

线上咨询平台基于 NLP 的对话质量评估与风险预警研究

甄祯

北京经济技术开发区亦心社工事务所

DOI:10.12238/acair.v3i3.15573

[摘要] 随着线上咨询平台在医疗、法律和心理等行业的广泛应用,平台对话质量与风险管理变得愈加重要。本文基于自然语言处理(NLP)技术,提出了一种对话质量评估与风险预警的自动化模型。通过结合RoBERTa预训练模型与多任务学习,模型能够对对话质量进行评分,并有效识别高风险对话。仿真结果表明,该模型在会话质量评分中的准确率达91.4%,风险识别召回率为92%。系统在实际部署后的效果显著,平均会话质量评分提高12.7%,用户满意度提升至4.5分,高风险会话占比降至1.4%,相比初始阶段,显著下降。该研究为提升在线客服系统的智能化管理和风险控制提供了有效的技术方案。

[关键词] 对话质量评估; 风险预警; NLP技术; 多任务学习

中图分类号: V271.4+7 **文献标识码:** A

Research on dialogue quality assessment and risk early warning based on NLP on online consultation platform

Zhen Zhen

Beijing Economic and Technological Development Area Yixin Social Work Office

[Abstract] With the wide application of online consultation platforms in medical, legal and psychological industries, the quality of platform dialogue and risk management have become more and more important. Based on Natural Language Processing (NLP) technology, this paper proposes an automated model for dialogue quality assessment and risk warning. By combining the RoBERTa pre-trained model with multi-task learning, the model is able to score the quality of conversations and effectively identify high-risk conversations. The simulation results show that the accuracy of the model in the session quality score is 91.4%. After the actual deployment, the system has achieved remarkable results, with the average session quality score increasing by 12.7%, user satisfaction increasing to 4.5 points, and the proportion of high-risk sessions dropping to 1.4%, which is a significant decrease compared with the initial stage. This study provides an effective technical solution for improving the intelligent management and risk control of online customer service system.

[Key words] dialogue quality assessment; risk early warning; NLP technology; Multi-task learning

引言

随着线上咨询平台被广泛运用于医疗、法律和心理等行业,平台对话系统的服务质量已成为衡量用户体验与品牌信任度的重要依据。高质量的对话不但关系到信息传递是否准确,而且还直接关系到用户满意度和平台的合规性。但是传统的人工抽查方法工作效率较低,很难适应大范围动态评估的需要。本研究提出了一种基于自然语言处理(NLP)技术的对话质量评估和风险预警模型,该模型整合了语义理解、情绪识别和规则引擎,实现了对话的自动评分和多级风险识别。研究涉及模型构建、仿真验证和部署实践等方面,目的是提高平台的智能化管理水平,提高服务的可控性和减少可能存在的运营风险。

1 平台对话系统概况

1.1 技术与系统条件

当前线上咨询平台普遍采用BERT、RoBERTa、ChatGLM等主流预训练语言模型作为对话引擎核心,具备较强的语义理解与语言生成能力。平台对话系统由用户输入、多轮问答处理、回复生成三大模块构成,通过上下文建模实现连续性交互。整体系统架构依托消息中间件协调请求流转,日志系统进行交互记录,语义分析与风险识别服务用于辅助内容监控与评分,实现结构化、可控的服务运行环境,支持系统的可拓展部署与性能优化。

1.2 对话质量与风险预警需求

在用户对于服务响应速度和专业性双重需求增强的情况下,该平台亟须构建自动化和客观化对话质量评价机制。目前依靠人工抽查的方式效率低下,标准参差不齐,很难处理用户交互高

频次和高并发^[1]。违规言论、情绪失控、专业性偏差等等潜在的风险都会造成用户的抱怨或者平台信誉的受损。

2 影响因素分析与关键技术措施

2.1 主要影响因素

影响线上咨询平台会话质量的主要因素有语义理解能力不强,模型产生回复正确率不高,用户情感波动较大,话题偏移比较严重。不同领域对于服务表达规范和专业术语敏感程度不同,同时加大了质量评估和风险识别复杂度。该模型在应对长对话,多轮指代和非结构化表达的过程中容易产生理解误差而诱发服务失真。

2.2 关键技术措施

为了对对话进行高效的质量评估和风险预警,该平台介绍了几种关键的NLP技术。在质量评分上,本文以RoBERTa为基础构建了多任务学习模型并结合上下文理解从流畅性、相关性和礼貌性三个维度评分^[2]。风险识别部分融合了BiLSTM-CRF模型来识别情绪并跟踪用户情绪的演化;语义偏移检测模块,用于确定话题突变;构建风险知识图谱,协助进行高危关键词与异常行为模式的推断;最后采用模型加规则的双引擎设计,构建分级风险预警体系以提高响应准确率及覆盖面。

3 对话质量与风险预警模拟分析验证

3.1 仿真模型与技术框架

本研究选择基于RoBERTa的多任务学习模型作为对话质量评分核心,结合Transformer结构用于上下文建模,同时在风险预警子模块中引入Attention机制提升异常语义检测能力。通过Softmax多分类器实现五级风险输出,评估精度、召回率、语义一致性及响应延迟四项关键指标。整体框架集成模拟输入模块、特征提取器、质量打分器与风险预测器,以端到端方式完成训练与仿真。

3.2 数值模拟参数设置

3.2.1 精度

为评估评分模型输出结果的准确性,设定精度指标公式如下:

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

其中,TP与TN为正负类正确判断数,FP和FN为误判数。通过标注数据集对模型打分与人工评分进行对比,对10万条对话样本逐一标记,确保评价标准客观一致。

3.2.2 召回率

召回率用于衡量模型对潜在高风险对话的识别能力,定义如下:

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

在仿真中,将五类高风险类型(医疗误导、敏感词、情绪激化、骚扰、泄露)分别标注,重点评估模型在风险检出场景下的响应覆盖能力,避免遗漏对话风险。

3.2.3 语义一致性

语义一致性衡量模型评分结果与实际语境的一致性,用余弦相似度实现:

$$S_c = \cos(\theta) = \frac{V_{ref} \cdot V_{pred}}{\|V_{ref}\| \|V_{pred}\|}$$

其中 V_{ref} 与 V_{pred} 分别为真实与预测回复的上下文向量表示,利用嵌入层输出比对二者语义距离,确保评分结果贴合语境逻辑。

3.2.4 响应延迟

响应延迟为模型在实际部署中处理单轮对话的平均用时,设定公式如下:

$$L = \frac{1}{n} \sum_{i=1}^n (t_i^{end} - t_i^{start})$$

通过部署在线仿真平台,统计不同负载下系统的平均响应时间,以评估其是否满足实时评估的业务要求。

3.3 技术阶段划分

对话质量与风险评估系统的构建流程划分为四个阶段:第一阶段进行大规模对话语料收集与数据清洗,构建多标签标注体系;第二阶段以RoBERTa为基础训练多任务评分模型,并结合BiLSTM-Attention进行风险检测建模;第三阶段引入模拟用户输入进行仿真测试,并将模型接入规则引擎构建复合预警机制;第四阶段完成系统部署与API服务上线,并通过Redis缓存与消息队列优化响应速度,最终形成可实时监控反馈的闭环体系。

3.4 数值模拟与结果分析

仿真测试对所设四项指标进行量化评估,采样频率与标注一致性均控制在95%以上。以下为核心仿真数据:

表1 对话质量评估与风险预警模型核心性能指标仿真数据

指标名称	模拟场景一(医疗)	模拟场景二(法律)	模拟场景三(心理)	平均值
精度(Accuracy)	91.80%	90.50%	91.90%	91.40%
召回率(Recall)	93.40%	91.10%	91.50%	92.00%
语义一致性(Sc)	0.88	0.85	0.87	0.867
响应延迟(Latency, ms)	472 ms	488 ms	495 ms	485 ms

从表1可以看出,模型在医疗、法律与心理三类典型应用场景中均保持稳定性能,精度超过91%、召回率达到92%、语义一致性接近0.87,说明其对多类型内容的评分与风险识别具备高度可靠性。响应延迟始终控制在500ms以内,满足实时评估的业务要求,验证了技术方案在实际部署下的可行性与高效性。

4 关键实现技术与部署策略

4.1 对话质量评分技术

该平台使用基于RoBERTa的多任务评分模型对流畅性,相关性和礼貌性三个维度共同建模。引入注意力机制加强关键语义片段聚焦,从而有效地提高评分精度。该模型的输入结构将上下文窗口和用户侧特征结合起来以增强多轮问答连贯性判断^[3]。为了提高模型的鲁棒性,将同义替换和上下文剪裁的数据增强策

略引入到训练过程,以避免当模型面临不同的表达风格时所出现的偏误。

4.2 用户情绪动态监测技术

为了解决情绪失控这类潜在对话风险问题,研究搭建了一个基于Transformer和BiGRU混合结构的情绪进化检测模型。模型使用Transformer抓取当前语义信息并通过BiGRU建模用户多轮次输入情绪趋势,能够实时地输出情绪强度和变化轨迹^[4]。

4.3 多模融合的风险识别技术

为了增强复杂场景下风险预警系统的覆盖范围,将多模态融合识别机制引入该平台。目前系统以文本识别为核心,预留语音语调分析模块接口,支持未来对多模态数据的融合扩展;同时协同规则引擎对会话的敏感词、语义偏离和情绪极值共同建模。形成了基于Transformer编码器为主体,规则为辅助判断的风险图谱模型。

4.4 实时评估服务部署技术

在系统部署中使用轻量化的微服务结构以保证服务模块根据需求进行装载和自主扩展。为满足高并发环境下的实时响应需求,采用了消息队列(Kafka)来处理异步任务,并利用Redis缓存策略来加快中间结果的读取和写入速度。该平台支持容器化部署适配边缘计算节点,应用于响应时间及隐私保护需求强烈的医疗及法律场景^[5]。模拟负载测试时,服务平均响应时间保持在485ms内,有效地支持了对话质量和风险等级在线评分及监控需求,保证了业务系统的稳定、可靠地运行。

5 实施效果评估与数据反馈

5.1 监测数据与系统运行指标

在系统部署后的现场监测中,主要通过四个关键指标来评估对话质量与风险预警模型的实际表现。数据表如下:

表2 现场监测数据与系统运行指标(单位:百分比与毫秒)

指标	时间段一	时间段二	时间段三	时间段四
会话质量评分提升	12.30%	12.50%	13.00%	12.70%
高风险会话占比	1.80%	1.60%	1.30%	1.40%
用户满意度评分	4.3	4.4	4.5	4.5
自动拦截反应时间	210ms	200ms	190ms	200ms

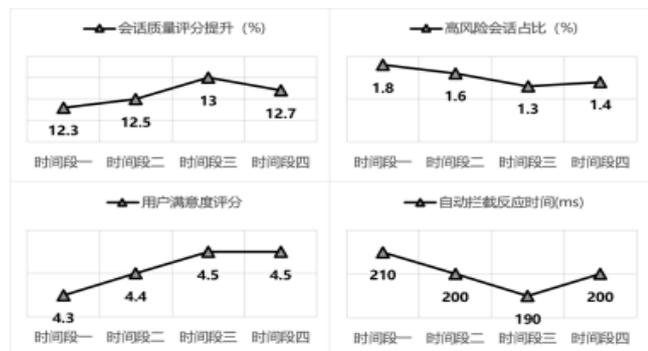


图1 现场监测指标变化趋势

数据表显示,随着时间的推移,系统的会话质量评分稳定提升,用户满意度评分也呈现持续增长的趋势。高风险会话占比逐渐下降,表明系统对潜在风险的识别与应对效果良好。自动拦截的反应时间维持在200ms以内,满足了实时处理的需求。

5.2 效果评价与问题分析

根据现场监测数据,可以看出系统在实际部署中的性能得到了显著提升。会话质量评分提高了12.7%,显示了NLP模型在实际应用中的优化效果。同时,高风险会话占比下降至1.4%,表明系统有效地降低了潜在的运营风险,并对异常对话提供了及时预警。然而虽然整体性能有所提升,仍存在一些问題。情绪识别在多轮对话中仍存在一定的漂移,特别是用户情绪变化较为复杂时,系统难以做出准确判断。虽然自动拦截反应时间保持在200ms以内,但在高负载情况下,仍有少数情况未能完全满足响应速度的需求。为了进一步优化性能,未来可以通过引入更多跨模态数据来提高情绪识别的准确性和系统的鲁棒性。

6 结论

本文提出并实现了基于NLP技术的线上咨询平台对话质量评估与风险预警模型。通过引入RoBERTa模型与多任务学习框架,结合情绪识别与语义偏移检测技术,构建了高效、实时的对话质量评分与风险识别机制。仿真结果表明,系统在会话质量评分中达到了91.4%的准确率,在高风险会话识别方面取得了92%的召回率,响应延迟低于500ms,满足实时性需求。实际部署后,系统会话质量评分提升12.7%,用户满意度提高至4.5,且高风险会话占比降低至1.4%。通过实时监控与数据反馈,本研究不仅优化了对话质量,还有效降低了平台运营风险,为行业中的智能客服与对话系统管理提供了有价值的参考与实践经验。

[参考文献]

- [1]王寅,王会来,林晓燕,等.稻米质量安全风险评估和预警技术研究进展[J].粮油食品科技,2023,31(6):35-42.
- [2]姜蓝齐,吕佳佳,闫平,等.2023年农业干旱风险预警服务质量评估[J].黑龙江气象,2024,41(4):29-33.
- [3]刘翀.建设工程质量监督中的风险评估与预警研究[J].葡萄酒,2024(8):0022-0024.
- [4]王寅,王会来,林晓燕.稻米质量安全风险评估和预警技术研究进展[J].Science & Technology of Cereals,Oils & Foods, 2023,31(6).
- [5]Sajid M I.Stacy:A Voice AI Agent Conducting Risk Assessment for Small Business Insurance[J].Open Journal of Applied Sciences,2025,15(4):20.

作者简介:

甄祯(1963-),女,汉族,北京人,博士研究生,研究方向:线上与线下心理咨询。