

基于少样本学习和思维链提示的知识概念抽取方法研究

余霖琳, 熊龙洋, 陆雪松

华东师范大学数据科学与工程学院, 上海 200062

摘要

知识概念抽取在教育、医疗、金融领域均有重要的应用价值。知识概念抽取属于命名实体识别的一个细分任务, 但是由于缺乏数据集和知识概念实体类型的特殊性, 直接将通用命名实体识别方法运用到知识概念抽取任务中, 往往效果不佳。鉴于上述挑战, 利用开源大语言模型, 提出了基于少样本学习和思维链提示的知识概念抽取方法。首先, 通过对比学习训练关注了实体语义的文本表征, 并采用 K -近邻算法提升检索到的少样本示例的相关性。其次, 采用思维链提示的方法展示样本, 以提升大语言模型在知识概念抽取任务中的推理能力。在多个数据集上的实验结果表明, 基于少样本学习和思维链提示的知识概念抽取方法总体上表现出了优于现有方法的效果。

关键词

知识概念抽取; 命名实体识别; 大语言模型

中图分类号: TP391.1

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2025033

Research on knowledge concept extraction method based on few-shot learning and chain-of-thought prompting

She Linlin, Xiong Longyang, Lu Xuesong

School of Data Science & Engineering, East China Normal University, Shanghai 200062, China

Abstract

Knowledge concept extraction has important application value in the fields of education, medical care, and finance. Knowledge concept extraction is a sub-task of named entity recognition. However, due to the lack of data sets and the particularity of knowledge concept entity types, directly applying general named entity recognition methods to knowledge concept extraction tasks often has poor results. In view of the above challenges, a method based on few-shot learning and chain-of-thought prompting for knowledge concept extraction was proposed, utilizing open-source large language models. Firstly, text representations focusing on entity semantics were trained through contrastive learning, and the relevance of the retrieved few-shot examples was enhanced using the K -nearest neighbors algorithm. Secondly, a method utilizing chain-of-thought prompting was adopted to present the samples, with the aim of improving the reasoning ability of large language models in knowledge concept extraction. Experimental results on multiple datasets demonstrate that the few-shot learning and chain-of-thought prompting for knowledge concept extraction method, on

the whole, has shown results superior over existing methods.

Key words

knowledge concept extraction, named entity recognition, large language model

0 引言

知识概念抽取是指从文本或数据集中提取出有意义的知识概念的过程，在学习资源标注、知识概念推荐、知识图谱构建等领域具有重要应用价值。例如，在慕课（MOOC）平台上，海量的教学资源需要标注相应的知识概念，以便教师组织授课内容、平台推荐学习资源^[1]。在医疗领域，知识概念抽取被用于从文本病历中提取临床信息，广泛应用于临床决策支持、护理质量改进等各个领域^[2]。

知识概念本质上是一种实体，因此，其抽取过程是自然语言处理中的命名实体识别（named entity recognition, NER）方法^[3-4]的一种应用。NER方法用于从文本中识别出具有特定意义的实体名称，并将其分类为预先定义好的类别。这些实体可以是人名、地名、组织机构名、时间、日期、货币、百分比等。例如，在“华为公司于2024年发布了三折叠屏手机MateXT非凡大师”句子中，“华为公司”会被识别为组织机构名，“2024年”会被识别为时间，“MateXT非凡大师”会被识别为产品名。常见的NER方法包括基于人工规则的方法^[5]、基于统计机器学习的方法^[6]和基于深度学习的方法^[4]。利用NER方法，研究人员可以从视频、课件、文本等教学资源中提取关键概念和术语，从而提高知识概念抽取的准确性。然而，知识概念抽取也面临有别于传统NER方法的挑战。一是研究人员已经为传统NER方法构

建了大量数据集，涵盖了金融、文化、新闻等众多领域，相反地，用于训练知识概念抽取的数据集非常有限，并且包含的学科类别也相对较少。二是知识概念抽取不只是提取知识概念实体，更重要的是提取的必须是与教学资源相关的关键知识概念实体，同时需要自动过滤无关的知识概念实体。例如，在心理学课程文本“我们在亲密关系那节课里讲到如何促进你和配偶之间的关系，今天我们就着重来讲一讲，积极心理学怎么样可以帮助你，培养出一个更好的孩子”中，“亲密关系”和“积极心理学”就是关键知识概念，“配偶”和“孩子”是实体但不是相关的知识概念。因此，应用传统NER方法完成知识概念抽取任务，其性能存在一定的局限。

近年来，大语言模型（large language model, LLM）^[7-8]逐渐成为完成自然语言处理任务的主流技术，其丰富的知识储备和出色的指令遵循能力，为知识概念抽取任务提供了另一种可行的方法，即利用提示工程提示LLM从数据中抽取知识概念。少样本学习的提出更是展现了LLM在不需要大样本数据训练的情况下通过少量示例就能出色完成任务的能力。知识概念抽取涉及多领域、多学科，其现有的标注数据仅涉及部分学科且数据量匮乏^[9]，难以进行大样本训练，使用LLM结合少样本学习成为适配该任务的解决方案。当前，基于LLM的代表性NER方法包括GPT-NER^[10]和PromptNER^[11]。前者通过基于少样本学习的提示工程将NER方法从序列标注任务转变为基于LLM的文本生成任务。后者在提示中加入模块化定义，在

示例中融入思维链 (chain-of-thought, CoT)^[12], 使 LLM 在 NER 任务上运用得更加灵活, 提升了识别准确率。上述方法尽管取得了不错的效果, 但其应用于知识概念抽取任务时仍然效果欠佳, 主要挑战在于如何选取最佳的示例用于少样本学习, 以及如何用 CoT 提升知识概念抽取的精度。

针对上述挑战, 本文提出一种基于少样本学习和思维链提示的知识概念抽取 (few-shot learning and chain-of-thought prompting for knowledge concept extraction, FCKCE) 方法, 以提升抽取的精度。首先, 本文采用对比学习训练关注了实体语义的文本表征, 用于从训练集中检索文本-知识概念对, 从而提升检索到的少样本示例的相关性。其次, 本文对 PromptNER 中的 CoT 提示进行了改进, 以期提升 LLM 在知识概念抽取任务上的理解与推理能力, 使其更关注与教学资源相关的关键知识概念。本文采用了学科数据集和通用 NER 数据集进行实验, 实验结果表明, FCKCE 方法在大多数数据集上取得了优于现有方法的效果, 并且这一效果在不同的大模型上都得以体现。

1 相关工作

1.1 知识概念抽取

知识概念抽取被广泛应用于 MOOC 平台, 其原因是课程数量的快速增长, 导致海量教学资源的维护愈发困难, 学生更容易陷入知识迷航^[13], 因此, 需要为学习资源标注知识概念, 构建知识图谱, 从而为学习资源推荐和学习路径规划提供便利^[14-15]。为了自动标注知识概念, 国内外学者及研究团队提出了许多从学习资源中

抽取知识概念的方法。早期的工作采用引入外部数据源知识的方法来提高知识概念抽取的精准度。例如, Medelyan 等^[16]构建了 KEA++ 系统, 利用从特定领域词库中收集的术语和短语的语义信息来增强包含知识概念的关键短语提取的性能。Rospocher 等^[17]从语料库中提取特定领域的关键概念列表, 按相关性对其进行排序, 然后应用各种评估指标来评估领域本体相对于关键概念列表的术语覆盖率。Berend 等^[18]专注于科学出版物的知识概念提取, 并提出了一组用于监督学习的新特征, 包含短语级、语料库级和基于维基百科的外部知识特征。上述方法都利用了外部源中包含的显式语义知识。随着机器学习和深度学习方法的兴起, 近年来, 知识概念抽取主要采用数据驱动的方法。Pan 等^[19]和 Lu 等^[20]首先采用嵌入方法计算知识概念的表征, 随后提出了基于图传播的方法, 对知识概念进行排序, 并将排序靠前的知识概念作为抽取的结果。Wu 等^[21]和 Albahr 等^[22]提出的基于统计排名的方法, 用于从 MOOC 视频中提取包含知识概念的关键短语, 并进行排名。

上述传统方法均取得了不错的效果, 但仍然存在噪声和不完整标注的问题, 无法有效解决未登录词 (out-of-vocabulary, OOV) 和域外 (out-of-domain, OOD) 的问题, 特别是在知识概念抽取任务上, 这些问题更加突出。

1.2 基于 LLM 的命名实体识别

GPT-4^[7]、Qwen^[8] 和 Llama 2^[23] 等 LLM 在广泛的自然语言处理任务中取得了重大进展, 为人工智能技术的发展带来了新的转折点。由于其巨大的搜索空间和广泛的训练数据, LLM 拥有大量的知识, 并

具备解决 OOV、OOD 问题的潜力。然而，LLM 虽然擅长理解上下文和会话语言生成，但在 NER 任务上的表现并不理想。近期的研究表明，LLM 具有优秀的上下文学习能力。上下文学习通过将输入-输出演示集成到提示文本中，使现成的模型不仅不需要额外的微调，且可以同时实现类似的优秀性能^[24]。然而，在 NER 任务上，尽管 LLM 采用了一些上下文样本作为提示，但与 SOTA (state of the art) 方法仍存在显著的性能差距，这可能是由于 LLM 缺乏特定领域的学习和对 NER 任务的理解欠佳。这种限制被归因为 LLM 的“缺乏专业性”。因此，使用 LLM 直接处理 NER 任务仍然是一项具有挑战性的工作。

GPT-NER^[10] 是最早应用 LLM 处理 NER 任务的方法之一，其将 NER 序列标注式任务在 LLM 上转变为文本生成式任务，在实体左右两侧添加特殊符号，使用少量上下文样本提示 GPT-3 生成包含特殊符号的文本，然后通过识别特殊符号提取命名实体，实现了与有监督学习基准模型相同性能的水平。随着 GPT-4 的推出，LLM 在垂直领域的表现以及进行多步推理的能力得到进一步提升。UniversalNER^[25] 提出了一种以任务为中心的针对性蒸馏技术，结合指令微调开发了一个在开放域上的 NER 方法。

研究发现，当面对复杂的推理任务时，LLM 仅使用端到端推理方法往往表现不佳，但通过添加逐步推理的过程，可以提高其推理能力，这被称为 CoT 提示。CoT 提示使模型能够更准确地理解问题的复杂性和推理过程。此外，这种方式让模型生成了一系列的推理步骤，使模型的推理过程有了一个更透明的呈现，进一步提高了可解释性^[24]。PromptNER^[11] 通过在提示中加入模块化定义以及在示例中融入 CoT

的方式，使 LLM 在 NER 任务上能够突破之前仅使用上下文学习的效果，将端到端的推理模式拆解为一步步推理的模式，展现了 LLM 灵活适应 NER 任务的潜力。

现有工作主要采用以下两种自回归输出格式。

- 容纳原始输入文本以包含实体标签信息，被称为增强语言形式。

- 直接使用定制的、易于解析的结构化格式输出所有实体类型标签和实体，被称为结构化注释形式。

这两种方式都存在某些缺陷，其中，增强语言形式需要复制所有原始输入文本，从而增加输出长度，导致推理效率低下。结构化注释形式以自回归的方式生成所有的标签和实体，意味着每个后生成的标签实体对都依赖于其之前生成的标签实体对，导致序列长度过长。本文针对这两种输出格式的缺点，突破语言形式输出的冗余与结构化注释形式输出的繁杂，提出融入 CoT 的简明高效的输出格式，以少样本示例的方式提示 LLM 按示例格式输出。

2 FCKCE 方法

本节主要介绍 FCKCE 方法的细节，FCKCE 方法的整体流程如图 1 所示。首先，本文采用对比学习训练待抽取文本（句子）的表征；随后，对于每一个待抽取的句子，采用 K-近邻算法，从训练集中选取与其表征最接近 K 的句子及其对应的知识概念抽取结果，作为少样本示例；最后，基于 CoT 构建少样本示例，并构建完整的提示模板，提示 LLM 生成知识概念。

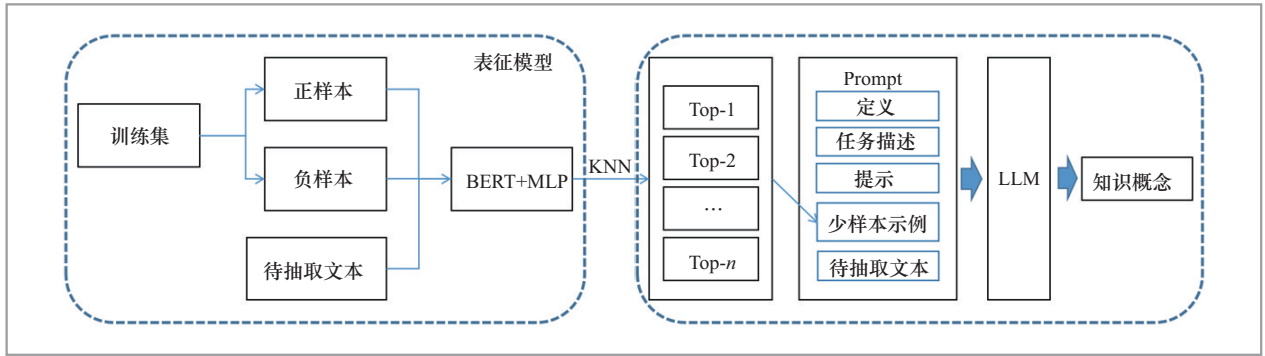


图1 FCKCE方法的整体流程

2.1 基于对比学习的文本表征学习

为了选取更相似的少样本示例，本文首先采用对比学习来训练待抽取知识概念的文本的表征，以期得到更有区分度的文本表征。本文使用双向编码器表示 (bidirectional encoder representations from transformer, BERT) 模型^[26]对输入文本进行嵌入，然后引入多层感知机 (multilayer perceptron, MLP) 对 BERT 模型生成的嵌入进行非线性变换，以增强文本嵌入的表达能。

具体来说，本文首先加载预训练的 BERT 模型和相应的分词器。对于长度为 n 的输入句子 $S = \{w_1, w_2, w_3, \dots, w_n\}$ ，将其转化为对应的词嵌入表示，如式 (1) 所示。其中， w_i 表示句子中的第 i 个词汇， x_i 表示词汇 w_i 对应的初始嵌入表示。

$$X = \{x_1, x_2, x_3, \dots, x_n\} \quad (1)$$

通过 BERT 模型生成其隐藏状态矩阵。

$$H = \text{BERT}(X) = \{h_1, h_2, h_3, \dots, h_n\} \quad (2)$$

随后对输出的隐藏状态进行平均池化，得到输入句子的全局嵌入向量，其中， h_i 表示词汇 w_i 对应的隐藏状态向量。

$$h_s = \frac{1}{n} \sum_{i=1}^n h_i \quad (3)$$

以往的研究主要采用 BERT 模型生成的第一个向量作为句子表征进行后续研究，这种方法主要关注句子的语义信息。本文则进行了改进，采用所有词输出的平均池化作为句子表征。这些表征不仅包含了句子的语义信息，还关注到了词语间的相似度，比以往的方法更能捕捉知识概念实体间的复杂关联。在 BERT 模型生成的嵌入基础上，MLP 作为一个非线性变换模块，由 5 层全连接神经网络组成，每层后接 ReLU 激活函数，其计算过程如下。

$$z_1 = \text{ReLU}(W_1 h_s + b_1) \quad (4)$$

$$z_2 = \text{ReLU}(W_2 z_1 + b_2) \quad (5)$$

$$z_3 = \text{ReLU}(W_3 z_2 + b_3) \quad (6)$$

$$z_4 = \text{ReLU}(W_4 z_3 + b_4) \quad (7)$$

$$e_s = W_5 z_4 + b_5 \quad (8)$$

其中， W_i 和 b_i 分别为每一层的权重和偏差参数，最终得到输入句子 S 的表征向量 e_s 。

在得到每个句子的表征后，本文采用对比学习来进一步提高其表征能力。具体

来说,对于训练集中的每一个句子 S_i ,选取 M 个与其表征最相似的句子作为正样本,选取 N 个与其表征最不相似的句子作为负样本。本文采用余弦相似度来计算句子表征之间的相似度。随后,本文采用最小化负对数似然函数作为句子 S_i 的对比损失,即最大化 e_{s_i} 与每一个正样本的表征 $e_{s_j}^+$ 的相似度,且最小化 e_{s_i} 与每一个负样本的表征 $e_{s_j}^-$ 的相似度,计算式如下。

$$L_i = -\ln \frac{\sum_{j=1}^M \exp(\text{sim}(e_{s_i}, e_{s_j}^+))}{\sum_{j=1}^M \exp(\text{sim}(e_{s_i}, e_{s_j}^+)) + \sum_{j=1}^N \exp(\text{sim}(e_{s_i}, e_{s_j}^-))} \quad (9)$$

其中, $\exp(\cdot)$ 为指数函数, $\text{sim}(\cdot)$ 为余弦相似度函数。训练完毕后,每一个句子的表征将会在语义和知识概念类型上,与其相似的句子更相似,而与其不同的句子更具有区分度。这种方式可以在后续少样本学习时,尽可能避免噪声样本的引入。

2.2 少样本示例选取

在提示中加入少样本示例,通常可以有效提升LLM的指令遵循能力,而如何选取样本示例是发挥少样本学习有效性的关键。在选取样本时,应尽可能挑选和输入与文本语义接近的示例,否则可能会降低LLM回复内容的准确性。为了挑选合适的少样本示例用于提示LLM进行知识概念抽取,本文采用基于句子表征的 K -近邻算法从验证集中选取示例。

具体来说,对于给定的测试句 S_i 和验证集中的候选句子 $\{S_1, S_2, \dots, S_m\}$,采用前文训练得到的表征模型获取 S_i 的表征向量 e_{s_i} 与候选句子集合的表征向量 $\{e_{s_1}, e_{s_2}, \dots, e_{s_m}\}$ 。随后,计算 e_{s_i} 与候选句子集合的每一个句子表征之间的余弦相似度。

$$\text{sim}(e_{s_i}, e_{s_j}) = \frac{e_{s_i} \cdot e_{s_j}}{\|e_{s_i}\| \cdot \|e_{s_j}\|}, i = 1, 2, \dots, m \quad (10)$$

最后,根据计算结果,选取与 e_{s_i} 最相似的 K 个表征对应的句子及其CoT推理过程作为提示LLM的少样本示例,以实现所选取的句子在语义和知识概念上都尽可能与测试句 S_i 相近。

2.3 CoT提示

本文首先构建了用于抽取知识概念的提示模板,该模板分为5个部分:知识概念定义、涉及学科、任务描述、提示(对需要排除的特殊情况的举例,如日期、代词),以及少样本示例。其中,少样本学习能够提升LLM的指令遵循能力,但在复杂任务上的推理能力仍然受限。知识概念抽取并非简单的推理任务,首先需要确定文本所属的学科,然后推理每个实体是不是该学科的一个知识概念,进一步判断其是不是文本的主要内容,最终确定其是不是文本的关键知识概念。鉴于上述要求,本文采用CoT提示的方式来抽取知识概念。

为了探索LLM的潜在推理能力,一个标准的方法是在CoT提示示例中引入不同的推理过程,并从中选取效果最佳的推理路径。通常,为了保证推理过程的多样性,可以采用较高性能的LLM(如GPT-4)来自动构建推理过程,并通过聚类方法选择具有代表性的路径生成示例。然而,由于未能融入人类价值,自动生成的CoT推理路径会遇到一些挑战,如质量差、存在幻觉以及与事实不符等问题。因此,本文采取手动与自动相结合的方式构建推理过程,先利用少量人工构建的推理过程示例来提示GPT-4生成更多不同的推理路径,随后进行人工检查,筛选出有效的推

理过程，并将其作为提示模板的重要组成部分。

基于上述方法，本文确定了CoT推理的过程，首先确定给定的短语属于什么学科，然后确定是否能认定其为知识概念。具体的结构要求为：输出的每一行首先是一个候选知识概念，紧接着以CoT的方式阐述该候选知识概念短语的特点，分析其是否满足知识概念的要求，最后得出是不是知识概念的结论。完整的知识概念抽取提示模板如图2所示。其中，第一段描述包含了知识概念的定义、待抽取概念的学科类别、知识概念抽取任务描述，以及关于注意事项的提示。后续是采用K-近邻算法选出的5个示例，其中，例1、3、4、5是正例，各包含了一段MOOC课堂教学视频的字幕文本，以及关于文本中知识概念的CoT推理过程；例2是一个反例，文本中不包含知识概念。最后输入待抽取知识概念的句子，要求LLM按照示例输出结果。

3 实验结果与分析

3.1 数据集

本文采用MOOCCube数据集^[27]验证FCKCE方法的效果。这是一个包含20个学科的中文数据集。包括由全国科学技术名词审定委员会审定的超过10万个知识概念的专家词典，以及来自315个课程的字幕语料库，平均每个课程包含167 496个未标记的字符序列。其训练集没有知识概念标注，验证集和测试集分别包含来自17门课程的272个和522个由专家标注的句子，包含15 375个与学科相关的知识概念。

由于缺乏其他包含知识概念的公开数据集，为了验证FCKCE方法的泛化能力，本文评估其在通用NER任务上的表现。本文选取了广泛使用的英文NER数据集CoNLL03^[28]和CrossNER^[29]。

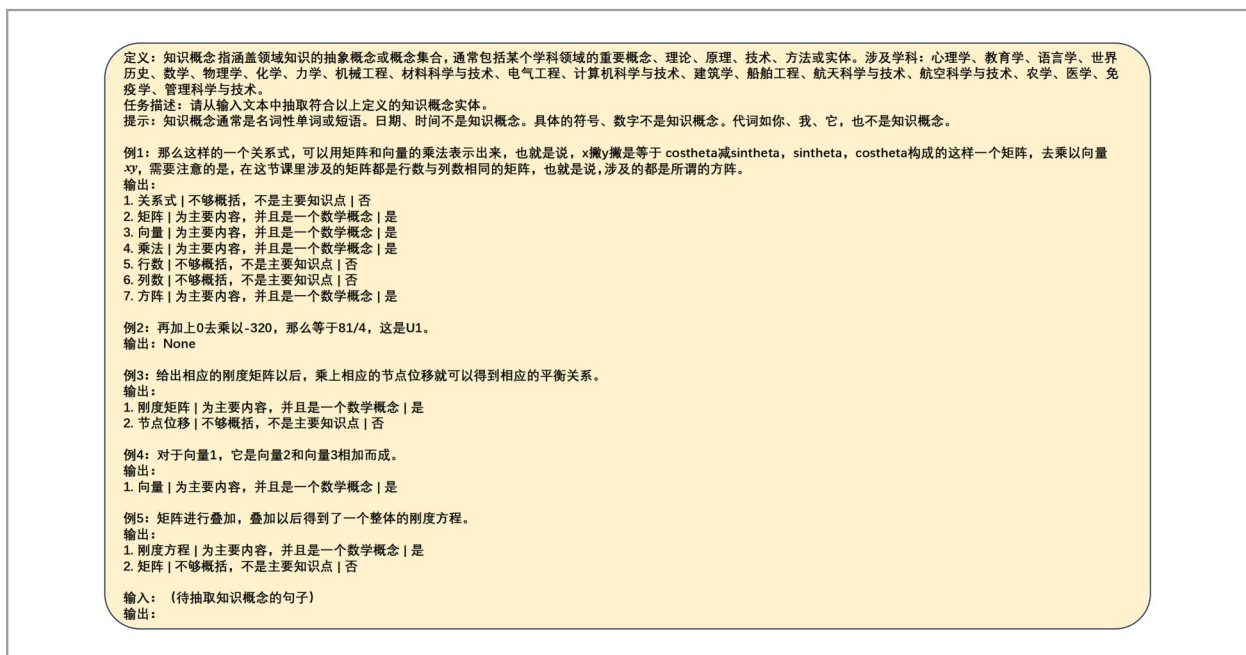


图2 完整的知识概念抽取提示模板

CoNLL03 包含多种语言下的 4 类命名实体：人物、地点、组织和不属于前 3 类的杂项实体名称。其中的英文数据取自路透社语料库，包含从 1996 年 8 月至 1997 年 8 月的路透社新闻报道。训练集和验证集由 1996 年 8 月底 10 天的数据构成，分别包含 14 987 个和 3 466 个句子。测试集由 1996 年 12 月的数据构成，包含 3 684 个句子。

CrossNER 涵盖政治、自然科学、音乐、文学和人工智能 5 个不同领域的句子，分别由该领域的专家提供实体类别。例如，音乐领域包含音乐流派、歌曲、乐队、专辑、音乐艺术家、乐器等；人工智能领域包含任务、产品、算法、研究员、指标、大学等。5 个领域的训练集、验证集和测试集均包含数百个已标注的句子，且数量略有不同。

3.2 实验设置

在对比实验中，对于每一个句子，本文选取 5 个正样本和 250 个负样本计算对比损失。此外，本文设置 BERT 模型的嵌入大小为 768，批大小为 32，学习率为 2×10^{-5} ，优化器为 Adam，训练 30 个轮次后收敛。

在知识概念抽取实验中，考虑到实验数据集涉及中英文两种语言，本文选取对中文、英文文本理解与生成能力都较强的开源大模型 Yi-34B^[30]和 Qwen1.5-32B^[8]，并分别设置 temperature 为 0.95 和 0.9，设置少样本示例的数量为 5 个。

3.3 对比模型和评价指标

对于每个大模型，本文分别采用 SimpleNER、GPT-NER^[10]、PromptNER^[11]和本文提出的 FCKCE 方法来比较性能。

其中，SimpleNER 采用简单的提示方法，仅仅给出知识概念抽取的任务描述，采用 5 个固定样本作为示例，每个样本包含句子和抽取的知识概念。GPT-NER 的提示模板中包含知识概念抽取任务和将该任务转换为文本生成任务的描述，采用 SimCSE^[31]模型对句子进行表征，并根据表征相似度选取最相似的 5 个句子和转化后的文本作为少样本示例。PromptNER 的提示模板和 FCKCE 方法类似，但是没有“涉及学科”部分，“提示”部分也不如 FCKCE 方法详细，仅提示了日期、时间、形容词和动词不属于实体。此外，PromptNER 采用固定的 5 个样本作为示例，每个样本都包含句子和提取知识概念的 CoT。与 FCKCE 方法不同，PromptNER 的 CoT 推理顺序与 FCKCE 方法的正好相反，即首先给出实体是不是知识概念的结论，随后阐述得出结论的理由。

本文采用准确率、召回率和 F1 分数作为评价指标。其中，准确率用于衡量所抽取的实体中确实是知识概念的比例；召回率用于衡量所有知识概念中被正确抽取的比例；F1 分数是综合考虑准确率和召回率的融合指标。这 3 个指标越高，说明模型性能越好。

3.4 知识概念抽取结果与分析

MOOCCube 的知识概念抽取结果见表 1。由表 1 可知，在 Yi 模型上，FCKCE 方法的表现超过了 GPT-NER 和 PromptNER，但是略低于 SimpleNER。本文推测，复杂提示反而效果欠佳的原因可能是过长的指令导致 Yi 理解过载。事实上，相关文献^[32-34]已经验证，过长的输入会使 LLM 的推理能力下降。尽管如此，FCKCE 方法的表现仍优于 GPT-NER 和 PromptNER，

这证明了本文设计的 K -近邻算法和 CoT 提示推理方法的有效性。这一结果在 Qwen 模型上更显著，FCKCE 方法在 Qwen 上取得了优于其他 3 种方法的最佳表现。

本文从 FCKCE 方法中删除对比学习模块，以及用固定示例代替 K -近邻算法选取少样本示例，构建了两个消融模型 FCKCE w/o CL 和 FCKCE w/o KNN，验证上述模块的作用。MOOCube 上 FCKCE 方法消融实验结果见表 2。由表 2 可知，在 Yi 和 Qwen 两个模型上，两个消融模型相较 FCKCE 方法的性能均大幅下降。这证明了本文设计的对比学习模块和 K -近邻算法模块的有效性。

本文利用全部验证集共 272 条标注数据以 LoRA^[35] 方式进行微调，以对比大样本学习与少样本学习效果，MOOCube 上 FCKCE 方法与大模型微调对比结果见表 3。由表 3 可知，在 Yi 和 Qwen 两个模型上，微调之后模型各项指标提升幅度仅为 1%~2%，而这部分提升需要耗费大量的标注数据，以及训练时间与计算资源。由此可以看出，在数据量缺乏的场景下，少样本学习更适合知识概念抽取任务，而本文提出的少样本学习配合 CoT 提示的方法使模型在只有少量标注数据时也能达到出色的效果。

3.5 命名实体识别结果与分析

为了验证 FCKCE 方法的泛化能力，本文在通用 NER 任务上对比前述方法的性能。对于 CoNLL03，本文直接在整个数据集上进行实验。对于 CrossNER，本文分领域进行实验。CoNLL03 和 CrossNER 上通用 NER 方法的 F1 分数的实验结果见表 4。由表 4 可知，FCKCE 方法在大多数情况下取得了最佳表现。其中，FCKCE 方法在 CoNLL03 和 CrossNER 的人工智能领

表1 MOOCube的知识概念抽取结果

模型	方法	准确率	召回率	F1 分数
Yi	SimpleNER	16.78%	30.29%	21.59%
	GPT-NER	1.12%	1.45%	1.26%
	PromptNER	12.49%	28.86%	17.43%
	FCKCE	14.88%	29.87%	19.53%
Qwen	SimpleNER	26.46%	32.68%	29.24%
	GPT-NER	28.40%	13.29%	18.10%
	PromptNER	21.56%	37.82%	27.46%
	FCKCE	35.45%	39.98%	37.30%

表2 MOOCube上 FCKCE方法消融实验结果

模型	方法	准确率	召回率	F1 分数
Yi	FCKCE	14.88%	29.87%	19.53%
	FCKCE w/o KNN	12.49%	28.86%	17.43%
	FCKCE w/o CL	9.92%	28.17%	14.67%
Qwen	FCKCE	35.45%	39.36%	37.30%
	FCKCE w/o KNN	21.56%	37.82%	27.46%
	FCKCE w/o CL	23.05%	38.26%	28.77%

表3 MOOCube上 FCKCE方法与大模型微调对比结果

模型	方法	样本量	准确率	召回率	F1 分数
Yi	FCKCE	5	14.88%	29.87%	19.53%
	LoRA 微调	272	15.29%	30.61%	21.34%
Qwen	FCKCE	5	35.45%	39.36%	37.30%
	LoRA 微调	272	36.76%	40.23%	38.08%

域上，表现略差于 SimpleNER，其原因可能是过长的提示模板导致 Yi 的推理性能下降。

3.6 案例分析

概率论与统计推断课程的文本案例 1

表4 CoNLL03和CrossNER上方法通用NER的F1分数实验结果

模型	方法	CoNLL03	CrossNER				
			政治	文学	音乐	人工智能	自然科学
Yi	SimpleNER	61.78%	60.75%	65.34%	60.33%	71.09%	60.28%
	GPT-NER	24.56%	60.89%	59.47%	62.80%	45.16%	57.51%
	PromptNER	51.98%	74.88%	32.00%	45.33%	24.72%	33.29%
	FCKCE	59.89%	75.47%	65.39%	74.02%	68.23%	70.42%
Qwen	SimpleNER	85.04%	75.04%	79.37%	78.06%	66.58%	75.81%
	GPT-NER	73.02%	72.55%	76.73%	71.49%	64.25%	74.68%
	PromptNER	85.03%	68.44%	78.15%	79.22%	68.13%	66.87%
	FCKCE	86.04%	81.56%	79.89%	85.65%	71.80%	85.65%

如图3所示，心理学课程的文本案例2如图4所示。图3和图4展示了MOOCCube中的两段文本，以及用前述4种方法提取的知识概念。为了便于观察，本文将正确提取的知识概念标注为灰色。FCKCE方法在这两个案例中均提取了最多的真实知识概念，并且提取的非知识概念实体也是最少的。第一段文本来自概率论与统计推断课程，由于这类课程的知识概念比较明确，因此4种方法均提取了较多的知识概念。第二段文本属于心理学课程，知识概念体系相对较为模糊，因此4种方法相较案例1

均有下降，尤其是GPT-NER未能提取任何正确的知识概念。FCKCE方法虽然提取了3个非知识概念实体，但是成功抽取了所有真实的知识概念。

4 结束语

针对缺乏数据集和知识概念实体类型特殊性的挑战，本文利用开源LLM，提出了一种基于少样本学习和思维链提示的FCKCE方法。一方面，本文用对比学习

待抽取文本：下面介绍贝叶斯网的推断，贝叶斯网络训练好之后，就能用来进行一些查询的工作，也就是说，通过一些已经观测到的属性变量，去推测其它未知的属性变量，这种通过已知变量观测值来推测待查询变量的过程就称为推断inference，已知变量的观测值称为证据Evidence，最理想的情况是直接根据贝叶斯网定义的联合概率分布来精确计算后验概率，然而，这样精确推断又被证明是NP难的问题，也就是说，当网络的结点较多的时候，连接很稠密，难以进行精确的推断，所以这里就要采用近似推断。通过降低精度要求，在有限的时间内求得近似解。

真实知识概念：精确推断，待查询，属性，近似推断，贝叶斯网络，evidence，联合概率分布，精度，后验概率，证据，观测值，inference，结点，NP难的问题，近似解，变量

SimpleNER：变量，推断，查询，贝叶斯网络，观测，证据，精确，稠密，近似推断，结点，联合概率分布，后验概率

GPT-NER：推断，证据，工作，概率，属性，结点，近似，贝叶斯，联合，稠密，变量

PromptNER：贝叶斯网络，精确，推断，联合概率分布，后验概率，NP难的问题，连接，稠密，有限的时间

FCKCE：贝叶斯网络，推断，待查询，属性，变量，后验概率，联合概率分布，NP难的问题，精确推断，近似推断，结点，稠密，近似解

图3 概率论与统计推断课程的文本案例1

待抽取文本：下面我们来讲一讲积极心理学，可以在家庭里面发挥的作用，此前啊我们在亲密关系那节课里呢，已经讲了很多，如何促进你和配偶之间的关系，所以今天呢我们就着重来讲一讲，积极心理学怎么样可以帮助你，培养出一个更好的孩子，那么此前的话其实我们讲过一种方法，叫做优势教养，而且我们也提到，优势教养的一个实证依据，就是当我们教父母怎么样进行优势教养之后呢，我们会发现孩子的行为问题减少了，而父母感受到的压力也是大幅的减少，那么这个具体是怎么做的呢，这个的话主要是澳大利亚墨尔本大学，leawaters沃德斯教授提出来的方法，她写过一本书，题目就叫优势开关，意思的话就是说，我们对于孩子，要打开我们意识中的一个优势开关，不仅仅只是像以前那样，总是在要求他修补缺点，总是看到你这儿不对那儿不对，我怎么样把他的这个问题给改掉，怎么样矫正他的这个缺点，而是要认识到发挥优势，比修补缺点更加重要，在实际操作的过程当中呢也确实发现，就是大部分父母啊，你让他去写孩子的优点。

真实知识概念：积极心理学，优势开关，优势，优势教养

SimpleNER：行为问题，压力，父母，家庭，**积极心理学**，**优势**，**优势教养**，缺点

GPT-NER：优点，心理学

PromptNER：**积极心理学**，家庭，亲密关系，夫妻关系，促进，孩子，**优势教养**，实证依据，行为问题，父母压力，墨尔本大学，leawaters沃德斯教授，**优势**，缺点，发挥优势，修补缺点，矫正，优点

FCKCE：**积极心理学**，亲密关系，**优势**，夫妻关系，**优势教养**，行为问题，**优势开关**

图4 心理学课程的文本案例2

训练关注了实体语义的文本表征，并采用K-近邻算法提升检索到的少样本示例的相关性；另一方面，本文采用CoT提示的方法展示样本，提升了LLM在知识概念抽取上的理解与推理能力。本文采用学科数据集和通用NER数据集进行实验，实验结果表明，FCKCE方法在大多数数据集上取得了优于现有方法的效果，并且这一效果在Yi和Qwen两个开源LLM上都得到了体现。

参考文献：

- [1] Gong J B, Wang S, Wang J L, et al. Attentional graph convolutional networks for knowledge concept recommendation in MOOCs in a heterogeneous view[C]// Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 79-88.
- [2] Fu S Y, Chen D, He H, et al. Clinical concept extraction: a methodology re-

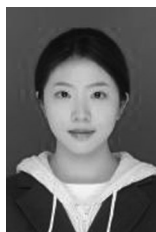
view[J]. Journal of Biomedical Informatics, 2020, 109: 103526.

- [3] Nadeau D, Sekine S. A survey of named entity recognition and classification[J]. Lingvisticae Investigationes, 2007, 30(1): 3-26.
- [4] Li J, Sun A X, Han J L, et al. A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 50-70.
- [5] Eftimov T, Koroušić Seljak B, Korošec P. A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations[J]. PLoS One, 2017, 12(6): e0179488.
- [6] Mansouri A, Affendey L S, Mamat A. Named entity recognition approaches[J]. International Journal of Computer Science and Network Security, 2008, 8(2): 339-344.
- [7] Achiam J, Adler S, Agarwal S, et al. GPT-4 technical report[PP]. arXiv preprint, 2023, arXiv: 2303.08774.

- [8] Bai J, Ba, S, Chu Y, et al. Qwen technical report[PP]. arXiv preprint, 2023, arXiv: 2309.16609.
- [9] Lu M Y, Wang Y Q, Yu J F, et al. Distantly supervised course concept extraction in MOOCs with academic discipline[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. [S.l.:s.n.], 2023.
- [10] Wang S, Sun X, Li X, et al. GPT-NER: named entity recognition via large language models[PP]. arXiv preprint, 2023, arXiv: 2304.10428.
- [11] Ashok D, Lipton Z C. PromptNER: prompting for named entity recognition[PP]. arXiv preprint, 2023, arXiv: 2305.15444.
- [12] Wei J, Wang X Z, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM, 2022: 24824–24837.
- [13] Chen C M. Intelligent web-based learning system with personalized learning path guidance[J]. Computers & Education, 2008, 51(2): 787–814.
- [14] Gu H N, Duan Z Y, Xie P, et al. Modeling balanced explicit and implicit relations with contrastive learning for knowledge concept recommendation in MOOCs[C]//Proceedings of the ACM Web Conference 2024. New York: ACM, 2024: 3712–3722.
- [15] Zhang H T, Shen S H, Xu B H, et al. Item-difficulty-aware learning path recommendation: from a real walking perspective[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM, 2024.
- [16] Medelyan O, Witten I H. Thesaurus based automatic keyphrase indexing[C]//Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries. New York: ACM, 2006: 296–302.
- [17] Rospocher M, Tonelli S, Serafini L, et al. Corpus-based terminological evaluation of ontologies[J]. Applied Ontology, 2012, 7(4): 429–448.
- [18] Berend G, Farkas R. Sztergak: feature engineering for keyphrase extraction[J]. International Workshop on Semantic Evaluation, 2010: 186–189.
- [19] Pan L, Wang X, Li C, et al. Course concept extraction in MOOCs via embedding-based graph propagation[C]//Proceedings of the Eighth International Joint Conference on Natural Language Processing. [S.l.:s.n.], 2017: 875–884.
- [20] Lu W M, Zhou Y F, Yu J L, et al. Concept extraction and prerequisite relation learning from educational data[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 9678–9685.
- [21] Wu Z J, Zhu J, Xu S, et al. LTWNN: a novel approach using sentence embeddings for extracting diverse concepts in MOOCs[M]//AI 2021: Advances in Artificial Intelligence. Cham: Springer International Publishing, 2022: 763–774.
- [22] Albahr A, Che D R, Albahr M. A novel cluster-based approach for keyphrase extraction from MOOC video lectures[J]. Knowledge and Information Systems, 2021, 63(7): 1663–1686.
- [23] Touvron H, Martin L, Stone K, et al. Llama 2: open foundation and fine-tuned chat models[PP]. arXiv preprint, 2023, arXiv: 2307.09288.
- [24] Chu Z, Chen J C, Chen Q L, et al. Navi-

- gate through enigmatic labyrinth a survey of chain of thought reasoning: advances, frontiers and future[C]//Annual Meeting of the Association for Computational Linguistics. [S.l.:s.n.], 2023.
- [25] Zhou W X, Zhang S, Gu Y, et al. UniversalNER: targeted distillation from large language models for open named entity recognition[PP]. arXiv preprint, 2023, arXiv: 2308.03279.
- [26] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2019. [S.l.: s. n.], 2019: 3498-4195.
- [27] Yu J F, Luo G, Xiao T, et al. MOOC-Cube: a large-scale data repository for NLP applications in MOOCs[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USAACL, 2020: 3135-3142.
- [28] Sang E F T K, De Meulder F. Introduction to the CoNLL-2003 shared task: language-independent named entity recognition[PP]. arXiv preprint, 2003, arXiv: cs/0306050.
- [29] Liu Z H, Xu Y, Yu T Z, et al. Cross-NER: evaluating cross-domain named entity recognition[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(15): 13452-13460.
- [30] Young A, Chen B, Li C, et al. Yi: open foundation models by 01. AI[PP]. arXiv preprint, 2024, arXiv: 2403.04652.
- [31] Gao T Y, Yao X C, Chen D Q. SimCSE: simple contrastive learning of sentence embeddings[PP]. arXiv preprint, 2021, arXiv: 2104.08821.
- [32] Li J Q, Wang M M, Zheng Z L, et al. LooGLE: can long-context language models understand long contexts? [PP]. arXiv preprint, 2023, arXiv: 2311.04939.
- [33] Bai Y S, Lv X, Zhang J J, et al. Long-Bench: a bilingual, multitask benchmark for long context understanding [PP]. arXiv preprint, 2023, arXiv: 2308.14508.
- [34] Levy M, Jacoby A, Goldberg Y. Same task, more tokens: the impact of input length on the reasoning performance of large language models[PP]. arXiv preprint, 2024, arxiv: 2402.14848v2.
- [35] Hu E J, Shen Y, Wallis P, et al. Lora: low-rank adaptation of large language models[PP]. arXiv preprint, 2021, arXiv: 2106.09685.

作者简介



余霖琳 (2000-), 女, 华东师范大学数据科学与工程学院硕士生, 主要研究方向为教育领域大模型应用。



熊龙洋 (2002-), 男, 华东师范大学数据科学与工程学院本科生, 主要研究方向为教育领域大模型应用。



陆雪松 (1985-), 男, 华东师范大学数据科学与工程学院副教授, 主要研究方向为数据驱动的计算教育学。

收稿日期: 2024-10-31

通信作者: 陆雪松, xslu@dase.ecnu.edu.cn

基金项目: 国家自然科学基金项目(No.62277017)

Foundation Item: The National Natural Science Foundation of China (No.62277017)