

基于大语言模型的事件常识知识图谱扩展方法^{*}

黄俏娟^{1,2}, 曹存根¹, 王亚³, 王石¹

¹(中国科学院 计算技术研究所 智能信息处理重点实验室, 北京 100190)

²(中国科学院大学, 北京 100049)

³(南阳师范学院 人工智能与软件工程学院, 河南 南阳 473061)

通信作者: 曹存根, E-mail: cgcao@ict.ac.cn



摘要: 常识知识通常不在自然语言中明确表述, 而是隐含在人类的认知中, 为机器提供常识知识一直是人工智能领域的长期目标之一。前期, 课题组成员手工构建了一个高精度的以事件为中心的中文种子常识知识图谱(ECKG), 包含了 26 606 个常识事件三元组, 覆盖了因果、时序、条件等多种常见的事件关系。尽管种子 ECKG 具有一定的价值, 但规模较小, 在实际应用中发挥的作用有限, 且大规模的事件常识知识图谱在现有研究中较为稀缺。为了应对这些挑战, 采用 GPT 系列的大语言模型来扩展种子 ECKG 中的因果、时序、条件和子事件这 4 种事件关系。扩展方法包括 3 个主要的步骤: 首先, 将种子 ECKG 中的事件结合 4 种关系定义设计了特定的事件知识提示(ek-prompt), 并使用 GPT-4-Turbo 模型生成相应的事件三元组。其次, 将种子 ECKG 的三元组与通过 ek-prompt 获取的正确三元组组合, 以构建特定的数据集, 并将 GPT-3.5-Turbo 模型在数据集上进行微调, 以生成更多具体的事件三元组和验证新三元组准确性。最后, 通过分析种子 ECKG 事件的相似性, 并引入事件共享机制, 将相同关系下的相似事件关联的事件互相共享, 以保持相似事件的三元组一致性。经过实验评估, 新获取的三元组具有高质量, 尤其是时序关系的三元组准确率最高, 达到了 98.2%。所提扩展方法最终为种子 ECKG 增补了 2 433 012 个常识事件三元组, 显著扩大了 ECKG 的规模, 可以为人工智能领域的许多应用提供了更为丰富的常识知识资源。

关键词: 事件常识知识图谱; 大语言模型; 微调技术; 事件三元组; 事件共享机制

中图法分类号: TP18

中文引用格式: 黄俏娟, 曹存根, 王亚, 王石. 基于大语言模型的事件常识知识图谱扩展方法. 软件学报, 2025, 36(9): 4153–4186. <http://www.jos.org.cn/1000-9825/7262.htm>

英文引用格式: Huang QJ, Cao CG, Wang Y, Wang S. Method for Expanding Event Commonsense Knowledge Graph Based on Large Language Models. Ruan Jian Xue Bao/Journal of Software, 2025, 36(9): 4153–4186 (in Chinese). <http://www.jos.org.cn/1000-9825/7262.htm>

Method for Expanding Event Commonsense Knowledge Graph Based on Large Language Models

HUANG Qiao-Juan^{1,2}, CAO Cun-Gen¹, WANG Ya³, WANG Shi¹

¹(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

³(School of Artificial Intelligence and Software Engineering, Nanyang Normal University, Nanyang 473061, China)

Abstract: Commonsense knowledge is usually not explicitly expressed in natural languages but is implicitly understood in human cognition. Providing machines with commonsense knowledge has been a longstanding aim in artificial intelligence. Initially, this study manually constructs a high-precision, event-centric commonsense knowledge graph (ECKG) for seed events in Chinese. It contains 26 606

* 基金项目: 国家重点研发计划 (2022YFC3302300); 国家科技重大专项 (7090201050307); 国家 242 信息安全计划 (2023A105)

收稿时间: 2024-04-19; 修改时间: 2024-06-10; 采用时间: 2024-07-19; jos 在线出版时间: 2024-12-31

CNKI 网络首发时间: 2025-01-02

commonsense event triples encompassing causal, temporal, conditional, and other common event relationships. Although the constructed ECKG holds considerable value, its limited scale curtails practical applications. Besides, large-scale event commonsense knowledge graphs are rare in current studies. To overcome these challenges, this paper uses large language models from the GPT series to expand the above-mentioned three event relationships and sub-events of the proposed ECKG. The expansion method involves three primary steps. Firstly, specific prompts for event knowledge (ek-prompts) are designed by combining the events in the ECKG with four relationships, and GPT-4-Turbo is used to generate corresponding event triples. Secondly, the triples of the ECKG are integrated with accurate triples obtained by ek-prompts to create a specialized dataset. Additionally, GPT-3.5-Turbo is fine-tuned on the dataset to generate more specific event triples and validate the accuracy of new triples. Lastly, by analyzing the similarities among events in the ECKG and implementing an event-sharing mechanism, similar events within the same relationship are interconnected, ensuring consistency across similar event triples. Experimental results show that the newly acquired triples are of high quality, particularly those of the temporal relationships, with an accuracy rate of 98.2%. Ultimately, the proposed expansion method appends 2 433 012 commonsense event triples to the original ECKG, significantly expanding its scale and providing more commonsense knowledge for many applications in artificial intelligence.

Key words: event commonsense knowledge graph; large language model (LLM); fine-tuning technique; event triple; event-sharing mechanism

常识知识是大多数人共同认可的关于世界的真实描述,如“苹果是一种水果”“走路时手臂会摆动”以及“人被赞美时通常会感到开心”等。这类知识广泛应用于自动问答^[1]、计算机视觉^[2]和情感分析^[3]等人工智能领域。由于常识知识具有共享性(人们共同认知的信息)、隐含性(通常不会在文本明确表达)和广泛性(遍布各个学科领域),使得获取常识知识成为一项具有挑战性的任务。

目前,许多著名的大型知识图谱主要关注实体相关的知识,比如 YAGO^[4]、DBpedia^[5]和 Wikidata^[6]等。然而,以事件为中心的知识图谱通过将事件与相关实体、时间和空间等因素关联起来,在自然语言处理^[7]、问答系统^[8]和事件预测^[9]等领域起到了至关重要的作用。但是,与事件相关的大规模的知识图谱较为稀缺。例如,ATOMIC^[10]专注于事件间的因果关系,却未涉及其他类型的事件关系;而 ConceptNet^[11]虽然提供了丰富的实体相关常识知识,但事件相关的常识知识较为有限。

在早期阶段,本课题组投入大量的时间及人力资源,从各种资源中筛选和整理出一定量的事件知识。经过多次严格的修正和校对,最终构建成一个包含 26 606 个事件三元组的高质量中文种子事件常识知识图谱(event commonsense knowledge graph, ECKG)^[12-15]。该种子图谱涵盖了因果、时序和条件等多种常见的事件关系,其中最为基本和主要的关系类型包括因果、时序、条件以及子事件关系。尽管 ECKG 在精度和实用性上具有显著优势,但规模及知识覆盖的局限性,影响了其在实际应用中的广泛性。

此外,我们注意到现有获取常识知识的方法包括人工^[10,13,16-18]、自动化^[1,19-22]和半自动化^[11,23-25]过程,这些方法通常需要复杂的数据预处理和大量人力。同时,在常识知识图谱补全方面,尽管已有研究通过规则、嵌入及神经网络等技术来提高补全的准确性并扩展知识的深度^[26-33],但依然面临数据稀缺和质量不高的问题。此外,LLM (large language model) 的出现为知识图谱补全带来新的机遇。已有研究表明 LLM 在补全与实体相关的知识方面显示出优势^[34-40],然而在事件常识知识图谱领域的研究相对较少。

针对上述的问题,本文提出了对种子 ECKG 进行扩展。主要是增强图谱中的因果、时序、条件和子事件这 4 种事件关系的知识库,通过增加每个事件(E)在关系 R 中的头三元组(<?, R, E>)和尾三元组(<E, R, ?>),以扩展整个 ECKG 的规模,同时保证高精度。

近年来,OpenAI 发布了 GPT-3.5 和 GPT-4 等一系列先进的大语言模型^[41]。这些模型基于庞大的预训练数据集,具有丰富的内部知识,能够广泛应用于文本生成、问答系统和知识图谱补全等多种任务。研究表明,GPT-3.5 和 GPT-4 具有强大的学习能力和快速的反应速度,可以通过简单的提示或微调来有效地生成知识。基于这些优势,本文采用 GPT 系列 LLM 来扩展种子 ECKG。具体地,本文将使用 GPT-3.5-Turbo 和 GPT-4-Turbo 模型的强大知识库,全面增强种子 ECKG 的 4 种事件关系的三元组。本文的主要贡献如下。

(1) 针对每种事件关系设计特定的事件知识提示(event knowledge prompt, ek-prompt),并利用 GPT-4-Turbo 模型生成新的事件三元组。

(2) 结合种子 ECKG 和通过 ek-prompt 生成的事件三元组,为每种事件关系构建特定的数据集,并用于微调

GPT-3.5-Turbo 模型, 以进一步丰富和扩展事件三元组。

(3) 为了确保大规模事件三元组的高精度, 对 GPT-3.5-Turbo 模型进行特别的微调, 专门用于评估新三元组的正确性。

(4) 引入 OpenAI 的嵌入模型及多维度语义的事件相似度计算和共享机制。这一机制独特地结合了语义角色标注、含句子成分的语义角色、语义角色标注结构与事件反义多重特征, 获得了精确的种子 ECKG 事件的相似度。在相同关系中, 将相似度高的 ECKG 事件关联的事件进行共享。该机制显著地解决了 LLM 在生成高度相似事件的三元组时出现的不一致问题, 极大地优化了获取新三元组的过程。

实验结果表明, 本文方法在扩展种子 ECKG 任务上表现出显著的精确度和有效性, 为常识事件三元组的获取提供了一种新的思路。扩展后的 ECKG 涵盖丰富的常识知识, 能够为多种人工智能领域的研究和应用提供了可靠的常识知识支持。

本文第 1 节探讨获取常识知识和知识补全的相关研究。第 2 节介绍扩展 ECKG 的基础知识。第 3 节讨论获取常识知识和扩展 ECKG 的具体方法。第 4 节通过实验评估本文方法的效果。第 5 节总结全文及未来工作的展望。

1 相关工作

1.1 常识知识图谱补全

常识知识可以通过人工、自动化和半自动化的方式获取。对于人工方式, ATOMIC^[10]使用众包的方式从事件的注解中提取“if-then”形式的常识知识。Wang 等人^[13]从多种数据源中收集和筛选常识相关的事件三元组, 构建了一个以事件为中心的高质量种子知识图谱。Cyc^[16]是首个大规模的常识知识库, 其内容由工程师手工编辑, 并采用 CycL 语言对这些常识知识进行形式化。WordNet^[17]由词典编辑者以文本的形式输入而构成的常识知识库, 其涵盖英文单词、短语和同义词。OMCS^[18]通过互联网从用户中搜集常识知识, 允许网络用户以预设模板或自由文本方式贡献常识知识。在自动获取常识知识的工作中, Nguyen 等人^[1]提出从网络文本中自动构建大规模常识知识库的方法, 主要利用搜索查询、句法分析和 RoBERTa 模型^[42]提取并丰富知识的语义信息。TransOMCS^[19]将大型语言知识库转换为常识知识库, 并使用排序模型选取高质量的知识。Romero 等人^[20]从网络资源中获取概念属性常识, 并通过搜索日志、问答论坛和统计的方法获取常识并验证正确性。曹亚男^[21]提出了从网络语料库中提取因果常识的方法, 包括识别文本中的因果关系和基于贝叶斯学习因果边界划分两个阶段。Bosselut 等人^[22]基于 ATOMIC 和 ConceptNet 开发了 COMET 系统, 能够自动生成较高精度的自然语言描述的常识知识。半自动获取方法结合人工与自动化技术, 例如 ConceptNet^[11]通过词法句法模式从 OMCS 知识库中提取常识知识三元组, 再结合人工规则扩展知识的覆盖率。WebChild^[23]基于 WordNet 获取概念属性的常识知识, 进而以模式匹配和半监督学习从网络文本中增强这些知识。Learner 系统^[24]采用类比方法提取常识知识, 要求用户提供概念相关的事实, 从而获取更多常识信息。曹聪等人^[25]提出了一个人机交互的方法, 利用机器动态生成常识性的问题, 用户通过回答问题贡献了他们大脑中的常识, 再通过人工筛选和清理进而获取到相关的常识知识。

将获取的常识知识进行补全常识知识图谱的工作中, 研究者们采用多种方法以提高知识图谱的规模 and 知识覆盖率。王亚^[26]通过设计常识性逻辑规则对图谱进行推理, 以发掘事件间的常识关系。此外, 还使用常识知识的上下位关系来生成新事件, 以扩大事件知识图谱的规模。邓鹏^[27]融合语义角色和 AMIE^[43]+关联规则挖掘技术, 通过语义增强和规则过滤剪枝, 提高了事件图谱补全的准确性。Meilicke 等人^[28]提出了一种结合规则和嵌入方法来进行知识图谱补全, 从而有效提升知识图谱补全的性能。针对 ConceptNet 的补全任务, Li 等人^[29]开发了一个神经网络模型, 该模型可以自动修改和生成新的三元组, 并对新三元组的正确性进行评分。Malaviya 等人^[30]设计了一个基于概念结构和语义上下文的常识知识库补全模型, 通过图卷积网络和图致密化技术将预训练语言模型的学习迁移到知识图谱上, 增强知识的上下文表示能力。Niu 等人^[31]构建了一个逻辑和常识指导的嵌入模型。该模型学习事件的时间敏感表示和时间独立表示, 有效捕捉事件的及时性和因果性, 并从常识的视角理解事件间的关系。此外, 他们引入了辅助的常识知识来提高事件可能性评估的准确性, 显著提高了时间知识图谱补全的性能提升。Wang 等

人^[32]介绍了一种归纳学习框架, 专为解决常识知识图谱补全中的归纳学习难题而设计. 该框架能够处理未见过的实体, 通过结合实体嵌入技术和图神经网络的结构信息, 有效提高了对新实体的表示精度. Bosselut 等人^[33]提出了一种针对零样本常识问答的方法. 该方法能够动态生成和推理常识知识图谱, 有效应对叙事文中隐含的情境相关的原因、结果和现状等知识的挑战.

上述方法依赖手工编写、自动化或半自动化从各种数据源筛选和验证常识知识, 依赖复杂的预处理过程, 导致高昂的人工成本, 并且难以保证知识的高质量. 不同于这些方法, 本文结合 LLM、设计特定的 **ek-prompt** 以及微调模型策略来生成大规模且高质量的事件三元组, 显著提高了获取常识知识的效率. 此外, 提出了事件共享机制, 通过在同种关系下将相似事件关联的事件互相共享, 这一机制有效地保证了相似事件三元组的一致性, 防止了知识碎片化. 整体而言, 本文提供了一种更为灵活和有效的常识知识获取方法来补全常识知识图谱.

1.2 LLM 补全知识图谱

近些年, 大语言模型已经广泛应用于多种自然语言处理任务. 例如, 针对知识图谱补全中长尾三元组 and 高质量数据的稀缺性, 张宁豫等人^[34]提出了一种基于知识协同微调的低资源知识图谱补全方法. 通过构建知识图谱补全提示并采用协同微调算法, 有效地使用了知识图谱的结构化知识和语言模型中的隐式事实知识, 适用于链接预测和关系抽取任务. 为了提升 LLM 预测知识图谱中缺失三元组的准确性, Zhang 等人^[35]整合了知识图谱的结构信息到 LLM 中, 通过转换现有 LLM 范式为结构感知设置, 并设计了知识前缀适配器来捕获实体与关系的结构信息, 将信息传送给 LLM 以提高预测精度. Wei 等人^[36]设计了一个名为 KICGPT 的知识图谱补全框架, 该框架通过集成 LLM 和基于三元组的知识图谱补全检索器, 并采用知识提示学习策略, 以指导 LLM 高效地进行知识图谱补全. Luo 等人^[37]专注于时间知识图谱补全任务, 基于 LLM 的生成能力和对时间知识图谱推理的调整, 通过对参数的微调和分析 LLM 内在的变异因素, 提高了整个图谱的补全性能. Yao 等人^[38]通过将知识图谱中的三元组与 LLM 结合, 使用实体和关系的描述作为输入提示, 并利用 LLM 的响应来预测事件. Pan 等人^[39]提出了一种将 LLM 和知识图谱进行统一的前瞻性路线图, 该路线图主要通过知识图谱增强 LLM、LLM 增强知识图谱, 以及 LLM 与知识图谱的协同工作, 通过数据和知识驱动的双向推理增强 LLM 和知识图谱. Li 等人^[40]为了解决知识图谱补全任务中遇到的文本语料的静态和噪音问题, 提出了一种上下文蒸馏方法. 该方法首先指导 LLM 赋予三元组丰富的上下文信息. 再引入重建和情景化辅助机制, 使得较小的知识图谱补全模型能够吸收这些丰富的三元组信息.

与其他使用 LLM 进行知识图谱补全的研究相比, 本文方法显得更为直接. 通过设计特定的 **ek-prompt** 和微调模型, 专门扩展以事件为中心的常识知识图谱. 特别地, 本文引入的事件共享机制解决了由于 LLM 偏差可能引起的信息不一致问题, 从而确保了图谱中相似事件的信息一致性. 这不仅有效扩展事件知识图谱的规模, 还为知识图谱的应用和深化开辟了新的路径.

2 基础知识

本节将介绍在扩展 ECKG 所需的准备工作, 包括主要的事件关系定义和如何辨别事件是否存在这些关系. 接下来, 将阐述相关的概念和基础知识.

在种子 ECKG 中, 本文定义了 4 种事件关系, 具体如下.

- (1) 因果关系: 一个事件的出现可能导致另一个事件的发生, 用三元组表示为 $\langle E_1, \text{导致}, E_2 \rangle$. 例如, 减肥导致变瘦.
- (2) 时序关系: 一个事件在时间上先于另一个事件发生, 三元组表示为 $\langle E_1, \text{先于}, E_2 \rangle$. 例如, 订婚先于结婚.
- (3) 条件关系: 一个事件的发生基于一定的前提条件基础上, 三元组表示为 $\langle E_1, \text{是条件}, E_2 \rangle$. 例如, 参军是参加战斗的前提条件.
- (4) 子事件关系: 一个过程事件发生时涉及一系列相关的子事件出现, 三元组表示为 $\langle E_1, \text{有子事件}, E_2 \rangle$. 例如, “找工作”这一事件包含“投简历、准备面试、参加面试、等待面试结果”等一系列的子事件.

从常识的角度理解, 事件之间的发生都存在因果、时间上的先后顺序, 以及基于特定的前提条件. 因此, 在扩

展种子 ECKG 时, 我们认为每一个事件都应该考虑这 3 种关系. 然而, 并非所有的事件都包含子事件. 当一个事件有子事件时, 该事件应该是包含一系列相关的活动或步骤的复杂过程事件. 例如, 事件“准备晚餐”可能涉及“购买食材、清洗食材、煮食材”等多个步骤.

为了有效地扩展种子 ECKG, 则需要预先识别哪些事件含有子事件. 本文根据常识的事件语义类别^[44,45]和事件本身的含义总结 11 种不包含子事件关系的事件特征, 如表 1 所示, 下划线标识的是关键特征的字词. 基于表 1 的特征, 本文标注一定的数据集, 并采用了支持向量机^[46]对 ECKG 中的事件进行分类, 进而判断出事件是否具有子事件关系. 对于筛选出包含有子事件关系的事件, 再使用第 3 节的方法扩展子事件关系的三元组.

表 1 不含子事件关系的事件特征

描述	实例
包含起止含义或表示最终结果的事件	开始工作、吃完食物、课程结束、完成任务、找到礼物、获得证书
“太+形容词”修饰的事件	吃太多苹果、喝酒太多、吃得太快、进展太慢、喝水太少
描述主体现状的事件	家庭贫穷、已婚、婚姻牢固、婚姻美满、身体健康、身体不健康
愿意类的事件	想吃香蕉、不想站立、愿意帮忙、不愿意锻炼
态度语义类的事件	上课认真、粗心、与他人亲密、司机热情、他国对祖国友好
主体自身在时间上进展的消散类事件	春天来临、怒气消失、社会进步、他人退步
特征语义类的事件	自己聪明、她漂亮、相貌帅气、厨艺好、名声不好、售价高
情感语义类的事件	朋友悲伤、厌恶他人、忌妒朋友、讨厌食物、喜欢锻炼身体、他人满意
像是类的事件	是记者、不是观众、像老师、不像明星
领属语义类的事件	有电话、拥有钱财、他人患有重病、没有房子、某地回归国家
感知语义类的事件	看到电影广告、理解父母、饥饿、感到冷、发现真相、知道重要消息

3 LLM 扩展 ECKG

本节将详细介绍使用 GPT 系列的 LLM 获取因果、时序、条件和子事件关系的事件三元组以扩展 ECKG 的方法. 首先, 在第 3.1 节中, 本文将探讨如何设计特定的 ek-prompt, 并基于 GPT-4-Turbo 模型扩展 ECKG 中的 4 种事件关系. 接着, 第 3.2 节描述了构建特定结构的数据集, 经过微调 GPT-3.5-Turbo 模型以生成和验证新三元组. 最后, 在第 3.3 节中, 引入一个事件共享机制. 通过计算 ECKG 中事件间的语义相似度, 将高度相似的事件在相同关系下关联的事件进行互相共享. 后文图 1 展示了扩展 ECKG 的整体框架, 其中 ft-gen 表示用于生成新三元组的微调模型, ft-ver 是用于验证新三元组正确性的微调模型.

3.1 设计 ek-prompt

由于 LLM 依赖于自然语言的输入才能展示更好的性能, 因此本文设计了 ek-prompt. 通过 ek-prompt 精确控制的提示来指导 LLM 学习 ECKG 的相关知识, 以生成特定类型的事件. 为 ECKG 中相同关系下的每个事件设置两种特定的 ek-prompt, 分别用于生成事件的头三元组和尾三元组. 例如, 对于事件 E, 在相同关系 R 中, 一种 ek-prompt 为了生成头三元组<?, R, E>, 另一种则是为了生成尾三元组<E, R, ?>.

GPT-4-Turbo 模型是 OpenAI 发布的先进大规模语言模型, 具有更新的数据集, 并且支持 128k 的上下文窗口^[47]. 因此, 本文基于 ek-prompt 通过 API 调用 GPT-4-Turbo 模型来扩展 ECKG 中的因果、时序、条件和子事件关系的三元组. 为了确保生成高质量的三元组, 经过多轮测试和优化, 最终确定了效果最佳的 ek-prompt. 每种事件关系的 ek-prompt 如表 2 所示, 每个 ek-prompt 主要包括事件关系的定义描述、举例说明 (可选)、提出问题和输出要求这 4 个部分. 以下是对每一部分的详细介绍.

(1) 事件关系的定义描述

在 ek-prompt 的第 1 部分, 通过对特定事件关系的明确定义, 为 GPT-4-Turbo 模型提供了清晰的背景信息框架, 以引导模型理解和生成相关的事件. 例如, 对于因果关系, 本文定义为“一个事件的发生可能导致另一个事件的产生”. 这种清晰的定义有助于模型学习事件生成的基础逻辑, 确保生成的三元组与预期的事件关系类型一致.

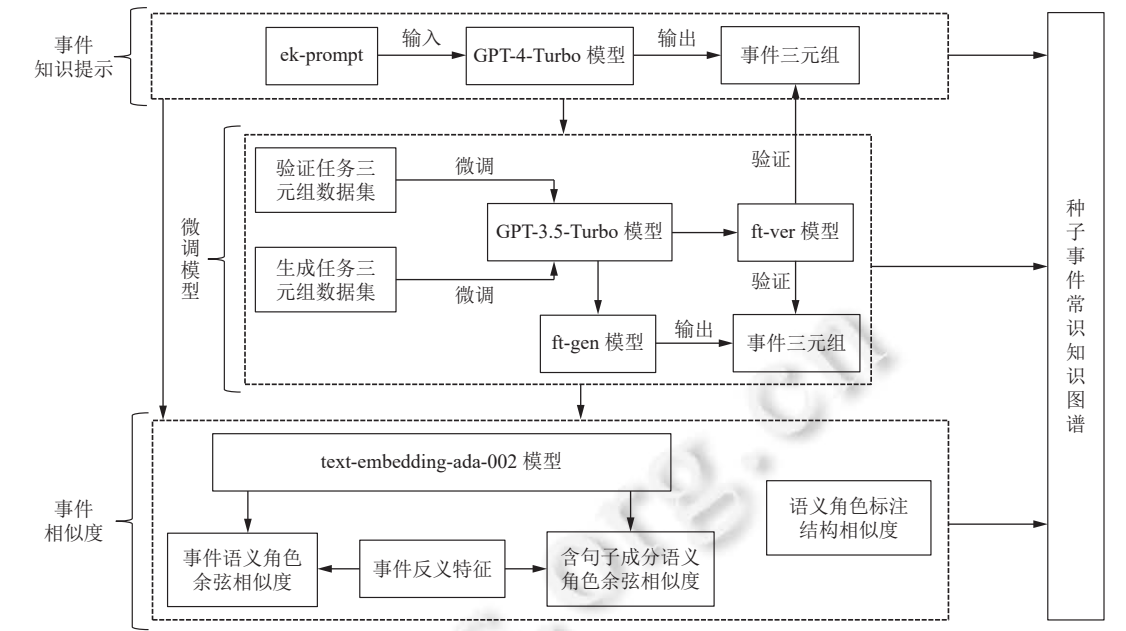


图 1 LLM 扩展 ECKG 的架构

表 2 ECKG 中不同事件关系的 ek-prompt

关系	三元组	ek-prompt			
		定义描述	举例说明	提出问题	输出要求
因果	<?, 导致, E>	因果关系是指一个事件的发生可能导致另一个事件的产生. 其中, 前一个事件是原因事件, 后一个事件是结果事件.	—	哪些原因会导致E?	
	<E, 导致, ?>	因果关系是指一个事件的发生可能导致另一个事件的产生.	—	E可能产生哪些后果?	
时序	<?, 先于, E>	时序关系是指事件按照一定的时间顺序发生.	接电话之前, 会按下接听键.	E之前, 会发生什么事情?	(1) 列出语义明确的事件.
	<E, 先于, ?>	时序关系是指事件按照一定的时间顺序发生.	撕开薯片袋子之后, 再取出薯片.	E之后, 再发生什么事情?	(2) 每个事件必须包含谓语.
条件	<?, 是条件, E>	有条件指一个事件发生所需要的条件.	接电话发生所需要的条件之一为按下接听键.	E事件发生所需要的条件有哪些?	(3) 事件的宾语应保持一致.
	<E, 是条件, ?>	是条件指前一个事件作为后一个事件发生的前提条件.	申请退学是学校批准退学的条件.	E是哪些事件的前提条件?	(4) 输出的三元组以空格分隔.
子事件	<?, 有子事件, E>	主事件是指包含多个子步骤的总步骤.	调整手指的位置是按压和弦的子步骤.	E是哪些总过程的子步骤?	
	<E, 有子事件, ?>	有子事件是指一个总体事件发生时, 会涉及的一系列互相关联的事件. 这些子事件共同构成了总体事件的发展过程.	—	总体事件E有哪些子事件?	

(2) 举例说明

根据 GPT-4-Turbo 模型输出事件的质量, 本文选择性地在 ek-prompt 中加入具体的实例. 该部分主要是为了给模型提供一个直观的示例, 帮助模型更好地理解特定事件关系, 并在这个基础上生成新三元组. 比如, 在扩展时序关系的尾事件时, 添加“撕开薯片袋子之后, 再取出薯片.”作为实例. 然而, 这部分不是必需的, 只有当模型的初步

输出需要进一步指导时才会被加入。比如,在扩充 ECKG 的因果关系时,由于模型生成该关系的三元组具有高的准确率,便省略了具体实例。

(3) 问题提出

问题的提出是 **ek-prompt** 的核心部分,它直接引导模型针对特定事件关系来生成新三元组。本文通过精心设计的问题,如生成时序关系的尾事件时提问“E 之后,再发生什么事情?”,以此激发模型按照特定的事件关系进行学习和生成。提出的问题不仅紧扣事件关系的定义,还促使模型探索并预测事件之间的逻辑联系。

(4) 输出要求

最后,设定具体的输出标准,指导模型以结构化形式展示生成的三元组。这些标准包括输出事件的一致性、包含必要信息和三元组的展示格式等。例如,要求模型输出的事件必须是语义清晰、包含谓语、宾语的一致性,以及规定以空格分隔的形式展示三元组,以确保生成的三元组不仅准确,也便于数据后处理。

通过以上 4 个部分的结合,**ek-prompt** 有效地优化并指导了模型生成事件三元组的过程,提高生成三元组的准确性与一致性。这种结构化的 **ek-prompt** 在提升模型理解复杂事件关系的性能,以及扩展 ECKG 任务中展现出显著的效果。

3.2 微调模型

尽管通过第 3.1 节的 **ek-prompt** 能在一定程度上扩展了 ECKG,但 **ek-prompt** 获得的事件依赖于模型的知识库及其对 **ek-prompt** 的理解,以至于生成较为抽象的三元组。例如,使用 **ek-prompt** 生成的“吃饼干”的子事件三元组为<吃饼干,有子事件,选择饼干>和<吃饼干,有子事件,享用饼干>。为了进一步丰富 ECKG,除了抽象的三元组外,还需要获取更为细致和具体的三元组。如“吃饼干”的子事件三元组还应包括<吃饼干,有子事件,张开嘴巴>、<吃饼干,有子事件,咀嚼饼干>和<吃饼干,有子事件,吞咽饼干>等。

GPT-3.5-Turbo 是由 OpenAI 开发可以进行微调的模型,该模型具有较高的响应速度和生成能力,在许多自然语言处理任务中展现了卓越的性能。因此,本文选择在特定的数据集上微调 GPT-3.5-Turbo 模型,使模型能够捕捉到更为细致的信息,从而更好适应 ECKG 扩展的需求。具体地,将 ECKG 的三元组以一种标准的格式传递给 GPT-3.5-Turbo 模型进行学习,以增强模型生成事件三元组的能力。对于微调 LLM 的任务,数据集的质量对于提高模型在各种下游任务中的表现至关重要。针对 GPT-3.5-Turbo 模型生成三元组的任务,本文为 ECKG 的每种事件关系标注一个高质量的数据集。与此同时,为了验证新三元组的准确性,还将为每种事件关系准备一个数据集来微调 GPT-3.5-Turbo 模型,以评估新三元组的正确性。下面将具体介绍生成新三元组与验证三元组任务的数据集构建过程。

(1) 生成三元组的数据集构建

在扩展 ECKG 的任务中,我们的目标是将 ECKG 的任意一个事件在特定关系下关联多个事件,即多对一和一对多的数据结构。如图 2 所示(图中蓝色的实线圆和箭头代表原始三元组,橙色的虚线圆和箭头表示新增的三元组),对于“吃饼干”事件,在扩展之前,其先于关系关联的事件有“有饼干”和“能张开嘴巴”,而子事件关系下的事件包括“咀嚼饼干”和“吞咽饼干”。在扩展之后,在先于关系下应增加“买饼干”“看见饼干”和“想吃饼干”事件,子事件关系下则应该增加“拿起饼干”“张开嘴巴”和“放进嘴里”事件。

因此,在构建用于生成三元组的数据集时,本文特别标注了多对一和一对多的数据结构。如图 2 中,为扩展“吃饼干”的头三元组标注的数据为<E,先于,吃饼干>;相似地,扩展的尾三元组标注为<吃饼干,有子事件,E>,其中 E 表示“吃饼干”事件在 ECKG 中的先于和子事件关系下对应的事件。

完成对每种事件关系三元组的数据集标注后,将其转换为 GPT-3.5-Turbo 模型可理解的 JSONL 格式,便于模型解析和处理。例如,将图 2 的“吃饼干”的子事件三元组进行格式化为以下结构。

```
{“messages”: [{“role”: “system”, “content”: “这是一个生成事件知识图谱中每个总事件对应的所有可能的子事件的助手。”}]}
```

```
{“role”: “user”, “content”: “‘吃饼干’的子事件有哪些?”}.
```


{“role”: “assistant”, “content”: “咀嚼饼干, 吞咽饼干”}}.

在格式化子事件关系数据集时, 整个数据集由一系列用户和助手之间的交互数据组成, 每条数据都涵盖一个总事件的问题描述和一组关联的子事件列表. 比如, 上述实例涉及总事件“吃饼干”, 以及子事件“咀嚼饼干”和“吞咽饼干”.

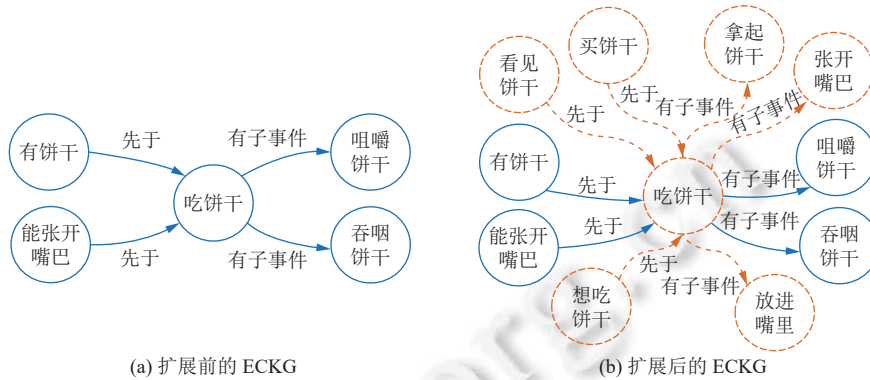


图2 扩展 ECKG 实例展示

(2) 验证三元组正确性的数据集构建

使用 LLM 生成新三元组后, 需要对三元组进行正确性评估. 考虑到 LLM 生成的是大规模三元组, 仅依靠人工评判并不现实. 因此, 本文采用专用于验证任务的微调后的 GPT-3.5-Turbo 模型来对新三元组进行自动评估.

对于评估任务, 本文构建了包含正样本和负样本数据集, 并对事件三元组进行格式化处理, 以便模型能够更全面地学习事件三元组的特征, 并具有广泛的适应性. 以子事件关系为例, 数据格式化形式如下.

(a) 正样本数据格式化:

```
{“messages”: [{“role”: “system”, “content”: “这是一个验证事件知识图谱中的有子事件关系的助手.”},
{“role”: “user”, “content”: “‘吃饼干’有子事件是‘咀嚼饼干’是否正确?”},
{“role”: “assistant”, “content”: “正确”}]}
```

(b) 负样本数据格式化:

```
{“messages”: [{“role”: “system”, “content”: “这是一个验证事件知识图谱中的有子事件关系的助手.”},
{“role”: “user”, “content”: “‘吃饼干’有子事件是‘不喜欢饼干’是否正确?”},
{“role”: “assistant”, “content”: “不正确”}]}
```

通过以上步骤, 本文完成了为生成和验证事件三元组任务所需的数据集预处理工作, 再将数据集按照 7:2:1 的比例划分为训练集、验证集和测试集. 随后, 将训练集和验证集上传到 OpenAI 的服务器进行模型训练, 而测试集则用来评估微调后模型的性能. 这一过程不仅增强了模型生成事件三元组的能力, 还提高了对三元组评估的准确性, 为扩展 ECKG 的工作提供了有效的支撑.

3.3 事件共享机制

由于 LLM 是基于大规模数据集预训练的模型, 具有生成多样化回答的能力. 即面对相似或相同的输入, LLM 能够利用其内部强大的语言模式和知识库, 提供多种不同并且合理的答案. 例如, 以图 3 为例 (实线圆表示共同拥有的事件, 蓝色与橙色圆角矩形是可以共享的事件, 虚线箭头连接共享事件组成新三元组), 当 LLM 扩展“帮助朋友”和“帮助他人”事件的时序关系时, “收获信任”在这两个事件的三元组中同时出现, 而“获得感激”与“收获感激”“强化关系”与“增强关系”在语义上相似. 目前, 图 3 中的圆角矩形框的事件分别是“帮助朋友”和“帮助他人”独有的事件, 但我们认为“帮助朋友”先于“解决难题、获得尊重”, 以及“帮助他人”先于“收到回报、传递正能量”也合理 (如图 3 中的虚线箭头所连接的事件组成的三元组).

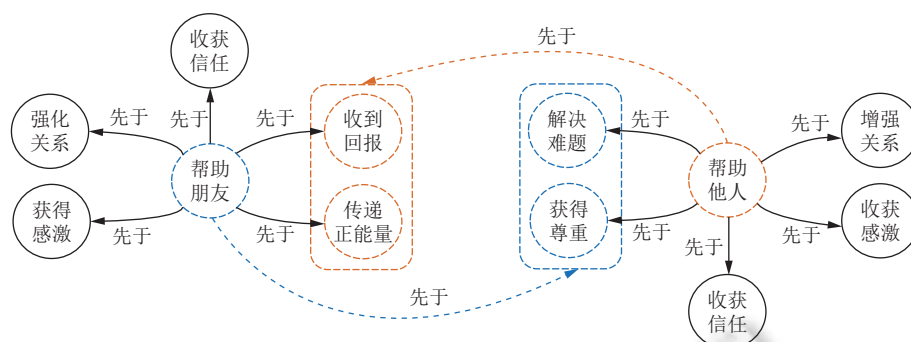


图3 事件共享示例图

为了克服相似事件生成的三元组不一致的问题, 本文提出事件共享机制。即在相同事件关系下, 将语义相似的 ECKG 事件所关联的事件进行互相共享。此外, 通过分析原始事件的语义相似度, 不仅可以增强原始事件的相似语义 (如图 3 中的“帮助朋友”和“帮助他人”这两个事件), 还有效解决了由 LLM 生成事件三元组时知识碎片化的问题。

在获取 ECKG 事件的语义相似度方面, 本文提出了一种融合多维度语义特征的嵌入式方法。对于获取事件的高质量嵌入向量, 本文选用 OpenAI 开发的 text-embedding-ada-002 模型。这是一个基于大型语料库预训练的模型, 能够将文本信息转换为 1536 维的向量, 进而捕捉文本的丰富语义内容, 这对于精确计算事件之间的相似度起到至关重要的作用。在语义相似度的计算过程中, 采用了余弦相似度作为度量标准。余弦相似度通过计算两个向量在多维空间中夹角的余弦值来评估它们之间的相似度, 其值的范围从 -1 (完全相反) 到 1 (完全相同), 0 表示两个向量互相独立。余弦相似度的计算公式为:

$$\text{CosSim}(v_1, v_2) = \frac{\sum_{i=1}^n v_{1i} \cdot v_{2i}}{\sqrt{\sum_{i=1}^n v_{1i}^2} \sqrt{\sum_{i=1}^n v_{2i}^2}} \quad (1)$$

其中, v_1 和 v_2 是两个非 0 的向量, v_{1i} 和 v_{2i} 分别代表向量 v_1 和 v_2 在第 i 个维度上学习到的特征值。

具体地, 本文通过融合涵盖实体类型的事件语义角色、关联句子成分的语义角色、语义角色的标注结构和事件反义多维度特征来计算事件余弦相似度, 以捕捉事件中的细微语义差异并精确量化它们之间的相似性。这 4 种维度的详细介绍如下。

(1) 基于实体类型的事件语义角色 (event semantic role, ESR)

包含实体类型的事件语义角色是指事件中的主体、时间和地点等成分扮演的语义角色, 以及每个成分对应的实体类型^[14,26]。在本文中, 采用了课题组提出的包含实体类型的事件语义角色分类体系^[14,45,48], 该分类体系已经应用于对 ECKG 中所有事件进行详细的语义角色标注。具体地, 该语义角色分类体系主要包含以下 3 个部分。

(a) 中枢语义角色: 表示一个事件的核心, 通常由谓词扮演, 并且每个事件仅有一个中枢语义角色。课题组归纳的中枢语义角色包括 8 种一级类别和 26 种更为详细的二级子类^[49,50]。表 3 展示了中枢语义角色的具体类别、定义和示例, 其中下划线字体为二级子类对应的中枢词。

(b) 周边语义角色: 指事件中枢所涉及的处于周边的语义角色, 一个事件可以包含多个周边语义角色。课题组对周边语义角色进行了细致的分类, 划分为 9 个一级类别和 32 个具体的二级子类 (即周边语义角色)^[49-57]。表 4 详细展示了周边语义角色的类别、定义和具体判断标准, 其中括号中的内容为例句, 下划线字体表示对应的周边语义角色。

(c) 实体类型: 事件中句子成分对应的实体类型。在上述 (a) 和 (b) 的语义角色标注过程中, 会存在同一个语义角色可能对应不同的实体类型的情形。例如, 在“我买衬衫”和“我买手机”这两个事件中, 语义角色标注均为“施事买受事”。然而, 衬衫属于“衣物”类, “手机”属于“通信电子设备”。因此, 为了获取事件更为精确的语义信息, 我们需

要对事件成分的实体类型进行标注 (课题组总结了 267 个实体类型^[14,58]). 表 5 展示了部分实体类型, 其中下划线内容代表实体类型相关的事件部分.

表 3 事件中枢语义角色的分类

一级类别	定义	二级子类	定义	例子
状态	事件主体静态的“存在”或“性质”	存在	主体的存在状态或相对静止的姿态	主席 <u>坐</u> 椅子上
		特征	事件主体呈现的非可控属性	哥哥 <u>高</u> 一米八三
		态度	有意志主体表现出的可控属性	父母 <u>温柔</u>
心理	主体的感知、情感、思考或愿望等心理活动	感知	有意志主体的感觉、感知等	王老师 <u>疲倦</u>
		情感	有意志主体的情绪或感情等	孙子 <u>想念</u> 奶奶
		思想	有意志主体的思考或愿望等	老师 <u>思考</u> 问题
关系	事物之间静态的相互联系	领属	事件中的领属或整体与部分关系	人 <u>有</u> 两只眼睛
		类同	事件中客观存在的类同关系以及主观认定的认同关系	妹妹 <u>像</u> 姐姐
		关联	主体跟其他系体间除了领属和类同之外的各种关系	窗户 <u>朝向</u> 西边
自动	主体不涉及外界其他事物的自身的进展或变化、移动或活动	自变	主体自身的各种变化	小蝌蚪 <u>变成</u> 青蛙
		进展	主体自身在时间上的进展	学生 <u>进步</u>
		自移	主体在空间的自身移动	总理 <u>前往</u> 村庄
关涉	主体仅关涉到其他事物但是不改变其他事物的行动	活动	主体不强调空间位移的自身活动或非可控的运动	地球 <u>自转</u>
		遭受	主体非可控的遭遇或得失客体	老板 <u>亏损</u> 很多钱
		对待	有意志主体以某种态度来对待或协助邻体的行动	警察 <u>保护</u> 人民
改动	主体改变客体或促使客体改变行动	寻求	有意志主体寻找或等候客体	孩子 <u>寻找</u> 妈妈
		支配	有意志主体碰撞、加工、管理、控制或养育等改变客体的行动	工人 正在 <u>施工</u>
		质变	主体导致客体发生化学物理或生理变化的行动	保姆 <u>加热</u> 牛奶
交际	事件主体探询信息、显示信息以及互相交流信息的行动	促使	主体使客体在时间上进展, 或使客体实现主体的意图的行动	司令 <u>派</u> 士兵 打探消息
		创建	主体创立、建造或产生新的客体的行动	叔叔 <u>创办</u> 公司
		探询	主体询问或观察以获取信息的行动	记者 <u>采访</u> 作家
转移	事件主体索取、给予、交换或搬移客体的行动	传播	主体传递信息或交流信息以及显示信息的活动	老师 <u>表扬</u> 学生
		给予	主体无偿向外转让客体所有权的行动	志愿者 <u>捐款</u>
		索取	有意志主体无偿向内转移客体所有权的行动	小偷 <u>偷了</u> 奶奶 手机
		交易	有意志主体有偿地买卖或互相交换客体的行动	哥哥 用奖金 <u>采购</u> 一台电视
		搬移	主体搬动客体或自移主体自身的活动	朋友 <u>挥动</u> 手臂

表 4 事件周边语义角色的分类

一级类别	定义	周边语义角色	判断标准和对应例子
主体	事件的中枢语义角色都必有的运动主体	施事	发出直接可控行为或思维活动的有意志的主体: ① 自身亲自执行可控行动的有意志的主体 (<u>运动员</u> 跑步); ② 发出指令指使客体完成而自身并未亲自执行该指令的主体 (<u>老大</u> 指使手下偷窃); ③ 自身在进行思维活动的有意志的主体 (<u>学生</u> 思考问题)
		雇施事	有意志的主体, 但它不是事件中行为的直接发起者, 而是只授予或雇用一些真正的实体来发起事件中的行为操作 (<u>学校</u> 新建了足球场)
		代施事	有意志的实体产生的衍生物, 它能够代替该有意志的实体本身发出相关行为, 该衍生物主要包括: ① 有意志的主体产生的思想、观点或著作 (<u>这项研究</u> 提出了新的理论框架); ② 有意志的主体产生的举措或行为 (<u>新政策</u> 改进了城市的交通堵塞); ③ 有意志的主体产生的具有某种特定功能的无意志的物体等 (<u>电脑病毒</u> 偷盗了信息)
		当事	发出非可控行为的主体或各种关系的主体: ① 发出非可控行为的主体 (<u>雨水</u> 下落); ② 表现自身状态的主体 (<u>宿舍楼</u> 很高); ③ 遭受等内向活动所涉及主体 (<u>朋友</u> 收到一束花); ④ 除领属关系之外, 各种关系的主体 (<u>学校</u> 靠近市区)

表4 事件周边语义角色的分类(续)

一级类别	定义	周边语义角色	判断标准和对应例子
主体	事件的中枢语义角色都必有的运动主体	感事	非可控心理活动的主体、主体的感知部位, 或具有观察检测感知能力的人造装置或者人造物部件: ① 各种感知体验的主体 (孩子听到音乐); ② 各种情感状态的主体 (国民热爱祖国); ③ 具有观察检测感知能力的人造装置 (雷达感知到物品); ④ 各种感知体验主体的感知器官 (眼睛感到疲劳)
		领事	领属关系的主体: ① 领有关系的主体 (小明有个新手机); ② 整体部分关系的主体 (手机自带摄像头)
客体	事件中的主体之外的第2个参与事件的实体	受事	事件涉及及改变的客体, 主要包括支配行动在: ① 形态: 通过碰撞或加工等行为所改变的直接客体 (武松打虎); ② 位置: 通过搬移行为造成位置有所改变的直接客体 (物业搬走了椅子); ③ 所有权: 通过给予、索取和交易等行为造成所有权有所改变的直接客体 (老师送给我一本书); ④ 使用权: 通过给予、索取和交易等行为造成使用权有所改变的直接客体 (我借了一本书); ⑤ 主体执行管理、控制等行为所涉及的客体 (警察处罚犯人)
		客事	事件涉及的未改变的客体, 主要包括: ① 感知所获得的信息 (孩子听到音乐); ② 情感所涉及的客体 (我爱母亲); ③ 思想所涉及的客体 (我猜测出了结果); ④ 传播所传递的信息 (法官宣读判决书); ⑤ 探询所得到的信息 (警方打听通缉犯的行踪); ⑥ 遭受等内向活动所关涉到的客体 (孩子被妈妈一顿骂)
		成事	主体创建或产生的新客体以及主体实施参与事件产生的结果: ① 有意志主体创建的新客体 (学生写了一篇文章); ② 有意志主体对客体进行加工形成另一种状态的新客体 (妈妈把面团揉成了面条); ③ 有意志主体实施或参与事件产生的结果 (学生考了100分); ④ 无意志主体产生的新客体 (蚕分泌一种脑激素)
		变事	致变类行为造成自身发生变化或在时间上有所进展的客体: ① 致变类行为导致自身发生改变的客体 (工人拧紧了螺丝); ② 致变类行为导致在时间上有所进展的客体 (他们阻碍了工作速度)
		致事	主体通过促使类行为实现自己意图所促使的客体, 或者在事物促使行为下自发行动或呈现某种状态的客体: ① 主体为实现自己的意图, 通过促使类行为导致客体去执行后继的行动 (老师安排学生做实验); ② 在事物的促使行为下, 客体自发行动或者呈现出某种状态 (冷战会加深矛盾)
邻体	事件中的主体和客体之外的第3个参与事件的实体	源事	事件中时空和过程的起点以及作为来源的邻体: ① 时间的起点: 事件中作为时间的起点 (班车从7点开到8点); ② 空间的起点: 事件中作为空间的起点 (我从天津回到北京); ③ 发展过程的起点: 事件中作为过程的起点 (半自动化方法从手工提取发展到深度学习); ④ 信息的来源: 事件中交际活动作为来源的邻体 (他拜访了老师); ⑤ 物品的起点: 事件中转移活动中作为来源的邻体 (小偷盗窃银行金钱); ⑥ 组织或者活动的起点: 事件中主体退出的组织或者活动 (队员退出了比赛)
		经事	事件中经历的空间或过程: ① 事件所经过的空间 (航班经过天津); ② 事件所经历的过程 (学生经过高考考上了大学)
		宿事	事件中的主体进行一系列活动后的归属方: ① 主体经过活动后已经到达的空间归宿点 (我从广州经过深圳回到香港); ② 主体经过活动后已经达到的时间归宿点 (爸爸加班到深夜); ③ 主体加入的一个组织 (我加入了中国共产党); ④ 主体参加的活动 (我参加音乐会); ⑤ 发展过程归宿点: 事件中事物发展过程的结束阶段 (购物方式从实体店购物发展到网上购物)
		向事	事件所“趋向”的邻体, 信息、转移物以及有意志或无意志主体行为所产生利益的接受者: ① 动态所趋向的邻体 (医生赶往疫情重灾区); ② 交际活动中的信息接收者 (校长向新生致开学词); ③ 转移活动中转移物的接收者 (学校为优秀学生颁发奖状); ④ 有意志或无意志主体行为所产生明显利益的接收者 (老师指导学生写论文)
伴体	事件所涉及的事物以及所凭借的事物	同事	事件中的伴随者、排除者或作为基准的邻体: ① 事件中的伴随者 (我陪妈妈去散步); ② 事件中的排除者 (除了中文之外我还会英语); ③ 主体比较或测量的基准 (爸爸比妈妈大两岁)
		用事	事件中所用到的方法、时耗、工具以及材料等. 按照客体用事种类的不同, 用事一般可分为: ① 方法类用事 (学生使用速记法记录了会议内容); ② 时耗类用事 (我用了半年时间写完了学位论文); ③ 工具类用事 (人们利用探测器寻找文物); ④ 材料类用事 (装修工人使用抗菌涂料粉刷墙壁)等

表 4 事件周边语义角色的分类 (续)

一级类别	定义	周边语义角色	判断标准和对应例子
系体	与事件的参与角色 有关系的事物	属事	领属关系中所属的: ① 亲属朋友 (阿姨有两个儿子); ② 敌人 (老鼠有很多天敌); ③ 财务 (朋友有一辆名车); ④ 构成部分 (手机自带摄像头)
		类事	事件中存在的类同者、类似者以及事件中主体最后变成的状态: ① 事件中客观存在或公认的类型关系的类同者 (豌豆是豆类蔬菜); ② 事件中的类似者 (白云酷似一朵棉花糖); ③ 有意志主体扮成的类同关系的类同者 (警察伪装成普通市民); ④ 有意志主体主观认同的类同者 (刘邦在咸阳自封为皇帝); ⑤ 与主体在数值方面相当的客体 (江苏省面积相当于欧洲小国卢森堡); ⑥ 事件中主体最后变成的状态 (水蒸气冷却成了水)
		连事	除领属关系和类同关系之外的关联关系的相关者, 主要包括: ① 临时性关系的相关者 (学校靠近市中心); ② 常识性关系的相关者 (广西壮族自治区毗邻越南); ③ 人际关系的相关者 (孙子与祖母关系很好); ④ 事件中的牵涉者 (国际油价上涨波及中国)
情由	引起事件的原因或 者目的	意图	事件所要达到的目的, 主要包括: ① 事件主体的意图 (她劝说儿子接管家族企业); ② 事件客体的意图 (我经常帮助同学复习功课)
		原因	造成某种结果或者另一件事情发生的条件 (孩子因为不按时睡觉而被妈妈批评)
		依据	事件发生所遵循的根据或标准 (员工按照老板的指示办事)
时空	事件所涉及的时间 和空间	时间	事件所发生的时间点或时期: ① 事件发生的时间点 (中央新闻联播于晚上 7 点播出); ② 事件发生的时期 (儿子于 2013 年出生)
		位事	事件发生所处的空间, 主要包括: ① 事件发生所处的物理空间 (我在实验室做实验); ② 事件发生所处的网络空间 (经理在微信发布公告); ③ 事件发生所处的组织空间 (他在培训班学习先进的技术); ④ 事件发生所处的思想空间 (小偷在我的梦里偷东西); ⑤ 事件发生所处的社会环境 (小明在困境想出了解决的方法)
		范围	事件所关涉的范围或限定的界限: ① 事件所涉及的范围 (校长就当前学校改革发表了意见); ② 事件所限定的界限 (计算机方面我很在行)
描述	对事件中的事物某 个方面或者特征的 阐述或说明	属性	事件中对象的某一个方面的信息, 通常与值事搭配使用 (华为 P40 手机的官网售价是 4 999 元)
		性质	事件中对象本身所具有的与其他对象不同的特征 (这种塑料具有很好的耐高温性)
度量	衡量或者计量事件 中的事物某方面属 性值的统称	物量	事件所涉及的物体的数量: ① 事件中物体的数量 (这件衣服的扣子有 6 个); ② 事件中度量衡的数值量 (广西到北京大约 2 000 公里)
		时量	事件本身持续的时段或迄今所经历的时段数值: ① 事件本身所持续经历的时段数值 (这个实验做了半个月还没做完); ② 事件结束后到现在所经历的时段 (这瓶牛奶已经过期了一个月)
		频量	事件中行动或变化所进行的次数以及以所用工具表示的动量: ① 事件中实体行动重复的次数 (我一年回家两次); ② 事件中事物变化重复的次数 (苹果每年可以结果 3 到 4 次); ③ 事件中所用工具表示的动量 (修理工敲了 3 下才修好机器)
		值事	事件中一个对象在某个属性上的取值, 通常与属性搭配使用 (华为 P40 手机的官网售价是 4 999 元)

总体而言, 当我们对一个事件进行语义角色标注时, 首先需要识别事件的中枢语义角色, 这一角色为事件的主干, 较为容易判断 (如表 3 所示). 接着, 根据表 4 的判断标准详细标注事件中的每个周边语义角色. 最后, 对事件中句子成分的所对应的实体类型进行标注 (如表 5 所示). 例如, 对于事件“招待客人”, 我们将其标注为“施事: 人 招待客人/向事: 人”其中, “招待”作为中枢语义角色; “施事、向事”是围绕中枢的周边语义角色; “人”表示实体的类型.

涵盖实体类型的事件语义角色的标注不仅有助于理解复杂的事件结构, 揭示事件中不同实体间关系, 还能深入挖掘事件的深层含义. 此外, 课题组构建的 ECKG 中包含实体类型的语义角色数据经过多方面的应用和验证, 证明了其准确性与可用性. 这些验证方法包括长期的多轮人工审核、设计了一个使用中枢语义角色注意力机制的语义角色识别系统^[59]、ECKG 的逻辑推理应用^[27]以及事件语义解析^[60]. 因此, 本文使用了 ECKG 中含实体类型的语义角色标注数据来计算事件的余弦相似度 (cosine similarity of ESR, $CosSim_ESR$), 下面以两个例子详细介绍 $CosSim_ESR$ 的计算过程:

事件 1: 招待客人.
事件 2: 接待客人.
在计算事件 1 和事件 2 的余弦相似度时, 首先从 ECKG 的语义角色标注数据中获取它们对应的标注数据:
事件 1 标注: 施事: 人 招待 客人/向事: 人.
事件 2 标注: 施事: 人 接待 他人/向事: 人.

表 5 实体类型的分类

实体类型	实体类型标注示例	实体类型	实体类型标注示例
人	<u>朋友/施事: 人</u> 哭泣	时长属性	<u>吃饭时间/当事: 时长属性</u> 短
人群	<u>父母/施事: 人群</u> 争吵	国家	<u>施事: 国家</u> 举办 奥运会/成事: 竞赛
物体	施事: 人 买 <u>材料/受事: 物体</u>	生物属性	<u>睡眠/当事: 生物属性</u> 不足
食物	施事: 人 吃 <u>饼干/受事: 食物</u>	信息传达过程	当事: 人 不能 <u>交流/动事: 信息传达过程</u>
量化属性	<u>二氧化碳浓度/当事: 量化属性</u> 上升	建筑物部件	施事: 人 跳 <u>窗/源事: 建筑物部件</u>
钱财	施事: 人 兑 <u>现金/受事: 钱财</u>	人造物	当事: 人 丢失 <u>物品/客事: 人造物</u>
动物	施事: 人 养 <u>宠物/客事: 动物</u>	植物躯体部件	施事: 人 点燃 <u>木头/变事: 植物躯体部件</u>
外观属性	施事: 人 保持 <u>身材/客事: 外观属性</u>	意向性过程	当事: 人 不能 <u>开车/动事: 意向性过程</u>
心理属性	领事: 人 有 <u>经济压力/属事: 心理属性</u>	电气器件	施事: 人 修 <u>电脑/受事: 电气器件</u>
人躯体部件	<u>手指/变事: 人躯体部件</u> 被 刺破	时间属性	施事: 人 填写 <u>日期/成事: 时间属性</u>
地理区域	当事: 灯 照亮 <u>周围环境/变事: 地理区域</u>	辐射过程	当事: 人 接触 <u>高压电/客事: 辐射过程</u>
生理属性	当事: 人 失去 <u>意识/客事: 生理属性</u>	交通工具部件	施事: 人 踩 <u>油门/受事: 交通工具部件</u>
情绪值	感事: 人 感到 <u>轻松/客事: 情绪值</u>	反常解剖结构	当事: 人 长 <u>皱纹/成事: 反常解剖结构</u>
交通工具	<u>汽车/当事: 交通工具</u> 抛锚	损害过程	当事: 人 出 <u>事故/客事: 损害过程</u>
空间区域	<u>天/位事: 空间区域</u> 下雨	蒸馏酒精饮料	施事: 人 买 <u>啤酒/受事: 蒸馏酒精饮料</u>

接下来, 剔除事件 1 标注和事件 2 标注中句子成分“客人”和“他人”, 得到对应的 ESR (包括中枢语义角色、周边语义角色和实体类型) 为:

ESR₁: 施事: 人 招待 向事: 人.
ESR₂: 施事: 人 招待 向事: 人.

然后, 将事件 1 和事件 2 的整个 ESR (在此例中两个 ESR 相同, 都为“施事: 人 招待 向事: 人”) 通过 text-embedding-ada-002 模型得到了 1536 维的嵌入向量. 由于 ESR₁ 和 ESR₂ 相同, 因此它们的嵌入向量也相同. 这些向量的前后 4 个维度值 (保留小数点后 4 位) 分别是:

$[-0.0200, -0.0163, 0.0249, -0.0093, \dots, 0.0157, -0.0094, -0.0017, -0.0040]$.

最后, 通过公式 (1) 计算这两个向量的余弦相似度, 从而得到含实体类型的语义角色标注 1 和标注 2 的余弦相似度值. 由上述得知这两个标注具有相同的嵌入向量, 因而它们的余弦相似度等于 1.

(2) 关联句子成分的语义角色标注 (semantic role labeling of associated sentence components, SRL-ASC)

通过分析事件的含实体类型语义角色相似度, 我们发觉尽管有些事件在语义角色上具有很高的相似度, 但它们的实际含义却有所差异. 以事件 3 和事件 4 为例进行说明:

事件 3: 父母爱孩子.
事件 4: 朋友爱不合适的人.
在 ECKG 中, 事件 3 和事件 4 的标注分别为:
事件 3 标注: 父母/施事: 人 爱 孩子/向事: 人.
事件 4 标注: 朋友/施事: 人 爱 不合适的人/向事: 人.

从事件 3 标注和事件 4 标注可以看出, 它们对应的含实体类型语义角色标注相同, 均为“施事: 人 爱 向事: 人”. 但事件主体分别为“父母”和“朋友”, 客体分别是“孩子”和“不合适的人”, 从而导致整个事件所表达的情感含义存在差异.

因此,为了解决含实体类型的语义角色高度相似,而忽视事件实际含义差异的问题,我们考虑了关联句子成分(如主语、宾语和补语等)的含实体类型语义角色标注.例如,在计算事件3与事件4的相似度时,实际上是在衡量这两个事件的包含句子成分标注的余弦相似度($CosSim_SRL-ASC$),下面以事件3和事件4为例介绍 $CosSim_SRL-ASC$ 的计算步骤.

首先,基于 text-embedding-ada-002 模型分别获取事件3标注(父母/施事:人爱孩子/向事:人)和事件4标注(朋友/施事:人爱不合适的人/向事:人)的1536维嵌入向量.这两个向量的前后4个维度值(保留小数点后4位)分别为:

SRL-ASC 嵌入向量 1: [0.0066, -0.0206, -0.0093, -0.0098, ..., 0.0077, 0.0024, 0.0020, -0.0213].

SRL-ASC 嵌入向量 2: [-0.0120, -0.0222, 0.0184, 0.0049, ..., 0.0123, -0.0138, -0.0020, -0.0273].

其次,使用公式(1)计算这两个向量余弦相似度,从而得到事件3和事件4的SRL-ASC的余弦相似度.

(3) 语义角色标注结构 (semantic role labeling structure, SRLS)

事件中包含的词组结构能够揭示事件参与者之间的相互作用,对深入理解事件之间的复杂关系起到了关键的作用.例如,在标注“施事:人把客事:信息告诉向事:人”中,通过“把...告诉...”的结构,指明了行为的方向与对象.这些结构特征在计算事件的语义相似度时提供了有力的指导.因此,本文将语义角色标注格式化为“施事-把-客事-告诉-向事”,进而计算语义角色标注结构的相似度(Sim_SRLS),以事件5和事件6具体阐述 Sim_SRLS 的计算过程:

事件5: 把秘密告诉他人.

事件6: 把物品归还给他人.

首先,在ECKG标注数据中获取事件5和事件6的标注分别为:

事件5标注: 施事:人把秘密/客事:信息告诉他人/向事:人.

事件6标注: 施事:人把物品/受事:人造物归还给他人/向事:人.

然后,去除事件5标注和事件6标注中的句子成分和实体类型,将中枢语义角色和周边语义角色通过连接符“-”进行连接,得到对应的语义角色标注结构为:

语义角色标注结构5: 施事-把-客事-告诉-向事.

语义角色标注结构6: 施事-把-受事-归还-给-向事.

最后,使用公式(2)计算语义角色标注结构5和语义角色标注结构6的相似度.公式(2)的定义如下.

假设,两个事件的语义角色标注结构为 S_1 和 S_2 其包含的语义角色列表可以表示为 $L_1 = [s_{11}, s_{12}, \dots, s_{1m}]$ 和 $L_2 = [s_{21}, s_{22}, \dots, s_{2n}]$,相似度计算公式如下:

$$Sim_SRLS(S_1, S_2) = \frac{\sum_{i=1}^{\min(m,n)} (s_{1i} = s_{2i})}{\max(m, n)} \quad (2)$$

其中, $Sim_SRLS(S_1, S_2)$ 表示两个事件语义角色标注结构的相似度, m 是 L_1 的元素数量, n 是 L_2 的元素数量, s_{1i} 和 s_{2i} 分别是 L_1 和 L_2 的第 i 个元素.

将语义角色标注结构5和语义角色标注结构6使用公式(2)计算时,对应的语义角色元素列表分别为:

语义角色标注结构5的元素列表: [施事, 把, 客事, 告诉, 向事].

语义角色标注结构6的元素列表: [施事, 把, 受事, 归还, 给, 向事].

这两个结构的语义角色元素个数分别为 $m=5$, $n=6$, 两个角色列表中相同位置上相等的元素数量为2, 从而计算出这两个语义角色结构的相似度约为0.3333.

(4) 事件反义

在ECKG中,反义事件是指含义相对立的事件.例如,“吃汉堡包”与“不吃汉堡包”这两个事件,尽管它们在行为(吃)和对象(汉堡包)上相同,但表达的语义截然相反.这种语义上的对立在计算事件相似度时尤为重要,能够影响人们对事件关系的理解.然而,传统的余弦相似度算法往往会将反义事件判定为高度相似.为了解决这一问题,本文首先使用预先定义的规则识别反义词.接着,在计算事件余弦相似度时,将这些互为反义的事件相似度进

行调整为 0, 以降低反义事件之间的相似度得分.

基于以上 4 个维度计算得到 ECKG 事件之间的相似度, 再进行加权平均后, 就能得到两个事件的综合相似度. 加权平均相似度的计算公式为:

$$\begin{aligned} Sim_{wt}(E_i, E_j) = & CosSim_ESR(v_{esri}, v_{esrj}) \times wt_CosSim_ESR \\ & + CosSim_SRL-ASC(v_{srl-asci}, v_{srl-ascj}) \times wt_CosSim_SRL-ASC \\ & + Sim_SRLS(S_i, S_j) \times wt_Sim_SRLS \end{aligned} \quad (3)$$

其中, $CosSim_ESR(v_{esri}, v_{esrj})$ 和 wt_CosSim_ESR 分别表示第 i 个与第 j 个含实体类型的语义角色嵌入向量的余弦相似度及其对应的权重; $CosSim_SRL-ASC(v_{srl-asci}, v_{srl-ascj})$ 和 $wt_CosSim_SRL-ASC$ 分别是第 i 个与第 j 个关联句子成分且含实体类型的语义角色嵌入向量的余弦相似度与对应的权重; $Sim_SRLS(S_i, S_j)$ 和 wt_Sim_SRLS 分别代表第 i 个与第 j 个语义角色标注结构的相似度及权重. 经过多轮的实验测试, 本文最终设置 wt_CosSim_ESR 、 $wt_CosSim_SRL-ASC$ 和 wt_Sim_SRLS 这 3 个权重的值分别为 0.6、0.2 和 0.2.

经过实验证明, 本文方法在准确性上实现了显著的提升, 证明了本文方法在评估计算事件相似度上的有效性. 算法 1 展示了计算 ECKG 事件相似度的整个流程.

算法 1. 计算 ECKG 事件相似度算法.

输入: 含实体类型的事件语义角色 (ESR)、关联句子成分的语义角色标注 (SRL-ASC) 和语义角色标注结构 (SRLS) 文件;

输出: 加权平均相似度.

1. 初始化嵌入向量大小 $embedding_size = 1536$
 2. 初始化空的 $embedding_df$ 存储嵌入向量
 3. 分别遍历 ESR 和 SRL-ASC 数据
 4. 通过 API 调用 OpenAI 的 $embedding$ 模型
 5. 获取 ESR 与 SRL-ASC 的嵌入向量 v_{esr} 和 $v_{srl-asc}$
 6. 引入反义词特征, 计算余弦相似度
 7. 当两个事件互为反义:
 8. $adjusted_similarity *= 0$
 9. 获取余弦相似度 $CosSim_ESR$ 和 $CosSim_SRL-ASC$
 10. 遍历 SRLS 数据
 11. 获取事件标注结构中相同的元素数量 num
 12. 计算结构的相似度 $Sim_SRLS = num/max$ (标注结构序列长度)
 13. 分配相似度权重 $wt_CosSim_ESR=0.6$, $wt_CosSim_SRL-ASC=0.2$, $wt_Sim_SRLS=0.2$
 14. 加权平均 $CosSim_ESR \times 0.6 + CosSim_SRL-ASC \times 0.2 + Sim_SRLS \times 0.2$
 15. 得到 ECKG 事件之间的相似度
-

在算法 1 中, 首先, 初始化嵌入向量的大小、存储 ESR 和 SRL-ASC 的嵌入向量数据 (第 1 和 2 行). 接下来, 遍历 ESR 和 SRL-ASC 的数据, 并通过 OpenAI 的嵌入模型分别获取对应的嵌入向量 (第 3–5 行). 在计算 ESR 和 SRL-ASC 余弦相似度时, 引入事件反义特征, 将互为反义的事件相似度调整为 0, 进而得到语义角色和关联句子成分的语义角色标注的相似度 (第 6–9 行). 然后, 引入事件标注结构特征, 并计算了结构的相似度 (第 10–12 行). 最后, 通过加权平均 ESR、SRL-ASC 和 SRLS 的余弦相似度, 得到了 ECKG 事件之间的综合相似度 (第 13–15 行).

4 实验与分析

本节展示具体的实验结果, 并进行了详细分析和讨论, 以证明本文方法在扩展事件常识知识图谱的实用性.

4.1 数据集

本文的数据集主要包括 ek-prompt 所使用的原始 ECKG 的事件节点, 以及微调 GPT-3.5-Turbo 模型所需的事件三元组.

4.1.1 ECKG 事件三元组

课题组成员从各类资源中手工收集了常识性的事件三元组, 并经过精细的筛选和修正, 得到了高精度的事件三元组, 进而构建成了 ECKG, 表 6 展示 ECKG 中各种事件关系的三元组数量分布情况. 经去重处理后, ECKG 共包含了 8466 个事件, 这些事件主要提供给 ek-prompt 扩展因果、时序和条件关系. 此外, 对于 ek-prompt 扩展子事件关系三元组所需的数据, 本文根据第 2 节的事件语义的分类, 筛选出具有子事件关系的事件为 4484 个.

表 6 ECKG 事件三元组

事件关系	三元组数量
因果	4810
时序	7821
条件	7316
子事件	1598
目标	3343
同现	186
推出	1532
总计	26606

4.1.2 微调模型数据集

在微调 GPT-3.5-Turbo 模型的过程中, 本文采用的数据集是由 ECKG 原始事件三元组与通过 ek-prompt 生成并且经过人工评估的新三元组共同组成. 微调模型主要用于生成和验证 4 种事件关系的新三元组, 对于生成任务, 需要分别为生成 ECKG 事件的头三元组和尾三元组构建数据集, 而验证任务只需要一个数据集. 每种事件关系的数据集至少包含 1 万条数据, 并按照 7:2:1 划分为训练集、验证集和测试集, 表 7 详细展示了不同任务中各种事件关系的数据集统计情况.

表 7 微调模型数据集统计

微调任务	事件关系	训练集	验证集	测试集	总计
生成头三元组	因果	8397	2399	1200	11996
	时序	7070	2020	1010	10100
	条件	7035	2010	1005	10050
	子事件	7655	2187	1094	10936
生成尾三元组	因果	7150	2043	1021	10214
	时序	7084	2024	1012	10120
	条件	7014	2004	1002	10020
	子事件	7632	2181	1090	10903
验证新三元组	因果	7056	2016	1008	10080
	时序	7119	2034	1017	10170
	条件	7030	2009	1004	10043
	子事件	9243	2641	1321	13205

4.2 实验设置

在扩展 ECKG 三元组的任务中, 本文实现了基于 GPT-4-Turbo 模型设计的 ek-prompt、微调 GPT-3.5-Turbo 模型, 以及使用事件共享机制 3 个扩展策略. 在使用 GPT-4-Turbo 模型时, 本文为 4 种事件关系分别设计高质量的 ek-prompt 模板. 对于微调模型生成事件三元组任务, 为了使 GPT-3.5-Turbo 模型更好地学习到每个 ECKG 事件的头三元组和尾三元组的特征, 以及它们之间的紧密联系, 本文对同一种事件关系下的生成头三元组和尾三元组

的任务共同微调同一个模型.

对于新三元组的正确性验证任务, 本文结合微调后的 GPT-3.5-Turbo 模型和人工审核的方法进行评估. 在模型微调阶段, 分别为每种事件关系微调一个模型, 以评估新三元组的正确性. 新三元组生成和验证过程中微调 GPT-3.5-Turbo 模型的参数设置如表 8 所示.

表 8 微调 GPT-3.5-Turbo 模型参数

微调任务	事件关系	n_epochs	batch_size	learning_rate_multiplier
生成三元组	因果	3	2	2
	时序	3	3	2
	条件	3	3	2
	子事件	3	1	2
验证三元组	因果	3	14	2
	时序	3	14	2
	条件	3	14	2
	子事件	2	12	2

针对事件共享机制扩展 ECKG 的任务, 首先按照第 3.3 节介绍的方法计算了 ECKG 事件之间的相似度. 接下来, 通过设置一定的阈值来筛选出高度相似的事件对. 在选择相似度阈值方面, 为了确保获取高精度和高质量的相似事件对, 本文通过人工审查不同相似度区间的事件对, 发现相似度在 $[0.95, 1)$ 区间的事件对具有较高的准确率 (如第 4.4.4 节的相似度实验结果所示). 因此, 本文选择 0.95 作为相似度阈值, 即保留相似度的大于或等于 0.95 的事件对. 尽管相似度低于 0.95 的事件对中可能存在一些被遗漏的相似事件对, 但这些语义相近的事件对所占比例较小, 影响不大. 如表 9 展示的相似度为 $[0.6, 0.95)$ 区间的语义相近事件对数量, 在表 9 中的每个区间随机抽取 200 个事件对进行考察的结果.

表 9 相似度区间 $[0.6, 0.95)$ 的语义相似事件对数量统计

统计项	$[0.6, 0.65)$	$[0.65, 0.7)$	$[0.7, 0.75)$	$[0.75, 0.8)$	$[0.8, 0.85)$	$[0.85, 0.9)$	$[0.9, 0.95)$
语义相近的事件对数量	0	0	0	0	0	1	5
语义相近的事件对占比 (%)	0	0	0	0	0	0.5	2.5

获取相似度区间 $[0.95, 1)$ 的所有事件对后, 本文经过人工筛选剔除了语义相似度较低的事件对, 仅保留了高度相似的事件对, 表 10 展示了 $[0.95, 1)$ 相似度区间的原始事件对及经过人工筛选后的事件对数量. 在人工筛选的过程中, 本文严格遵循一定的判断标准, 确保筛选出高质量的相似事件对. 具体地, 根据事件本身的含义、事件的语义角色以及在相同事件关系下所关联事件的相似性来判断每一个事件对是否语义相近. 最终, 在相同的事件关系中, 将人工筛选保留的相似事件对所关联的事件进行互相共享.

表 10 人工筛选相似事件对的数量统计

相似度区间	初始事件对数量	人工筛选后事件对数量	相似事件对保留比例
$[0.95, 1)$	20 071	18 848	93.91%

此外, 本文进一步对使用事件共享机制获取的新三元组与原始 ECKG 三元组、ek-prompt 以及微调模型生成的新三元组进行了相似度去重处理. 具体来说, 使用 WoBERT 预训练模型计算这些三元组的相似度, 并设置阈值为 0.82. 当相似度大于或等于这个阈值时, 将剔除该三元组. 而相似度小于 0.82 的三元组则为通过事件共享策略最终得到的新三元组. 该过程的实验在装备有 TITAN RTX 显卡的 Ubuntu Linux 操作系统环境中执行.

4.3 评估指标

4.3.1 自动评估

为了评估获取的新三元组的质量, 本文采用了 BERTScore^[61]方法计算新三元组的得分. BERTScore 是一种利用 BERT 模型^[62]比较生成文本与参考文本之间的语义相似度的方法, 提供了精确度 (P_{BERT})、召回率 (R_{BERT})、 $F1$

分数 (F_{BERT}) 这 3 个评估指标, 有效地衡量系统生成文本的整体性能. 这 3 个指标的介绍如下.

P_{BERT} 衡量了新三元组中的每个三元组找到的最相似参考三元组的相似度平均值, 以反映新三元组的整体质量. 计算公式为:

$$P_{\text{BERT}} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} X_i^T \hat{x}_j \quad (4)$$

其中, $\hat{x}$ 表示新三元组数据集, $|\hat{x}|$ 表示新三元组的总数量, $\hat{x}_j$ 表示第 j 个新三元组, x 表示参考三元组数据集, x_i 表示第 i 个参考三元组, $X_i^T \hat{x}_j$ 表示新三元组与参考三元组的余弦相似度.

R_{BERT} 测量了每个参考三元组找到的最相似新三元组的相似度平均值, 反映了参考三元组在新三元组中的覆盖度. 计算公式为:

$$R_{\text{BERT}} = \frac{1}{|x|} \sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} X_i^T \hat{x}_j \quad (5)$$

其中, $|x|$ 表示参考三元组的总数量.

F_{BERT} 是 P_{BERT} 和 R_{BERT} 的调和平均, 综合考量了新三元组的精确性和完整性. 计算公式为:

$$F_{\text{BERT}} = 2 \cdot \frac{P_{\text{BERT}} \cdot R_{\text{BERT}}}{P_{\text{BERT}} + R_{\text{BERT}}} \quad (6)$$

同时, 针对验证任务的微调模型, 选用 Alpaydin^[63] 的准确率、精确率、召回率和 $F1$ 值作为衡量指标, 以评估微调后模型的性能. 这些评估指标的详细说明如下.

准确率: 表示在新生成的事件三元组中, 正确事件三元组所占的比例. 计算公式为:

$$\text{准确率} = \frac{\text{正确的事件三元组数量}}{\text{事件三元组总数量}} \quad (7)$$

精确率: 表示在所有被标记为正确的事件三元组中, 实际为正确的比例. 计算公式如:

$$\text{精确率} = \frac{\text{正确识别的事件三元组数量}}{\text{标记为正确的事件三元组总数量}} \quad (8)$$

召回率: 表示在所有实际为正确的事件三元组中, 被正确标记出来的比例. 计算公式为:

$$\text{召回率} = \frac{\text{正确识别的事件三元组数量}}{\text{实际为正确的事件三元组总数量}} \quad (9)$$

$F1$ 值: 精确率和召回率的调和平均. 计算公式为:

$$F1 = 2 \times \frac{\text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}} \quad (10)$$

4.3.2 人工评估

为了进一步验证新三元组的准确性, 本文还组织了 4 名具备相关领域专业知识的研究生 (包括两名博士生和两名硕士生) 进行人工评估, 并计算了评估结果的准确率. 在评估时, 评估小组按照以下标准和步骤进行审查.

(1) 评估人员根据事件三元组的语义和逻辑关系是否符合常识、事件三元组表达的语义是否完整以及是否与现有的 ECKG 事件三元组的表达形式一致对测试集生成的所有事件三元组进行评估, 并记录每一个事件三元组的评估结果.

(2) 如果评估人员认为一个事件三元组是正确的, 则标记为“正确”; 若认为是错误的, 则标记为“错误”; 对于不确定一个事件三元组的准确性, 则标记为“不确定”.

(3) 所有评估人员的评估结果整合后, 对于评审结果一致的事件三元组, 将其认为最终结果.

(4) 对于审查结果不一致的事件三元组, 评估小组将再次讨论和审核, 直到至少有 3 位评估人员的评估结果一致, 才将这个结果作为该事件三元组的最终审查结果.

4.3.3 实际应用评估

此外, 本文基于课题组常识知识库的推理方法^[27]将新获取的事件三元组应用于逻辑推理任务, 以验证本文扩展的 ECKG 在具体应用中的实际表现. 具体来说, 使用事件语义角色和头尾事件约束谓词的翻译方法将新获取的

事件三元进行翻译, 并将翻译结果转换为 SWI-Prolog^[64]的规则形式进行常识推理. 其中, 头尾事件约束谓词是连接头尾事件语义角色的谓词, 能够加强事件之间的语义衔接. 以因果关系的事件三元组为例说明翻译方法和推理规则转换的具体过程如下.

(1) 翻译过程

已知 ECKG 中的数据为:

事件三元组: < 打排球, 导致, 出汗>.

事件语义角色标注: 施事: 人 打 排球/受事: 竞赛用品, 当事: 人 出汗.

头尾事件约束谓词: 等同 (施事 1, 当事 2).

在翻译过程中, 首先头尾事件的每个周边语义角色被抽象化为变量, 并使用上下位谓词“是一个”描述这些变量的值. 其次, 事件的中枢语义角色被翻译为作用在这些变量之间的逻辑谓词. 接着, 头尾事件约束谓词翻译成连接两个事件中具有一定语义联系的变量. 最后, 基于事件约束谓词的定义将同一个周边语义角色产生的不同变量统一为同一个变量. 根据这些定义, 这两个事件的翻译为:

周边语义角色标注的翻译: 是一个 (X_{11} , 人), 是一个 (X_{12} , 排球), 是一个 (X_{21} , 人).

中枢语义角色标注的翻译: 打 (X_{11} , X_{12}), 出汗 (X_{21}).

约束谓词的翻译: 等同 (X_{11} , X_{21}),

其中, X_{11} 和 X_{12} 分别表示头事件的第 1 个和第 2 个周边语义角色, X_{21} 是尾事件的第 1 个周边语义角色. 从上述的两个事件约束谓词可以得知, “施事 1 (对应变量 X_{11})”与“当事 2 (对应变量 X_{21})”指向同一个人, 将变量统一后, 事件“出汗”翻译为“是一个 (X_{11} , 人), 出汗 (X_{11})”. 最终, 事件三元组“< 打排球, 导致, 出汗>”的翻译结果为:

头事件的翻译: 是一个 (X_{11} , 人), 是一个 (X_{12} , 排球), 打 (X_{11} , X_{12}).

尾事件的翻译: 是一个 (X_{11} , 人), 出汗 (X_{11}).

(2) 推理规则转换的过程

当事件关系为因果、时序、条件、子事件关系和上下位时, 把事件中枢的翻译放于规则头, 其余内容放于规则体. 例如, 事件三元组“< 打排球, 导致, 出汗>”的翻译结果转换为推理规则形式为:

出汗 (X_{11}):- 是一个 (X_{11} , 人), 是一个 (X_{12} , 排球), 打 (X_{11} , X_{12}),

其中, “:-”表示推理规则的连接符号, 该符号左边的内容为规则头, 右边的内容为规则体.

当事件关系为目标类的关系时, 除了以上规则转换处理外, 还需要识别主体对应的语义角色, 并增加约束谓词“愿望”的翻译“愿望 (X_{ij} , 对应的中枢翻译)”. 例如, 事件三元组“< 是公司股民, 当事 1 有目标, 公司健康发展>”的语义角色标注为:

头事件标注: 当事: 人 是 公司股民/类事: 人.

事件关系: 当事 1 有目标.

尾事件标注: 公司/当事: 公司 健康发展.

该三元组对应的推理规则形式为:

愿望 (X_{11} , 健康发展 (X_{21})): 扮演 (X_{11} , 公司股民, 公司), 是一个 (X_{11} , 人), 是一个 (X_{21} , 公司),

其中, “扮演”是事件约束谓词, 该约束谓词的翻译形式为“扮演 (A_{ij} , $*W_{ij}$ 对应的事元, W_{ij} 对应的场合)”. A_{ij} 是事件成分对应的语义角色标注; “ $*W_{ij}$ 对应的事元标注”表示事件成分对应的具体值; “ W_{ij} 对应的场合”中所使用的场合是课题组总结的 342 个场合, 包括“学校”“家庭”和“公司”等.

通过上述的自动评估、人工评估和实际应用评估方法, 我们从多个维度对新三元组的质量及验证模型的性能进行了全面的评估, 从而确保了获取的三元组具有高质量和可靠性.

4.4 实验结果

本节将详细分析本方法的实验结果与对比实验. 主要包括基于设计的 ek-prompt 生成的事件三元组、使用微调模型生成的事件三元组、用于验证三元组正确性的微调模型性能、通过事件共享机制获取的事件三元组、扩展后 ECKG 的 4 种事件关系三元组的统计情况以及获取的新事件三元组用于推理应用.

4.4.1 分析 ek-prompt 生成结果

除了将 ek-prompt 应用于 GPT-4-Turbo 模型扩展 ECKG 外, 同时本文还与其他 3 个 LLM 在同一个测试集上进行了对比实验, 包括 Anthropic 的 Claude 3 模型^[65]、科大讯飞的星火认知大模型 V3.5^[66]和阿里的通义千问大模型^[67]. 使用 BERTScore 评价指标来衡量这些 LLM 生成事件的头三元组和尾三元组的性能, 图 4 展示了不同模型在 4 种事件生成头三元组和尾三元组的 P_{BERT} 、 R_{BERT} 和 F_{BERT} 结果. 从图 4 中我们观察到以下几点.

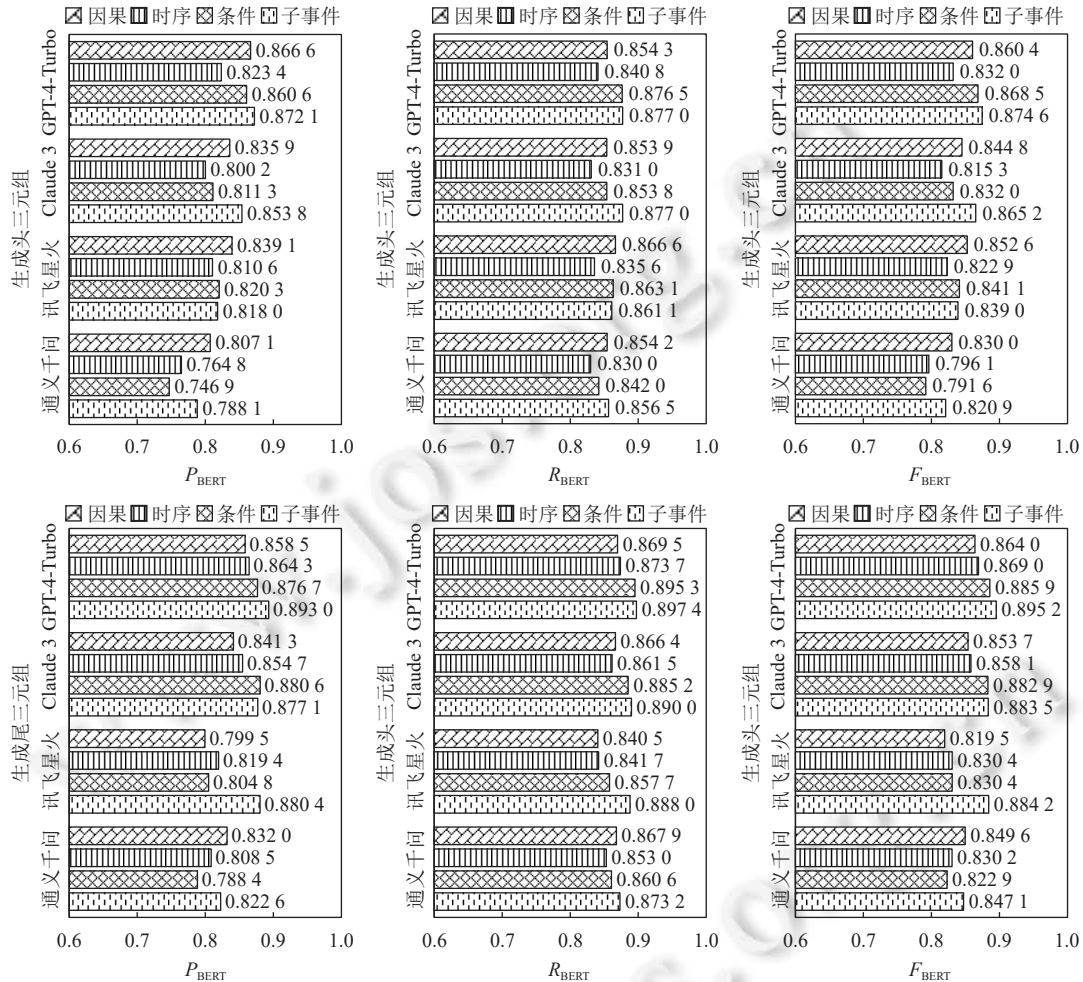


图 4 比较不同 LLM 使用 ek-prompt 生成事件头尾三元组的 BERTScore

(1) 在生成 4 种事件关系的头三元组方面, 所有 LLM 在生成子事件的头三元组和尾三元组上均表现较优和较稳定的性能, 而在处理时序关系时表现略差.

(2) 在所有 BERTScore 得分中, GPT-4-Turbo 模型整体上表现最为出色. 尤其是在扩展条件和子事件关系时, 生成子事件关系的尾三元组的 R_{BERT} 得到最高分 (0.8974). 体现了 GPT-4-Turbo 模型强大的语言理解能力, 能够生成高质量的事件三元组.

(3) Claude 3 模型在多个指标上紧随 GPT-4-Turbo 模型之后, 特别是在生成条件关系的尾三元组时展现出较好的性能, P_{BERT} 取得 0.8806 的得分, 领先于其他模型, 表明了 Claude 3 模型具有较强的事件关系理解能力.

(4) 讯飞星火在扩展 4 种事件关系的任务中, 生成因果关系的头三元组比尾三元组有更好的表现, 而生成子事件关系的尾三元组比头三元组得分更高, 显示了该模型在特定领域的优势.

(5) 通义千问模型扩展子事件关系的尾三元组表现还不错, R_{BERT} 取得了 0.8732 分。然而, 在生成时序和条件关系的三元组上的性能相对较差, 特别是生成条件关系的头三元组时 P_{BERT} 达到最低分 (0.7469)。

本文统计了图 4 中 4 个 LLM 在生成同种事件关系头尾三元组的平均 BERTScore 得分, 以综合评估每个模型生成各种事件关系的整体性能。平均 BERTScore 得分如图 5 所示。从图 5 中可以看出 GPT-4-Turbo 模型总体表现最佳, 在各种事件关系均展现出较为稳定的性能。4 种事件关系的平均 F_{BERT} 均超过 0.85, 尤其在子事件上表现最好, 平均 F_{BERT} 获得了 0.8849 分。Claude 3 模型在 4 种事件关系的平均 F_{BERT} 与 GPT-4-Turbo 模型和讯飞星火模型的较为接近, 但 Claude 3 模型的平均 P_{BERT} 略低于 GPT-4-Turbo 模型, 却优于讯飞星火模型。相比之下, 通义千问模型是 4 个模型中得分最低, 但在子事件的平均 R_{BERT} 表现还不错。

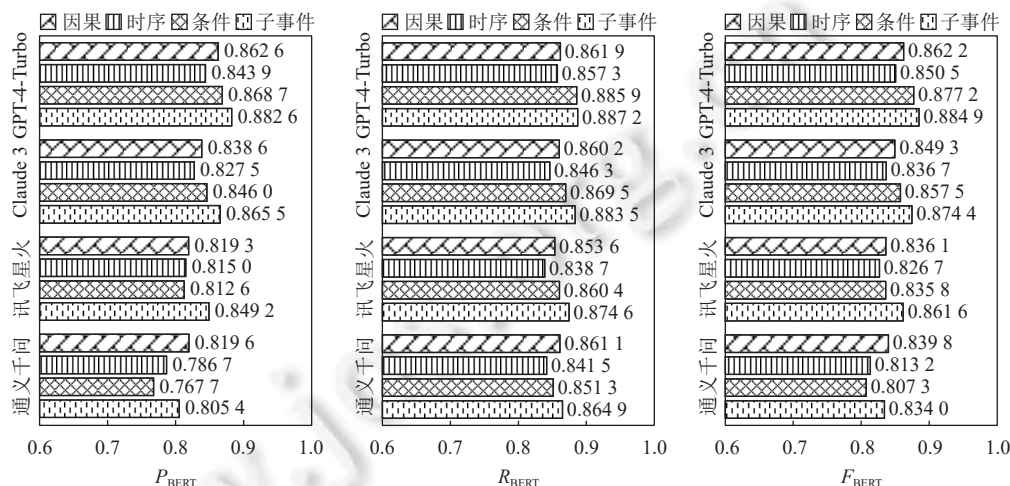


图 5 比较不同 LLM 使用 ek-prompt 生成事件头尾三元组的平均 BERTScore

除此以外, 我们 4 人评审团从常识知识的角度对各个 LLM 生成事件三元组进行准确地评估。在评估的过程中, 对于意见不一致的三元组进行二次审核, 以确保每个三元组的评价达到一致性。本文统计了不同 LLM 生成每个事件头尾三元组的准确率, 详细结果如图 6 所示。通过分析图 6 中的数据, 得出如下结论。

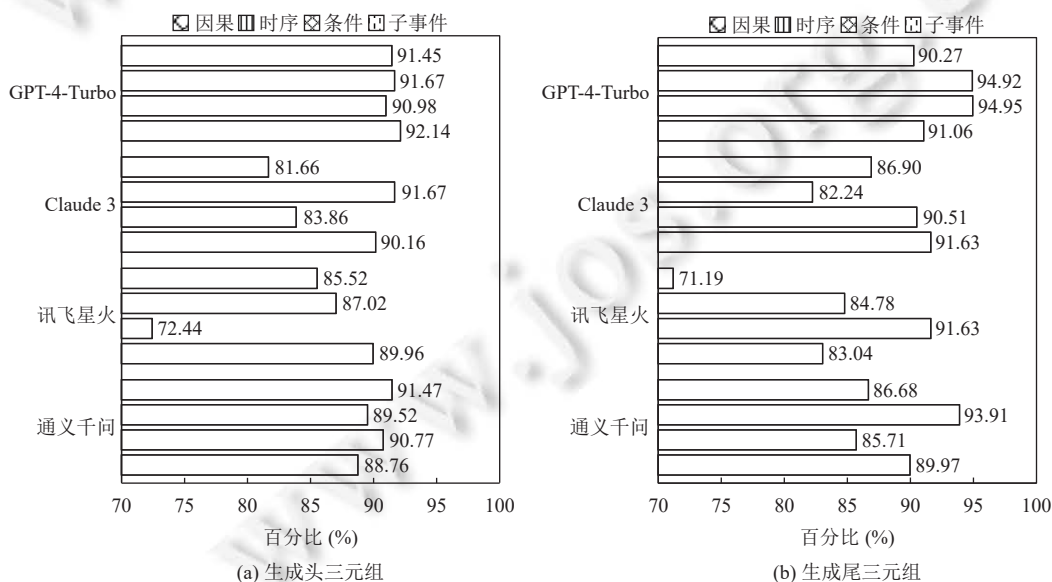


图 6 人工评估不同 LLM 使用 ek-prompt 生成事件头尾三元组的准确率

(1) GPT-4-Turbo 模型在所有模型中生成的事件三元组准确率最高, 并且每种关系的表现较为稳定, 准确率均超过 90%. 其中时序和条件关系的尾三元组准确率分别高达 94.92% 和 94.95%.

(2) Claude 3 模型在生成子事件关系的三元组表现较好且均衡, 头尾三元组的准确率都在 90% 以上. 与之相反的是, 在扩展其他 3 种关系时性能波动较大, 最为明显的是时序和条件关系. 如生成时序关系的头三元组准确率高达 91.67%, 而尾三元组的准确率仅为 82.24%.

(3) 讯飞星火模型在扩展 4 种事件关系的头尾三元组任务中表现均不平衡, 尤其是在生成条件关系的尾三元组是准确率较高, 但生成因果关系尾三元组的准确率在所有模型中最低 (71.19%).

(4) 通义千问模型扩展 4 种事件关系的头三元组准确率比尾三元组更为稳定, 在尾事件三元组中, 时序关系的准确率较高, 达到 93.91%, 但在扩展条件关系时表现略差.

通过将图 6 的人工评估各种事件关系头尾三元组的准确率进行平均处理, 结果如图 7 所示. 图 7 中清晰展现了 GPT-4-Turbo 模型在扩展所有事件关系的优越性能, 都获得了最高的平均准确率. 紧接着是通义千问模型表现还不错, 但略微不足的是在子事件关系时, 该模型的平均准确率略低于 Claude 3 模型. 然而, Claude 3 模型在其他 3 种事件关系的整体表现都不及通义千问模型. 相比之下, 讯飞星火模型的整体性能相对较差且不稳定, 尽管子事件关系取得较高的平均准确率 (86.5%), 但因果关系的平均准确率仅为 78.36%.

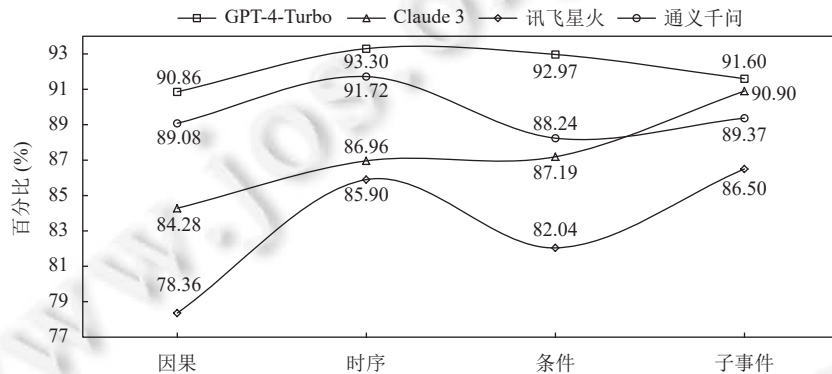


图 7 人工评估不同 LLM 使用 ek-prompt 生成事件头尾三元组的平均准确率

总体来说, 通过对比不同 LLM 在生成事件三元组的表现, 实验结果揭示了各个 LLM 在处理复杂事件关系的性能差异, 同时也突出了 GPT-4-Turbo 模型在理解和生成高精度事件三元组的卓越能力, 这对于扩展 ECKG 具有重要的价值.

4.4.2 评估生成三元组的微调模型性能

本文通过微调 GPT-3.5-Turbo 模型生成更多具体的事件三元组外, 还与未经过微调的 GPT-3.5-Turbo 模型和百度的文心一言模型 V4.0^[68]在相同的测试集上进行了详细的对比实验, 并使用了 P_{BERT} 、 R_{BERT} 和 F_{BERT} 对这 3 个 LLM 生成的事件三元组进行评估. 实验结果如图 8 所示, 图中的 ft-gen 表示经过微调的模型. 通过对图 8 的细心观察, 得到以下的结论.

(1) ft-gen 模型与未经过微调的 LLM 相比, ft-gen 模型在 BERTScore 所有的指标上有一定的提升, 并且在扩展 4 种事件关系的性能表现出较高的稳定性. 尤其在生成子事件的尾三元组中表现更为突出, R_{BERT} 取得了最高分 0.9129, F_{BERT} 为 0.909 5.

(2) 与 ft-gen 模型相比, GPT-3.5-Turbo 模型在扩展 4 种事件关系的头三元组表现较为不均衡. 在 F_{BERT} 得分中, 时序关系为 0.823, 而子事件关系达到 0.874. 但该模型在扩展尾事件三元组的整体表现相对稳定, 各种事件关系的得分均在 0.86 和 0.88 之间.

(3) 文心一言模型在每种事件关系尾事件三元组的 BERTScore 评价上较为平衡, 得分都在 0.85–0.88 之间. 然而, 在扩展头三元组任务中, 性能的稳定性有所下降. 例如, 在生成子事件头三元组时表现较好, 所有的 BERTScore

指标得分都超过 0.88 分。相反, 在条件关系时表现较差, P_{BERT} 仅为 0.796 8。

本文对图 8 中每种事件关系的头尾三元组在 BERTScore 各个指标上进行处理, 结果如图 9 所示。从图 9 中可以看出, 3 个 LLM 在扩展 ECKG 任务中都表现不错, F_{BERT} 得分都超过 0.84。具体来说, 微调后的模型表现最佳, 各项得分都高于原始 LLM, 尤其在子事件关系上, R_{BERT} 取的最高分, 达到 0.901 8。相比之下, 未经过微调的 GPT-3.5-Turbo 和文心一言模型表现较为相近, 但 GPT-3.5-Turbo 模型在扩展时序关系的得分更低。

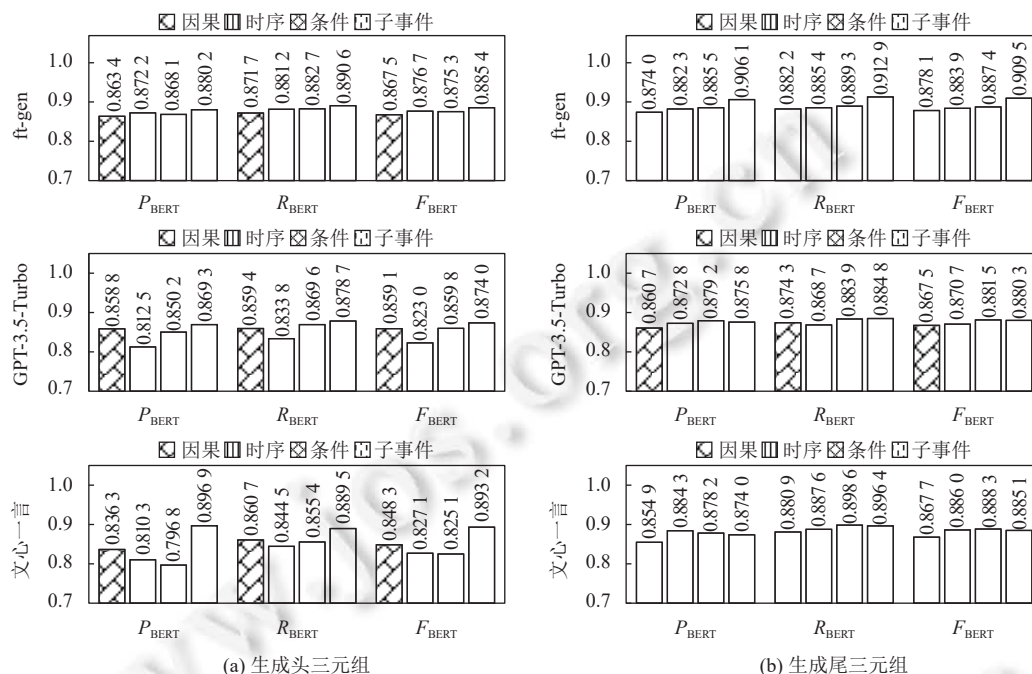


图 8 微调与原始 LLM 生成事件头尾三元组的 BERTScore

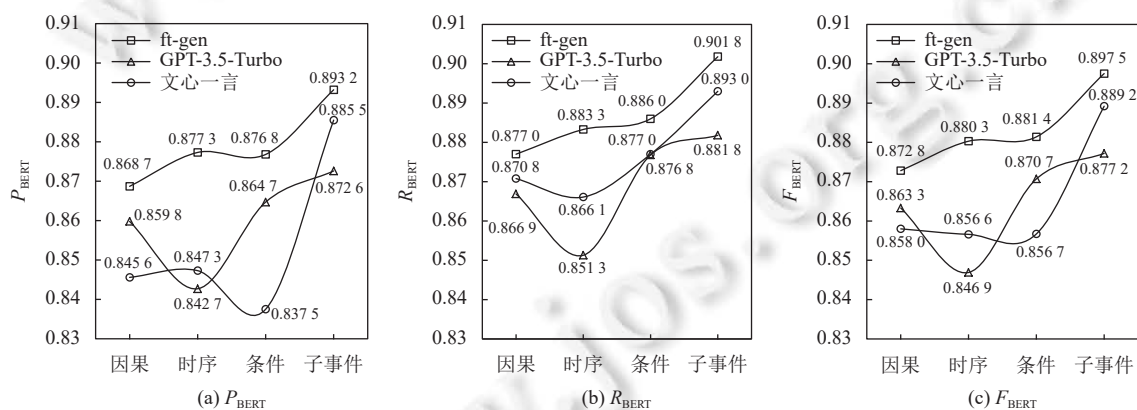


图 9 微调与原始 LLM 生成事件头尾三元组的平均 BERTScore

与此同时, 本文人工评估了微调后和原始的 LLM 生成的头尾三元组的准确率, 详细的实验结果在图 10 中展示。从图 10 中, 可以观察到以下信息。

- (1) 在扩展头尾三元组中, ft-gen 模型在所有事件关系下均获得最高的准确率, 且具有较好的平衡性。每种事件关系的准确率均超过 92%, 其中条件关系的头尾三元组的准确率尤为突出, 分别达到了 97.04% 和 97.46%。
- (2) GPT-3.5-Turbo 与文心一言模型相比, 在头三元组的表现中, GPT-3.5-Turbo 模型在因果关系上优于文心一

言模型,但在其他 3 种事件关系上表现略差.在尾三元组中,GPT-3.5-Turbo 模型只在条件关系的表现高于文心一言模型.值得注意的是,这两个模型的性能表现极为不平衡,特别是在增强尾三元组时,GPT-3.5-Turbo 模型在因果关系的准确率低至 82.99%,而在条件关系下又提高到 92.89%.同时,文心一言模型在时序关系的最高准确率达 92.59%,而条件关系最低,仅为 82.17%.

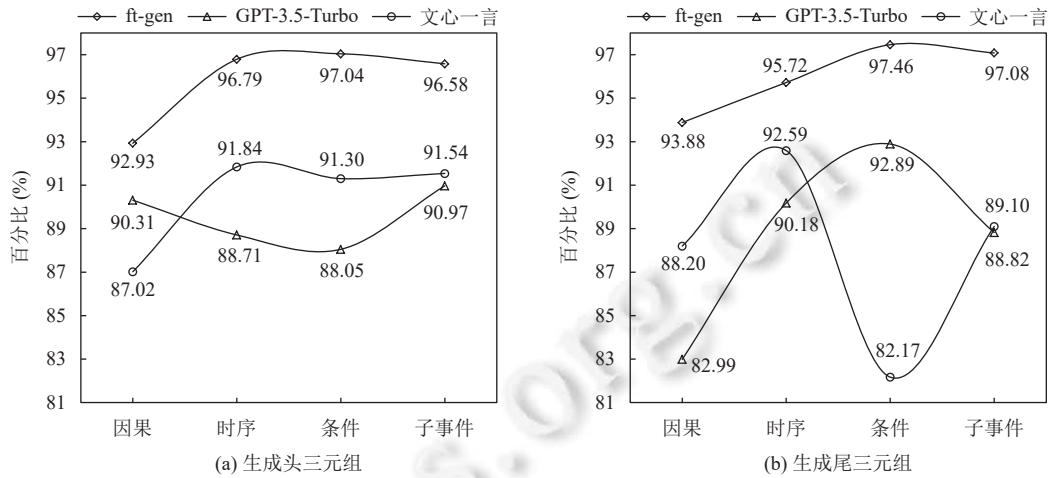


图 10 人工评估微调与原始 LLM 生成事件头尾三元组的准确率

此外,为了更直观地展示微调模型与原始 LLM 在不同事件关系上的性能差异,本文对图 10 中同种关系的头尾三元组的准确率进行平均处理,结果如图 11 所示.从图中可以看出,ft-gen 模型明显优于原始 LLM,特别是条件关系的头尾三元组平均准确率最高,达到了 97.25%,相比于 GPT-3.5-Turbo 和文心一言模型分别提升了 7.49% 和 12.12%.相比之下,GPT-3.5-Turbo 和文心一言模型性能较为落后,尤其是 GPT-3.5-Turbo 模型在因果关系下表现最为不佳.同时,文心一言模型在扩展不同事件关系下的平均准确率波动较大,虽然在时序和子事件关系下表现较好,但在因果和条件关系表现较差.

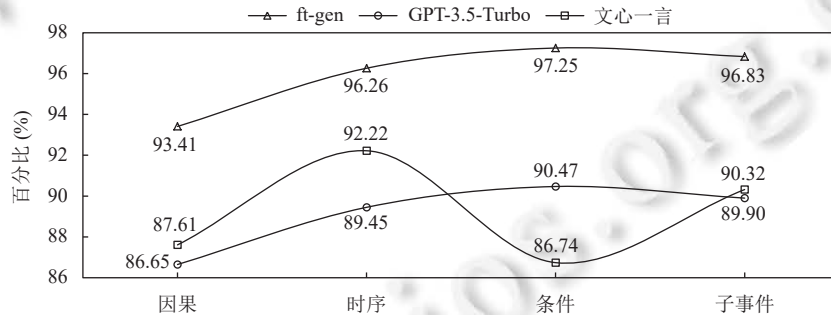


图 11 人工评估微调与原始 LLM 生成事件头尾三元组的平均准确率

从对微调与未经过微调的 LLM 的实验结果来看,在扩展 ECKG 的 4 种事件关系任务中,微调后的模型在生成每种事件的头尾三元组的性能上均有所提升,并且整体表现较为平衡.相比之下,未经微调的 LLM 性能较为不稳定,有些事件关系表现优异,但有些略显不足.然而,对于扩展 ECKG 任务,考虑到我们的目标是确保每个事件的头尾三元组都具有高精度,以保持整个知识图谱的高质量.因此,经过微调的模型显得更加适用.

4.4.3 评估验证三元组的微调模型性能

本文使用 **ek-prompt** 和微调后 LLM 生成的事件三元组,并在特定的测试集上进行了评估,验证了这两种方法生成三元组的有效性.为了进一步提升扩展后的 ECKG 三元组的精度,对所有获取的新事件三元组进行了一个全

面的评估. 由于数据规模庞大, 仅依靠人工审核的方法显然不现实. 因此, 本文采用了微调后的 GPT-3.5-Turbo 模型来验证三元组的正确性.

在验证任务中, 本文使用了准确率、精确率、召回率和 $F1$ 值来评估微调模型的性能. 此外, 还与未经过微调的 GPT-3.5-Turbo、GPT-4-Turbo、讯飞星火和文心一言模型进行了对比实验. 图 12 展示了实验结果, 其中 ft-ver 表示用于验证任务的微调模型. 通过对图 12 中实验数据的分析, 得到以下结论.

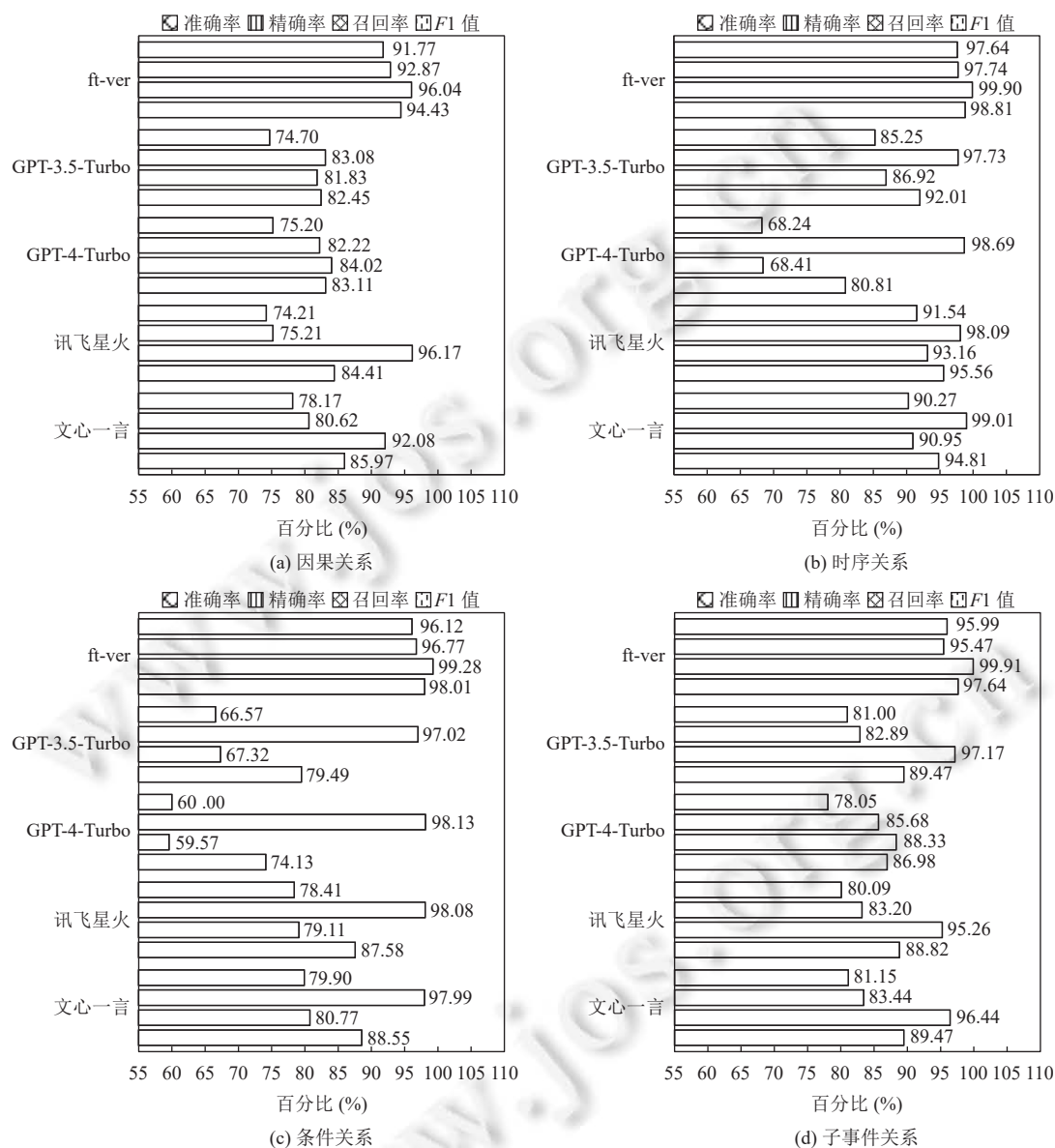


图 12 不同 LLM 验证各事件关系三元组的性能对比

(1) 在所有测试的 LLM 中, ft-ver 模型在验证 4 种不同事件关系的任务中表现出较好且稳定的性能. 特别是在时序、条件和子事件的召回率高达 99% 以上. 最低的准确率体现在因果关系中, 但仍然保持在 91% 以上, 也是相当不错的表现.

(2) GPT-3.5-Turbo 模型在时序和条件关系的精确率以及子事件关系的召回率表现优异, 均在 97% 以上. 然而,

其他的关系和指标均表现不佳,尤其是条件关系的准确率和召回率最低,分别只有 66.57% 和 67.32%.

(3) GPT-4-Turbo 模型在时序关系和条件关系的精确率表现较好,都超过 98%,但在准确率和召回率上表现较弱. 时序关系的准确率和召回率分别为 68.24% 和 68.41%,条件关系表现更为不足,准确率和召回率分别只有 60% 和 59.57%.

(4) 讯飞星火和文心一言模型在 4 种事件关系中表现相近,尤其在时序关系中表现最为出色,各项指标都超过 90%. 在因果关系中,这两个模型的召回率较高,均超过 92%,但其他 3 个指标表现一般,准确率最低为 74.21%. 在条件关系中,虽然精确率在 97% 以上,但其他指标均不足 89%. 在子事件关系中,召回率超过 95%,但其他指标相对较弱.

总体来说, ft-ver 模型能够深入学习事件三元组的特征,有效地验证新生成三元组的正确性,为评估大规模的因果、时序、条件和子事件关系的三元组提供了一种新的途径.

4.4.4 ECKG 事件相似度的实验结果

本文结合嵌入技术与多维度语义特征方法获取了 ECKG 事件的相似度,并对相似度结果经过阈值的筛选、去重等预处理后,最终得到 18 848 对相似事件. 此外,为了验证本文计算相似度方法的有效性,对没有经过语义角色标注的原始事件进行相似度计算.

在这两种方法中,本文首先筛选了相似度值在 [0.95, 1) 区间内的数据,并分为 5 个子区间进行分析,即 [0.95, 0.96), [0.96, 0.97), [0.97, 0.98), [0.98, 0.99) 和 [0.99, 1). 其次,在每个区间随机选择 200 个事件对进行人工评估正确性. 相关的结果如表 11 所示.

表 11 不同相似度计算方法在各相似度区间正确的事件对数量和准确率

方法	统计项	[0.95, 0.96)	[0.96, 0.97)	[0.97, 0.98)	[0.98, 0.99)	[0.99, 1)	[0.95, 1)
多维度事件语义角色相似度	正确的事件对数量	187	190	195	196	198	966
	准确率 (%)	93.5	95	97.5	98	99	96.6
原始事件相似度	正确的事件对数量	171	156	163	166	185	841
	准确率 (%)	85.5	78	81.5	83	92.5	84.1

从表 11 中的数据,可以总结出以下结论.

(1) 使用多维度特征计算的方法,相似事件的准确率随着相似度增加逐渐提高,在 [0.99, 1) 区间时准确率达到 99%,最低的准确率为 93.5%. 在 5 个相似度区间中,1 000 个事件对的平均准确率达到 96.6%.

(2) 相比多维度特征的计算方法,仅基于事件本身的计算方法准确率有所下降,最高准确率为 92.5%,最低准确率在 [0.96, 0.97) 区间,仅为 78%,整体平均准确率为 84.1%.

除了使用多维度语义方法和原始事件计算事件相似度之外,本文还采用了支持中文的抽象语义表示^[69]工具 HanLP^[70]来解析事件和句法依存树^[71]进行事件相似度的计算. 本文详细考察了这 4 种相似度计算方法的具体实例,并对比了不同方法计算得到的同一个事件的前 3 个最相似事件 (Top 3). 这些结果如表 12 所示,其中每个事件用顿号分隔. 关于表 12 的实例,分析如下.

(1) 在“招待客人”事件中,使用原始事件计算方法得到最相似事件是“邀请客人”,而通过事件抽象语义表示和句法依存树计算得出最相似的事件均为“想招待客人”. 本文提出的方法计算出“招待他人”为最相似事件. 根据本文所提及中枢语义角色是事件的核心成分,在判断事件相似度时,中枢语义角色扮演着极为重要的作用. 因而在“邀请客人”“想招待客人”和“招待他人”这 3 个相似事件中,与“招待客人”的中枢语义最为一致的是“招待他人”. 这体现了事件语义角色在提高事件相似度计算准确性中的重要性.

(2) 对于事件“对方爱自己”,本文方法计算出最相似的事件为“他人爱自己”,这两个事件的宾语一致. 相比之下,基于原始事件的计算结果为“爱不合适的人”,从常识的角度分析,事件“他人爱自己”与“对方爱自己”在情感的表达上相似度更高. 同时,通过抽象语义表示和句法依存树计算得到的前 3 个最相似事件也与“对方爱自己”事件较为相似. 这些结果证明了句子成分在事件相似度计算中的重要性.

表 12 不同方法计算事件相似度的示例

原始事件	基于原始事件 计算相似事件 Top 3	抽象语义表示 计算相似事件 Top 3	句法依存树 计算相似事件 Top 3	多维度特征方法 计算相似事件 Top 3
招待客人	邀请客人、 想招待客人、 迎接顾客	想招待客人、 招待他人、 邀请客人	想招待客人、 主人在餐厅宴请客人、 朋友接待客人	招待他人、 接待他人、 朋友接待客人
帮助他人	能帮助他人、 决定帮助他人、 支持他人	他人帮助、 能帮助他人、 他人帮助自己	他人帮助、 寻求他人帮助、 愿意帮助他人	帮助朋友、 帮助同学、 他人帮助自己
对方爱自己	爱不合适的人、 对方重视自己、 他人爱自己	对方忠于自己、 对方体贴自己、 对方父母喜欢自己	他人爱自己、 对方不欺骗自己、 他人像父母一样爱自己	他人爱自己、 爱他人、 他人喜欢自己
知道事实	知道他人的喜好、 知道真相、 发现真相	不知道事实、 不知道真相、 不知道信息	知道真相、 知道候选人、 不知道真相	知道真相、 知道信息、 知道秘密
把物品归还给他人	把物品借给他人、 他人把物品归还自己、 把物品卖给他人	他人把物品归还自己、 归还物品、 把物品卖给他人	把物品借给他人、 他人把物品归还自己、 把物品卖给他人	把失物归还给乘客、 把书籍归还给图书馆、 他人把物品归还自己
将杯子放在桌子上	杯子在桌子上、 杯子被放在桌子上、 把杯子放在桌子上	杯子被放在桌子上、 把杯子放在桌子上、 杯子在桌子上	把杯子放在桌子上、 杯子被放在桌子上、 杯子在桌子上	把杯子放在桌子上、 把鱼饵放在鱼钩上、 把手放在手刹上
不吃汉堡包	吃汉堡包、 吃太多汉堡包、 没有汉堡包	吃汉堡包、 不吃食物、 想吃汉堡包	不吃薯条、 不吃食物、 想吃汉堡包	不吃食物、 不吃薯条、 不吃垃圾食物
冰箱里没有食物	冰箱有食物、 冰箱装满食物、 食品储藏室里没有食物	冰箱有食物、 胃里没有食物、 没有食物	食品储藏室里没有食物、 没有食物、 胃里没有食物	食品储藏室里没有食物、 胃里没有食物、 没有食物

(3) 对于事件“把物品归还给他人”，本文方法得到的前 3 个事件都包含“把...归还...”的结构。而另外 3 种方法在各自的前 3 个最相似事件中仅有一个事件包含该结构，剩余的事件都是其他类型的结构。例如，与“把物品借给他人”事件相比，“把...归还...”和“把...借给...”在表达的含义上有显著区别，反映出事件包含的结构不相同导致相似度更低。这表明了事件结构在相似度计算中的重要影响。

(4) 通过引入事件的反义特征，能够更准确地计算事件相似度。例如，“不吃汉堡包”与“不吃食物”的相似度明显高于“吃汉堡包”的相似度。“冰箱里没有食物”与“食品储藏室里没有食物”的相似度高于“冰箱有食物”的相似度。这些结果表明了使用原始事件和抽象语义表示计算事件相似度的结果较差，而句法依存树和本文方法计算结果较好，反映了句法依存树在处理反义词方面表现还不错。同时，这也证明了本文引入反义特征有效地改进了其他两种计算事件相似度方法的局限。

通过分析相似度结果，证实了基于语义角色的多维度事件相似度计算方法的有效性。该方法能够从多方面精确地捕捉了事件的深层语义，从而显著提高相似度计算的可靠性。因此，这一方法增强了利用事件相似度进行事件共享来扩展 ECKG 的实用性。

4.4.5 ECKG 事件三元组扩展结果

本文将 3 个扩展 ECKG 方法获得的常识知识进行剔除无关文本和符号，以及将事件三元组经过相似度去重等预处理后，每种方法得到事件三元组的数量如表 13 所示。

为了确保扩展后的 ECKG 具有高质量，本文对表 13 中的所有事件三元组进行验证。验证过程包括两个主要步骤：首先，使用微调过的 LLM 进行初次验证。其次，基于 LLM 验证的结果，随机抽取不同方法生成的每种事件关系下的 500 个三元组，进行了人工审核的二次验证。表 14 展示了验证结果的准确率，通过这些数据可以得出以下结论。

(1) 对于使用 ek-prompt 和微调模型生成的事件三元组，微调模型验证 4 种事件关系三元组的准确率均超过 93%。其中，时序关系的三元组准确率最高，达到了 99.9%。这一结果不仅说明了本文设计的 ek-prompt 有效地引导

模型生成常识事件三元组, 还显示了专门为生成更为具体的事件三元组微调的模型, 充分学习到了训练数据中的特征, 能够生成了高质量的事件三元组.

表 13 不同方法获取的三元组数量

方法	事件关系	三元组数量	总计	三元组总数量
ek-prompt	因果	226 845	798 838	
	时序	187 192		
	条件	212 103		
	子事件	172 698		
微调模型生成	因果	234 609	783 091	2 806 155
	时序	163 253		
	条件	211 086		
	子事件	174 143		
事件共享机制	因果	395 495	1 224 226	
	时序	265 596		
	条件	316 446		
	子事件	246 689		

表 14 微调模型与人工验证事件三元组的准确率 (%)

方法	事件关系	微调模型验证准确率	人工验证准确率
ek-prompt	因果	93.04	96.2
	时序	99.9	97.8
	条件	98.21	97.6
	子事件	96.08	96.6
微调模型生成	因果	95.76	96.8
	时序	99.88	98
	条件	97.44	97.8
	子事件	93.72	97.4
事件共享机制	因果	82.94	95.6
	时序	93.8	98.2
	条件	84.75	96
	子事件	86.25	97.8

(2) 通过事件共享机制获取的新三元组, 虽然准确率略低于基于 ek-prompt 和微调模型生成的三元组, 但整体表现仍然不错, 特别是在时序关系中, 准确率达到到了 93.8%. 这体现了事件共享机制扩展 ECKG 的可行性, 确保了相似事件三元组信息一致性, 并获取了一定数量的事件三元组.

(3) 人工抽样审核的结果显示, 3 种扩展方法中的 4 种事件关系的准确率均在 95% 以上. 这表明了本文微调的 LLM 用于验证事件三元组的有效性和可靠性, 证明它可以作为验证大规模事件三元组正确性的有力工具.

总体而言, 上述的验证结果充分证明了本文方法在获取事件三元组方面具有高精度和可靠性, 有效地支持了扩展 ECKG 的工作.

基于事件三元组正确性的验证结果, 本文统计了微调模型评估为正确的三元组数量, 具体的数据见表 15. 从表 15 可以看出, 在 3 种扩展 ECKG 的方法中, 利用 ek-prompt 和微调模型生成的事件三元组, 评估为正确的总三元组数量较为接近. 值得一提的是, 通过引入事件共享机制获取的正确三元组数量最多. 这一现象表明了事件共享机制扩展 ECKG 的可行性和有效性, 同时也促进了相似事件三元组的一致性.

与此同时, 本文对表 15 中的所有事件三元组进行了合并和去重处理, 从而获得最终增补到 ECKG 的总三元组数量, 如表 16 所示. 这些数据表明, 本文方法有效地扩大了 ECKG 的规模并丰富了其知识库, 能够为使用 ECKG 进行更广泛的应用和研究提供了可靠的常识知识资源.

表 15 不同方法获取的正确三元组数量

方法	事件关系	三元组数量	总计	三元组总数量
ek-prompt	因果	211 056	772 295	
	时序	187 005		
	条件	208 306		
	子事件	165 928		
微调模型生成	因果	224 661	756 607	2 587 011
	时序	163 057		
	条件	205 682		
	子事件	163 207		
事件共享机制	因果	328 023	1 058 109	
	时序	249 129		
	条件	268 188		
	子事件	212 769		

表 16 增补到 ECKG 的三元组数量

事件关系	三元组数量
因果	666 193
时序	592 691
条件	653 605
子事件	520 523
总计	2 433 012

4.4.6 扩展后 ECKG 的推理应用

扩展后 ECKG 的事件三元组通过翻译和规则转换为常识知识库, 并使用 SWI-Prolog 逻辑编程语言实现单个知识库的推理和多个知识库的推理. 推理过程主要包含 4 个关键部分: “事实”(从 ECKG 获取的事件三元组信息)、“规则”(事件三元组翻译结果转换为 SWI-Prolog 形式的规则)、“table”(SWI-Prolog 内置谓词)以及查询(从知识库中查询并获得推理结果). 下面将介绍推理过程.

(1) 单个知识库的推理

单知识库的推理是指在一个独立的知识库中进行推理. 例如, 在因果知识库中, 已知“张三因为工资低导致罢工”的场景. 使用已知的因果事实、规则和内置谓词“table”, 通过输入“导致(工资低, Y)”进行推理. 这个推理可以确定 Y 的实体类型为“事件”, 具体的值为“罢工”(如图 13 所示).

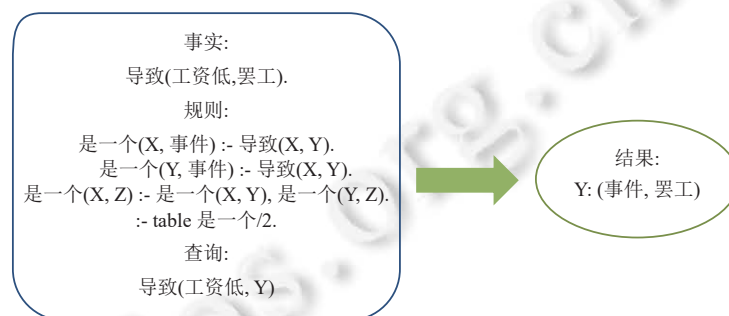


图 13 因果知识库的推理

上下位知识库推理场景: 已知苹果是一种含维生素 C 的果类, 维生素 C 的果类属于水果, 那么可以得出什么结论? 推理的结果如图 14 所示, 根据上下位知识的事实和推理规则, 最后可以推理得出苹果是一个实体、是一种含维生素 C 的果类, 也是一种水果.

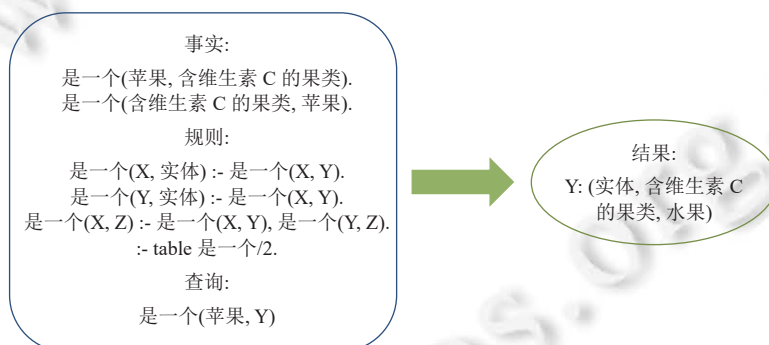


图 14 上下位知识库的推理

(2) 多个知识库的推理

多知识库的推理通过整合不同的知识库进行协同推理, 有效解决了单一知识库因缺失信息无法完成的推理任务, 从而支持更复杂的推理. 图 15 展示了上下位知识库与目标知识库的推理结果. 在该图中, 已知相关事实为“两个主体是家人关系”, 但是主体的具体实体类型未知. 因此, 首先将场景信息提供给上下位知识库, 在这个知识库中得出主体对应的实体类型是“人”. 接着, 目标知识库从上下位知识库获得实体类型后, 进一步推理出“李红希望张姨能够康复”的结论. 这种通过不同知识库联合推理的方法, 极大地增强了推理的准确性和适用性.

通过上述常识知识库的推理结果说明, 扩展后 ECKG 的事件三元组通过知识翻译可以进行推理任务, 有效验证了 ECKG 在实际应用中的实用性.

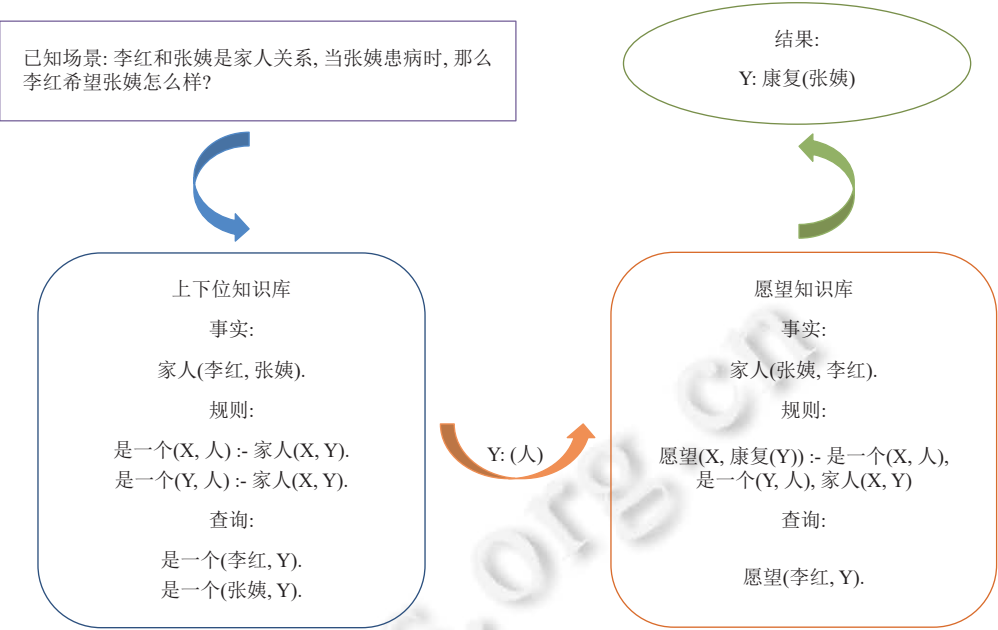


图 15 多知识库推理

5 总 结

在研究初期, 课题组成员投入了大量的人力资源构建了一个高精度的种子事件常识知识图谱 (ECKG). 由于其规模较小, 在实际应用中发挥的作用受限. 在现有研究中, 以事件为中心的大规模中文常识知识图谱相对较少. 此外, 常识知识通常不会在自然语言中直接表述, 更多地反映在人们的日常生活和潜意识中. 针对这些问题, 本文提出了一种基于 LLM 扩展 ECKG 的方法. 本文主要是通过利用 GPT 系列 LLM 的强大数据处理和知识生成能力, 获取更多关于 ECKG 的事件三元组, 以扩大其规模.

具体来说, 本文首先设计了 4 种事件关系的高质量 **ek-prompt**, 并调用了 GPT-4-Turbo 模型初步生成事件三元组. 其次, 通过特定的三元组数据集对 GPT-3.5-Turbo 模型进行微调, 并使用微调后的模型进一步生成和验证新的事件三元组. 特别地, 为了解决 LLM 生成的相似事件三元组中存在不一致性的问题, 本文基于语义角色多维度特征获得了精确的事件相似度, 并引入了事件共享机制. 这一策略极大地扩展了 ECKG 的规模, 同时也提高了相似事件三元组的一致性.

通过实验证明, 本文的方法能够高效地获取大规模的常识事件三元组, 并且新获取的三元组经过自动化评估、人工评估以及推理应用评估, 证实了具有高质量和实用性. 这不仅保持了 ECKG 的高精度特性, 还显著扩大了其规模, 丰富了常识知识库.

在未来, 仍有很多有意义的工作值得我们去探索. 首先, 计划对 ECKG 进行更多层次的自动扩展, 以进一步扩大常识知识图谱. 其次, 将实现基于人类反馈的增强学习 (reinforcement learning from human feedback, RLHF) 方法^[72], 通过广泛收集用户的反馈来设计奖励机制以优化现有的 LLM 学习能力, 提升事件三元组生成和验证的准确性及有效性. 最后, 为了验证和提升 ECKG 的可用性和应用价值, 计划探索多方面的应用, 包括开发一个常识知识的自动问答系统、事件预测以及大规模常识知识库的常识推理等应用. 通过这些未来研究, 期望能够显著提升 ECKG 的质量与应用范围, 使其在自然语言处理和人工智能领域发挥更大的作用和价值.

References:

[1] Nguyen TP, Razniewski S, Weikum G. Advanced semantics for commonsense knowledge extraction. In: Proc. of the 30th Web Conf.

- Ljubljana: ACM, 2021. 2636–2647. [doi: [10.1145/3442381.3449827](https://doi.org/10.1145/3442381.3449827)]
- [2] Zellers R, Bisk Y, Farhadi A, Choi Y. From recognition to cognition: Visual commonsense reasoning. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 6713–6724. [doi: [10.1109/CVPR.2019.00688](https://doi.org/10.1109/CVPR.2019.00688)]
 - [3] Vijayaraghavan P, Roy D. Modeling human motives and emotions from personal narratives using external knowledge and entity tracking. In: Proc. of the 30th Web Conf. Ljubljana: ACM, 2021. 529–540. [doi: [10.1145/3442381.3449997](https://doi.org/10.1145/3442381.3449997)]
 - [4] Suchanek FM, Kasneci G, Weikum G. YAGO: A core of semantic knowledge. In: Proc. of the 16th Int'l Conf. on World Wide Web. 2007. 697–706. [doi: [10.1145/1242572.1242667](https://doi.org/10.1145/1242572.1242667)]
 - [5] Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z. DBpedia: A nucleus for a Web of open data. In: Proc. of the 6th Int'l Semantic Web Conf., the 2nd Asian Semantic Web Conf. Busan: Springer, 2007. 722–735. [doi: [10.1007/978-3-540-76298-0_52](https://doi.org/10.1007/978-3-540-76298-0_52)]
 - [6] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase. Communications of the ACM, 2014, 57(10): 78–85. [doi: [10.1145/2629489](https://doi.org/10.1145/2629489)]
 - [7] Glavaš G, Šnajder J. Event graphs for information retrieval and multi-document summarization. Expert Systems with Applications, 2014, 41(15): 6904–6916. [doi: [10.1016/j.eswa.2014.04.004](https://doi.org/10.1016/j.eswa.2014.04.004)]
 - [8] Zhang HM, Liu X, Pan HJ, Song YQ, Leung CWK. ASER: A large-scale eventuality knowledge graph. In: Proc. of the 29th Web Conf. Taipei: ACM, 2020. 201–211. [doi: [10.1145/3366423.3380107](https://doi.org/10.1145/3366423.3380107)]
 - [9] Ding X, Li Z, Liu T, Liao K. ELG: An event logic graph. arXiv:1907.08015, 2019.
 - [10] Sap M, Le Bras R, Allaway E, Bhagavatula C, Lourie N, Rashkin H, Roof B, Smith NA, Choi Y. ATOMIC: An atlas of machine commonsense for if-then reasoning. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence. Honolulu: AAAI, 2019. 3027–3035. [doi: [10.1609/aaai.v33i01.33013027](https://doi.org/10.1609/aaai.v33i01.33013027)]
 - [11] Liu H, Singh P. ConceptNet—A practical commonsense reasoning tool-kit. BT Technology Journal, 2004, 22(4): 211–226. [doi: [10.1023/B:BTJ.0000047600.45421.6d](https://doi.org/10.1023/B:BTJ.0000047600.45421.6d)]
 - [12] Wang Y, Cao CG, Cao YT, Wang S. A property-based method for acquiring commonsense knowledge. In: Proc. of the 14th Int'l Conf. on KSEM Science, Engineering and Management. Tokyo: Springer, 2021. 53–65. [doi: [10.1007/978-3-030-82136-4_5](https://doi.org/10.1007/978-3-030-82136-4_5)]
 - [13] Wang Y, Cao CG, Chen ZW, Wang S. ECCKG: An eventuality-centric commonsense knowledge graph. In: Proc. of the 15th Int'l Conf. on Knowledge Science, Engineering and Management. Singapore: Springer, 2022. 568–584. [doi: [10.1007/978-3-031-10983-6_44](https://doi.org/10.1007/978-3-031-10983-6_44)]
 - [14] Song H. Research on Chinese semantic parsing and knowledge acquisition method based on semantic roles [Ph.D. Thesis]. Beijing: University of Chinese Academy of Sciences, 2022 (in Chinese with English abstract).
 - [15] Wang Y, Cao CG, Wang S. Commonsense knowledge acquisition via semantic roles in events. Journal of Chinese Information Processing, 2023, 37(6): 77–88 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2023.06.008](https://doi.org/10.3969/j.issn.1003-0077.2023.06.008)]
 - [16] Lenat DB. CYC: A large-scale investment in knowledge infrastructure. Communications of the ACM, 1995, 38(11): 33–38. [doi: [10.1145/219717.219745](https://doi.org/10.1145/219717.219745)]
 - [17] Fellbaum C, Miller G. WordNet: An Electronic Lexical Database. Cambridge: MIT Press, 1998.
 - [18] Singh P, Lin T, Mueller ET, Lim G, Perkins T, Zhu WL. Open mind common sense: Knowledge acquisition from the general public. In: Meersman R, Tari Z, eds. On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE: Confederated Int'l Conf. CoopIS, DOA, and ODBASE 2002 Proc. Berlin: Springer, 2002. 1223–1237. [doi: [10.1007/3-540-36124-3_77](https://doi.org/10.1007/3-540-36124-3_77)]
 - [19] Zhang HM, Khashabi D, Song YQ, Roth D. TransOMCS: From linguistic graphs to commonsense knowledge. In: Proc. of the 29th Int'l Joint Conf. on Artificial Intelligence. Yokohama, 2021. 554.
 - [20] Romero J, Razniewski S, Pal K, Pan JZ, Sakhadeo A, Weikum G. Commonsense properties from query logs and question answering forums. In: Proc. of the 28th ACM Int'l Conf. on Information and Knowledge Management. Beijing: ACM, 2019. 1411–1420. [doi: [10.1145/3357384.3357955](https://doi.org/10.1145/3357384.3357955)]
 - [21] Cao YN. Research on causal knowledge acquisition [Ph.D. Thesis]. Beijing: University of Chinese Academy of Sciences, 2012 (in Chinese with English abstract).
 - [22] Bosselut A, Rashkin H, Sap M, Malaviya C, Celikyilmaz A, Choi Y. COMET: Commonsense Transformers for automatic knowledge graph construction. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 4762–4779.
 - [23] Tandon N, De Melo G, Suchanek F, Weikum G. WebChild: Harvesting and organizing commonsense knowledge from the Web. In: Proc. of the 7th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM, 2014. 523–532. [doi: [10.1145/2556195.2556245](https://doi.org/10.1145/2556195.2556245)]
 - [24] Chklovski T. Learner: A system for acquiring commonsense knowledge by analogy. In: Proc. of the 2nd Int'l Conf. on Knowledge Capture. Sanibel Island: ACM, 2003. 4–12. [doi: [10.1145/945645.945650](https://doi.org/10.1145/945645.945650)]
 - [25] Cao C, Cao CG, Zang LJ, Wang S. An interactive method for acquiring event-based commonsense knowledge. Journal of Chinese

- Information Processing, 2016, 30(3): 125–132 (in Chinese with English abstract).
- [26] Wang Y. Research on commonsense knowledge acquisition [Ph.D. Thesis]. Beijing: University of Chinese Academy of Sciences, 2022 (in Chinese with English abstract).
 - [27] Deng P. A study of event graph completion and reasoning based on semantic enhancement [MS. Thesis]. Beijing: Institute of Computer Technology, Chinese Academy of Sciences, 2023 (in Chinese with English abstract).
 - [28] Meilicke C, Fink M, Wang YJ, Ruffinelli D, Gemulla R, Stuckenschmidt H. Fine-grained evaluation of rule- and embedding-based systems for knowledge graph completion. In: Proc. of the 17th Int'l Semantic Web Conf. Monterey: Springer, 2018. 3–20. [doi: [10.1007/978-3-030-00671-6_1](https://doi.org/10.1007/978-3-030-00671-6_1)]
 - [29] Li X, Taheri A, Tu LF, Gimpel K. Commonsense knowledge base completion. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACM, 2016. 1445–1455. [doi: [10.18653/v1/P16-1137](https://doi.org/10.18653/v1/P16-1137)]
 - [30] Malaviya C, Bhagavatula C, Bosselut A, Choi Y. Commonsense knowledge base completion with structural and semantic context. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 2925–2933. [doi: [10.1609/aaai.v34i03.5684](https://doi.org/10.1609/aaai.v34i03.5684)]
 - [31] Niu GL, Li B. Logic and commonsense-guided temporal knowledge graph completion. In: Proc. of the 37th AAAI Conf. on Artificial Intelligence. Washington: AAAI, 2023. 4569–4577. [doi: [10.1609/aaai.v37i4.25579](https://doi.org/10.1609/aaai.v37i4.25579)]
 - [32] Wang B, Wang GT, Huang J, You JX, Leskovec J, Jay Kuo CC. Inductive learning on commonsense knowledge graph completion. In: Proc. of the 2021 Int'l Joint Conf. on Neural Networks. Shenzhen: IEEE, 2021. 1–8. [doi: [10.1109/IJCNN52387.2021.9534355](https://doi.org/10.1109/IJCNN52387.2021.9534355)]
 - [33] Bosselut A, Le Bras R, Choi Y. Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. AAAI, 2021. 4923–4931. [doi: [10.1609/aaai.v35i6.16625](https://doi.org/10.1609/aaai.v35i6.16625)]
 - [34] Zhang NY, Xie X, Chen X, Deng SM, Ye HB, Chen HJ. Knowledge collaborative fine-tuning for low-resource knowledge graph completion. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3531–3545 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6628.htm> [doi: [10.13328/j.cnki.jos.006628](https://doi.org/10.13328/j.cnki.jos.006628)]
 - [35] Zhang YC, Chen Z, Guo LB, Xu YJ, Zhang W, Chen HJ. Making large language models perform better in knowledge graph completion. arXiv:2310.06671, 2023.
 - [36] Wei YB, Huang QS, Zhang Y, Kwok JT. KICGPT: Large language model with knowledge in context for knowledge graph completion. In: Proc. of the 2024 Findings of the Association for Computational Linguistics. Singapore: ACL, 2024. 8667–8683. [doi: [10.18653/v1/2023.findings-emnlp.580](https://doi.org/10.18653/v1/2023.findings-emnlp.580)]
 - [37] Luo RL, Gu TL, Li HL, Li JZ, Lin ZC, Li JY, Yang YJ. Chain of history: Learning and forecasting with LLMs for temporal knowledge graph completion. arXiv:2401.06072, 2024.
 - [38] Yao L, Peng JZ, Mao CS, Luo Y. Exploring large language models for knowledge graph completion. arXiv:2308.13916, 2023.
 - [39] Pan SR, Luo LH, Wang YF, Chen C, Wang JP, Wu XD. Unifying large language models and knowledge graphs: A roadmap. IEEE Trans. on Knowledge and Data Engineering, 2024, 36(7): 3580–3599. [doi: [10.1109/TKDE.2024.3352100](https://doi.org/10.1109/TKDE.2024.3352100)]
 - [40] Li DW, Tan Z, Chen TL, Liu H. Contextualization distillation from large language model for knowledge graph completion. In: Proc. of the 2024 Findings of the Association for Computational Linguistics. St. Julian's: ACL, 2024. 458–477.
 - [41] OpenAI. GPT-4 technical report. arXiv:2303.08774, 2023.
 - [42] Liu YH, Ott M, Goyal N, Du JF, Joshi M, Chen DQ, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692, 2019.
 - [43] Galárraga LA, Teflioudi C, Hose K, Suchanek F. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases. In: Proc. of the 22nd Int'l Conf. on World Wide Web. Rio de Janeiro: ACM, 2013. 413–422. [doi: [10.1145/2488388.2488425](https://doi.org/10.1145/2488388.2488425)]
 - [44] Wang Y, Cao CG. Research on categorization of events based on event attributes. Journal of Chinese Information Processing, 2020, 34(10): 39–50 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2020.10.006](https://doi.org/10.3969/j.issn.1003-0077.2020.10.006)]
 - [45] Song H, Cao CG, Wang Y, Wang S. A fine-grained annotated dataset for Chinese semantic-role labeling. Journal of Chinese Information Processing, 2023, 37(1): 16–32 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2023.01.002](https://doi.org/10.3969/j.issn.1003-0077.2023.01.002)]
 - [46] Cortes C, Vapnik V. Support-vector networks. Machine Learning, 1995, 20(3): 273–297. [doi: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018)]
 - [47] OpenAI. Introduction to GPT-4 and GPT-4-Turbo. 2024. <https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>
 - [48] Song H, Cao CG, Wang Y, Wang S. Construction of a finely-grained training dataset for Chinese semantic-role labeling. Journal of Chinese Information Processing, 2022, 36(12): 52–66, 73 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2022.12.006](https://doi.org/10.3969/j.issn.1003-0077.2022.12.006)]
 - [49] Lu C. Linguistics for Knowledge Engineering. Beijing: Tsinghua University Press, 2010 (in Chinese).
 - [50] Wang Y. Research on common sense knowledge acquisition methods based on semantic classification [MS. Thesis]. Guilin: Guangxi Normal University, 2015 (in Chinese with English abstract).
 - [51] Zhu XY. Studies on Semantic Structure Patterns of Sentences in Modern Chinese. Beijing: Peking University Press, 2001 (in Chinese).

- [52] Yuan YL. Research on Chinese Computational Linguistics Based on Cognition. Beijing: Peking University Press, 2008 (in Chinese).
- [53] Liu MF, Hu HJ. Research on Event Semantics Based on Cognition and Computation. Beijing: Science Press, 2013 (in Chinese).
- [54] Zang LJ, Wang WM, Wang Y, Fang F, Cao C, Wu XL, Zhou Q, Lu Y, Li TY, Cao CG. A Chinese framework of semantic taxonomy and description: Preliminary experimental evaluation using web information extraction. In: Proc. of the 8th Int'l Conf. on Knowledge Science, Engineering and Management. Chongqing: Springer, 2015. 275–286. [doi: 10.1007/978-3-319-25159-2_25]
- [55] Bai XP. Adding global semantic roles to the Chinese proposition bank to facilitate Chinese language studies. Linguistic Sciences, 2017, 16(5): 481–492 (in Chinese with English abstract). [doi: 10.7509/j.linsci.201705.031178]
- [56] Wang CW, Qian QQ, Xun ED, Xing D, Li M, Rao GQ. Construction of semantic role bank for Chinese verbs from the perspective of ternary collocation. Journal of Chinese Information Processing, 2020, 34(9): 19–27 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2020.09.004]
- [57] Wang LJ. Research on Chinese semantic dependency analysis [MS. Thesis]. Harbin: Harbin Institute of Technology, 2010 (in Chinese with English abstract).
- [58] SUMO. 2023. <http://www.ontologyportal.org>
- [59] Song H, Wang S, Liu Y, Wang Y. Predicate-attention neural model for Chinese semantic role labeling. Computers and Electrical Engineering, 2022, 99: 107741. [doi: 10.1016/j.compeleceng.2022.107741]
- [60] Chen Z W. A Study of event graph construction method based on semantic parsing [MS. Thesis]. Beijing: Institute of Computer Technology, Chinese Academy of Sciences, 2023 (in Chinese with English abstract).
- [61] Zhang TY, Kishore V, Wu F, Weinberger KQ, Artzi Y. BERTScore: Evaluating text generation with BERT. In: Proc. of the 8th Int'l Conf. on Learning Representations. Addis Ababa: OpenReview.net, 2020.
- [62] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: ACL, 2019. 4171–4186. [doi: 10.18653/v1/N19-1423]
- [63] Alpaydm E. Introduction to Machine Learning. 4th ed., Cambridge: MIT Press, 2020. 2–12.
- [64] Wielemaker J, Schrijvers T, Triska M, Lager T. SWI-Prolog. Theory and Practice of Logic Programming, 2012, 12(1–2): 67–96. [doi: 10.1017/S147106841000494]
- [65] Anthropic. 2024. <https://www.anthropic.com/claude>
- [66] iFLYTEK. 2024. <https://www.xfyun.cn/doc/spark/Web.html>
- [67] Alibaba Cloud. 2024. <https://tongyi.aliyun.com/qianwen>
- [68] Baidu. 2024. <https://yiyan.baidu.com/>
- [69] Banarescu L, Bonial C, Cai S, Georgescu M, Griffitt K, Hermjakob U, Knight K, Koehn P, Palmer M, Schneider N. Abstract meaning representation for sembanking. In: Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse. Sofia: ACL, 2013. 178–186.
- [70] He H, Choi JD. The stem cell hypothesis: Dilemma behind multi-task learning with Transformer encoders. arXiv:2109.06939, 2021.
- [71] Nivre J. Dependency grammar and dependency parsing. MSI Report, 2005, 5133(1959): 1–32.
- [72] Chaudhari S, Aggarwal P, Murahari V, Rajpurohit T, Kalyan A, Narasimhan K, Deshpande A, da Silva BC. RLHF Deciphered: A critical analysis of reinforcement learning from human feedback for LLMs. arXiv:2404.08555, 2024.

附中文参考文献:

- [14] 宋衡. 基于语义角色的中文语义解析与知识获取方法研究 [博士学位论文]. 北京: 中国科学院大学, 2022.
- [15] 王亚, 曹存根, 王石. 基于事件语义角色的常识知识获取. 中文信息学报, 2023, 37(6): 77–88. [doi: 10.3969/j.issn.1003-0077.2023.06.008]
- [21] 曹亚男. 面向 Web 语料的因果知识获取研究 [博士学位论文]. 北京: 中国科学院大学, 2012.
- [25] 曹聪, 曹存根, 臧良俊, 王石. 一种交互式事件常识知识的获取方法. 中文信息学报, 2016, 30(3): 125–132.
- [26] 王亚. 常识知识获取研究 [博士学位论文]. 北京: 中国科学院大学, 2022.
- [27] 邓鹏. 基于语义增强的事件图谱补全和推理研究 [硕士学位论文]. 北京: 中国科学院计算技术研究所, 2023.
- [34] 张宁豫, 谢辛, 陈想, 邓淑敏, 叶宏彬, 陈华钧. 基于知识协同微调的低资源知识图谱补全方法. 软件学报, 2022, 33(10): 3531–3545. <http://www.jos.org.cn/1000-9825/6628.htm> [doi: 10.13328/j.cnki.jos.006628]
- [44] 王亚, 曹存根. 基于事件属性的事件分类研究. 中文信息学报, 2020, 34(10): 39–50. [doi: 10.3969/j.issn.1003-0077.2020.10.006]
- [45] 宋衡, 曹存根, 王亚, 王石. 一种改进的汉语语义角色分类体系与标注实践. 中文信息学报, 2023, 37(1): 16–32. [doi: 10.3969/j.

[issn.1003-0077.2023.01.002](#)]

[48] 宋衡, 曹存根, 王亚, 王石. 一种细粒度的汉语语义角色标注数据集的构建方法. 中文信息学报, 2022, 36(12): 52–66, 73. [doi: [10.3969/j.issn.1003-0077.2022.12.006](#)]

[49] 鲁川. 知识工程语言学. 北京: 清华大学出版社, 2010.

[50] 王亚. 基于语义分类的常识知识获取方法研究 [硕士学位论文]. 桂林: 广西师范大学, 2015.

[51] 朱晓亚. 现代汉语句模研究. 北京: 北京大学出版, 2001.

[52] 袁毓林. 基于认知的汉语计算语言学研究. 北京: 北京大学出版社, 2008.

[53] 刘茂福, 胡慧君. 基于认知与计算的事件语义学研究. 北京: 科学出版社, 2013.

[55] 柏晓鹏. 中文命题库的全局性语义角色标注及其对汉语研究的影响. 语言科学, 2017, 16(5): 481–492. [doi: [10.7509/j.linsci.201705.031178](#)]

[56] 王诚文, 钱青青, 荀恩东, 邢丹, 李梦, 饶高琦. 三元搭配视角下的汉语动词语义角色知识库构建. 中文信息学报, 2020, 34(9): 19–27. [doi: [10.3969/j.issn.1003-0077.2020.09.004](#)]

[57] 王丽杰. 汉语语义依存分析研究 [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2010.

[60] 陈志文. 基于语义解析的事件图谱构建方法研究 [硕士学位论文]. 北京: 中国科学院计算技术研究所, 2023.



黄俏娟(1992—), 女, 博士生, CCF 学生会员, 主要研究领域为知识获取, 事件知识图谱补全, 人工智能.



王亚(1988—), 女, 博士, 主要研究领域为常识获取, 人工智能.



曹存根(1964—), 男, 博士, 研究员, 博士生导师, 主要研究领域为文本挖掘, 大数据知识处理.



王石(1981—), 男, 博士, 副研究员, CCF 专业会员, 主要研究领域为知识图谱, 语义分析, 自然语言处理.