

文章编号: 1003-0077(2025)11-0050-16

动词驱动事件的共指关系中文语料库构建及大模型评测

王皓泽¹, 辛欣^{1,2}

(1. 北京理工大学 计算机学院, 北京 100081;

2. 北京理工大学 北京市海量语言信息处理与云计算应用工程技术研究中心, 北京 100081)

摘要: 事件共指关系研究是自然语言处理领域的一项基础任务。现有工作主要在封闭域上对事件共指进行研究。为了在开放域上研究事件共指关系, 该文提出直接从事件的表达载体出发定义事件共指的研究思路, 并具体讨论了动词驱动的事件之间的共指关系。该文构建了动词驱动事件共指关系的数据规范, 并将共指关系细分为强共指与弱共指。基于数据规范, 人工标注构建了 38 万余字的中文语料库, 含 1 000 篇文档, 29 250 个动词, 1 858 条强共指链, 994 个弱共指动词对。并进行了流水线共指消解实验: 提及检测阶段, 采用成分句法分析算法提取动词成分, F_1 值为 96.9%; 共指关系预测阶段, 分别采用提示工程、指令微调、知识蒸馏评测了大语言模型在该任务上的表现, 比较模型预测的平均 F_1 值, 其中, 提示工程方法为 50.0%, 指令微调方法为 64.4%, 知识蒸馏方法为 71.7%。将经指令微调的大模型与经有监督微调的小模型进行对比, 小模型的推理表现略低于大模型, 推理效率远高于大模型。

关键词: 事件共指关系; 中文语料库; 大语言模型

中图分类号: TP391

文献标识码: A

Chinese Coreference Corpus from Verb-Driven Events and Its Application for LLM Evaluation

WANG Haoze¹, XIN Xin^{1,2}

(1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China;

2. Beijing Engineering Research Center of High Volume Language Information Processing and Cloud Computing Applications, Beijing Institute of Technology, Beijing 100081, China)

Abstract: The study of event coreference is a fundamental task in the field of natural language processing. To investigate coreference between all events in open domains, an approach that defines event coreference directly from the event expressions is proposed, with a specific focus on verb-driven events. A guideline for verb-driven event coreference is further composed, which divides event coreference into strong and weak coreference. Then, a manually annotated Chinese corpus consisting of over 380 thousand characters is constructed, comprising 1,000 documents, 29,250 verbs, 1,858 strong coreference chains and 994 weak coreference verb pairs. A pipeline experiments for coreference resolution is conducted, which demonstrate an F_1 score of 96.9 in mention detection phase, and average F_1 scores of 50.0, 64.4, and 71.7 for coreference prediction via prompt engineering, instruction tuning and knowledge distillation of LLM, respectively.

Keywords: event coreference; Chinese corpus; large language model

0 引言

事件共指, 是文本中多个提及指代相同事件的

语言现象, 而事件共指消解是自然语言处理领域的一项基础任务^[1]。对事件共指关系以及共指消解的研究具有现实意义: 在知识工程领域, 该任务可以促进对事件内在联系和演化规律的把握, 为知识库的

收稿日期: 2024-05-08 定稿日期: 2024-08-29

基金项目: 国家自然科学基金(62172044)

扩充、事理图谱的构建等工作提供有效支撑；在语言理解与生成领域，该任务可以帮助语言模型捕捉并利用事件之间的结构化信息，有助于提升模型的深层次理解能力和连贯生成能力。

现有对事件共指的研究大多在特定类型的事件上进行，属于封闭域，如图 1(a)所示。例如，ACE 2005^[2]将事件分类为 8 个大类和 33 个子类；ECB +^[3]数据集讨论了 14 类事件；KAC-KBP^[4]限定的事件类型为 38 类；MAVEN-ERE^[5]讨论了 168 类事件。这些工作基于有限事件类型的事件体系，研究特定事件之间的指代关系。

封闭域事件共指研究存在的问题是，它不能覆盖所有类型的事件。首先，事件种类繁多，从宏观社会政治变化到个人行动与心理状态，范畴相当广泛，如何在有限的类别中覆盖所有现实事件，同时保证类别的互斥性与完备性，是一个艰巨的任务；其次，不同领域对事件的关注并不相同，一个统一的分类体系难以满足所有领域需求；再者，随着社会的发展和现象的出现，新的事件类型会不断涌现，分类体系需要不断更新。

为了解决封闭域事件共指研究覆盖事件不全面的问题，本文开展开放域事件共指研究，目的是对现实世界中的事件进行完整而全面的结构化构建。由上文分析可知，在开放域上，把所有的事件类型都覆盖比较困难。但事件共指是一种事件与另一种事件之间的关系，事件间的关系与事件类型原本就属于不同的概念，定义事件的关系不一定需要遍历所有的事件类型。对于共指关系来说，其关注的是两个事件对应的动作（或复杂动作的集合）是否相同，这样的定义可在不同的事件类型上进行泛化。例如，假定当前定义的事件类型体系中仅有 ACE 2005^[2]的 33 个事件。在图 1 中，“建设”与“兴建”所描述的事件并不在该体系中。但当我们根据上面对共指关系的描述，仍然能够分辨出这里的“建设”与“兴建”

指的是相同的动作，是共指的，即使该体系并未覆盖“建设”这一事件。因此，本文的主要想法是直接去定义事件的共指关系，使得该定义不依赖于事件类型的定义，从而使事件共指所涉及的事件类型从传统的封闭域扩展为开放域。具体的做法是直接通过事件的表达载体对事件共指进行定义，如图 1(b)所示。一般来说，事件通过句子中的动词、名词等触发，其角色可以通过句法语义分析获得。如果定义了动词和动词的共指、动词和名词的共指、名词和名词的共指等，就可以覆盖开放域事件的共指关系。这种关系在不同的事件类型下具有泛化性。本文主要关注动词和动词之间的共指关系，动词和名词、名词和名词等其他共指关系将在未来工作中展开。

定义开放域中动词驱动事件的共指关系存在的挑战是：同一事件在不同的动词载体中存在复杂的逻辑关系，如动词词义包含、角色包含和角色变更，这为共指关系的定义带来困难。直观上，如果两个动词所指事件是共指的，那么它们应当具有相同的含义，拥有相同的参与实体、发生时间、地点等，如表 1-例 1 所示；而互不共指的动词在动词词义、角色等方面存在差异，如表 1-例 2 所示。但是，存在一些特殊情况，例如：表 1-例 3 中，动词“任免”和“任”的动作含义存在蕴含的关系，根据上下文可知，这两个动词所指事件的参与者、时间等均存在重叠的部分；表 1-例 4 中，动词“获得”的主体角色均为“居里夫人”，但两个动词的发生时间、客体角色等并不完全相同，而是存在包含的关系。例 3 和例 4 中，动词所指的事件并不完全相同，但两组事件之间又分别存在着关联关系：可以认为一个动词所指事件包含了另一个动词所指事件；表 1-例 5 中，虽然两个动词“展开”都对应着“他们展开实验”这件事，但这两个动词的发生时间又不完全一致，事件提及对应着同一个事件的不同发生情况。

基于以上所提出的挑战，本文对事件共指关系做了更为细致的分类：将动词驱动事件的共指关系分为强共指与弱共指。其中，强共指讨论的是动词所指事件相同的情况，其中要求动词的含义、各类语义角色均相互一致；弱共指讨论的是不完全满足强共指条件，但动词所指事件存在含义上、语义角色上的包含或者变更，动词所指事件存在相关关系的情形。

针对上述任务背景，本文在分析 20 万余字中文语料的基础上制定了动词驱动事件的共指关系数据

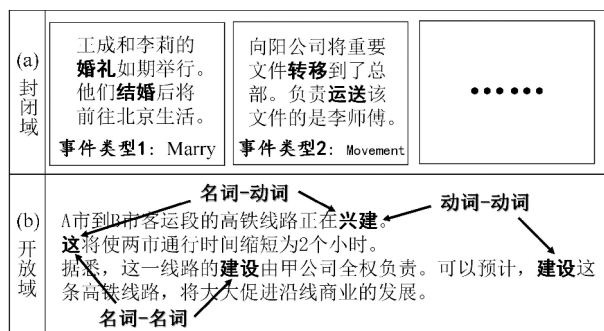


图 1 封闭域与开放域事件共指研究对比

表 1 共指关系句例及解释

	句例	解释
强共指	例 1: 某校共派出 20 名运动员参加 ¹ 比赛。 这些参赛 ² 的运动员均已做好充分的准备	“参加 ¹ ”和“参赛 ² ”均指代同一个具体的事件,即“某校的 20 名运动员参加比赛”
不共指	例 2: 我国成功举办 ¹ 2008 年夏季奥运会。 我国成功举办 ² 2022 年冬季奥运会	“举办 ¹ ”指代“我国举办 2008 年奥运会”, “举办 ² ”指代“我国举办 2022 年奥运会”
弱共指	词义包含 例 3: 该企业任免 ¹ 一批员工。 其中,任 ² 小张为人事部经理	“任免 ¹ ”指代“企业任命、免职了一批员工”的事件,其包含了动词“任 ² ”所指代的“企业任命小张”这一事件
	角色包含 例 4: 居里夫人分别获得 ¹ 诺贝尔物理学奖和诺贝尔化学奖。 1911 年,居里夫人因发现元素钋和镭获得 ² 诺贝尔化学奖	“获得 ¹ ”指代“居里夫人两次获得诺贝尔奖”的事件,其包含动词“获得 ² ”所指代的“居里夫人第二次获得诺贝尔奖”这一事件
	角色变更 例 5: 他们原定于下午展开 ¹ 实验,但因为意外,推迟至明日展开 ²	“展开 ¹ ”和“展开 ² ”分别指代了“他们展开实验”这一事件的不同发生情况

规范,并人工标注构建了中文语料库;为进一步分析当前语言模型在解决该任务上的难点与挑战,并在中文通用大语言模型上进行了实验和分析。综上所述,本文的主要贡献有以下 3 点:

(1) 提出动词驱动事件的共指关系数据规范:规范中,事件共指分为强共指与弱共指。弱共指中,又包括动词词义包含、动词角色包含、动词角色变更等三类情况。

(2) 构建动词驱动事件的共指关系中文语料库:该语料库包含 1 000 篇文档,共计 38 万余字,1 858 条强共指链,994 个弱共指动词对。本文所构建的语料资源已公布于 <https://github.com/NP-NET-research/Reference-Resolution-for-Chinese-Verbs>。

(3) 大语言模型评测:本文进行了流水线的共指消解^[6-7]实验:提及检测阶段采用成分句法分析^[8]抽取文本中的动词;共指关系预测阶段分别采用提示工程^[9]、指令微调^[10]、知识蒸馏^[11]三种方式在该语料库上对智谱清言模型^[12]、ChatGLM3-6B 模型^[13]、QWen-7B-Chat^[14]模型进行了实验,并分析了现有大模型在解决该任务时存在的问题。此外,本文还将大模型与有监督微调的小模型^[15]在推理表现和效率上进行了综合对比。

1 相关研究

1.1 事件共指研究

目前,对事件共指关系的研究主要集中在封闭

域中,是在封闭域事件抽取的基础上针对所抽取的特定类型的事件进行的研究。其中有些工作并不针对特定场景,如 ACE^[2]、ECB+^[3]等,还有一些工作将事件限定在军事^[16]、金融^[17]等特定场景,并定义了该场景中常见事件的类型。由于这些工作将事件类型限定在封闭域上,均难以覆盖现实世界的全部动作。

有少部分工作没有限定事件的类型。OntoNotes^[18]提出了综合实体与事件共指标注的数据集,针对动词成分,该工作规定,动词被标注,当且仅当该动词与其他名词短语构成共指关系,这遗漏了动词与动词之间可能存在的共指。此外,该工作也忽略了那些不能被锚定到时间轴上的一般性事件。RED^[19]将事件定义为任何发生、动作、过程等,基于该定义依然没能全面覆盖现实事件,例如,它忽略了“拥有”“认为”等表示静止不变状态的动作所代表的事件。

本工作在开放域上展开,不限定事件的特定类型,主要关注中文篇章中动词等表达载体所指的事件。所构建的共指数据既包括能被锚定到现实时间点上的具体事件,也包括不能被锚定到具体时间点上的一般事件,同时,也涉及表示静态概念的静态事件。表 2 展示了当前事件共指研究工作与本文所做工作在事件上的覆盖情况区别。

1.2 事件共指数据规范制定

目前被广泛使用的几种事件共指语料在数据规范制定方面对事件共指的介绍多停留在概念性声明,

表 2 事件共指研究对事件覆盖情况对比

数据集	动态事件		静态事件
	具体事件	一般事件	
示例	李华参观了 非遗博物馆	他经常跑步	我拥有这本书
ACE	✓		
OntoNotes	✓	—	
ECB+	✓		—
KAC-KBP	✓		
RED	✓	✓	
MAVEN-ERE	✓	✓	
本文工作	✓	✓	✓

缺乏详细的定义讨论。例如,在 ACE 中文事件标注规范^[2]中,仅用一页篇幅对事件共指进行了介绍,内容仅包括事件共指的概念声明,并列举了事件共指的两条反例。同样地,OntoNotes^[18]在规范中对事件共指的规定也仅限于简单的概念层面。ECB+^[3]提出从动作、人类参与者、非人类参与者、地点、时间等事件属性出发讨论事件共指,但未对这四类属性的一致性分别进行更细致地介绍。RED^[19]提出了一个综合实体、事件、时间属性及其关系的统一规范,在事件共指方面,将事件共指分为 IDENTITY、SET-MEMBER、BRIDGING 三类,并具体举例说明,但并未从事件角色等要素方面对这三类事件的关系进行详细的形式论述。

对此,本文对事件共指进行了更为详细的论述。由于本文重点关注动词驱动的事件共指情况,除了给出概念说明外,本文还结合动词词义与语义角色特征对动词的共指情况进行了分类讨论。在动词含义的讨论中,以 HowNet 义原体系^[20]为参照;在语义角色的讨论中,则引用了袁毓林构建的论元角色体系^[21],细致探讨动词的各种角色在动词驱动的事件共指关系中的作用与影响。

1.3 事件共指语言资源建设

当前,事件共指语言资源建设主要集中在英文领域,并已经积累了丰富的资源^[2-5]。相比之下,中文事件共指语料的构建,特别是在开放域方面的工作较为匮乏。尽管部分工作针对中文数据构建了封闭域事件抽取语料^[22-23],但这些工作中尚未进行事件共指任务。目前,对中文事件共指的算法研究与模型评估^[24-26]主要依赖于 ACE 2005 数据^[2],该数据集覆盖的事件类型有限,未能充分反映中文文本

中复杂多样的现实事件。

为丰富中文语言资源建设,本文基于人民日报标注语料库^[27-28]这一高质量中文语料^①,构建了开放域动词驱动事件共指中文语料库,以推动中文语料下事件共指消解的方法研究。

2 动词驱动事件共指关系规范

2.1 总体概述

根据指代相同或相关事件的动词在含义与动词角色方面的一致性与相关性,本文将动词驱动事件共指关系分为强共指与弱共指。其中,“强共指”表示文本中多个动词分别驱动完全相同的事件;“弱共指”表示多个动词并非驱动完全相同的事件,但这些事件存在次关联关系。本文将“弱共指”细分为三类:“词义包含”“角色包含”与“角色变更”。其中,“词义包含”的动词含义存在包含关系,所驱动事件存在包含关系;“角色包含”动词的角色间存在包含关系,且所驱动事件存在包含关系;“角色变更”的动词驱动的事件根据不同的动词角色条件展现出不同的发生情境。动词驱动事件共指关系数据规范框架如图 2 所示。本节对规范体系进行介绍,内容安排如下。

(1) 共指关系影响要素:2.2 节对影响动词驱动事件共指关系的词义要素和语义角色要素进行介绍讨论。

(2) 强共指关系:2.3 节对强共指关系(A)进行介绍。其中,分别讨论主(客)体角色(A1)、时间角色(A2)、空间角色(A3)、命题角色(A4)对强共指关系的影响。

(3) 弱共指关系:2.4 节对弱共指关系进行介绍,分别于 2.4.1 节论述动词词义包含(B)的弱共指关系、2.4.2 节论述动词角色包含(C)的弱共指关系、2.4.3 节论述动词角色变更(D)的弱共指关系,其中分别论述主(客)体角色(D1)、时间角色(D2)、空间角色(D3)、命题角色(D4)对弱共指关系的影响。

2.2 动词驱动事件共指关系影响要素

2.2.1 动词词义影响共指关系

一般来说,共指事件对应的动词的词义应当是

① 语料资源下载自 <https://klcl.pku.edu.cn/gxzy/231686.htm>

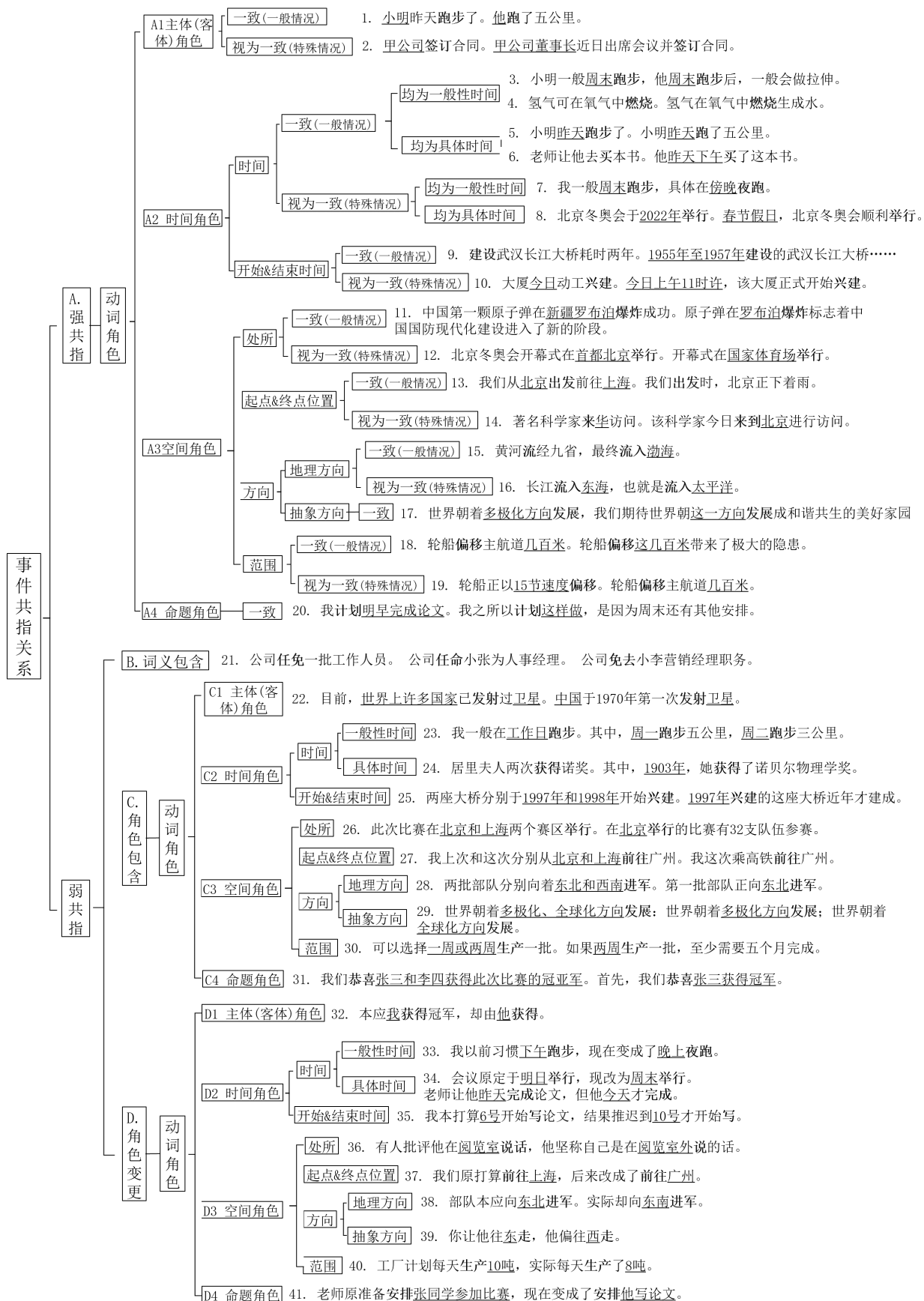


图2 数据规范总体架构

例句中, 动词加粗标识, 相应角色下划线标识。

相同的;但也存在特殊情况,即共指事件对应的动词拥有不同的词义。

一般情况 共指事件对应动词的词义相同。

直观上,由于动词含义直接反映了该动词所驱动事件的含义,所以可以用动词的词义来表达其指代事件的含义。共指事件对应的动词在文本中一般具有相同或相近的动词词义。“参加会议”和“举行会议”中,动词“参加”和“举行”的含义不同,因此,二者并不共指。

特殊情况 共指事件对应动词的词义不同。

当一个动词在文本中被赋予了隐喻或延伸意义时,它可能会与另一个字面意义相异的动词共同指代同一事件,这时我们可以探讨这两者之间是否存在共指关系。比如,在“产业界与学术界协作”与“产业界与学术界‘**联姻**’”这两个表达中,动词“联姻”通过隐喻手法实际上意指“协作”,表明两者描述的是同一个事件。

在动词词义一致性的判别过程中,本文引用了HowNet^[20]动词义原标签。判定两个动词是否具有共指关系时,考虑两个动词的所有候选义原,如果二者的词义相同,那么它们一般具有相同的候选义原,进而进行后续共指关系的分析。对于被赋予隐含意义的动词,我们不能按照其字面意义进行分析,而应当保持其在文本中的特定含义。

此外,对于多义动词,应考虑其在上下文中的具体含义,并对其加以区分。例如,句子“我**燃烧**¹了一把干柴”和“干柴正在**燃烧**²”中,动词“**燃烧**¹”的含义为点燃的行为,动词“**燃烧**²”的含义为“燃烧”这种氧化反应。

2.2.2 动词语义角色影响共指关系

动词的语义角色描述了该动词所指事件的参与者、时间、空间属性、原因、结果等要素。直观上,不同事件的这些要素存在差异,因此动词角色可以被用于区分不同事件。其中,部分角色对共指关系具有显著影响,它们之间的异同决定着事件的异同,另外的一些角色对共指关系并无显著影响,实际标注中通常不会关注这些角色。本文对这两类动词角色进行举例分析。

对共指关系有影响的角 一般而言,主客体、环境(处所、时间、范围)、命题等动词角色的一致性与事件共指密切相关,如果两个动词的这些角色不一致,那么二者一般驱动不同的事件,用表 3 中例句说明。实际数据标注过程中,应当对这些角色做重点关注。

表 3 动词角色对事件共指的影响举例说明

影响类型	角色	例句
对共指关系有影响	主体	(非共指) <u>我</u> 昨天跑步了, <u>他</u> 也跑步了
	客体	(非共指) 他 准备 了饺子,也 准备 了汤圆
	处所	(非共指) 我获得两次冠军:在 <u>北京</u> 获得、在 <u>上海</u> 获得
	时间	(非共指) 张三 <u>昨天</u> 吃了米饭, <u>前天</u> 也吃了米饭
	范围	(非共指) 我跑了 <u>两公里</u> ;我跑了 <u>五公里</u>
	命题	(非共指) <u>他选择辞职</u> 影响了公司的发展; <u>合作失败</u> 也影响了公司的发展
对共指关系无影响	工具	(共指) 我 完成 这个项目用了 <u>一台服务器</u> ,还用 <u>一张显卡</u> 辅助 完成
	材料	(共指) 北京奥运会金牌由 <u>镀金银</u> 制成,制作金牌还使用了 <u>产自青海的昆仑玉</u>
	方式	(共指) 他轻轻地 走 ,慢慢地 走
	原因	(共指) 他为 <u>谋生</u> 而工作,也为 <u>社会责任</u> 而工作

对共指关系无影响的角 一般而言,工具、材料、方式、原因等角色的一致性与共指关系并不密切相关。即使两个动词的这些角色不一致,二者仍可能指代相同的事件,用表 3 中例句说明。在数据标注过程中,一般不对这些动词角色做重点关注。

在后文规范说明中,所讨论的动词角色均为对共指关系有影响的角

2.3 强共指关系规范(A)

强共指,描述的是文本中两个动词驱动完全相同的事件的情况。

一般来说,两个事件为强共指,那么它们所对应的动词含义应当是相同的。在此基础上,二者的语义角色也应保持一致。也有一种例外,即使两个动词具有不同的动词角色,二者仍可能驱动同一个事件。本节对动词的各类角色进行讨论。

A1 主体(客体)角 一般来说,在强共指下,事件对应动词的主、客体角色,包括施事、感事、受事、与事等等,是需要保持一致的,即对应的主客体角色指代同一个实体。例如,在图 2 例句 1 中,“小明”与“他”指代了同一个实体。此外,如果其中一个动词的角色是集合,那么另一个动词的相应角色也应是相等的集合。

除此之外,也有一些特殊情况。虽然主客体角

色并不相同,但仍然可以将其归纳为强共指。例如,图 2 例句 2 中,“甲公司”和“甲公司董事长”并不指代同一实体,但动词“签订”却驱动了同一个事件。这类情况中,两个实体一般存在着间接回指关系,也可以认为,“甲公司董事长”是对“甲公司”的递进与补充说明。

A2 时间角色 一般来说,如果两个动词所驱动事件为强共指,那么这两个事件会具有相同的发生时间,对应到动词上,动词的时间角色,包括一般性时间、具体时间、开始时间、结束时间等,是需要保持一致的。按照时间是否被锚定到现实的时间段或时间点上,可将时间分为具体时间和一般性时间。图 2 例句 3 中,两个时间“周末”都是一般性时间,虽然无法将其锚定到时间轴上,但仍可看出二者表示的是同一时间。图 2 例句 4 中动词“燃烧”对应着一个客观常识事件,本身无时间概念,本文将这类情况也归为一般性时间。对于具体时间,则需要考察这两个时间在时间轴上的位置、长度是否一致。图 2 例句 5 中,根据上下文不难分析出,动词“跑步”和“跑”具有相同的发生时间;图 2 例句 6 中,即使文中未提及第一个动词“买”的发生时间,但仍可认为它与第二个动词“买”会同时发生。开始、结束时间的分析与具体时间的分析类似。需要特别说明的是,具有一般性时间的动词驱动了一般性的事件,而具有具体时间的动词驱动了具体事件,这两种事件不存在共指关系。

除此之外,也有一种特殊情况,即便两个动词的时间角色不同,二者仍可能驱动强共指事件。图 2 例句 7 和 8 中,“周末”与“傍晚”时间不同,“2022 年”和“春假假日”时间不同,但是可以认为第二个时间描述是对第一个时间描述的补充和详尽说明,可以认为两个动词的发生时间是一致的。

A3 空间角色 一般来说,在强共指下,事件所对应动词的空间角色,包括处所、起点、终点、方向、范围等,应当是一致的。图 2 例句 11、13 中,地理处所“新疆罗布泊”和“罗布泊”指代了相同的处所,两个动词“出发”的地理起点均为“北京”,地理终点均为“上海”。图 2 例句 15 中,动词“流”和“流入”的地理方向一致,图 2 例句 18 中,两个动词“偏移”的范围也相互一致。此外,动词的抽象方向,通常描述事物的发展趋向,例如图 2 例句 17 中的“多极化方向”和“这一方向”,也应当保持一致。显然,由于地理方向和抽象方向的概念本身不同,所以不能将一个动词的地理方向角色同另一个动词的抽象方向角色进

行对比。

除此之外,存在一种特殊情况:即使两个动词的空间角色不同,它们所驱动事件仍可以是强共指的。如图 2 例句 12 所示,处所“首都北京”和“国家体育场”并不一致;终点“华”和“北京”并不一致;方向“东海”和“太平洋”也并不一致,但是可以认为第二个动词的处所角色是对第一个动词相应角色的补充说明,两个角色并不存在冲突。

A4 命题角色 一般来说,在强共指下,动词的命题角色(一般为另一个事件)也应当是共指的。如图 2 例句 20 中,“明早完成论文”和“这样做”指代了同一个事件。

本节对每种角色进行讨论时,均涉及角色不同但事件强共指的这类特殊情况,这与不共指均具有角色不同的共性。对二者进行对比:强共指时,虽然事件对应动词的角色不同,但它们一般不构成冲突,可以认为一个动词的角色是对另一个动词相应角色的补充递进说明。例如“某公司”与“某公司董事长”、“春天”与“4 月 20 号”;而不共指的动词的角色之间通常明显存在冲突,例如“甲公司”与“乙公司”、“4 月 20 号”和“4 月 30 号”。

2.4 弱共指关系规范(B-D)

弱共指,描述的是指代相关性事件、但对应动词的角色并不完全一致、所指事件也并非强共指的情形。按照动词词义与动词语义角色间的关系,又可以将弱共指分为以下三类:

词义包含 动词 A 的含义包含了动词 B 的含义。进一步地,可以认为动词 A 所指事件包含动词 B 所指事件。

角色包含 动词 A 的若干动词角色包含动词 B 的若干对应角色,进一步地,可以认为动词 A 所指事件包含动词 B 所指事件。

角色变更 动词 A 与动词 B 指代同一类事件,但部分角色之间存在差异,它们指代同一事件在不同角色条件下的不同发生情况。下面对这几类弱共指进行介绍。

2.4.1 词义包含(B)

直观上,动词的词义反映了该动词所驱动事件的核心意义。如果动词 A 的词义包含了动词 B 的词义,则可以推测动词 A 所驱动事件的含义也可能包含动词 B 所驱动事件的含义。如图 2 例句 21 中,动词“任免”含有动词“任命”和“免除”的含义,类似的例子还有“进出口”与“进口”,“开合”与“打开”

等。在这一条件的基础上,词义包含的弱共指进一步要求,两个动词所指事件也存在包含关系。例如图 2 例句 21 中,“任免”指代“任免一批工作人员”这件事,这事件自然地覆盖动词“任命”所指代的“任命小张”这一子事件。

2.4.2 角色包含(C)

在动词词义相同的前提下,两个动词的某几个角色之间如果存在包含关系,例如,“我们昨天跑步了,其中,我跑了五公里。”中,“跑步”的主体角色“我们”包含了“跑”的主体角色“我”,我们可以认为,“跑步”所指事件包含“跑”所指事件。本文把这类情况归纳为角色包含的弱共指。以下对各类角色的包含情况加以讨论。

C1 主体(客体)角色 主体(客体)角色一般是实体类型,它们之间的包含关系通常有“集合-元素”关系或“集合-子集”关系两种形式。符号描述为:

记三个实体分别为 e_1, e_2, e_3 , 两种形式的角色包含关系分别是:

“集合-元素”: $\{e_1, e_2, e_3\}$ 包含 e_1 ;

“集合-子集”: $\{e_1, e_2, e_3\}$ 包含 $\{e_1, e_2\}$ 。

在中文语料中,常见以下两种主体(客体)角色包含表述方式:

① 由连接词体现出的包含,如实体“沃森和克里克”包含实体“克里克”。

② 实体概念所蕴含的包含,如图 2 例句 22,实体“世界上许多国家”包含实体“中国”。

特别说明:“整体-部分”包含关系,如“人体”和“手指”;“类型-实例”包含关系,如“水”和“这瓶水”。这类包含关系不属本文考虑范畴。

C2 时间角色 时间角色的包含存在以下三类典型情况:

① 时间段包含时间段:如“2022 年至 2024 年”包含“2023 年全年”。

② 时间段包含时间点:如“2023 年全年”包含“神舟十七号载人飞船点火发射时”。

③ 时间点的集合包含时间点(的集合):符号描述为:记 4 个时间点: t_1, t_2, t_3, t_4 , 有 $\{t_1, t_2\}$ 包含 t_1 , 或 $\{t_2, t_3, t_4\}$ 包含 $\{t_3, t_4\}$ 。

特别说明:“周末”和“上周六”这种一般性时间与具体时间的关系不属该部分讨论范畴。

C3 空间角色 两个动词的空间角色可以存在包含关系,例如图 2 例句 26、28、30 中,“北京和上海”包含更小的地理单位“北京”;同样地,“东北和西南”这一复合方向包含方向“西南”;“一周或两周”这

一范围的表述包含“两周”。再次强调,需明确区分地理方向与抽象方向,两者不宜直接进行比较,因为它们属于不同的维度范畴。

C4 命题角色 动词的命题角色(通常指代事件)之间可以存在包含关系。如图 2 例句 31 中,事件“张三和李四获得此次比赛的冠亚军”包含事件“张三获得冠军”。命题角色的包含可以是本文提到的“词义包含”和“角色包含”,也可以是在其他事件表达载体配对中所体现的事件包含关系。

在分析两个动词的角色时,可能会识别出多对包含关系。以句子“在过去的几天里,同学们跑步了,其中,小张昨天跑了五公里。”为例,主体角色“同学们”包含实体“小张”;时间角色“过去的几天”包含时间“昨天”,动词“跑步”所驱动的事件包含动词“跑”所驱动的事件。显然,要确立两个动词之间存在角色包含的弱共指关系,需要保证动词 A 的至少一个动词角色包含动词 B 的相应角色,且不存在动词 B 的任意一个角色包含动词 A 的相应角色的情况。

2.4.3 角色变更(D)

在描述那些发生情况不确定的事件时,对应动词的角色可能会存在差异,本文将此现象标注为“角色变更”。例如,图 2 例句 32 中,尽管两个动词“获得”均用来叙述“获得冠军”这一事件,却各自关联了不同的主体。若将这二者归类于强共指,则会违反强共指对角色一致性的要求。然而,鉴于这两个表达载体实质上指向同一客观事件,只是因动词角色的不同而展现出事件发生情况的差异,因此,本文将这类情形归纳为“角色变更”的弱共指关系。

角色变更与不共指均具有角色不同的共性,对二者进行对比:首先,不共指的动词通常驱动两个不同的事件;而角色变更的动词驱动了同一事件的不同发生情况。一般地,如果两个动词指代同一事件的不同发生情况,那么这两种发生情况不应该全部真实发生,而指代两个不同事件的动词可以在现实中一并发生。通过例句说明:

例句 1:(不共指)“小明获得了金牌。小红获得了银牌。”

例句 2:(角色变更)“小明获得了这次比赛的金牌,虽然本应小红获得。”

例句 1 中,动词“获得”分别指代两个不同的事件,不共指,这两个事件均发生;例句 2 中,动词“获得”指代“获得金牌”这一事件的不同发生情况:其主体角色分别为“小明”和“小红”,两个事件不应该共

同发生,否则会产生矛盾。

3 数据规范与语料库构建

3.1 语料选择

本文采用的语料源于人民日报标注语料库^[27-28]。我们从该语料库中挑选了 1 000 篇篇幅适中、消息文体的新闻文章,并在标注开始前,对该语料库的动词标注进行了初筛,对其中误标、漏标的动词进行了勘误。

3.2 数据规范与语料库构建过程

图 3 展示了数据规范及语料库构建的详细流程。在数据规范制定环节,采取增量式标签体系构建策略,由两位研究者共同参与规范制定。在明确动词驱动事件共指关系的标注任务下,借鉴了先前研究的数据规范,初步设计了一套事件共指标注规范,并对数据实施预标注,在此过程中,不断积累并归纳新发现的特例,对原有规范进行了必要的调整与补充。预标注初期,规范聚焦于“强共指”关系。随着标注体系的逐步成熟,我们注意到语料中存在一些虽不构成典型共指但依然存在关联性的实例。对此,将“弱共指”关系纳入分析范畴,并依据动词词义与语义角色特点进一步细分,最终细分为三种弱共指类型。

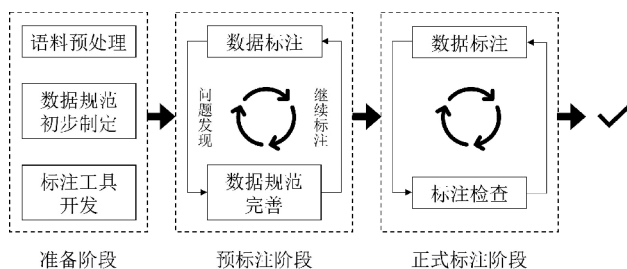


图 3 规范与语料库构建流程

数据标注工作由本规范的主要制定者单独完成,该标注者对标注规范具有深入的理解和熟练的掌握。标注任务的复杂度为 $O(N^2)$, 其中, N 代表每个篇章中动词的数量,正式标注历时 60 天。为了确保数据标注的准确性和可靠性,从全集中随机抽样了 40% 的篇章进行了标注检查,对篇章进行两次标注,标注结果完全一致的篇章所占比率为 94.75%。

正式标注阶段,本文充分利用强共指关系的等价特性及词义包含、角色包含弱共指关系的偏序性

质。具体地,构建强共指链时,本文将具有强共指关系的动词归并至同一集合中;对于词义包含与角色包含的弱共指标注,则充分利用偏序关系的传递性:一旦确定了“动词 1-包含-动词 2”及“动词 2-包含-动词 3”的标注,便自然推导出“动词 1-包含-动词 3”,无须重复标注。此外,在同一强共指链内部,任两动词与链外其他动词之间的关系保持一致。例如,假定动词 1、2、3 构成了一个强共指链,并已知“动词 1-弱共指-动词 4”,根据此逻辑,可以推导出“动词 2-弱共指-动词 4”“动词 3-弱共指-动词 4”。类似地,若动词 1、2、3 位于第一条强共指链,而动词 4、5 位于第二条强共指链,且已知“动词 1-弱共指-动词 4”,则可推广得出两条链上的所有动词两两之间均存在弱共指关系。该标注策略既保证了各类关系的逻辑合理性,又有效提升了标注效率。

3.3 语料库统计

语料库数据规模为:1 000 篇文档,共计 38 万余字,其中涉及两万余动词。

强共指关系统计结果为:1 858 条长度大于 1 的强共指链,涉及 4 251 个动词(占全部动词 14.53%)。强共指链平均长度为 2.29,链长分布统计如表 4 所示。统计强共指动词在文档中的平均字符间距,分别以强共指链中相邻共指动词对和全部共指动词对计。相邻强共指动词对的平均字符间距为 129 个字符、中位数为 83 个字符;全部强共指动词对的平均字符间距为 162 个字符、中位数为 108 个字符;累积分布图、概率分布图如图 4 所示。由此可知,在本语料库上,近一半的强共指动词文本间距小于 100 个字符;若将动词附近 400 个字符作为截断窗口来指定候选先行词的范围,即可召回近 90% 的强共指动词。

表 4 强共指链长度情况统计

链长	1	2	3	4	5	>5
数目	24 999	1 480	273	74	20	11
占比/%	93.08	5.51	1.02	0.28	0.07	0.04

按照动词配对进行统计,强共指对、弱共指对的统计结果如表 5 所示。不共指对、强共指对、弱共指对数目比为 575.09 : 3.21 : 1。

图 5 展示存在共指关系的动词词频统计。可见,在中文通用领域新闻文本中,“举行”“增长”“会见”“进行”等动词出现共指的频率更高。

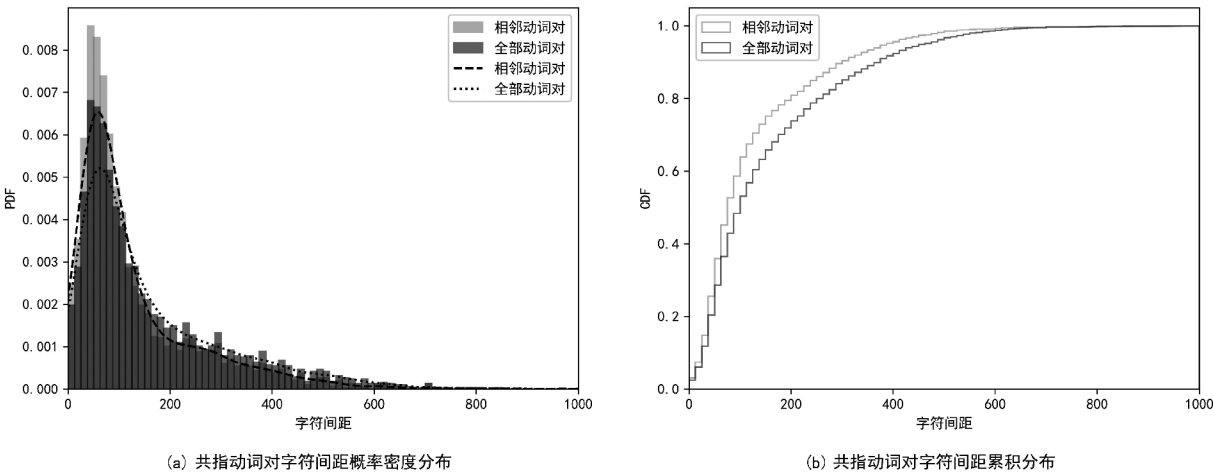


图 4 共指动词对字符间距分布情况

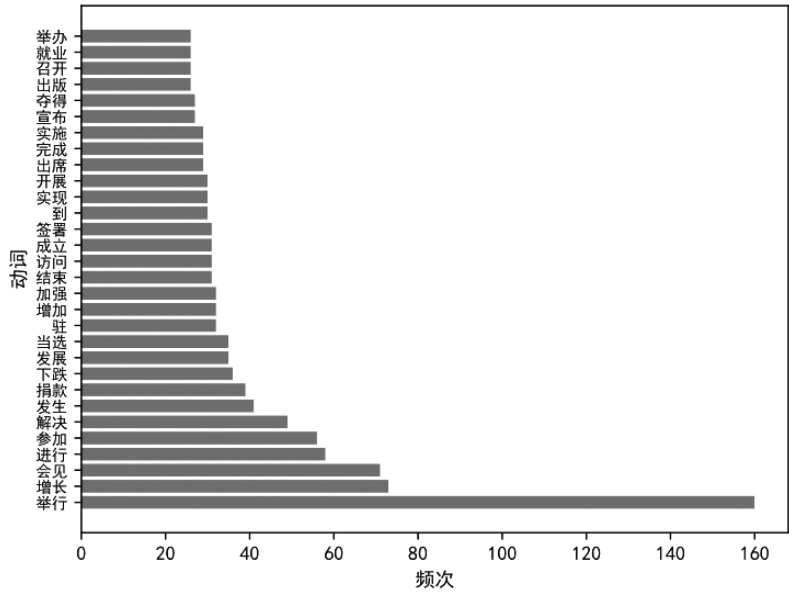


图 5 存在共指关系的动词词频统计(前 30)

表 5 动词对共指关系类型数量统计

共指情况	强共指对	弱共指对		
		词义包含	角色包含	角色变更
数量	3 190	16	959	19
占总共指关系对比/%	76.243	0.382	22.921	0.454
占总动词对比/%	0.554	0.003	0.167	0.003

4 实验

4.1 实验设置

流水线的共指消解方法^[6-7]涉及提及检测和共指关系预测两个关键子任务。提及检测阶段,需从文段中提取所有提及;共指关系预测阶段,则需对

所提取的提及进行共指关系的预测。

4.1.1 提及检测

本文从动词出发探究动词所驱动事件之间的共指关系,因此,动词即为提及检测阶段的提取对象。该阶段,本文采取成分句法分析算法^[8]分析句子成分,构建成分树,并从中提取动词。具体地,提取 CTB 树库标准中标签为 VV 和 VE 的动词,涵盖中文普通动词及存现动词。

4.1.2 共指关系预测

共指关系预测阶段,本文基于大语言模型展开实验,分别采用提示工程、指令微调、知识蒸馏等方式探究了当前通用大语言模型在该任务上的性能表现。此外,从预测性能和推理效率的综合角度出发,进行了大语言模型与小语言模型的对比分析,实验设置如图 6 所示。

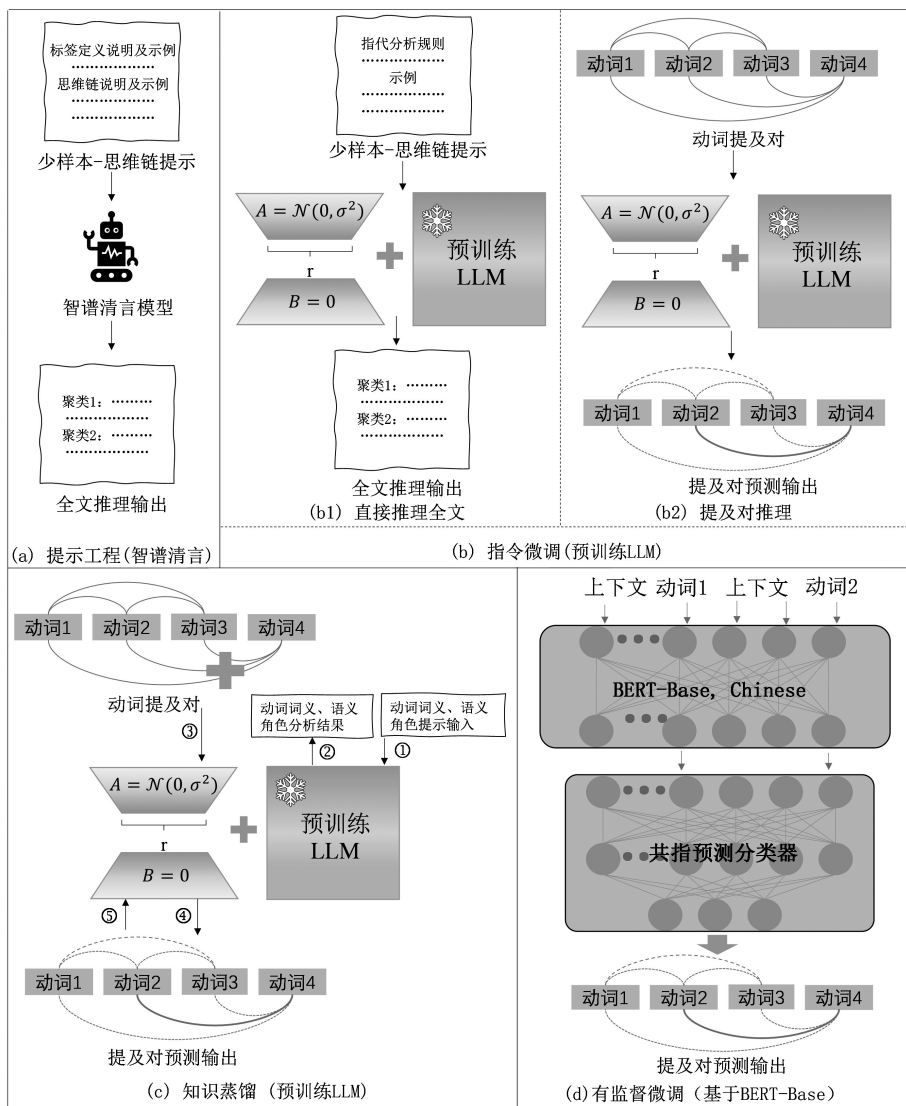


图6 共指关系预测实验设置

提示工程 如图6(a)所示。本文在智谱清言^[12]模型上采用提示工程^[9]方法,设计了少样本-思维链^[29]提示,采用直接推理全文的实验设置,评测了该模型对本语料库任务的推理表现。提示文本如图7所示,具体包含①标签定义及示例、②思维链说明及示例两部分内容。

指令微调 如图6(b)所示。ChatGLM3-6B^[13]和QWen-7B-Chat^[14]均为通用领域开源大模型,本文在两个模型上采用低秩适配器微调(LoRA)^[30],分别进行直接推理全文和提及对推理的实验设置,指令微调并评测了该模型对本任务的推理表现。提示文本如图8所示,具体包含①任务描述、②共指关系分析示例两部分内容。

知识蒸馏 如图6(c)所示。鉴于动词词义和角色与动词所指事件共指关系密切相关,本文在提

及对推理的设置下,进行了基于知识蒸馏^[11]的指令微调实验。首先设计少样本提示,令未在该任务上进行过指令微调的模型对动词词义与语义角色进行分析,获得分析结果,然后将该分析结果作为特征,与指令微调中的提示词一并送入模型,用于协助大模型完成对动词共指关系的分析。该实验采用提及对推理设置完成,其中包含以下五个步骤:

- ① 采用少样本提示,指导基座大模型生成对动词在文本中具体词义及其语义角色的分析;
- ② 基座大模型生成词义、语义角色分析结果;
- ③ 将共指分析任务指令及模型生成特征文本拼接,输入大模型;
- ④ 大模型结合任务描述与所生成的特征,分析文本中动词的指代关系,输出分类结果;
- ⑤ 分类结果用于反向传播,更新适配器参数。

标签定义及示例	动词指代关系共3类：共指关系、相关关系、不具指代关系。定义如下： 【共指关系】：两个词义相同的动词指代同一个事件，动词的主体、客体、时间、地点、命题语义角色分别一致而不相冲突。举例说明： 例1：小明昨天[跑步] (#1) 了，[跑] (#2) 了五公里。 [跑步] (#1)、[跑] (#2) 词义相同，主体、客体、时间、地点语义角色均一致，且均指代“小明昨天跑步”这一事件，所以[跑步] (#1)、[跑] (#2) 为共指关系 【相关关系】：两个动词在字面词义上存在包含关系，或相应角色存在“集合-元素”或“集合-子集”的包含关系或存在变更关系。举例说明： 例1：我国去年[进出口] (#1) 总额庞大。我国去年[进口] (#2) 大幅增长。[进出口] (#1) 中，字“进”“口”与[进口] (#2) 词义相同，[进出口] (#1) 的词义包含[进口] (#2) 的词义，[进出口] (#1) 所指事件包含[进口] (#2) 所指事件，所以[进出口] (#1)、[进口] (#2) 为动词字面意义包含的相关关系 【不具指代关系】：动词的词义不同；或者词义相同，但不满足动词共指、角色包含、动词字面词义包含关系、语义角色变更关系，举例说明： 例1：我[喜欢] (#1) 书法。我昨天[练习] (#2) 了书法。[喜欢] (#1)、[练习] (#2) 词义不同，所以[喜欢] (#1)、[练习] (#2) 不具指代关系
	分析文本中动词的指代关系，其中动词用 [] 标出，其后 () 为该动词在文本中的序号，分为2步： 1. 将文本中相同词义的动词聚成一类；2. 指代关系分析：在每个聚类内部，比较动词的主体、客体、时间、地点、命题等语义角色以及所指代的事件。示例如下： 示例1： 科学家近日[发现] (#1) 远古节肢动物安吉古蜥。此次新[发现] (#2) 的安吉古蜥，形态比以往[发现] (#3) 的多种板足蜥类动物更加圆润。 分析： 1. 将文本中相同词义的动词聚成一类，得到聚类： 聚类1：[发现] (#1)、[发现] (#2)、[发现] (#3) 2. 指代关系分析 聚类1：[发现] (#1)、[发现] (#2) 词义相同，主体、客体、时间、地点语义角色均一致，且均指代“科学家近日发现安吉古蜥”这一事件，所以[发现] (#1)、[发现] (#2) 为共指关系；[发现] (#1)、[发现] (#3) 词义相同，但[发现] (#1) 的客体语义角色“安吉古蜥”和[发现] (#3) 的客体语义角色“多种板足蜥类动物”不一致，动词指代两个不同事件，所以[发现] (#1)、[发现] (#3) 不具指代关系
	思维链说明及示例 请参照定义与示例，对以下文本完成动词指代关系分析： 【文本】：.....

图7 提示工程少样本-思维链提示

基于以上实验设置,得到4组对比实验:不添加特征、添加词义特征、添加语义角色特征、添加词义特征和语义角色特征。

有监督微调 如图6(d)所示。BERT^[15]是一种编码器架构的预训练语言模型,已被广泛应用于自然语言理解任务。本文使用BERT-Base预训练模型提取特征,以此为基础构建了一个用于动词提及对共指关系预测的分类模型,对其进行有监督微调,同以上几类大模型在推理表现和效率上进行综合对比。

在以上几组实验中,由于采取了提及对推理的设置,不共指动词对占比极大,本文采用两种方式解决此问题:① 欠采样:本文从不共指动词对中随机抽取一部分样本,并通过动词义原标签匹配^[20]、词向量^[31]相似度计算等规则随机剔除训练集中词义差异过大的动词对,在最终得到的训练集中,不共指对、强共指对、弱共指对数目比为36:3:1;② 损失

直接推理全文	任务描述 请你按照以下规则,完成文本分析任务。 【规则】:动词指代关系有3类:A.共指;B.非共指,但是相关;C.无关。它们的定义如下: A.共指:动词的词义相同、主体、客体、时间、空间等语义角色均相互一致,动词指代同一事件; B.非共指,但是相关:动词的词义相同,主体、客体、时间、空间等语义角色不同,但是存在包含、变更等关系,动词指代的事件存在包含、变更等相关关系; C.无关:动词的词义不同,或主体、客体、时间、空间等语义角色存在冲突,不指代同一事件且不具有相关关系。 【任务】:给你一段文本,其中所有的动词均以“[动词] (#动词编号)”格式标出,请你分析这些动词之间的指代关系。 【示例】: 示例1:小明昨天[跑步] (#0) 了,他[跑] (#1) 了五公里,他还去[游泳] (#2) 了。 分析1:动词集合[[跑步] (#0), [跑] (#1)]的词义相同,主体、客体、时间、空间等语义角色均互相一致,指代同一事件,因此,A.共指;其他动词间均为C.无关
	推理 请你对以下文本进行分析: 【文本】:.....
	提及对推理 任务描述 请你按照以下规则,完成对文本中所标注的2个动词间指代关系的分析 【规则】:动词指代关系分为3类:A.共指;B.非共指,但是相关;C.无关。它们的定义如下: A.共指:动词在文本中的含义相同、主体、客体、时间、空间等语义角色均相互一致,动词指代同一事件; B.非共指,但是相关:动词在文本中的含义相同,主体、客体、时间、空间等语义角色不同,但是存在包含、变更等关系,动词指代的事件存在包含、变更等相关关系; C.无关:动词的词义不同,或主体、客体、时间、空间等语义角色存在冲突,不指代同一事件且不具有相关关系。 【任务】:给你一段【文本】(其中的2个动词用“[动词] (#编号)”标注),请你基于【文本】来分析这两个动词的指代关系: 【文本】: 注意:请你基于【文本】,仅分析【文本】所标注的2个动词的指代关系,仅输出这2个动词指代关系的选项

图8 指令微调提示

权重调整:依据各类标签在样本中的实际占比,对交叉熵损失函数进行权重调整。

评测指标 在本文设置下,动词所驱动事件对的关系有三类,分别为强共指、弱共指与不共指。本文将一个篇章内部全部动词分别进行配对,每个动词对作为一个样本,评测模型对这三个标签的预测精确率(Precision)、召回率(Recall)、 F_1 值、宏平均 F_1 值(Macro- F_1)。本文重点关注模型对强共指与弱共指标签的预测结果。对模型推理效率的评估中,以平均每篇推理时长为评价指标,并对模型实施五次独立的推理实验,基于观测数据计算平均值,以此作为最终的推理效率度量。

4.2 实验结果

4.2.1 提及检测

提及检测阶段结果如表6所示。

表6 提及检测实验结果 (单位: %)

精确率	召回率	F_1
97.5	96.3	96.9

4.2.2 共指关系预测

以智谱清言为基座的提示工程方法、以 Chat-

GLM3-6B 和 QWen-7B-Chat 为基座的指令微调方法、知识蒸馏方法、小模型的有监督微调方法的实验结果如表 7 所示。

表 7 模型推理表现 (单位: %)

模型方法 推理设置	小模型 有监督微调 提及对	智谱清言 提示工程 全文	ChatGLM3-6B			QWen-7B-Chat		
			指令微调		知识蒸馏	指令微调		知识蒸馏
			全文	提及对	提及对	全文	提及对	提及对
Macro- F_1	55.0	50.0	52.0	59.1	65.3	60.0	64.4	71.7
强共指- F_1	52.3	39.4	49.8	64.5	66.7	62.3	68.9	71.8
弱共指- F_1	13.3	11.1	6.4	13.0	29.4	17.9	24.7	43.4

智谱清言模型的预测表现最差,尽管本文设置了较为详细的少样本-思维链提示,但该模型依旧较难完成该分类任务,这证实了当前通用大语言模型被直接用于构建该语料库的困难。

在 ChatGLM3-6B 和 QWen-7B-Chat 模型上所进行的直接推理全文和提及对推理两种实验设置中,直接推理全文方法计算效率更高,但是分类准确率低于提及对推理的实验设置,在两个模型上,平均 F_1 值分别相差 7.1% 和 4.4%。以上指标均高于智谱清言模型相应指标。由此可见,通过引入本文所构建的数据,即使在参数规模仅 6B 或

7B 的模型上,依然能有效提升对该任务的推理能力。

本文对该任务所涉及的动词词义、语义角色特征进行知识蒸馏,结果显示,引入以上特征均有助于提升模型的推理表现,通过知识蒸馏,ChatGLM3-6B 与 QWen-7B-Chat 模型的平均 F_1 值分别提升 6.2% 和 7.3%。表 8 展示分别引入词义、语义角色特征对模型表现的影响:单独引入任意一类特征与引入全部特征均有助于提升模型表现。其中,引入特征对弱共指预测的提升相比强共指更为显著,分别引入两类特征对模型影响近似。

表 8 大模型引入各类特征的知识蒸馏实验结果 (单位: %)

模型	特征设置	Macro- F_1	强共指- F_1	弱共指- F_1
ChatGLM3-6B	原始	59.1	64.5	13.0
	+ 词义特征	62.3 +3.2	66.0 +1.5	21.1 +8.1
	+ 角色特征	63.6 +4.5	65.8 +1.3	25.4 +12.4
	+ 词义特征 + 角色特征	65.3 +6.2	66.7 +2.2	29.4 +16.4
QWen-7B-Chat	原始	64.4	68.9	24.7
	+ 词义特征	70.6 +6.2	69.7 +0.8	42.3 +17.6
	+ 角色特征	68.7 +4.3	70.9 +2.0	35.6 +10.9
	+ 词义特征 + 角色特征	71.7 +7.3	71.8 +2.9	43.4 +18.7

对比 ChatGLM3-6B 与 QWen-7B-Chat 模型,在全部实验设置下,QWen-7B-Chat 的所有评测指标均高于 ChatGLM3-6B 的对应指标。这可能说明,随着参数量的增加,经过指令微调,模型在该任务上的表现也随之提升。

对比小模型与大模型的实验结果显示,经过微调的小模型的预测表现略微逊色于经指令微调而未经知识蒸馏的大模型。表 9 列出不同模型在不同推理配置下的推理效率比较。其中,小模型具有最高的推理效率。尽管采用大模型针对每对提及项进

表 9 模型推理效率对比

模型	推理设置	用时/(s/篇)
小模型	提及对	0.03
ChatGLM3-6B	提及对	98.65
	全文	2.275
QWen-7B-Chat	提及对	141.25
	全文	2.41

行单独推理能够得到最佳的推理结果,但这一方法

在效率方面最低,其所需时间大致相当于大模型直接推理全文耗时的 50 倍、小模型的 3 000 至 4 000 倍。比较两个大模型, Qwen-7B-Chat 因参数量大于 ChatGLM3-6B,在推理过程中消耗更长的时间。

4.3 错误分析

为进一步探究当前语言模型在动词驱动事件共指消解任务上的能力边界,本文对模型在该任务上的推理结果进行了错误分析。

在提及检测子任务中,有以下几类错误。

① 错误检测为动词:

a) 词性错误:将其他词性的词元错误地预测为动词;

b) 词元错误:所预测动词未出现在评测语料的分词中。

② 动词检测遗漏:

a) 词性错误:将词元错误预测为其他词性;

b) 词元错误:不在所预测的分词中,即动词不对应成分树叶子节点中的任意一词。

错误分析统计如图 9 所示。错误预测为动词主要是词元错误所致;而动词检测遗漏主要是词性错误所致。在词性错误问题中,占大比例的是对名词(n, NN)和介词(p, P)的错误预测。

在共指关系预测子任务中,有以下几类错误。

① 误分类为共指:

a) 词义类型错误:动词在文本内的具体含义不一致,动词本应指代两个不同的事件,但模型预测为

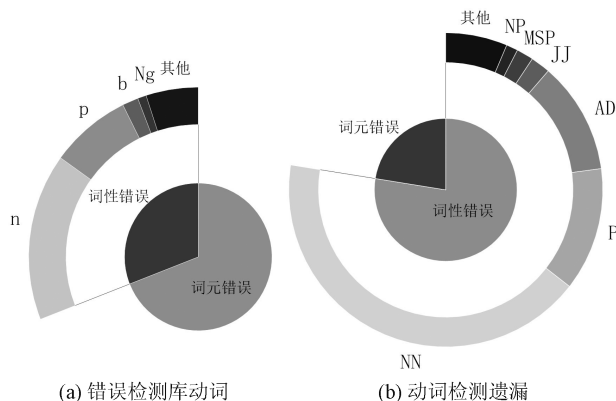


图 9 提及检测错误情况统计

强共指/弱共指;

b) 语义角色矛盾:动词的语义角色存在冲突,不满足强共指条件,但被模型预测为强共指;或不存在包含等关系,不满足弱共指条件,但被模型预测为弱共指;

② 误分类为不共指:

a) 动词字符序列不相同:这可能导致模型将动词判断为含义不同,因此预测为不共指;

b) 动词角色辨识困难:动词所在句子的句式不同、动词的语义角色出现位置距动词较远或被省略而难以分析,可能导致模型在分析二者语义角色的关系时产生错误,因此预测为不共指。

错误分析统计如表 10 所示。结果表明,对于全部实验设置,动词语义角色的辨识错误为动词共指关系判别错误的主要因素,其占比均高于对动词词义的判别错误。

表 10 共指关系预测错误情况统计

错误预测	模型方法 推理设置	小模型 微调 提及对	智谱清言 提示工程 全文	ChatGLM3-6B			QWen-7B-Chat		
				指令微调		知识蒸馏	指令微调		知识蒸馏
				全文	提及对	提及对	全文	提及对	提及对
强共指	词义类型不同	25.19	61.54	29.37	28.68	29.63	10.10	29.41	20.83
	语义角色矛盾	98.78	94.08	100.00	99.22	97.04	98.99	94.96	95.83
弱共指	词义类型不同	6.29	69.23	12.50	8.33	7.41	8.69	12.82	0.00
	语义角色矛盾	100.00	100.00	100.00	100.00	100.00	100.00	97.44	100.00
不共指	字符序列不同	36.33	32.37	39.02	42.37	33.69	37.24	42.53	32.58
	角色辨识困难	99.38	87.92	89.63	94.06	93.48	88.96	93.10	95.51

5 总结

本文提出从事件表达载体出发研究开放域事件

共指关系的研究思路,并围绕动词驱动的事件的共指关系进行了研究。在分析大规模中文语料的基础上,制定了动词驱动事件的共指关系数据规范,其中涵盖强共指与弱共指,并依此人工标注构建了中文

语料库。本文采用流水线的事件共指消解方法,在提及检测阶段,用成分句法分析算法抽取文本内的动词;在共指预测阶段,分别采用提示工程、指令微调、知识蒸馏方法在中文大语言模型上对该任务展开实验,并设置了小模型的有监督微调实验。在几类大模型中,经本语料库微调的模型在当前任务上的表现优于采用提示工程方法模型的表现,采用知识蒸馏、引入相关特征后,模型性能得到进一步提升。小模型方法的表现略差于指令微调后的大模型,但具有最高的计算效率。本文分析验证了当前语言模型对动词共指关系的认知和推理所存在的不足,其中,模型对动词角色关系的判别错误是当前模型在该任务上推理能力的主要限制因素。本文也呼吁广大研究者关注该类问题。

未来,在事件共指研究方面,我们将从名词短语、动词短语等其他表达载体出发,对开放域事件共指关系进行深入探究;在规范的推广与应用方面,将尝试把数据规范应用于当前已有的事件抽取数据集中,从而发挥其在事件抽取等任务上的潜在作用。

参考文献

- [1] AHN D. The stages of event extraction[C]//Proceedings of the Workshop on Annotating and Reasoning about Time and Events, 2006: 1-8.
- [2] WALKER C, STRASSEL S, MEDERO J. ACE 2005 multilingual training corpus [DS]. Philadelphia: Linguistic Data Consortium, 2006.
- [3] CYBULSKA A, VOSSSEN P. Using a sledgehammer to crack a nut?: Lexical diversity and event coreference resolution[C]//Proceedings of the 9th International Conference on Language Resources and Evaluation, 2014: 4545-4552.
- [4] ELLIS J, GETMAN J, STRASSELS. Overview of linguistic resource for the TAC KBP evaluations: Planning, execution, and results[C]//Proceedings of TAC KBP Workshop, 2014: 17-18.
- [5] WANG X, CHEN Y, DING N. MAVEN-ERE: A unified large-scale dataset for event coreference, temporal, causal, and subevent relation extraction[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2022: 926-941.
- [6] RAGHUNATHAN K, LEE H, RANGARAJAN S. A multi-pass sieve for coreference resolution[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2010: 492-501.
- [7] 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, 2016: 643-653.
- [8] XIN X, LI J, TAN Z. N-ary constituent tree parsing with recursive semi-Markov model[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021: 2631-2642.
- [9] BROWN T B, MANN B, RYDER N. Language models are few-shot learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020: 1877-1901.
- [10] WEI J, BOSMA M, ZHAO V Y. Finetuned language models are zero-shot learners[C]//Proceedings of the 10th International Conference on Learning Representations, 2022.
- [11] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J/OL]. arXiv preprint arXiv:1503.02531, 2015.
- [12] 北京智谱华章科技有限公司. 智谱清言[CP/OL]. [2023-12-13]. <https://chatglm.cn/>
- [13] DU Z, QIAN Y, LIU X. GLM: General language model pretraining with autoregressive blank infilling [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022: 320-335.
- [14] BAI J, BAI S, CHU Y. QWen technical report[J/OL]. arXiv preprint arXiv:2309.16609, 2023.
- [15] DEVLIN J, CHANG M W, LEE K. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, 2019: 4171-4186.
- [16] 黄宏斌, 孙皎, 韦晖. 基于开源军事新闻的领域事件数据集[J]. 中国科学数据(中英文网络版), 2023, 8(1): 339-348.
- [17] Tianchi. CCKS2022 金融领域 FEW-SHOT 事件抽取数据集[DS/OL][2024-03-01]. <https://tianchi.aliyun.com/dataset/136800>
- [18] PRADHAN S, RAMSHAW L, WEISCHEDEL R. Unrestricted coreference: Identifying entities and events in OntoNotes[C]//Proceedings of International Conference on Semantic Computing, 2007: 446-453.
- [19] O' GORMAN T, WRIGHT-BETTNER K, PALMER M. Richer event description: Integrating event coreference with temporal, causal and bridging annotation[C]//Proceedings of the 2nd Workshop on Computing News Storylines, 2016: 47-56.
- [20] DONG Z, DONG Q, HAO C. HowNet and its com-

- putation of meaning[C]//Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, 2010: 53-56.
- [21] 袁毓林. 论元角色的层级关系和语义特征[J]. 世界汉语教学, 2002, 16(3): 11-22.
- [22] 上海大学语义智能实验室. CEC-Corpus[DS/OL]. [2024-03-28]. <https://github.com/shijiebei2009/CEC-Corpus>
- [23] Li X, Li F, PAN L. DUEE: A large-scale dataset for chinese event extraction in real-world scenarios[C]//Proceedings of Natural Language Processing and Chinese Computing, 2020.
- [24] 滕佳月, 李培峰, 朱巧明. 基于全局优化的中文事件同指消解方法[J]. 北京大学学报. 自然科学版, 2016 (01): 97-103.
- [25] 宦敏, 程昊熠, 李培峰. 基于结构化表示的中文事件同指消解方法. 中文信息学报. 2019, 33(12): 19-27.
- [26] 环志刚, 蒋国权, 张玉健. 门控机制融合多种特征的中文事件共指消解[J]. 计算机科学, 2023, 50(3): 291-297.
- [27] YUS, DUAN H, SWEN B. Specification for corpus processing at peking university: Word segmentation, pos tagging and phonetic notation[J]. Journal of Chinese Language and Computing, 2003; 3(2): 121-158.
- [28] 俞士汶, 朱学锋, 段慧明. 大规模现代汉语标注语料库的加工规范[J]. 中文信息学报, 2000, 14(6): 58-64.
- [29] WEI J, WANG X, SCHUURMANS D. Chain-of-thought prompting elicits reasoning in large language models[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 24824-24837.
- [30] HU E J, SHEN Y, WALLIS P. LoRA: Low-rank adaptation of large language models[C]//Proceedings of the 10th International Conference on Learning Representations, 2022.
- [31] MIKOLOV T, CHEN K, CORRADO G. Efficient estimation of word representations in vector space [C]//Proceedings the 1st International Conference on Learning Representations, 2013.



王皓泽(2001—), 硕士研究生, 主要研究领域为基于大语言模型的推理技术。
E-mail: w_haoze@bit.edu.cn



辛欣(1984—), 通信作者, 博士, 副教授, 主要研究领域为基于大语言模型的多模态信息检索与推理技术。
E-mail: xxin@bit.edu.cn