专利技术交底书样本

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 客户名称 |  | | | |
| 发明名称 | ​多模态语言身份聚合引擎 | | | |
| 技术联系人 | 郭鸿宏 | 邮 箱 | weiwang@szlanyou.com | |
| 手 机 |  | | 专利类型 | 发 明  实用新型 |
| 固定电话 |  | |

一、技术背景

当前自然语言处理技术已具备多模态数据处理能力，从语音转文本（STT）、多源内容爬取到大模型微调均存在成熟解决方案，但各模块的独立运作导致个性化语言风格建模效率低下。音频处理模块虽能提取音视频中的原始文本，却忽略语调节奏对用词风格的影响；多搜索引擎爬取的关键词扩展虽能关联个人高频术语，但静态规则清洗易抹去口语化表达特征；而传统微调模型依赖固定训练集，无法通过动态知识库融合目标对象的最新语言特征（如社交媒体实时更新的表达习惯）。这种割裂式技术堆叠使得生成的文本难以复现个体语言的全貌，仅保留碎片化的风格痕迹。

因此，亟需一种能够贯通音视频解析、多源数据融合、动态知识演进的协同系统，来突破现有技术孤岛：既需要将语调波形特征（如语速起伏）与文本句式结构建立隐性关联，使音频韵律转化为可量化的风格参数；又需要设计多级过滤机制，在清洗语法错误的同时保留个人标志性表达（如特定网络用语）；更需要构建内容自动回流通道，使目标对象的新生成内容实时强化训练数据，确保语言风格持续进化。唯有通过此类深度整合，方能实现用词习惯、逻辑特征、情感倾向的多维度对齐，最终生成与真实个体高度一致的拟真交互文本，满足虚拟博主、数字分身、个性化教育等场景的工业化落地需求。

二、发明目的

本发明旨在开发一种基于多模态大模型的具有个人语言风格的大模型，提供一种可以高度模仿某个人物说话风格（比如口头禅，语气词等）的生成式AI。系统通过搜集到个人公布的视频和音频等内容，结合STT（Speech-to-Text）模块将所有搜集到的内容转换为文本内容，总结提取出个人的语言风格和文本数据作为模型训练的数据集，同时将其储存到知识库，最终实现具有个人语言特色的个人风格大模型。此外，结合RAG知识库检索、多引擎爬虫搜索和工具调用模块，系统充分利用丰富的历史数据和实时信息，加强了交互的相关性与准确性。简洁高效地搜集和处理各种信息，增加了系统答复的速率、准确率以及和目标人物语言风格的匹配程度。

系统可以根据具体需求（关键词），实时地进行多搜索引擎爬取相关的数据，利用高效地解析和清洗规则和LLM对数据清洗的双重清洗筛选，提炼出与用户需求（关键词）相关联的纯净资料。交付给已经训练好的个人风格大模型，大模型根据提炼出的纯净数据，最终写出具有目标人物语言风格、与用户需求（关键词）密切相关的答复（文章）。本系统基于强大且灵活的多智能体模型平台构建，具备出色的适应性和扩展性，适合多种硬件和软件环境。虽然智能体数量较多，但合理的架构设计有效避免了链路的+复杂性，确保用户拥有高效自然的交互体验。系统能够根据具体应用需求快速重组和调整（比如作诗或作曲），使产品具有高度的可移植性和市场适应性，可以满足各种行业用户多样化的需求。

三、技术方案

本发明所要解决的技术问题是通过以下技术方案实现的，一种基于多模态大模型为核心的具有个人风格的模型，包括以下步骤：

步骤1：指定目标人物（比如某个博主、名人），训练出具有该人物语言风格的LLM，同时构建对应的RAG知识库。具体步骤如下：

1. 搜集和处理对应目标人物的媒体数据（曾经公布的视频，发布的歌曲、音频、电台、文章等）。

Ⅰ. 数据搜集和视频文件处理

通过定制化网络爬虫工具对目标人物的公开内容进行全平台抓取，覆盖视频、音频、文本等多种格式。爬虫脚本基于目标平台的公开接口或页面解析协议，自动识别并下载新增内容，同时记录文件的元数据（如发布时间、文件格式、原始链接）至临时存储区。视频文件会被送入音频提取模块，该模块调用预设参数的音轨分离算法，确保输出音频的采样率、声道数适配后续语音识别引擎的输入规范。若音频长度超过阈值，系统自动按固定时长切割分段，避免单次处理超时。

Ⅱ. 音频处理

音频内容通过预训练的语音识别模型进行转写，模型基于深度学习架构，支持多语言与口音自适应。引擎内置噪声抑制与说话人分离功能，可区分目标人物与其他干扰声源，提升核心内容的转写准确率。转写结果以时间戳对齐的文本流形式输出，同时生成置信度评分，低置信度片段触发自动重试或标记待人工复审。原始文本经过正则表达式规则库清洗，剔除平台水印、重复语气词等无关信息，随后输入自然语言处理流水线，完成分词、实体识别与语义标签生成（如情感极性、主题聚类）。

Ⅲ. 监控

流程中设置了多层质量控制机制。在音频转写阶段，系统通过预设的置信度阈值筛选低质量结果（如背景噪声干扰片段），自动触发重转写或标记为待审核状态；文本清洗环节则结合规则引擎（正则表达式过滤广告文本）与统计模型（基于上下文语义校验逻辑性）进行双重验证。失败任务默认进入自动重试队列，超过最大重试次数后推送至人工处理通道，确保流程容错性。数据库层面采用版本快照与增量备份策略，同时通过角色权限控制（RBAC）限制数据访问范围，并对敏感字段进行脱敏处理（如哈希替换真实姓名），以满足GDPR（通用数据保护条例）等合规要求。系统定期收集错误样本与用户反馈数据，迭代更新文本清洗规则库并微调语音识别模型参数，形成闭环优化机制。

1. 对得到的文本数据进行格式化处理（得到可以体现目标说话风格的数据集）。

为适配RAG与大模型微调需求，需将原始文本转化为兼具语义与风格特征的结构化数据。首先采用BERT对文本进行嵌入编码，重点利用中间层捕获长距离语义依赖和修辞特征，结合LDA识别内容的核心议题，形成多维度标签。处理后的文本按段落切割，最终存储为JSON格式，便于后续检索与模型学习。

1. 构建RAG知识库同时微调大模型

Ⅰ. 利用得到的数据集构建RAG知识库

采用滑动窗口法对原始文本分块，窗口大小为512字符，步长128字符，确保语义连续性，使用预训练的Sentence-BERT模型（all-mpnet-base-v2版本）对每个文本块生成768维向量。通过FAISS库构建索引，采用IVF（倒排文件）分簇算法，设置分簇数为256，加速相似性检索。

为每个文本块附加时间衰减权重，计算公式为：

LaTex: weight=\frac{1}{1+\log\_{}{\left ( 当前时间戳-发布时间戳 +1\right ) } }

同时，通过LDA主题模型提取每个文本块的前3个主题词作为标签（如“科技”“教育”“生活”），用于检索过滤。

Ⅱ. 微调大模型

a. 在Llama-2-7B模型的注意力层注入LoRA低秩适配器，秩数设为8，仅训练适配器参数并冻结原模型权重。

b. 主损失为交叉熵损失，风格损失通过计算生成文本与目标人物词频分布的KL散度获得，加权系数为0.3。

c. 数据集按9:1划分为训练集与验证集，使用AdamW优化器，学习率3e-4，权重衰减0.01，批量大小32，训练3轮，每轮评估风格一致性得分与困惑度。

步骤2：用户选择具有指定人物风格大模型进行对话，在指定要生成的内容的关键词后，系统根据用户提供的关键词实时多引擎爬取信息，进行多重清洗筛选后，交付与步骤1中训练出的具有个人风格的LLM，LLM一方面根据被提供的数据资料，一方面不断从步骤1建立的RAG知识库检索出与数据资料关联度高的信息。

1. 数据爬取

部署SearXNG框架并配置启用Google、Bing、百度搜索引擎，禁用广告类数据源。通过SearXNG的API接口/search提交用户关键词，设置超时时间为10秒，每页返回20条结果。请求分发时集成开源代理池proxy-pool，每个代理IP使用上限为5次，请求间隔随机延迟1-3秒以规避反爬机制。解析返回的JSON数据提取标题、正文片段、URL及发布时间，对正文内容计算64位SimHash指纹，若与已有数据指纹的汉明距离≤2则判定为重复内容并丢弃，得到与关键词相关的网页内容。

1. 清洗

第一层规则清洗调用readability-lxml库提取网页正文，删除广告、导航栏及脚本标签。通过正则表达式清洗连续换行符、火星文符号及非常规标点，统一日期格式为“YYYY-MM”。第二层LLM清洗通过DeepSeek的API接口执行，温度参数设为0.1，输入Prompt指令为“删除无关段落并修正拼写错误，保留口语化表达，仅输出清洗后文本”，输出结果保留与关键词强相关的段落。

1. 生成与关键词相关的文章

将清洗后的文本与用户关键词拼接为查询语句，输入RAG知识库执行混合检索，按综合得分召回Top 3相关文档并截取最相关段落。生成输入格式为“[背景资料]{纯净资料}[历史观点]{检索段落}请以目标人物风格撰写文章。”，模型推理参数设置为温度0.7、重复惩罚系数1.2、最大长度1000字符，确保生成文本兼具创意与风格一致性。

四、实施示例

为了进一步描述本发明的技术特点和效果，以下结合附图和具体实施方式对本发明做进一步描述。