

北京交通大学

硕士学位论文

基于图卷积的关系抽取方法研究

Research on Relation Extraction Based on Graph Convolutions

作者：赵 义

导师：万怀宇

北京交通大学

2020 年 6 月

学位论文版权使用授权书

本学位论文作者完全了解北京交通大学有关保留、使用学位论文的规定。特授权北京交通大学可以将学位论文的全部或部分内容编入有关数据库进行检索，提供阅览服务，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：

赵义

导师签名：

万冰宇

签字日期：2020年6月15日

签字日期：2020年6月15日

学校代码：10004

密级：公开

北京交通大学

硕士学位论文

基于图卷积的关系抽取方法研究

Research on Relation Extraction Based on Graph Convolutions

作者姓名：赵 义

学 号：17120452

导师姓名：万怀宇

职 称：副教授

学位类别：工学

学位级别：硕士

学科专业：计算机科学与技术

研究方向：数据与知识工程

北京交通大学

2020 年 6 月

致谢

交大七年，岁月青葱。回首这七年时光，有欢笑也不乏泪水，有收获也伴随失去，有成长更不乏遗憾。在交大的这七年将会是我人生中最宝贵的一段时光，交大也记录着我的成长，从一个初入校园时懵懵懂懂到如今的成熟自信。在毕业之际，我想首先对亲爱的母校说一声感谢！百廿时光，你迎来送走无数莘莘学子，你伴随着每一名交大人的成长，交大人也见证你的变迁，愿母校再创辉煌，我将永远以作为一名交大人作为荣！交大其实就是由一群可爱可敬的交大人所构成的，正是这些可爱可敬的交大人伴随着我的七年交大时光，给我帮助和鞭策，在此我衷心地向所有指导过我的老师们、陪伴和帮助我的同学们和朋友们表示深深的感谢，谢谢你们的教导和帮助让我鞭策着我不断地成长和进步！

感谢我的导师万怀宇老师！研究生阶段的学习和科研任务并不轻松，对于转专业的我来说尤其如此，感谢万老师在我初入实验室阶段对我的细心指导帮助我快速找到方向、进入角色；也是万老师的信任和包容让我能够快速融入实验室的项目之中，收获宝贵的实战经验；更是万老师的以身作则，其勤勉的工作态度，一直鞭策和激励着我不断学习进步。“为人师者，必先正其身”，正是您的言传身教，让我更明确地认识自我，找到自己的方向，并满怀为之奋斗的动力；“师者，所以传道授业解惑也”，正是您通过一次次的耐心的研讨，为我指明了研究的方向，教会了我科研的方法，传递给我科研的态度。传道授业解惑的背后展现的是您对学生的爱心、耐心和责任心。在此，郑重地感谢万老师在工作学习和生活等方方面面对我的关心和照顾，您教会我的一切将使我受益终生。

感谢实验室的其他各位老师对我工作和学习上的指导和帮助。感谢林友芳老师，从项目的立项到实验成果的交付，从论文的选题立意到最终定稿，您高瞻远瞩的指导使我深感钦佩。感谢武志昊老师和王晶老师，在我的学习过程中给予了许多醍醐灌顶的指导，让我得以掌握更多专业知识。感谢张硕老师在论文撰写阶段为我提供的宝贵建议，并督促我保质保量地完成论文的撰写。感谢实验室提供的良好学习和工作环境，让我能够顺利地完成研究生学业与课题。

“学贵得师，亦贵得友”，良师必然有益友相伴。感谢张东雷、马语丹等师兄师姐，是你们在我初入实验室时给我指导和包容，在我遇到困难之时基于帮助和建议。感谢高建伟、宋龙泽、郭倩、孙盼等师弟师妹，在科研项目和学习中，对我的帮助和支持，让我们可以一起完成一个又一个项目指标和学习任务。感谢纪宇泽、宋超、曹仲、金婧、辛颖等同窗，因为有你们的朝夕相处和互帮互助，让我度过了充实而快乐的研究生时光。感谢健身小队的所有小伙伴，是你们的陪伴，让我在研究生期间能够在学业上有所收获的同时，也拥有了一副健康的体魄。

此外，我还要特别感谢计算机学院篮球队的所有成员，不仅仅是每一位队员，还有给予球队大力支持的院领导和团委老师们、背后默默付出的工作人员以及所有为球队呐喊过的计算机人，每一个为计算机学院篮球队付出过心血的人。篮球队承载着我们每一个人的努力和心血，我们一起经历过低谷，也攀上过山峰；一起流过汗，也抹过泪。它是我们的共同记忆，是我们归属感的来源，是它让我们的大学生活更加精彩。在此，我衷心希望计算机篮球队能够将这份归属感和凝聚力传承下去，再创辉煌！

感谢我的父母与家人，感谢你们对我的养育之恩和无微不至的关怀，感谢你们一直以来对我的默默付出和鼓励，让我拥有了离乡背井展翅高飞的勇气，有了追求梦想坚持不懈的决心。

最后，特别感谢百忙之中参加论文评审和答辩的各位专家和老师们，感谢你们为本文提出的宝贵意见。

摘要

得益于信息产生和传播的成本降低, 各类信息在网络上迅速增加并且广泛传播, 人们获取信息的便利性大大提高。然而, 随之而来的信息爆炸问题却开始影响人们检索信息的效率, 如何从互联网上海量的信息中精准地获取用户的目标信息成为了亟待解决的问题。为了解决这一问题, 信息抽取技术应运而生, 并在近些年获得了研究学者的广泛关注。

关系抽取是信息抽取领域中的一项关键任务。现有的关系抽取方法主要依靠外部信息和背景知识来提升关系抽取的效果, 忽略了实体与实体之间以及实体对与实体对之间天然存在的关联关系。这些关联关系可以有效地辅助实体之间的关系抽取任务。本文使用图的形式来表示上述关联关系, 并设计了新颖的基于图的神经网络模型, 采用注意力机制和图卷积等技术来捕获图中的拓扑关系, 并结合关系实例中包含的上下文语义特征, 实现有效的关系抽取。

首先, 本文提出了基于实体关联图和注意力机制的关系抽取模型 (entity relational graph based attentional neural network, ERANN)。考虑到实体在语料集中的关联实体信息能辅助目标实体的关系抽取, 本文提出了实体关联图的概念。依据实体之间的潜在关联关系构建的实体关联图, 并通过注意力机制为实体的关联实体计算关联权重, 构建关联实体特征, 然后使用结合自注意力机制的双向门控循环神经网络提取句子语义特征, 最后结合关联特征和句子语义特征进行关系抽取。

其次, 本文进一步提出了基于实体对图和图卷积的关系抽取模型 (entity pair graph based neural network, EPGNN)。针对 ERANN 模型以独立实体为对象, 导致模型丧失实体对中整体关联性的问题, 本文将实体对作为节点, 将公共实体作为边, 构建得到实体对图代替实体关联图。鉴于关联关系在实体之间的传递性, 仅仅考虑实体的一阶关联关系不够全面, 而 ERANN 模型受限于注意力机制的特点, 对多阶实体关联关系无能为力。为了克服这一不足, EPGNN 模型使用多层图卷积网络捕获目标实体对的拓扑特征来表征多阶关联关系, 并使用预训练语言模型 BERT 提取句子语义特征, 最后结合实体对图拓扑特征和句子语义特征进行关系抽取。

最后, 本文以当前主流的关系抽取方法作为基准、在两个本领域的常用公开数据集上对提出两个模型进行了对比实验和结果分析。ERANN 模型通过构建实体关联图, 充分挖掘目标实体对的关联关系, 在句子语义特征的基础上补充了目标实体对的全局性结构特征, 取得了比现有方法更好的效果; 经过实体关联图构建和特征构建方法改进后的 EPGNN 模型, 能更有效地捕获多阶实体对拓扑特征和文本语义特征, 在两个公开数据集上取得了当前最优的关系抽取效果。

关键词: 关系抽取; 实体对图; 图卷积网络; 预训练语言模型; 拓扑特征

ABSTRACT

With the development of information technology and the popularization of the Internet, there are mass of information been generated and widely spread on the Internet. Because of the reduction of the cost of information generation and dissemination, the convenience of people to obtain information has greatly improved. However, the subsequent information explosion problem has started to affect the efficiency of people's information retrieval. How to obtain the information users need form the amount of information on the Internet quickly and accurately has become an urgent problem. In order to solve this problem, information extraction technology has emerged as the times require, and has attracted widespread attention from research scholars in recent years.

Relation extraction is one of the most important tasks in the field of information extraction. Existing relation extraction approaches mainly rely on exploiting external resources and background knowledge to improve the performance and ignore the correlations between entity pairs which are helpful for relation extraction. We present the concept of entity pair graph to represent the correlations between entity pairs and proposed a novel entity pairs graph based neural network model, relying on graph convolutional network to capture the topological features of an entity pair graph. EPGNN combines sentence semantic features generated by pre-trained BERT model with graph topological features for relation extraction.

We proposed an entity association graph and attention based model for relation extraction. Considering that the information of the related entities in the corpus can assist the relation extraction of the target entity, we proposed the concept of entity association graph. An entity association graph constructed according to the co-occurrence relation between entities. The weights of the related entities are calculated by the attention mechanism. Then the features of associated entities are formed by the weighted sum of the features of all related entities. The combination of self-attention and Bi-GRU is used to extract sentence semantic features. Combining the features of associated entities and sentence semantic features for relation extraction.

Based on the ERANN model, we proposed an entity pair graph based model (EPGNN). For the ERANN model, the independent entity is the object, resulting in the loss of the entity's overall association, we take entity pairs as nodes and common entities as edges to construct entity pair graph. In view of the transitivity of the relationship between entities, it is not comprehensive enough to consider only the directly related

entities of the entity. The ERANN model is limited by its own characteristics of the attention mechanism, and is powerless for multi-level entity association relationships. To overcome this deficiency, EPGNN uses a multi-layer graph convolutional network to capture the topological features of the target entity pair. A pre-trained language model BERT is used to extract the sentence semantic features. EPGNN combines graph topological features and sentence semantic features for relation extraction.

To verify the performance of our proposed model, the current mainstream relation extraction methods are used as benchmarks, and comparative experiments and results analysis are performed with the proposed ERANN model and EPGNN model on two public data sets commonly used in the fields of relation extraction. The ERANN model fully exploits the association relationship of target entity pairs by constructing entity association graphs, and then complements the global characteristics of target entity pairs based on sentence semantic features, and achieves better results than existing methods. The improved EPGNN model of feature construction method can more effectively capture the topological features and text semantic features of multi-order entities, and has achieved the current optimal relationship extraction effect on two public data sets.

KEYWORDS: relation extraction; entity-to-graph; graph convolutional network; pre-trained language model; topological features

目录

摘要	iv
ABSTRACT	v
1 引言	1
1.1 研究背景及意义	1
1.2 研究内容及创新点	2
1.3 论文组织结构	4
2 理论基础与研究现状	5
2.1 相关理论基础	5
2.1.1 注意力机制	5
2.1.2 预训练语言模型	7
2.1.3 图卷积网络	9
2.2 关系抽取数据集	11
2.2.1 ACE 2005 数据集	11
2.2.2 SemEval 2010 数据集	12
2.3 评价标准	12
2.4 研究现状	14
2.4.1 基于特征工程的关系抽取方法	14
2.4.2 基于神经网络的关系抽取方法	15
2.5 本章小结	16
3 基于实体关联图与注意力机制的关系抽取	17
3.1 实体关联图定义	17
3.2 模型总体框架	18
3.3 文本语义特征抽取	18
3.3.1 词向量表示	18
3.3.2 文本语义特征构建	20
3.4 实体关联特征抽取	21
3.4.1 实体关联图节点嵌入表示	21
3.4.2 邻居节点注意力机制	22
3.5 关系抽取	23
3.6 实验设置与结果分析	23
3.6.1 实验参数设置	24

3.6.2	对比实验设置	24
3.6.3	实验结果分析	25
3.7	本章小结	28
4	基于实体对图与图卷积的关系抽取	29
4.1	实体对图定义	29
4.2	模型总体框架	30
4.3	文本语义特征抽取	31
4.3.1	基于预训练语言模型 BERT 的文本表示	32
4.3.2	文本语义特征构建	33
4.4	实体对图拓扑特征抽取	34
4.4.1	实体对图节点嵌入表示	34
4.4.2	多层图卷积网络	35
4.5	关系抽取	37
4.6	实验设置与结果分析	37
4.6.1	实验设置	37
4.6.2	对比实验设置	38
4.6.3	实验结果分析	39
4.7	本章小结	43
5	结论	44
5.1	工作总结	44
5.2	未来工作展望	45
	参考文献	46
	作者简历及攻读硕士学位期间取得的研究成果	50
	独创性声明	51
	学位论文数据集	52

1 引言

本章首先对实体关系抽取的研究背景及意义进行简要介绍；然后对本文的研究内容及创新点展开详细阐述；最后对本文的内容构成进行介绍。

1.1 研究背景及意义

信息为人类决策提供有效数据，是人类活动中不可或缺的一种资源，一直以来都是学者们的重点研究对象。信息论中信息的定义为：消除随机不确定性的东西。然而，随着互联网的快速发展，传统信息技术中的时间和空间障碍得以克服，使得信息的产生和传播成本大大降低。与此同时，信息本身的不确定性也随着增加，大量的虚假信息以及低价值信息充斥着网络环境，形成了互联网时代信息爆炸的问题。信息的获取变得更加容易，但是获取高质量信息的成本却不降反升。在这种背景下，如何有效地利用互联网上海量的信息资源，快速准确地从其中抽取出真正有价值的信息，已经成为信息技术领域亟待解决的问题。

语言文字是人类知识传播的主要载体之一。自然语言处理（Natural Language Processing, NLP），作为人工智能领域的重要研究方向，其目标是使计算机通过自然语言理解人的思维逻辑。信息的形式多种多样，如今主要以文本、音频和视频等形式存在于互联网上，其中又以文本形式的信息为主。互联网时代海量的文本数据为自然语言处理领域的技术发展提供了充足的资源。如何从互联网中海量的文本资源中抽取得到有价值的信息，从而更有效地利用资源，成为了自然语言处理领域的重要研究课题。信息抽取技术就此应运而生，其任务主要从非结构化文本数据中抽取得到命名实体、实体间关系以及事件这三类关键信息。完成命名实体识别后，以句子或段落为单位对其中包含的命名实体之间的关系进行分类，完成关系抽取；或者通过选择触发词串联文本中包含的实体元素，完成事件发现。本文将主要对其中的关系抽取任务进行研究。

《自然语言处理综述》^[1]将关系抽取任务定义为：对于给定句子中的两个实体，通过上下文信息找到两个实体之间的语义关系，并对这些关系进行分类。关系抽取作为信息抽取领域的关键任务，其成果不仅可以服务于知识图谱中的实体关系网络构建这种显式关系需求任务，同样可以应用于问答系统和机器翻译等需要对非结构化文本进行关系理解的自然语言处理系统^[2]中。因此，关系抽取结果的准确与否将直接影响下游任务的结果。关系抽取任务可以根据是否预先设置实体关系类型的标签空间，分为有监督和无监督两种关系抽取方法^[3]。有监督关系抽取方法利

用标注数据中的监督信息,将关系抽取视为分类任务,通过计算目标实体对属于各类关系的概率大小来确定其关系类型;无监督关系抽取方法不依赖于标注数据,而需要更大规模的语料进行学习。本文主要针对有监督关系抽取方法进行研究。

1.2 研究内容及创新点

实体关系抽取近些年来一直是自然语言领域学者的研究重点。早期的关系抽取方法多为基于特征工程的分类方法,其主要思想是从关系实例的上下文中抽取得到诸如词法、句法和语法等语义特征,基于上述语义特征构造特征向量,并使用诸如模式匹配^{[4][5][6]}、支持向量机^{[7][8]}和条件随机场^[9]等传统机器学习方法完成关系分类。这类基于特征工程的关系抽取方法^{[10][11][12]}往往依赖于依存关系分析器、词法句法分析器等自然语言处理工具,其关系分类准确率也很大程度上受到这些自然语言处理工具的影响。同时由于特征工程往往具有数据集特异性,因此这些关系抽取模型的可扩展性和复用性较差。近几年,得益于神经网络模型的广泛应用,关系抽取领域得以飞速发展,提出了诸如基于卷积神经网络(CNN)和基于循环神经网络(RNN)的等关系抽取模型^{[13][14][15]}。相较于传统的实体关系抽取方法,基于神经网络的方法无需大量的人工特征工程,模型可以端到端地学习特征表示,减少了对自然语言处理工具的依赖,大大削弱了文本预处理结果对结果的影响,提升了关系抽取模型的稳定性、复用性和扩展性。由于仅从关系实例中通过神经网络模型提取的特征无法带来较高的关系分类准确率,研究人员开始利用通过自然语言处理工具提取得到的结构特征和实体背景知识^{[16][17]}补充信息,辅助关系抽取任务。但是,这些模型仍然存在一些不足之处。例如,使用自然语言处理工具可能导致误差在网络中的传播和累计;使用实体背景知识导致模型过于局限于特定语料,丧失了模型的复用性和扩展性。同时,当前大多数的关系抽取模型将目标实体对作为独立对象^{[16][18]},仅通过神经网络从关系实例中提取上下文语义特征,忽略了在语料集层次的实体关联关系。

基于上述问题和不足之处,本文考虑整个语料集中包含的所有实体对,认为与目标实体对包含公共实体的实体对可以用于辅助目标实体对的关系抽取。具体来说,一个实体大概率不仅仅只在当前关系实例中出现,或者只与某一个实体之间存在关系。在语料足够丰富的情况下,一个实体可能与多个实体之间存在潜在关系。将整个语料集中的所有实体之间的关联关系提取出来,以不同的形式表示就可以分别形成以实体为节点的实体关联图和以实体对为节点的实体对图。为了有效地捕获实体之间的关联关系,构建全局特征辅助关系抽取任务,本文分别提出了基于注意力机制和基于多层图卷积的全局结构特征抽取方法,并将之分别与基于词向

量^[19]和双向门控循环神经网络 (Bidirectional Gated Recurrent Unit Network, Bi-GRU) 以及预训练语言模型 BERT^[20]学习得到的文本语义特征结合, 形成了基于实体关联图与注意力机制的关系抽取方法和基于实体对图与图卷积的关系抽取方法。本文在两个公开关系抽取数据集 ACE 2005^[21]以及 SemEval 2010 Task 8^[22]上进行了大量对比实验以及模块孤立实验, 验证了所提出的模型的效果。本文的主要研究内容和贡献可以总结为以下几点:

(1) 实体关联关系图构建

本文认为在实体关系抽取任务中, 目标实体对的关系类型不仅能够由包含目标实体对的关系实例表示, 与目标实体对包含公共实体的实体对也可以为目标实体对的关系抽取提供有价值的信息。因此, 本文对整个语料集中的实体关联关系进行提取, 采用两种不同的方式构建图结构, 分别是以实体为节点, 以关联关系为边的实体关联图, 以及以实体对为节点, 以公共实体为边的实体对图。根据两种实体关联关系图各自的特点, 设计并实现了两种特征抽取方法, 并通过实验对本文提出的两个基于图的关系抽取模型进行对比。

(2) 实体关联关系特征提取

实体关联关系图构建完成后, 目标实体对的关系可以由实体关联关系图的拓扑结构来表示, 本文针对两种不同的实体关联关系图分别设计了拓扑特征抽取方法。在基于实体关联图与注意力机制的关系抽取方法中, 采用注意力机制计算目标实体邻居节点的关系权重, 使用带权邻居节点表示和目标实体节点表示构建目标实体对的拓扑特征; 在基于实体对图与图卷积的关系抽取方法中, 采用多层图卷积操作融合目标节点表示与关联节点表示, 使用图卷积操作得到的目标节点表示作为拓扑特征。本文通过模块孤立实验验证拓扑特征给模型带来的提升。

(3) 基于图的关系抽取

针对实体关联图和实体对图分别设计了两种基于图的关系抽取方法。在基于实体关联图与注意力机制的关系抽取方法中, 使用 Bi-GRU 提取上下文语义信息特征与实体关联图拓扑特征结合, 综合考虑实体关联信息与上下文语义信息, 完成实体关系抽取; 再基于实体对图与多层图卷积的关系抽取方法中, 考虑到关系抽取任务对于单个关系实例可以视为句子分类任务, 本文采用当前在文本分类等自然语言处理下游任务中表现突出的预训练语言模型 BERT 学习句子表示和目标实体表示, 使用多层图卷积网络捕获实体对图上的多阶实体关联信息, 并将之与句子上下文信息进行结合, 完成关系抽取。在两个公开数据集上与当前主流方法的对比实验结果表明, 本文提出的基于图的关系抽取方法取得了当前最优的结果。

与当前关系抽取领域的其他方法相比, 本文所提出的基于实体关联关系图的关系抽取方法不依赖外部信息补充, 打破了传统关系抽取模型仅从单个句子中提

取文本语义特征的思路,从整个语料集的角度出发,使用目标实体对在其中的关联实体作为补充信息,通过构建实体关联关系图的形式将实体的关联关系特征应用到关系抽取任务之中。在基于实体关联图与注意力机制的关系抽取方法中,本文构建了以实体作为节点的实体关联图来表征实体关联关系,并通过注意力机制对实体关联图中目标实体的邻居实体关联权重进行计算,以目标实体和其关联实体的带权表示作为实体关联特征;在基于实体对图与图卷积的关系抽取方法中,本文构建了以实体对作为节点的实体对图来表征实体关联关系,通过多层图卷积操作将目标实体对的表示与其关联实体对表示进行融合,融合后的目标实体对表示作为其拓扑特征。在两个公开数据集上的对比实验结果表明,本文提出的基于实体关联关系图的关系抽取方法在不适用任何外部信息的情况下,取得了当前领域内领先的结果;同时,通过在基础模型上引入拓扑特征,进行模块孤立实验来证明引入实体对图拓扑特征能够有效提升模型的效果。

1.3 论文组织结构

本文一共分为五章,其中第一章为引言,主要介绍了课题的研究背景及意义,并提出了本文的研究内容和创新点。本文的剩余部分组织如下:

第二章为实体关系抽取研究综述,主要介绍本文研究内容中涉及的相关基础理论,包括预训练语言模型、注意力机制和图卷积的相关知识,接着对关系抽取领域内广泛使用的公开数据集进行详细介绍,然后对关系抽取模型的评价标准和研究现状进行简要描述。

第三章为基于实体关联图与注意力机制的关系抽取方法,首先对实体关联图的概念进行阐述,然后对本章提出的关系抽取模型整体框架进行详细介绍,接下来分别对实体关联特征抽取模块、文本语义特征抽取模块和关系分类模块进行解释说明,最后对实验设置、对比方法及实验结果分别予以阐述。

第四章为基于实体对图与图卷积的关系抽取方法,首先对实体对图的概念进行阐述,然后对本章提出的关系抽取模型整体框架进行详细介绍,接下来分别对实体对图拓扑特征抽取模块、文本语义特征抽取模块和关系分类模块进行解释说明,最后对实验设置和实验结果进行说明,在本章中还将 ERANN 模型与 EPGNN 模型进行对比,用于进一步说明基于图卷积提取的实体对图拓扑特征在关系抽取任务中的有效性。

第五章为总结和展望,将会对本文的工作内容进行总结,并对未来的研究方向和研究内容进行分析和展望。

2 理论基础与研究现状

本章将介绍本文的理论基础与研究现状。首先对本文提出的关系抽取方法所涉及的相关理论进行介绍，包括预训练语言模型、注意力机制和图卷积网络；其次对本文实验所使用的，关系抽取领域常用的两个公开数据集进行介绍；然后对关系抽取领域常用的评价指标进行说明；最后对关系抽取领域的研究现状进行阐述。

2.1 相关理论基础

本节将对本文所提出的基于实体关联图与注意力机制的关系抽取方法以及基于实体对图与多层图卷积的关系抽取方法所涉及的关键技术的相关理论基础进行详细阐述，包括用于提取关系实例中的文本语义信息的预训练语言模型，用于学习实体关联图中目标实体表示的注意力机制，以及用于捕获实体对图拓扑结构特征的图卷积网络。

2.1.1 注意力机制

注意力机制（Attention Mechanism, AM）最初在计算机视觉领域提出^[23]，近年来在深度学习的各个领域大放异彩，取得了重大成功。注意力机制逐步成为了人工智能领域的研究热点，在自然语言处理、计算机视觉以及语音识别等人工智能领域越来越多优秀的基于注意力机制的神经网络模型涌现而出^{[24][25][26]}。

注意力机制参考人类的视觉系统，在观察时有选择性地侧重高区分度信息，同时削弱冗余信息的影响^[27]。视觉系统这种对于信息重要程度的区分能力是基于以往的经验学习得到的，被这种对信息的区分能力所启发，研究人员提出了注意力机制的概念，对模型获取的信息进行权重分配。注意力机制在神经网络领域得以快速发展的主要原因有以下三点。首先，在如机器翻译、问答系统、情感分析和对话系统等众多任务中，引入注意力机制的模型都取得了当时最好的结果，证明注意力机制的有效性；其次，除了提升模型性能以外，注意力机制还能够提升神经网络模型的可解释性；最后，注意力机制还有助于克服循环神经网络对输入长度以及输入顺序过于敏感的问题。综合上述三点优势，注意力机制在人工智能领域取得了长足的发展。

在自然语言处理领域，Bahdanau 等^[28]首先在神经网络机器翻译（Neural Machine Translation, NMT）模型中引入注意力机制，通过将翻译和对齐联合学习，

将编码器所学习到的源语言序列的每个词的语义向量与目标语言序列中的每个待翻译词建立联系，有效地提升了模型的翻译效果。Loung 等^[29]为神经机器翻译提出了两种简单有效的注意力机制，全局注意力机制和局部注意力机制，并通过实验分析说明了两种不同的注意力模型所适用的场景。此外，能够在单个句子中捕获深度上下文信息的自注意力机制也在机器翻译和文本表示等自然语言处理关键任务上取得了突出表现^{[30][31]}。

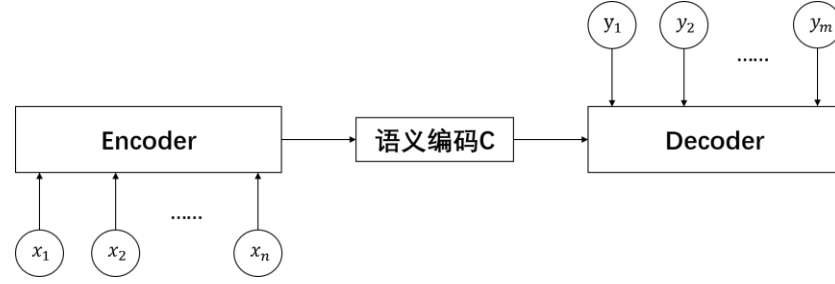


图 2-1 Encoder-Decoder 模型框架

Figure 2-1 Architecture of Encoder-Decoder

注意力机制在自然语言领域多是基于编码-解码器（Encoder-Decoder）框架实现的，框架如图 2-1 所示。Encoder 通过非线性变换将输入不定长序列 $X = (x_1, x_2, \dots, x_n)$ 转换成语义编码 $C = f(x_1, x_2, \dots, x_n)$ ，Decoder 在 i 时刻结合 Encoder 生成的语义编码 C 与 Decoder 的历史输出来预测 $y_i = g(y_1, y_2, \dots, y_{i-1}, C)$ ，最终输出不定长序列 $Y = (y_1, y_2, \dots, y_m)$ ，其中 $f(\cdot)$ 和 $g(\cdot)$ 均为非线性函数。Encoder-Decoder 基础上引入注意力机制，其本质是改变了中间语义表示 C 的计算方式，使得输入序列中单词对预测序列中每个单词的影响有所区别，对于 t 时刻的目标词 y_t 其对应的中间语义表示 C_t 可由式（2-1）计算得到：

$$C_t = \sum_{j=1}^T a_{t,j} h_j \quad (2-1)$$

其中， T 是输入序列的长度； $a_{t,j}$ 为 Encoder 端的第 j 个词对当前目标词 y_t 的注意力权重系数； h_j 表示 Encoder 端第 j 个词的隐向量。 $a_{t,j}$ 表示一个对齐模型，用于衡量输入序列中第 j 个词，对于输出序列 t 时刻的目标词的重要性，其计算公式如式（2-2）和式（2-3）所示：

$$a_{t,j} = \frac{\exp(a_{t,j})}{\sum_{j=1}^T \exp(a_{t,j})} \quad (2-2)$$

$$a_{t,j} = \text{score}(s_{t-1}, h_j) \quad (2-3)$$

其中， s_{t-1} 为 Decoder 端在 $t-1$ 时刻的隐状态。显然， $a_{t,j}$ 概率值之和为 1，即 $\sum_j a_{t,j} = 1$ 。

目前主流的对齐方式 $score()$ 有加性、点乘等 4 种计算方法, 具体如式 (2-4) 所示:

$$score(s_{t-1}, h_j) = \begin{cases} s_{t-1}^T h_j & dot \\ s_{t-1}^T W_a h_j & general \\ W_a[s_{t-1}; h_j] & concat \\ v_a^T \tanh(W_a s_{t-1} + U_a h_j) & perceptron \end{cases} \quad (2-4)$$

其中, W_a , U_a 以及 v_a 均为权重矩阵。

自注意力机制不同于传统的注意力机制, 仅面向一个对象进行计算, 只从该对象自身提取相关信息, 而不使用其他额外信息。在自注意力机制中, 权重计算发生在源端内部元素之间或者目标端内部元素之间, 其计算公式如式 (2-5) 所示:

$$A = \text{softmax}(v_a \tanh(W_a H^T)) \quad (2-5)$$

其中, $H = (h_1, h_2, \dots, h_T)$ 为输入序列的隐向量; W_a 和 v_a 为权重矩阵; A 为最终得到的注意力矩阵。由此可见, 自注意力机制计算过程中会直接将句子中任意两个单词的联系通过一个计算步骤直接联系起来, 所以远距离依赖特征之间的距离被极大缩短, 有利于有效地利用这些特征。除此以外, 自注意力机制对于增加计算的并行性也有直接帮助作用, 这也是为何自注意力机制逐渐被广泛使用的主要原因。

2.1.2 预训练语言模型

近几年来在自然语言处理领域中, 预训练语言模型^{[32][33]}在多项 NLP 任务中都取得了令人振奋的成果。语言模型 (Language Model, LM) 是一串词序列的概率分布, 它的作用是为一个长度为 m 的文本确定一个概率分布 P , 表示这段文本存在的可能性。但是, 在实践中如果文本长度较长, $P(w_i | w_1, w_2, \dots, w_{i-1})$ 的估算会非常困难。为了简化这一任务, 研究人员只使用当前词的前 $n-1$ 个词来估算条件概率, 提出 n 元 (n-gram) 模型。在 n-gram 模型中, 传统的方法一般采用频率计算的比例来估算 n-gram 条件概率, 其计算公式如式 (2-6) 所示。

$$P(w_i | w_{i-(n-1)}, \dots, w_{i-1}) = \frac{\text{count}(w_{i-(n-1)}, \dots, w_{i-1}, w_i)}{\text{count}(w_{i-(n-1)}, \dots, w_{i-1})} \quad (2-6)$$

当 n 较大时, 会存在数据稀疏的问题, 导致估算结果不准确。因此, 一般采用三元模型。神经网络语言模型就是为了解决数据稀疏这一问题而提出的。代表性工作是 Bengio 等^[34]提出的神经网络语言模型, 该语言模型使用了一个三层前馈神经网络来进行建模, 实验过程中发现其第一层参数作为词表示具有维度低、语义蕴含丰富等优点, 为后续词向量的发展打下了基础。语言模型是根据上下文去预测下一个词, 因此不需要人工标注数据, 可以无限制地从大规模语料中学习到丰富的语义知识。

目前神经网络在进行训练的时候基本是基于反向传播 (Back Propagation, BP)

算法, 通过对网络模型参数进行随机初始化, 然后通过 BP 算法和优化算法去优化模型参数。而预训练模型与之的区别在于, 其模型的参数不再是随机初始化, 而是先由面向一个任务训练得到一套模型参数, 然后利用这套参数对模型进行初始化, 再进行训练。本节将详细介绍本文在模型中所使用的预训练语言模型 BERT。

BERT 是 Google 团队提出的预训练语言模型, 它的全称是 “Bidirectional Encoder Representation from Transformer”, 其模型构成是基于 Transformer 的双向编码器表示。BERT 旨在通过联合调节所有层中的上下文来预先训练深度双向表示。因此, 预训练的 BERT 表示可以通过一个额外的输出层进行微调, 无需针对具体任务做大幅框架改动就可以适用于广泛的自然语言处理任务, 如问答系统和语义推理。

Transformer 是 Google 机器翻译团队^[35]提出的神经网络模型, 在该模型中使用之注意力机制代替传统的 RNN 来实现 Encoder-Decoder, 提升了模型的并行程度的同时还大大提升了模型特征抽取的能力。模型是典型的 Encoder-Decoder 结构, 其中 Encoder 部分由多个相同的组件堆叠而成, 每个组件由多头注意力机制和前馈神经网络组成; Decoder 部分也是由多个相同的组件堆叠而成, 每个组件在 Encoder 组件的基础上加入了包含掩码的多头注意力机制层。其中注意力机制的计算公式如式 (2-7) 所示, 多头注意力机制的由式 (2-8) 和 (2-9) 计算得到:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (2-7)$$

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^o \quad (2-8)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2-9)$$

在 Encoder 中, Q 表示 query 向量, K 表示 key 向量, V 表示 value 向量, 三者均为模型输入; d_k 表示 key 向量的维数, 引入 $\sqrt{d_k}$ 的目的是为了防止由于 key 向量维度过大, 导致点积结果过大, $softmax()$ 函数进入梯度极小的范围而难以训练。在 Decoder 中, Q 为 Decoder 每一层的输出; K 和 V 均为 Encoder 的最终输出。在 Decoder 中加入 mask 操作是由于文本生成的顺序是从前往后的, 每个词只能观察到之前的词。除了 Encoder-Decoder 中的自注意力机制之外, Encoder 与 Decoder 之间也使用了传统的注意力机制, 来提升模型特征抽取的效果。

BERT 得益于 Transformer 强大的特征抽取能力, 能够有效地提取深度语义信息。同时, BERT 所提出的遮罩 (Mask) 是预训练任务, 使得该模型能够做到上下文信息的深度融合。引入了 Mask 的预训练任务在训练时随机地使用 [MASK] 替换输入单词, 其任务目标是输出原始输入序列中被随机替换的词, 通过这种操作避免了双向结构在传统语言模型任务中会从反向输入中获取待预测词的问题。同时为了避免在后期微调阶段出现预训练阶段没有出现过的词, BERT 在预训练阶段的

Mask 操作并不会替换所有词，而是将其中 10% 的词替换成语料中的其他词；另外 10% 的词保留；80% 的词替换成[MASK]。除此之外，BERT 还使用了预测句子对是否连续作为预训练任务。具体操作为：输入句子对进行训练，随机选取语料中的句子替换句子对中的句子，使其不具备上下文联系，使用模型最后一层首位置输出的分类标签（[CLS]）来判断输入句子对是否具备上下文联系。通过这一任务，BERT 模型可以学习到有句子级别的语义信息。

2.1.3 图卷积网络

近几年来卷积神经网络凭借其强大的特征提取能力在图像、语音以及自然语言处理等领域取得了成功^{[36][37][38]}。上述这些数据类型都属于欧氏空间数据，这些类型的数据具有平移不变性^[39]，只有拥有这种特性的数据才能在输入数据空间定义全局共享的卷积核，因此，这种平移不变性是传统卷积操作能够实现的必要条件。但是，除了图像、语音以及自然语言这些欧式空间数据以外，我们生活中也存在着大量的非欧氏空间数据，这些数据大多以图或者网络的形式存在，例如社交网络、交通路网以及知识图谱等。在图数据中欧氏空间数据所具备的平移不变性不再存在，因此传统的卷积核无法使用，同时现实生活中的图数据规模较大，这也给图数据的处理和分析带来了极大的挑战^[40]。

随着图数据规模不断扩大，对有效处理及分析图数据的需求越来越强，研究人员们不断尝试将卷积神经网络应用到图数据的处理和分析之中，开发出一种具有通用性的网络模型。Bruna 等^[41]基于图谱理论从卷积定理出发，首先提出了图卷积网络，该模型中的图卷积操作是定义在谱空间，为后续图卷积领域的谱方法奠基；Defferrard 等^[42]在传统卷积神经网络的基础上引入切比雪夫多项式，通过这种方式学习得到自由参数来近似表示谱方法中的滤波器来学习图上的局部、固定以及组合特征；Kipf 等^[43]对谱方法中的卷积核进行参数化，简化了传统的图卷积网络，大大提升了模型的训练效率和精度。本节将详细介绍本文在实验中所使用的图卷积网络（Graph Convolutional Network, GCN）模型。

GCN 是通过一个谱图卷积的局部一阶近似得到的图卷积结构，该模型的复杂度与图中边的规模呈线性相关，所学习到的隐藏层可以对图的节点

和局部图结构同时进行编码。该方法由谱图卷积的一阶近似启发，提出了一个直接应用在无向图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 上的神经网络 $f(X, A)$ ，其中 $f(\cdot)$ 是一个神经网络之类的可微分函数； X 是节点特征向量 X_i 组成的矩阵； $A \in \mathbb{R}^{N \times N}$ 表示邻接矩阵。同时，该方法还提出了一个在多层图卷积中简单且高效的神经网络层传播规则，形式如式（2-10）所示：

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (2-10)$$

其中, $\tilde{A} = A + I_N$ 是无向图 \mathcal{G} 添加自连接之后的邻接矩阵; I_N 是单位阵; $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$; $W^{(l)}$ 是一个针对层训练的权重矩阵; $\sigma(\cdot)$ 表示一个激活函数, 例如 $ReLU(\cdot) = \max(0, \cdot)$; $H^{(l)} \in \mathbb{R}^{N \times D}$ 是第 l 层的激活矩阵; 初始激活矩阵为输入信号特征 X , 即 $H^{(0)} = X$ 。接下来将详细说明这规则是如何由图结构上的一阶局部谱滤波器启发推导得到。

首先, 定义图上的谱图卷积为滤波器 g_θ 和输入信号 $x \in \mathbb{R}^N$ 的乘积, 其中滤波器 $g_\theta = \text{diag}(\theta)$, 由参数 $\theta \in \mathbb{R}^N$ 在傅里叶域表示。图卷积的定义如式 (2-11) 所示:

$$g_\theta * x = U g_\theta U^T x \quad (2-11)$$

其中, U 是归一化的拉普拉斯矩阵 $L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \Lambda U^T$ 的特征向量组成的矩阵, Λ 是拉普拉斯矩阵的特征值组成的对角阵, $U^T x$ 表示输入信号 x 的图傅里叶变换。可以认为 g_θ 是关于拉普拉斯矩阵的特征值的函数, 即 $g_\theta(\Lambda)$ 。对于规模较大的图来说, 对 L 进行特征值分解是一件计算量非常大的事。通过使用 K 阶切比雪夫多项式 $T_k(x)$ 来近似表示 $g_\theta(\Lambda)$, 能够在降低计算复杂度的同时保证结果的准确性, 切比雪夫多项式的定义以及近似后的谱图卷积形式如式 (2-12) 和 (2-13) 所示:

$$T_k(x) = \begin{cases} 1 & k = 0 \\ x & k = 1 \\ 2xT_{k-1}(x) - T_{k-2}(x) & k > 1 \end{cases} \quad (2-12)$$

$$g_\theta' \approx \sum_{k=0}^K \theta'_k T_k(\tilde{\Lambda}) \quad (2-13)$$

其中, $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - I_N$, λ_{\max} 表示 L 的最大特征值, $\theta' \in \mathbb{R}^K$ 是切比雪夫系数向量。综合式 (2-11) 和 (2-13) 可得图卷积的定义如式 (2-14) 所示:

$$g_\theta' * x \approx \sum_{k=0}^K \theta'_k T_k(\tilde{L})x \quad (2-14)$$

其中, $\tilde{L} = 2L/\lambda_{\max} - I_N$ 。式 (2-14) 是拉普拉斯矩阵的 K 阶多项式, 也就说明是从中心节点向外最多走 K 步得到的 K 阶邻居, 其时间复杂度为 $O(|\mathcal{E}|)$, 与局部图结构中边的规模呈线性相关。基于图卷积的神经网络模型可以通过堆叠多个式 (2-14) 所示的卷积层来实现。现在假设 $\lambda_{\max} = 2$ 且 $K = 1$, 式 (2-14) 也就是对 L 线性相关的函数, 这样就可以摆脱切比雪夫多项式带来的过多参数, 简化后的图卷积定义如式 (2-15) 所示:

$$g_\theta' * x \approx \theta'_0 x + \theta'_1 (L - I_N)x = \theta'_0 x - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \quad (2-15)$$

其中两个滤波器参数 θ'_0 和 θ'_1 可以在整个图上共享。通过限制参数的数量可以进一步的解决过拟合的问题, 并且最小化每层的操作数量, 得到如下表达式:

$$g_\theta * x \approx \theta (I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}})x \quad (2-16)$$

此时式 (2-16) 中只包含一个参数 $\theta = \theta'_0 = -\theta'_1$ ，且 $I_N + D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 的特征值在 $[0, 2]$ 之间，在深层模型中重复这个操作会导致梯度消失、梯度爆炸以及数值不稳定等问题。为了减轻这一问题，引入了如下再正则化的技巧： $I_N + D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \rightarrow \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ ， $\tilde{A} = A + I_N$ ， $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ 。将此定义泛化到一个有着 C 个通道的信号 $X \in \mathbb{R}^{N \times C}$ 上，其中 N 为节点个数， C 为节点特征向量维度。对于 F 滤波器的图卷积定义如下：

$$Z = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X\Theta \quad (2-17)$$

其中 $\Theta \in \mathbb{R}^{C \times F}$ 是一个滤波器的参数矩阵， $Z \in \mathbb{R}^{N \times F}$ 是经卷积后的信号矩阵，由于 $\tilde{A}X$ 可以通过一个稀疏矩阵和稠密矩阵的乘积实现，此卷积操作的时间复杂度为 $O(|\mathcal{E}|FC)$ 。

GCN 作为卷积神经网络在非欧氏空间数据上的扩展，能够端到端地学习图中所有节点的特征表示，同时可以学习到数据整体的结构信息以及节点之间的关联关系。在自然语言处理领域，图结构的引入可以简化复杂语义关系的表示，相比传统的自然语言处理序列建模方式，GCN 能够挖掘出更丰富的复杂语义关系。

2.2 关系抽取数据集

本节将对本文实验部分使用到的两个公开数据集进行详细介绍，包括 ACE 2005 关系抽取数据集和 SemEval 2010 Task 8 关系抽取数据集。上述两个数据集均为关系抽取领域常用的公开数据集，当前关系抽取领域的最新研究成果也在这两个数据集上展开实验，本文将在这两个数据集上与目前领域内先进的方法进行对比实验，验证所提出模型的效果。

2.2.1 ACE 2005 数据集

ACE 2005 数据集是由语言资源联盟 (Linguistic Data Consortium, LDC) 收集、标注、整理得到，于 2005 年在自动内容抽取评测 (Automatic Content Extraction, ACE) 发布的实体关系抽取数据集，其语言涵盖英语、中文和阿拉伯语，本文实验部分使用的是其中的英语数据集，也是领域内广泛使用的数据集之一。该数据集使用包含“其他”标签在内的 7 种关系类型标注了 10573 条关系实例。该数据集虽然关系类型有限，但其关系实例的数据来源较广、长度较长且分布不均匀，这些因素给基于该数据集的关系抽取任务带来了不小的挑战。该数据集中各种关系类型的样本数量统计如表 1 所示。

表 1 ACE 2005 数据集中各类标签样本数量

Table 1 Number of samples of each label in the ACE 2005

关系类型	样本数量
ORG-AFF	3084
PHYS	2313
PART-WHOLE	1542
PER-SOC	1322
GEN-AFF	1211
ART	1101
总计	10573

2.2.2 SemEval 2010 数据集

SemEval 2010 数据集是关系抽取领域使用频率最高的数据集之一，该数据集由语义测评大赛（Semantic Evaluation, SemEval）在 2010 年发布于其大赛的第八项任务：Multi-Way Classification of Semantic Relations Between Pairs of Nominals。该数据集使用包括“其他”标签在内的 9 种关系类型标注了 10717 条关系实例，其中包含 8000 条训练样本，2717 条测试样本。在该数据集中，除了表示目标实体对之间关系不属于标注范围的“其他”类型以外，每种关系类型都是需要考虑头尾实体顺序的，也就是说每种关系都是有向的。例如，头实体与尾实体之间存在 Cause-Effect 关系与，尾实体与头实体之间存在 Cause-Effect 关系，在该数据集中为两种关系类型，分别表示为 Cause-Effect(e_1, e_2) 和 Cause-Effect(e_2, e_1)。因此，该数据集在关系抽取过程中实际包含 19 种关系类型。该数据集的标注质量高，关系实例的长度较为适中且集中，关系类型全面，同时具有关系方向判断这一挑战，是关系抽取领域最常用的公开数据集之一。该数据集中各种关系类型的样本数量统计如表 2 所示。

2.3 评价标准

在关系抽取领域，针对不同类型的数据集设计多种评测指标，常用的指标包括：准确率、召回率、F1 值、ROC 曲线以及 P-R 曲线等^[3]。其中准确率、召回率以及 F1 值根据计算方式的不同还可以细分为宏观和微观两种评价指标。P-R 曲线是针对基于远程监督的关系抽取模型的评价指标。本文所使用的两个关系抽取领域公开数据集都规定了所适用的评价指标，其中 ACE 2005 数据集采用微观 F1 值作为

评测指标；SemEval 2010 Task 8 采用宏观 F1 值作为评测指标，本文也遵循数据集的要求，使用上述两种评测指标衡量所提出模型的性能。

表 2 SemEval 2010 数据集中各类标签样本数量

Table 2 Number of samples of each label in the SemEval 2010

关系类别	样本数量
Cause-Effect	1331
Instrument-Agency	660
Product-Producer	948
Content-Container	732
Entity-Origin	974
Entity-Destination	1137
Component-Whole	1253
Member-Collection	923
Message-Topic	895
Others	1864
总计	10717

对分类任务而言，准确率、召回率和 F1 值是最为常见的评测指标。对于二分类问题而言，我们可以根据样例的真实标签和分类器预测得到的标签将样本分为真正例（True Positive, TP）、假正例（False Positive, FP）、真负例（True Negative, TN）、假负例（False Negative, FN）这四种类型。其中真正例表示真实标签为正且预测标签同样为正的样本，假正例表示真实标签为负但预测标签为正的样本，真负例和假负例以此类推。显然样本总数为上述四类样本之和，分类器的准确率（Precision, P）、召回率（Recall, R）由式（2-18）、（2-19）定义：

$$P = \frac{TP}{TP + FP} \quad (2-18)$$

$$R = \frac{TP}{TP + FN} \quad (2-19)$$

准确率表示在分类器预测为正例的样本中，真实标签也为真的样本所占比例；召回率表示被分类器预测为正样本的样例占有所有真实样本的比例。准确率和召回率指标通常难以兼顾，F1 值是综合考虑准确率和召回率的测评指标，由式（2-20）定义：

$$F1 = \frac{2 \times P \times R}{P + R} \quad (2-20)$$

对于多分类问题而言, 准确率、召回率和 F1 值根据计算方式的不同可以细分为宏观和微观两种评价指标。宏观的计算方法分别计算各个类的准确率、召回率和 F1 值, 再分别取平均得到宏观准确率 (marco-P)、宏观召回率 (marco-R) 和宏观 F1 值 (macro-F1)。而微观的计算方法是先分别对各类别的 TP、FP、TN 和 FN 计算平均值, 再基于这些平均值计算微观准确率 (micro-P)、微观召回率 (micro-R) 和微观 F1 值 (micro-F1)。本文在 ACE 2005 数据集上使用微观 F1 值进行评价, 再 SemEval 2010 Task 8 数据集上使用微观准确率作为评测指标。

2.4 研究现状

纵观近几年实体关系抽取领域的技术发展, 除了少数几种无监督的关系抽取方法^{[44][45]}外, 大多数现有的关系抽取方法都将此任务作为有监督的多分类任务处理。接下来将在本章中对现有关系抽取方法进行分类介绍。

2.4.1 基于特征工程的关系抽取方法

早期的实体关系抽取方法可以分为基于特征工程的方法和基于核函数的方法。其中, 基于特征工程的关系抽取方法依赖于由语言学预处理步骤得到的显式特征。Zhou 等^[46]通过结多种词汇、句法和语义知识构建 SVM 的输入特征, 并使用 WordNet 和 Name List 等语义信息进一步提升模型性能; Suchanek 等^[47]为了验证深层语法结构在关系抽取中的有效性, 通过分析目标实体对之间潜在关系的链接语法模式, 并基于潜在关系模式构建 kNN 分类器的输入特征; Jiang 等^[48]更深入地挖掘了多种语义特征, 并通过实验表明关键的单一的特征比冗余的复杂特征更为有效; Nguyen 和 Grishman^[49]探索了多种词嵌入在关系抽取任务中的应用方法, 通过结合词嵌入和词簇来捕获不同的文本信息, 并将两者组合进一步提高了关系抽取模型的适应性。基于特征工程的关系抽取方法虽然深度挖掘了关系实例中包含的语义信息, 但是对于目标实体对的上下文结构信息的提取不够充分。基于核函数的关系抽取方法, 不需要专家经验构造显式特征, 直接以文本字符串或者句法解析树的形式输入, 能够很好地挖掘目标实体对的上下文结构信息。Culotta 和 Sorensen^[50]使用依存关系分析树作为 SVM 的输入特征, 通过添加词性、实体类型等语义特征, 提升关系抽取的效果; Bunescu 等^[51]通过比较最短依存路径上相同节点的个数, 构建最短依存路径核函数作为分类器输入, 取得了更好的效果; Qiao 等^[52]基于卷积核函数改进得到动态依存关系树来表征关系实例, 综合考虑了目标实体对的上下文信息, 有效提升了模型性能。早期的关系抽取方法依赖专家经验人工设计的特征

或由预训练的 NLP 工具及其他外部资源构建的核函数。尽管这些特征或者核函数可以有效地提取关系实例中的语义和机构特征，但是外部 NLP 工具和资源包含潜在误差，这些误差将影响关系抽取模型的性能；同时缺乏对于新增关系实例的鲁棒性，以及对于其他领域和形式文本语料的扩展性。

2.4.2 基于神经网络的关系抽取方法

近几年关系抽取领域的研究表明，神经网络模型在关系抽取任务中效果拔群。与早期基于特征工程以及核方法的模型相比，基于神经网络的关系抽取模型可以自动学习潜在的语义特征，减少了对语言学文本预处理技术的依赖，大大削弱了外部 NLP 工具带来的负面影响。Socher 等^[53]使用 RNN 来学习句法分析树中实体之间路径的表示，并将该表示作为分类器输入进行关系分类；Zeng 等^[13]首先将 CNN 应用于关系抽取任务中，将人工设计的特征作为输入进行关系编码；Santos 等^[14]提出的 Rank-CNN 模型可以减少对人工设计特征的依赖，并通过设计新的目标函数极大地提高了模型性能；Miwa 和 Bansal 等^[16]发现，基于长短期记忆网络 (LSTM) 的模型可以捕获长单词序列的语言和句法特征，其性能要优于基于 CNN 的模型。随着注意力机制在自然语言领域的广泛应用，学者们也将其应用于关系抽取任务之中，来捕获关系实例中最有价值的信息。Zhou 等^[54]和 Wang 等^[55]分别提出了结合注意力机制的 RNN 和 CNN 的关系抽取模型，在各自数据集上均取得了当时最优的效果；Ji 等^[17]通过引入 Freebase 和 Wikipedia 的实体描述信息作为背景知识补充，并提出了句子级别的注意力机制模型来选取有效的关系实例，取得了不错的效果；Nguyen 和 Grishman 等^[18]通过联合使用 RNN 和 CNN，并借鉴传统特征工程方法构建输入特征，同时利用三种方法的优势。融合依存树的神经网络模型也在关系抽取领域取得了显著的成果，这类模型通过捕获长距离关系来提升模型效果。Xu 等^[15]通过在实体之间的最短依赖路径上应用 LSTM 实现两种方法的融合，对长距离关系实例的关系抽取准确率得到明显提升；Miwa 和 Bansal^[16]在实体关系联合抽取任务中使用融合了依存树的 LSTM 模型，相比其他联合抽取方法有较大提升。

近几年图神经网络模型^{[56][57]}开始被应用于关系抽取任务，并取得了显著的效果。Zhang 等^[58]首次将图卷积网络 (GCN) 应用于关系抽取，其模型可以有效地在任意依赖结构上并行的抽取信息。Christopoulou 等^[59]提出了一种基于图的神经网络模型，该模型可以同时处理句子中的多对实体，综合考虑实体对之间的关系，为目标实体对的关系抽取提供更丰富的上下文依据。

上述基于神经网络的关系抽取方法通过预训练得到的词向量构建神经网络的输入特征，利用神经网络为目标实体对自动抽取关系实例中上下文信息的深层语

义信息与结构特征用于关系分类,在大大减少了前期的数据预处理工作,弱化了需要专家经验的特征设计工作的同时,还取得了与传统关系抽取方法相当,甚至更好的结果,其模型的鲁棒性和通用性也更强。但是这些关系抽取方法仍然存在一些不足之处,例如,使用外部 NLP 工具进行预处理的模型存在着外部工具误差在网络中传播和累计的风险;利用外部背景知识和资源的模型又牺牲了模型对于新增关系实例的鲁棒性和复用性。当前的关系抽取模型大多是以目标实体作为对象,单独研究,忽略了实体与实体之间蕴含的关联关系。

针对上述问题,本文从语料集的层次出发,通过构建实体关联图和实体对图并提取其中的拓扑关系,对整个语料集中的实体关联关系进行充分挖掘。本文提出了基于实体关联图与注意力机制的关系抽取方法和基于实体对图与多层图卷积的关系抽取方法,充分挖掘了实体关联关系图中的拓扑特征,并结合由预训练语言模型生成的文本语义特征,完成关系抽取任务。

2.5 本章小结

本章为关系抽取研究综述部分,简要介绍了与本文研究相关的理论技术、数据集、评价标准和技术现状。首先介绍了本文所提出的关系抽取模型涉及技术的相关理论基础,包含预训练语言模型、注意力机制和图卷积网络;然后展开介绍了本文实验部分使用的关系抽取领域常用的两个公开数据集以及评价关系抽取模型性能的技术指标;最后对关系抽取领域的技术发展和现状进行阐述。

3 基于实体关联图与注意力机制的关系抽取

本章将详细介绍基于实体关联图与注意力机制的关系抽取方法。该关系抽取模型由三个主要模块组成,包括实体关联关系特征抽取模块、文本语义特征抽取模块以及关系分类模块。本章将从本方法提出的思路以及具体的模型结构两个方面展开阐述,通过给出方法核心概念实体关联图的定义阐明本方法的思路,再对模型的整体框架进行介绍,接着对其中的各个模块分别展开介绍。为了证明该关系抽取方法的有效性,将对对比实验设置以及实验结果进行详细阐述和分析。

3.1 实体关联图定义

关系抽取领域现有的方法大多将一个关系实例中的目标实体对作为独立对象处理,仅从其对应的关系实例中挖掘目标实体对之间的语言关系,这种只依赖关系实例文本信息的方式存在许多不足之处,诸如,对长距离关系抽取效果不好,单一关系实例中包含的信息不足等问题。由于仅从关系实例中提取的特征无法满足高精度关系抽取的需求,研究人员们开始利用自然语言处理工具提取的结构特征和实体背景知识对上下文信息进行补充。例如使用依存句法分析工具获取关系实例中的句法分析树作为输入辅助关系抽取,或者直接使用目标实体类别以及实体描述信息作为补充。这些方法固然能在一定程度上提升关系抽取模型的效果,但是仍然存在一些问题,如自然语言处理工具带来的误差会在网络中传播并累积,对于许多目标实体对外部信息可能难以获取。综合考虑现在关系抽取方法存在的不足,本文更加深入挖掘数据集中天然包含的实体关联信息,用于提升关系抽取模型效果。

实体关系抽取任务的目标是通过上下文进行分析,判断目标实体对之间的关系,并根据预设的关系类型空间对潜在关系进行分类。实际上,此处的上下文信息不仅仅包含关系实例中的文本信息,还包括目标实体对在整个语料集层次包含的关联关系,在语料充分的情况下,目标实体大概率出现在多个关系实例中,且与多个实体存在关联关系。这种实体之间的关联关系从关系抽取任务的角度来看,代表了两个实体共享同一关系实例的上下文信息,这种实体关联关系可以在一定程度上丰富目标实体对的上下文信息,辅助其完成关系抽取。明确了语料级别的实体关联关系对于关系抽取任务具有补充信息的作用之后,进一步思考如何对这种实体关联关系进行表示,从而应用到关系抽取模型之中。本文采用图结构来表示上述实体关联关系。具体来说,对于实体对图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, 图中节点 \mathcal{V} 表示语料集中的实体,图中的边 \mathcal{E} 实体与实体之间的关联关系。在实体关联图中,目标实体对应节点

的邻居节点就是与目标实体包含关联关系的实体，通过学习实体关联图中节点表示，并综合考虑目标实体与其关联实体就能够将语料级别的实体关联关系应用到关系抽取模型之中，达到深度挖掘数据集自身特征来提升关系抽取模型效果的目的。

3.2 模型总体框架

ERANN 模型的总体框架如图 3-1 所示。文本语义特征抽取模块基于 Word2Vec 对输入关系实例进行嵌入表示，得到词向量表示并在其中加入位置嵌入表示，考虑到目标实体在关系抽取任务中的重要性，使用结合自注意力机制的双向门控循环神经网络对关系实例中的上下文语义特征进行提取，完成文本语义特征抽取；而在实体关联特征抽取模块中，通过统计语料集中实体共现关系构建实体关联关系图，使用文本语义特征抽取模块中学习得到的实体向量表示作为图中节点的表示；使用注意力机制计算目标实体与其所有关联实体之间的注意力分数后进行带权融合，从而做到更高效地利用关联实体信息，完成实体关联特征抽取；将文本语义特征与实体关联特征进行融合，输入最后的关系分类模块，完成关系抽取任务。

3.3 文本语义特征抽取

本节将对 ERANN 模型中的文本语义特征抽取模块进行详细介绍，其中包括，基于 Word2Vec 的词嵌入表示以及位置嵌入表示方法；以及基于双向门控循环神经网络和自注意力机制的语义特征构建方法。

3.3.1 词向量表示

目前主流的神经网络关系抽取方法中一般采用词向量的形式来对关系实例的上下文语义特征进行嵌入表示，这种表示方法相较于传统的基于词法或者句法分析器的特征抽取方法，对于关系实例中的深层语义信息的挖掘更充分，同时对于各种语料的扩展性也更强。因此，在 ERANN 模型中，我们采用被广泛使用的词嵌入表示模型 Word2Vec 对语料集中的词向量进行嵌入表示。

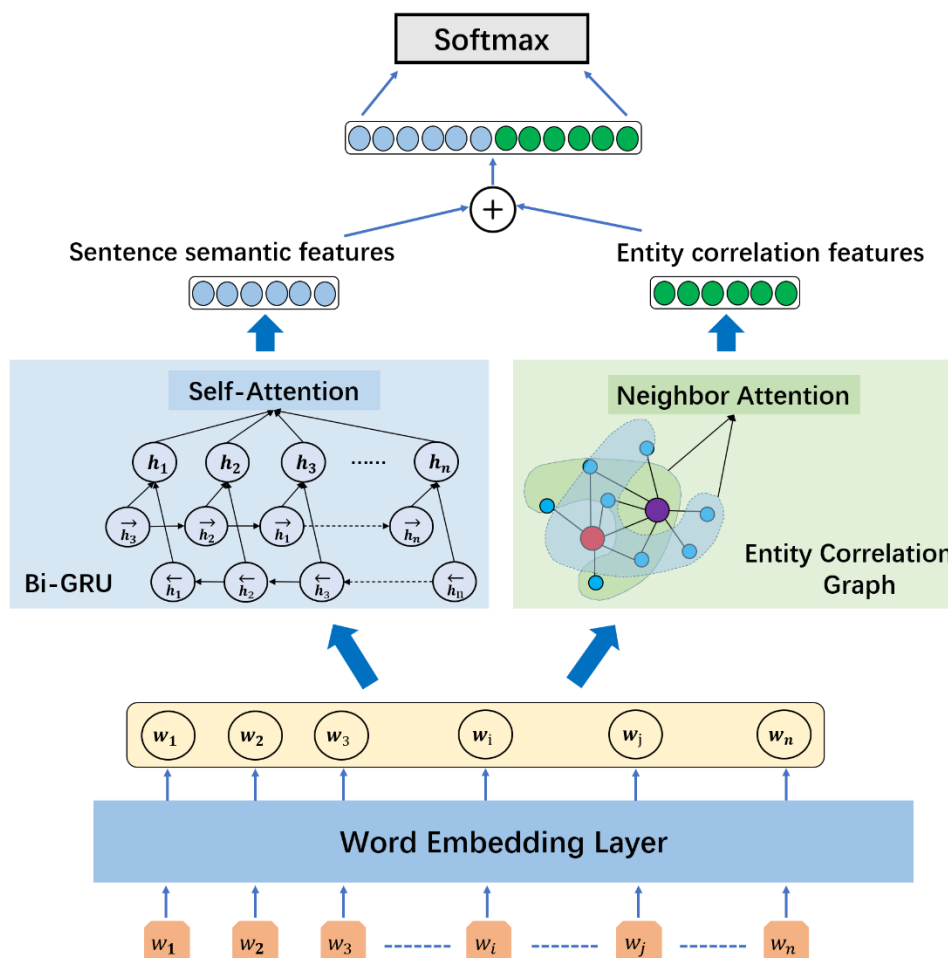


图 3-1 ERANN 模型框架图

Figure 3-1 Overall architecture of the ERANN model

Word2Vec 的训练目标是语言模型，词向量是隐层的输出特征。Word2Vec 由两种不同的语言模型训练目标：连续词袋模型（Continuous Bag of Words, CBOW）以及跳字模型（Skip-gram）。CBOW 的任务设置是通过滑动窗口内上下文来预测中心的概率；Skip-gram 则与之相反，是通过中心词来预测滑动窗口内上下文的概率。虽然 Word2Vec 模型将上下文内容限制在滑动窗口范围内，但是当训练语料的规模较大时，其中条件概率的计算难题仍然是限制模型广泛应用的重要约束。因此，Word2Vec 提出了两种简化模型求解的方法：层次分类方法（Hierarchical Softmax）和负采样（Negative Sampling）。其中，Hierarchical Softmax 采用构建 Huffman 树的形式简化模型的条件概率计算过程；而 Negative Sampling 则使用的是随机负采样的机制来提升训练速度，这两种简化策略均在很大程度上提升了模型的效率。ERANN 模型使用的是基于 Skip-gram 和 Hierarchical Softmax 的 Word2Vec 模型。

对于关系实例序列 $S = \{w_1, w_2, \dots, w_r\}$ ，基于 Skip-gram 的 Word2Vec 模型的任务是最大化给定中心词下窗口内上下文的概率，具体形式如式（3-1）所示：

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t) \quad (3-1)$$

其中, c 为滑动窗口大小。

Hierarchical Softmax 方法以各个词在语料中的词频作为权值构建 Huffman 树, 该 Huffman 树的叶子节点表示词, 对于语料中的任意词 w , Huffman 树中必存在唯一一条从根节点到代表 w 的叶子节点的长度为 l^w 的路径, 该路径上存在 $l^w - 1$ 个分支, 将每个分支看作一个二分类, 所有分支概率连乘的结果即代表了 w 的概率。在 Skip-gram 任务中具体来说, 假定中心词为 w , 上下文为 $Context(w)$, 上下文种任意词 $u \in Context(w)$, 则其求解目标可由式 (3-2)、(3-3) 和 (3-4) 计算得到:

$$p(Context(w) | w) = \prod_{u \in Context(w)} p(u | w) \quad (3-2)$$

$$p(u | w) = \prod_{j=1}^{l^u} p(d_j^u | v(w), \theta_{j-1}^u) \quad (3-3)$$

$$p(d_j^u | v(w), \theta_{j-1}^u) = [\sigma(v(w)^T \theta_{j-1}^u)]^{1-d_j^u} \cdot [1 - \sigma(v(w)^T \theta_{j-1}^u)]^{d_j^u} \quad (3-4)$$

其中 $d_j^u \in \{0, 1\}$ 表示词 u 在树中对应路径的第 j 个节点对应的编码值, θ_{j-1}^u 代表词 u 在树中对应路径的第 $j-1$ 个节点的向量表示, $v(w)$ 表示中心词 w 的词向量表示。

使用上述模型对语料集进行训练之后, 得到语料中所有词的词向量矩阵 $L \in \mathbb{R}^{d \times |V|}$, 其中 d 为预设的词向量维数, $|V|$ 为语料构建的词典大小。词向量矩阵中的每一列为词表中的一个词对应的向量表示 $w^d \in \mathbb{R}^d$ 。同时, 为了突出目标实体对在关系抽取任务中的重要性, ERANN 在上述词向量的基础上引入了位置嵌入表示, 来补充关系实例中的词与目标实体之间的相对距离。具体做法为, 将关系实例中的每个词与首尾实体之间的相对位置映射成向量 $w_h^p \in \mathbb{R}^{d_p}$ 和 $w_t^p \in \mathbb{R}^{d_p}$, 作为位置嵌入表示, 与词向量表示进行连结得到 $w = [w^d; w_h^p; w_t^p]$, $w \in \mathbb{R}^{d_w}$, $d_w = d + 2d_p$, 作为该词在关系实例中的表示。在后续文本语义特征构建模块, ERANN 模型使用添加位置嵌入表示后的关系实例中词的向量表示构建得到关系实例向量矩阵。

3.3.2 文本语义特征构建

关系抽取任务与一般的文本分类任务不同, 其目标不是对单个句子分类, 而是对句子这个关系实例中包含的目标实体对分类, 句子中每个词对于关系抽取任务的重要程度都不相同, 显然目标实体应该比其他的词更加重要, 因此不能单纯使用 3.3.1 小节中计算得到的关系实例向量矩阵作为关系分类特征。为了更深入地挖掘关系实例中的语义关系, 同时对关系实例中的每个词赋予合理的权重, 本文在 ERANN 模型中使用双向门控循环单元提取关系实例中的深层语义特征, 同时在关系实例内部使用自注意力机制计算其中每个词的影响力权重, 从而完成文本语义

特征构建工作。

对于输入关系实例序列 $S = \{w_1, w_2, \dots, w_T\}$ 其中的第 i 个词 w_i ，在添加位置嵌入表示之后， w_i 在该关系实例中的向量表示为 $w_i = [w_i^d; w_i^p; w_i^p]$ 。则关系实例序列 S 可以表示为向量矩阵 $S \in \mathbb{R}^{T \times d_w}$ ，作为输入特征输入 Bi-GRU，那么词 w_i 对应的隐藏单元输出可以表示为式 (3-5) 所示形式：

$$\begin{aligned}\vec{h}_i &= \overrightarrow{GRU}(w_i, \vec{h}_{i-1}) \\ \overleftarrow{h}_i &= \overleftarrow{GRU}(w_i, \overleftarrow{h}_{i-1})\end{aligned}\quad (3-5)$$

将 Bi-GRU 的隐层输出作为新的关系实例表示，如式 (3-6) 所示：

$$\begin{aligned}H &= (h_1, h_2, \dots, h_T) \\ h_i &= [\vec{h}_i; \overleftarrow{h}_i]\end{aligned}\quad (3-6)$$

在新的关系实例表示基础上，使用自注意力机制计算其内部每个词之间的权重分配情况，关系实例的内部权重分配矩阵 A 如式 (3-7) 所示：

$$A = \text{softmax}(W_2 \tanh(W_1 H^T)) \quad (3-7)$$

其中， $W_1 \in \mathbb{R}^{d_a \times 2u}$ 和 $W_2 \in \mathbb{R}^{t \times d_a}$ 均为权重矩阵， t 代表通过自注意力机制得到的向量维度， u 为 Bi-GRU 隐层神经元数量， d_a 为自注意力机制中间结果矩阵维度，后三者均为模型中的超参数。

最后，ERANN 模型对 Bi-GRU 隐层输入与权重分配矩阵相乘的结果使用最大池化操作，得到该关系实例的文本语义特征 $f^s \in \mathbb{R}^{d_s}$ ，如式 (3-8)，(3-9) 所示：

$$M = AH \quad (3-8)$$

$$f^s = \max M(i, \cdot) \quad 0 \leq i \leq t \quad (3-9)$$

其中， $d_s = 2u$ ， $M(i, \cdot)$ 表示矩阵 M 第 i 行的最大值。

3.4 实体关联特征抽取

本节将对 ERANN 模型中的实体关联特征抽取模块进行详细介绍，实体关联特征抽取模块是 ERANN 模型最重要的模块，其中，实体关联图的提出和构建和从图中提取目标实体关联特征的方法是本模型的主要贡献。实体关联图中的节点嵌入表示方式是本节的后续工作的基础。接下来介绍在完成实体关联图节点表示之后，如何基于邻居节点的注意力机制，抽取目标节点的拓扑特征，完成 ERANN 模型的实体关联特征抽取工作。

3.4.1 实体关联图节点嵌入表示

ERANN 模型构建了实体关联图，并将目标实体在图中对应节点的拓扑特征应

用到关系抽取的特征构建工作中, 通过这种方式将目标实体在语料级层次的实体关联关系融合到关系抽取模型之中, 提升模型的效果。无论是目标实体本身的信息, 还是其关联实体的信息, 在实体关联图中都是以节点嵌入表示的形式进行传递的。因此, 实体关联图中的节点嵌入表示是否合适, 将很大程度上影响模型的表现。

实体关联图中节点代表一个实体, 若两个实体之间有边相连, 则说明实体之间存在潜在关联关系。因此, 实体关联图中节点的表示应该能够表征其对应的实体的特征。鉴于这一特点, 在 ERANN 模型中我们使用 3.3.1 小节中训练得到的实体对应的词向量来构建实体关联图的节点嵌入表示, 由于此时的实体独立于关系实例, 不需要添加位置嵌入表示。考虑到直接使用节点对应实体的词向量表示会与语义特征提取部分存在重复, 我们使用实体向量表示进行非线性变换之后得到的向量表示作为实体关联图中节点的嵌入表示。具体地, 对于一个实体 e , 其在实体关联图中对应节点的嵌入表示 f_e^N 具体形式如式 (3-10):

$$f_e^N = \sigma(W_3[w_e^d]^T + b_3) \quad (3-10)$$

其中, $W_3 \in \mathbb{R}^{d_N \times d}$ 为网络权重矩阵, $b_3 \in \mathbb{R}^{d_N}$ 为偏置, $\sigma(\cdot)$ 为激活函数。

3.4.2 邻居节点注意力机制

ERANN 模型的出发点是将目标实体在语料集层次包含的关联实体信息应用到关系抽取任务中。在语料足够丰富的情况下, 一个实体可能与多个实体之间存在关联关系, 本方法通过构建实体关联图的形式表示这种实体关联关系。在实体关联图中, 一个实体的所有邻居实体都是其关联实体, 关联实体的信息可以在目标实体的关系抽取中起到补充作用。然而, 并不是每一个关联实体对于当前关系实例中的关系抽取起到的辅助作用都是相同的, 目标实体与其所有关联实体之间的关系权重各不相同。因此, 我们使用注意力机制来计算各个关联实体与目标实体之间的关系权重, 从而更有效地利用关联实体信息。

具体地, 在实体关联图中, 对于目标实体对应节点 e , 其邻居节点的集合可以表示为 $\Gamma_e = \{c_1, c_2, \dots, c_{|\Gamma_e|}\}$, 则节点 e 的邻居节点特征 f_e^c 可由式 (3-11) 计算得到:

$$f_e^c = \sum_{i=1}^{|\Gamma_e|} g_i^e \times f_i^N \quad (3-11)$$

其中 g_i^e 为目标实体对应节点 e 与邻居节点 c_i 之间的权重, f_i^N 为邻居节点 c_i 的节点表示。ERANN 模型使用注意力机制计算目标节点与其每个邻居节点之间的关联程度值, 计算公式如式 (3-12) 所示:

$$a_i^e = \tanh(W_4[f_e^N; f_i^N] + b_4) \quad (3-12)$$

其中 f_e^N 和 f_i^N 分别为目标实体 e 和其邻居节点 c_i 的节点表示, $W_4 \in \mathbb{R}^{1 \times 2d_N}$ 为网络权

重矩阵, $\mathbf{b}_4 \in \mathbb{R}^{1 \times 1}$ 为偏置。根据式 (3-12) 计算得到目标节点与其每个邻居节点之间的关联程度值后, 对其进行归一化操作使其所有邻居得关系权重和为 1, 从而得到目标节点与各个邻居节点的关系权重, 其计算公式如式 (3-13) 所示:

$$\mathbf{g}_i^e = \frac{\exp(\mathbf{a}_i^e)}{\sum_{i=1}^{|\Gamma_e|} \exp(\mathbf{a}_i^e)} \quad (3-13)$$

3.5 关系抽取

在完成文本语义特征抽取和实体关联特征抽取之后, ERANN 模型基于文本语义特征 \mathbf{f}^s 和实体关联特征 \mathbf{f}^c , 使用 softmax 分类器对输入关系实例进行关系分类。softmax 分类器由单层全连接网络构成, 并在输出层使用 softmax 归一化操作。对于输入关系实例 S , ERANN 模型将文本语义特征与实体关联特征通过特征向量拼接的形式进行融合, 使用 $\mathbf{f}^s \in \mathbb{R}^{d_s}$ 和 $\mathbf{f}^c \in \mathbb{R}^{d_c}$ 的连结作为 softmax 分类器的输入, 具体形式如式 (3-14) 所示:

$$p(y|S) = \text{softmax}(\mathbf{W}_0[\mathbf{f}^s; \mathbf{f}^c]^T + \mathbf{b}_0) \quad (3-14)$$

其中, $\mathbf{W}_0 \in \mathbb{R}^{r \times (d_s + d_c)}$ 为权重矩阵, $\mathbf{b}_0 \in \mathbb{R}^r$ 为偏置, r 代表预设的关系类型数量, 为模型超参数。

接着 ERANN 模型使用 softmax 分类器的输出对输入关系实例中的目标实体对之间的关系类型进行预测, 如式 (3-15) 所示:

$$\hat{y} = \arg \max_y p(y|S) \quad (3-15)$$

完成上述步骤得到输入关系实例的预测标签之后, softmax 分类器使用交叉熵 (Cross-Entropy) 损失函数, 如式 (3-16) 所示:

$$\mathcal{L} = -\sum_i \sum_j \text{label}_{ij} \log p(y_{ij} | S_i, \theta) \quad (3-16)$$

其中, θ 代表 ERANN 模型中的所有参数的集合, y_{ij} 表示样本 i 的预测标签为 j , label_{ij} 为关于样本 S_i 的 one-hot 向量, 当其真实标签为 j 时 label_{ij} 为 1, 否则为 0。

在 ERANN 模型中, 我们使用 Adam 优化算法求解参数, 并将网络参数使用 Xavier 初始化方法完成初始化。

3.6 实验设置与结果分析

本节将对 ERANN 模型在 ACE 2005 和 SemEval 2010 Task 8 两个数据集的实验进行详细介绍, 首先说明本实验中的参数设置情况, 然后简单介绍作为对比实验方法的各类主流关系抽取方法, 最后对实验结果进行介绍并深入分析。

3.6.1 实验参数设置

ERANN 模型使用 Word2Vec 模型对关系实例以及目标实体的编码表示, 通过构建实体关联图的形式将语料集层次的实体关联关系以图结构进行表示, 并使用注意力机制将目标实体和其关联实体信息进行融合用于关系抽取任务。在 ERANN 模型训练阶段, 我们使用 Xavier 初始化方法初始化网络参数, 使用 Adam 优化算法对模型参数进行求解, 同时为了防止模型出现过拟合, 在实体关联特征抽取部分加入了 dropout 操作。ERANN 模型中包含如下主要超参数, 分别是 Bi-GRU 隐层神经元数量 u ; 实体关联特征向量维数 d_N ; 自注意力机制中间结果矩阵大小 d_a 和焦点数量 t ; batchsize 大小; 模型的学习率 λ ; dropout 率 dp 以及 adam 优化器参数 β_1 、 β_2 和 ϵ 。为了更加科学准确地进行超参数的选择, 本文在实验部分使用 5 折交叉验证的方法进行参数选取, 所得出的最优超参数设置如表 3 所示。

表 3 ERANN 模型超参数设置

Table 3 Hyper-parameter setting of ERANN model

参数	数值
u	100
d_N	64
d_a	100
t	5
batchsize	64
λ	5e-6
dp	0.3
β_1	0.9
β_2	0.999
ϵ	10e-8

3.6.2 对比实验设置

为了验证 ERANN 模型在关系抽取任务上的效果, 我们分别在 ACE 2005 和 SemEval 2010 Task 8 这两个公开数据集上, 与当前各类主流的关系抽取方法进行对比实验。其中, 由于 SemEval 2010 Task 8 数据集免费开放, 人工标注质量更高并且关系实例长度适中且分布均匀, 近些年来很多的主流方法是基于这一数据集提出的。因此, 本文主要在 SemEval 2010 数据集上, 与当前在该数据集上取得较

好结果的方法进行对比试验。ACE 2005 数据集具有关系实例长度更长且分布不均匀等挑战,模型在该数据集上更难以捕获目标实体之间的长距离语义关系。本文在 ACE 2005 数据集上与当前在该数据集上取得较好结果的方法进行对比,从而进一步说明本文在 ERANN 模型中所使用的实体关联特征能够为目标实体的关系抽取补充上下文信息。

本文从特征信息来源、模型网络结构和特征构建方法等方面,将当前主流的关系抽取模型分为两类:首先是不使用任何外部自然语言处理工具和背景知识,仅使用神经网络模型端到端地从关系实例上下文提取特征的端到端的关系抽取方法(End-to-End Methods),这类模型大多使用 CNN 作为主要的网络结构;通过与该类型方法进行对比实验可以验证本文提出的模型从关系实例中抽取特征的能力强弱;第二类是使用句法依存分析器等自然语言处理工具挖掘关系实例中的依存关系,并将之作为特征与关系实例文本语义特征共同用于关系抽取的方法(Dependency Methods),这类模型主要基于 RNN 构建神经网络进行特征抽取,通过与这类模型比较进一步验证本文提出的模型在不使用外部自然语言处理工具和背景知识的前提下能否提取到足够的信息进行关系抽取。具体的对比实验设置如表 4 所示:

表 4 ERANN 对比实验设置

Table 4 Comparison setting of ERANN model

类型	模型	应用数据集
<i>End-to-End Methods</i>	CNN+PF ^[13]	SemEval
	CR-CNN ^[14]	SemEval
	Att-Pooling CNN ^[55]	SemEval
<i>Dependency Methods</i>	SDP-LSTM ^[15]	SemEval
	DRNNs ^[60]	SemEval
	SPTree ^[61]	ACE/SemEval
	walk-based model ^[59]	ACE

3.6.3 实验结果分析

本节将在介绍本文提出的 ERANN 模型在 ACE 2005 和 SemEval 2010 数据集上的对比实验结果,并对实验结果进行全面分析,证明本文提出的 ERANN 模型的有效性,主要对其中的结合 Bi-GRU 和自注意力机制的文本语义特征抽取模块功效,以及实体关联特征的有效性进行验证。

本文提出的 ERANN 模型在 SemEval 2010 数据集上与其他主流关系抽取模型的对比结果如表 5 所示。在该数据集上，我们按照任务要求使用宏观 F1 值作为评价指标，使用宏观 F1 值作为评测指标更注重模型对于所有类别的平均分类能力，在关系类别数不大的情况下，如果有的类别分类效果较差会造成该测评指标的大幅度下降。表 5 中的结果表明，在本文提出的 ERANN 模型在该数据集上的对比实验中取得了次优的结果，这一结果从宏观角度说明 ERANN 模型在处理 SemEval 2010 数据集这类关系实例长度适中，且分布均匀的样本时能够有效提取其中的有效信息准确地进行关系抽取。

表 5 ERANN 在 SemEval 数据集对比实验结果

Table 5 Comparison of ERANN model on SemEval 2010

类型	模型	Macro-F1
<i>End-to-End Methods</i>	CNN+PF	82.7
	CR-CNN	84.1
	Att-Pooling CNN	88.0
	Bi-GRU	83.3
<i>Dependency Methods</i>	SDP-LSTM	83.7
	DRNNs	85.8
	SPTree	84.5
	C-GCN	84.8
<i>Our Methods</i>	ERANN without ER	84.7
	ERANN without SAT	85.1
	ERANN	86.6

在两大类对比实验方法中，取得最优结果的是 Wang 等^[55]提出的 Att-Pooling CNN 模型，该模型基于两种不同层次的注意力机制仅从关系实例中捕获异构上下文特征，其结果要明显优于使用外部自然语言处理工具的方法。这一结果说明这种仅从关系实例中提取特征，不依靠外部自然语言处理资源和先验知识的方法，能够取得很好的关系抽取效果，甚至优于使用了上述资源的方法；同时，该模型依靠两种不同层次注意力机制的网络结构取得如此好的关系抽取效果也体现了注意力机制在关系抽取任务中的有效性；另外，这一结果还证明了本文思路的出发点：目标实体关系实例中包含丰富的语义和结构特征，不依赖外部资源，仅从关系实例中挖掘信息就能够取得很好的关系抽取效果。此外，在基于依存分析器的对比方法中，Xu 等^[60]提出的 DRNNs 模型使用斯坦福依存分析器（Stanford parser）解析关系实

例得到其依存关系,并将其输入深层 RNN 网络进行特征提取,取得了较好的效果,这一结果说明了深度更深的模型能够更有效地捕获关系实例中的语义和结构特征。从总体上看,使用了外部自然语言处理资源的方法其 F1 值要高于简单从关系实例中端到端地抽取特征的 CNN 方法,这一点证明了引入外部资源的有效性。但是,ERANN 模型在不使用任何外部资源的情况下取得了优于上述使用了外部信息的方法,这进一步验证了本文思路的出发点,即语料集自身包含有丰富的语义和结构信息,同时也证明了实体关联关系能够有效提升关系抽取的效果。

为了更好地验证 ERANN 模型中各个模块的作用,我们在对比实验中加入了 ERANN 各模块的孤立实验,从表 5 中结果我们可以看出,仅仅使用简单 Bi-GRU,没有加入句内自注意力机制和实体关联关系的模型(Bi-GRU)结果一般,与基础的引入位置嵌入表示 CNN 模型的结果相当,说明单纯的门控单元无法充分挖掘关系实例中的语义和结构特征。在 Bi-GRU 的基础上引入句内自注意力机制的方法,也就是在 ERANN 模型的基础上删除实体关联特征抽取模块(ERANN without ER),其 F1 值提升了 1.4%,这证明了句内自注意力机制通过给句内不同的词赋予各异的权重,侧重对实体关系抽取更重要的信息的提取,提升了关系抽取模型的效果;此外,在 Bi-GRU 的基础上引入实体关联关系的方法,也就是在 ERANN 模型的基础上删除句内自注意力机制(ERANN without SAT),其 F1 值提升了 1.8%,这一结果一方面证明了本文提出的实体关联特征能够有效提升模型的效果,同时还验证了实体关联特征相比于句内自注意力机制,对于实体关系抽取任务更加有效。

为了进一步验证 ERANN 模型在面对不同类型数据时的表现如何,我们在 ACE 2005 数据集上与两类典型方法进行了对比实验。该数据集相对于 SemEval 2010 数据集,其关系实例来源更为丰富,长度更长且分布不均匀,在该数据集上进行实验能够有效验证模型在处理长距离实体对的能力以及对于不同来源数据的鲁棒性如何。表 6 展示了本文提出的 ERANN 模型与其他三种模型在 ACE 2005 数据集上的对比实验结果。在该数据集上,我们按照任务要求使用微观 F1 值作为评价指标,不同于宏观 F1 值,微观 F1 值更加侧重模型对于每个样本的分类能力。在对比实验设置方面,由于该数据上样本长度较长且长度分布不均匀,基于 CNN 的模型在该数据集上的效果普遍较差,因此没有选取基于 CNN 的方法作对比实验,具体的对比方法有以下两种:Miwa 等^[61]提出的 SPTree 为基于依存分析器的方法;Christopoulou 等提出的 Walk-based Model 在依存关系的基础上引入了同一样本中不同实体对之间的关联关系。

从表 6 的结果可以看出, ERANN 模型在 ACE 2005 数据集上取得了最优的结果,进一步证明 ERANN 模型在关系抽取任务上的高效性以及对于不同来源数据的鲁棒性。Walk-based Model 与 SPTree 方法同样使用句法依存树作为特征,但

是前者在此基础引入了同一样本中不同实体对之间的关联特征，其 F1 值也高出后者 4%，这部分提升主要由实体之间的关联特征带来，验证了不同实体之间的关联关系对于关系抽取任务的促进作用。ERANN 在 Walk-based Model 中使用的特征的基础上更进一步，将实体之间的关联关系从同一样例扩展到整个语料集，信息更加丰富，在相关实验也取得了更好的结果。

表 6 ERANN 在 ACE 数据集对比实验结果

Table 6 Comparison of ERANN model on ACE 2005

模型	Micro-F1
SPTree	68.4
Walk-based Model	72.9
ERANN	74.1

综合两个数据集上的多组对比实验结果，我们可以看出 ERANN 模型在两个不同类型的关系抽取数据集上均取得了最优的结果，证明了模型的高效性以及鲁棒性。通过对多组对比实验的进一步分析，以及模块孤立实验，验证了 ERANN 模型中结合 Bi-GRU 和自注意力机制的文本语义特征抽取模块以及实体关联特征的有效性。其中，实体关联特征抽取模块对于结果的提升更为明显，验证了本文思路的出发点，即从语料集中深入挖掘有助于关系抽取任务的信息。

3.7 本章小结

本章对基于实体关联图和注意力机制的关系抽取方法进行了详细的阐述。首先通过对当前关系抽取方法的概述，引出了现有关系抽取方法中存在的问题，并由此引出了实体关联关系和实体关联图的概念；然后对基于实体关联图和注意力机制的关系抽取方法的三个模块：文本语义特征抽取模块、实体关联特征抽取模块以及关系分类模块进行了详细介绍；最后对实验中的参数和对比实验设置进行说明，展示了实验结果，并对实验结果进行了深入分析。

4 基于实体对图与图卷积的关系抽取

本章将详细介绍基于实体对图与图卷积的关系抽取方法。该关系抽取模型由三个主要模块组成,包括文本语义特征抽取模块、实体对图拓扑特征抽取模块以及关系分类模块。本章将从本方法提出的思路以及具体的模型结构两个方面展开阐述,通过给出方法核心概念实体对图的定义阐明本方法的思路,再对模型的整体框架进行介绍,接着对其中的各个模块分别展开介绍。为了证明该关系抽取方法的有效性,将对对比实验设置以及实验结果进行详细阐述和分析。

4.1 实体对图定义

ERANN 模型将语料集层次的实体关联关系以实体关联图的形式表示,并使用注意力机制融合其上目标节点与其邻居节点的信息,将实体关联关系应用于关系抽取任务中。与其他主流模型的对比实验结果表明,语料集层次的实体关联关系能够有效地补充目标实体对的上下文信息。但是,ERANN 模型在对实体关联关系的应用上还存在两点不足。首先,关系抽取任务的目标是基于关系实例上下文信息对目标实体对之间的关系进行分类,其目标是实体对而不是单一实体,因此模型应该更加注重实体对的整体性,从实体对的角度出发提取特征,而不是将目标实体对拆分为两个单独的实体,这样就丧失了其中的部分关联性;其次,实体关联关系是多阶的,目标实体对之间的关系可能由其二阶或者更高阶的实体对隐含,而在 ERANN 模型中我们受限于实体关联图的构造方式和注意力机制的自身特点,对于多阶实体关联关系无能为力。基于利用实体关联关系辅助关系抽取任务这一思路,针对上述两点问题,本文提出了 EPGNN 模型,该模型使用以实体对作为节点的实体对图代替 ERANN 模型中以单个实体为节点的实体关联图,同时利用多层图卷积网络捕获实体对图的拓扑特征,克服了 ERANN 中注意力机制在计算多阶邻居特征时过于复杂的限制,能够自动地捕获目标实体对在实体对图上的拓扑特征,用于构建其多阶实体关联特征。

与 ERANN 模型中提出的实体关联图不同,实体对图直接将在关系抽取任务中至关重要的目标实体对作为节点,而不是单独将其中的一个实体作为节点;若两个实体对之间包含公共实体,我们就认为这两个实体对之间存在实体关联关系,那么在实体对图中,代表这两个实体对的节点之间有一条无向边相连。当前大多数主流的关系抽取模型都是将目标实体对作为独立的对象来单独处理的,但事实上,目标实体对之间的关系可以由与其包含公共实体的实体对隐含。如果关系抽取模型

能够对这种实体对之间的关联关系进行建模，将这些关联实体对信息引入目标实体对的特征构建过程中，对于模型效果的提升会有所帮助。图 4-1 使用 SemEval 2010 数据集中的关系实例作为例子，展示了这种实体关联关系在目标实体对的关系抽取任务中是如何起作用的。其中，红色圆圈标注出了包含目标实体对，两个黑色圆圈标注出了目标实体对的关联实体对，它们均与目标实体对包含公共实体，这种关联以边的形式在图中展示。对于目标实体对 $\langle \text{money}, \text{companies} \rangle$ ，其包含的 Entity-Destination 关系不仅可以直接从关系实例中提取，还可以通过它的关联实体信息挖掘得到：我们可以从 $\langle \text{money}, \text{funds} \rangle$ 以及 $\langle \text{funds}, \text{companies} \rangle$ 包含 Entity-Destination 这一事实推断出目标实体对 $\langle \text{money}, \text{companies} \rangle$ 也为 Entity-Destination 关系。本文提出的 EPGNN 模型将上述这种实体对之间的关联关系以实体对图的形式加以表示，并通过后续的拓扑特征抽取模块进行特征提取，将之应用到关系抽取任务中，相较于 ERANN 模型，更加充分且合理地挖掘了语料集层次的实体关联关系，起到提升关系模型关系抽取效果的作用。

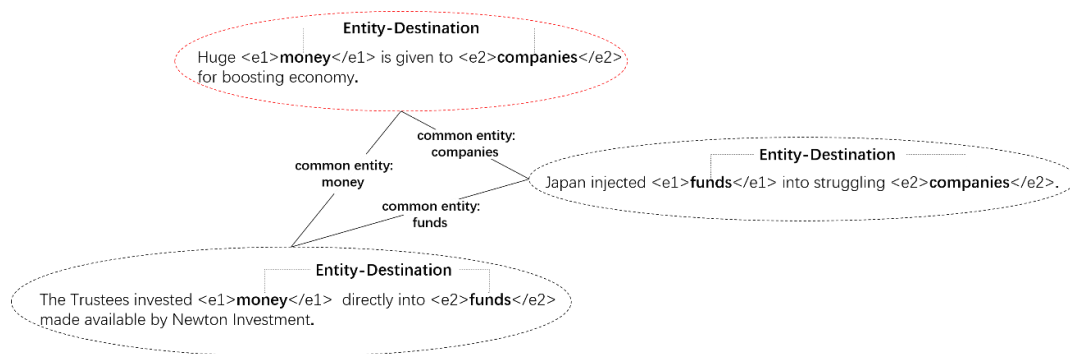


图 4-1 实体关联关系示意图

Figure 4-1 Illustration of the correlation between entity pairs.

4.2 模型总体框架

EPGNN 模型的总体框架图 4-2 所示。文本语义特征抽取模块基于预训练语言模型 BERT 对输入关系实例进行嵌入表示，得到关系实例向量表示和目标实体对向量表示，考虑到目标实体在关系抽取任务中的重要性，进一步将关系实例向量与目标实体对向量进行融合，完成文本语义特征抽取；而在实体对图拓扑特征抽取模块中，通过统计语料集中实体对之间的关联关系构建实体对图，设计实体分类学习任务，基于 BERT 学习得到实体表示用于构建实体对图节点初始表示；基于多层图卷积网络捕获实体对图的拓扑特征，高效地完成多阶实体关联关系信息抽取，应用于后续的关系分类特征构建；将文本语义特征与实体对图拓扑特征进行融合，输入

最后的关系分类模块，完成关系抽取任务。

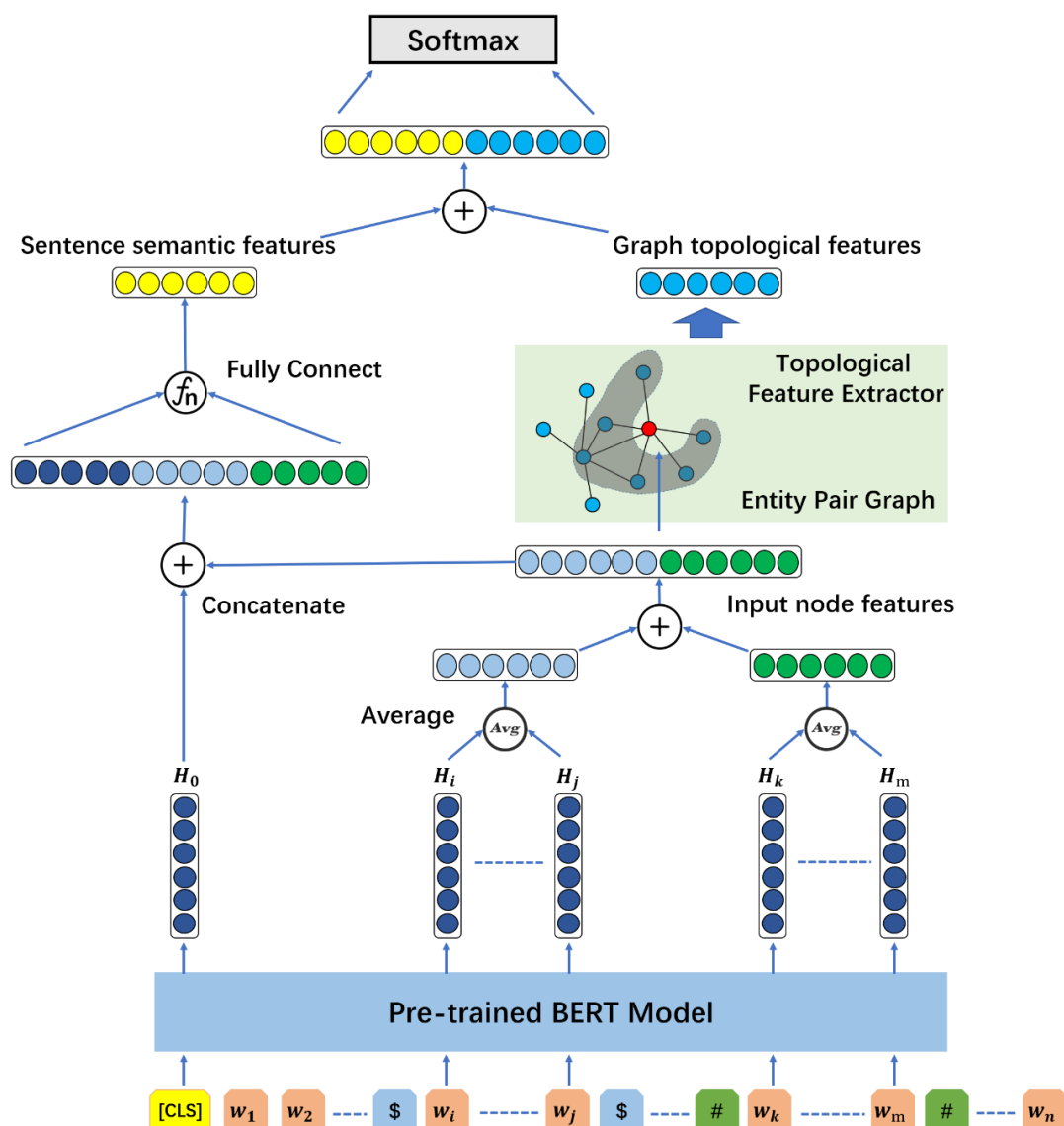


图 4-2 EPGNN 模型框架图

Figure 4-2 Overall architecture of the EPGNN model

4.3 文本语义特征抽取

本节将对 EPGNN 模型中的文本语义特征抽取模块进行详细介绍，首先介绍基于预训练语言模型 BERT 的文本表示，其中包括了关系实例向量表示和目标实体对向量表示方法；最后将阐述如何基于关系实例向量表示和目标实体对向量表示构建文本语义特征。

4.3.1 基于预训练语言模型 BERT 的文本表示

目前主流的关系抽取方法中一般采用, 诸如 Word2Vec 等文本向量表示方法来学习关系实例中词的向量表示, 结合位置向量等补充信息, 实现关系实例的文本表示。这种基于词嵌入的方法, 相较于传统关系抽取模型中依赖外部自然语言处理工具提取词法、句法以及语义特征的方法, 有着不依赖外部工具, 扩展性更强等优势。但是这种词向量模型是静态的, 虽然通用性强, 但是无法针对特定任务做动态优化; 同时词和向量之间是一对一的映射关系, 所以仍然存在一词多义的问题。针对上述两点问题, 本文提出的 EPGNN 模型基于预训练语言模型 BERT 进行关系实例的文本表示学习。使用 BERT 主要有以下三点优势: 首先, BERT 是基于英文维基百科以及 BooksCorpus 预训练得到的, 它的训练语料要远比传统的词嵌入表示方法丰富, 因此对于关系实例的文本语义表示效果将会更好; 其次, BERT 模型在训练过程中, 通过 Mask 的技术手段实现了双向的上下文语义信息学习, 使用 BERT 进行关系实例文本表示更能够捕获目标实体的上下文特征; 最后 BERT 是基于微调 (fine-tuning) 的预训练方法, 其生成的文本表示不是固定不变的将会适配下游任务, 因此可以解决传统词嵌入模型无法解决的一词多义问题。Google 提供了两种规模的预训练模型, 包括基础版 BERT_{BASE} 以及更大规模的 BERT_{LARGE}, 两种模型的区别在于所使用的 Transformer 层数、Transformer 中多头注意力机制头数以及最终隐藏层输出维数。ERANN 模型使用的是基础版本的 BERT 模型。接下来将介绍在 ERANN 模型中, 如何使用 BERT 进行文本表示。

BERT 模型的输入形式如图 4-3 所示, 由 token 嵌入、分句嵌入和位置嵌入三个部分组成。在 token 嵌入中, BERT 使用了 WordPiece 表示, 基于贪心的思想对词进行切片, 从而实现使用最少的 token 表示所有词汇的目的。BERT 所使用的 token 词汇表的规模为 30000。由于 BERT 所设计的预训练任务中包含了句之间的关联关系判断, 因此 BERT 的输入可能为两句话, 需要加入分句嵌入进行区分。由于 BERT 模型使用的 Transformer 框架是基于自注意力机制实现的特征抽取模型, 无法对词与目标实体之间的位置信息进行编码, BERT 在输入中加入了学习得到的位置向量。每个输入的第一个 token 都是用于下游分类任务的[CLS], 同时每一个输入的两个句子之间都是用[SEP]进行分隔。最终的输入是上述三种嵌入表示对位相加的结果。考虑到目标实体是关系抽取任务中至关重要的信息, 我们用特殊标记“\$”在首尾标识出头实体位置, 类似的用特殊标记“#”在在首尾标识出尾实体位置。通过这种方式赋予目标实体对特殊性同时对首尾实体进行区分, 使得模型的学习更加具有针对性。举例来说, 经过上述操作之后, EPGNN 模型文本表示模块的输入为如下形式:

“Huge \$ money \$ is given to # companies # for boosting economy.”

对于一个关系实例，其中包含两个目标实体 e_1 和 e_2 。经过上述预处理将其转换为 EPGNN 模型文本表示模块的输入形式，得到输入序列 $S = \{w_1, w_2, \dots, w_n\}$ 。将 S 输入 BERT 模型，在其内部完成词切片，分类标识和句分割标识添加，位置嵌入添加等操作之后通过预训练模型得到最终隐层输出，也就是该关系实例的文本表示 $\mathcal{H} \in \mathbb{R}^{(n+1) \times d_B}$ ，其中 n 为输入序列长度， d_B 为 BERT 模型隐藏单元数量，在此处所使用的基础版 BERT 模型中 $d_B = 768$ 。

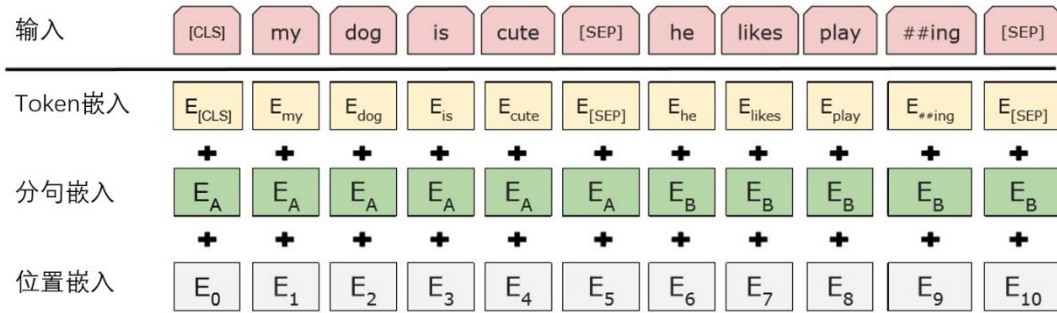


图 4-3 BERT 模型输入表示

Figure 4-3 BERT input representation.

4.3.2 文本语义特征构建

不同于一般的文本分类任务，关系抽取任务的目标是挖掘目标实体对之间的语义关系，而不仅仅是对关系实例进行文本分类。因此在关系抽取任务中，直接使用关系实例的文本表示进行关系分类往往难以取得好的效果，需要针对关系抽取任务进一步构建文本语义特征用于关系分类。在 EPGNN 模型中，我们针对关系抽取任务，以加强目标实体对信息为出发点构建文本语义特征。

在完成输入关系实例 S 的文本表示之后，我们得到其向量表示 $\mathcal{H} \in \mathbb{R}^{(n+1) \times d_B}$ ， \mathcal{H} 是由每个 token 的向量表示组成的矩阵。 \mathcal{H} 中的第一行对应的是 BERT 设计的文本分类标识，[CLS] 的向量表示，因此在 EPGNN 模型中，我们沿用这一设计，使用 [CLS] 的向量表示 H_0 作为输入关系实例的上下文向量表示，记为 $f^{cls} \in \mathbb{R}^{d_B}$ 。

同时，为了进一步加强目标实体对的文本信息，ERANN 模型结合该输入序列的分类标识向量表示，以及首尾实体的向量表示来构建文本语义特征。由于目标实体可能由多个 token 构成，假设头实体开始于 w_i ，结束于 w_j ，我们使用组成头实体的所有 token 对应的 BERT 隐层输出向量的平均来表示；类似的，假设尾实体开始于 w_k ，结束于 w_m ，则尾实体的表示由其组成 token 对应的向量表示的平均来表

得到包含关系词的句子向量表示

示。由于上述 token 向量表示在后续文本特征构建过程中会再次使用，此处我们在首尾实体向量生成过程中加入了全连接层，对特征进行非线性变换。至此完成了首尾实体的向量表示，头实体向量表示 f^{head} 和尾实体向量表示 f^{tail} 分别由式 (4-1) 和式 (4-2) 表示：

$$f^{head} = \sigma(W_1[\frac{\sum_{t=i}^j H_t}{j-i+1}]^T + b_1) \quad (4-1)$$

$$f^{tail} = \sigma(W_2[\frac{\sum_{t=k}^m H_t}{m-k+1}]^T + b_2) \quad (4-2)$$

其中， $W_1 \in \mathbb{R}^{d_B \times d_B}$ 和 $W_2 \in \mathbb{R}^{d_B \times d_B}$ 为权重矩阵， $b_1 \in \mathbb{R}^{d_B}$ 以及 $b_2 \in \mathbb{R}^{d_B}$ 为偏置。在得到关系实例上下文向量表示 f^{cls} 以及目标实体对向量表示 f^{head} 和 f^{tail} 之后，EPGNN 使用三者的连结作为输入关系实例的文本语义特征 $f^s \in \mathbb{R}^{d_s}$ ，其中 d_s 为 EPGNN 文本语义特征的维数。 f^s 表示如式 (4-3) 所示：

$$f^s = \sigma(W_3[f^{cls}; f^{head}; f^{tail}]^T + b_3) \quad (4-3)$$

其中， $\sigma(\cdot)$ 为激活函数，例如 $\text{ReLU}(\cdot)$ ； $W_3 \in \mathbb{R}^{d_s \times 3d_B}$ 为权重矩阵， $b_3 \in \mathbb{R}^{d_s}$ 为偏置。根据上式计算得到文本语义特征 f^s ，完成文本语义特征抽取。

4.4 实体对图拓扑特征抽取

本节将对 EPGNN 模型中的实体对图拓扑特征抽取模块进行详细介绍。实体对图拓扑特征抽取模块是 EPGNN 模型中的核心模块，是本文思路的出发点的体现，即不依赖外部资源，从语料集自身深入挖掘语义和结构信息用于关系抽取任务。其中实体对图节点的嵌入表示是信息有效传递的基础，如何在 ERANN 模型实体关联关系的基础上更进一步，使用多层图卷积网络挖掘多阶实体关联关系是本节的重点介绍内容。

4.4.1 实体对图节点嵌入表示

本文提出的 EPGNN 模型构建了实体对图，并将目标实体对在图中对应节点的拓扑特征应用到关系抽取的特征构建工作中，通过这种方式将目标实体对在语料集层次的实体关联关系融合到关系抽取模型之中，提升模型的效果。无论是目标实体对本身的信息，还是其关联实体对的信息，在实体对图中都是以节点嵌入表示的形式进行传递的。因此，实体对图中的节点嵌入表示是否合适，将很大程度上影响模型的表现。

实体对图是以实体对为节点，以公共实体为边构建的图结构。因此，其中的节

点表示应该由对应实体对的信息来表示, 使用目标实体对所包含的两个实体的表示来构造节点表示是常见做法。对于使用诸如 Word2Vec 等传统的文本嵌入表示方法的模型, 因为其词向量表示是固定, 这种方式是可行的。本文在 EPGNN 模型中使用的是预训练模型 BERT, 其词向量表示将会根据下游任务进行微调, 不是固定的, 因此不能直接使用对应实体的向量表示进行实体对图中的节点表示。为了解决这一问题, 本文设计了一个实体分类的训练任务作为预训练语言模型 BERT 的下游任务, 以此来学习得到实体的表示来构建实体对图中的节点表示。具体任务设置如下, 出于实体表示应该最能反映实体自身特点这一考虑, 我们将关系实例输入预训练语言模型, 采用与 4.3.2 小节中相同的方法, 根据式 (4-1) 和 (4-2) 分别计算得到对应实体的表示, 并将其作为特征输入 softmax 分类器进行分类。在整个语料集中, 假设一个实体 e_i 会出现在 K 关系实例中, 对第 j 关系实例 BERT 模型会学习得到一个向量表示 $f_j^i \in \mathbb{R}^{d_E}$, 我们选取模型预测属于正确类型概率最大的实体表示作为该实体的向量表示, 再使用目标实体中两个实体的向量表示进行连结, 构建实体对图中对应节点的向量表示。即对于一个实体对 $\langle e_i, e_j \rangle$, 对应实体对图中编号为 k 的节点, 假设其中实体 e_i 经过 BERT 输出的向量表示的集合为 $\mathcal{F}_i = \{f_1^i, f_2^i, \dots, f_{|\mathcal{F}_i|}^i\}$, 实体 e_j 的向量表示集合 \mathcal{F}_j 同理, 最终目标实体对 $\langle e_i, e_j \rangle$ 对应节点的向量表示 $f_k^N \in \mathbb{R}^{d_N}$, $d_N = 2d_E$, 可由式 (4-4)、(4-5) 和 (4-6) 计算得到:

$$f_k^N = [f_i^E; f_j^E] \quad (4-4)$$

$$f_i^E = \arg \max_{f^i} p(\text{label} | f^i) \quad (4-5)$$

$$p(\text{label} | f^i) = \text{softmax}(\mathbf{W}_4[f^i]^T + \mathbf{b}_4) \quad (4-6)$$

其中, label 为该实体的类型标签, $f^i \in \mathcal{F}_i$ 为关系实例对应的实体向量表示, f_i^E 为实体 e_i 的向量表示; $p(\text{label} | f^i)$ 表示当前实体 e_i 表示预测得到标签 label 的概率, $\mathbf{W}_4 \in \mathbb{R}^{r_e \times d_E}$ 和 $\mathbf{b}_4 \in \mathbb{R}^{r_e}$, r_e 为预设的实体类别数量。至此, EPGNN 模型中实体关联图的节点嵌入表示工作完成。

4.4.2 多层图卷积网络

本文提出的 ERANN 模型通过构建实体关联图, 并使用注意力机制构建实体关联特征的形式将实体关联特征融入实体关系抽取任务, 但是这种方式只能获取一阶实体关联关系。显然, 实体之间的多阶关联关系对于关系抽取任务价值更大, 因此, 在 EPGNN 模型中我们更进一步, 挖掘多阶实体关联关系特征作为模型的补充信息。EPGNN 模型通过构建实体对图的形式来表征语料集层次的实体关联关系。正如图 4-1 中所展示的, 实体对图的拓扑特征可以作为关系抽取任务中文本语义特征补充, 辅助于目标实体的关系抽取。为了深入挖掘语料集层次的多阶实体关联关

系, EPGNN 模型使用多层图卷积网络来编码实体对图中的拓扑结构, 其中包含了多阶实体关联信息, 将之作为补充特征输出到最终的关系分类模块。

GCNs 是工作在图结构上的神经网络模型, 它巧妙地集成了局部节点特征和图拓扑特征。Li 等^[10]通过实验证明, GCNs 模型是拉普拉斯平滑的一种特殊形式, 它将节点及其附近的邻居节点的特征融合在一起。因此, 经过该操作之后, 同一簇中节点的特征将变得更加相似, 显然, 这将有助于节点分类任务。在实体对图中, 节点对应于目标实体对, 节点分类即为目标实体对分类, 也就是关系抽取任务的目标。由此可见, 使用 GCNs 模型在原理上能够辅助于实体关系抽取任务。因此, 我们在 EPGNN 模型中, 使用多层 GCNs 提取实体对图拓扑特征。GCNs 的主流实现方法有两种: 空间方法和频谱方法。在 EPGNN 模型中, 我们使用谱图卷积方法实现实体对图的拓扑特征提取。谱图卷积通过在频域上分解图信号, 然后在其上应用频谱滤波器来定义卷积。

具体地, 对于给定的实体对图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, 其中包含 n 个节点, 该图结构可以使用邻接矩阵 $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ 表示, 其中, 当节点 i 和节点 j 之间有边相连时 $a_{ij} = a_{ji} = 1$ 。为了考虑目标节点自身, 我们在邻接矩阵中添加了节点的自连接操作, 此时, 我们可以得到新的邻接矩阵 $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$, 其中 $\mathbf{I}_N \in \mathbb{R}^{n \times n}$ 为单位阵。图 \mathcal{G} 对应的度矩阵 $\tilde{\mathbf{D}}$ 可以根据邻接矩阵 $\tilde{\mathbf{A}}$ 计算得到, 表示为 $\tilde{\mathbf{D}} = \text{diag}\{d_1, d_2, \dots, d_n\}$, 其中 $d_i = \sum_j \tilde{a}_{ij}$ 表示节点 i 的度。经过上述计算, 我们可以得到图 \mathcal{G} 对应的邻接矩阵 $\tilde{\mathbf{A}}$ 和度矩阵 $\tilde{\mathbf{D}}$ 。

对于图 \mathcal{G} , 其特征由两部分构成: 结构特征和节点特征。其中结构特征可以由邻接矩阵 $\tilde{\mathbf{A}}$ 和度矩阵 $\tilde{\mathbf{D}}$ 表示, 而节点特征则使用 4.4.2 小节中基于预训练模型 BERT 训练得到的节点表示构建的信号矩阵 $\mathbf{X} \in \mathbb{R}^{n \times d_N}$ 表示, 其中第 i 行代表实体对图中编号为 i 的节点对应的节点表示 $\mathbf{f}_i^N \in \mathbb{R}^{d_N}$ 。根据谱图卷积规定的特征传播规则, 具有 d_f 个滤波器的谱图卷积操作可以在数学上形式化为式 (4-7) 所示:

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \Theta^{(l)}) \quad (4-7)$$

其中, $\mathbf{H}^{(l)}$ 为第 l 层的节点特征矩阵, $\mathbf{H}^{(0)} = \mathbf{X}$; $\Theta \in \mathbb{R}^{d_N \times d_f}$ 为第 l 层的权重矩阵; $\sigma(\cdot)$ 为激活函数, 如 $\text{ReLU}(\cdot)$ 。

在 EPGNN 模型中, 我们使用 l 层图卷积操作捕获实体对图的拓扑特征, 其中图卷积网络层数 l 为超参数。举例来说, 两层实体对图拓扑特征 $\mathbf{f}^g \in \mathbb{R}^{d_g}$ 可由式 (4-8) 计算得到:

$$\mathbf{f}^g = \hat{\mathbf{A}} \sigma(\hat{\mathbf{A}} \mathbf{X} \mathbf{W}^{(0)}) \mathbf{W}^{(1)} \quad (4-8)$$

其中, $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$, $\mathbf{W}^{(0)} \in \mathbb{R}^{d_N \times d_h}$ 和 $\mathbf{W}^{(1)} \in \mathbb{R}^{d_h \times d_g}$ 均为权重矩阵, d_h 为隐层特征维度, d_g 为实体对图拓扑特征维度, d_h 和 d_g 均为 EPGNN 模型的超参数。

4.5 关系抽取

在完成文本语义特征抽取和实体对图拓扑特征抽取之后，EPGNN 模型基于文本语义特征 \mathbf{f}^s 和实体对图拓扑特征 \mathbf{f}^g ，使用 softmax 分类器对输入目标实体对进行关系分类。softmax 分类器由单层全连接网络构成，并在输出层使用 softmax 归一化操作。对于输入关系实例 S ，EPGNN 模型将文本语义特征与实体关联特征通过特征向量拼接的形式融合，使用 $\mathbf{f}^s \in \mathbb{R}^{d_s}$ 和 $\mathbf{f}^g \in \mathbb{R}^{d_g}$ 的连结作为 softmax 分类器的输入，具体形式如式 (4-9) 所示：

$$p(y|S) = \text{softmax}(\mathbf{W}_0[\mathbf{f}^s; \mathbf{f}^g]^T + \mathbf{b}_0) \quad (4-9)$$

其中， $\mathbf{W}_0 \in \mathbb{R}^{r \times (d_s + d_g)}$ 为权重矩阵， $\mathbf{b}_0 \in \mathbb{R}^r$ 为偏置， r 代表预设的关系类型数量，为模型超参数。

接着 EPGNN 模型使用 softmax 分类器的输出对输入关系实例中的目标实体对之间的关系类型进行预测，如式 (4-10) 所示：

$$\hat{y} = \arg \max_y p(y|S) \quad (4-10)$$

完成上述步骤得到输入关系实例的预测标签之后，softmax 分类器使用交叉熵（Cross-Entropy）损失函数，如式 (4-11) 所示：

$$\mathcal{L} = -\sum_i \sum_j \text{label}_{ij} \log p(y_{ij} | S_i, \theta) \quad (4-11)$$

其中， θ 代表 EPGNN 模型中的所有参数的集合， y_{ij} 表示样本 i 的预测标签为 j ， label_{ij} 为关于样本 S_i 的 one-hot 向量，当其真实标签为 j 时 label_{ij} 为 1，否则为 0。

在 EPGNN 模型中，我们使用 Adam 优化算法求解参数，并将网络参数使用 Xavier 初始化方法进行初始化。

4.6 实验设置与结果分析

本节将对 EPGNN 模型在 ACE 2005 和 SemEval 2010 数据集上的对比实验进行介绍，包括模型的超参数设置、对比实验设置以及实验结果。同时，还将针对不同类型的代表方法与 EPGNN 对比实验结果，对 EPGNN 模型的性能进行深入分析；并补充实验说明实体对拓扑特征抽取模块的即插即用特性。

4.6.1 实验设置

EPGNN 模型使用 $BERT_{BASE}$ 预训练模型对关系实例以及目标实体进行编码；通过构建实体对图的形式将语料集层次的实体关联关系以图结构的形式表示；使用

多层图卷积网络捕获实体对图的拓扑结构,对多阶实体关联特征进行表征,应用于实体关系抽取任务中。在 EPGNN 模型训练阶段,我们使用 Xavier 初始化方法完成网络参数的初始化工作,并使用 adam 优化算法求解模型参数。我们实体对图拓扑特征抽取模块加入了 dropout 操作来防止网络过拟合。EPGNN 模型中主要包括如下超参数:文本语义特征向量维度 d_s ; 实体对图节点向量维度 d_N ; 实体对图拓扑特征维度 d_g ; 图卷积网络层数 l ; 图卷积隐层特征维度 d_h ; batchsize 大小; 模型学习率 λ ; dropout 率 dp 以及 adam 优化器参数 β_1 、 β_2 和 ϵ 。为了更加科学准确地进行超参数的选择,本文在实验部分使用 5 折交叉验证的方法进行参数选取,所得出的最优超参数设置如表 7 所示。

表 7 EPGNN 模型超参数设置

Table 7 Hyper-parameter setting of EPGNN model

参数	数值
d_s	256
d_N	64
d_g	64
d_h	128
l	3
batchsize	16
λ	5e-6
dp	0.3
β_1	0.9
β_2	0.999
ϵ	10e-8

4.6.2 对比实验设置

在对比实验设置方面,除了沿用在 ERANN 模型对比实验中使用到两大类的主流关系抽取方法之外,考虑到 EPGNN 模型基于预训练语言模型进行文本语义特征抽取,我们在对比实验中加入了基于预训练语言模型的关系抽取方法 (Pre-Trained Methods)。添加这一类型的方法一方面可以验证使用预训练语言模型进行文本表示这一手段在实体关系抽取任务中的有效性;另一方面,通过与这类方法的对比,可以验证 EPGNN 模型中的实体对拓扑特征是否对模型的关系抽取效果有所提升,具体的对比实验设置如表 8 所示。

表 8 EPGNN 对比实验设置

Table 8 Comparison setting of EPGNN model

类型	模型	应用数据集
<i>End-to-End Methods</i>	CNN+PF	SemEval
	CR-CNN	SemEval
	Att-Pooling CNN	SemEval
<i>Dependency Methods</i>	SDP-LSTM	SemEval
	DRNNs	SemEval
	SPTree	ACE/SemEval
	C-GCN	SemEval
	walk-based model	ACE
<i>Pre-Trained Methods</i>	TRE ^[62]	SemEval
	R-BERT ^[63]	ACE/SemEval
<i>Our Methods</i>	ERANN	ACE/SemEval

4.6.3 实验结果分析

本节将介绍 EPGNN 模型在两个实验数据集上的对比实验结果,并通过对实验结果的全面分析,证明本文提出的 EPGNN 模型的有效性,主要说明基于预训练语言模型 BERT 的文本语义抽取模块以及实体对图拓扑特征抽取模块在关系抽取任务中起到的促进作用。同时,通过在集中典型的基础关系抽取模型中加入本文提出的实体对拓扑特征抽取模块,并与原始模型进行对比实验,在验证该模块有效性的同时,对其即插即用的特征进行阐述。

本文提出的 EPGNN 模型在 SemEval 2010 数据集上与对比方法的实验结果如表 9 所示。表 9 显示,本文提出的 EPGNN 模型在该数据集上取得了 90.2%的宏观 F1 值,这一结果高于其他所有对比实验方法。其中,相较于 End-to-End 类型的关系抽取方法的最先进方法 Att-Pooling CNN,EPGNN 模型提升了 2%的 F1 值;在与基于依存分析的关系抽取方法的对比中,我们模型的优势更为明显,比其最优方法结果高出 4%;考虑到基于 BERT 的方法已经取得了优秀的关系抽取结果,EPGNN 在此基础上将 F1 值提升了 1%,仍然证明了本文提出方法的有效性。

在 3 大类对比实验方法中,Wu 等^[63]提出的基于预训练模型 BERT 的关系方法 R-BERT 取得了最好的效果,这一点证明了,使用诸如 BERT 等预训练语言模型代替传统的词嵌入表示方法,能够更有效地提取关系实例的上下文信息,达到提升模

型关系抽取效果的目的。EPGNN 模型相比于单纯使用预训练模型的 R-BERT 方法 F1 值提升了近一个百分点，这部分的提升说明本文提出的实体对图拓扑特征能够补充目标实体的上下文信息，辅助其更好地进行关系抽取。此外，在 ERANN 模型对比实验中取得最优效果的 Att-Pooling CNN 模型，其 F1 值结果虽然略低于 EPGNN 模型，但在不依赖外部资源且未引入预训练语言模型的基础上，仅从关系实例中挖掘语义和结构信息就能够与基于预训练模型的 R-BERT 方法结果持平，进一步说明了关系实例本身蕴含着丰富的信息。

表 9 EPGNN 在 SemEval 数据集对比实验结果

Table 9 Comparison of EPGNN model on SemEval 2010

类型	模型	Macro-F1
<i>End-to-End Methods</i>	CNN+PF	82.7
	CR-CNN	84.1
	Att-Pooling CNN	88.0
<i>Dependency Methods</i>	SDP-LSTM	83.7
	DRNNs	85.8
	SPTree	84.5
	C-GCN	84.8
<i>Pre-Trained Methods</i>	TRE	87.1
	R-BERT	89.2
<i>Our Methods</i>	ERANN	86.6
	EPGNN	90.2

为了验证 EPGNN 模型在面对不同类型数据时的表现如何，我 ACE 2005 数据集上与三类典型方法进行了对比实验，实验结果如表 10 所示。从表 10 的结果可以看出，本文提出的 EPGNN 模型在 ACE 2005 数据集上仍然取得了最优的结果，这一结果结合 ERANN 模型的实验结果说明本文提出的语料集层次实体关联关系，对于实体关系抽取任务切实有效，并且对于不同类型的数据集具有不错的鲁棒性。此外，基于预训练模型的 R-BERT 方法明显优于除 EPGNN 以外的其他方法，这一结果证明了预训练模型的引入对于关系抽取任务提升明显。与另外两大类代表方法的对比实验可以得出与 ERANN 模型实验中的相同结论，实体之间的关联关系对于实体关系抽取任务具有促进作用，同时这种关联关系不仅限于一个关系实例内，还存在与关系实例之间，也就是语料集层次，合理表征这种实体之间的关联关系能够将之应用到关系抽取任务中，能够有效地提升实体关系抽取模型的效果。

表 10 EPGNN 在 ACE 数据集对比实验结果

Table 10 Comparison of EPGNN model on ACE 2005

模型	Micro-F1
SPTree	68.4
Walk-based Model	72.9
R-BERT	76.2
ERANN	74.1
EPGNN	77.1

EPGNN 模型与 ERANN 模型都将语料集层次的实体关联关系引入关系抽取任务中，两者的不同点在于对实体关联关系的建模方式和特征抽取方法。从表 9 和表 10 中的结果来看，EPGNN 模型的效果要明显优于 ERANN 模型，为了进一步对比两种实体关联关系的建模方式和特征抽取方法，我们在 Bi-GRU 模型的基础上分别使用两种不同的实体关联关系表征和特征抽取方式，在 SemEval 数据集上进行对比实验，其结果如表 11 所示。从表中所示结果可以看出，在 Bi-GRU 模型的基础上，两种引入实体关联关系特征的方式都带来了可观的提升，这再次验证了实体关联关系能够补充实体关系抽取任务所需的语义和结构信息。这两种处理方法中，基于实体对图和多层图卷的方法要优于基于实体关联图和注意力机制的方法，这一结果说明多阶实体关联特征相较于一阶特征更有助于目标实体对的关系抽取，且前者的训练也更加高效。

表 11 实体关联特征构建方法对比

Table 11 Comparison of entity correlation features

模型	Macro-F1
Bi-GRU	83.3
Bi-GRU+ER	85.1
Bi-GRU+GCN	85.9

上述实验结果验证了本文提出 EPGNN 模型中的实体对图拓扑特征在关系抽取任务上的有效性，除此之外，这一模块还可以即插即用地应用于其他现有的关系抽取模型，并在不需要更改原始特征构建方法的情况下提升模型的性能。在实体对图构建完成的前提下，完成拓扑特征提取模块的引入只需对现有模型进行两处修改，首先需要使用当前模型的嵌入表示方法生成目标实体对的表示，并将其输入拓扑特征提取模块以生成实体对图拓扑特征；最后，使用原始分类特征与图拓扑特征的

连结作为新的分类特征进行关系分类，具体模型结构如图 4-4 所示。

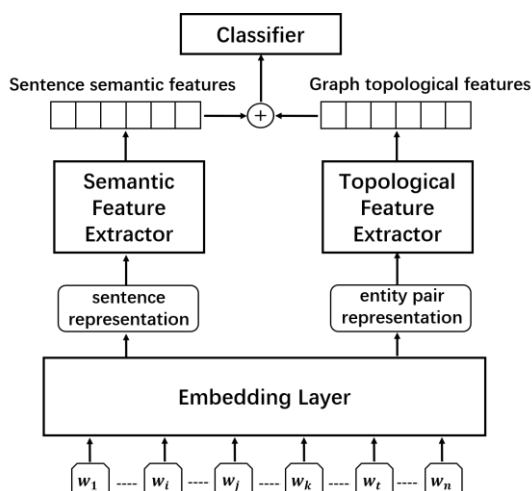


图 4-4 基于实体对图拓扑特征增强关系抽取模型框架图

Figure 4-4 The framework of entity pair graph enhanced relation classifier

为了证明这一特性，我们基于三种代表性的神经网络模型在 SemEval 数据集上进行实验：基于 CNN 的模型、基于 LSTM 的模型以及基于 BERT 的模型，实验结果如表 12 所示。从表中结果可以看出，基于 CNN 和基于 LSTM 的模型在引入实体对图拓扑特征后 F1 值都取得了 2% 左右的提升；对于已经取得较好分类结果的基于 BERT 的模型，添加拓扑特征仍然能够带来性能提升。结果表明，添加实体对图拓扑特征可以提升现有关系抽取模型的效果。拓扑特征抽取模块的核心组件是多层 GCN，它是拉普拉斯平滑的一种特殊形式，通过计算节点的新特征作为其自身及其邻居的加权平均值，使连接的节点更相似，从而更易于分类。现有的大多数关系抽取方法忽略了这种语料集层次的实体关联信息。因此，添加实体对图的拓扑特征对于实体关系抽取具有很大的价值。

表 12 基于 SemEval 数据集的 GCN 模块孤立实验

Table 12 Ablation with and without the GCN module on SemEval 10.

类型	模型	Macro-F1
CNN based Methods	without GCN	75.3
	with GCN	77.6
LSTM based Methods	without GCN	79.6
	With GCN	82.1
BERT based Methods	without GCN	89.2
	with GCN	90.2

4.7 本章小结

本章对基于实体对图和多层图卷积的关系抽取方法进行了详细阐述。通过分析在第 3 章提出的 ERANN 模型存在的不足之处，引出新的语料集层次实体关联关系特征的建模形式：实体对图，对实体对图的概念、优势和构建方式进行详细介绍。对 EPGNN 模型的整体架构进行介绍，并展开介绍了其中的基于预训练语言模型 BERT 的文本语义特征构建模块、基于多层图卷积实体对图拓扑特征抽取模块以及关系分类模块。最后对模型超参数设置和对实验结果进行详细说明和深入分析，通过于现有主流模型的对比实验验证 EPGNN 模型的有效性；与 ERANN 模型对比实验阐明实体对图这种建模方式的优势；最后针对实体对图拓扑特征抽取模块的即插即用特性进行实验和结果分析。

5 结论

5.1 工作总结

信息抽取技术能够帮助人们从海量的数据中快速准确的抽取真正需要的有价值信息,在这个数据爆炸式增长的信息时代扮演着越来越重要的角色。实体关系抽取在整个信息抽取流程中起到串联和整合的作用。实体关系抽取技术能够基于目标实体对上下文中的语义和结构特征,挖掘出两者之间的语义关系,其结果能够应用于问答系统、知识图谱等需要对文本进行关联关系理解的自然语言处理系统。传统的关系抽取方法大多将目标实体对作为独立对象,仅从包含目标实体对的关系实例中提取上下文信息来构建特征。本文将目标实体对的上下文信息从单个关系实例扩展到整个语料集,提出了语料集层次的实体关联关系,并通过两种不同的图结构来对这种关联关系进行建模,在此基础上对实体关联特征的抽取方法展开探索。基于上述思路,本文提出了基于实体关联图与注意力机制的关系抽取方法以及基于实体对图与多层图卷积的关系抽取方法,并与当前主流方法展开对比实验,验证本文提出方法的有效性。本文所做的工作可以总结为如下几点:

(1) 设计并实现基于实体关联图与注意力机制的关系抽取方法。一方面,通过构建实体关联图的形式对语料集层次的实体关联关系进行建模,并通过注意力机制对实体关联图中目标实体所对应节点的邻居节点信息赋予权重,通过将邻居节点信息与中心节点带权融合,实现语料集层次实体关联特征的抽取;另一方面基于 Bi-GRU 和句内自注意力机制完成文本语义特征抽取。将实体关联特征作为上下文特征的补充与文本语义信息融合用于关系抽取,提升了模型在关系抽取任务上的效果,验证了实体关联关系在该任务上的有效性。

(2) 设计并实现基于实体对图与多层图卷积的关系抽取方法。提出了以目标实体对为节点的实体对图,代替实体关联图对语料集层次实体关联关系进行建模;使用预训练语言模型 BERT 替代传统的词嵌入表示方法对关系实例及目标实体进行表示,基于关系实例的分类向量和目标实体对向量表示构建文本语义特征;针对预训练模型的特点,设计并实现了基于实体分类的实体对图节点表示构建方法;通过多层图卷积高效地捕获实体对图的拓扑特征,将 ERANN 模型中使用的一阶实体关联特征扩展到多阶特征。融合实体对图拓扑特征与文本语义特征进行关系抽取,在不依赖任何外部自然语言处理工具和背景知识的情况下,在两个公开关系抽取数据集上取得了当前最优的结果,验证了语料集层次实体关联特征在关系抽取

任务中的有效性；同时，将实体对图拓扑特征抽取模块应用到其他主流关系抽取模型中，其结果得到了不同程度的提升，验证了实体关联特征对于不同模型结构的鲁棒性和扩展性，可以作为即插即用的模块来提升关系抽取的效果。

5.2 未来工作展望

本文所提出的关系抽取方法的基本思想是在不依赖外部资源的前提下，从数据集中挖掘关系抽取任务所需的上下文语义特征和结构特征。基于这一思路，本文提出了语料集层次实体关联信息的概念，并对融合实体关联信息的方法做出了两次尝试，提出了两种基于实体关联特征的关系抽取方法，并在两个公开数据集上对模型的效果进行了充分的验证。虽然基于实体对图和多层图卷积的方法取得了当前最优的结果，但是这是对从本文基本思想的初步尝试，仍然存在很多不足，有待进一步的研究和探索：

（1）本文提出的基于实体对图和多层图卷积的关系抽取方法，由于其中 GCN 模型特性的天然限制，对于新增实体的处理较为复杂，因此如何选取更加合适的图神经网络模型，从而同时具备快速准确捕获图拓扑特征的能力以及对于新增节点的快速处理能力，将会是重点研究方向之一。

（2）目前实体关系抽取任务发展的热潮在基于远程监督的关系抽取以及基于小样本学习的关系抽取。无论是基于远程监督还是基于小样本学习的关系抽取研究，都致力于减少这一任务对标注数据的依赖性，如何将本文提出的关系抽取方法迁移应用到上述两个关系抽取的分支任务中，将会是接下来的研究重点。

参考文献

- [1] Jurafsky D, Martin J. Speech and Language Processing[M]. Beijing: Publishing House of Electronics Industry, 2018. 593.
- [2] Yu M, Yin W, Hasan K S, et al. Improved neural relation detection for knowledge base question answering[J]. arXiv preprint, 2017, ArXiv:1704.06194.
- [3] 刘绍毓, 李弼程, 郭志刚, 等. 实体关系抽取研究综述[J]. 信息工程大学学报, 2016, 17(5): 541-547.
- [4] Appelt D E, Hobbs J R, Bear J, et al. SRI International FASTUS system: MUC-6 test results and analysis[C]. //Proceedings of the 6th conference on Message understanding. Association for Computational Linguistics, 1995: 237-248.
- [5] McDonald D, Chen H, Su H, et al. Extracting gene pathway relations using a hybrid grammar: the Arizona Relation Parser[J]. Bioinformatics, 2004, 20(18): 3370-3378.
- [6] Jayram T S, Krishnamurthy R, Raghavan S, et al. Avatar information extraction system[J]. IEEE data engineering bulletin, 2006, 29(1): 40-48.
- [7] Zhao S, Grishman R. Extracting relations with integrated information using kernel methods[C]. //Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005: 419-426.
- [8] Guo D Z, Jian S, Jie Z, et al. Exploring various knowledge in relation extraction[C]. //Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005: 427-434.
- [9] Xi B, Qian L, Zhou G, et al. The Application of Combined Linguistic Features in Semantic Relation Extraction [J]. Journal of Chinese Information Processing, 2008, 22: 44-49.
- [10] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction[C]. //Proceedings of the ACL Interactive Poster and Demonstration Sessions. 2004: 178-181.
- [11] Suchanek F M, Ifrim G, Weikum G. Combining linguistic and statistical analysis to extract relations from web documents[C]. //Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. 2006: 712-717.
- [12] Nguyen T H, Grishman R. Employing word representations and regularization for domain adaptation of relation extraction[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2014: 68-74.
- [13] Zeng D J, Liu K, Lai S W, et al. Relation classification via convolutional deep neural network[C]. //Proceedings of the 25th International Conference on Computational Linguistics, 2014. 23-29.
- [14] dos Santos C, Xiang B, Zhou B. Classifying Relations by Ranking with Convolutional Neural Networks[C]. //Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015. 1: 626-634.
- [15] Xu Y, Mou L, Li G, et al. Classifying relations via long short term memory networks along

- shortest dependency paths[C]. //Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 1785-1794.
- [16] 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, 2016. 1105-1116.
- [17] Ji G L, Liu K, He S Z, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions[C]. //Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017. 3060-3066.
- [18] Nguyen T H, Grishman R. Relation extraction: Perspective from convolutional neural networks[C]. //Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing. 2015: 39-48.
- [19] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint, 2013, ArXiv:1301.3781.
- [20] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, 2018, ArXiv:1810.04805.
- [21] Walker C, Strassel S, Medero J, et al. ACE 2005 multilingual training corpus[J]. Linguistic Data Consortium, Philadelphia, 2006, 57.
- [22] Hendrickx I, Su N K, Kozareva Z, et al. SemEval-2010 task 8: multi-way classification of semantic relations between pairs of nominals[C]. //Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions, 2009. 94-99.
- [23] Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]. //Advances in Neural Information Processing Systems. 2014. 2204-2212.
- [24] Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network[J]. arXiv preprint, 2016, ArXiv:1605.08900.
- [25] Wang L L, Cao Z, Melo G D, et al. Relation classification via multi-level attention CNNs[C]. //Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016. 1298-1307.
- [26] Yin W, Schutze H, Xiang B, et al. ABCNN: attention-based convolutional neural network for modeling sentence pairs[J]. Transactions of the Association for Computational Linguistics, 2016, 4(1): 259-272.
- [27] 朱张莉, 饶元, 吴渊, 等. 注意力机制在深度学习中的研究进展[J]. 中文信息学报, 2019, 33(6): 1-11.
- [28] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint, 2014, ArXiv:1409.0473.
- [29] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint, 2015, ArXiv:1508.04025.
- [30] Yu A W, Dohan D, Luong M T, et al. Qanet: Combining local convolution with global self-attention for reading comprehension[J]. arXiv preprint, 2018, ArXiv:1804.09541.
- [31] Golub D, He X. Character-level question answering with attention[J]. arXiv preprint, 2016, ArXiv:1604.00727.
- [32] Howard J, Ruder S. Universal language model fine-tuning for text classification[C]. //Meeting of the Association for Computational Linguistics, 2018: 328-339.

- [33] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding with unsupervised learning[J]. Technical report, OpenAI, 2018.
- [34] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3: 1137-1155.
- [35] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. //Advances in Neural Information Processing Systems. 2017: 5998-6008.
- [36] Hu B, Lu Z, Li H, et al. Convolutional neural network architectures for matching natural language sentences[C]. //Advances in neural information processing systems. 2014: 2042-2050.
- [37] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. //Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [38] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal processing magazine, 2012, 29(6): 82-97.
- [39] Shuman D I, Narang S K, Frossard P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains[J]. IEEE signal processing magazine, 2013, 30(3): 83-98.
- [40] 徐冰冰, 岑科廷, 黄俊杰, 沈华伟, 程学旗. 图卷积神经网络综述[J/OL]. 计算机学报, 2019: 1-31.
- [41] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[J]. arXiv preprint, 2013, ArXiv:1312.6203.
- [42] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[C]. //Advances in neural information processing systems. 2016: 3844-3852.
- [43] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint, 2016, ArXiv:1609.02907.
- [44] Hasegawa T, Sekine S, Grishman R. Discovering relations among named entities from large corpora[C]. //Proceedings of the 42nd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2004: 415.
- [45] Chen J, Tandon N, De Melo G. Neural word representations from large-scale commonsense knowledge[C]. //2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). IEEE, 2015, 1: 225-228.
- [46] Guo D Z, Jian S, Jie Z, et al. Exploring various knowledge in relation extraction[C]. //Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005: 427-434.
- [47] Suchanek F M, Ifrim G, Weikum G. Combining linguistic and statistical analysis to extract relations from web documents[C]. //Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. 2006: 712-717.
- [48] Jiang J, Zhai C X. A systematic exploration of the feature space for relation extraction[C]. //Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics. 2007: 113-120.
- [49] Nguyen T V T, Moschitti A. End-to-end relation extraction using distant supervision from

- external semantic repositories[C]. //Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. Association for Computational Linguistics, 2011: 277-282.
- [50] Culotta A, Sorensen J. Dependency tree kernels for relation extraction[C]. //Proceedings of the 42nd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2004: 423.
- [51] Bunescu R C, Mooney R J. A shortest path dependency kernel for relation extraction[C]. //Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005: 724-731.
- [52] Qian L, Zhou G, Kong F, et al. Exploiting constituent dependencies for tree kernel-based semantic relation extraction[C]. //Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008). 2008: 697-704.
- [53] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces[C]. //Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. Association for Computational Linguistics, 2012: 1201-1211.
- [54] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]. //Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers). 2016: 207-212.
- [55] Wang L, Cao Z, De Melo G, et al. Relation classification via multi-level attention cnns[C]. //Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 1298-1307.
- [56] Gong C, Tao D, et al. Label propagation via teaching-to-learn and learning-to-teach[J]. IEEE transactions on neural networks and learning systems, 2016, 28(6): 1452-1465.
- [57] Gong C, Tao D, Chang X, et al. Ensemble teaching for hybrid label propagation[J]. IEEE transactions on cybernetics, 2017, 49(2): 388-402.
- [58] Zhang Y, Qi P, Manning C D. Graph convolution over pruned dependency trees improves relation extraction[J]. arXiv preprint, 2018, ArXiv:1809.10185.
- [59] Christopoulou F, Miwa M, Ananiadou S. A walk-based model on entity graphs for relation extraction[J]. arXiv preprint, 2019, ArXiv:1902.07023.
- [60] Xu Y, Jia R, Mou L, et al. Improved relation classification by deep recurrent neural networks with data augmentation[J]. arXiv preprint, 2016, ArXiv:1601.03651.
- [61] Miwa M, Bansal M. End-to-end relation extraction using lstms on sequences and tree structures[J]. arXiv preprint, 2016, ArXiv:1601.00770.
- [62] Alt C, Hübner M, Hennig L. Improving relation extraction by pre-trained language representations[J]. arXiv preprint, 2019, ArXiv:1906.03088.
- [63] Wu S, He Y. Enriching pre-trained language model with entity information for relation classification[C]. //Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 2361-2364.
- [64] Li Q, Han Z, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]. //Thirty-Second AAAI Conference on Artificial Intelligence. 2018: 3538-3545.

作者简历及攻读硕士学位期间取得的研究成果

一、作者简历

赵义（1995-）系北京交通大学计算机与信息技术学院计算机科学与技术专业在读硕士研究生，其教育经历如下：

- [1] 2013.09-2017.06，本科阶段，就读于北京交通大学计算机与信息技术学院
- [2] 2017.09-2020.06，研究生阶段，就读于北京交通大学计算机与信息技术学院

二、发表论文

[1] Zhao, Yi, Huaiyu Wan, Jianwei Gao, and Youfang Lin. Improving Relation Classification by Entity Pair Graph. In Asian Conference on Machine Learning, 2019, pp. 1156-1171.

[2] 马语丹，赵义，金婧，万怀宇. 结合实体共现信息与句子语义特征的关系抽取方法. 中国科学：信息科学，2018，48(11)：1533-1545.

三、参与科研项目

- [1] 国家重点研发计划（课题）：诉讼财产保全智能评估及预警技术研究.
- [2] 企业委托研究课题：CSDN 社交网络构建及用户技能与兴趣挖掘研究.

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京交通大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志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：赵义

签字日期：2020 年 6 月 15 日

学位论文数据集

表 1.1: 数据集页

关键词*	密级*	中图分类号	UDC	论文资助
关系抽取; 实体对图; 图卷积网络; 预训练模型; 拓扑特征	公开			国家重点研发计划 (2018YFC0830200)
学位授予单位名称*		学位授予单位代 码*	学位类别*	学位级别*
北京交通大学		10004	工学	硕士
论文题名*		并列题名		论文语种*
基于图卷积的关系抽取方法研究				中文
作者姓名*	赵义		学号*	17120452
培养单位名称*		培养单位代码*	培养单位地址	邮编
北京交通大学		10004	北京市海淀区西直 门外上园村 3 号	100044
学科专业*		研究方向*	学制*	学位授予年*
计算机科学与技术		数据与知识工程	3 年	2020
论文提交日期*	2020 年 6 月			
导师姓名*	万怀宇		职称*	副教授
评阅人	答辩委员会主席*		答辩委员会成员	
	黄慧芳		李宗辉 武志昊	
电子版论文提交格式 文本 (✓) 图像 () 视频 () 音频 () 多媒体 () 其他 () 推荐格式: application/msword; application/pdf				
电子版论文出版 (发布) 者		电子版论文出版 (发布) 地		权限声明
论文总页数*	52			
共 33 项, 其中带*为必填数据, 为 21 项。				