

中文引用格式:杨艺,王东麟,陈振颂. 基于大语言模型智能体的汽车安全风险判别:以投诉文本为驱动[J]. 中国安全科学学报, 2025, 35(12):204-212.

英文引用格式:YANG Yi, WANG Donglin, CHEN Zhensong. Research on automobile safety risk discrimination utilizing LLMs-based agents: driven by complaint text [J]. China Safety Science Journal, 2025, 35(12):204-212.

基于大语言模型智能体的汽车安全风险判别: 以投诉文本为驱动*

杨艺^{1,2}副教授, 王东麟¹, 陈振颂^{**3}副教授

(1 湖南工商大学 前沿交叉学院, 湖南 长沙 410205; 2 湘江实验室, 湖南 长沙 410205;

3. 武汉大学 土木建筑工程学院, 湖北 武汉 430072)

中图分类号:X913

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2025.12.0065

资助项目:国家自然科学基金资助(72201097);湘江实验室重大项目(25XJ01001;24XJJCYY01);湖南省青年科技人才计划基金资助(2023RC3182);湖南省研究生科研创新项目(LXBZZ2024344)。

【摘要】 为解决汽车安全风险管理中依赖人工提取投诉文本知识的局限性,应用大语言模型(LLMs)智能判别事件风险。首先,采集包含发动机等8大子系统的5万余条投诉文本,基于双语与跨语言嵌入(BCEmbedding)模型与社区发现算法,提出演示采样方法,构建多样化高质量示例集知识库;然后,从场景、技能以及示例等方面设计LLMs提示词,构建实现风险词抽取与扩充、子系统风险归类的智能体;最后,分析召回文本的推理知识,运用思维链技术构建召回文本风险等级知识库,设计基于共识寻求的LLMs多智能体(MA)系统,建立汽车安全风险等级判别模型。结果表明:该模型在节约人工成本的同时具有较好的准确性和效率,能够有效支持投诉事件中的风险词抽取、风险归类和风险等级判别,提升安全风险管理水平。

【关键词】 大语言模型(LLMs); 智能体; 汽车安全风险; 风险判别; 投诉文本; 提示词

Research on automobile safety risk discrimination utilizing LLMs-based agents: driven by complaint text

YANG Yi^{1,2}, WANG Donglin¹, CHEN Zhensong³

(1 School of Advanced Interdisciplinary Studies, Hunan University of Technology and Business, Changsha Hunan 410205, China; 2 Xiangjiang Laboratory, Changsha Hunan 410205, China;

3 School of Civil Engineering, Wuhan University, Wuhan Hubei 430072, China)

Abstract: To address the limitations of relying on manual extraction of knowledge from complaint texts in automotive safety risk management, this study employed LLMs for automated risk discrimination. First, over 50 000 complaint texts covering eight major subsystems such as the engine were collected. A demonstration sampling method based on Bilingual and Crosslingual Embedding (BCEmbedding) model and community detection algorithm was proposed to construct a diverse and high-quality example knowledge base. Secondly, prompts were designed from perspectives such as scenarios, skills, and examples to develop agents capable of performing risk keyword extraction and expansion, as well as subsystem risk categorization. Finally, by analyzing reasoning knowledge from retrieved texts and utilizing

* 文章编号:1003-3033(2025)12-0204-09; 收稿日期:2025-06-16; 修稿日期:2025-09-29

** 通信作者:陈振颂(1988—),男,福建三明人,博士,副教授,主要从事博弈论与决策分析研究。E-mail: zschen@whu.edu.cn。

chain of thought techniques, a risk level knowledge base for retrieved texts was established. A consensus-seeking multi-agent(MA) system based on LLMs was designed, resulting in a discriminative model for automotive safety risk levels. The results show that the model not only reduces labor costs but also achieves high accuracy and efficiency. It effectively supports risk term extraction, risk categorization, and risk level assessment in complaint incidents, thereby enhancing safety risk management.

Keywords: large language models (LLMs); agent; automobile safety risk; risk discrimination; complaint text; prompt

0 引言

当前,在人工智能领域的研究进展中,对话生成式预训练变换器 (Generative Pre-trained Transformer, ChatGPT) 等大语言模型 (Large Language Models, LLMs)因其卓越的语言理解、涌现性及生成能力,已成为学术界和产业界关注的焦点^[1]。LLMs 对信息抽取技术产生了革命性的影响,其能够从非结构化文本中高效、精确地自动提取结构化知识^[2]。另一方面,基于 LLMs 的多智能体 (LLMs-based Multi-Agent, LLMs-MA) 在推理、规划、反思、记忆等智能行为方面展现出显著的优势^[3],能够在复杂问题解决和现实世界环境模拟等领域发挥重要作用。

LLMs 在信息抽取、风险辨识以及安全事件模拟等维度展现出广阔的应用潜力。在信息抽取领域,学者们通过微调 LLMs 并结合提示工程技术,实现了从科学文档中提取结构化数据的流程化操作^[4-5],证实了 LLMs 在知识抽取中的实际应用价值。在安全领域,学者们运用 LLMs 深度挖掘安全事故报告,以识别事故诱因,并据此构建基于事故知识图谱和知识库的问答模型^[6-9],为安全事故管理提供了科学的数据分析与决策信息支持。目前,学者们以汽车在线平台数据为对象,依托监督式主题模型、词频-逆文档频率算法等人工智能模型,结合 Bert 等文本分析算法,识别和改进质量问题,为制造商提供质量优化策略^[10-14]。上述研究主要将人工智能应用于信息挖掘与分析,依赖于高成本的数据标注,在效率和精度方面也有待提升。

鉴于此,笔者拟综合考虑构建 LLMs-MA 提取结构信息的并识别汽车在线投诉数据风险程度,结合 LLMs-MA 的信息抽取、文本分析等强大性能,分析汽车在线投诉数据构建知识库,构建基于 LLMs 的质量风险等级判别模型,并提出共识集成方法,以期在投诉数据分析、风险管理和优化服务等方面改进系统性,更好地防范汽车安全风险,提升汽车安全领

域分析、研判和管理的水平。

1 基于聚类的演示采样

1.1 数据准备

从车质网等平台爬取汽车投诉数据作为评估依据,聚焦投诉文本和典型质量问题,共获取 52 300 条投诉记录。通过检查网页 Headers 识别关键统一资源定位符 (Uniform Resource Locator, URL) 参数,借助 requests 库模拟请求,使用正则表达式 (Regular Expressions, RE) 提取内容,存储为结构化数据,包括品牌、问题、投诉文本、风险词与风险类别。

首先,以《汽车产品安全—风险评估与风险控制指南》^[15]中的相关要求为依据,参考严重性等级说明来设计具体的风险程度评判提示词。

1.2 演示采样

构建汽车安全风险 (含发动机/电动机、变速器、离合器、转向系统、制动系统、轮胎、前后桥及悬挂系统、车身附件及电器 8 大子系统的问题和风险词)和召回文本 (含原因和可能导致的结果) 知识库。基于聚类的文本划分结果和演示采样框架分别如图 1 和图 2 所示。

为了更好地处理涉及多类质量问题的文本,并考虑到部分文本差异或同质影响,将文本进行聚类。首先,运用双语与跨语言嵌入 (Bilingual and Crosslingual Embedding, BCEmbedding) 模型实现文本向量化^[16];然后,执行文本的向量相似度计算,基于相似度值分布特征设立阈值,构建以相似度为关系的无向网络;最后,运用 Louvain 算法实现文本划分。

采用余弦相似度 S 计算文本的向量相似度,公式为:

$$S(X,Y)=\frac{X\cdot Y}{\|X\|\|Y\|}=\frac{\sum_{i=1}^zX_i\cdot Y_i}{\sqrt{\sum_{i=1}^zX_i^2}\cdot\sqrt{\sum_{i=1}^zY_i^2}}\tag{1}$$

式中: X 和 Y 为基于开源 BCEmbedding 模型下得到

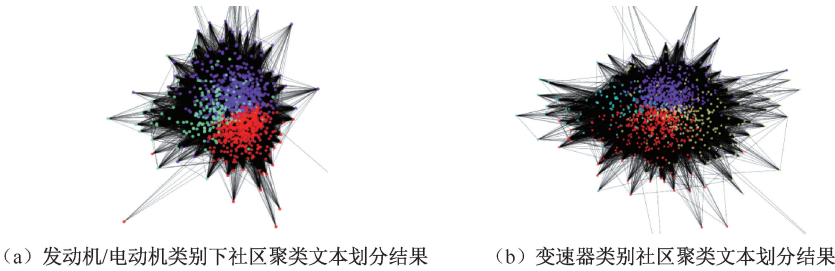


图1 社区聚类的演示采样结果

Fig. 1 Demonstration sampling results of community clustering

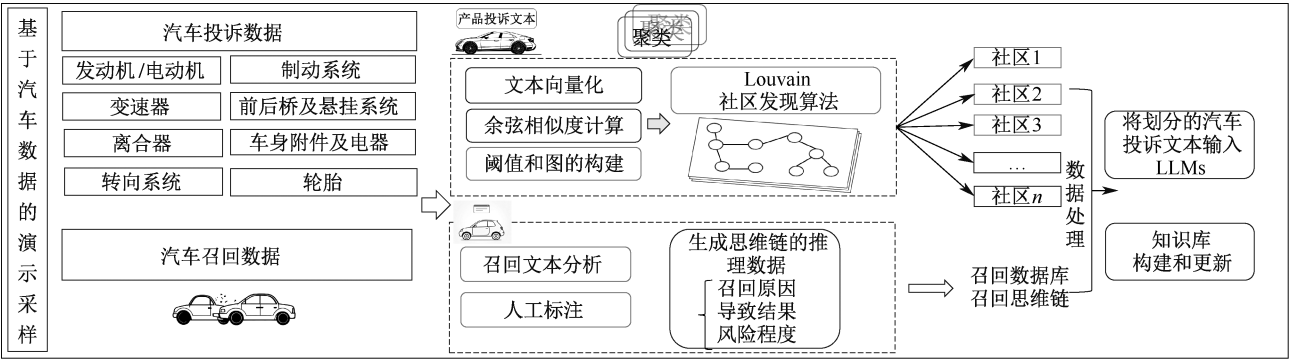


图2 基于聚类的汽车文本演示采样

Fig. 2 Clustering based sampling of automotive text demonstrations

的文本向量; X_i 和 Y_i 为文本向量 X 和 Y 的第 i 个元素; z 为文本向量的维度。

上述 Louvain 算法是一种社区发现算法,通过迭代优化最大化模块度来识别社区结构,无需预先设定社区数量,可以自行确定最佳的聚类结果。模块度 Q 是衡量社区划分质量的指标,公式为^[17]:

$$Q = \frac{1}{2m} \left(\sum_{ij} \vartheta_{ij} - \frac{1}{2m} \sum_i k_i \sum_j k_j \right) \delta(e_i, e_j) = \frac{1}{2m} \sum_e \left(\alpha - \frac{1}{2m} \beta^2 \right) \quad (2)$$

其中, $\delta(e_i, e_j) = \begin{cases} 1 & \text{节点 } i \text{ 和节点 } j \text{ 属于同一社区} \\ 0 & \text{否则} \end{cases}$

式中: ϑ_{ij} 为节点 i 和节点 j 之间边的权重,在网络不是带权图时,所有边的权重为 1; m 为所有边的权重之和(边的数目); k_i 为所有与节点 i 相连的边的权重之和; e_i 为节点 i 所属的社区; α 为社区 e 内的边的权重之和; β 为与社区 e 内的节点相连的边的权重之和,包括社区内的边和社区外部的边; δ 为指示函数,节点 i 和节点 j 同一社区时取值为 1,否则为 0。

进一步来说,Louvain 算法定义了模块度的增益来衡量将节点移动到相邻社区后模块度的增加量。该增益标准常用于决定是否应该进行移动,即把节点 i 分配到邻居节点 j 所在的社区 e 时模块度增益为^[17]:

$$\Delta Q = \left[\frac{\alpha + k_{i,\alpha}}{2m} - \left(\frac{\beta + k_i}{2m} \right)^2 \right] - \left[\frac{\alpha}{2m} - \left(\frac{\beta}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (3)$$

式中 $k_{i,\alpha}$ 为社区 e 内所有与节点 i 相连的边的权重之和。

文中以发动机/电动机这一质量问题类别为例,通过聚类采样以及提取的主题词语来具体示例分析,部分结果见表 1。

表1 发动机/电动机聚类类别特征示例

Table 1 Example of clustering category features for engine/electric motor

社区聚类质量问题类别	社区划分	投诉文本问题主题词语统计	投诉文本平均字数
发动机/电动机	1	发动机故障,4S 店,更换,保养,售后	238.47
	2	发动机故障,行驶,异响,机油	95.21
	3	模式,质量,无法启动	156.13

接着,结合知识库训练和社区聚类的结果,将划分出高质量的投诉文本,输入 LLMs 中采样分析,划分出每一类中代表性的汽车投诉文本进行后续处理。图 1 所示的 2 个聚类批次下,每一不同颜色的点表示投诉文本被划分的不同社区聚类类别。

2 基于 LLMs 的质量信息抽取智能体构建

2.1 知识库构建

文中汽车投诉文本来源于消费者在平台上的反馈,涵盖各种车型和故障类型。需要知识库的构建和更新来提升信息抽取能力,步骤如下:

- 1) 数据处理。包括特征分析、向量化、相似度计算与社区聚类等,形成高质量数据集。
- 2) 知识库构建。依据投诉数据中的实体关系与属性进行划分和提取,并存入知识库。
- 3) 知识库优化与更新。运用 LLMs 提取类别与风险词等数据,通过智能体或工作流机制持续扩充知识库,实现动态优化。

因此,为解决 LLMs 幻觉等问题,文中通过高质

量知识库支撑 LLMs 实现汽车质量信息抽取,文中的知识库通过智谱清言、文心一言和通义千问等 LLMs 综合建立。

2.2 提示词设计

为克服训练语料的不足,减少 LLMs 幻觉,确保回答的真实和准确,采用提示工程引入相关的知识片段,并将预测目标与相关内容结合,以辅助提取信息。相较于模型微调,提示工程成本低且易操作。文中的提示设计,通过将预设角色和规则知识注入信息抽取及风险识别的场景来实现^[8]。因此,文中对 LLMs 智能体引入知识库学习,接着根据汽车投诉数据进行提示工程,将其输入 LLMs 智能体中进行处理,并根据提示词要求结构化得到输出结果,主要信息提取和风险识别的提示模板见表 2。

表 2 提示词设计

Table 2 Design of prompt

场景	你是专业的汽车投诉文本风险信息抽取智能体,你会接收 1 个输入,结合{知识库}并根据你所定义的技能进行处理。{{input}} 这个是用户输入需要标注的文本
技能	技能 1:读取文本。读取{{input}}输入的文本,设置存储变量{{text}},用于存储文本内容; 技能 2:系统学习和识别知识库内容。即在每次的执行中必须结合知识库(风险词、召回原因等)内容{{zsk, recall_zsk}}来进行风险词提取等任务; 技能 3:风险词提取与存储。设置风险词变量{{risk_word}},提取文本中每个风险词,可能有多,每个单独存储,多个风险词构成 1 列。每个文本的存储至 1 个数组,多个文本的数组存储到 1 个大数组中; 技能 4:风险等级标注。设置风险等级变量{{risk_level}},根据汽车安全风险等级规范的 5 类风险等级的标准,判定{{risk_word}}中每个风险词的风险等级
JavaScript 对象表示法 (JavaScript Object Notation, JSON)输出	输出结果严格按照以下格式输出: { "text": "(文本)", "risk_word": "(风险词)", "risk_level": "(风险等级)", "category": "(归属类别)" }
信息提取示例	问:正常行驶动力突然中断,无法正常加速。 输出结果:{ "text": ["正常行驶动力突然中断,无法正常加速。"], "risk_word": ["动力中断", "无法加速"], "risk_level": ["5", "4"], "category": ["发动机/电动机", "发动机/电动机"] }

2.3 智能体构建

构建基于 LLMs 的质量信息抽取智能体时,首先,对 LLMs 定制化训练,使其能够理解汽车行业的专业术语和投诉语境;其次,通过提示词设计信息抽取框架;并规定模型输出 JSON 格式;最后,将智能体部署集成(如通过应用程序编程接口(Application Programming Interface, API)调用部署)到现有的数据处理系统中使其能自动化地处理投诉文本,并输出结构化的信息报告。通过上述过程,可得到能够有效服务于汽车风险质量识别和信息提取的智能体,其主要框架设计具体如图 3 所示。

3 基于 LLMs-MA 的风险等级判别模型

3.1 基于思维链技术的提示工程

思维链(Chain of Thought, CoT)通过引导 LLMs 将复杂推理分解为多个中间步骤,显式展示其逻辑推理过程,从而提升模型在复杂任务中的思维性能。与传统的基于(输入,输出)二元组的上下文学习不同,CoT 通过插入中间推理环节,构建从输入到输出的完整逻辑链。文中基于召回文本的 CoT 设计如图 4 所示。

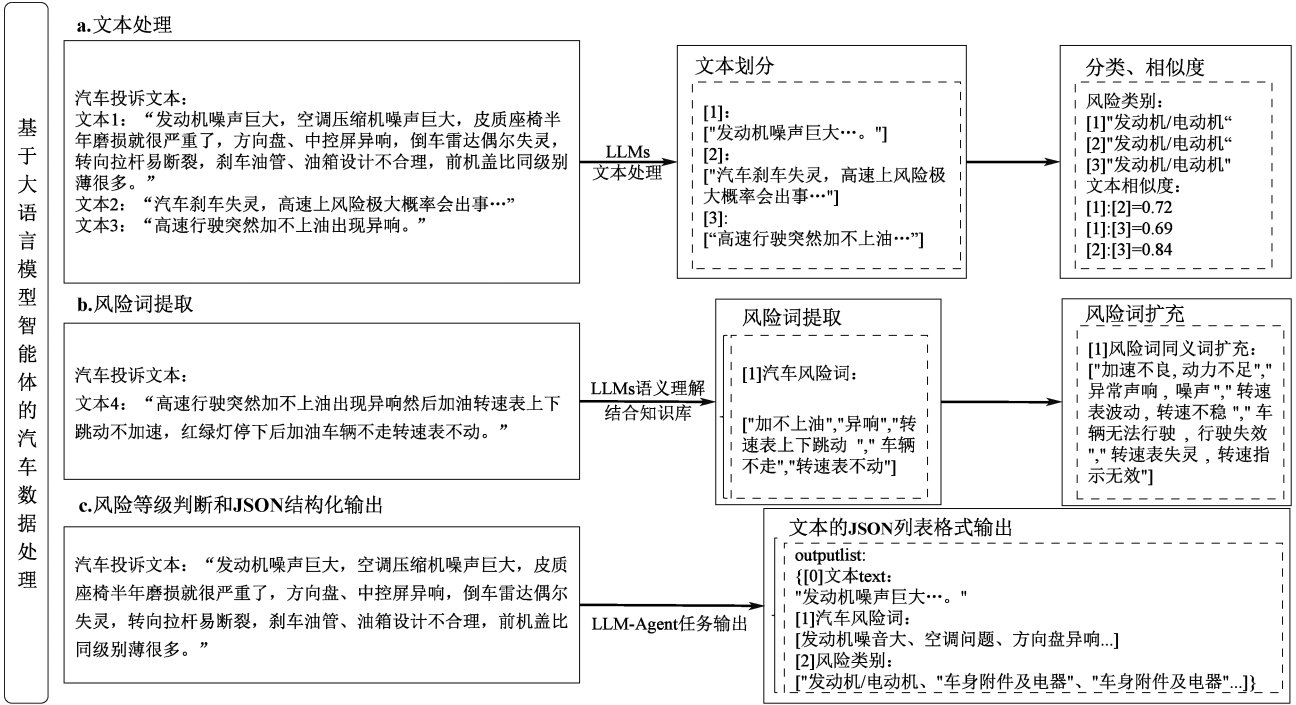


图3 基于LLMs智能体的结构化信息抽取(风险词知识增强)

Fig. 3 Structured information extraction based on LLMs agent (risk word knowledge enhancement)

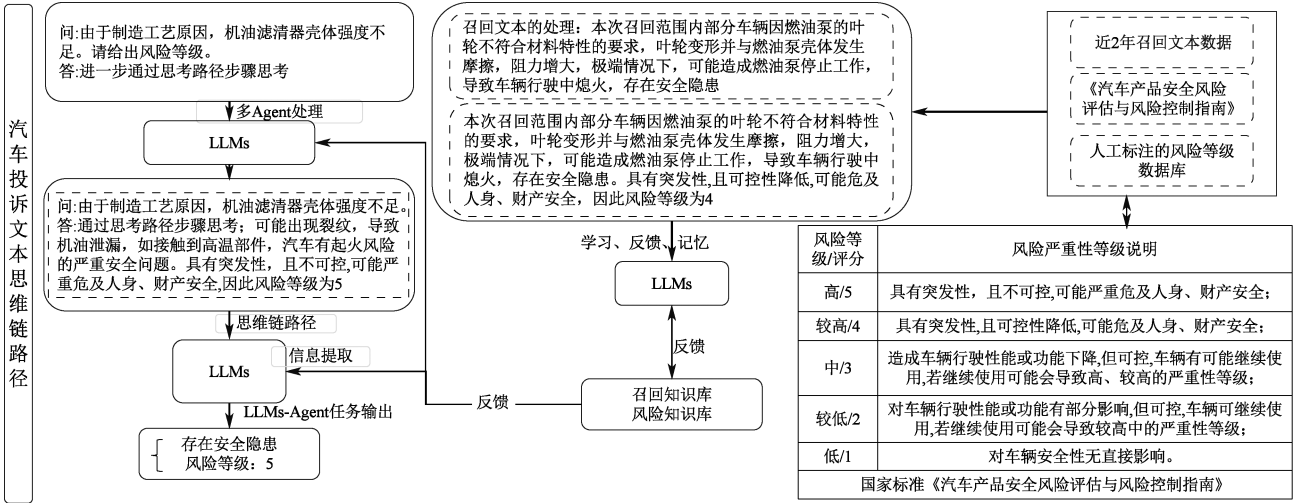


图4 基于召回文本与CoT技术的风险知识库

Fig. 4 Risk knowledge base based on recall text and CoT technology

3.2 召回知识库构建

将召回文本结合 LLMs 的思维链技术, 引入汽车召回知识库, 以增强对汽车风险的深度语义分析及 MA 协同处理能力。召回知识库以近 2 年国家缺陷产品召回技术中心发布的月度召回汇总数据为基础, 通过 LLMs 多模态信息提取方法获取结构化文本, 并经人工标注与思维链处理补充, 涵盖车辆信息、召回原因、思维提示与风险等级的高质量召回语料。

3.3 风险等级判别智能体构建

单个智能体具备决策思维、工具使用、记忆和反馈等功能, 通过多个智能体的协作, 可以解决更动态和复杂的任务。具体步骤如下:

1) 步骤 1: LLMs 输入处理。首先, 通过智能体解析输入的文本(支持多文本输入), 运用换行分析或 LLMs 按输入顺序划分文本数据, 且相关输入输出流程可简化如下:

$$\underbrace{\text{LLM}}_{\text{大模型}}(\underbrace{I}_{\text{任务描述}}, \underbrace{f(u_1, v_1), \dots, f(u_l, v_l)}_{\text{示例}}, \underbrace{f(u_{l+1}, \quad)}_{\text{输入}} \underbrace{\quad)}_{\text{答案}}) \rightarrow \hat{v}_{l+1} \quad (4)$$

2) 步骤 2:知识库的融入。智能体选择相应的 LLMs 之后,给定提示词和思维链需要,导入设计的知识库模块并结合分析,再进行风险词和风险等级的分析与预测输出。

3) 步骤 3: LLMs-MA 的风险等级判别任务执行。MA 协同技术上,定义好相关 MA 应用的规则,考虑共识寻求的机制,采用 MA 分层化、集中化和去中心化部署来适应不同场景,更好评估和解决问题,从而综合全面的执行风险等级判别任务。

3.4 基于共识寻求的 MA 集成方法

基于召回的思维链、风险词和知识库,运用多个 LLMs 处理投诉文本,开展风险等级评估,执行共识寻求算法来集成多个评估结果。步骤如下:

1) 步骤 1:工作流 Agent 设计。设置包含 $n+1$ 个 LLMs 型、代码、文本处理、数据库等结构的 MA 系统的工作流;结合设计好的高质量知识库、人工标注召回库和思维链内容分析。将投诉文本数据输给 n 个智能体 $[g_1, g_2, \dots, g_n]$ 进行信息的结构化提取和评估预测,得到前 n 个智能体的 JSON 格式输出向量: { 文本、风险类别、风险词、风险等级 }。

2) 步骤 2:多模型评估。将投诉本文引入 n 个智能体 $[g_1, g_2, \dots, g_n]$ 中,执行 MA 信息提取和风险识别任务,将输出内容引入后续 LLMs 和代码流进行共识迭代特定任务处理,得到 n 个智能体 $[g_1, g_2, \dots, g_n]$ 输出的风险等级和评估内容。

3) 步骤 3: 共识寻求输出。结合智能体 Agent_{n+1} 和知识库、代码等工作流模块获取前 n 个智能体的风险等级内容进行排序评判,执行表 3 中共识寻求迭代算法,输出对于该文本风险词的最优风险等级向量 C 。

表 3 基于共识寻求的 MA 集成算法的伪代码

Table 3 Pseudo-code for MA consensus seeking

输入:汽车质量问题投诉文本
输出:基于 MA 的风险识别向量 C
1. Initialize agents $G = [g_1, g_2, \dots, g_n]$
2. Define level $C = [c_1, c_2, \dots, c_n]$ // $c_i \in [1, 5]$ 为 g_i 输出的风险等级
3. text = "投诉文本内容"
4. $C = []$ // 风险等级
5. function get_risk(G , txt):

续表 3

输入:汽车质量问题投诉文本
return agent. eval(txt)
6. for g in G : // 遍历所有智能体得到所有风险等级的结果
level = get_risk(g , text)
$C.append(level)$
7. mean_level = sum(C) / len(C) // 求风险等级均值
8. outliers = [g for i, g in enumerate(G) if abs($c_i - \text{mean_level}$) > 1.5] // 通过阈值筛选出离群智能体
9. for g in outliers:
10. closest = argmin_ $ g \in G \mid \text{abs}(\text{get_risk}(g, \text{text}) - \text{mean_level}) $ // 获取距离均值最小的智能体
11. prompt, think = closest. get_chain() // 获取所有离均距离最小的智能体思维链
12. new_level = get_risk(g , text + [prompt + think]) // 离群智能体重新评估风险等级
$C[G.index(g)] = \text{new_level}$
13. final_ $C = [c_i$ for i in C if abs($c_i - \text{mean_level}$) < = 1.5] // 遍历 MA 得到最终风险结果

基于共识寻求的 MA 风险等级输出流程图和样例,如图 5 所示。

因此,为更好进行后续试验指导和分析,文中基于上述流程构建 MA 的汽车投诉数据信息提取和风险识别流程总框架。总结模型特征,并设计总体框架结构,如图 6 所示。

4 汽车安全风险等级判别试验

汽车安全风险等级判别试验聚焦于车质网用户投诉文本的具体内容、典型问题上,结合前文数据预处理可得风险文本内容、风险类别和风险词数据,构造出的数据矩阵主要包括投诉车辆品牌、投诉问题、投诉文本、风险词和风险类别等。具体示例见表 4。

文中基于划分的高质量投诉数据作为 MA 的输入,引入具有召回知识库、思维链提示的 MA 中,具体以 5 个智能体、多个代码节点的工作流为试验示例。采用自然语言处理中广泛使用的分类评价指标,包括正确率 A (Accuracy)、准确率 P (Precision)、召回率 R (Recall) 和 F_1 分数。 A 是指所有预测中正确的比例, P 指预测为正的样本中,实际为正确的比例, R 是指在所有实际为正的样本中,被正确预测为正确的比例, F_1 为精确度和召回率的调和平均值。

$$A = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

$$P = \frac{TP}{TP + FP} \quad (6)$$

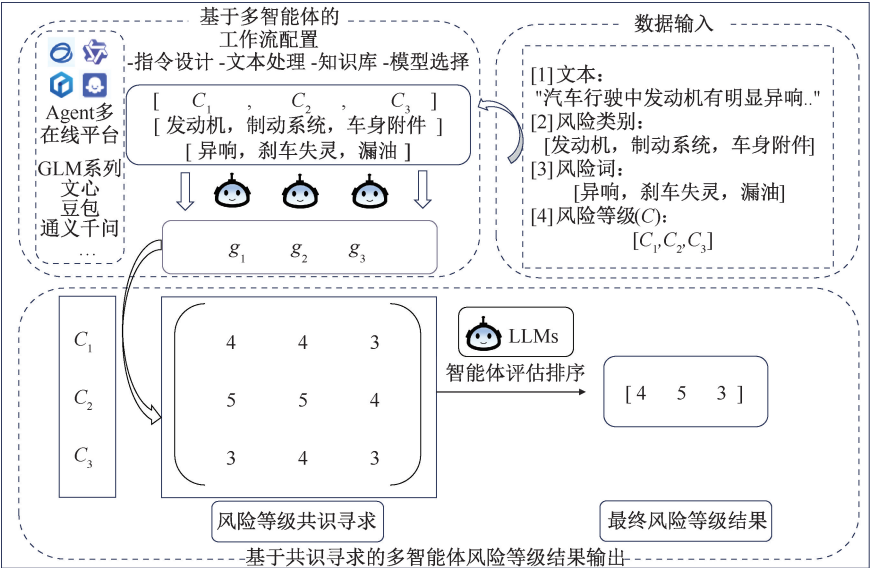


图 5 基于 LLMs-MA 的共识寻求样例

Fig. 5 Example of consensus seeking based on LLMs-MA

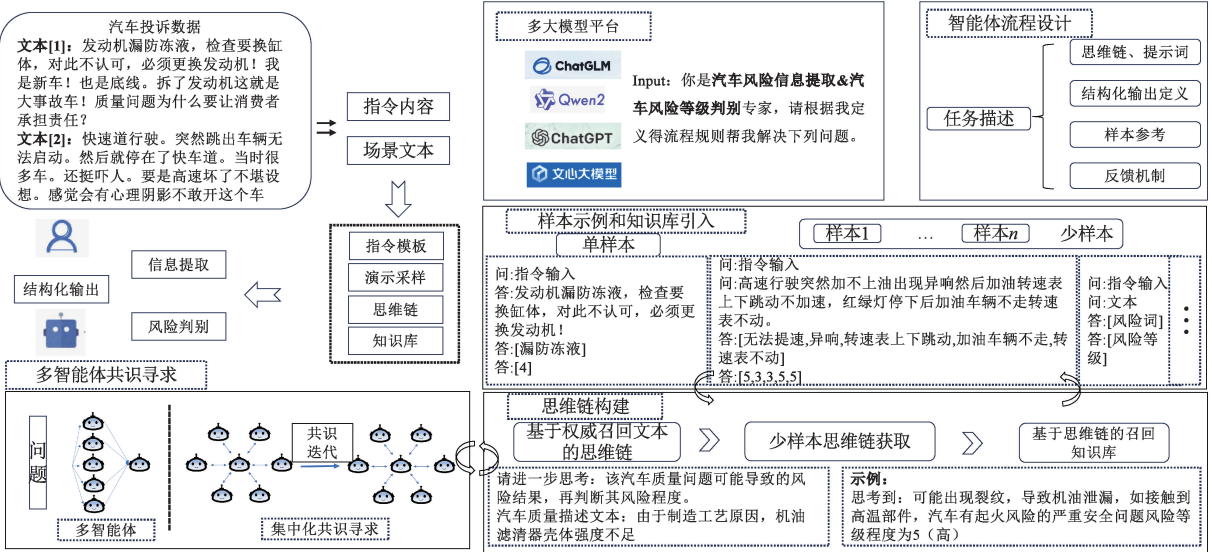


图 6 基于 LLMs-MA 的汽车安全风险等级判别模型框架

Fig. 6 A framework for identifying automotive safety risks based on LLMs-MA

表 4 汽车投诉数据信息示例

Table 4 Example of car complaint data information

投诉信息	案例 1	投诉信息	案例 2
投诉时间	2024-10-13	投诉时间	2024-09-01
投诉品牌-车系	A 品牌-a 车系	投诉品牌-车系	B 品牌-b 车系
投诉车型	2024 款 X 车型	投诉车型	2023 款 Y 车型
投诉文本内容	提车第 2 天在红绿灯路口就显示动力受限,无法行驶,拖车回 4 S 店做检查,结果未出,要求处理	投诉文本内容	发动机漏防冻液,检查要换缸体,对此不认可,必须更换发动机!我是新车!也是底线。拆了发动机这就是大事故车!质量问题为什么要让消费者承担责任
风险词	动力受限,无法行驶	风险词	漏防冻液
风险类别	发动机/电动机,发动机/电动机	风险类别	发动机/电动机
风险等级	5,5	风险等级	4

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \tag{8}$$

式中:TP (True positive) 和 FP (False positive) 分别为真、假正例;TN (True negative) 和 FN (False negative) 分别为真、假负例。

因此,随机选择(数据集中未成为演示样本的)8 大风险类别中 600 个质量问题样本,进行人工标注,建立基准集,并与文中结合了知识增强的智能体模型比较分析,得到 MA 以及 Doubao-pro、通用语言模型 (General Language Model, GLM)-4、文心-基于知识融合的增强表征 (Enhanced Representation through Knowledge Integration, ERNIE)-4、Qwen-max 的单智能体和零样本模型风险判别的部分对比结果。以部分文本为测试样本,根据火山引擎用量统计和计费规则概述等,给出 MA 的性能测试(词元量、资源点量、响应时间)和成本分析参考,分布见表 5—表 7 和如图 7 所示。

表 5 知识增强下的模型评价

Table 5 Model evaluation under Knowledge-enhanced %				
LLMs	<i>A</i>	<i>P</i>	<i>R</i>	<i>F</i> ₁
MA	95	96	92	93
GLM-4	90	91	85	90
Doubao-pro	91	90	84	89
文心-ERNIE-4	89	89	85	88
Qwen-max	90	91	84	89

表 6 零样本下的模型评价

Table 6 Model evaluation under zero samples %				
LLMs	<i>A</i>	<i>P</i>	<i>R</i>	<i>F</i> ₁
MA	90	90	89	91
GLM-4	88	87	82	89
Doubao-pro	87	87	84	89
文心-ERNIE-4	85	88	85	88
Qwen-max	84	86	82	85

表 7 LLMs-MA 性能响应和成本参考

Table 7 LLMs-MA performance response and cost reference	
性能参数	平均每条文本成本计量
智能体数量	<i>n</i>
API 平均响应时间/s	10~15
词元量	2 000 · <i>n</i> (不含插件、知识库等工作流)
资源点消耗(输入资源点 <i>D</i> 和输出资源点 <i>O</i>)	(0.8 × <i>D</i> + 2 × <i>O</i>) × <i>n</i>
智能体资源/元	(0.008+0.02) × <i>n</i>

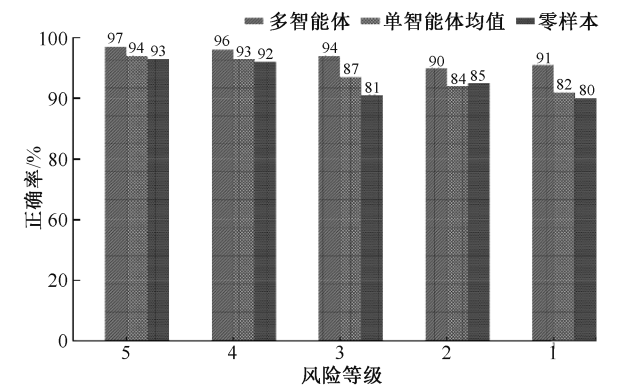


图 7 不同模型方法下风险等级判别效益对比

Fig. 7 Comparison of risk level discrimination benefits under different model methods

5 结 论

1) 通过分析 LLMs 智能体结合知识库与相关技术工具以及利用其开展汽车质量与投诉数据的信息提取和汽车安全风险判别任务,可有效降低人工决策成本。CoT 优化模型的应用,有助于提升 LLMs-MA 在汽车安全风险识别中的准确率与效率,在信息提取与风险判别任务中,文中所提方法较传统方案与零样本大模型展现出更高的可靠性。

2) 基于 LLMs-MA 共识寻求的集成方法可有效提升汽车投诉数据风险等级预测准确性,减少传统决策分析依赖和增强系统稳健性。基于权威召回文本的知识库结合思维链设计,有助于厘清汽车质量原因和风险识别决策,也为其他应用场景提供示范,有较强普适性。

3) 随着我国汽车产业与智驾技术的快速发展,相关车机系统配件等服务和质量问题也逐渐受到消费者重点关注,汽车投诉数据的智能化信息提取和处理需求增多,未来研究可进一步规范信息提取流程,引入概率分布的决策预测,并考虑模型异构性以完善 MA 共识能力,优化 workflow 效率与成本,进一步提升模型性能。

参考文献

[1] ZHAO Wanyu, ZHOU Kun, LI Jian, et al. A survey of large language models[J]. arXiv, 2023; DOI: 10.48550/arXiv.2303.18223.

[2] XU Derong, CHEN Wei, PENG Wenjun, et al. Large language models for generative information extraction: a survey[J]. Frontiers of Computer Science, 2024, 18(6): DOI: 10.1007/s11704-024-40555.

[3] GUO Taicheng, CHEN Xiang, WANG Yang, et al. Large language model based multi-agents: a survey of progress and challenges[J]. arXiv, 2024; DOI: 10.48550/arXiv, 2402.01680.

[4] POLAK P M, MORGAN D. Extracting accurate materials data from research papers with conversational language models and prompt engineering[J]. Nature communications, 2024, 15(1): DOI: 10.1038/s41467-024-45910-y.

[5] DAGDELEN J, DUNN A, LEE S, et al. Structured information extraction from scientific text with large language models[J]. Nature communications, 2024, 15(1): DOI: 10.1038/s41467-024-45631-2.

[6] 江耀森, 杨超宇, 刘晓蕾. 基于 ChatGLM3-6B 的煤矿事故知识图谱构建及应用研究[J]. 中国安全生产科学技术, 2024, 20(10): 12-21.

JIANG Yaosen, YANG Chaoyu, LIU Xiaolei. Research on construction and application of knowledge graph for coal mine accidents based on ChatGLM3-6B[J]. Journal of Safety Science and Technology, 2024, 20(10): 12-21.

[7] 王喆, 杨栋梁, 况星园, 等. 考虑提示学习的洪涝灾害应急决策自动问答模型研究[J]. 中国安全生产科学技术, 2022, 18(11): 12-18.

WANG Zhe, YANG Dongliang, KUANG Xingyuan, et al. Research on automatic question answering model of flood disaster emergency decision-making considering Prompt-learning[J]. Journal of Safety Science and Technology, 2022, 18(11): 12-18.

[8] 唐海洋, 刘振翼, 陈东平, 等. ChatSOS: 基于大语言模型的安全工程知识问答系统[J]. 中国安全科学学报, 2024, 34(8): 178-185.

TANG Haiyang, LIU Zhenyi, CHEN Dongping, et al. ChatSOS: large language model-based knowledge Q & A system for safety engineering [J]. China Safety Science Journal, 2024, 34(8): 178-185.

[9] 宋敦江, 杨霖, 钟少波. 基于 BERT 的灾害三元组信息抽取优化研究[J]. 中国安全科学学报, 2022, 32(2): 115-120.

SONG Dunjiang, YANG Lin, ZHONG Shaobo. Research on optimization of disaster triplet information extraction based on BERT[J]. China Safety Science Journal, 2022, 32(2): 115-120.

[10] 施文, 渠玉杰, 王小双. 考虑文本结构特征的产品召回监督主题模型及应用研究[J/OL]. 中国管理科学: 1-19. [2025-05-14]. <https://doi.org/10.16381/j.cnki.issn1003-207x.2023.1637>.

SHI Wen, QU Yujie, WANG Xiaoshuang. Supervised topic modeling and application research on product recall considering textual structural features [J/OL]. Chinese Journal of Management Science: 1-19. [2025-05-14]. <https://doi.org/10.16381/j.cnki.issn1003-207x.2023.1637>.

[11] 施文, 渠玉杰, 蒋国银. 基于随机 Kriging 的汽车品牌质量序贯主题比较研究[J]. 中国管理科学, 2023, 31(11): 114-127.

SHI Wen, QU Yujie, JIANG Guoyin. Comparative study on sequential topics of automobile brand quality based on stochastic Kriging[J]. Chinese Journal of Management Science, 2023, 31(11): 114-127.

[12] 王余行, 党延忠, 徐照光. 针对论坛数据特点的汽车质量问题挖掘[J]. 中国管理科学, 2021, 29(9): 201-212.

WANG Yuhang, DANG Yanzhong, XU Zhaoguang. Mining automobile quality problems based on the characteristics of forum data [J]. Chinese Journal of Management Science, 2021, 29(9): 201-212.

[13] 王晗, 宋赞. 在线投诉文本驱动的汽车质量监控方法研究[J]. 数理统计与管理, 2024, 43(5): 789-799.

WANG Han, SONG Zhi. Research on automobile quality monitoring method driven by online complaint text[J]. Journal of Applied Statistics and Management, 2024, 43(5): 789-799.

[14] 刘心报, 胡俊迎, 陆少军, 等. 新一代信息技术环境下的全生命周期质量管理[J]. 管理科学学报, 2022, 25(7): 2-11.

LIU Xinbao, HU Junying, LU Shaojun, et al. The entire life cycle quality management in the new generation of information technology environment[J]. Journal of Management Sciences in China, 2022, 25(7): 2-11.

[15] GB/T 34402—2017, 汽车产品安全风险评估与风险控制指南[S].

GB/T 34402-2017, Safety of motor vehicle product: guidelines for risk assessment and risk control[S].

[16] NETEASE Y D. BCEmbedding: bilingual and crosslingual embedding for RAG[Z]. [2024-12-07]. <https://github.com/netease-youdao/BCEmbedding>.

[17] BLONDEL V, GUILLAUME L J, LAMBIOTTE R. Fast unfolding of communities in large networks: 15 years later[J]. Journal of Statistical Mechanics: Theory and Experiment, 2024, 2024(10): DOI: 10.1088/1742-5468/AD6139.

作者简介: 杨艺 (1990—), 男, 湖南邵阳人, 博士, 副教授, 主要从事大语言模型智能体、多层复杂网络、拓扑聚类、在线服务平台大数据挖掘等方面的研究。E-mail: yangyihnuc2018@163.com。

