泛化能力，对于不同来源和质量的命名实体识别输入都能带来普适性的性能提升效果。进一步对性能差距进行细致分析发现，TransRE 相对于基模型的优势随着输入命名实体识别质量的提高而呈现扩大趋势，尤其是在综合性最强的 Rel+ 指标上：当使用 Binder NER 作为输入时，TransRE 领先基模型 0.4%。当使用 PL-Marker NER 作为输入时，领先优势扩大到 0.6%。而

当使用性能最佳的 SEE NER 作为输入时，领先优势最为显著，达到了 1.4%。Rel 和 Ent 指标也表现出相似的规律。这表明，尽管更高质量的命名实体识别结果就能提升下游关系抽取模型的性能，但本章提出的知识迁移方法能够使得模型从这些高质量输入中获得更为明显的相对增益。这有力地证明了该知识迁移方法能与更准确、更丰富的上游实体信息产生强大的协同效应。

**表 4.6** 不同命名实体识别结果在 ACE2005 数据集上的性能对比

| 命名实体识别模型 | P(%) | R(%) | F1(%) |
|---|---|---|---|
| PL-Marker NER | 89.3 | 89.9 | 89.6 |
| BINDER NER | 88.9 | 90.1 | 89.5 |
| SEE NER | 90.1 | 90.2 | 90.2 |

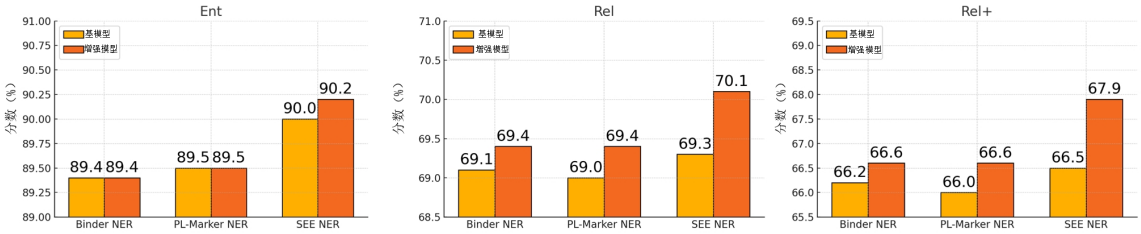



**图 4.4** 数据敏感性分析结果

随后本章探究将迁移过来的类别嵌入信息有效融入关系抽取模型的最佳方式，对不同的信息融合策略进行了比较实验。在保持其他组件（如迁移的 BERTT 编码器）不变的条件下，对比了使用多层感知机、残差连接以及门控机制这三种常见的融合方法对模型性能的影响。实验使用 SEE 模型预测结果作为输入，结果如表 4.7 所示，采用多层感知机进行融合的方式在所有三个评估指标上均取得了最佳性能：Ent 达到 90.20%，Rel 达到了 70.1%，最关键的 Rel+ 指标也取得了最高值，高达 67.9%。该结果表明，通过一个可学习的变换来结合实体表示和注入的类别信息，能够最有效地利用这两种特征，从而增强关系抽取过程中的类别感知能力。相比之下，门控机制和残差连接的表现在 Rel 上相同且略低于多层感知机。在 Rel+ 指标上，门控机制高于残差连接 0.2%，这可能意味着动态地调整类别信息流比简单的特征相加对于需要精确类别判断的严格关系评估更有利。在 Ent 指标上，三种方法的差异也相对较小，但多层感知机仍然领先。

经过分析，多层感知机在此任务中表现最佳，可能源于其更强的特征融合与适

配能力。基于预测概率加权的类别嵌入的类别感知特征本身提供了高质量的类别先验，而多层感知机作为一个参数化的非线性映射模块，能够更灵活地学习如何将这种类别特征与原有的实体上下文表示进行最优的交互与整合，以更好地服务于复杂的关系分类决策。相对而言，残差连接或相对门控机制在特征融合方面会损失一部分信息，未能像多层感知机那样充分发掘和利用类别引导信息的全部潜力。

**表** 4.7　不同类别信息融合方式在 ACE05-R 数据集上的定量结果

| 融合方式 | Ent (%) | Rel (%) | Rel+ (%) |
|---|---|---|---|
| 残差连接 | 90.1 | 69.7 | 66.8 |
| 门控机制 | 90.1 | 69.7 | 67.0 |
| 多层感知机 | 90.2 | 70.1 | 67.9 |

此外，表 4.8 展示了在 ACE05-R 数据集上，不同损失参数 $\lambda$ 设置对模型性能的影响。该参数用于平衡关系抽取损失与辅助实体类别预测损失之间的权重。从表中结果可以看出，模型在各项指标上的性能波动相对较小，这表明本章提出的 TransRE 模型在一定程度上对损失函数的权重分配不高度敏感，展现了较好的鲁棒性。最终，模型在 $\lambda = 0.3$ 时在 Rel 和 Rel+ 指标上同时达到了最优性能，分别达到了 70.1% 和 67.9%，并且 Ent 指标也保持在较高水平。综合考量各项评估指标的表现，最终模型选择将损失参数 $\lambda$ 设置为 0.3。

**表** 4.8　不同损失参数 $\lambda$ 设置在 ACE05-R 数据集上的定量结果

| 损失参数 $\lambda$ 取值 | Ent(%) | Rel(%) | Rel+(%) |
|---|---|---|---|
| 0.1 | 90.2 | 69.9 | 67.7 |
| 0.2 | 90.2 | 69.8 | 67.5 |
| 0.3 | 90.2 | 70.1 | 67.9 |
| 0.4 | 90.1 | 70.0 | 67.9 |
| 0.5 | 90.1 | 69.9 | 67.9 |
| 0.6 | 90.1 | 69.9 | 67.7 |
| 0.7 | 90.1 | 69.8 | 67.5 |
| 0.8 | 90.1 | 69.8 | 67.7 |
| 0.9 | 90.1 | 70.0 | 67.9 |

在之前的消融实验中，已经初步证明在关系抽取模型中引入 ETAF 模块能够在关系预测过程提供实体类别先验知识，进而有效提升关系对类别敏感关系的预测。为

进一步探究在此模块中迁移知识的重要性，针对类别嵌入的不同初始化策略进行了专项研究。在该组实验中，保留了迁移的文本编码器设置并采用 SEE NER 的预测结果作为输入，比较了三种不同的类别信息处理方式：（1）无类别信息，不使用 ETAF 模块；（2）正态分布，ETAF 模块中类别嵌入层权重在关系抽取训练开始时通过正态分布随机初始化；（3）嵌入迁移，ETAF 模块中类别嵌入层加载第三章 SEE 模型训练得到的权重，结果如表 4.9 所示。

实验结果清晰地表明，不同的类别嵌入初始化处理方式对关系抽取模型性能具有显著影响。首先，将从命名实体识别阶段得到的类别编码器进行知识迁移的模型表现显著优于其他两种方式。在 Rel+ 指标上，迁移嵌入达到了 67.9%，相比于正态分布随机初始化的 66.9%，提升了 1%；在 Ent 和 Rel 指标上，同样分别优于随机初始化 0.2% 和 0.5%。其次，随机正态分布初始化的类别嵌入在 Rel+ 和 Rel 指标上甚至略低于不添加类别感知融合模块的模型，该模块的添加给模型带来了负面影响。

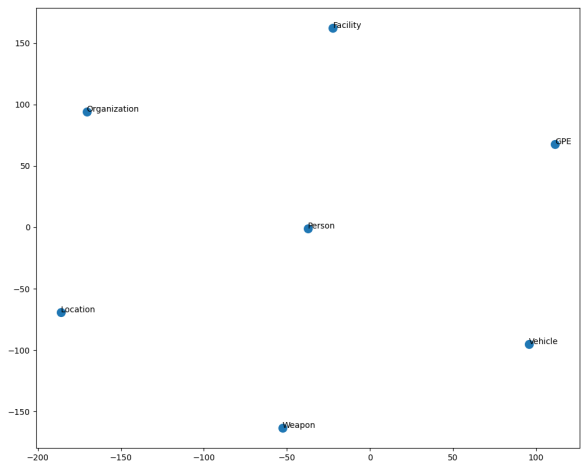
**表** 4.9　不同类别嵌入初始化方式在 ACE05-R 上的定量结果

| 类别嵌入 | Ent (%) | Rel (%) | Rel+ (%) |
|---|---|---|---|
| 无类别嵌入 | 90.0 | 69.8 | 67.0 |
| 正态分布随机初始化 | 90.0 | 69.6 | 66.9 |
| 类别编码器知识迁移 | 90.2 | 70.1 | 67.9 |

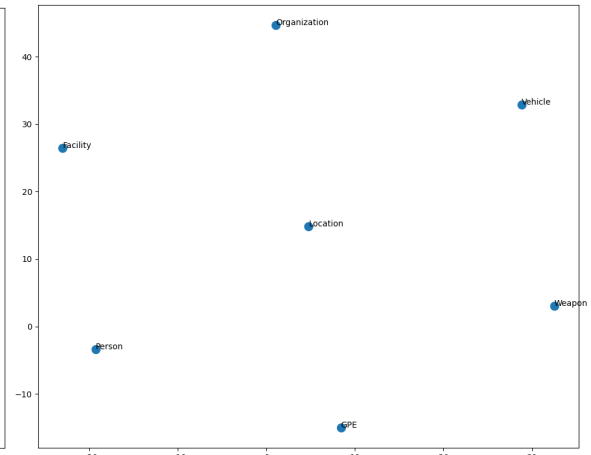
为了进一步分析原因，本章对两种初始化的类别嵌入特征进行了降维可视化。图 4.5 展示了利用 t-SNE 降维技术得到的 ACE05-R 数据集中 7 种实体类别嵌入向量的二维分布情况以及这些类别嵌入向量之间的余弦相似度矩阵热力图。

从特征分布可以清晰地看到，不论是正态分布随机初始化还是类别编码器知识迁移，不同的实体类别嵌入在二维空间中形成了相对独立且彼此分离的点簇。这确保了类别嵌入具有良好的整体可区分性，可以为 ETAF 模块提供更有判别性的实体类别特征。

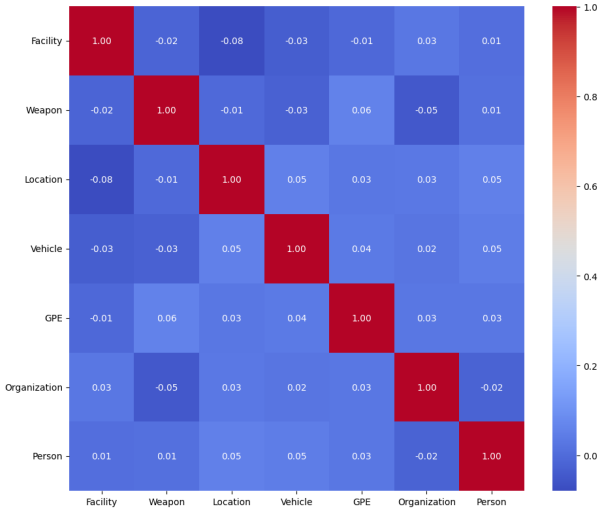
但是从余弦相似度矩阵热力图可以清晰地看出两者的区别，在正态分布随机初始化的类别嵌入余弦相似度矩阵中，代表不同实体类别间相似度的非对角线元素数值普遍较低且分布相对均匀，未能清晰显示出某些类别队之间有显著高于其他类别对的相似性模式。这与采用类别编码器知识迁移初始化的类别嵌入所对应的热力图存在明显差异。在后者的热力图中，虽然多数非对角线相似度值也较低，但可以观察



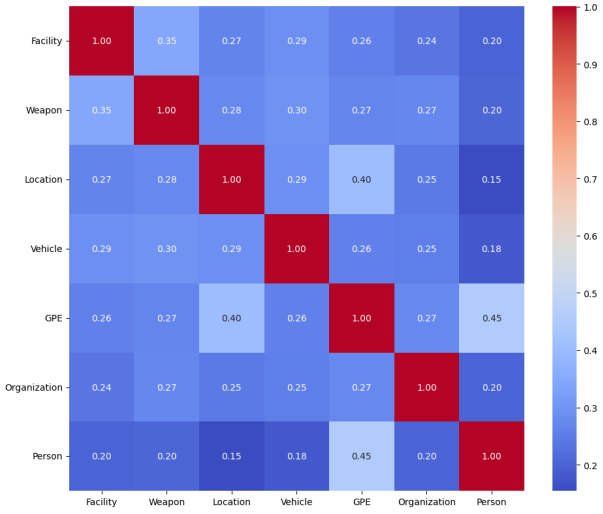

(a) 随机初始化的类别嵌入特征分布




(b) 知识迁移的类别嵌入特征分布




(c) 随机初始化的类别嵌入相似度矩阵




(d) 知识迁移的类别嵌入相似度矩阵

**图 4.5** ACE05-R 数据集中类别嵌入特征的二维分布及相似度矩阵可视化

到部分非对角线元素显示出相对较高的数值，这些高相似性区域对应于语义上相关的类别对，例如：Person (人) 与 GPE (地缘政治实体) 之间的相似度在所有异类配对中最高，这符合人与国家/地区概念紧密关联的常识；GPE 与 Location (地理位置) 之间也具有高相似度，体现了两者在地理空间概念上的联系；Weapon (武器) 与 Facility (设施) 的相似度也相对较高，这可能反映了它们同为人造物理实体的共性，或是它们在特定场景下的潜在关联。这些观察结果表明，知识迁移使得类别嵌入的相似性结构能够反映类别间的语义关系。

综合定量分析和可视化分析，结果有力地表明类别编码器知识迁移是 ETAF 模块不可或缺的一部分，通过对比学习训练得到的类别嵌入不仅实现了有效的类别区分，而且成功地将类别间的语义远近关系编码到了嵌入空间的结构中。这种具备区分度且蕴含语义结构的类别表征，相比随机初始化或仅能区分的嵌入，能够为下游的关系抽取任务提供更丰富的先验知识，从而有力地支持了知识迁移策略的有效性。

### 4.2.5 对比实验

为了全面评估本论文提出的基于知识迁移的关系抽取模型 TransRE 的效果，本节将该模型与一系列具有代表性的、涵盖不同技术范式的先进基线模型进行了全面的比较。这些基线模型包括：联合抽取模型 SPTree[89]、DYGIE[5]、Multi-turn QA[90]、OneIE[91]、DYGIE++[92]、TriMF[93]、PFN[42]、UniRE[94]、ATG[95]和 CEFF[96]；以及同样基于标记的流水线关系抽取模型 PURE[50]和 PL-Marker[46]。

本章选取了三个广泛使用的关系抽取基准数据集 ACE05-R、ACE04-R 和 SciERC 进行系统的对比实验。这些数据集覆盖了通用新闻领域和科学文献领域，是评估关系抽取模型性能的常用数据集。具体的实验结果与各基线模型的详细比较呈现在表 4.10 和 4.11 中。在该结果表格中，遵循了通用的性能标注约定：最优结果以**粗体标识**，次优结果以下划线标识。为了清晰展示各方法所依赖的基础模型，表格中也列出了所使用的编码器类别:LSTM 、ELMo 、BERT-large（Bl）、BERT-base（Bb）和 Scibert(Scib)。

在 ACE05-R 数据集上，TransRE 在最严格、也是最能反映模型综合能力的 Rel+ 指标上取得了 67.9% 的 F1 分数，超越了该表中所列的大多数基线模型，相较于次优的 CEFF 模型提升了 0.8%，并且显著优于同为基于标记的关系抽取模型 PURE 和 PL-Marker 等。在 Rel 指标上，TransRE 达到了 70.1% 的 F1 分数，与 CEFF 表现相

当，并优于 PL-Marker 。在 Ent 指标上，得益于上游高质量的命名实体识别输入和关系抽取阶段的知识协同，TransRE 也达到了 90.2% 的 F1 分数，处于顶尖水平。

在 ACE04-R 数据集结果中，TransRE 在 Rel+ 指标上以 63.7% 的 F1 分数取得了最佳结果，显著领先于次优的 PL-Marker 达 1.1%。在 Rel 指标上，TransRE 获得了 67.3% 的 F1 分数，优于 PL-Marker 和 PURE 。在 Ent 指标上，TransRE 获得了 89.5% 的 F1 分数，同样具有很强的竞争力。

**表** 4.10　不同模型在 ACE05-R 和 ACE04-R 数据集上的定量结果对比

| 方法 | 编码器 | ACE05 | | | ACE04 | | |
|---|---|---|---|---|---|---|---|
| | | Ent(%) | Rel(%) | Rel+(%) | Ent(%) | Rel(%) | Rel+(%) |
| SPTree[89] | LSTM | 83.4 | - | 55.6 | 81.8 | - | 48.4 |
| DYGIE[5] | ELMo | 88.4 | 63.2 | - | 87.4 | 59.7 | - |
| Multi-turn QA[90] | Bl | 84.8 | - | 60.2 | 83.6 | - | 49.4 |
| OneIE[91] | | 88.8 | 67.5 | - | - | - | - |
| DYGIE++[92] | Bb | 88.6 | 63.4 | - | - | - | - |
| TriMF[93] | | 87.6 | 66.5 | 62.8 | - | - | - |
| UniRE[94] | | 88.8 | - | 64.3 | 87.7 | - | 60.0 |
| PFN[42] | | 89.0 | - | 66.8 | 89.3 | - | 62.5 |
| PURE[45] | | 90.1 | 67.7 | 64.8 | 89.2 | 63.9 | 60.1 |
| PL-Marker[46] | | 89.8 | 69.0 | 66.5 | 88.8 | 66.7 | 62.6 |
| ATG[95] | | 90.1 | 68.7 | 66.2 | - | - | - |
| CEFF[96] | | **90.3** | **70.2** | 67.1 | - | - | - |
| TransRE(ours) | Bb | 90.2 | 70.1 | **67.9** | **89.5** | **67.3** | **63.7** |

在科学领域的 SciERC 数据集上，TransRE 同样展现了竞争力。在 Rel+ 指标上，TransRE 取得了 43.7% 的 F1 分数，这一结果微弱领先于 CEFF 的 43.5%，并显著超过 PL-Marker。在 Rel 指标上，TransRE 达到了 54.2% 的 F1 分数，与当前先进模型 CEFF 持平，并优于 PL-Marker。在 Ent 指标上，TransRE 也取得了 70.9% 的 F1 分数。在与通用领域有显著差异的科学文献数据集 SciERC 上取得具有竞争力的结果，有力地证明了 TransRE 在跨领域的良好泛化性，而非仅限于特定领域。

综合三个数据集的对比结果，TransRE 模型展现出有竞争力的整体性能，尤其在衡量关系抽取任务完整性能的 Rel+ 指标上，展现出稳定且显著的性能优势。这种跨数据集的性能一致性提升，有力地验证了本章方法的有效性。

**表** 4.11　不同模型在 SciERC 数据集上的定量结果

| 方法 | 编码器 | SciERC | | |
|---|---|---|---|---|
| | | Ent(%) | Rel(%) | Rel+(%) |
| Multi-turn QA[90] | Bl | 65.2 | 41.6 | - |
| TriMF[93] | Scib | 70.2 | 52.4 | - |
| UniRE[94] | Scib | 68.4 | - | 36.9 |
| PFN[42] | Scib | 66.8 | - | 38.4 |
| PURE[45] | Scib | 68.9 | 50.1 | 36.8 |
| PL-Marker[50] | Scib | 69.9 | 53.2 | 41.6 |
| ATG[95] | Scib | 69.7 | 51.1 | 38.6 |
| CEFF[96] | Scib | 70.6 | **54.2** | 43.5 |
| TransRE(ours) | Scib | **70.9** | **54.2** | **43.7** |

## 4.3　本章小结

本章聚焦于解决信息抽取中关系抽取方法对命名实体识别知识利用不足的核心问题，探索并提出了一种基于知识迁移的 TransRE 模型。TransRE 的核心在于将上一章基于对比学习优化的命名实体识别模型 SEE 所获得的深层知识有效迁移到关系抽取模型中，具体通过迁移预训练文本编码器和设计类别感知融合模块这两项关键知识迁移策略实现。在多个标准数据集上的系统性实验充分验证了 TransRE 模型的性能，结果显示其显著提升了关系抽取的性能，特别是在衡量完整性能的 Rel+ 指标上表现优异。这项研究证明了所提出知识迁移机制的有效性与优越性，实现了关系抽取对命名实体识别阶段更深层次知识的利用，为克服信息抽取子任务交互不足的局限性提供了新的有效途径。

# 第五章　总结与展望

## 5.1　总结

信息抽取作为从海量非结构化文本中获取结构化知识的关键技术，其核心子任务包括命名实体识别和关系抽取。然而，信息抽取流水线中存在命名实体识别与关系抽取交互不足的固有难题，导致关系抽取未能充分利用命名实体识别阶段的丰富语义信息。针对上述困境，并结合近年来对比学习在表示学习领域的成功应用，本论文旨在探索一条通过增强命名实体识别阶段学习到的知识并向关系抽取阶段迁移，来提升信息抽取整体性能的新路径。本论文的主要研究工作和创新性贡献体现在以下方面：

（1）提出了一种基于偶数卷积的跨度表示增强模块 ECNN，有效解决了传统跨度表示方法易忽略内部语义信息的局限性。该模块被集成到本文提出的对比学习命名实体识别模型 SEE 中，利用偶数卷积核感受野的特性，侧重于捕捉子跨度的信息，从而补充实体跨度内部被传统边界表示法所忽视的丰富语义。第三章通过详尽的消融实验和跨度长度性能分析，证明了 ECNN 模块能够显著增强实体跨度的表达能力，有效提升了命名实体识别模型的性能，特别是在处理长跨度实体方面效果显著。

（2）提出了一种结合实体跨度维度和实体类别维度的双重对比学习损失函数，通过在两个维度上进行对比优化，在特征空间拉进了实体跨度及相应实体类别的距离。同时，为了提升模型效率，本文采用轻量级的可学习嵌入层代替原有的预训练模型来表示实体类型，在保证模型性能的同时，大幅减少了参数量并提高了运算速度。第三章的实验充分验证了这些优化策略的有效性和高效性。

（3）将优化后的命名实体识别模型知识有效迁移至下游关系抽取模型，旨在实现信息抽取任务内部更深层次的知识共享。为此，本文提出了 TransRE 模型，其核心在于利用前述 SEE 模型学习到的高质量表征来增强下游关系抽取任务。TransRE 模型包含文本编码器和类别编码器迁移这两个关键机制。具体而言，通过复用命名实体识别模型文本编码器的权重，模型能够更有效地编码和感知实体上下文信息；同时，引入基于类别编码器迁移的类别感知融合模块，增强了模型在关系抽取时对实体类别的感知能力。第四章的系统实验有力地证明该知识迁移机制能够提升关系抽

取模型的性能。

本文研究为信息抽取任务提供了一种新的思路和优化方向。通过证明命名实体识别阶段学习到的深层表示可以被有效迁移并用于增强下游关系抽取任务，本文方法突破了信息抽取任务中关系抽取模型和命名实体识别模型仅在结果层面传递信息的局限性，实现了模型内部更深层次的知识共享与交互。因此，本论文提出的知识迁移策略不仅仅是一种性能提升的技术，更代表了对如何挖掘和利用上游任务知识弥合信息抽取各阶段隔阂的一种探索，为提升信息抽取系统的整体效能提供了有效的解决方案。

## 5.2 展望

尽管本论文的研究在提升信息抽取性能方面取得了一系列积极的成果，但仍清晰地认识到其中存在一些局限性：

(1) 尽管本研究取得了积极成果，但仍需认识到实验设计方面存在的局限性。首先，本文的验证主要在几个标准的、数据量充足的数据集上进行，未能充分探究所提方法在更具挑战性的场景下的性能边界，例如长尾类别、低资源数据集、类别不均衡等更具挑战性的实验场景。

(2) 知识迁移策略的适用范围有限：本文提出的知识迁移策略，无论是文本编码器的复用还是类别信息的注入，其设计和有效性验证主要是在基于特定标记的关系抽取模型上完成的。因此，本文提出的迁移方法对于其他类型的关系抽取架构，其普适性和具体效果尚不明确，需要进一步的实验研究和针对性的适配调整才能确定。

(3) 上下游任务学习范式与数据集特性的交互影响：本研究中作为知识源头的命名实体识别模型采用了对比学习训练范式，该范式依赖于数据中正负样本的有效构建来优化表示空间，其效果可能受到数据集自身特性的影响。而下游的关系抽取任务通常采用基于标注数据的全监督分类范式。这两种范式在学习目标和对数据分布的敏感度上存在固有差异。这种范式差距与数据集特性的交互可能导致观察到的现象：在某些特定数据集上，如果其特性（例如缺乏足够丰富或区分性的正负样本对）导致基于对比学习的命名实体识别模型表现不佳，那么在该数据集上学习到的特征可能也难以有效满足下游全监督关系抽取任务的需求，从而影响了知识迁移在该特定数据集上的最终效果。这种训练范式、数据集特性与迁移效果之间的复杂关系，给

知识迁移的稳定性和效果带来了一定的挑战。

针对上述研究中存在的局限性，并结合本研究的核心发现，未来可以围绕以下几个方面展开进一步的工作：

（1）扩展与适配知识迁移策略：将本文的知识迁移思路推广到更广泛的、不同结构的关系抽取模型中，检验其通用性并进行必要的适配优化。

（2）深入研究范式差异与数据集适应性：进一步探索对比学习与全监督学习范式差异对跨任务知识迁移的具体影响机制，并研究如何提升对比学习范式对不同数据集特性的鲁棒性，或者设计能更好桥接范式差异的迁移或微调技术。

（3）探索对比学习在关系抽取中的直接应用：将对比学习的思想直接应用于关系抽取任务本身。研究如何为关系抽取任务设计有效的正负样本对构建策略和对比学习目标，构建基于对比学习的关系抽取模型，并评估其性能。

（4）面向更复杂场景的应用：将验证有效的知识迁移技术应用于低资源、跨领域、多语言等更具挑战性的信息抽取场景，比如探索其在事件抽取、共指消解等相关自然语言处理任务中的应用潜力。

# 参考文献

[1] ZHANG S, CHENG H, GAO J, et al. Optimizing bi-encoder for named entity recognition via contrastive learning[C]//The Eleventh International Conference on Learning Representations. 2022.

[2] SHEN Y, WANG X, TAN Z, et al. Parallel instance query network for named entity recognition[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022: 947-961.

[3] SHEN Y, SONG K, TAN X, et al. DiffusionNER: Boundary diffusion for named entity recognition[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2023: 3875-3890.

[4] EBERTS M, ULGES A. Span-based joint entity and relation extraction with transformer pre-training[M]//ECAI 2020. IOS Press, 2020: 2006-2013.

[5] LUAN Y, WADDEN D, HE L, et al. A general framework for information extraction using dynamic span graphs[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019: 3036-3046.

[6] BHATTACHARJEE A, KARAMI M, LIU H. Text transformations in contrastive self-supervised learning: A review[C]//31st International Joint Conference on Artificial Intelligence, IJCAI 2022. International Joint Conferences on Artificial Intelligence, 2022: 5394-5401.

[7] KIM J H, WOODLAND P C. A rule-based named entity recognition system for speech input.[C]//INTERSPEECH. 2000: 528-531.

[8] ZHANG S, ELHADAD N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts[J]. Journal of biomedical informatics, 2013, 46(6): 1088-1098.

[9] 张华平, 刘群. 基于角色标注的中国人名自动识别研究[J]. 计算机学报, 2004 (01): 85-91.

[10] 俞鸿魁, 张华平, 刘群, 等. 基于层叠隐马尔可夫模型的中文命名实体识别[J]. 通信学报, 2006(02): 87-94.

[11] BIKEL D M, MILLER S, SCHWARTZ R, et al. Nymble: a high-performance learning name-finder[C]//Proceedings of the Fifth Conference on Applied Natural Language Processing. Association for Computational Linguistics, 1997: 194–201.

[12] BIKEL D M, SCHWARTZ R, WEISCHEDEL R M. An algorithm that learns what's in a name[J]. Machine learning, 1999, 34: 211-231.

[13] 冯元勇, 孙乐, 张大鲲, 等. 基于小规模尾字特征的中文命名实体识别研究[J]. 电子学报, 2008(09): 1833-1838.

[14] SZARVAS G, FARKAS R, KOCSOR A. A multilingual named entity recognition system using boosting and c4. 5 decision tree learning algorithms[C]//Discovery Science. Springer, 2006: 267.

[15] MCCALLUM A, LI W. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons[C]//Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003. 2003: 188-191.

[16] 邬伦, 刘磊, 李浩然, 等. 基于条件随机场的中文地名识别方法[J]. 武汉大学学报(信息科学版), 2017, 42(02): 150-156.

[17] HUANG Z, XU W, YU K. Bidirectional lstm-crf models for sequence tagging[A]. 2015. arXiv: 1508.01991.

[18] MA X, HOVY E. End-to-end sequence labeling via bi-directional lstm-cnns-crf[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 1064-1074.

[19] 罗凌, 杨志豪, 宋雅文, 等. 基于笔画 ELMo 和多任务学习的中文电子病历命名实体识别研究[J]. 计算机学报, 2020, 43(10): 1943-1957.

[20] YAN H, DENG B, LI X, et al. Tener: adapting transformer encoder for named entity recognition[A]. 2019. arXiv: 1911.04474.

[21] YAN H, GUI T, DAI J, et al. A unified generative framework for various ner subtasks [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 5808-5822.

[22] TAN Z, SHEN Y, ZHANG S, et al. A sequence-to-set network for nested named entity recognition[C]//Proceedings of the 30th International Joint Conference on Artificial Intelligence. 2021: 3936-3942.

[23] LI X, FENG J, MENG Y, et al. A unified mrc framework for named entity recognition [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 5849-5859.

[24] YU J, BOHNET B, POESIO M. Named entity recognition as dependency parsing [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 6470-6476.

[25] DOZAT T, MANNING C D. Deep biaffine attention for neural dependency parsing [C]//International Conference on Learning Representations. 2017.

[26] WAN J, RU D, ZHANG W, et al. Nested named entity recognition with span-level graphs[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 892-903.

[27] LI J, FEI H, LIU J, et al. Unified named entity recognition as word-word relation classification[C]//proceedings of the AAAI conference on artificial intelligence: Vol. 36. 2022: 10965-10973.

[28] YUAN Z, TAN C, HUANG S, et al. Fusing heterogeneous factors with triaffine mechanism for nested named entity recognition[A]. 2021. arXiv: 2110.07480.

[29] GHOSH S, KUMAR S, KUMAR Y, et al. Span classification with structured information for disfluency detection in spoken utterances[C]//Proc. Interspeech 2022. 2022: 3998-4002.

[30] MAO H, MAO X L, TANG H, et al. Span graph transformer for document-level named entity recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 38. 2024: 18769-18777.

[31] YAN H, SUN Y, LI X, et al. An embarrassingly easy but strong baseline for nested named entity recognition[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2023: 1442-1452.

[32] ZHU E, LI J. Boundary smoothing for named entity recognition[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 7096-7108.

[33] RILOFF E. Automatically constructing a dictionary for information extraction tasks [C]//Proceedings of the eleventh national conference on Artificial intelligence. 1993: 811-816.

[34] APPELT D E, HOBBS J R, BEAR J, et al. Fastus: A finite-state processor for information extraction from real-world text[C]//Ijcai: Vol. 93. 1993: 1172-1178.

[35] AGICHTEIN E, GRAVANO L. Snowball: Extracting relations from large plain-text collections[C]//Proceedings of the fifth ACM conference on Digital libraries. 2000: 85-94.

[36] 刘克彬, 李芳, 刘磊, 等. 基于核函数中文关系自动抽取系统的实现[J]. 计算机研究与发展, 2007(08): 1406-1411.

[37] ZENG D, LIU K, LAI S, et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers. 2014: 2335-2344.

[38] 武文雅, 陈钰枫, 徐金安, 等. 基于高层语义注意力机制的中文实体关系抽取[J]. 广西师范大学学报 (自然科学版), 2019, 37(01): 32-41.

[39] CHEN D, MANNING C D. A fast and accurate dependency parser using neural networks[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 740-750.

[40] WEI Z, SU J, WANG Y, et al. A novel cascade binary tagging framework for relational triple extraction[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 1476-1488.

[41] WANG Y, YU B, ZHANG Y, et al. Tplinker: Single-stage joint extraction of entities and relations through token pair linking[C]//Proceedings of the 28th International Conference on Computational Linguistics. 2020: 1572-1582.

[42] YAN Z, ZHANG C, FU J, et al. A partition filter network for joint entity and relation extraction[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 185-197.

[43] SUI D, ZENG X, CHEN Y, et al. Joint entity and relation extraction with set prediction networks[J]. IEEE transactions on neural networks and learning systems, 2023: 12784-12795.

[44] 李丽双, 王泽昊, 秦雪洋, 等. 基于平行交互注意力网络的中文电子病历实体及关系联合抽取[J]. 中文信息学报, 2024, 38(06): 108-118.

[45] ZHONG Z, CHEN D. A frustratingly easy approach for entity and relation extraction[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021: 50-61.

[46] YE D, LIN Y, LI P, et al. Packed levitated marker for entity and relation extraction [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 4904-4917.

[47] FENG Y, CHEN Q, WU Q, et al. Sure: Mutually visible objects and self-generated candidate labels for relation extraction[C]//Proceedings of the 31st International Conference on Computational Linguistics. 2025: 459-468.

[48] 李希, 刘喜平, 李旺才, 等. 对比学习研究综述[J]. 小型微型计算机系统, 2023, 44 (04): 787-797.

[49] 张重生, 陈杰, 李岐龙, 等. 深度对比学习综述[J]. 自动化学报, 2023, 49(01): 15-39.

[50] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PmLR, 2020: 1597-1607.

[51] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9729-9738.

[52] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent-a new approach to self-supervised learning[J]. Advances in neural information processing systems, 2020, 33: 21271-21284.

[53] REIMERS N, GUREVYCH I. Sentence-bert: Sentence embeddings using siamese bert-networks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 3982-3992.

[54] GAO T, YAO X, CHEN D. Simcse: Simple contrastive learning of sentence embeddings[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 6894-6910.

[55] SU Y, LIU F, MENG Z, et al. Tacl: improving bert pre-training with token-aware contrastive learning[C]//North American Association for Computational Linguistics 2022: NAACL 2022. Association for Computational Linguistics (ACL), 2022: 2497-2507.

[56] WANG A, SINGH A, MICHAEL J, et al. Glue: A multi-task benchmark and analysis platform for natural language understanding[C]//International Conference on Learning Representations. 2019.

[57] RAJPURKAR P, ZHANG J, LOPYREV K, et al. Squad: 100,000+ questions for machine comprehension of text[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 2383-2392.

[58] KARPUKHIN V, OGUZ B, MIN S, et al. Dense passage retrieval for open-domain question answering[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 6769-6781.

[59] GRISHMAN R, SUNDHEIM B M. Message understanding conference-6: A brief history[C]//COLING 1996 volume 1: The 16th international conference on computational linguistics. 1996.

[60] LI J, SUN A, HAN J, et al. A survey on deep learning for named entity recognition [J]. IEEE transactions on knowledge and data engineering, 2020, 34(1): 50-70.

[61] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019: 4171-4186.

[62] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. Curran Associates, Inc., 2017.

[63] LIU Y, OTT M, GOYAL N, et al. Roberta: A robustly optimized bert pretraining approach[A]. 2019. arXiv: 1907.11692.

[64] ZHU Y, KIROS R, ZEMEL R, et al. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books[C]//Proceedings of the IEEE international conference on computer vision. 2015: 19-27.

[65] HAMBORG F, MEUSCHKE N, BREITINGER C, et al. news-please: A generic news crawler and extractor[C]//Proceedings of the 15th International Symposium of Information Science. 2017: 218-223.

[66] GOKASLAN A, COHEN V, PAVLICK E, et al. Openwebtext corpus[EB/OL]. 2019. http://Skylion007.github.io/OpenWebTextCorpus.

[67] TRINH T H, LE Q V. A simple method for commonsense reasoning[A]. 2018. arXiv: 1806.02847.

[68] RAJPURKAR P, JIA R, LIANG P. Know what you don't know: Unanswerable questions for squad[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2018: 784-789.

[69] GU Y, TINN R, CHENG H, et al. Domain-specific language model pretraining for biomedical natural language processing[J]. ACM Transactions on Computing for Healthcare (HEALTH), 2021, 3(1): 1-23.

[70] BELTAGY I, LO K, COHAN A. Scibert: A pretrained language model for scientific text[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 3615-3620.

[71] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[72] HENDRYCKS D, GIMPEL K. Gaussian error linear units (gelus)[A]. 2016. arXiv: 1606.08415.

[73] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[A]. 2014. arXiv: 1409.0473.

[74] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PMLR, 2021: 8748-8763.

[75] WU S, WANG G, TANG P, et al. Convolution with even-sized kernels and symmetric padding[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019: 1194-1205.

[76] WALKER C, et al. Ace 2005 multilingual training corpus: LDC2006T06[R]. Philadelphia: Linguistic Data Consortium, 2006.

[77] MITCHELL A, et al. Ace 2004 multilingual training corpus: LDC2005T09[R]. Philadelphia: Linguistic Data Consortium, 2005.

[78] KIM J D, OHTA T, TATEISI Y, et al. Genia corpus—a semantically annotated corpus for bio-textmining[J]. Bioinformatics, 2003, 19: i180-i182.

[79] PRADHAN S, MOSCHITTI A, XUE N, et al. Towards robust linguistic analysis using ontonotes[C]//Proceedings of the Seventeenth Conference on Computational Natural Language Learning. 2013: 143-152.

[80] TJONG KIM SANG E F, DE MEULDER F. Introduction to the conll-2003 shared task: language-independent named entity recognition[C]//Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. 2003: 142-147.

[81] LUAN Y, HE L, OSTENDORF M, et al. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 3219-3232.

[82] LU W, ROTH D. Joint mention extraction and classification with mention hypergraphs [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 857-867.

[83] MUIS A O, LU W. Labeling gaps between words: Recognizing overlapping mentions with mention separators[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 2608-2618.

[84] FINKEL J R, MANNING C D. Nested named entity recognition[C]//Proceedings of the 2009 conference on empirical methods in natural language processing. 2009: 141-150.

[85] WANG J, SHOU L, CHEN K, et al. Pyramid: A layered model for nested named entity recognition[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. 2020: 5918-5928.

[86] ZHANG S, SHEN Y, TAN Z, et al. De-bias for generative extraction in unified ner task [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 808-818.

[87] WANG S, SUN X, LI X, et al. Gpt-ner: Named entity recognition via large language models[A]. 2023. arXiv: 2304.10428.

[88] LU J, WANG Y, YANG Z, et al. Padellm-ner: parallel decoding in large language models for named entity recognition[J]. Advances in Neural Information Processing Systems, 2024, 37: 117853-117880.

[89] MIWA M, BANSAL M. End-to-end relation extraction using lstms on sequences and tree structures[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 1105-1116.

[90] LI X, YIN F, SUN Z, et al. Entity-relation extraction as multi-turn question answering [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 1340-1350.

[91] LIN Y, JI H, HUANG F, et al. A joint neural model for information extraction with global features[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. 2020: 7999-8009.

[92] WADDEN D, WENNBERG U, LUAN Y, et al. Entity, relation, and event extraction with contextualized span representations[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 5784-5789.

[93] SHEN Y, MA X, TANG Y, et al. A trigger-sense memory flow framework for joint entity and relation extraction[C]//Proceedings of the web conference 2021. 2021: 1704-1715.

[94] WANG Y, SUN C, WU Y, et al. Unire: A unified label space for entity relation extraction[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 220-231.

[95] ZARATIANA U, TOMEH N, HOLAT P, et al. An autoregressive text-to-graph framework for joint entity and relation extraction[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 38. 2024: 19477-19487.

[96] LI Q, YAO N, ZHOU N, et al. Entity and relationship extraction based on span contribution evaluation and focusing framework[J]. Computer Speech & Language, 2025, 90: 101744.

# 攻读硕士学位期间取得的研究成果

## 一、论文

[1] Zhang R, Ling C, Chen Q,et al. ACAM: An Adaptive Context-aware Attention Module for Nested Named Entity Recognition. 2025 6th International Conference on Computer Information and Big Data Applications （EI 会议，已录用，导师一作，本人二作）

[2] Zhang R, Ling C, Han Y,et al. Improving Relation Extraction with Contrastive Learning-based Named Entity Recognition. 2025 6th International Conference on Computer Engineering and Application（EI 会议，已录用，导师一作，本人二作）

[3] Zhang R, Ling C, Han Y,et al. Compensating Semantic Information for Span Embedding by Even-sized Kernel Convolution. （修改中）

## 二、软著

[1] 软件名称：基于嵌套命名实体识别的通用知识挖掘和统计分析平台 V1.0，开发人：韩越兴、凌晨帆、张瑞。登记号：2024R11L1545295，登记日期：2024.10.21，申请人：上海大学。

# 致　谢

时光飞逝，数载求学之路即将画上句点。回望这段漫长而充实的旅程，心中充满无限感慨与感激。此刻，笔落至此，谨以寥寥数语，向所有曾给予我温暖与力量的人们表达我最深切的谢意。

首先我要感谢我的导师张瑞老师。犹记初入校园之际，寻找导师的过程中屡次碰壁，我心中充满了迷茫与不安，是张老师向我伸出了橄榄枝，让我有幸成为课题组这个温暖大家庭中的一员。在科研中，无论是实验设计的反复推敲，数据处理的一丝不苟，还是论文措辞的再三斟酌，无不体现着张老师严谨求实的治学风范。这种精益求精的精神，潜移默化地影响了我，使我逐渐养成了求真务实的学术品格。在生活中，张老师是一位平易近人的长辈。每当我因科研压力或生活琐事感到挫败时，张老师总是不吝时间，耐心开导，帮助我一次次走出阴霾。此外，我还要诚挚感谢韩越兴老师和陈侨川老师。在论文的撰写过程中，二位老师给予了宝贵的指导，提出了富有启发性的思路，并对论文进行了悉心的斧正。谨此，祝愿老师们身体健康，学术常青。

这段漫长而充实的求学之路，我并非一人独行。最庆幸的是，有一群志同道合的伙伴与我并肩前行。首先，我要感谢课题组的伙伴们，在研究生的这三年里，我们互相勉励、相互扶持度过了重重困难，我们曾一同在实验室奋战至深夜，也曾在健身房中锤炼体魄，还曾并肩站在真正的高山之巅，俯瞰壮丽风景。这段共同成长的经历将是我一生宝贵的财富。我也要感谢包括我室友和 106 实验室在内的同学们。在紧张的科研生活之外，是他们的陪伴让我的闲暇时光同样丰富多彩。我们曾在球场上挥洒汗水，也曾在饭桌上畅谈未来。正是这些苦中作乐的时光，为枯燥的科研生活注入了活力与温度，也让我拥有了继续前行的动力。

在所有的支持中，最深沉、最持久的，永远来自我的父母。他们无私的奉献与毫无保留的爱，是我能够完成学业最根本的保障。在经济上，他们为我免去所有后顾之忧；在精神上，他们的理解与支持是我克服困难、坚持到底的力量源泉。

数载求学路，终点亦是新的起点。再次向所有给予我无私帮助与温暖关怀的师长、亲友致以最真挚的谢意。愿自己常怀感恩之心，在未来的道路砥砺前行。