分类号 密级

UDC



学 位 论 文

面向英汉互译文本的共指消解技术及其关键问题研究

|  |  |  |  |
| --- | --- | --- | --- |
| 作者姓名： |  | | |
| 指导教师： |  | | |
|  |  | | |
| 申请学位级别： |  | 学科类别： |  |
| 学科专业名称： |  | |  |
| 论文提交日期： | 2021年6月 | 论文答辩日期： | 2021年6月 |
| 学位授予日期： | 2021年6月 | 答辩委员会主席： |  |
| 评阅人： |  | | |

东 北 大 学

2021年6月

**A Thesis in Control Engineering**

Research on the Technology and Key Problems

of Coreference Resolution for

English and Chinese Intertranslation Text

By

Supervisor:

**Northeastern University**

**December 2020**

独创性声明

本人声明，所呈交的学位论文是在导师的指导下完成的。论文中取得的研究成果除加以标注和致谢的地方外，不包含其他人已经发表或撰写过的研究成果，也不包括本人为获得其他学位而使用过的材料。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谢意。

学位论文作者签名：

日 期： 年 月 日

学位论文版权使用授权书

本学位论文作者和指导教师完全了解东北大学有关保留、使用学位论文的规定：即学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人同意东北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、交流。

作者和导师同意网上交流的时间为作者获得学位后：

半年 □ 一年□ 一年半□ 两年□

学位论文作者签名： 导师签名：

签字日期： 签字日期：

摘 要

随着科学技术的发展，自然语言处理技术被广泛应用于人们的日常生活。共指消解作为理解文本中指代关系的重要技术，受到了学者的广泛关注。由于不同语言系统具有一致性和互补性的特点，基于多语言系统的文本分析使得对文本的理解更全面。因此，近年来，对跨语言自然语言处理技术的研究引起了学术界的重视。针对英汉互译文本的共指消解技术及其关键问题的研究对于获取文本中完善的指代关系并服务于其他跨语言自然语言处理技术的研究有重大意义。因此，本论文围绕英汉互译文本的共指消解任务，着眼于英文共指消解问题和中文零指代词消解问题展开研究并提出了相关解决方案。

对于英文共指消解问题的研究，本文提出了一种新颖的英文共指消解模型。在该模型中，我们采用多种词嵌入模型代替单一词嵌入模型对原始语料库进行联合向量化表示，并提出了一种多通道字段向量化表示方法。为提升模型效率，该模型基于一种交替非线性评分函数对真正先行语的搜索空间由粗粒度到细粒度进行了有效的修剪。此外，我们提出了一种串并形结构的特征提取网络模块 (SPNet) 来构建一种高阶英文共指消解器，SPNet展示了从多个尺度对表述对的语义特征进行有效的提取和融合。最后，本文还探索了表述联合检测和消解问题，提出了一种联合优化目标函数来提升整体模型的精确度。

对于中文零指代消解问题的研究，本文提出了一种前序零指代词消解信息提取模块 (ZPR-ENet) 对前序零指代词的消解信息进行编码。结合SPNet和ZPR-ENet，本文提出了一种零指代词消解模块 (ZP-SPNet)，并以此来构建中文零指代词消解器。此外，基于强化学习的启发并为了进一步优化整体模型性能，本文以平均F1值作为消解器判断的“奖励”，并以回报期望最大化策略对中文零指代词消解模型进行训练。

对于以上提出的模型，本文分别设计实验来验证其有效性。基于同样的数据集和模型性能评价方法，相较于基准模型，本文提出的英文共指消解模型和中文零指代消解模型分别有0.3%和1.63%的提升。此外，为了验证本文提出的各个模块的效果，本文设计了一系列的消融实验来证明这些模块对整体模型性能提升的贡献度。

**关键词：**自然语言处理；深度学习；英文共指消解；中文零指代消解；SPNet；ZP-SPNet

**Abstract**

With the development of science and technology, natural language processing technology is widely used in our daily life. Coreference resolution, as an important technology to understand the anaphora relationships in a text, has been widely concerned by scholars. Because different language systems have the characteristics of consistency and complementarity, text analysis based on a multilingual system makes the understanding of text more comprehensive. Therefore, in recent years, the research on cross-language natural language processing technology has attracted the attention of the academic community. The study of coreference resolution technology and its key problems for English and Chinese intertranslation texts are of great significance for obtaining perfect reference relationships of texts and serving the research of other cross-language natural language processing technologies. Therefore, this paper focuses on the task of coreference resolution to English and Chinese intertranslation texts, focusing on the problems of English coreference resolution and Chinese zero pronoun resolution, and puts forward relevant solutions.

For the research of English coreference resolution, we propose a novel model. In this study, our paper uses multiple word embedding technology instead of one to represent the original corpus and proposes a multi-channel span representation method. In order to improve the efficiency of the model, we effectively trim the search space of real antecedent from coarse to fine based on an alternate bilinear scoring function. What’s more, a Serial-Parallel feature extraction module (SPNet) is proposed to construct a higher-order English coreference resolver. SPNet shows the effective extraction and fusion of semantic features of mention pair from multiple scales. Meanwhile, the problem of joint detection and resolution of the mention is explored in this study, and we propose a joint optimization objective function to improve the accuracy of our model.

For the research of Chinese zero pronoun resolution. We propose a module, called ZPR-ENet, to extract the information of predicted zero pronoun coreference chains. Then, in this paper, SPNet and ZPR-ENet are combined to construct a zero pronoun resolution module, named ZP-SPNet, and uses it to build a Chinese zero pronoun resolver. In addition, inspired by the reinforcement learning, the average F1 value is used as the "reward" for the predictions of resolver, and the strategy of maximizing return expectation is employed to optimize the Chinese zero pronoun resolution model.

For the models proposed above, we conduct a series of experiments to verify their effectiveness. Based on the same datasets and evaluation methods, compared with the benchmarks, the English coreference resolution model and Chinese zero pronoun resolution model have been improved by 0.3% and 1.63% respectively. In addition, in order to verify the effects of the various modules that proposed in this paper, we design a series of ablation experiments to prove the contributions of these proposed modules.

**Keywords**：Natural Language Process; Deep Learning; English Coreference Resolution; Chinese Zero Pronoun Resolution; SPNet; ZP-SPNet

目 录

[独创性声明 I](#_Toc72398148)

[摘 要 II](#_Toc72398149)

[Abstract III](#_Toc72398150)

[第1章 绪 论 1](#_Toc72398151)

[1.1 研究背景 1](#_Toc72398152)

[1.2 国内外研究现状 2](#_Toc72398153)

[1.2.1 英文共指消解研究现状 2](#_Toc72398154)

[1.2.2 中文零指代消解研究现状 3](#_Toc72398155)

[1.3 研究内容及意义 4](#_Toc72398156)

[1.4 内容组织安排 6](#_Toc72398157)

[1.5 本章小结 7](#_Toc72398158)

[第2章 自然语言处理基础模型概述 9](#_Toc72398159)

[2.1 词嵌入技术介绍 9](#_Toc72398160)

[2.1.1 Glove模型 9](#_Toc72398161)

[2.1.2 ELMo模型 10](#_Toc72398162)

[2.1.3 BERT模型 12](#_Toc72398163)

[2.2 深度学习方法介绍 14](#_Toc72398164)

[2.2.1 RNN 14](#_Toc72398165)

[2.2.2 LSTM 15](#_Toc72398166)

[2.2.3 BiLSTM 16](#_Toc72398167)

[2.2.4 注意力机制 17](#_Toc72398168)

[2.3 强化学习方法介绍 19](#_Toc72398169)

[2.4 本章小结 20](#_Toc72398170)

[第3章 基于SPNet的英文共指消解模型 23](#_Toc72398171)

[3.1 英文共指消解任务定义 23](#_Toc72398172)

[3.2 英文共指消解整体框架 24](#_Toc72398173)

[3.3 原始文本向量化表示层 26](#_Toc72398174)

[3.4 表述检测层 29](#_Toc72398175)

[3.5 搜索空间修剪层 31](#_Toc72398176)

[3.6 消解器设计 32](#_Toc72398177)

[3.6.1 SPNet模块设计 33](#_Toc72398178)

[3.6.2 字段表示更新融合层 35](#_Toc72398179)

[3.6.3 T-SPNet模块设计 36](#_Toc72398180)

[3.6.4 预测层 37](#_Toc72398181)

[3.7 本章小结 38](#_Toc72398182)

[第4章 基于ZP-SPNet的中文零指代消解模型 39](#_Toc72398183)

[4.1 中文零指代消解任务定义 39](#_Toc72398184)

[4.2 中文零指代消解模型整体框架 40](#_Toc72398185)

[4.3 中文零指代词表示层 41](#_Toc72398186)

[4.4 中文零指代词消解器 42](#_Toc72398187)

[4.5 前序零指代词消解信息编码层 44](#_Toc72398188)

[4.6 预训练模型 45](#_Toc72398189)

[4.7 本章小结 45](#_Toc72398190)

[第5章 实验与结果分析 47](#_Toc72398191)

[5.1 数据集介绍 47](#_Toc72398192)

[5.2 英文共指消解模型实验设计与结果分析 48](#_Toc72398193)

[5.2.1 模型评价指标 48](#_Toc72398194)

[5.2.2 模型参数设计 49](#_Toc72398195)

[5.2.3 实验结果分析 50](#_Toc72398196)

[5.3 中文零指代消解模型实验设计与结果分析 55](#_Toc72398197)

[5.3.1 模型评价指标 55](#_Toc72398198)

[5.3.2 模型参数设计 56](#_Toc72398199)

[5.3.3 实验结果分析 56](#_Toc72398200)

[5.4 本章小结 59](#_Toc72398201)

[第6章 结果与展望 61](#_Toc72398202)

[6.1 论文工作总结 61](#_Toc72398203)

[6.2 研究方向展望 62](#_Toc72398204)

[参考文献 65](#_Toc72398205)

[致 谢 71](#_Toc72398206)

[攻硕期间的科研成果及获奖情况 73](#_Toc72398207)

1. 绪 论

随着人工智能技术的成熟和计算机软件硬件的发展，人工智能在某些工作上通过人机交互可以辅助人类做出更加正确的判断，甚至能够超越人类。人工智能旨在使得机器可以如人类一样可以实现在线的视觉感知、语言符号理解以及动作输出等工作，如何使得机器可以理解语言符号并帮助人类正确完成某些任务，这是自然语言处理研究的一个重要出发点。人类的语言非常复杂，尤其是具有众多指代性表述的复杂文本可能更令人费解。如果文本中的表示之间相互的指代关系能比较清楚的被标记，对于人类理解一段文本非常重要。并且基于该文本实现下游的任务的解决，如关系抽取[1]、知识图谱构建[2]、自动摘要[3]等具有非凡的意义。通过人工智能技术分析文本中的指代关系，能在一定程度上就帮助人类甚至机器更好、更快、更准确的理解文本内容。于是，本文针对面向英汉互译文本的共指消解技术及其关键问题进行了深入的研究。本章节主要从研究背景和国内外现有的研究手段分析对共指消解研究的必要性。基于前人研究的启发，本文也开创性的提出了自己的研究方案并设计实验进行了验证。

* 1. 研究背景

信息时代的发展促进了计算机软硬件技术的发展，计算机相关技术的发展使得文本信息急剧的积累。由于人类文化传统的不一致以及某些文本信息的复杂性较高，使得一些有用的文本信息无法得到很好的传播和理解。因此如何有效的自动化的筛选有用信息，统一不同地域之间有用的文本信息显得及其重要，众多自然语言处理技术就是在这种背景下产生的。

关系抽取、知识图谱以及自动摘要等作为自然语言处理领域的一些重要研究方向，旨在从多种文本信息提取重要的信息进行概括性描述。诸如关系抽取旨在从非结构化的文本中抽取出不同的实体，并判断实体之间的关系。这些研究方向不局限于从单语言系统中提取重要信息，多种语言系统之间的一致性和互补性对于某些自然语言处理领域相关任务的深入研究更为重要[4-6]。值得注意的是英汉互译文本序列本身在语言上就存在一致性和互补性。此外，共指消解技术作为自然语言处理领域的一个研究方向，其对文本中指代关系的解析对于自然语言处理的众多下游任务的研究至关重要[7, 8]。因此，对英汉互译文本中的共指消解技术及其关键问题的研究具有非常重要的意义和价值。

对于共指消解问题，从语法表达形式上对英文系统和中文系统的语法规则进行解析，英文的表达上更体现出通过多个名词、短语等对现实世界中的同一个参照物进行解释或者描述的特点，这一特点就是共指消解现象。根据Pradhan等人[9]的统计数据，共指消解更多的在英文的表达中体现出来，共指消解早期的研究大都集中于对英文语言系统中共指现象的研究。近年来，国际上也先后出现了MUC，ACE，CoNLL等共指消解相关的国际测评，这足以体现共指消解研究的价值所在。零指代消解是一种特殊的共指消解现象，区别于共指消解是判断多个具有实际表达形式的表述是否指向现实世界中的同一参照体，零指代消解是判断“零指代词”和实际表述是否具有相同的参照体。共指消解个零指代消解都是预测多个共指链，同一条共指链中的零指代词和表述之间都指向现实世界中的同一个参照体。在下游任务中，零指代链可以作为共指链的一种补充来生成表述之间更完整的消解关系，因此对零指代消解的研究是共指消解研究中的重要一环。零指代消解现象更集中的体现在中文语言系统中，根据Kim等人[10]的统计，英文中出现句法成份省略的现象不超过4%，但是在中文中句法成份省略现象却高达36%。那么分别得到英文系统中的共指链和中文系统中的零指代链，进而通过英汉互译文本之间的一致性和互补性为媒介，使用英文共指消解和中文零指代消解分别生成的共指链来分析文本中的指代关系对于下游任务的完成具有重大的意义。

基于以上研究背景，本文对于英汉互译文本共指消解技术及其关键问题的研究主要集中于两个方面：英文系统的共指消解研究和中文系统的零指代消解研究。这两点的研究对于获取文本中更完整的指代关系并服务于后续的跨语言关系抽取、知识图谱等自然语言处理任务的研究具有重大意义。

* 1. 国内外研究现状
     1. 英文共指消解研究现状

针对英文共指消解的研究起源相对较早，早期的英文共指消解模型主要是基于规则来构建。Hobbs最早基于Hobbs理论[11]构建句法分析树来解决指代消解的问题；Grosz等人[12]提出了以角色为载体的中心理论，其基本思想是将表达模式视为语篇的基本组成单元，通过识别表达式中的实体，可以获得当前和后续语篇中的关注中心，根据语义的局部连贯性和显著性，就可以在语篇中跟踪受关注的实体进而得到表述之间的消解关系；以上模型依赖的规则过于冗长，并且需要大量的人工特征。这使得模型的效率非常低下，泛化性不高，从模型的效果来看也有待提升。

随着对共指消解研究的深入以及机器学习的发展，Raghunathan等人[13]提出了一种基于多重过滤框架的共指消解模型，该框架摆脱了传统的单一消解器进行共指性判断的模式，以迭代式方式由高到低构建了多个筛子模型来为待消解的表述选择可能的候选先行语并对表述对之间的共指性进行判断；Ma等人[14]发现英文文章中具有指代关系的两个实体描的文本距离比较近，对于较远的候选实体可以进行修剪，以提高计算的效率，同时把模型总的损失函数分解为了多个进行联合训练；Wiseman等人[15]提出了一种非线性表述排序模型来学习文本原始特征的表达以取代人工特征并用来构建共指消解模型；Peng等人[16]发现表述头在共指消解的实体边缘检测中很重要，因此把检测头单词的模块作为一种辅助结构来构建共指消解模型；Clark等人[17]在前人研究的基础上提出了一个辅助网络来获取文本中实体的分布概率来指导共指消解，根据此概率分布模型可以对表述之间的共指关系做初步的预测。

近年来，随着深度学习的普及，研究这发现深度学习技术可以更好的捕捉表述对深层的特征、语义依赖以及语义相似性。更大规模、更规范的英文指代消解数据集的发布以及词向量化技术的进步也促进了英文共指消解的发展。Lee等人[18]提出了一个端到端的共指消解模型，该模型用Glove词嵌入模型对原始语料库进行向量化表示，并使用前馈神经网络构建消解器来判断表述之间共指性；Kundu等人[19]提出了一种跨语言的共指消解模型，以英语作为训练集来训练模型，而以中文作为测试集来测试，这是首次共指消解的跨语言尝试；Rui等人[20]第一次把Biaffine注意力机制引入到了共指消解模型，基于该注意力机制对每个可能是真正先行语的表述在后续给以更多的关注；Hongming等人[21]发现外部知识库对于表述之间的共指性的共指性判断很很重要，因此将待消解表述对链接到外部智库以帮助共指性判断；Lee等人[22]提出来了一种名为“Coarse-to-fine”的候选先行语修剪方法为待消解表述选择候选先行语，该方法在提升模型精确度的同时进一步降低了模型的复杂度提升了模型效率。

此外，词向量技术的进步也促进了共指消解模型性能的提升，相较于word2vec[23]和Glove[24]词向量，ELMo[25]和BERT[26]模型能更好的动态的构建词的上下文信息。当相同的词具有不同的上下文时，ELMo和BERT模型能够根据上下文得到不同的表示词向量。Knator等人[27]首次把BERT词嵌入模型引入英文共指消解中，并联合Glove模型对文本进行向量化表示。Joshi等人[28]对BERT进行了微调并融合到Lee等人[18]提出的端到端共指消解模型中，该模型在当时达到了英文共指消解领域性能最优。

* + 1. 中文零指代消解研究现状

零代词本身是一种特殊的表述，其没有实际的表示形式，在中文语言系统中比较常见。因此，对中文零指代消解的研究引起了广泛的关注[29, 30]。早期，大都基于规则对中文零指代消解进行研究和模型设计。Converse等人[31]首先对文本的完全句法分析数进行研究，然后基于经典的Hobbs算法[11]提出了一种规则中文零指代词消解方法。近些年来，随着人工智能相关技术的发展以及优秀的人工标注的语料库的产生，机器学习算法被越来越多的应用到中文零指代消解任务中。Fang等人[32]以提出了一种以内容敏感卷积数为核函数的支持向量机，该模型先对句法树进行剪枝，从树中获取句法结构信息，最后训练一个支持向量机对零指代词进行检测和消解；Zhao和 Ng[33]首次将机器学习算法应用零指代消解的问题上，提出了一种基于特征抽取的零指代消解系统； 传统的有监督学习方法需要大量标注数据集，Chen和Ng[34]等人使用无监督方法来解决零指代消解任务，基于最大期望算法，提出使用在人工标注的语料库上训练一个的无监督模型来完成零指代消解任务。

而后，随着深度学习出现，Chen和Ng[35]为了解决传统零指代消解模型需要大量手工特征的问题，从词嵌入模型出发，首次使用深度卷积神经网络来进行特征提取，并在得分层计算对表述对的共指性进行评分；基于Chen的研究成果，Yin等人[36]提出了一种ZPSNN网络，该网络主要探索更好的利用上下文语义信息来对零指代词进行表示，使得向量化后的零指代词具有更丰富的特征；Liu等人[37]提出一种使用伪数据集对原始语料库进行扩充的方法来提升模型的性能，同时其将零指代消解问题转化为Q/A问题来对零指代词进行消解；Yin等人[38]基于LSTM模型[39]提出了ZP-Centered LSTM模型来对中文零指代词进行向量化表示，LSTM能更好的获得零指代词上下文的文本特征；Fang等人[40]基于Chain-to-Chain的方式来解决中文零指代消解问题，该模型中把连续的零指代词聚集在一条共同引用链上，同时每个零指代词也都是一个独立的个体。这样每个零指代词就可以以共指链上的其他零指代词作为桥梁与文本距离较远的候选先行语更好的判断其是否具有可消解关系。Yin等人[41]基于强化学习的思想将中文零指代词的消解抽象为马尔科夫决策过程，抽象出状态、动作和奖励三要素来解决中文零指代消解问题，并且达到了该领域的最优效果。

* 1. 研究内容及意义

基于上述研究背景，本论文主要对英汉互译文本中的共指消解现象及其关键问题进行了深入探索，着眼于英文共指消解和中文零指代消解两个关键问题并提出了相关解决方案并进行实验验证，本文具体的研究内容如下。

(1) 英文共指消解

基于前人的研究，整个英文共指消解任务可以分为三个阶段：原始语料库的向量化表示、表述检测以及共指消解器的设计，其中的重点是共指消解器的设计。本文针对英文共指消解的研究内容包含以下几点内容。

1) 原始语料库的多词嵌入模型联合表示。考虑到静态词嵌入模型和预训练词嵌入模型各自的特点，本文联合多种词嵌入技术对原始语料库进行向量化表示，并词嵌入层后使用编码层对文本的上下文进一步编码。

2) 候选先行语的搜索空间修剪。为了提高模型的效率并且并非所有的表述都是可消解的表述，本文基于表述检测结果对待消解表述空间进行了修剪。为进一步提升模型效率，本文基于一种交替非线性评分函数对当前待消解表述的真正先行语的搜索空间进行了由粗粒度到精粒度的修剪。

3) 共指消解器设计。本文提出了两种多尺度文本表述对语义特征提取模块：SPNet (Serial and Parallel Convolutional Neural Network) 和T-SPNet (Two Channels Serial and Parallel Convolutional Neural Network)。基于此两者本文构建了一种高阶英文共指消解器。该消解器针对输入的表述对可以在不同的尺度对表述对的深层次语义特征、语义依赖以及语义相似性等特征进行提取并用于共指性判断。

4) 字段向量化表示更新。之前阶段的预测结果对当前的预测具有一定的指导性，本文结合注意力机制在消解器的中间阶段对字段的向量化表示进行更新并结合原始的表示形式构建一种字段的多通道表示形式。

5) 表述检测模块和共指消解器的联合优化。传统的英文共指消解模型并没有对表述检测模块的优化给予单独的指导，但是表述检测模块的判断为消解器的共指性预测提供了部分先验信息。因此，本文针对表述检测模块单独设计目标函数并结合消解器的目标函数进行联合优化。

(2) 中文零指代消解

中文零指代消解和英文共指消解在本质上相同，只是消解的对象存在区别。基于对英文共指消解的研究，本文对中文零指代消解的研究内容包含以下几点。

1) 零指代词的表示。不同于英文共指消解中的表述对，零指代词本身没有实际的表达形式，本文基于ZP-LSTM-center模型的启发对中文零指代词进行向量化表示。

2) 前序零指代词消解信息的编码。前序零指代词共指链的信息为当前零指代词-表述对的共指性预测提供了重要的线索，为此，本文提出一种前序零指代词消解信息编码网络 (ZPR-ENet) 对前序共指链消解信息进行编码并作为后续预测的线索。

3) 中文零指代消解器设计。中文零指代消解器本质上是预测零指代词-表述之间的共指性。本文结合SPNet和ZPR-ENet提出了ZP-SPNet来构建中文零指代消解器。同时，基于强化学习的启发，本文以F1值作为消解器预测结果的回报，通过最大化回报期望的优化方法来优化整个模型。

4) 预训练模型的设计。使用预训练模型初始化中文零指代消解器的参数，这种方法使得模型可以更稳定的收敛，性能也得到更好的提升。

以上就是本文研究的主要内容。指代问题在人类的语言系统中是非常常见的一种语言现象，使得机器能够理解指代关系是自然语言处理技术更好服务社会的关键一步。现在的自然语言处理技术，如智能问答、知识图谱、关系抽取、机器翻译等都在人类的日常生活中得以体现，但是效果还有待提升。共指消解作为自然语言处理方向的一个基础问题，其模型性能的提升能够有效的提升其下游模型的性能。跨语言共指消解技术及其关键问题的研究也促进着自然语言处理技术更好的服务于社会大众。同时，本文的研究方案对于其他自然语言处理任务的解决也具有很好借鉴意义。

* 1. 内容组织安排

本文从针对共指消解技术的研究展开，分析了英汉互译文本的共指消解技术中的两个关键问题：英文共指消解和中文零指代消解。在本文中，针对以上两个关键问题提出了解决相关问题的模型框架。针对英文共指消解主要从文本的向量化表示，表述的检测以及消解器的构建三个方面进行研究。同时，本文基于英文共指消解的研究，提出了一种新颖的中文零指代消解模型。针对以上两个模型，后续分别设计实验来验证模型的有效性，并在相同的指标下对比前人的优秀模型。全文章节安排如下。

第一章，绪论部分。该部分主要介绍了英汉互译文本共指消解研究的背景和意义。聚焦于英文共指消解和中文零指代消解两个关键问题，提出了自己的研究内容以及相关解决方案。

第二章，自然语言处理基本模型概述部分。该部分主要介绍了自然语言处理技术中常用的词嵌入技术以及常用的深度学习模型，包括常用RNN[42]、LSTM、BiLSTM[43]的模型结构。此外，本章节还对自然语言处理中常用的注意力机制的原理进行介绍。最后针对强化学习的基本原理做了概述。

第三章，基于SPNet的英文共指消解模型框架部分。该部分主要是对英文共指消解模型整体框架的详细介绍，分层次分别介绍了文本的向量化表示、表述检测模块的设计、表述的表示、先行语搜索空间的修剪、表述对的构建、高阶英文共指消解器的原理和结构。详细介绍了高阶英文消解器中的SPNet、T-SPNet以及基于注意力机制对字段向量化表示的更新融合过程。最后，该章节对表述检测和消解的联合损失函数的设计进行了详细的介绍。

第四章，基于ZP-SPNet的中文零指代消解模型框架部分。该部分基于共指消解的研究成果提出了ZP-SPNet，并以此构建中文零指代词消解器。主要介绍了本文提出的中文零指代消解模型的框架、零指代词的表示、前序零指代词消解信息的编码、模型目标函数的设计以及预训练模型目标函数的设计。

第五章，实验和结果分析部分。该部分主要针对上述两种模型的性能进行验证。两种模型都使用OntoNotes 5.0数据集进行模型的训练和测试。为了证明模型的有效性，本文基于相同的评价手段将；两个模型与同领域的优秀模型进行了对比。此外，为了验证提出的模块的效果，本文设计了一系列的消融实验。

第六章，总结与展望部分。本文在该部分主要总结了本文的工作内容，同时分析了两种模型的不足和以后的改进研究方向。

* 1. 本章小结

本章主要介绍了面向英汉互译文本的共指消解技术及其关键问题的研究背景和相关价值。作为一个具有重大研究价值，并且是大部分自然语言处理技术的上游任务，本文着眼于英文共指消解和中文零指代消解两个共指消解的关键问题，对其历史研究的方法和成果进行了梳理，并且提出了自己的研究思路以及具体的实践手段。此外，本章节对论文整体的内容组织安排做了详细的介绍。

1. 自然语言处理基础模型概述

本章主要介绍本文模型设计中涉及到的基本自然语言处理技术，具体介绍的内容包括以下几个方面。1) 具体阐述三种词嵌入技术的原理；2) 针对自然语言处理领域常用的深度学习模型进行详细介绍；3) 介绍了比较常用的注意力机制；4) 对强化学习的基本概念做出介绍。

* 1. 词嵌入技术介绍

自然语言处理中里程碑式的技术之一就是词嵌入技术。词嵌入技术从形式上是对文本进行向量化表示，把每个词嵌入到固定维度的空间内使每个词语都占据词嵌入空间的某一点，而该点对应原点的坐标向量即为这个词语的词嵌入。现有的词嵌入技术通常是以当前词的其上下文信息为训练数据来对词嵌入模型进行训练，词嵌入模型的副产物一般就是当前词的词嵌入。对于上下文语义近似的单词，他们所对应的词嵌入也是相近的。此外，词嵌入具有可加性，词嵌入的可加性可以表示为词的可加性，例如：“King-Man+Woman=Queen”等。早期的基础词嵌入技术有one-hot、word2vec等，但是后来的研究表明这些词嵌入技术在上下文的语义表达或者区分多义词等方面表现不佳，因此产生了基于共现矩阵的Glove词嵌入技术以及预训练语言模型等。接下来，本小节将对本文使用到的词嵌入技术的原理做出详细的介绍。

* + 1. Glove模型

Glove是一个基于全局词频统计的方式训练得到词向量的词表征工具，它可以把一个单词表达成一个由实数组成的向量，这些向量捕捉到了单词之间一些语义特性，比如相似性、类比性等。我们通过对向量的运算，比如欧几里得距离或者余弦相似度，可以计算出两个单词之间的语义相似性。相对于之前的one-hot、word2vec等词嵌入技术，Glove模型只训练共现矩阵中的非零元素，这有效的利用了原始文本的全局统计信息。

Glove模型的实现主要包括3部分内容。首先，建立一个共现矩阵*X*，该矩阵*X*就是对原始语料库的词频统计。在固定窗口大小下，*X*中的每一个元素的值表示当前单词*i*和上文或者下文单词*j*共同出现的次数。其次，构建词嵌入和共现矩阵*X*之间的近似关系，式 (2.1) 可以近似的表达两者之间的关系。

 (2.1)

上式中，和是最终要求解的词向量；和分别是两个标量。此外，针对Glove模型的损失函数如式 (2.2) 所示。

 (2.2)

上式中，表示权重矩阵，*N*表示语料库词汇表的大小。其中权重函数如式 (2.3) 所示。

 (2.3)

最后训练Glove模型。模型的训练采用了AdaGrad[44]的梯度下降算法，对矩阵*X*中的所有非零元素进行随机采样，学习率设为0.05。最终学习得到的是两个矩阵和就是Glove模型的词嵌入。

* + 1. ELMo模型

word2vec和Glove词嵌入技术的缺点是在一个固定语料库中，对于每一个形式相同但是上下文不同的单词都是用唯一的一个词嵌入表示，也就是在训练完成以后，一个文本中相同的词就具有相同的词向量表征。这本质上是一种静态词嵌入的方式，但是单词的意思又和上下文相关，上下文的不同使得语义信息并不相同，显然静态词嵌入技术对于多义词的处理不符合实际。为了解决这个问题，预训练语言模型被提了出来。



图2.1 ELMo模型结构

Fig. 2.1 ELMo structure

ELMo模型就是一种预训练预压模型，他的做法是只预训练语言模型，但是词嵌入是在使用过程中针对输入的文本序列输出每个词序列对应的词向量，这样词的意思就是上下文相关的了，很大程度上缓解了单词歧义的发生。基础ELMo模型的设计结构主要包括三个部分，此三部分分别是输入层，两层BiLSTM层，模型结构如图2.1所示。

该模型首先使用字符级卷积神经网络来将文本中的词转换成原始词向量，这些原始词向量被输入双向语言模型中第一层。BiLSTM的前向迭代中包含了该词以及该词之前的一些词汇或语境的信息，后向迭代包含了该词之后的信息，这两种迭代的信息组成了中间词向量。这些中间词向量被输入到模型的下一层来得到下一层的中间词向量，最终的词嵌入就是原始词向量和两个中间词向量的加权和。

在模型训练的过程中，假设给定一个长度为*N*的句子*S*=[*t1,t2,…,tN*]，BiLSTM可以通过前向序列和后向序列信息得到当前出现的概率。前向计算方法如式 (2.4) 所示。

 (2.4)

后向计算方法与前向计算方法相同，如式 (2.5) 所示。

 (2.5)

BiLSTM层的损失函数如式 (2.6) 所示。

 (2.6)

上式中，和表示前向LSTM和反向LSTM的参数；是词向量参数；是softmax参数，和两个参数在各个BiLSTM层中是共享的，BiLSMT层的训练就是最大化上述目标函数。ELMo针对每一个*tk*，通过一个*L*层的BiLSTM计算出*2L+1*个表示。过程如式 (2.7) 所示。

 (2.7)

上式中，是对句子中的词直接进行编码的结果；代表。是每个BiLSTM层输出的结果。那么在ELMo模型中得到词嵌入的过程就是把各个BiLSTM层的中间层词向量进行加权求和得到最后唯一的一个词向量表示当前词向量。最终的词向量的表示方式如式 (2.8) 所示。

 (2.8)

上式中，是一个缩放因子，用于把所有的BiLSTM层进行加权求和，最后得到文本序列中所有单词的唯一词向量。

在实际使用ELMo词嵌入技术的过程中，是把ELMo作为一个预训练模型来使用。针对一个自然语言处理任务中文本的向量化，是在本任务解决的过程中直接使用ELMo来得到对应的词向量，同一个单词的上下文不同时，由ELMo预训练模型得到的词向量也是不相同的，这种词向量可以从本质上解决一词多义的问题。并且由于ELMo是一个轻量级的预训练模型，在使用的过程中也是比较方便的并且能够保证比较好的效果。

* + 1. BERT模型

BERT模型的出现是词嵌入技术的一个里程碑，因为BERT的出现，大部分自然语言处理任务的基准水平得到较高的提升。对于BERT，其和ELMo一样都是预训练语言模型，都能很好的解决一词多义的问题，但是由于BERT的结构更加复杂、参数量更大以及训练的语料库更大使得BERT的效果相对于ELMo要更好。BERT模型的设计是基于Transformer[45]的编码器部分进行设计和实现的。BERT的模型类似于ELMo是一个双向语言模型，区别在于ELMo的基础单元是LSTM，而BERT的基础单元是Transformer中的编码器。BERT的模型整体结构如下图2.2所示。其中“Trm”表示BERT的基本组成单元，其本质就是Transformer的编码器。



图2.2 BERT结构

Fig. 2.2 BERT structure

BERT的整个预训练可以分成两部分：Mask Language Model 和 Next Sentence Prediction。对于Mask Language Model，这是BERT模型的一种训练的技巧，可以理解为完形填空，在训练的过程中，要掩盖每一个句子中15%的词，用其上下文来做预测。例如：my cat is white → my cat is [mask]，这个句子中用 [mask] 将white进行替换，然后需要模型对被掩盖的词进行预测其真实值，这就是Mask Language Model的训练过程。此外为了提升模型的效果，在训练的模型的时候并非一直使用 [mask] 来替换目标词，而是在训练模型时采用如下的方式对替换的部分做处理。

1) 80%的时间是采用 [mask]，my cat is white → my cat is [mask]

2) 10%的时间是随机取一个词来代替mask的词，my cat is white → my cat is dog

3) 10%的时间保持不变，my cat is white → my cat is white

该任务的损失函数定义如式 (2.9) 所示。

 (2.9)

上式中，表示被掩盖的词的总数；表示BERT中编码器部分的参数；表示Mask Language Model任务中在编码器接上的输出层的参数；表示整个语料库中词的数量。对于Next Sentence Prediction任务，该任务是给定 [A,B] 两个句子，使用BERT模型来预测这两个句子是否具有逻辑上的上下文关系，也就是B是否是A的下一句。在向量化的过程中会在文本序列的头部添加 [CLS] 标志位，该位上的输出作为预测的结果，然后对整个BERT模型进行训练，该部分的损失函数如式 (2.10) 所示。

 (2.10)

上式中，表示句子对的数量；表示句子预测任务中在编码器接上的分类器的参数；当两个句子具有上下文逻辑上的关系，也就是B是A的下一句时有，否则。以上两个任务可以视为是两个独立的任务，并且可以独立训练，但是BERT模型在训练的时候是把两者结合起来进行训练的。在输入BERT模型 [A,B] 文本序列中，可以使用 [mask] 对 [A,B] 文本序列中的部分单词按照Mask Language Model中的方式进行替换，这样就可以使用联合两个任务共同训练整个BERT模型。BERT的总的损失函数如式 (2.11) 所示。

 (2.11)

上述损失函数是把以上两个子任务的损失函数进行相加。在训练的过程中对两个任务联合优化。此外，BERT模型在处理输入的文本序列时，不同于之前的词嵌入技术的。BERT模型的输入可以是单一的一个句子或者是句子对，但是其实际的输入值包括了三个部分，分别是当前词的词向量；句向量，每个句子有个句子整体的词嵌入对应给每个输入的单词；还有位置向量，这三个部分相加形成了最终BERT模型的输入向量。这种输入方式保证了BERT模型能够更好的结合其上下文，位置和其句法结构得到信息表达更丰富的词向量。

BERT词嵌入技术的出现，使得大部分的自然语言处理任务基线模型性能得到较好的提升。这主要得益于BERT模型能够更好的对当前词的上下文信息进行编码。同时，由于BERT模型是一个预训练语言模型，在使用的时候可以对其进行微调。由于BERT模型比较庞大，但是其每层结构学习得到的文本序列的信息是不一样的，在较低的层次BERT学到的更多的是句法和语法信息，在较高的层次学到的更多的是语义信息，因此在实际使用的时候也可以根据不同的需求调整BERT的结构。

* 1. 深度学习方法介绍

人类理解一段文本中某些单词或者句子实际表达的含义时，通常会联合当前句子的前文或者下文来推敲当前单词或句子要表达的含义，那么计算机对文本的处理如果能够借鉴当前文本序列的前文或者后文信息进行编码，对下游任务的处理将会有很大的益处[46]。那么针对上述问题，RNN给出了比较好的解决办法，但是RNN本身有其不足点，后续又提出了LSTM网模型和BiLSTM模型，这两种模型在一定程度上针对RNN的缺陷提供了解决方案。本节将对这几种模型的结构和原理做详细的介绍。

* + 1. RNN

RNN是较早提出的一种对文本序列信息进行编码并且效果比较好的网络。对于输入的向量化文本序列，RNN能根据时间顺序对每个时序结点的信息进行编码，由于RNN结构的特殊性，能把这些编码后的信息前后链接。由于RNN这种独特的处理时序特征的方式，使得RNN在机器翻译[47, 48]、自动摘要等自然语言处理任务上得到广泛的应用。

随着自然语言处理技术和深度学习的进步，基于RNN设计的模型也比较多。RNN的自身结构更像每个处理单元之间的串联，前后贯通。这种结构使得时序信息在整个RNN模型中可以由前至后依次传递。RNN的基本结构如图2.3示。

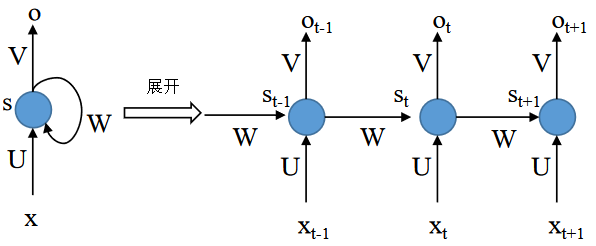


图2.3 RNN结构

Fig. 2.3 RNN structure

整个RNN的前向传播计算相对比较简单明了，在RNN中各个展开结点的关系如下列各式所示。

 (2.12)

 (2.13)

上式中，表示当前时序输入；表示上一个结点的隐藏层状态；表示当前结点的输出；表示共享参数矩阵；表示RNN的激活函数，一般为tanh激活函数；为线性关系的偏倚。

从RNN的结构和前向传播原理来看，RNN非常适合处理时序信息。但是RNN本身的结构也有一定的缺陷。由于RNN的激活函数为，在训练的后期，梯度会变得比较小，如果几个趋于0的值相乘的话，乘积就会变得非常小，就会出现梯度消失现象。在文本信息中的影响就是当RNN的输入时序序列非常长时，起初的文本序列信息在后面序列中的“影响”就会变得很小。此外，RNN本身只能使用前向的文本序列信，但是这并不符合人类的阅读习惯，因为后向的文本序列信息也很重要。针对以上两个问题，后续相继出现了LSTM和BiLSTM模型。

* + 1. LSTM

由于RNN模型存在梯度消失的问题，LSTM被提出来解决RNN的梯度消失的问题。LSTM模型结构主要包括三个门：输入门、遗忘门和输出门。这种门机制[49]也在一定程度上解决了RNN的梯度消失问题。LSTM的整体结构如图2.4所示。



图2.4 LSTM结构

Fig. 2.4 LSTM structure

从整个LSTM的模型上结构看，LSTM除了三个门之外，还有由细胞状态构建的一条“高速公路”贯穿模型始终，前向编码信息在这一条路上只有“乘”和“加”两种操作，这相对于其他门的内部结构简单很多。在LSTM的各个门中，遗忘门的作用主要以一定的概率控制是否遗忘上一层的细胞状态。主要通过遗忘门计算得到遗忘因子，以作为忘记门控来控制上一层的细胞状态对当前细胞层状态的影响。

 (2.14)

上式中，表示sigmoid激活函数，由于sigmoid函数的值域是 [0, 1]，这很好的表示了门的“开”和“闭”，以此来控制之前序列信息对当前细胞状态的影响；

输入门主要负责处理当前序列位置的输入。输入门由两部分组成，第一部分使用了sigmoid激活函数，输出为，第二部分使用了tanh激活函数，输出为，两者的结果后面会相乘再去更新细胞状态。细胞状态由两部分组成，第一部分是和遗忘门输出的乘积，第二部分是输入门的和的乘积。

 (2.15)

 (2.16)

 (2.17)

上式中，代表当前细胞状态。

输出门的作用主要是考虑到有多少细胞中的信息被加入到当前的输出状态中。输出的计算过程可以由式 (2.18) 和 (2.19) 所示。

 (2.18)

 (2.19)

LTSM主要是通过细胞单元构建的“高速公路”来解决梯度消失的问题。在LSTM内部，高速公路贯穿始终，并且在这条路径上只有加和乘操作，梯度是比较稳定的。LSTM内部其他路径上的梯度依旧会发生梯度消失的问题，但是当其他路径发生梯度消失的时候，高速公路上的梯度没有消失，那么远距离的梯度就没有消失，也就缓解了梯度消失的问题。

* + 1. BiLSTM

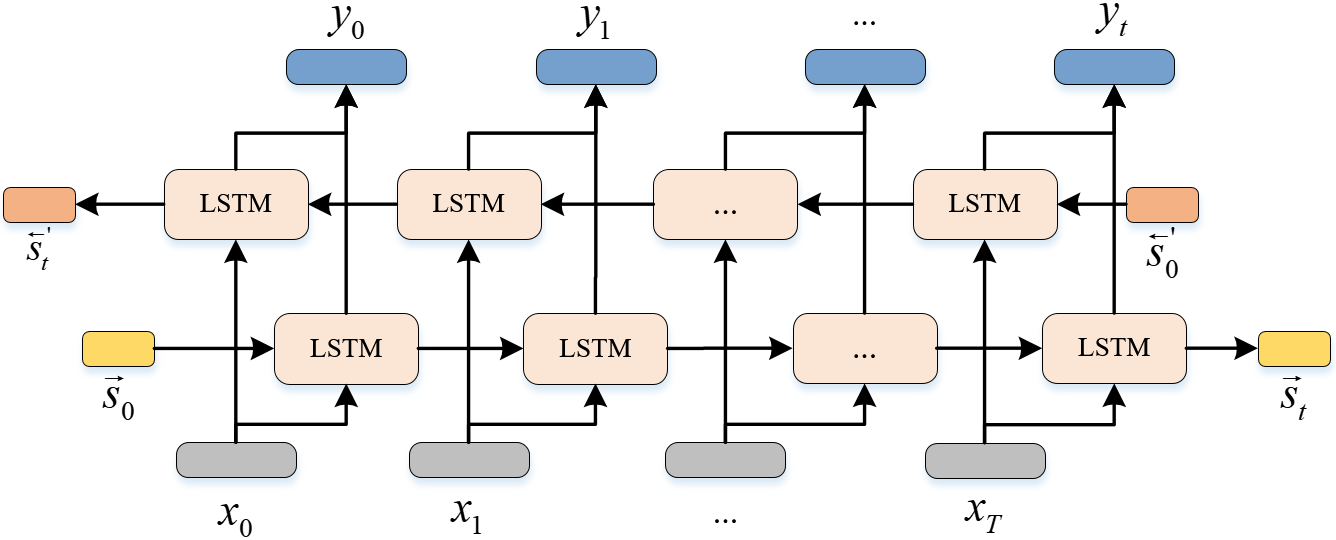


图2.5 BiLSTM结构

Fig. 2.5 BiLSTM structure

LSTM和RNN都只是按照时间序列由前至后对时序信息进行编码，这种编码方式决定着对当前结点编码时参考前序信息，但是后序信息对当前结点信息的编码也有很重要的影响。因此BiLSTM在使用LSTM的正向编码的基础上，把时序信息反向输入到LSTM中，用反向LSTM来得到后序的信息。那么使用BiLSTM网络对当前结点信息进行编码时，其隐藏层状态就由正向LSTM编码的隐藏层状态和反向LSTM编码的隐藏层状态拼接得到。BiLSTM网络的结构图如图2.5所示。

在图2.5可以看出，BiLSTM的正反向结构和LSTM没有什么区别，但是在某一时刻的输出数据是由正向LSTM和反向LSTM的输出共同决定的。在BiLSTM内部，数据的传输形式和LSTM相同，只不过对于其传输方向存在差异。在进行文本信息的处理时，BiLSTM会对文本的特征进行正向提取和反向提取，这样使得提取得到的文本信息更完备并且更加精确。

* + 1. 注意力机制

注意力机制的提出是基于人类观察事物的时候会有针对性的更多关注自己感兴趣的事务，这些被特别关注的事物也就是全局最主要的内容。那么借鉴人类的这种机制，注意力机制在早期被广泛应用于图像处理工作。但是在文本序列的处理中同样也有对注意力需求。比如在机器翻译中把“**Tom** Chase Jerry”翻译为“**汤姆**追逐杰瑞”，那么“Tom”对“汤姆”的翻译至关重要，这就需要机器翻译模型在翻译“汤姆”时更关注“Tom”的作用而不是“Chase”和“Jerry”。随着注意力机制各种变体的出现，注意力机制在自然语言处理中应用也越来越广，并且对模型效果的提升有很大的帮助[50, 51]。

相对于输入，传统的注意力机制是为输入分配一些的权重参数。针对高纬度的输入数据，通过注意力机制可以有效的选取子集以降低输入数据的维度，进而可以有效的降低计算量，让系统能够提取输入的文本序列中与当前输出信息相关度更高的特征，从而能够得到更精确的输出，进而提升模型整体的性能。

传统的注意力机制一般是在编码器-解码器框架中实现。为了探索解码器的输出和编码器的输入之间的关联性，在这个过程中，注意力机制可以被理解为一个寻址的过程，如图2.6所示。



图2.6 注意力机制

Fig. 2.6 Attention mechanism

在计算注意力的过程中，给定一个查询向量*q* (Query)，可以计算与*k* (Key) 的注意力分布并且把注意力分布添加在*v* (Value) 上，这样就可以计算注意力值。在这个过程中，注意力机制在计算输入和输出相关性上也进一步通过注意力值降低了神经网络模型的复杂度，也在一定程度上降低了计算量。通过注意力机制不再需要神经网络对所有的输入做计算，只需要根据注意力分布，从输入的序列中选择和要完成的任务最相关的信息给神经网络来计算。实际使用过程中，注意力机制的计算一般分为如下步骤。

(1) 是原始序列信息的输入。假设输入的信息有*N*个，这里使用*X*来表示。

(2) 是计算注意力分布。如图2.6中的*Key*=[*k1,k2,…,kN*]和*Value*=[*v1,v2,…,vN*]，令*Key*=*Value*=*X*，注意力分布的计算如式 (2.20) 所示，其中称为注意力分布。为注意力打分机制。

 (2.20)

其中打分机制有以下几种。

1) 加性模型

 (2.21)

2) 点积模型

 (2.22)

3) 双线性模型

 (2.23)

4) 缩放点积模型

 (2.24)

(3) 是对原始序列信息加权平均。注意力分布本质上就是在上下文的信息中查询时，第个输入信息被关注的程度。因此再使用信息选择机制对输入的信息*X*进行编码，其计算过程如下式所示。

 (2.25)

注意力机制的本质就是对与输出相关性大的输入信息给予更多的关注，这样可以使的得到的输入更精确。

* 1. 强化学习方法介绍

强化学习[52]主要是确定一个或者多个智能体 (Agent)，智能体在和环境不断的交互过程中，来收获奖励和惩罚，在进行多轮的探索之后，不断的改变原有的策略，进而能够得到一个最优的策略，也就是使得智能体可以学的一个从环境到动作的一个最优的映射，当智能体按照这个策略执行动作时，到达目标点可以获得最大的奖励。强化学习的奖赏信息主要来自环境对智能体所选择动作的好坏给予一种评价，可能是奖赏或者是惩罚。对于智能体来说，一般不会对环境有一个全面的感知，而是在不断的交互过程中去探索环境，根据获得的奖赏或者惩罚信号来改变自己的行为策略，进而更好的去适应环境。强化学习的和环境交互的过程可以由图2.7来表示。



图2.7 强化学习示意图

Fig. 2.7 Sketch map of reinforcement learning

强化学习模型的基本要素包括策略、值函数、回报函数和环境模型。策略，也就是智能体的行为选择概率分布模型，智能体的执行策略决定了智能体在某个状态下应该执行哪种动作才是最优的动作。强化学习模型中策略如式 (2.26) 表示。

 (2.26)

在强化学习模型中，策略仅和当前的状态相关，与历史信息无关。某一个确定的策略是静态的，和时间无关。

回报函数是在智能体在与环境的交互过程中的，环境对智能体的动作好坏的一个评价。对强化学习模型的搭建，如何衡量一个智能体动作的好坏，也就是如何设计回报函数是强化学习模型的一个关键。

对于回报函数考虑的是当前的即时的回报，而值函数考虑的则是一个更加长远的回报，它主要考虑到了智能体选择动作的不确定性和目标的长远过程。而值函数又分为状态值函数和行为值函数。行为价值函数是指某一个状态的价值可以用该状态下所有动作的价值表述。而状态价值函数是指某一个动作的价值可以用该状态后续状态的价值表达。

状态价值函数如式 (2.27) 所示。

 (2.27)

上式中，为折扣因子；和分别为*t*时刻的立即奖赏和环境状态。对于任意策略，*t*时刻系统在状态的值函数如式 (2.28) 所示。

 (2.28)

行为价值函数如式 (2.29) 所示。

 (2.29)

强化学习本身存在诸多的问题，其中模型不能很好的收敛在诸多问题比较突出。尽管强化学习存在诸多的问题，但是强化学习的诸多变体在自然语言处理领域也发挥出了不错的性能，尤其是深度强化学习的出现，使得强化学习的思想在自然语言处理领域中有这诸多的应用[53-55]。在英文指代消解领域中，Clark 和Manning提出使用强化学习来优化Mention-Ranking模型[56]，并且模型达到了当时的领域最优效果；在关系抽取领域，Feng等人[57]基于强化学习的思想构建实例选择器，实例选择器可以更好的选择高质量的句子，并将这些句子输入到关系分类器中，关系分类器对实例选择器进行句子级预测并给予奖励。这两个模块共同训练，以优化实例选择和关系分类过程。整个模型应用在具有噪声的文本序列中达到了比较好的效果；在对话生成领域，Li等人[58]基于强化学习中奖励函数的启发，把整个模型最为智能体。并且利用不同的奖励方式对智能体的动作进行奖励从而学习相应的策略。此外本文采用博弈论的方式设置多个智能体进行博弈学习进而模拟人的真实对话生成新的训练集，这进一步提升了模型的性能；在知识图谱领域，Xiong等人[59]基于借鉴强化学习中的策略梯度来对知识图谱的推理网络进行训练。该模型能够对关系路径中的元素进行预测，并将路径的一些特征编码到模型的奖励中，进而使得模型能够达到更好的推理效果。尽管强化学习本身有一些问题，但是强化学习的相关思想在一定程度上促进了自然语言处理相关任务模型性能的提升。

* 1. 本章小结

本章节主要针对自然语言处理基础技术做了一个比较详细的介绍。本章节首先对自然语言处理中常用的词嵌入技术的原理以及各个词嵌入技术的优势和缺陷做了一个概述。然后，对自然语言处理中常用的模型RNN、LSTM和BiLSTM的模型结构、原理和计算过程做了比较详细的介绍；接着，本章节介绍了自然语言处理技术中比较重要注意力机制的原理；最后，本章节对强化学习的基本概念做了概述。本文针对英文共指消解和中文零指代消解的模型设计就是基于这些常用的自然语言处理技术设计的。接下来两章，本文将提出的模型的结构和原理进行详细的介绍。

1. 基于SPNet的英文共指消解模型

本章节的模型是面向英文共指消解任务而提出。该章节主要对整个英文共指消解模型由整体到局部做出详细的介绍。首先本章节对英文共指消解任务做出了明确定义。其次，本章节详细描述了整个英文共指消解模型框架，然后整个模型框架的各个模块的结构和原理进行了详细的介绍，其中着重介绍了基于SPNet构建的英文共指消解器模块；最后，本章节介绍了整个英文共指消解模型的联合损失函数的设计。

* 1. 英文共指消解任务定义

英文共指消解任务本身是自然语言处理的一个基础性任务，它为机器翻译、关系抽取和智能问答等自然语言语言处理任务提供了表述之间的可消解关系指导。



图3.1 英文共指消解举例

Fig. 3.1 The example of English coreference resolution

图3.1为一段英文文本，具有共指关系的字段被标注了相同的颜色。在上例中，Drug Emporium Inc、Gary Wilber、this drugstore chain、He、the company和company’s都是具体的表述。这些表述的表达形式不同，但是如 [Drug Emporium Inc, this drugstore chain, the company, company’s] 这四种表述指向的是真实世界中同一个参照体“Drug Emporium Inc”；[Gary Wilber, He] 的同一参照体都是“Gary Wilber”，这就是英文中的共指现象。英文共指消解的过程即为文本中的表述确定其在真实世界中所指向的实体的过程[7]，其目的就是使得阅读者能够理解文本中不同的表述之间实际上具有相同的含义。

假设给定一段待消解文本，该文本包含*W*个单词。在给定最大字段跨度的条件下，可以确定文本中所有的字段，每一个字段是一个N-Gram的文本序列，并且每个字段由起始词和结束词进行联合表示。在所有的字段中，某些字段的起始位置可能是相同的，但是他们的结束位置不同，以此来区分字段之间的不同。在确定所有的字段的表示后，通过表述检测模块来检测所有可能是真正表述的字段。为了提升模型的效率和速度，本文提出的模型没有把所有被判断可能是表述的字段作为待消解的对象进行消解。相反，本文设定了一个系数*k*来决定需要消解的表述占所有字段的比例。确定待消解的表述后，需要确定每个待消解表述的候选先行语，假设当前的待消解表述为，每一个都会被分配一个候选先行语，其中，式中表示虚拟的候选先行语，如果当前待消解表述不是可消解项或者当前候选先行语中没有其真正的先行语，那么有，否则有。

但是，如果基于规则把当前待消解表述文本序列之前的表述全部设置为候选先行语，将极大的增加模型的计算消耗降低模型的效率，因此需要为先行语的搜索空间做修剪后确定每个待消解表述的候选先行语。此外，基于前人研究的启发，部分人工特征在共指消解中起着决定性的作用，因此本文把部分人工特征整合到待消解表述-候选先行语对 (，其中) 的表示中对其联合表示。

* 1. 英文共指消解整体框架

给定待消解文本，本文提出的英文共指消解模型可以实现对文本中表述的检测以及表述之间的共指性的判断，最后得到*N*个共指链集合，每个集合包含了在真实世界中指向同一个参照体的*M*个表述。模型整体框架如图3.2所示。



图3.2 英文共指消解模型整体框架

Fig. 3.2 Framework of our English coreference resolution model

首先是原始训练语料的向量化表示。不同于其他语言系统，英文可以直接使用两个字符串中的空格做分词处理，本文使用多种词嵌入技术联合对训练语料库中原始的文本序列进行向量化表示，以使得得到的词嵌入中包含的信息更精确。为了更好的对当前词的上下文信息进行编码，本文在上下文信息表示层中使用BiLSTM网络对当前单词的上下文信息来进一步编码。

其次是真正表述的检测。根据上一阶段得到得向量化文本，基于规则可以得到所有的字段。字段的向量化表示使用其起始位置词的词嵌入、结束位置的词嵌入联合表示。但是，字段的中心表示对字段内部的信息表示更为丰富，结合头词词库并基于自注意力机制可以得到字段的中心表示。本文在此阶利用对得到的字段中心表示联合原始字段表示方法对字段表示重新定义并得到其联合表示方式。为了判断文本中真正表述的概率分布，根据得到向量化的字段，使用表述检测模块来判断当前字段是真正表述的概率。模型中的文本向量化表示层和表述检测层部分如图3.3所示。



图3.3 文本向量化表示及表述检测

Fig. 3.3 Framework of text vectorization representation and mention detection

框架最后是表述对中两个表述之间的共指性判断。表述对包含待消解表述和候选先行语两个字段，该两个字段的可以消解的条件有两个：1) 待消解字段以及候选先行语字段两者都必须为真正表述；2) 待消解字段和候选先行语字段必须存在可消解关系。以上述条件来构建表述对并对表述对进行共指性判断。但是并非所有的表述都是可消解的表述，为了提高模型的效率，要对待消解表述以及其真正先行语的搜索空间进行修剪。对于待消解表述选择，本文从表述检测模块的判断结果中选择分数最高的个字段作为待消解的表述。对于真正先行语搜索空间的修剪，采用一种交替非线性评分函数对候选先行语进行选择以缩小真正先行语的搜索空间。对真正先行语的搜索空间进行修剪后，为每个待消解的表述分配个候选先行语并组成待消解表述-候选先行语对输入共指消解器来判断其共指性。本文提出的消解器的框架如图3.4所示。

本文提出的英文共指消解器是一个高阶消解器。整个消解器框架可以分成三个部分来分析。首先是第一部分，在该部分把经过修剪模块得到的表述-候选先行语对输入到一阶模型中。在该部分，本文提出了SPNet模块，SPNet模块又包含多尺度特征提取模块 (PNet) 和高级语义特征提取模块 (SNet) 两个子模块。SPNet可以对待消解表述对进行内部语义特征抽取并用于后续的评分器。其次是字段表示更新部分。在该部分，本文结合注意力机制利用第一阶的预测结果对字段初始的词嵌入矩阵进行更新。为了更好的保留原有表述对内部的特征信息，本文融合了更新后的词嵌入矩阵和初始化的词嵌入矩阵组成多通道向量化字段表示方法来重新构建表述对。最后是第三部分，该部分主要是以T-SPNet来构建消解器的第二阶，T-SPNet的内部结构和SPNet是相似的，只是该网络的通道数发生了相应的变化。最后此部分的输出得到的就是对共指链的一个精确预测。为了进一步保证得到的共指链的精确性，本文采用四部分的评价得分来判断一个表述对是否可消解。



图3.4 英文共指消解器结构

Fig. 3.4 English coreference resolver structure

以上内容是对本文提出的英文共指消解框架的一个概述，接下来本文将对每个模块的细节进行详细的介绍。

* 1. 原始文本向量化表示层

本文使用的是公开的数据集OntoNotes 5.0对模型进行训练和测试，原始数据集在结构上无法构建语料库，因此在使用该公开数据集之前，本文先把该数据集转化为json格式的语料库，转化后的语料库其内容如图3.5所示。

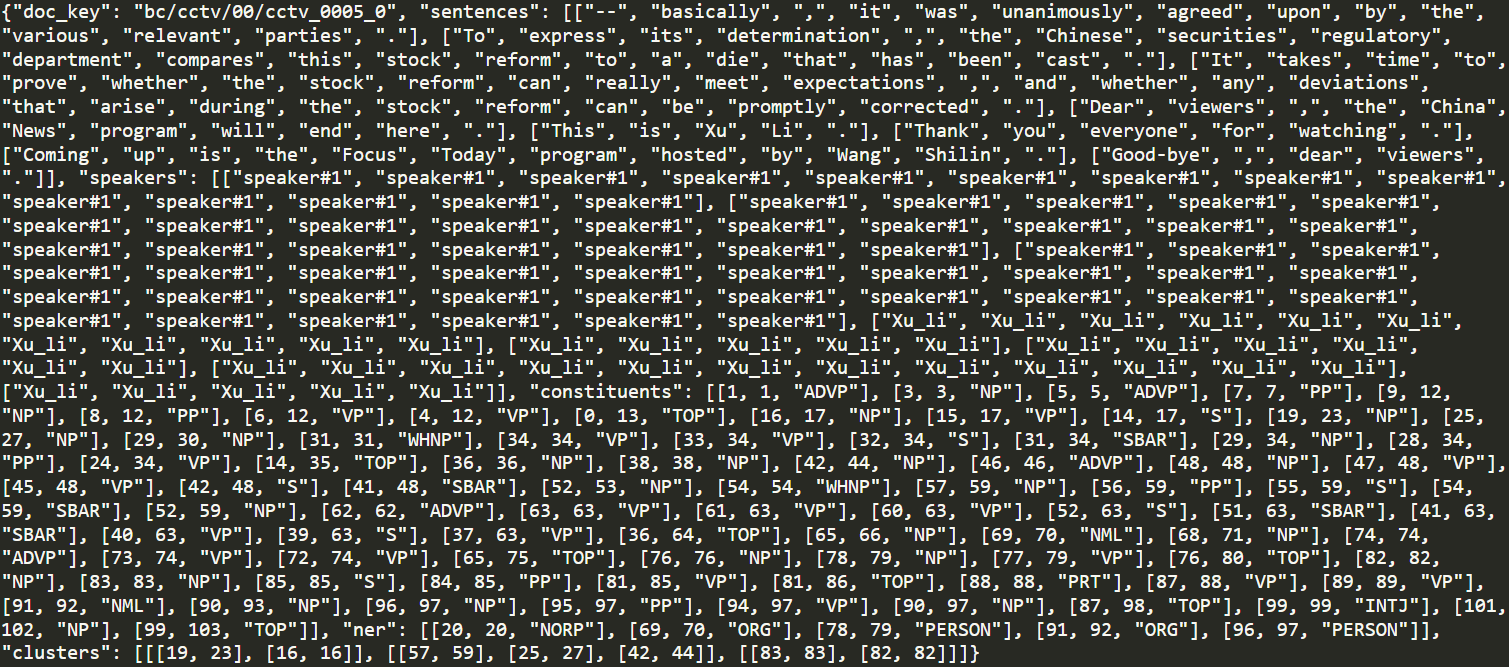


图3.5 格式化后的语料库

Fig. 3.5 JSON format corpus

原始数据集的处理后生成训练集、测试集、交叉验证集三部分语料库。以训练集为例，该部分语料库一共包含2802个字典，每一个字典是原始数据集中对应的一个文本段落。每个字典包含文本路径、文本序列、讲述人信息、表述词性以及具有共指性的表述索引集合标签五部分内容。由于英文的分词是以空格进行分词，语料库中的文本序列是已经分词后生成的文本列表。

本文使用Glove、ELMo和BERT三种词嵌入模型来联合对文本序列中的单词进行向量化表示。首先，对于Glove词嵌入模型，由于Glove本身不是一个预训练模型，使用文本的语料库对Glove模型重新训练的效果并不如原始Glove模型的训练结果。因为原始训练语料库的大小在很大程度上决定了Glove模型生成的词向量在空间上的信息完整性。并且针对Glove模型，无论当前单词的上下文是否相同，其对应的词向量并没有任何形式上的改变，因此本文直接使用预训练的Glove词向量表初始化语料库中的文本序列。假设一个文本序列包含*T*个单词，那么得到的词嵌入表示如式 (3.1) 所示。

 (3.1)

对于ELMo模型，其本身是一个预训练词嵌入模型。本文使用ELMo模型提供的预训练模型参数结合本文使用的语料库重新生成其对应的词向量。本文生成的ELMo词向量存储为一个.hdf5文件，在模型中加载使用ELMo向量时可以利用查询的方式得到每个句子中每个词对应的ELMo词向量，其得到的词向量如式 (3.2) 所示。

 (3.2)

对于BERT词嵌入模型，本文采用和ELMo相似的方法生成语料库对应的词向量。相对于ELMo，本文使用的是BERT-Large预训练模型，该模型的原始语料库更为全面，并且BERT-large本身的结构也保证了生成的词向量包含的上下文信息更加的丰富，在空间中对应的词向量信息更精确。语料库中单词得到的BERT词向量如式 (3.3) 所示。

 (3.3)

最终的原始语料库单词对应的词向量是把Glove、ELMo和BERT分别得到的单词的向量化表示形式进行拼接，其对应的向量化表示形式为，如式 (3.4) 所示。

 (3.4)

在模型框架中的文本向量化表示层的具体结构如图3.6所示。

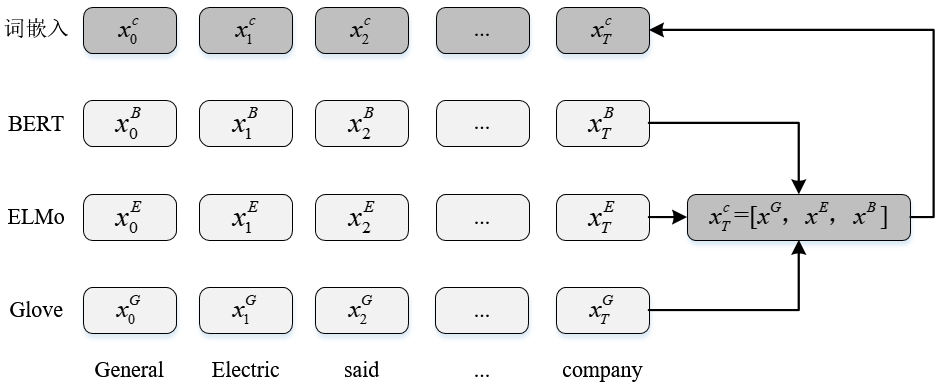


图3.6 基于多种词嵌入模型的联合文本向量化表示

Fig. 3.6 Text vectorization based on multiple word embedding model

在英文共指消解模型中，首先需要确定文本中的表述。对于一个文本序列，该序列中的第一个单词对于确定该文本序列是否是头词具有很大的辅助性作用[16]，本文在对每一个单词获取其有Glove、ELMo和BERT组合得到的词向量的同时，也使用头词库通过检索的方式得到每个词对应的头词中心表达形式。

 (3.5)

使用多种词嵌入模型对原始语料库中的文本序列进行联合向量化表示后，每个单词的相关信息被嵌入到向量空间，为了保证每个单词在其所处的上下文环境中的词嵌入信息更加完整，本文使用BiLSTM对得到的词向量进一步编码。过程如下列各式所示。

 (3.6)

 (3.7)

 (3.8)

 (3.9)

 (3.10)

 (3.11)

上式中，，其中1表示正向LSTM，-1表示反向LSTM；表示激活函数sigmoid；表示经过BiLSTM编码后的文本序列中每个词的词向量。

* 1. 表述检测层

针对语料库中所有文本序列中的单词进行向量化表示后，需要检测文本序列中真正可消解的表述。本文首先基于规则得到一个文本序列中包含的所有字段。为了保证模型的效率以及精确性，本文对一个字段包含的最大的单词数量限制在一定的词跨度范围以内。从文本的句法上看，一个真正表述的组成元素不可能位于两个句子中。因此本文的前提假设是所有字段的组成元素需要是同一个句子中的单词。对于一个字段的表示，本文使用一个字段的起始位置单词的词向量表示*starti*和结束位置单词的词向量表示*endi*以及其他特征进行联合表示。

表述和字段之间的判定关系是一个充分非必要条件，所有的表述也并非都是可消解的表述。判断一个字段是否为真正的表述，字段的中心表示信息具有非常重要的价值。结合头词词库以及自注意力机制[45]可以捕获字段内部的特征对联合对字段进行表示。基于自注意力机制，本文采用一种头部注意力机制来计算一个单词标准化的重要性得分，然后通过对当前字段所有组成元素的得分进行加权求和以学习当前字段的中心表示。假设组成当前字段*i*的文本序列起始位置词的词向量表示为，结束位置词的词向量表示为。结合头部注意力机制，可以得到一个向量，该向量代表了当前字段*i*的中心表示，计算公式如下所示。

 (3.12)

 (3.13)

 (3.14)

上式中，表示本文采用的前馈神经网络；表示自动学习得到的权重，他的定义和传统的头词的定义密切相关；表示当前字段的每个单词在头词语料库中对应的词向量；表示当前字段*i*中每个单词向量化表示的加权和，也是其中心表示信息。此外，针对字段的表示，本文考虑到部分人工特征和字段本身特征的重要性，结合上述内容，本文使用字段的起始位置词的词向量、结束位置词的词向量、表述概率以及其他相关特征来联合表示当前的字段*i*。

 (3.15)

上式中，表示当前字段的向量化表示。如何表示字段在英文共指消解中起着决定性的作用，但是字段是基于规则的方式来提取，其中基于规则得到的字段中并不是都是表述，本文采用前馈神经网络 (FFNN) 构建评分函数，来为当前字段*i*是否为表述做出判断。

 (3.16)

上式中，是判断当前字段是否为一个表述的评分；是由前馈神经网络构建的表述检测模块；是网络的参数。对于本文采用的前馈神经网络，其结构相对比较简单。本文使用两层隐藏层组建前馈神经网络，每一层包含150个神经单元，输出层同样采用一层线性层组成。

并非所有的表述都是真正的可消解表述，为了提高模型的效率，根据表述检测模块的结果对待消解表述空间进行修剪是有必要的。假设输入的某个连续文本序列包含*T*个单词，字段最大跨度设为*T*，那么得到的可能的字段的数量将是，假设所有的字段都为表述，这个数量级的表述的消解对于整个模型的性能以及效率将是一个极大的挑战。但英文文本中存在共指消解语言现象的表述在整个文本中的分布相对比较稀疏，因此对所有的字段进行消解并没有现实的意义，并且为了尽可能提升共指消解模型的效率，本文根据表述检测的结果选择部分表述作为待消解的表述，使用系数来控制待消解的表述占整个文本得到的字段的比例，使用表述检测模块计算的来确定待消解表述的数量。下式中，*K*表示修剪后总的待消解表述的数量。

 (3.17)

根据以上公式确定的待消解表述的数量以及来选择得分最高的字段作为待消解表述进行消解。从上述内容可知，表述检测模块的目的是从基于规则提取的字段中尽可能精确的提取出可能的表述。共指消解器的目的是判断两个表述是否具有共指消解性，保证消解器精确度的前提是保证消解器的输入是真正的表述。因此表述检测模块在一定程度上决定了整个共指消解模型的精确性。基于以上所述，本文提出把表述检测模块和英文共指消解器联合优化。根据数据集提供的标准表述，本文使用表述检测模块的预测结果设计的交叉熵损失函数如式 (3.19) 所示。

 (3.18)

 (3.19)

上式中，表示所有可能的字段的数量；表示表述检测模块的预测值；是数据集提供的标准表述标签，如果是数据集中提供的标准表述，；否则有。该损失函数与英文共指消解器的损失函数加权相加作为整个模型总的优化目标函数。

* 1. 搜索空间修剪层

假设当前待消解句子中确定了*K*个待消解表述，如果考虑所有的表述为当前表述的候选先行语，那么对于消解器要计算的张量的量级将会达到。当待消解的文本序列较长时，基于规则确定的字段的数量也会增加，那么张量的计算量级会大幅增加，这将使得模型的计算效率急剧下降。因此，为了降低复杂度并提高模型的效率。 本文采用一种交替非线性评分函数对候选先行语的搜索空间进行修剪。利用修剪后得到的候选先行语构建待消解表述-候选先行语对，这种操作在一定程度上提升了模型的效率。假设为每个待消解的表述分配*C*个候选先行语，其中当表述的数量少于预分配到候选先行语的数量时，本模型将考虑当前表述前的所有表述为其候选先行语。反之本文采用一种交替非线性评分函数的方法为每个待消解表述提取可能的*C*个候选先行语。

 (3.20)

该交替非线性评分函数如式 (3.21) 所示。

 (3.21)

上式中，是候选先行语修剪过程中可以学习的权重矩阵；是对当前字段是否为候选先行语的评分；表示由表述检测模块判断为真正表述的字段。是粗粒度的候选先行语计算方法，根据除去最不可能是先行语的表述，得到一个细粒度的候选先行语集合，该集合中的表述和待消解表述组成待消解表述对，由消解器判断两个表述的共指性。

利用经过搜索空间修剪后得到的候选先行语与待消解的表述构建待消解表述-候选先行语对，每个待消解的表述和其候选先行语可以表示为待消解表述集合，如下列各式所示。

 (3.22)

 (3.23)

上式中，表示当前批次中所有待消解的表述-候选先行语对；表示当前待消解表述及其候选先行语构建的待消解表述对集合。对于表述对的表示，本文采用式 (3.24) 所示的表示方法。

 (3.24)

上式中，和分别表示当前待消解表述及其候选先行语的向量化表示；表示逐元素相乘算子，以计算当前待消解表述和其候选先行语的相似性；表示对一些人工特征 (说话者ID、字段跨度、字段距离以及原数据类型) 的编码向量矩阵。

* 1. 消解器设计

整个英文共指消解模型的核心内容是消解器的设计，消解器的输出精确的表示了两个表述之间是否指向真实世界中的同一个参照体。从模型本身出发，消解器是旨在学习一个条件概率分布以预测文本序列中具有共指性的表述集合，也就是共指链。具体如式 (3.25) 所示。

 (3.25)

上式中，表示消解器对文本序列*D*中的字段*i*和字段 *j*之间存在共指性概率的精确预测；表示当前表述的候选先行语集合。基于以上学习目标，本文提出了一种比较新颖的英文共指消解器。该英文共指消解器是基于卷积神经网络构建的二阶消解器，旨在能够挖掘待消解表述-候选先行语对表示中的更多高级特征辅助进行表述对之间的共指性判断。为了能够挖掘更多的潜在高级特征，基于传统卷积神经网络的启发，本文提出了SPNet。该模块宏观上看是一个双流网络，能够对不同尺度的文本特征进行提取并进行特征融合。由于词嵌入本身可以看做是一个单通道的输入，SPNet内部结构都保持为对单通道输入的计算。本文提出的消解器是一个二阶消解器，整个SPNet是消解器的一阶子模块。基于SPNet，本文提出了T-SPNet模块来构建消解器的二阶子模块，T-SPNet的结构和SPNet保持一致。根据一阶段的输出，本文结合注意力机制对表述的词向量表示做出更新，并结合原始表述的词向量和更新后的词向量共同组成多通道词向量表示表述对。整个消解器的结构如图3.7所示。



图3.7 英文共指消解器

Fig. 3.7 English coreference resolver

该英文共指消解器第二阶的输出是对输入的表述对之间共指性的一个精确预测，但是为了保证预测结果的精确性，本文结合之前表述检测模块的输出来使得最后的预测结果更准确。此外，对于整个模型的目标函数设计，本文结合表述检测模块的目标函数和英文共指消解器的优化函数来组成整个模型的目标优化函数。从最后的实验结果来看，这种方式的目标函数的设计在一定程度上有助于模型性能的提升。接下来本章节将对消解器的各个模块进行详细的介绍。

* + 1. SPNet模块设计

对于SPNet模块，其构成了本文提出的消解器的第一部分，同时也是消解器的第一阶。对于SPNet 模块本身的结构，其包含两个子模块：1) 多尺度特征提取模块 (PNet)；2) 高级语义特征提取模块 (SNet)。PNet是多个卷积层以并行的方式对原始输入进行特征提取，其卷积核的不同也决定了其提取特征的粒度不同；SNet是多个卷积层以串行的方式对输入的特征进行挖局，尽可能的提取输入的高级特征。SPNet的整体结构如图3.8所示。接下来，本章节将深入介绍SPNet各个模块的结构以及作用原理。



图3.8 SPNet结构

Fig. 3.8 SPNet structure

首选对于PNet，其本身是一个多尺度语义特征提取网络，本文输入的待消解表述对的向量化表示主要是由四部分组成，分别是：1) 待消解表述的词向量表示; 2)候选先行语的词向量表示；3) 待消解表述及其候选先行语的相似性度量；4) 其他人工特征。利用上述四部分，本文得到待消解表述-候选先行语对的向量化表示。类比于图像处理中通道的概念，词向量本身可是一个只有一个通道的数据，那么有，这也就是第一阶的输入向量。考虑到不同特征粒度包含的特征信息不同，从而对最后预测结果有不同的影响。为了更好的提取低级特征信息和高级特征信息，PNet使用多个不同尺寸的卷积核对原始输入做特征提取。

从图3.8可以看出，PNet模块使用多个不同尺寸的卷积核对原始输入进行多尺度的特征提取，出于提升模型效率的考虑，PNet主线主要使用3层卷积操作，3层最大池化操作。在本文设计的PNet中，每个卷积层后面都添加了ReLU激活函数来加速网络的训练，同时也有效的防止了梯度消失的问题。为了联合使用多尺度信息来提升模型预测的准确性，PNet提取了每一层最大池化层的输出特征图并进行拼接得到当前特征图。

 (3.26)

 (3.27)

上式中，表示最大池化操作；表示卷积操作；表示不同尺度下得到的特征图；表示多尺度融合特征图。在得到多尺度融合特征图后，PNet中使用一层线性层对其高级语义特征进一步提取得到最后的特征图。式 (3.28) 中，表示ReLU激活函数。

 (3.28)

受卷积神经网络在图像处理中的应用的启发，高级语义特征对于判断待消解表述-候选先行语对中两个表述之间的共指性具有至关重要的作用，本文在使用PNet提取输入原始词向量多尺度特征的基础上，SNet使用多层卷积层串行对提取原始输入的高级特征。类似于PNet，SNet在每一层的卷积层后使用ReLU函数来加速网络的训练，同时防止梯度消失的发生。在每层的卷积层后SNet同样使用了最大池化。在最后一层池化层后，进一步使用线性网络对特征进行提取。

 (3.29)

 (3.30)

 (3.31)

对于SNet和PNet得到的特征，本文采用直接相加的形式对两者进行融合，融合的结果就是SPNet的输出。

 (3.32)

以上就是SPNet模块的原理。再从模块的结构上看，SNet与PNet以一种双流网络的形式组成SPNet，保证了特征提取的完整性，这对于共指性的判断至关重要，后续的实验也验证了这一想法的正确性。

* + 1. 字段表示更新融合层

单词向量化表示的更新在一定程度上能更使得后续的预测模块得到更精准的预测结果。基于SPNet的输出，结合注意力机制，本文在第一阶后对候选先行语的词向量表示以及当前待消解先行语的词向量表示进行更新。SPNet模块本质上是一个评分机制模块，其输出是用一个分数来初步的确定待消解表述和候选先行语之间的共指性，但是第一阶的输出是一个不精确的输出，但是其在词嵌入空间中也初步确定了一个粗略的共指链，这条共指链是为后续的共指消解链的预测提供了部分先验信息，所以对这部分先验信息的编码是十分重要的。本文结合注意力机制来使用这部分先验信息对原有的词向量信息进行更新以使得后续的预测能够更多的关注粗略的共指消解链中先行语的概率分布。针对SPNet模块的输出，本文使用一层评分层先得到粗略的共指链概率分布信息。

 (3.33)

上式中，本身是一个分数，它表示基于SPNet由评分层得到的粗略的共指概率。针对该评分层的得到的分数，基于注意力机制，本文将其转化为概率权重并将这种概率分布信息嵌入到待消解表述的词嵌入空间中。在模型训练的每迭代的过程中，该模型相当于先去预测得到的每一个待消解表述真正的先行语，那么在接下的预测任务中，如果候选先行语和当前待消解表述在特征上取得一致性，模型就能准确的预测当前候选先行语是当前待消解表述的真正先行语。

 (3.34)

 (3.35)

上式中，表示得到的注意力权重；表示当前待消解实体的候选先行语的个数；表示带有粗略共指消解信息的表述词嵌入，也是期望的真正先行语的词向量表示。在得到期望的先行语词嵌入表示后，该模型结合注意力机制和对原有的待消解实体的词向量进行更新。在更新模块中，本文使用前馈神经网络再一次验证当前待消解表述和候选先行语的表述概率分布。该前馈神经网络和表述检测中使用的前馈神经网络共享参数，以保证当前消解的两个字段是真正的表述。并且文本以该前馈神经网络的预测结果作为参数对当前的原始词嵌入向量进行更新。整个字段表示的更新过程如下列各式所示。

 (3.36)

 (3.37)

上式中，表示更新后的表述词向量。本文考虑到在词向量更新的过程中，可能会忽略原始词向量空间中所包含的部分特征，因此本文考虑使用两通道的方式来表示检测到的表述，一个通道是原始的表述词向量，另一个通道是更新后的表述词向量。更新后的字段表示的定义如式 (3.38)，更新后的待消解表述-候选先行语对的向量化表示的定义如式 (3.39) 所示。

 (3.38)

 (3.39)

* + 1. T-SPNet模块设计

对于T-SPNet，其构成了本文提出的消解器的第三部分，同时也是消解器的第二阶。T-SPNet在结构上和SPNet的机构是相同的，相对于SPNet的输入的实体表示是单通道的表示，T-SPNet的输入是双通道的表示，因此T-SPNet尽在模型架构的通道数上做出了改变。T-SPNet本身也是由两个自模块组成，分别是T-SNet和T-PNet。T-SPNet对输入进行的特征提取操作本文使用T-SPNet进行表示，如式 (3.40) 所示。

 (3.40)

 (3.41)

上式中，表示经过T-SPNet网络得到的高级语义特征；表示基于由评分层对当前待消解表述和候选先行语之间共指性的一个精确评分。

* + 1. 预测层

对一个表述对中的两个表述的共指性进行判断的前提是两者都必须是表述，由这点考虑出发，同时也是为了进一步提升模型的精确性，该共指消解模型判断一个候选先行语是否为当前待消解表述的真正的先行语的指标由四部分组成：1) 当前待消解对象是否是是一个表述 ()；2) 当前候选先行语是否是一个表述 ()；3) 由粗粒度到细粒度对先行语搜索空间进行修剪时对当前表述是否为先行语的初步判断 ()； 4)消解器的最终判断结果 ()。因此，最后的判断分数组成如式 (3.42) 所示。

 (3.42)

上式中，表示当前候选先行语是真正先行语的得分。在上式中，如果当前的待消解表述是一个不可消解的对象时，有，且。在预测层中，模型将选择得分最高的候选先行语作为当前待消解表述的真正先行语，从而生成最终的共指链。

针对整个模型的优化过程，本文采取对消解器和表述检测模块联合优化的策略。表述检测模块优化目标函数的设计如式 (3.43) 所示。那么针对消解器损失函数的设计，本文采用对数似然函数值来设计其优化目标函数，如式 (3.34) 所示。

 (3.43)

上式中，*N*表示待消解表述的个数，表示在数据集提供的标签中和字段*i*具有共指性的字段*j*的集合。本文对表述检测的损失函数和消解器的损失函数加权求和得到总的优化目标函数如式 (3.44) 所示。

 (3.44)

上式中，*L*表示总的优化损失函数；表示权值

* 1. 本章小结

在本章从整体框架到局部模块详细介绍了本文提出的英文共指消解模型。本章节首先介绍了英文共指消解任务的定义以及模型的整体框架结构。首先，3.3节介绍了基于多种词嵌入技术对原始语料库的联合向量化表示方式；3.4节和3.5节详细介绍了表述检测模块、先行语搜索空间修剪的模型细节以及工作原理；3.6节详细介绍了共指消解器的构建，并重点介绍了SPNet和T-SPNet的作用原理。此外，该章节也详细介绍了一种有效的字段表示的更新方式。在后续章节的实验部分，本文对该英文共指消解模型的整体性能以及各模块的有效性进行了充分的验证。

1. 基于ZP-SPNet的中文零指代消解模型

中文零指代消解是一种特殊的共指消解任务。相较于英文共指消解任务中待消解的表述对中两个字段都为具有真实形态的表述，中文零指代消解中的表述对中的两个字段一个为表述，另一个为没有实际表达形式的零指代词。在英文共指消解的研究基础之上，本文提出了一种新的网络结构ZP-SPNet，并基于ZP-SPNet提出了一种新颖且有效的中文零指代消解模型。首先，本章节对中文零指代消解任务做出明确定义；其次，本文详细描述了本文提出的整个中文零指代消解模型框架。最后，本章节对模型的设计和作用原理做了详细介绍。本章核心的内容主要包括三点：1) 本文提出了一种前序零指代词消解信息编码网络ZPR-ENet来对前序的中文零指代词消解信息进行编码并应用于后续的中文零指代词可消解性判断；2) 基于SPNet并结合ZPR-ENet设计中文零指代消解器；3) 基于强化学习的启发，使用最大化回报期望的策略来设计模型目标函数并训练模型以提升模型性能。

* 1. 中文零指代消解任务定义

完整的中文零指代消解任务主体包含两方面的任务，首先是中文零指代词的检测和表示；其次是中文零指代词的消解。但是，中文零指代词的消解更为关键[36]。本文也更注重中文零指代词的消解工作。零指代消解是一种特殊的共指消解，并且这种语言现象在中文语言系统中更普遍。图4.1中列举的两个例子展示了中文零指代消解这种现象。



图4.1 中文零指代消解举例

Fig. 4.1 The example of Chinese zero pronoun resolution

在图4.1的两个句子序列中，、和表示文本中的三处零指代词。从文本的表达形式上看，上述句子中的零指代词并没有具体的文本表示形式。针对 [例1]，结合该句子的上下文，可以比较容易的理解出和 [警方] 是指向现实世界中的同一参照体，那么和 [警方] 就构成了中文零指代消解。同理在 [例2] 中，和[俄罗斯]都表示真实世界中的“俄罗斯”。相对于和，在语义理解上也是有所指代，比如“警局”等实际实体，但是在其所在为文本序列中并没有同其他表述构成消解关系。以上例子中、和都称之为零指代词，他们和句中的某些表述指向真实世界中的同一个参照体。

中文零指代词的消解任务和英文共指消解任务的目的相同，只是消解的对象存在区别。对于中文零指代消解，假设给定一个可消解的中文零指代词及其候选先行语，其目的就是在其候选先行语中找到当前零指代词的真正先行语。对于整个文本，可能存在*N*个共指链。针对以上定义，本文提出了一种新颖且性能较为卓越的中文零指代消解模型。

* 1. 中文零指代消解模型整体框架

针对中文零指代消解的研究，在模型结构上本文主要关注两个方面的内容，分别是：1) 中文零指代词的表示；2) 中文零指代消解器的构建。模型整体框架如图4.2所示。



图4.2 中文零指代消解模型框架

Fig. 4.2 Chinese zero pronoun resolution model structure

假定给定可消解中文零指代词以及文本中可能的表述。由于中文零指代词本身没有具体表示形式，本文采用其上下文对零指代词进行向量化表示。在上述框架中，本文首先使用word2vec词嵌入模型对原始的中文训练语料进行向量化表示。然后结合零指代词的上下文对零指代词具体化，在该过程中，本文基于LSTM网络对其前文正向编码，对其下文反向编码，对得到的结果拼接以完成对中文零指代词的向量化表示。这也就是中文零指代词表示层的工作。

在候选先行语表示层，主要是对原始语料库中的表述进行表示，基于英文共指消解的研究，本文也采用其起始位置词的词向量、结束位置词的词向量以及其他人工特征构建表述的向量化表示形式。

对于共指链的预测。首先，本文基于英文共指消解研究中提出来的SPNet模块来构建了中文零指代消解器。在研究过程中我们发现前序得到的中文零指代词消解信息，对于后续的中文零指代词-候选先行语对的共指性的判断提供了重要的线索。因此，本文提出了ZPR-ENet对前序中文零指代词消解信息进行编码并加入到后续的共指链预测中，并结合SPNet和ZPR-ENet构建了ZP-SPNet模块来设计中文零指代消解模型的主体。此外，基于强化学习的启发，本文使用F1值作为消解器预测的奖励，并采用梯度优化策略使得奖励期望最大化来优化整个中文零指代消解模型。

接下来，本章节将对上述内容设计到的模块进行详细的介绍。

* 1. 中文零指代词表示层

给定待消解零指代词及其上下文序列。首先，基于word2vec词嵌入模型得到向量化表示的原始文本序列，接下来是对零指代词进行表示。零指代词本身是文本中的一个空隙，并没有实际的表达形式，只在人类的语言的理解层面存在其真实的表示，但是要实现对零指代词的消解，首先要得到其真实的表达形式才能由消解器判断其真正的先行语。基于Yin 等人[36]的启发，本文采用LSTM对中文零指代词进行表示。假设给定中文零指代词的上下文文本序列包含*N*个字以及一个零指代词，那么该段文本经过词嵌入模型得到的向量化的表示为。使用LSTM对当前零指代词的上文和下文分别进行正向和反向编码得到当前待消解零指代词的表示，对中文零指代词的上下文进行编码的模型结构如图4.3所示。



图4.3 中文零指代词表示模块

Fig. 4.3 Chinese zero pronoun representation module

在上述中文零指代词表示模型中，使用LSTM分别对当前零指代词的前文正向编码，对其下文反向编码。针对编码的结果进行拼接得到当前零指代词的表示。零指代词本身的意义是存在于上下文中，使用上下文的信息来表示当前零指代词是比较合理的一种方式。在本文的中文零指代词表示中，选择零指代词的*N*个连续上文文本序列以及*N*个连续的下文文本序列对其进行编码并联合表示中文零指代词。

 (4.1)

 (4.2)

 (4.3)

在上述公式中，*wi*表示零指代词的上文；*wj*表示零指代词的下文。对上下文的编码结果和进行拼接得到最后的中文的零指代词向量化表示。确定当前零指代词的表达形式后，本文使用Chen和Ng[35]提出的规则确定当前零指代词的候选先行语。该候选先行语的位置位于当前零指代词之前，并且与当前零指代词的距离不超过两个句子的文本序列距离。

* 1. 中文零指代词消解器

中文零指代消解器旨在针对给定的中文零指代词及其候选先行语，从候选先行语中选择零指代词真正的先行语。在以往的共指消解器设计中，都是只考虑当前待消解中文零指代词-候选先行语对存在共指性的概率，而忽略了之前预测得到的共指链对当前决策的影响。但是对于消解器的设计应该能尽可能使用潜在的信息来辅助当前的决策。那么之前预测得到的共指链对当前的判断更有意义，因此一个理想的零指代消解器在尽可能挖掘输入词对本文特征外，应该能更好的结合之前预测得到的共指链来帮助当前共指性的推断，并且当前的预测结果也能编码到决策链中对后续的消解预测提供必要的信息。



图4.4 中文零指代词消解器框架

Fig. 4.4 Chinese zero pronoun resolver structure

为了解决以上问题，本文基于在英文共指消解中提出的SPNet模块并结合ZPR-ENet来设计中文零指代消解器，该中文零指代词消解器的结构如图4.4所示。在此研究的基础上，基于强化学习的启发，本文引入强化学习的最大化回报期望的相关理论来优化整个中文零指代消解器的目标函数使得模型达到了更优的效果。

中文零指代消解的目的是给定可消解的中文零指代词以及其候选先行语集合，从其候选先行语中选择其最可能的真正先行语。在上述中文零指代消解模型中，中文零指代词消解器本质上作为一个评分模块，在每一时刻为给定零指代词-候选先行语对做出一个预测来表示两者具有共指性的概率。也就是从词嵌入空间到共指概率空间的映射，其中表示当前的待消解零指代词-候选先行语对的向量化表示矩阵，表示是否具有共指性的预测结果，表示当前模型的参数。

首先，对于消解器的输入。在每一个时刻，给定一个待消解的中文零指代词及其候选先行语序列 ，利用当前的零指代词及一个候选先行语序列构建词对  作为待消解的目标。但是在每次的消解判断中只依赖这两者进行是否可消解判断，信息相对有限。但是之前预测得到的共指链对当前的预测提供丰富的信息。比如，在句子“一直在指导这个学生的整个比赛，他的知识非常扎实，我充分相信答辩可以进行的很顺利。”中，的真正的先行语是“这个学生”和“他”。但是由于在训练集中大部分的零指代词的先行语是“我”，并且“我”距离的文本距离更近，很容易使得消解器判断“我”为的真正的先行语。但是如果在之前的共指链中已经确定“那个学生”是的真正先行语，并且把这个消解结果编码到当前消解器的输入中，这就可以有效的帮助当前的消解器对 [，“我”] 进行共指消解判断时得出两者不存在共指性的预测。因此，本文提出使用ZPR-ENet对之前的中文零指代词消解信息进行编码得到。此外，少量的人工特征 (字段跨度、字段距离以及原数据类型等) 对提升模型的效果具有比较大的帮助。据上所述，每个词对的共指性预测依赖四部分的信息，这四部分信息分别是：中文零指代词的词嵌入表示、候选先行语的词嵌入表示、前序零指代词消解信息以及少量人工特征*vf*。在每一时刻的的定义如式 (4.4) 所示。

 (4.4)

上式中，由ZPR-ENet对前序零指代词消解信息进行编码得到。在得到后，SPNet可以把映射到共指概率空间。对于每个中文零指代词-候选先行语对输入，由输入到共指概率空间的映射只包含两种，一种是表示词对中的零指代词和候选先行语具有共指性，他们指向现实世界中的相同参照物，可以被消解；一种表示两者不具有共指性。该过程如式 (4.5) 所示。

 (4.5)

上式中，表示softmax函数。在SPNet的输出层后本文添加了softmax非线性变换来得到最后的共指概率空间。根据得到共指概率信息，利用该概率信息去更新候选先行语信息后对其进行编码。进而使得在后续的编码过程中可以利用前序的先行语信息做出更准确的预测。

基于强化学习的启发，本文采用“奖励”来最大化该模型的概率。模型性能的好坏是由预测结果反映的，那么中文零指代消解模型的性能评价指标采用F1值来衡量，因此，在奖励函数设计时，本文同样采用F1值作为奖励函数来作为对模型做出预测的回报，进而通过模型的多轮次训练以使得整个模型的奖励期望最大化从而不断优化模型。根据以上所述并基于Williams等人[60]提出的策略梯度算法训练模型，也就是使模型的回报的期望最大化，最后的优化函数如式 (4.6) 所示。

 (4.6)

上式中，*T*表示当前零指代词所有候选先行语的数量。根据Clark和 Manning 的研究[56]，该目标函数在优化过程中并不稳定，因此需要对奖励进行缩减，也就是添加一个回报基准以减少方差进而使得模型在训练过程中更加稳定，对于回报基准的取值一般是计算在*t*时刻的回报期望得到。那么最后的优化函数如式 (4.7) 所示。

 (4.7)

* 1. 前序零指代词消解信息编码层

前序零指代词消解信息对当前中文零指代词-候选先行语的消解非常重要，本文提出一种前序零指代词消解信息编码网络ZPR-ENet对当前时刻之前的候选先行语及其历史消解结果进行编码得到。整个ZPR-ENet的结构如图4.5所示。



图4.5 ZPR-ENet结构

Fig. 4.5 ZPR-ENet structure

假设在对当前中文零指代词-候选先行语对进行共指性判断时，已经在之前的阶段确定了当前中文零指代词的*M*个真正的先行语，该*M*个先行语可以表示为。本文采用ZPR-ENet对此*M*个先行语进行编码并加入到后续的零指代消解判断中。对于ZPR-ENet的设计，本文首先使用最大池化操作和平均池化操作并行对前零指代词消解信息进行特征提取，并得到两种不同的特征图后，对特征图进行拼接。如式 (4.8) 所示。

 (4.8)

上式中，*max*表示最大池化操作；*avg*表示平均池化操作；表示中间特征。为了对高级特征进一步挖掘并基于TextCNN[61]的启发，本文使用3层卷积操作对进一步进行高级特征提取。如图4.4所示，本文使用3层卷积操作串行的进行特征提取，在每一层卷积操作后链接一层最大池化操作。同时，在每层最大池化层后使用ReLU激活函数来增强模型的非线性因素，也保证模型在训练的过程中在一定程度上避免梯度消失问题。在最后该模块通过多层感知机对特征维度进行降维度来获得最终的前序共指链信息。

 (4.9)

上式中，表示对前序共指链的编码结果；*ZPR-ENet*表示前序零指代词消解信息编码网络中的所有操作。

* 1. 预训练模型

预训练模型在提升模型的性能上很重要同时基于前人[60]研究的启发。本论文设计了预训练模型来初始化整个ZP-SPNet的网络参数，最后的实验结果表示该预训练模型有效的提升了本文模型的整体效果。本文的预训练模型的优化目标函数采用交叉熵损失函数来设计，具体的优化函数如式 (4.10) 所示。

 (4.10)

上式中，表示整个模型的参数，表示当前零指代词的候选先行语集合，表示当前待消解的零指代词。在上式中，当前零指代词和候选先行语可消解时，；如果两者不可消解时，。

* 1. 本章小结

本章从整体框架到局部模块详细介绍了本文提出的中文零指代消解模型。由于中文零指代词的特殊性，本文基于给定可消解零指代词的条件下研究中文零指代消解模型的设计和验证。首先本文介绍了中文零指代消解的定义和整体模型框架。接下来本章节对模型的细节进行了详细介绍。4.3节首先介绍了本文对原始中文零指代词的向量化表示方法，本文基于LSTM联合当前零指代词的上下文对当前零指代词进行表示；4.4节详细介绍了中文零指代器的设计，该中文零指代消解器基于SPNet并结合ZPR-ENet进行构建。此外，本文基于强化学习的启发，使用F1值作为整个消解器的奖励，优化模型的过程就是使得奖励最大化的过程。在4.5节本文主要介绍了前序零指代词词消解信息编码网络ZPR-ENet。为了提升模型的性能以及保证模型的可收敛性，本文在4.6节介绍了本文使用的预训练模型。在后续章节的实验部分，本文对该中文零指代消解模型的整体性能以及各模块的有效性进行了充分的验证。

1. 实验与结果分析

本文在第3章和第4章分别对英文共指消解和中文零指代消解两个任务进行了定义并提出了相应的解决模型，还对模型的结构和原理做了详细的介绍。本章节将针对以上提出来的模型的性能进行详细评价。为了验证两种模型的有效性，基于前人提出的评价手段本文对模型进行量化评估并和历史优秀模型进行了对比。同时，为了验证本文提出的方案的有效性，本文进行了相关消融实验方案进行深入验证。

* 1. 数据集介绍

目前广泛应用于共指消解的公开语料库包括以下3种：MUC语料库、ACE语料库以及OntoNotes语料库。针对本文提出的英文共指消解模型和中文零指代消解模型，本论文使用OntoNotes 5.0数据集来对其进行训练和性能评估。

OntoNotes 5.0是一个包含了三种语言的语料库，分别是：英语、汉语和阿拉伯语。针对每种语言类型，该语料库又包含新闻、新闻组的广播谈话、对话中的电话语音文稿等7种来源的数据。该数据集为共指消解类任务提供了标注比较精确的训练集、测试集和交叉验证集，该数据集也是有史以来针对共指消解被认为是最经典也是最标准的数据集。并且该数据集收集的语料相对较多，来源比较广泛。最近基于深度学习，针对英文共指消解任务和中文零指代消解任务的研究任务均使用OntoNotes 5.0 数据集来对模型进行训练并进行性能验证。

表5.1 OntoNotes 5.0语料库英语各类数据大小统计

Table 5.1 The statistics of English various data size in OntoNotes 5.0 corpus

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 语言类型 | 类型 | 训练集 | 验证集 | 测试集 |
| 英语 | 单词数 | 1.3M | 160K | 170K |
| 文档数 | 2.8K | 0.3K | 0.3K |
| 实体数 | 35.1K | 4.5K | 4.5K |
| 共指链 | 120.4K | 14.6K | 15.2K |
| 表述数 | 155.5K | 19.1K | 19.7K |

对于英文共指消解任务，本文使用了OntoNotes 5.0数据集中的英语部分对模型进行训练和性能验证。表5.1是本文对OntoNotes 5.0数据集中英语部分关于共指消解任务的统计数据做的相关整理，该统计数据主要包括英语数据集中训练集、测试集和交叉验证集中的单词数、文档数、实体数等数据信息。从表5.1可以看出，该数据集中的真正表述的数量和共指链数量在训练集、测试集以及交叉验证集中的占比先对较多且一致，这对训练一个优秀的英文共指消解模型非常必要。

表5.2 OntoNotes 5.0语料库汉语各类数据大小统计

Table 5.2 The statistics of Chinese various data size in Ontonotes 5.0

|  |  |  |  |
| --- | --- | --- | --- |
| 语言类型 | 类型 | 训练集 | 测试集 |
| 汉语 | 文档数 | 1.4K | 0.2K |
| 句子数 | 36.5K | 6.1K |
| 表述数 | 178.1K | 25.9K |
| AZPs | 12.1K | 1.7K |
| ZPs | 23.1K | 3.7K |

针对中文零指代消解，这种语言现象相对比较特殊，在英语中出现该语言现象的文本较少，但是在汉语中较多。根据Chen和Ng对OntoNotes 5.0中英语数据集和汉语数据集的统计，仅有4%的英语句子中存在的零指代消解现象，但是在汉语中却达到了36%因此本文使用该数据集的汉语部分训练和验证中文零指代消解模型。汉语中零指代消解现象相对英语的占比虽然较大，但是相对于数据集中所有的汉语文本，零指代词的数量比较有限。并且数据集的组织者仅对OntoNotes 5.0汉语数据集的训练集和交叉验证集中的零指代词进行了标注。因此，本文使用交叉验证集作为测试集来验证中文零指代消解模型。表5.2对OntoNotes 5.0汉语数据集中存在零指代消解的文件相关数据进行了统计。主要包括文件数目、零指代词数目 (ZPs) 、可消解零指代词数目 (AZPs) 以及表述数目 (NPs) 等数据项。

* 1. 英文共指消解模型实验设计与结果分析

本小节主要验证本文提出的英文共指消解模型的性能。首先介绍了英文共指消解模型相关的几种评价指标；然后本文基于OntoNotes 5.0的测试集得到本文模型的性能分析结果，同时本论文把前人的一些优秀模型以及基准方法和本文提出的模型进行对比以验证本文模型的有效性。此外，为了验证本文提出的模块的有效性，本文设计了消融实验来验证各个模块对模型整体性能提升的有效性。

* + 1. 模型评价指标

模型的评估方法在一定程度上决定了模型的性能。对共指消解模型的评价最直观的方法是把模型预测得到的共指链和数据集中的标准共指链做对比，来评价模型性能的好坏。目前，随着共指消解任务研究的深入，英文共指消解模型性能的评价已经存在多种规范化的指标。基于前人的研究，本文采用和前人相同的三种模型评价方法来评价本论文提出的英文共指消解模型，这三种评价指标分别是：、和。针对这三种评价标准，每一种都采用三种衡量的标准来进行描述，分别是：准确率、召回率和F1值，此三种评价手段的详细介绍如下。

(1) 评价方法

****评测方法是Vilain[62]等人为共指消解模型提出的第一个规范化的评价指标，该评价指标是基于链接 (link-based)，每个表述最对和另外的两个表述链接。在该评价指标中定义了“*key*”和“*response*”，其中“*key*”表示数据集提供的真实共指链，*response*表示提出的共指消解模型预测的共指链。该评价指标主要统计*key*和*response*的交集，其准确率是这两者的交集和真实值得比值。

(2) 评价方法

该评价指标是对的一种改进方法。不同于评价指标，该评价指标是直接对每个表述的准确率进行计算，然后对所有的描述加权求和得到最终的准确率和召回率。根据模型的预测结果，****会给以不同的错误类型以不同的权重惩罚，其中一般使用一个表述的数量的倒数作为权值。但是该方法也有一定的缺点，一个实体有可能被重复多次匹配，这是一个不合理的现象。

(3) 评价方法

该评价指标是由Luo等人[63]提出的一种比较复杂的共指消解模型评价指标。该方法会计算*key*和*response*之间的相似性来得到一个一对一的映射关系，然后使用自相似性计算准确率和召回率。这反映的是正确指向实体的表述占所有预测结果中所有表述的百分比。

在英文共指消解模型的评测中，其评价方法的优劣一直没有一个清晰的区分，为了模型的准确性以及可靠性，本文使用以上三种评价指标对模型进行综合评测。对该三种方法，本文都计算了其对应的准确率、召回率和F1值，最后取三种评价方法的平均F1值来衡量模型的最终性能。

* + 1. 模型参数设计

受基准模型的启发以及交叉验证集对模型性能多轮次训练的验证，本文主要的参数设计如下。对于文本向量化表示，本文是采用Glove、ELMo和BERT三种词嵌入方式来表示本文的字段。Glove词嵌入模型得到的词向量维度与基准方法的保持一致，为300维。ELMo和BERT得到的词向量的维度为1024。训练数据中的词在向量化表示后使用LSTM网络来对当前词的上下文信息进一步编码，该LSTM的隐藏层单元设计为200。在表述检测模块中，本文使用的前馈神经网络拥有两层隐藏层，每一层包含150个单元，并且在该模块中激活函数采用ReLU激活函数。SPNet模块主要包含PNet和SNet两个子模块，该两个模块的设计是基于卷积神经网络。对于PNet的设计，本文采用三种大小的卷积核来构建平行卷积层进行特征提取，卷积核的大小为 {1, 3, 5}，其中每个卷积核的数量设置为50。SNet模块的设计同样是基于卷积神经网络，卷积核的大小分别为 {1, 3, 3}，每个卷积核的数量同样设置为50。在PNet和SNet模块的每层卷积层后，本文都采用ReLU激活函数来提升模型的训练速度，并在一定程度上防止梯度消失现象的发生。此外，为了防止模型在训练过程中过拟合，本文在部分全连接层后添加了dopout层，系数设置为0.5。

在本文提出的英文共指消解模型中，字段是基于规则得到，为了提高模型的效率，首先对每个字段的跨度做了限定，本文限定每个字段的最大文本跨度为30个单词。而后使用表述检测模块对每个字段是否是真正的表述进行判断。为了提升模型的性能，本文仅保留35%具有最高分数的字段作为待消解表述，每个待消解表述分配30个表述作为其候选先行语。对于英文共指消解模型的优化过程，本文采用Adam优化器来优化整个模型，初始的学习率设置为0.01。在本论文中，对实体表述检测模块和共指消解器进行联合优化，对两者的损失函数进行加权求和作为最终的优化目标函数，其中设置为0.01时，能够得到最优的模型效果。对于模型训练的批次大小，本文和Lee等人[18]提出的模型保持一致，每次以一个文件作为一个训练批次。

* + 1. 实验结果分析

(1) 模型性能测试及结果分析

本文基于本章节上述的三种评价方法对本文提出的英文共指消解模型的性能进行验证，结果如表5.3所示。为了增强本文提出的英文共指消解模型的说服力，我们在该表5.3中同时展示了近年来在共指消解领域的一些优秀模型。其中，Kantor 和Globerson提出的模型[27]为本论文提出的英文共指消解模型的基准模型。在表5.3中所涉及到的模型的性能评价指标均和本文使用的评价指标相同，并且其模型均是使用OntoNotes 5.0数据集对模型进行训练和性能验证。

从表5.3可以看出，和前人的优秀模型的性能相比，本文提出的共指消解模型在测试集上的表现更为优秀。本文提出的英文共指消解模型在最优的参数设计下使用测试集对模型的性能进行验证，该模型在三种评价方法对应的F1值分别达到了83.5%、75.1%以及72.1%，该结果明显优于基准模型的83.4%、74.7%以及71.8%。从平均F1值来看，我们的模型的平均F1值达到了76.9%，其相对于基准模型的76.6%提升了0.3%。

本文基准模型的原始语料库向量化表示方法也是采用了多种词嵌入联合向量化表示的方式，本文相对于其添加了一种ELMo词嵌入模型。为了更好的体现本文设计的消解器的性能，我们去掉ELMo词嵌入模型，采用和基准模型相同的原始语料库向量化表示方法。测试结果如表5.3的最后一行所示，相对于基准模型，我们的模型在****和两种评价方法上仍具有优势，相对于基准模型在此两个评价指标的F1值上分别有0.3%和0.2%的提升。从平均F1值看，相对于基准模型，我们的模型也仍有0.2%的提升。这可以证明本文提出的方案的有效性。

表5.3 英文共指消解模型实验结果

Table 5.3 Experimental results of English coreference resolution model

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 |  | | |  | | |  | | |  |
| P | R | F1 | P | R | F1 | P | R | F1 | Avg.F1 |
| Martschat and Strube[64] | 76.7 | 68.1 | 72.2 | 66.1 | 54.2 | 59.6 | 59.5 | 52.3 | 55.7 | 62.5 |
| Clark and Manning[65] | 76.1 | 69.4 | 72.6 | 65.6 | 56.0 | 60.4 | 59.4 | 53.0 | 56.0 | 63.0 |
| Wiseman et al.[15] | 76.2 | 69.3 | 72.6 | 66.2 | 55.8 | 60.5 | 59.4 | 54.9 | 57.1 | 63.4 |
| Wiseman et al.[66] | 77.5 | 69.8 | 73.4 | 66.8 | 57.0 | 61.5 | 62.1 | 53.9 | 57.7 | 64.2 |
| Clark and Manning[56] | 79.2 | 70.4 | 74.6 | 69.9 | 58.0 | 63.4 | 63.5 | 55.5 | 59.2 | 65.7 |
| Lee et al.[18] | 78.4 | 73.4 | 75.8 | 68.6 | 61.8 | 65.0 | 62.7 | 59.0 | 60.8 | 67.2 |
| Lee et al.[22] | 81.4 | 79.5 | 80.4 | 72.2 | 69.5 | 70.8 | 68.2 | 67.1 | 67.6 | 73.0 |
| Fei et al.[67] | **85.4** | 77.9 | 81.4 | 77.9 | 66.4 | 71.7 | 70.6 | 66.3 | 68.4 | 73.8 |
| Kantor and Globerson[27] | 82.6 | 84.1 | 83.4 | 73.3 | 76.2 | 74.7 | 72.4 | 71.1 | 71.8 | 76.6 |
| Joshi et al.[28] | 84.7 | 82.4 | **83.5** | **76.5** | 74.0 | **75.3** | **74.1** | 69.8 | 71.9 | **76.9** |
| **Our Model** | 82.8 | **84.2** | **83.5** | 74.0 | **76.2** | 75.1 | 72.0 | **72.1** | **72.1** | **76.9** |
| -ELMo | 82.8 | 83.9 | 83.3 | 74.3 | 75.7 | 75.0 | 71.8 | **72.2** | 72.0 | 76.8 |

相较于基准模型及其之前的英文共指消解模型，本文提出的模型在精确度上的优势主要来源于四个方面。第一点，本文采用多种词嵌入技术联合对原始文本进行向量表示，并且在词嵌入层后添加了编码层对当前词向量的上下文信息进一步编码，这使得本文每个词对应的词向量包含更多的文本信息；第二点，本文对表述检测和消解器进行联合优化，这可以保证表述检测模块的检测结果尽可能的是真正的表述，这也就使得当消解器消解的对象尽可能正确，而不是对非表述间的共指性进行判断；第三点，文本提出的SPNet模块相较于传统的使用简单前馈神经网络构建消解器更有效，并且本文构建的消解器是一个二阶消解器，在第二阶使用两通道的词向量对表述进行表示，这使用的消解器可以更好的对原始先行语信息进行挖掘；第四点，文本在消解的第一阶和第二阶之间对词向量进行更新。消解器的第一阶已经对表述对的共指性做出了初步的判断，使用这个前序判断信息结合注意力机制对词向量进行更新，可以保证在后续的判断过程中更好的考虑前一阶段的判断的结果，并且为了保留原始先行语分布信息，本文采用原始词向量和更新后的词向量共同对表述进行表示，这进一步提升了本文提出模型的性能。

本文提出的模型相较于Joshi等人[28]提出的模型在平均F1值上相同，但是本文提出的模型在评价指标上优于该模型。本文提出的模型指标的F1值上达到了72.1%，这相对于Joshi等人提出的模型提升了0.2%。但是值得注意的是，Joshi等人提出的英文共指消解模型更注重的是文本的向量化表示而非消解器的设计。Joshi等人对BERT进行了微调以生成文本中的词对应的词向量，后续的消解器是直接使用的Lee等人[18]提出的消解器，以此来构建一种端到端的英文共指消解模型。而本文更注重的是共指消解器本身的设计以及性能。此外，本文提出的模型是一个相对效率较高对硬件设备需求较少的模型，相对于Joshi等人提出的英文共指消解模型需要4块12GB内存的GPU进行训练的最低硬件需求，本文提出的模型仅需1块12GB的GPU就可以完成该模型的训练。并且本文提出的模型在训练时间上有比较大的优势。在最优参数设计下达到最好的模型效果仅需要36小时的训练。

本文提出的模型的效率优势主要来自于以下几个方面。首先，相对于我们的基线模型以及Joshi等人提出的模型，本文对待消解的表述和每个表述分配的候选先行语的数量进行了限制。基线模型以及Joshi等人提出的模型每次选择所有字段的40%作为待消解表述，每个表述分配50个候选先行语来构建表述对，但是本文提出的模型仅对35%的字段进行消解，并且每个待消解表述仅分配30个候选先行语来构建表述对；其次，相对去前人使用参数量巨大的全连接层构建消解器，本文使用基于卷积神经网络构建的SPNet以及T-SPNet来构建消解器在一定程度上压缩了模型的参数量。以上两点在一定程度上促进了模型在效率上的提升。

(2) 消融实验及结果分析

表5.4消融实验结果

Table 5.4 Results of ablation experiments

|  |  |  |
| --- | --- | --- |
| 模型 | Avg.F1 | Δ |
| -SPNet | 72.44 | 4.47 |
| -字段表示更新 | 76.52 | 0.39 |
| -多通道字段表示 | 76.64 | 0.27 |
| -表述检测联合优化 | 76.79 | 0.12 |

为了进一步证明本文提出的各个方案对模型性能的贡献，本文设计了一系列的消融实验进行验证，实验结果如表5.4所示，表中的Δ表示当前模型相对于完整模型性能的下降的绝对值。本文设计的消融实验是使用相同的参数以及相同的测试集进行验证。该消融实验分别移除SPNet模块、字段表示更新模块、多通道字段表示以及表述的联合检测后，平均F1值相对最优值分别下降了4.47%、0.39%、0.27%以及0.12%。从表5.4可以看出，当移除SPNet时，模型效果有非常大幅度的下降，这证明本文提出的方案以及各个模块对提升英文共指消解模型性能上有比较大的促进作用。

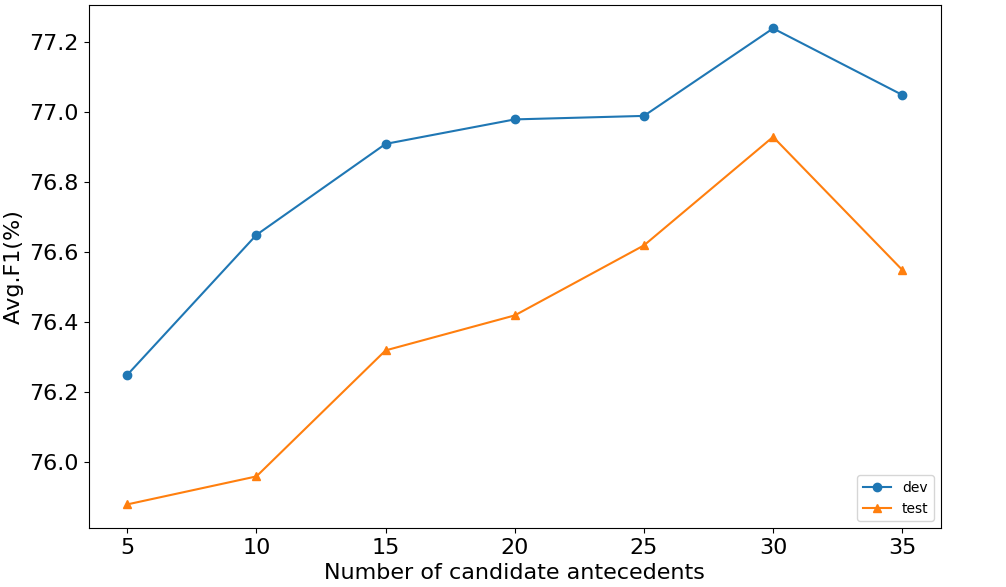


图5.1 候选先行语数量对模型性能的影响

Fig. 5.1 The influence of number of candidate antecedents on our model

本文基于一种交替非线性评分函数采用由粗粒度到细粒度的筛选方法来确定每个待消解表述的候选先行语，候选先行语的数量对模型性能具有比较大的影响。为探索候选先行语数量对模型性能的影响，分别设置不同的候选先行语集进行实验，实验结果如图5.1所示。从图5.1可以看出，随着候选先行语数量的增加，模型在测试集和交叉验证集上的性能表现先升高后降低。候选先行语集的数量的增加在一定程度上可以对模型性能的提升起到促进做用，但是当候选先行语集超过一定的阈值时，就会使得模型的性能下降。候选先行语集在增大的过程，会使得当前待消解表述的真正的先行语出现在候选先行语集中的概率增大，提升模型的准确性。但是当候选先行语集过大时，模型做出正确判断的概率会下降。

从以上消融实验结果可以看出，本文提出的相应方案对英文共指消解模型性能的提升是有一定的促进作用的。尤其是基于SPNet构建的英文共指消解器为模型整体性能的提升带来了巨大的优势。SPNet本身是一个多尺度的特征抽取模块，该模块可以更好的提取表述对内部的低级和高级语义特征并用于共指链的预测。此外，字段表示的更新和多通道表示方法使得模型能够充分利用之前阶段得到的共指链信息并没有削弱原始词向量中的共指信息，这避免了一部分词向量更新过程中对后续模型预测的错误引导。对于实体表述检测和消解器联合优化，这提升了实体表述检测模块的准确性，尽可能的保证消解器要消解的表述对中两个字段是真正的表述。消融实验的结果也证明了上述方案可以为共指消解模型性能带来一定的提升。

(3) 模型效果展示

本文针对该英文共指消解模型的实际效果进行了测试。英文共指消解模型旨在输入一个待消解的文本序列，可以对该文本序列中指向真实世界中相同实体的表述聚集到相同的集合中。待消解的文本序列如图5.2所示。在待消解文本中，一共存在五条共指链，本文使用不同的文本颜色人工对共指链做了标注，其中的相同颜色文本的表述输入同一条共指链。此外，本文为待消解的文本中每个句子的开头添加上了起始的索引，方便后续对预测结果的解码操作。



图5.2 待消解文本序列

Fig. 5.2 Text sequence to be resolved



图5.3 预测得到共指链

Fig. 5.3 The predicted coreference chains

将图5.2中所示待消解文本输入到本文提出的英文共指消解模型中，模型的输出是一系列的共指链集合，如图5.3所示。图5.3中每一行表示预测得到的一条共指链，每一条共指链的包含若干个被判断为表述的字段。每一个字段由其起始位置单词的索引和结束位置的单词索引表示。针对图5.2输入的文本序列，本文提出的模型同样得到五条共指链，对预测的共指链集合进行解码并在原文中进行标注如后图5.4所示。在图5.4中，每一种颜色的文本表示一条共指链中表述的集合。



图5.4 共指链在原文中对应的表述

Fig. 5.4 The mentions corresponding to the coreference chains in original text

对预测得到的共指链进行分析，对比图5.2和图5.4可以看出，本文提出的模型能够对文本中的共指链进行比较精确的预测。相对于图5.2所示的标准共指链中 [acquired, the transactions] 这一条共指链，本文提出的模型预测到了“acquired”和“the transactions”存在共指关系，但是错误的预测了额外的一个表述也存在于该条共指链中。除去该错误的共指链之外，其余的共指链均和标准共指链相同。这表明我们提出的模型可以比较好的处理实际的英文共指消解问题。

* 1. 中文零指代消解模型实验设计与结果分析

本小节主要验证本文提出的中文零指代消解模型的性能。首先介绍了中文零指代消解领域的模型性能评价指标；然后本文基于OntoNotes 5.0数据集进行模型性能测试和验证分析。为了验证该中文零指代消解模型的性能，本文使用相同的评价指标和前人的模型进行了对比。同时为了验证本文中提出的相关模块的重要性，本文设计了消融实验进行验证。

* + 1. 模型评价指标

中文零指代消解的研究相对较少。根据数据集提供的数据以及常用的模型评价指标，在中文零指代消解领域常用的是使用召回率 (R)、准确率 (P) 和F1值对模型性能进行评价。在中文零指代消解的数据集中提供了6种不同种类来源的测试数据集，本文分别对该子数据集计算其召回率、准确率和F1值，最后使用平均F1值作为最终的性能评价指标。根据Chen和Ng提出的系统中对中文零指代消解模型评价指标的定义，召回率、准确率和F1值的计算方式如下列各式所示。

 (5.1)

 (5.2)

 (5.3)

在上式中，“Res Hit”表示当前待消解中文零指代词被判断为可消解的候选先行语，在OntoNotes 5.0数据集中，一个中文零指代词可能具有多个真正的先行语；“AZP in Key”表示具有先行语的中文零指代词的统计数量；“AZP in Prediction”表示被消解了的零指代词的数目。由于只是对中文零指代消解器进行研究，因此本文只考虑可消解的中文零指代词。

* + 1. 模型参数设计

在本文的参数设计中，SPNet的基本超参数设计和英文共指消解中的SPNet的超参数设计保持一致。本文使用到中到的中文词向量维度设置为100。本文整个模型的训练分为预训练模型和正式训练模型两部分。首先对于预训练模型，训练中的批次大小设置为256，训练的轮次设置为50。预训练模型中使用了dropout层，dropout概率设置为0.5。使用AdaGrad优化算法来优化目标函数，学习率设置为0.003。对于完成模型的训练，其和预训练模型的超参数设计仅有两点不同。在完整模型的训练中，dropout概率设置为0.5，学习率设置为0.00009。完整模型的训练轮次设置为90时，效果达到最优。

* + 1. 实验结果分析

(1) 模型性能测试及结果分析

基于上述的中文零指代消解模型评价指标对本文提出模型的性能进行验证，测试结果如表5.4所示。表5.4前6列是使用六种不同来源的测试数据对模型进行测试，每列中的数字是各个测试集对模型进行测试得到的F1值。在最后一列 (Overall)，本文使用整体计算了六种测试数据F1值的平均值，这个指标能更好的展示该中文零指代消解模型的性能。此外，为了增强模型的说服力，在该表中同时展示了近年来 (2015年至今) 在中文零指代消解领域表现优秀的消解模型在各个测试集上的表现。表中所列举的模型的性能均使用相同的评价指标进行衡量，其中Yin等人[37]提出的模型达到了57.2%，同时该模型也是本文的基准模型。

从表5.4可以看出，本文提出的中文零指代消解模型相对表中所列模型表现优异。基于50轮的预训练模型对最终模型的参数进行初始化，我们对整个模型进行多轮次的训练。在我们的模型训练50轮时平均F1值达到了58.54%，相较于基准模型提升了1.34%。当本文的模型进行90轮的训练后，模型的性能达到最优，其平均F1值相对于基准模型提升了1.63%。从六种不同的测试数据集来看，本文提出的模型在MZ、WB、BN以及TC四个测试数据集上相较于我们的基准模型的表现有明显的提升。

表5.4 中文零指代消解模型测试结果

Table 5.4 Experimental results of Chinese zero pronoun resolution model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | NW | MZ | WB | BN | BC | TC | Overall |
| Zhao and Ng[33] | 40.5 | 28.4 | 40.1 | 43.1 | 44.7 | 42.8 | 41.5 |
| Chen and Ng[68] | 46.4 | 39.0 | 51.8 | 53.8 | 49.4 | 52.7 | 50.2 |
| Chen and Ng[35] | 48.8 | 41.5 | 56.3 | 55.4 | 50.8 | 53.1 | 52.2 |
| Yin et al.[36] | 50.0 | 45.0 | 55.9 | 53.3 | 55.3 | 54.4 | 53.6 |
| Yin et al.[38] | 48.8 | 46.3 | 59.8 | 58.4 | 53.2 | 54.8 | 54.9 |
| Yin et al.[41] | 63.1 | 50.2 | 63.1 | 56.7 | 57.5 | 54.0 | 57.2 |
| Yin et al.[69] | **64.3** | 52.5 | 62.0 | **58.5** | 57.6 | 53.2 | 57.3 |
| Our Model (50 Epoch) | 58.33 | **53.87** | 63.14 | 58.46 | 56.86 | **57.98** | 58.54 |
| Our Model (90 Epoch) | 59.52 | 53.25 | 63.84 | 58.46 | 56.07 | **57.98** | **58.83** |
| -RL (90 Epoch) | 55.95 | **53.87** | **65.26** | 57.95 | **59.23** | 57.25 | 58.36 |

此外，本文提出的中文零指代消解模型是基于英文共指消解研究中的SPNet模块并结合进行构建。此外，模型的训练方法区别于传统的使用交叉熵来设计优化目标函数。为了探索SPNet在中文零指代词消解中的表现，本文使用交叉熵损失函数代替以最大化回报期望的方式设计的目标函数，进而对SPNet在中文零指代词消解任务中的表现进行测试结果如表5.4中的“-RL”行所示。从实验结果来看，相对于基准模型，我们的模型在MZ、WB、BN、BC以及TC子测试集中的测试结果明显优于基准模型，平均F1值相对于基准模型也有1.16%的提升。这表明SPNet不仅可以有效的解决英文共指消解任务，在中文零指代消解领域也有很好的效果。该实验结果相对于最优的模型性能下降了0.18%，这也表明了采用最大化回报期望设计目标函数的方法对模型性能的提升有一定的贡献度。

(2) 消融实验及结果分析

表5.5 消融实验结果

Table 5.5 Results of ablation experiments

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | NW | MZ | WB | BN | BC | TC | Overall | Δ |
| -ZPR-ENet | 51.19 | 52.01 | 62.08 | 56.41 | 57.06 | 56.15 | 57.42 | 1.42 |
| -Pretraining | 35.71 | 27.86 | 40.21 | 39.49 | 43.63 | 49.17 | 40.64 | 18.79 |

为了验证ZPR-ENet和预训练模型对整体模型的贡献度，本文设计了相关消融实验来对模型进行测试，消融实验结果如表5.5所示，其中Δ表示模型性能相对于最优值下降的绝对值。首先，为了验证本文提出的ZPR-ENet模块对整个模型性能的影响，在消融实验中屏蔽ZPR-ENet对模型的影响后，对模型的性能进行测试。从表5.5第一行可以看出，在去掉ZPR-ENet模块后，模型的性能下降1.42%，因此前序信息编码模块提取的前序零指代词消解信息对当前零指代词的消解非常重要。其次，本文采用预训练模型对消解器的参数进行初始化，为了探索预训练模型对模型整体性能的影响。本文去掉预训练模型的影响，同样训练90轮次后得到的模型的精度相对于采用预训练模型的完整模型精度下降了18.79%。这足以证明预训练模型在中文零指代消解模型中的贡献度。并且在实验阶段，由于预训练模型的存在，使得中文零指代消解模型整体更容易收敛。

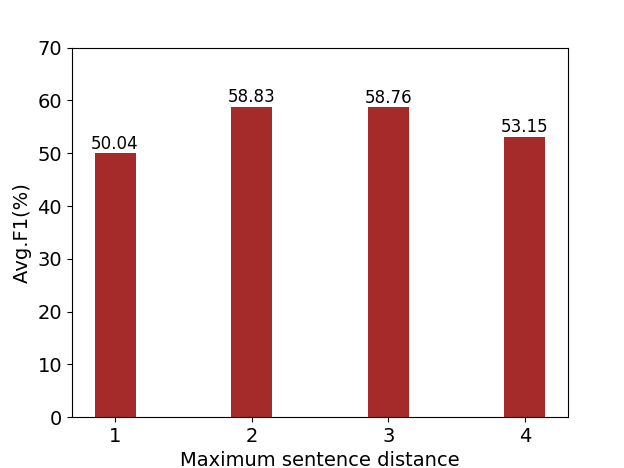


图5.5 不同候选先行语集对模型性能的影响

Fig. 5.5 The influence of the different set of

candidate antecedents on the performance of the model

在模型的设计过程中，对候选先行语集的定义是仅考虑距离当前零指代词一定句子距离内的表述作为当前零指代词的候选先行语。为了探究候选先行语于当前零指代词的距离对模型性能的影响，本文设计的如下消融实验。针对当前零指代词，分别设定距离当前零指代为1、2、3、4个句子距离范围内的表述作为其候选先行语。当设定的范围不同时，对模型的影响如图5.5所示。从图5.5可以看出，当文本句子距离的设定由少增多时，模型性能也由低到高增加。当在合理给定搜索空间中选择零指代词的候选先行语时，模型往往能有更合理的表现。但是随着搜索空间的增大，模型的误判率将会增加，这也将导致模型整体性能的下降。

此外，本文对预训练模型和完整的模型进行了多轮次的训练以验证整体模型的性能，实验结果如图5.6所示。从图5.6可以看出，无论对于预训练模型还是最后完整的模型，当训练轮次逐渐增加的过程中，其在测试集上得到的平均F1值都在起伏中先增加后减少。当训练轮次为60时，预训练模型的平均F1值达到最大，为58.53%；当训练轮次为90时，完整模型的平均F1值达到最大，为58.83%。该训练结果，相对于本文采用的基准方法的最优性能，一直都处于优势。

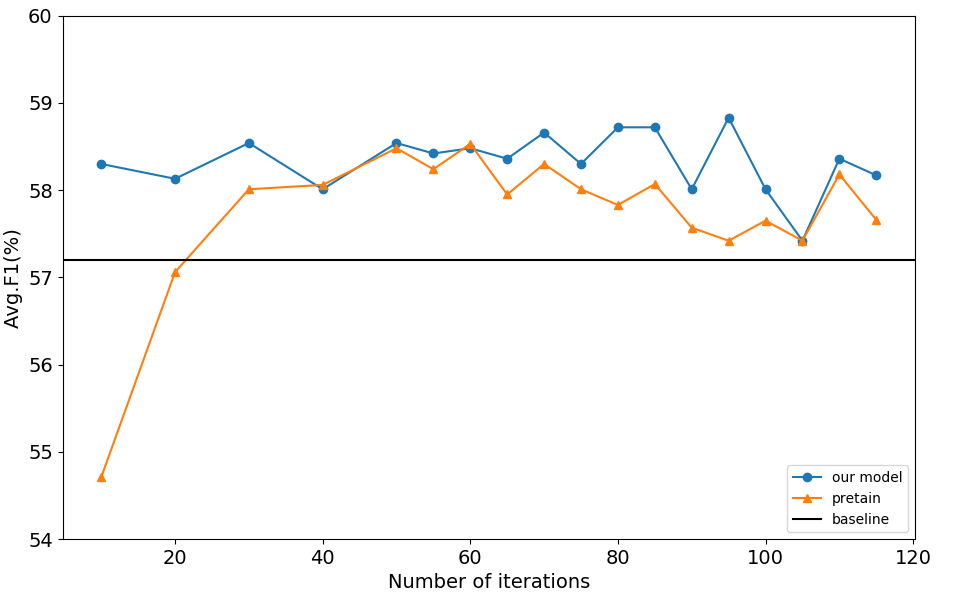


图5.6 不同训练轮次下模型的性能

Fig.5.6 Performance of our model under different training epoch

相较于之前提出的中文零指代消解模型，本文提出的中文零指代消解模型的性能提升主要来源于以下三点：1) 本文提出使用ZPR-ENet对之前的共指链信息进行编码提取，并将该特征信息用于后续共指性的判断，为后续的判断提供线索；2) 本文基于SPNet并结合ZPR-ENet提出了ZP-SPNet，并以此来构建中文零指代消解器，ZP-SPNet模块在进行共指性判断时能更好的提取潜在的高级特征，这对最后词对之间共指性的判断更有帮助；3) 基于强化学习的启发，本文使用F1值作为奖励函数对消解器做出的判断进行奖励，并且整个模型的优化目标就是最大化回报的期望。此外，为提升模型的收敛性，本文设计了预训练模型来对模型的参数进行初始化。以上共同提升了中文零指代消解模型的整体性能。

* 1. 本章小结

本章主要对本文提出的英文共指消解模型和中文零指代消解模型的性能进行验证和分析。首先，本文在5.1节对模型训练和测试的数据集进行介绍；在5.2节对英文共指消解模型的评价指标、模型参数设计以及实验结果进行分析，同时本文对英文共指消解模型的实际使用进行了展示，输入结构化文本，可以得到文本中存在共指性的描述；5.3节对中文零指代消解的模型性能进行分析，针对其评价指标和模型参数进行了详细的介绍。

1. 结果与展望
   1. 论文工作总结

共指消解是自然语言处理技术中的一个重要研究方向，并且是其他多种自然语言处理技术的上游任务。由于人类语言系统之间具有一定的一致性和互补性，基于多种语言的自然语言处理技术能够更全面的对文本进行深入分析。因此，近年来，跨语言的自然语言处理技术被相关学者广泛关注，那么对多种语言的共指消解技术中的关键问题进行深入研究也变得更具有实际意义。这种研究使得共指消解技术能够更全面的分析原始文本中的指代关系并更好的服务于其他跨语言自然语言处理任务。基于以上考虑，本文致力于英汉互译文本中共指消解技术及其关键问题的研究，聚焦于英文共指消解和中文零指代消解两个关键问题，创新性的提出了相应的解决方案并设计实验进行了验证。

(1) 英文共指消解模型

对于英文共指消解的研究，基于端到端的思想，本文提出了一种新颖且性能较好的英文共指消解模型。本文研究的内容主要分为三个部分：1) 原始语料库的向量化表示；2) 待消解表述的检测；3) 英文共指消解器的设计。整个模型的核心内容是基于本文提出的SPNet和T-SPNet构建的高阶英文共指消解器。在整个模型设计的过程中，首先，在将原始语料映射到词嵌入空间阶段，本文联合多种词嵌入模型对原始语料库中的单词进行向量化表示以使得字段的中心表达信息更精准。其次，在待消解表述检测阶段，基于联合表述检测和消解思想，本文将表述检测模型和消解模型联合优化以使得待消解表述检测模块更好的学习到真正表述在原始文本中的空间概率分布。此外，为提升模型效率，基于一种交替非线性评分函数本文对选先行语的搜索空间进行了由粗粒度到细粒度的修剪，对最不可能是先行语的表述予以修剪。最后，在英文共指消解器构建中，本文提出了SPNet和T-SPNet并以此构建了一种高阶英文共指消解器。在消解器对表述对的共指性判断过程中，本文结合注意力机制对字段的向量化表示进行更新并采用多通道表示方法对字段进行向量化表示。本文在公开的数据集OntoNotes 5.0中选择英文数据集对模型性能进行验证。基于相同的评价手段，本文提出的模型相对于基准模型有着0.3%的提升，这证明了我们提出的英文共指消解模型的有效性。此外，本文设计了一系列的消融实验来验证各个模块的贡献度，消融实验结果表明了各个模块对整体模型性能的提升有较高的贡献度。

(2) 中文零指代消解模型

对于中文零指代消解的研究，本文的研究内容主要分为两部分：1) 中文零指代词的表示；2) 中文零指代词消解器的设计。中文零指代消解和英文共指消解在本质上具有相似性，主要的区别在于前者的消解对象之一是没有实际表示形式的零指代词。对于本文的整个中文零指代词消解模型设计的过程中，首先是中文零指代词的向量化表示，本文基于LSTM对给定零指代词的上下文进行编码以得到零指代词的实际表示形式，对其上文使用正向LSTM进行编码，下文使用反向LSTM进行编码，得到的结果进行拼接以表示当前中文零指代词。其次，前序的零指代词消解信息对后序的中文零指代词可消解性判断极为重要，本文提出了一种前序零指代词消解信息编码器ZPR-ENet对该信息进行编码以作为后序中文零指代词的可消解性判断的“线索”，然后结合SPNet和ZPR-ENet得到ZP-SPNet并以此构建最终的中文零指代词消解器。最后，本文基于强化学习的启发，使用平均F1值作为中文零指代消解器判断结果的奖励来设计目标优化函数，以最大化回报的期望做为目标来优化整体模型。本文基于公共数据集OntoNotes 5.0的汉语部分对模型性能进行验证。首先在没有融入强化学习相关优化方法的条件下单独验证了基于ZP-SPNet构建的中文零指代消解器的性能，相对于基准模型得到1.16%性能的提升。然后使用最大化回报期望的方式对该模型进行训练，本文模型的性能相对于基准模型提升了1.63%。这证明了本文提出解决方案的有效性。

* 1. 研究方向展望

本文聚焦于英文共指消解任务和中文零指代消解任务，基于前人的研究提出了自己的研究方案并设计相关实验进行验证。从实验结果来看，相较于前人的算法本文提出的模型虽然在性能上实现了一定幅度的提升，但是仍然存在一些可以改进之处。

(1) 对于英文共指消解，本文并没有考虑表述之间的重叠问题，一个文本序列中存在的多个相同的表述被重复作为待消解词进行消解，这无疑造成了大量计算资源的浪费，使得模型的性能受到了限制。同时，由于相同的表述被重复消解进而产生了重复的共指链，在共指链聚合阶段也浪费了大量计算资源。因此可以从表述本身的匹配入手，避免已经被消解的表述被作为待消解表述重复消解。

(2) 本文提出的英文共指消解模型的效率相对基准模型虽然有一些相对的提升，但是仍然较低。本文在确定可能的表述时是基于规则提取所有可能的字段来检测其是否为表述，尽管在进行表述对构建时对待消解表述的数量进行了修剪，并且也对先行语的搜索空间进行了修剪，但是待消解表述的检测无疑大大增加了模型的复杂度。因此，如何使用复杂度更低且有效的方法从原始语料库中确定待消解项是提升模型效率非常关键的一点。

(3) 在中文零指代消解模型设计的过程，本文仅考虑在给定可消解中文零指代词及其候选先行语的条件下来确定中文零指代词所在的共指链。这种先验信息在实际的中文零指代消解任务中并不能很好的获取，因此如何对一个句子中的零指代词的位置精准的检测对于中文零指代消解的实际应用具有重大意义。

(4) 在对中文零指代的研究过程中，我们发现先验知识对于零指代词的消解非常重要。在确定零指代词后，将零指代词及其候选先行语同外部知识库进行链接，在一定程度上不仅仅可以使得模型的可解释性更强，也对提升模型的性能也有一定的帮助。

参考文献

1. Han X, Gao T, Lin Y, et al. More Data, More Relations, More Context and More Openness: A Review and Outlook for Relation Extraction[C]//Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing. 2020: 745-758.
2. Ji S, Pan S, Cambria E, et al. A Survey on Knowledge Graphs: Representation, Acquisition and Applications[J]. arXiv e-prints, 2020: arXiv: 2002.00388.
3. Ma S, Deng Z H, Yang Y. An unsupervised multi-document summarization framework based on neural document model[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 1514-1523.
4. Lin Y, Liu Z, Sun M. Neural relation extraction with multi-lingual attention[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017: 34-43..
5. 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589.
6. Yao J, Wan X, Xiao J. Phrase-based compressive cross-language summarization[C]//Proceedings of the 2015 conference on empirical methods in natural language processing. 2015: 118-127.
7. 宋洋, 王厚峰. 共指消解研究方法综述[J]. 中文信息学报, 2015, 29(1): 1-12.
8. 王厚峰. 指代消解的基本方法和实现技术[J]. 中文信息学报, 2002, 16(6): 10-18.
9. Pradhan S, Moschitti A, Xue N, et al. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes[C]//Joint Conference on EMNLP and CoNLL-Shared Task. 2012: 1-40.
10. Kim Y-J. Subject/Object Drop in the Acquisition of Korean: A Cross-Linguistic Comparison [J]. Journal of East Asian Linguistics. 2000, 9 (4): 325–351.
11. Hobbs J R. Resolving pronoun references[J]. Lingua, 1978, 44(4): 311-338.
12. B. Grosz, A. Joshi, S. Weinstein. Centering: A framework for modelling the local coherence of discourse[J]. Journal of Computational Linguistics, 1995, 21(2): 203-225.
13. Raghunathan K, Lee H, Rangarajan S, et al. A multi-pass sieve for coreference resolution[C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. 2010: 492-501.
14. Ma C, Doppa J R, Orr J W, et al. Prune-and-score: Learning for greedy coreference resolution[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014: 2115-2126.
15. Wiseman S, Rush A M, Shieber S M, et al. Learning Anaphoricity and Antecedent Ranking Features for Coreference Resolution[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015: 1416-1426.
16. Peng H, Chang K W, Roth D. A joint framework for coreference resolution and mention head detection[C]//Proceedings of the Nineteenth Conference on Computational Natural Language Learning. 2015: 12-21.
17. Clark K, Manning C D. Improving Coreference Resolution by Learning Entity-Level Distributed Representations[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 643-653.
18. Lee K, He L, Lewis M, et al. End-to-end neural coreference resolution[J]. arXiv preprint arXiv:1707.07045, 2017.
19. Kundu G, Sil A, Florian R, et al. Neural cross-lingual coreference resolution and its application to entity linking[J]. arXiv preprint arXiv:1806.10201, 2018.
20. Zhang R, dos Santos C, Yasunaga M, et al. Neural Coreference Resolution with Deep Biaffine Attention by Joint Mention Detection and Mention Clustering[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2018: 102-107.
21. Zhang H, Song Y, Song Y. Incorporating Context and External Knowledge for Pronoun Coreference Resolution[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: 872-881.
22. Lee K, He L, Zettlemoyer L. Higher-Order Coreference Resolution with Coarse-to-Fine Inference[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). 2018: 687-692.
23. Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
24. Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C] //Proceedings of the 2014 conference on empirical methods in natural language processing. 2014: 1532-1543.
25. Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. arXiv preprint arXiv:1802.05365, 2018.
26. Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
27. Kantor B, Globerson A. Coreference resolution with entity equalization[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 673-677.
28. Joshi M, Levy O, Weld D S, et al. BERT for coreference resolution: Baselines and analysis[J]. arXiv preprint arXiv:1908.09091, 2019
29. 杨国庆,孔芳,朱巧明等.基于规则的中文缺省识别研究[J].计算机科学, 2011, 38(12): 255-257.
30. 宋洋,王厚峰.基于马尔可夫逻辑的中文零指代消解[J].计算机研究与展, 2015, 52(9): 2114-2122.
31. Converse S P, Palmer M S. Pronominal anaphora resolution in Chinese[M]. University of Pennsylvania, 2006.
32. Kong F, Zhou G. A Tree Kernel-Based Unified Framework for Chinese Zero Anaphora Resolution[C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. 2010: 882-891.
33. Zhao S, Ng H T.Identification and Resolution of Chinese Zero Pronouns: A Machine Learning Approach[C]//EMNLP-CoNLL 2007,Proceedings of the 2007Joint Conference on Empirical Methods in Natural Language Processing and Com-putational Natural Language Learning, June 28-30, 2007, Prague, Czech Republic.2007 : 541 -550.
34. Chen C, Ng V. Chinese zero pronoun resolution: an unsupervised approach combining ranking and integer linear programming[C]//Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. 2014: 1622-1628.
35. Chen C, Ng V. Chinese zero pronoun resolution with deep neural networks[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 778-788.
36. Yin Q, Zhang W, Zhang Y, et al. A Deep Neural Network for Chinese Zero Pronoun Resolution[J]. arXiv e-prints, 2016: arXiv: 1604.05800.
37. Liu T, Cui Y, Yin Q, et al. Generating and exploiting large-scale pseudo training data for zero pronoun resolution[J]. arXiv preprint arXiv:1606.01603, 2016.
38. Yin Q, Zhang Y, Zhang W, et al. Chinese zero pronoun resolution with deep memory network[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. 2017: 1309-1318.
39. Hochreiter S, Schmidhuber J. Long short-term memory.[J]. Neural Computation, 1997, 9(8): 1735-1780.
40. Kong F, Zhang M, Zhou G. Chinese zero pronoun resolution: A chain-to-chain approach[J]. ACM Transactions on Asian and Low-Resource Language Information Processing, 2019, 19(1): 1-21.
41. Yin Q, Zhang Y, Zhang W, et al. Deep Reinforcement Learning for Chinese Zero Pronoun Resolution[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 569-578.
42. Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv: 1409.2329, 2014.
43. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks, 2005, 18(5-6): 602-610.
44. Duchi J, Hazan E, Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization[J]. Journal of Machine Learning Research, 2011, 12: 2121-2159.
45. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.
46. Elman J L. Finding Structure in Time[J]. Cognitive Science, 1990, 14(2):179-211.
47. Li X, Wu X. Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2015: 4520-4524.
48. Sak H, Senior A, Beaufays F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition[J]. arXiv preprint arXiv:1402.1128, 2014.
49. Gers F A, Schraudolph N N, Schmidhuber J. Learning precise timing with LSTM recurrent networks[J]. Journal of machine learning research, 2002, 3(Aug): 115-143.
50. Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[J]. arXiv preprint arXiv:1609.08144,2016
51. 赵富, 杨洋, 蒋瑞, 等. 融合词性的双注意力 BiLSTM 情感分析[J]. 计算机应用, 2018, 38(S2): 108-111.
52. Arulkumaran K, Deisenroth M P, Brundage M, et al. A brief survey of deep reinforcement learning[J]. arXiv preprint arXiv:1708.05866, 2017.
53. Branavan SR K, Silver D, Barzilay R. Learning to Win by Reading Manuals in aMonte-Carlo Framework [J]. J. Artif. Intell. Res., 2012,43:661-704.
54. Narasimhan K, KulkarniT D, Barzilay R. Language Understanding for Text-based Games using Deep Reinforcement Learning[C]//Proceedings of the 2015 Confer-ence on Empirical Methods in Natural Language Processing, EMNLP 2015, Lis-bon, Portugal, September 17-21,2015.2015 : 1-11.
55. Li J, Monroe W, Ritter A, et al. Deep Reinforcement Learning for Dialogue Gen-eration[C] //Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016,Austin,Texas,USA, November 1-4, 2016.2016: 1192-1202.
56. Clark K, Manning C D. Deep Reinforcement Learning for Mention-Ranking Coreference Models[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 2256-2262.
57. Feng J, Huang M, Zhao L, et al. Reinforcement Learning for Relation Classification from Noisy Data[J]. arXiv preprint arXiv:1808.08013, 2018.
58. Li J, Monroe W, Ritter A, et al. Deep Reinforcement Learning for Dialogue Gen-eration[C] //Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016,Austin,Texas,USA, November 1-4, 2016.2016: 1192-1202.
59. Xiong W, Hoang T, Wang W Y. Deep Path: A Reinforcement Learning Method for Knowledge Graph Reasoning[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017.2017: 564-573.
60. Williams R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Machine learning, 1992, 8(3-4): 229-256.
61. Zhang Y, Wallace B C. A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification[C]//Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2017: 253-263.
62. Vilain M, Burger J, Aberdeen J, et al. A model-theoretic coreference scoring scheme[C]//Proceedings of the 6th conference on Message understanding. 1995: 45-52.
63. Luo X. On coreference resolution performance metrics[C]//Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. 2005: 25-32.
64. Martschat S, Strube M. Latent structures for coreference resolution[J]. Transactions of the Association for Computational Linguistics, 2015, 3: 405-418.
65. Clark K, Manning C D. Entity-Centric Coreference Resolution with Model Stacking[C] //Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015: 1405-1415.
66. Wiseman S, Rush A M, Shieber S M. Learning Global Features for Coreference Resolution[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 994-1004.
67. Fei H, Li X, Li D, et al. End-to-end deep reinforcement learning based coreference resolution[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 660-665.
68. Chen C, Ng V. Chinese zero pronoun resolution: A joint unsupervised discourse-aware model rivaling state-of-the-art resolvers[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). 2015: 320-326.
69. Yin Q, Zhang Y, Zhang W, et al. Zero pronoun resolution with attention-based neural network[C]//Proceedings of the 27th international conference on computational linguistics. 2018: 13-23.

致 谢

攻硕期间的科研成果及获奖情况