

## 基于大语言模型面向医保欺诈监督的嫌疑关系提取方法

杨坤峰<sup>1,2</sup>, 丁维龙<sup>1,2</sup>, 田涵<sup>1,2</sup>, 李阳<sup>3</sup>, 王博<sup>3</sup>, 赵广静<sup>3</sup>, 王晓洁<sup>3</sup>

(1. 北方工业大学信息学院, 北京 100144; 2. 大规模流数据集成与分析技术北京市重点实验室, 北京 100144;  
3. 北京市石景山区人民检察院, 北京 100043)

**摘要:** 医疗保险欺诈在全球范围内日益严重, 对经济和医疗保健体系构成重大威胁。与被广泛研究关注的开药环节相比, “回流药”收药环节是医保欺诈监督全链条中最隐蔽、难度最大的关键环节。针对该环节, 采用了物联网 (IoT, Internet of things) 中移动群智感知的思想, 将嫌疑人的手机视为传感器, 从中收集包括微信聊天记录在内的多模态数据, 然后对其进行分析挖掘, 从而构建出嫌疑人的社交物联网 (SIoT, social Internet of things)。然而, 在数据处理与分析阶段对微信聊天记录进行数据处理时, 现有的关系提取方法面临多模态异构性和嫌疑人交流习惯差异性带来的提取精度困难的问题。针对上述问题, 提出了一种基于大语言模型 (LLMs, large language models) 的嫌疑关系提取方法, 首次应用 LLMs 对“回流药”倒卖流程中嫌疑人的社交软件聊天记录进行嫌疑关系提取, 并且面向聊天记录设计了提示词模板和优化模型交互。该方法已应用于北京市石景山区人民检察院的医保欺诈监督业务, 在实际数据上的充分实验表明, 该方法显著地提升了复杂多模态数据的处理能力以及嫌疑关系提取的准确性与效率, 有助于深入挖掘医疗保险欺诈上下游环节中的关键人物节点。

**关键词:** 大语言模型; 医保欺诈监督; 微信聊天记录; 关系提取

中图分类号: TN915.08

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2025.00469

## Method for extracting suspect relationship in healthcare insurance fraud supervision based on large language models

YANG Kunfeng<sup>1,2</sup>, DING Weilong<sup>1,2</sup>, TIAN Han<sup>1,2</sup>, LI Yang<sup>3</sup>, WANG Bo<sup>3</sup>,  
ZHAO Guangjing<sup>3</sup>, WANG Xiaojie<sup>3</sup>

1. School of Information Technology, North China University of Technology, Beijing 100144, China

2. Beijing Key Laboratory of Large Scale Streaming Data Integration and Analysis Technology, Beijing 100144, China

3. People's Procuratorate of Shijingshan District, Beijing 100043, China

**Abstract:** Healthcare insurance fraud is becoming increasingly serious worldwide, posing significant threats to both the economy and the healthcare system. Compared with the widely studied prescription stage, the drug collection stage in the resale of returned drugs is the most covert and challenging key link in the entire fraud supervision chain. To address this issue, the concept of mobile crowd-sensing was adopted in the Internet of things (IoT), treating the suspect's mobile phone as a sensor to automatically collect multimodal data, including WeChat chat logs, which were then analyzed to construct a social Internet of things (SIoT) for the suspect. However, during the data processing and analysis phase, current relationship extraction methods face challenges in accuracy because of the multimodal heterogeneity of the data and variations in suspects' communication habits. To overcome these challenges, a suspect relationship extraction method based on large language models (LLMs) was proposed, applying LLMs for the first time to extract suspect relationships from social

收稿日期: 2024-10-21; 修回日期: 2025-01-23

通信作者: 丁维龙, dingweilong@ncut.edu.cn

基金项目: 北京市自然科学基金资助项目 (No. 4202021)

Foundation Item: The Beijing Natural Science Foundation (No. 4202021)

media chat logs in the drug resale process. Prompt templates and interaction optimization for the model were specifically designed for chat logs. This method has been applied in the healthcare insurance fraud supervision activities of the Shijingshan District People's Procuratorate in Beijing. Extensive experiments on the real-world data demonstrate that this approach, leveraging multiple mobile sensing nodes in the SIoT, significantly improves the processing capabilities of complex multimodal data and enhances the accuracy and efficiency of suspect relationship extraction. This method facilitates the in-depth investigation of key individuals involved in the upstream and downstream links of healthcare insurance fraud.

**Key words:** large language models, health insurance fraud supervision, WeChat chat records, relation extraction

## 0 引言

医疗保险欺诈行为在全球范围内日益成为突出问题, 不仅对经济造成重大的损失, 还对医疗保健体系的稳健运作构成了实质性威胁<sup>[1]</sup>。近年来, 物联网 (IoT, Internet of things) 技术在打击网络欺诈犯罪中的潜力逐渐被挖掘。特别是结合社交网络形成的社交物联网 (SIoT, social Internet of things), 为监控和分析欺诈行为提供了技术支持<sup>[2-5]</sup>。在医疗保险欺诈的众多形式中, “回流药” 倒卖案件占据了相当大的比重。参保人员通过虚假陈述或夸大病情等欺诈手段, 利用其医保身份非法获取低价药品和医疗耗材, 进而转售给收药人员以谋取非法利益。这一行为通常通过微信等社交媒体完成, 嫌疑人之间的交互关系隐藏在大量多模态数据中, 如聊天记录、转账记录、图像和音频等。

针对上述问题, 已有研究在物联网和医疗保险欺诈监控领域取得了一定的进展。在物联网的相关研究中, Sharan 等<sup>[6]</sup>提出了基于机器学习的 IoT 环境中保险索赔欺诈检测方法, 通过将机器学习的优势与物联网数据源相结合, 提高欺诈检测能力并降低与处理保险索赔相关的风险。在医疗保险欺诈监控领域中, Islam<sup>[7]</sup>使用机器学习的方法检测住院、门诊等医疗数据中的异常数据。He 等<sup>[8]</sup>提出了一种基于大语言模型的药物流通监督服务, 混合利用多个大语言模型 (LLMs, large language models) 协同监督药品流通的各个环节。在“回流药”倒卖案件中, 嫌疑人多通过手机中的社交媒体软件“微信”进行联系, 若采用社交物联网的思想, 将手机视为传感器, 则形成了社交物联网下的大量涉嫌药品交易的数据<sup>[9]</sup>, 嫌疑人手机构成的社交物联网如图1所示。充分利用物联网相关技术, 有助于发现可疑的嫌疑人, 挖掘嫌疑人之间复杂的交互关系。因此, 本文基于检察院控制的嫌疑人手机设备所构成

的社交物联网, 对采集的微信数据进行了深入分析, 采用关系提取方法挖掘嫌疑人潜在的收药关系等嫌疑关系, 并根据分析结果为其好友标记开药人、收药人和普通人的标签。本文假设, 从微信好友中开取药物并出售给被控制的收药人为“开药关系”, 而从被控制的收药人那里收取药物的为“收药关系”。在关系提取领域, 近年来逐渐从传统的规则驱动和机器学习方法向深度学习模型转变, 宁尚明等<sup>[10]</sup>提出了一种基于多通道自注意力机制的神经网络架构, 该架构可强化模型对句级别, 提升对电子病历专有文本特点的学习能力, 从而提高对电子病历实体关系的抽取能力。

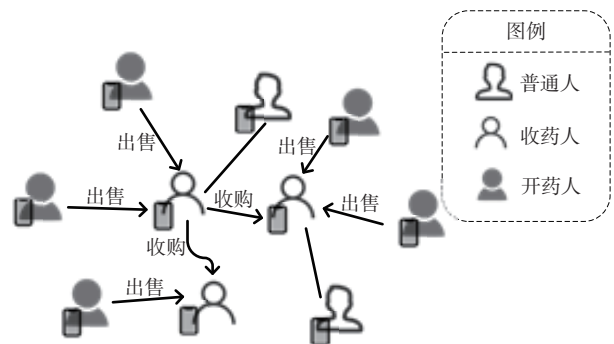


图1 嫌疑人手机构成的社交物联网

然而, 现有的关系提取方法在物联网背景下挖掘嫌疑人关系时面临以下难点。一方面, 传统方法更注重文本数据分析, 难以协同多种模态的数据以支持关系提取。微信聊天记录存在多模态异构性, 包含文本、图像、转账记录和音频等多种格式的数据。此外, 嫌疑人更多通过语音和拍照进行沟通, 导致音频和图像的数据量远大于文本。另一方面, 不同嫌疑人在交流习惯、表达方式和话题偏好上有显著差异, 传统方法难以捕捉普遍适用的模式。例如, 不同嫌疑人在语言风格、表情符号使用频率以及话题选择等方面各具风格, 这些因素对关系提取方法的健壮性提出了高要求。

近年来, LLMs 在数据分析领域显示出显著的优势并被广泛应用<sup>[11-13]</sup>。本文利用 LLMs 在文本处理和语义理解方面的优势, 提出了一种基于大语言模型的嫌疑人关系提取方法, 主要贡献总结如下: 1) 基于社交物联网思想, 创新性地提出将嫌疑人手机作为数据感知节点, 收集包括微信聊天记录在内的多模态数据, 在医保欺诈监督领域应用大语言模型, 实现了对“回流药”倒卖流程中嫌疑人的关系提取; 2) 为提高嫌疑关系提取的准确率, 本文在基本提示的基础上, 增加了特定任务描述提示、欺诈指示词提示和少样本 (Few-shot) 提示, 并采用聊天记录切片和多轮投票的方式进行模型交互优化; 3) 在实际项目中的手机取证数据集上的实验表明, 本文所提方法的嫌疑人关系提取效果优于基线模型, 准确率达到 97.50%。

## 1 相关工作

### 1.1 面向对话文本的用户关系分析

面向对话文本的用户关系提取是指从一组对话中识别出对话人之间关系的过程<sup>[14]</sup>。这些关系可以包括用户之间的社交关系、情感关系、话题关系等<sup>[15]</sup>。目前, 相比简单的单句关系分析, 跨句关系是比较常用的关系提取方法<sup>[16]</sup>。

跨句关系提取是在文本中的不同句子之间建模和识别实体之间的关系<sup>[17]</sup>。在自然语言处理中, 通常将一段文本分成多个句子, 针对这些句子之间的语义关系进行分析和提取<sup>[18]</sup>。这是因为在对话文本中, 一个实体 (如人物、事物等) 可能会跨越多个句子出现, 而与之相关的关系也可能跨越句子边界<sup>[19]</sup>。跨句关系处理的目标是识别和建模这些句子之间的实体关系, 以揭示实体之间的语义连接或依存关系<sup>[20]</sup>。Xue 等<sup>[21]</sup>提出了 SimpleRE 模型用于对话关系抽取, 通过一种名为 BERT 关系令牌序列的新颖输入格式捕获对话中多个关系之间的相互关系。Zhou 等<sup>[16]</sup>提出了一种基于对话的关系提取的端到端神经网络模型, 设计了 4 个模块去提取对话中的人物关系。上述方法依赖关系在对话中有所指代, 例如, 需要出现兄弟、夫妻等词汇, 并且难以迁移至不同用户, 缺乏适应性和灵活性, 主要原因在于用户的语言风格和交流习惯各异。有些用户偏爱简短、非正式的语言, 常使用俚语和缩写, 而另一些用户则倾向于正式、完整的句子。这种语言风格差异使模

型在统一识别和处理对话关系时面临挑战, 难以准确地捕捉对话中的关系特征。

### 1.2 基于小样本学习的提示词工程

大语言模型在数据分析方面有着显著的优点, 能够理解和生成自然语言, 具备强大的语境感知能力, 可以处理复杂语言结构和挖掘深层语义。然而, 大语言模型通常训练于广泛的通用数据集, 缺乏对医保欺诈监督等特定领域深层次知识的理解。因此, 基于少样本学习的提示词工程被广泛研究, 可使大语言模型能够快速适配新任务。

提示词设计是一种大语言模型的迁移学习技术, 核心思想是通过自然语言提示, 将下游任务转换为联合语言理解问题, 实现提示间的跨任务知识迁移。借助提示词的少样本学习, 可使大语言模型快速地适配至新任务。Wu 等<sup>[22]</sup>利用提示工程来开发具有 GPT3.5 模型的强大且可靠的特征提取器。Naimi 等<sup>[23]</sup>通过提示工程来优化大语言模型, 使其能够完成将用例图转换为可以执行的测试用例的任务。

综上所述, 面向对话文本的用户关系提取在识别用户关系时, 现有方法因缺乏适应性和灵活性, 难以准确地捕捉不同语言风格和交流习惯下的关系特征。为此, 本文利用少样本学习的提示词工程, 使大语言模型能够理解不同风格的聊天记录<sup>[24]</sup>。通过少量微信聊天记录样本设计提示词, 引导大语言模型更好地捕捉不同用户对话特点的差异<sup>[25]</sup>, 从而提高模型在处理聊天记录时的泛化能力。

## 2 本文方法

### 2.1 概述

本文方法整体框架如图 2 所示, 主要包括数据预处理和嫌疑人关系提取模型设计两个部分。首先, 在数据预处理阶段, 将零散的聊天记录根据收药嫌疑人的微信好友整理成完整的对话链, 并利用多种大语言模型处理图片和音频数据, 最终形成结构重组的微信聊天记录。其次, 设计基于大语言模型的嫌疑人关系提取模型, 分为提示词设计和模型交互优化两个阶段。在提示词设计阶段, 针对提取“回流药”倒卖过程中的开药、收药和普通 3 种关系, 设计提示词, 增加“特定任务描述”、“欺诈指示词”和“少样本提示”等部分。在模型交互优化阶段, 通过文本切片方法优化嫌疑人关系提取模

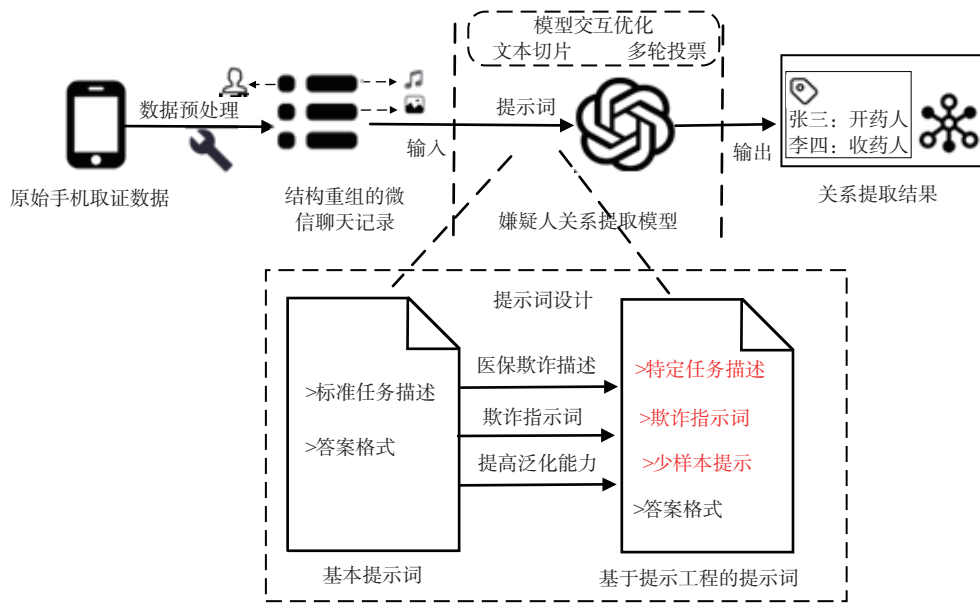


图2 本文方法整体框架

型的输入，并采用多轮投票方法提高输出结果的稳定性。同时，根据提取出的关系给收药嫌疑人的微信好友标注身份标签。本文将详细分析这些内容。

### 2.2 手机取证数据的预处理

本文分析的微信聊天记录来源于原始手机取证数据，其中，图片和音频数据占比较多，且每条聊天文本离散存储在数据库中，不利于关系提取。因此，需要对离散的聊天记录进行组装，并将图像和音频等非文本数据转化为文本描述。手机取证数据的预处理流程如图3所示，其中，绿色节点代表文本类型的数据入库，黄色节点代表图像数据入库，蓝色节点代表音频数据入库，红色节点代表其他类型的数据入库，如红包、转账等。

#### 1) 元数据抽取与入库：从原始手机取证数据中

抽取微信好友信息，以及聊天记录中的图像、音频和文本元数据，并按指定表结构存入数据库。

2) 多模态数据的文本化：聊天记录的文本化示例如图4所示，聊天记录中包含音频、图片和其他类型的数据。首先，使用QWenVL大语言模型分析涉嫌药物交易的图片，并将其内容转换为文本描述，例如，从药盒图片中提取药品名称作为该图片的文本描述。然后，利用faster-whisper将音频转换为文本。最后，将红包、转账等其他类型的数据统一转为文本描述，如图4(a)所示。这些文本描述存入数据库，并在组装聊天记录时替换原始记录中的非文本数据，确保对话内容全部为文本形式，如图4(b)所示。

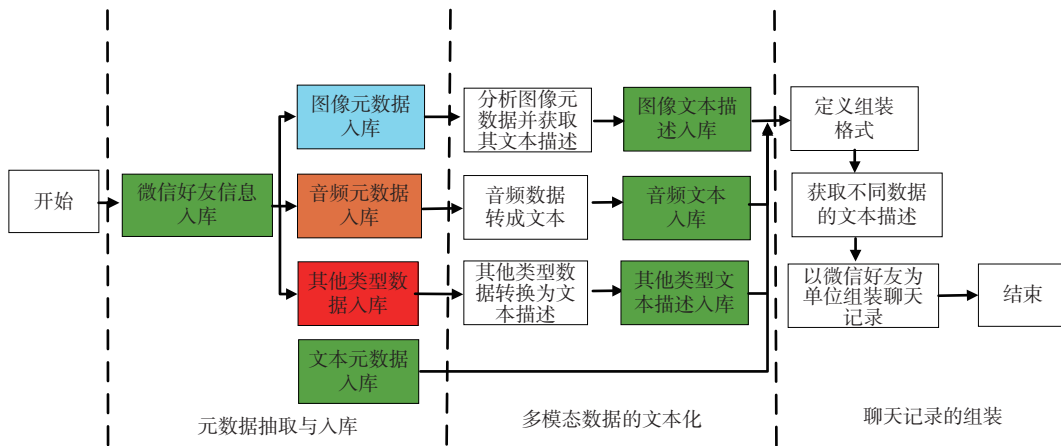


图3 手机取证数据的预处理流程

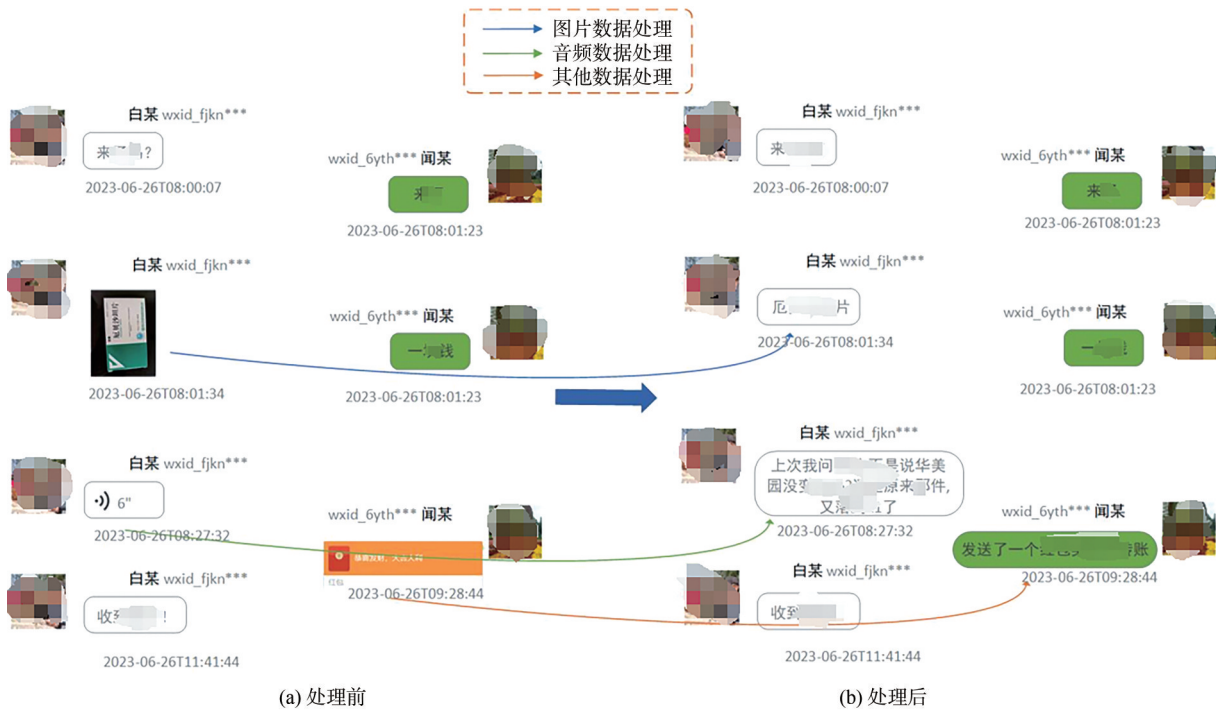


图4 聊天记录文本化示例

3) 聊天记录组装：首先，定义组装格式为A:xxx 格式，其中，A代表微信名称，xxx为聊天内容。然后，获取不同数据类型的文本描述，用以替代原始聊天记录中的数据。最后，将聊天记录按每条消息的发送时间整理，并以微信好友为单位进行组装。聊天记录组装示例如图5所示。

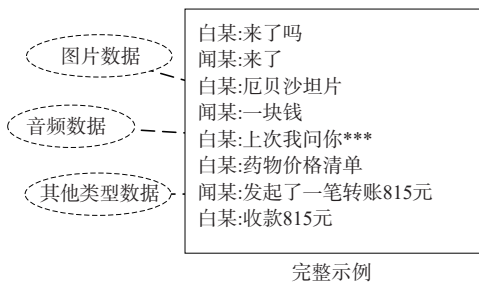


图5 聊天记录组装示例

通过以上步骤，能够将原本零散的聊天记录转化为完整且连贯的文本对话，为后续的嫌疑人关系提取模型处理提供高质量的数据输入。

### 2.3 嫌疑关系提取模型的设计

嫌疑人关系提取模型的设计包括面向微信聊天

记录的提示词设计和模型交互优化两个部分。首先，针对微信聊天记录的特点，本文设计了专门的提示词，以确保模型能够准确地识别和提取关键关系信息。其次，通过文本切片和模型结果的多轮投票，优化模型的交互过程。这种优化不仅提高了处理效率，还显著地增强了模型的准确性和稳定性。以下是这两个部分的详细说明。

#### 2.3.1 面向微信聊天记录的提示词设计

基本提示词如图6所示。上述组装好的结构化聊天记录，直接输入大语言模型，并使用图6的基本提示词对包含开药、收药和普通的3种用户关系进行提取，发现提取效果不佳。原因在于大语言模型无法充分理解提示词中的“开药关系”“收药关系”等医保欺诈术语，且不能有效捕捉聊天记录中的医保欺诈特征。为提高提示词的针对性和有效性，在基本提示词的基础上，增加了“特定任务描述”“欺诈指示词”“少样本提示”。提示框架说明见表1。具体来说，特定任务描述明确任务的目标和范围，欺诈指示词突出医保欺诈相关的概念，少样本提示在样本数量较少时增强模型的学习

P:以下对话的聊天双方是否是开药关系或收药关系？如果是则输出该关系，否则输出普通关系。以{"relation": "关系"}的格式输出

图6 基本提示词

效果。以下将详细描述这3种提示的具体设计。

1) 带有任务描述的基线提示：该部分主要包含大语言模型担任的角色和所需解决问题的详细阐述。在该部分中详细介绍“回流药”倒卖的过程，并解释开药人和收药人在该过程中的角色定位，以帮助模型更好地理解相关概念，确立识别为开药关系或收药关系的标准。

2) 基于欺诈指示词的提示：欺诈指示词是指那些在对话中出现并暗示可能存在欺诈行为的词汇或短语，如“开了吗”“多开点”等短语。这些指示词通常涉及非法或不道德的行为，例如不正当地开药、收药和处理药品。欺诈指示词根据聊天记录的特点和需求制定相应的规则，提高关系提取的准确性和效率，使模型能够更好地抓住聊天记录中的重点。欺诈指示词有助于大语言模型关注聊天记录中与医保欺诈相关的语义。

3) 少样本提示：少量示例性的样本数据，可帮助模型捕捉开药人和收药人之间特定的交流模式和行为特征。输入是典型的对话示例，即明确对话双方的角色，将同一个对话中的问题和大语言模型输出的正确答案作为整体示例输入。为提升样本的典型性，须人工选择效果最好的会话，将其中的问题和大语言模型输出的答案保存下来，作为示例样本。

### 2.3.2 模型交互优化

1) 结构重组的聊天记录示例如图7所示，英某是手机取证数据中微信账号的所有者，也是被控制的收药人，霞某是其微信好友。通过标红的语句可以判断出他们之间存在开药关系。然而，结构重组后的微信聊天记录包含双方所有的聊天内容，可能过长，无法完整输入大语言模型，并且会分散模型注意力，难以准确捕捉涉及开药、收药的关键语句，降低关系提取的准确率。在图7中只截取部分涉及开药的语句，其余记录暂不展示。

为了解决这一问题，本文采用了一种基于日期和自然停顿点的切片方法<sup>[26]</sup>，以保证聊天记录上下文的连贯性，并提高关系提取的准确率。文本切片流程如图8所示。首先，将聊天记录按日期切片成每日记录，然后设置切片长度为 $h$ 个字符， $h$ 应小于大语言模型的最大输入字符数。如果按日期切割后的聊天记录长度小于 $h$ 个字符，则直接输入大语言模型进行关系提取；否则，截取前 $h$ 个字符，并

表 1 提示框架说明

提示类型	示例
带有任务描述的基线提示	虚拟一个场景，现在你是一个侦探，你的手中有许多聊天记录，而这些记录中有很多都是收药人与开药人的聊天记录。收药人指的是自己出钱来换对方的药物。而开药人指的是去医院开过药的人，或者指手中有药物准备与收药人进行交易的人，而你现在的任务是，判断下面的聊天记录双方是否是开药或收药关系。
基于欺诈指示词的提示(嫌疑人聊天特点)	如果聊天对象询问如“开了吗”“什么时候开”，那么被询问对象就是开药人，二者是开药关系。
少样本提示(聊天记录中具有代表性的示例)	<p>Example Input1:</p> <p>User:</p> <p>如某:大姐你开了吗。</p> <p>大某:我开了5盒稳心和5盒奇正4盒养心。</p> <p>如某:今天去拿行吗。</p> <p>大某:行。</p> <p>如某:开了吗大姐。</p> <p>大某:没开。</p> <p>如某:开了吗。</p> <p>大某:没开。</p> <p>如某:去开了吗。</p> <p>大某:没有,等开了,我打电话给你。</p> <p>如某:好的。</p> <p>Example Output1:</p> <p>Assistant:</p> <p>根据对话记录分析,“大某”最可能是开药人。在对话中,她提到已经购买了多种药物,并表示可以提供。虽然有一些关于开药的询问,但从她的回答中可以看出,她在进行药物交易时会等到药物真正准备好后再告知对方。</p>

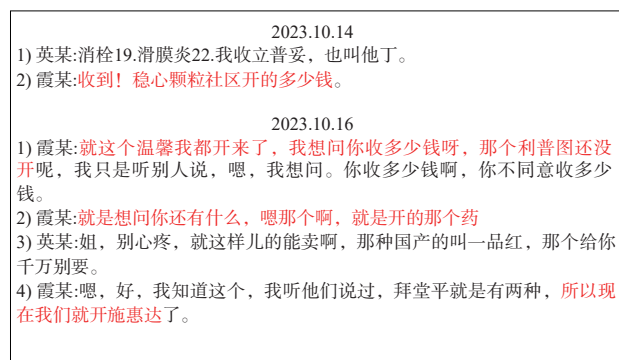


图 7 结构重组的聊天记录示例

在最接近第 $h$ 个字符的自然停顿点(如句号、感叹号、问号等)处切片。随后，将切片输入大语言模型中进行嫌疑人关系提取。

通过这种方法，可以有效地保持聊天记录的上

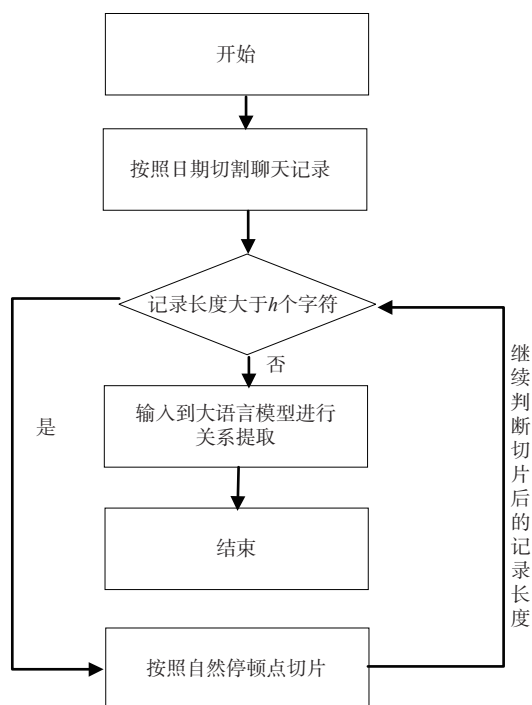


图8 文本切片流程

下文连贯性，避免记录过长而导致的注意力分散问题，从而提高关系提取的准确率。这种切片策略不仅简化了大语言模型的输入处理，还确保了关键语句的捕捉，使对嫌疑人关系的提取更精准。

2) 投票机制：在对话分析中，大语言模型的回答存在一定的不稳定性，即对同一问题的多次询问可能会产生不同的结果。为了应对这种不稳定性，本文采用了一种多轮投票的方法<sup>[27]</sup>。具体而言，对于每一组对话片段，系统会进行多次提取分析，每次分析都会生成一个预测结果。这些预测结果可能有所不同，因此，通过投票机制来确定最终的结果变得非常重要。

多轮投票方法的基本思路是，通过多次分析同一组对话片段，累积各次分析结果中的多数意见，最终以得票最多的结果作为最终的预测结果。这种方法可以有效地减少个别异常结果对最终判断的影响，提高预测的稳定性和准确性。

具体操作上，本文对每一组对话片段进行  $m$  轮分析， $m$  为奇数。在每一轮中，大语言模型都会生成一个关系标签预测。然后，将所有  $m$  轮的预测结果汇总，统计每个标签的出现频次。通过这种方式，可以识别出在多次分析中出现频率最高的标签，并将其作为最终的关系标签。这一过程类似民主投票，通过多次投票来确保最终决策的可靠性和一致性。

例如，对图7中对话记录进行分析可得出，霞某开取药物并出售给英某，因此二者是开药关系。然而，由于英某多次提及“开”字，大语言模型可能误认为英某是开药人，而霞某从英某处获取药品。并且根据之前的关系定义，将二者的关系误定义为收药关系。因此，需要进行多轮询问，以减少误判带来的准确率影响。

通过这种多轮投票机制，本文在对话分析中不仅提高了关系提取结果的稳定性，还增强了对异常情况的抵抗力，确保最终的关系标签更可靠。这种方法有效地利用了大语言模型的多次分析能力，将其不稳定性转化为稳定、准确的关系提取结果，从而提升了关系提取结果的整体性能和可信度。

### 3 评估分析

#### 3.1 环境与设置

##### 3.1.1 数据集

本文使用的数据集来源于北京市石景山区人民检察院，包含9名嫌疑人和12部手机的取证数据，共计超过98 000条个人聊天记录。从检察业务视角，选择了具有代表性的嫌疑人李某和闻某，将其手机取证数据作为实验数据集，实验数据集信息见表2。二人的微信好友数多，聊天记录数量大，具有较高的业务价值。

表2 实验数据集信息

机主	微信好友/个	聊天记录/条	嫌疑人/个	普通人/个
李某	69	8 419	54	15
闻某	40	7 642	21	19

##### 3.1.2 实验环境

实验环境见表3。本文医保欺诈嫌疑人关系提取实验所采用的大语言模型为QWEN-72B和Chat-GLM3等本地部署的模型，大语言模型API参数设置见表4。其中， $n$ 表示大语言模型的回答数量， $temperature$ 用于调整随机从生成模型中抽样的程度，将其设置为0能够有效地降低模型生成结果的随机性。

##### 3.1.3 基线模型

本文选取基线模型的标准从任务适配性、模型的简洁性与有效性、上下文建模能力等方面考虑，选取两种基线模型。本文方法的模型名称为SRE。

1) Albert<sup>[28]</sup>：Albert是一种轻量级的预训练语

表3 实验环境

名称	配置信息
操作系统	Ubuntu 20.04
开发语言	Python3.10+JAVA
CPU	Intel® Xeon® Silver 4214R
GPU	2*NVIDIA GeForce RTX 3090
开发环境	CUDA: 12.2

表4 大语言模型API参数设置

模型	$n$ (生成回答数量)	Temperature(回答随机度)
QWEN-72B	1	0
ChatGLM3	1	0

言模型，通过分解参数矩阵和跨层参数共享机制显著地减少了模型参数量，同时保持了良好的语义建模能力。

2) BERT+GRU<sup>[29]</sup>: BERT+GRU方法通过BERT提取对话中的上下文语义信息，再结合GRU处理对话序列的时间特征，从而提高对对话关系建模的能力。

3) SRE-ChatGLM3: 该模型是本文方法以ChatGLM3为基座的模型。ChatGLM3是一个基于GLM架构的自然语言处理模型，能用于各种文本生成任务。

4) SRE-QWEN-72B: 该模型是本文方法以QWEN-72B为基座的模型。QWEN-72B是基于QWEN架构的大规模预训练语言模型，采用了深度学习技术，能够处理包括文本生成、对话系统、问答、情感分析等多种自然语言处理(NLP)任务。

### 3.1.4 评价指标

可以将嫌疑人关系提取任务视为多分类任务。为了评估所提方法的有效性，本文使用精确率(precision)、召回率(recall)和F1分数作为评价指标。在关系提取任务中，精确率衡量的是方法在预测某种关系的准确性；召回率反映了模型在多大程度上能够覆盖所有的真实关系，高召回率意味着模型能发现大部分的真实药品交易关系，漏报率低；F1分数提供了一种权衡精确率和召回率的方法。一个模型可能具有很高的精确率但低召回率，或者高召回率但低精确率。F1分数在这两者之间找到平衡点，给出一个整体的性能评价。后续实验重点关注以下5个问题。

1) 与传统方法的性能相比如何?(RQ1)

2) 在不同手机取证数据上的健壮性表现如何?(RQ2)

3) 不同提示下方法的性能表现如何?(RQ3)

4) 聊天记录中图片和音频数据对提取结果的影响?(RQ4)

5)  $N$ -shot示例对方法性能的影响?(RQ5)

## 3.2 实验分析

在本文的实验分析中，考虑QWEN-72B相较于ChatGLM3的表现更突出，本文选择将QWEN-72B作为主要实验模型，并分别在李某和闻某的数据上进行实验对比，以验证本方法的健壮性。与此同时，SRE-ChatGLM3主要用于基座模型的对比实验，以确保实验重点集中在模型的核心性能对比上。因此，闻某的手机取证数据主要用于关系提取性能实验，旨在评估本方法在不同数据集上的健壮性，而其余实验所使用的数据集均为李某的手机取证数据。

### 3.2.1 关系提取性能

将本文提出的嫌疑人关系提取模型与Albert、BERT+GRU等传统方法分别在李某和闻某的手机取证数据上进行实验对比，以探讨本文模型在不同手机取证数据上的健壮性，从而回答RQ1和RQ2。使用不同模型在李某和闻某的数据集上进行嫌疑人关系提取实验，关系提取性能结果见表5。传统方法由于无法有效地捕捉聊天记录中的欺诈指示词，表现较差。而本文所提的嫌疑人关系提取模型能够有效地捕捉涉及开药、收药的关键语句，因此效果较好。

同时可以看出，QWEN-72B的表现明显优于其他大语言模型，如ChatGLM3。这是因为QWEN-72B的训练参数远多于ChatGLM3，能够更有效地从给定的聊天示例中学习得到医保欺诈的聊天特征，从而在关系提取任务中表现出色。

从表5可以看出，本文提出的嫌疑人关系提取模型在两个手机取证数据上的准确率几乎没有变化，表现出高度的稳定性。而Albert、BERT+GRU等传统方法的准确率则均出现了不同程度的下降。这是因为传统方法在训练过程中，通常依赖于特定的数据集进行优化。这意味着它们在面对与训练集有显著差异的新数据时，缺乏足够的泛化能力。李某和闻某具有不同的语言风格、交流习惯和语境，导致传统方法难以适应这些变化，从而导致准确率下降。同时也表明，本文模型在不同数据集上的健

表5 关系提取性能结果

模型	李某			闻某		
	精确率	召回率	F1	精确率	召回率	F1
Albert	0.818	0.642	0.720	0.698	0.653	0.674
BERT+GRU	0.782	0.680	0.727	0.702	0.643	0.670
SRE-ChatGLM3	0.834	0.702	0.762	—	—	—
SRE-QWEN-72B	<b>0.975</b>	<b>0.829</b>	<b>0.896</b>	<b>0.985</b>	<b>0.759</b>	<b>0.863</b>

壮性明显优于传统方法。

### 3.2.2 消融实验

为了回答 RQ3 和 RQ4，本文进行了系列消融实验，分析大语言模型提示文本中不同部分及图片、音频数据对关系提取效果的影响。不同提示策略的消融实验结果见表6，详细介绍了 SRE-QWEN-72B 和 SRE-ChatGLM3 在不同提示下的关系提取任务性能评估，其中，少样本提示包括3个示例，BaselinePrompt 为标准提示词模板，仅包含简要任务描述和答案输出格式。(1)、(2)、(3)分别为带有任务描述的基线提示、基于欺诈指示词的提示和少样本提示。结果显示，集成了带有任务描述的基线提示、基于欺诈指示词的提示和少样本提示的指令后，两个模型的性能指标均有所改善，且这3个组件对 SRE-QWEN-72B 的影响比对 SRE-ChatGLM3 的影响更显著。这表明，大语言模型训练参数越多，从提示中学习的越好。

表6 不同提示策略的消融实验结果

模型	提示策略	精确率	召回率	F1
SRE-ChatGLM3	基本提示词	0.736 8	0.595 7	0.658 8
	(1)+(2)	0.571 4	0.666 6	0.612 0
	(1)+(2)+(3)	0.804 9	0.702 1	0.750 0
SRE-QWEN-72B	基本提示词	0.891 9	0.702 1	0.785 7
	(1)+(2)	0.944 4	0.739 1	0.829 3
	(1)+(2)+(3)	<b>0.975 0</b>	<b>0.829 8</b>	<b>0.896 6</b>

为研究音频和图片数据对关系提取结果的影响，首先去除聊天记录中的音频和图片数据，然后将处理后的文本输入 SRE-QWEN-72B 进行实验。多模态输入的消融实验见表7，去除音频和图片数据后，关系提取效果明显下降。由于部分微信用户多通过拍照和语音进行沟通，去除这些数据后，聊天记录中的关键信息减少，导致大语言模型难以识别出开药或收药的关系。

本文在上述实验的基础上，修改少样本提示中的示例数量，分析样本数量的不同是否会造成影

表7 多模态输入的消融实验

数据集	精确率	召回率	F1
包含音频与图片	<b>0.975 0</b>	<b>0.829 8</b>	<b>0.896 6</b>
不包含音频与图片	0.906 2	0.537 0	0.674 4

响，从而回答 RQ5。样本数量的消融实验结果见表8，给出了不同数量的样本示例之间的性能，该实验中包括所有的提示组件。通常，包含更多的示例会使关系提取任务的效果更好，3-shot 和所有提示的组合产生了 SRE-QWEN-72B 的最佳结果。SRE-ChatGLM3 在使用 1-shot 的时候，召回率出现了降低，但 SRE-QWEN-72B 却是增加的，这是因为 SRE-ChatGLM3 只包含 6 亿个参数<sup>[30]</sup>，但是 SRE-QWEN-72B 的参数量为 175 亿，所以当只有一个示例的时候，SRE-ChatGLM3 不能总结出示例的对话特点，当示例数量增加的时候，SRE-ChatGLM3 能够从多个示例中总结出医保欺诈类型的即时通信数据的对话特点，所以才会出现示例增加到3个的时候，SRE-ChatGLM3 的召回率指数上升。

表8 样本数量的消融实验结果

模型	提示策略	精确率	召回率	F1
SRE-ChatGLM3	(1)+(2)+0-shot	0.571 4	0.666 6	0.612 0
	(1)+(2)+1-shot	0.658 5	0.587 0	0.620 7
	(1)+(2)+3-shot	0.804 9	0.702 1	0.750 0
SRE-QWEN-72B	(1)+(2)+0-shot	0.944 4	0.739 1	0.829 3
	(1)+(2)+1-shot	0.837 2	0.782 6	0.809 0
	(1)+(2)+3-shot	<b>0.975 0</b>	<b>0.829 8</b>	<b>0.896 6</b>

### 3.3 案例分析

本文所提的嫌疑人关系提取方法已在北京市石景山区人民检察院的医保欺诈监督业务中得到应用。该方法通过分析手机取证数据的微信聊天记录，将微信好友分类为开药人、收药人和普通人，并以图形化网络展示，医保欺诈嫌疑人识别的展示如图9所示。在图中，黄色节点代表被控制的收药人，紫色节点表示被判定为开药人，粉色节点则代表被判定为收药人。

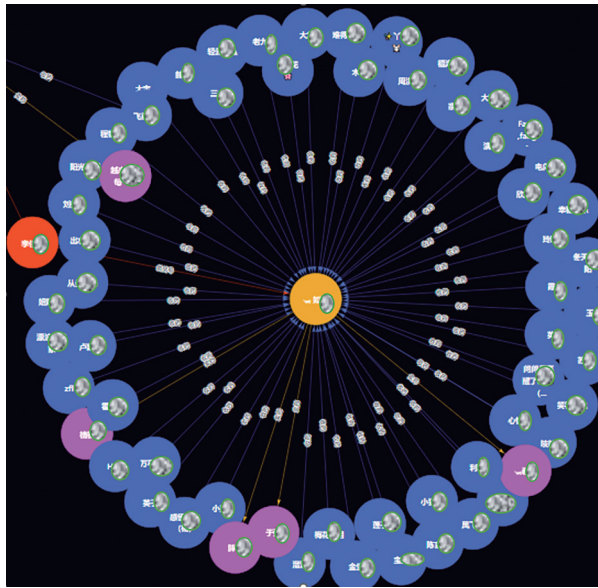


图9 医保欺诈嫌疑人识别的展示

这种关系网络图不仅直观地展示了参与医保欺诈的各方之间的复杂交互，还帮助揭示了在识别欺诈行为时可能遇到的问题。例如，节点之间可能存在多重关系，一个人既是开药人，又在其他时段作为收药人参与交易，这种交叉关系需要精细的分析来区分不同的角色和行为。此外，网络中的节点之间的连接模式和频率也能反映出不同的关系强度和交易频率，进一步地深化了对嫌疑人之间关系的理解。在实际的手机取证数据集上的测试结果显示，该方法的嫌疑人关系提取准确率达到了97.50%。

### 4 结束语

在医保欺诈监督领域，微信聊天记录的多模态异构性和嫌疑人交流习惯的差异性对传统方法提出了较大的挑战。针对这些问题，本文提出了一种创新的解决方案，利用社交物联网的思想，将嫌疑人的手机作为数据感知的关键节点，收集了包括微信聊天记录在内的多模态数据，并通过大语言模型进行分析。与传统方法相比，本文所提方法不仅能有效地处理不同数据源之间的异构性，还能弥补不同嫌疑人的交流习惯差异带来的困扰。

具体而言，本文提出了一种基于大语言模型的嫌疑人关系提取方法，通过精心设计的提示词框架来引导模型从聊天记录中提取具有价值的嫌疑人关系信息。这一方法有效地解决了传统模型在面对复杂聊天内容时准确性低、健壮性差的问题，尤其在处理含糊或复杂情境中的对话时，表现出更强的鲁

棒性和灵活性。

在实际项目中，本文所提方法在不同数据集上的实验结果表明，相较于传统基线模型，显著地提高了嫌疑人关系提取的精度，准确率达到了97.50%。这一成果不仅为医保欺诈监督提供了强有力的数据支持，也展示了大语言模型在实际法律案件分析中的广泛应用前景。

本文方法目前仅针对微信单聊数据，后续将会针对更全面的数据进行分析，增强关系提取的能力。

### 参考文献：

- [1] CUI H Y, LI Q Z, LI H, et al. Healthcare fraud detection based on trustworthiness of doctors[C]//Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA. Piscataway: IEEE Press, 2017: 74-81.
- [2] 弭宝瞳, 梁循, 张树森. 社交物联网研究综述[J]. 计算机学报, 2018, 41(7): 1448-1475.
- [3] MI B T, LIANG X, ZHANG S S. A survey on social Internet of things[J]. Chinese Journal of Computers, 2018, 41(7): 1448-1475.
- [4] VAIBHAVA LAKSHMI R, DEEPAK G, SANTHANAVIJAYAN A, et al. Search for social smart objects constituting sensor ontology, social IoT and social network interaction[C]//Proceedings of the 2022 6th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). Piscataway: IEEE Press, 2022: 60-65.
- [5] DHELMIM S, NING H S, FARHA F, et al. IoT-enabled social relationships meet artificial social intelligence[J]. IEEE Internet of Things Journal, 2021, 8(24): 17817-17828.
- [6] AFZAL B, UMAIR M, ASADULLAH SHAH G, et al. Enabling IoT platforms for social IoT applications: vision, feature mapping, and challenges[J]. Future Generation Computer Systems, 2019, 92: 718-731.
- [7] SHARAN B, HASSAN M, VANI V D, et al. Machine learning-based fraud detection system for insurance claims in IoT environment[C]//Proceedings of the 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). Piscataway: IEEE Press, 2024: 1-5.
- [8] ISLAM PROVA N N. Healthcare fraud detection using machine learning[C]//Proceedings of the 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI). Piscataway: IEEE Press, 2024: 1119-1123.
- [9] HE H Y, DING W L, ZHANG H, et al. A LLMs-based procuratorial service for detecting drug trafficking on digital forensics data[C]//Proceedings of the Service-Oriented Computing - ICSOC 2024 Workshops. Singapore: Springer, 2026: 254-258.
- [9] YSTGAARD K F, ATZORI L, PALMA D, et al. Review of the theory, principles, and design requirements of human-centric Internet of Things (IoT)[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(3): 2827-2859.

- [11] 宁尚明, 滕飞, 李天瑞. 基于多通道自注意力机制的电子病历实体关系抽取[J]. 计算机学报, 2020, 43(5): 916-929.  
NING S M, TENG F, LI T R. Multi-channel self-attention mechanism for relation extraction in clinical records[J]. Chinese Journal of Computers, 2020, 43(5): 916-929.
- [12] AGRAWAL M, HEGSELMANN S, LANG H, et al. Large language models are few-shot clinical information extractors[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: ACL, 2022: 1998-2022.
- [13] WADHWA S, AMIR S, WALLACE B. Revisiting relation extraction in the era of large language models[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA, USA: ACL, 2023: 15566-15589.
- [14] WEI J Q, ZHU W J, ZHU C, et al. AxialRE: axial attention for dialogue relation extraction[C]//Proceedings of the 2023 International Conference on Intelligent Management and Software Engineering (IMSE). Piscataway: IEEE Press, 2024: 25-29.
- [15] WANG Y T, PAN Y H, YAN M, et al. A survey on ChatGPT: AI-generated contents, challenges, and solutions[J]. IEEE Open Journal of the Computer Society, 2023, 4: 280-302.
- [16] FANG F, HU X G, SHU J H, et al. Text classification model based on multi-head self-attention mechanism and BiGRU[C]//Proceedings of the 2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS). Piscataway: IEEE Press, 2022: 357-361.
- [17] ZHOU M J, JI D H, LI F. Relation extraction in dialogues: a deep learning model based on the generality and specialty of dialogue text[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 2015-2026.
- [18] ZHAO L L, XU W R, GAO S, et al. Utilizing graph neural networks to improving dialogue-based relation extraction[J]. Neurocomputing, 2021, 456: 299-311.
- [19] ZHANG S Z, YE J Y, WANG Q X. Spa-L Transformer: sparse-self attention model of long short-term memory positional encoding based on long text classification[C]//Proceedings of the 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Piscataway: IEEE Press, 2023: 618-623.
- [20] GUO M H, WU F, JIANG J L, et al. Investigations on scientific literature meta information extraction using large language models[C]//Proceedings of the 2023 IEEE International Conference on Knowledge Graph (ICKG). Piscataway: IEEE Press, 2024: 249-254.
- [21] WANG Z Y, ZHOU Q, ZHAO J F, et al. A knowledge-enhanced medical named entity recognition method that integrates pre-trained language models[C]//Proceedings of the 2023 IEEE International Conference on Medical Artificial Intelligence (MedAI). Piscataway: IEEE Press, 2024: 296-301.
- [22] XUE F Z, SUN A X, ZHANG H, et al. An embarrassingly simple model for dialogue relation extraction[C]//Proceedings of the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 6707-6711.
- [23] WU G Y, WU W J, LIU X H, et al. Cheap-fake detection with LLM using prompt engineering[C]//Proceedings of the 2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). Piscataway: IEEE Press, 2023: 105-109.
- [24] NAIMI L, BOUZIANE E M, MANAOUCH M, et al. A new approach for automatic test case generation from use case diagram using LLMs and prompt engineering[C]//Proceedings of the 2024 International Conference on Circuit, Systems and Communication (ICCSC). Piscataway: IEEE Press, 2024: 1-5.
- [25] LIU X, JI K X, FU Y C, et al. P-tuning: prompt tuning can be comparable to fine-tuning across scales and tasks[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Stroudsburg, PA, USA: ACL, 2022: 61-68.
- [26] NORHEIM J J, REBENTISCH E. Structuring natural language requirements with large language models[C]//Proceedings of the 2024 IEEE 32nd International Requirements Engineering Conference Workshops (REW). Piscataway: IEEE Press, 2024: 68-71.
- [27] ZHANG C Y, LIU H, ZENG J T, et al. Prompt-enhanced software vulnerability detection using ChatGPT[C]//Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion). Piscataway: IEEE Press, 2024: 276-277.
- [28] HU Y, CHEN Q Y, DU J C, et al. Improving large language models for clinical named entity recognition via prompt engineering[J]. Journal of the American Medical Informatics Association, 2024, 31(9): 1812-1820.
- [29] LAN Z Z, CHEN M D, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations[EB]. arXiv preprint, 2019, arXiv: 1909.11942.
- [30] 赵丹丹, 黄德根, 孟佳娜, 等. 基于BERT-GRU-ATT模型的中文实体关系分类[J]. 计算机科学, 2022, 49(6): 319-325.  
ZHAO D D, HUANG D G, MENG J N, et al. Chinese entity relations classification based on BERT-GRU-ATT[J]. Computer Science, 2022, 49(6): 319-325.
- [31] YANG F, XU S, LI P F, et al. Prompt-based Chinese event temporal relation extraction on LLM predictive information[C]//Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2024: 1-8.

#### [作者简介]



杨坤峰(2000-), 男, 北方工业大学信息学院硕士生, 主要研究方向为大语言模型、大数据。



丁维龙(1983-), 男, 博士, 北方工业大学信息学院副教授, 主要研究方向为时空大数据、多模态大数据人工智能技术和服务计算。



王博(1987-), 男, 北京市石景山区人民检察院三级主任科员, 主要研究方向为法律影像表征、意图分析等。



田涵(1998-), 男, 北方工业大学信息学院硕士生, 主要研究方向为大数据、 workflow 技术、人工智能。



赵广静(1982-), 女, 北京市石景山区人民检察院四级高级检察官, 主要研究方向为刑事检察等。



李阳(1988-), 女, 北京市石景山区人民检察院四级高级检察官, 主要研究方向为刑事检察。



王晓洁(1991-), 女, 北京市石景山区人民检察院检察官助理, 主要研究方向为法学。