

基于 NLP 的园区多语言智能服务系统与方言保护融合实践

沈晓琴

杭州商旗科技有限公司, 浙江杭州, 310000;

摘要: 随着国家新型智慧城市与数字园区建设的深入推进, 多语言服务已成为智慧园区国际化、包容性和服务精细化的重要方向。自然语言处理(NLP)技术的快速发展为园区多语种智能服务系统提供了强有力的技术支撑。与此同时, 区域方言作为文化多样性的载体, 正面临加速消亡的风险。本文从多语言服务与方言保护的双重需求出发, 探讨基于NLP的园区智能服务系统在实现语言平权、跨文化沟通与文化传承中的融合策略。通过语义识别、情境翻译、语音识别与生成等关键技术实践, 提出一套多语言覆盖与方言数字化共建的新型路径, 推动技术与文化的双向赋能。

关键词: 自然语言处理; 智慧园区; 多语言服务; 方言保护; 语言数字化; 语义计算

DOI: 10.69979/3041-0673.25.10.054

1 智慧园区多语言智能服务系统的构建逻辑

1.1 多语种用户需求驱动下的服务体系重构

随着园区朝着“国际化”以及“数字化”的方向加快推进进程, 多语种的使用场景早已不再仅仅局限于那一小部分语言学习者或者外籍人员了, 其已经广泛涉及到技术外包、跨境电商、国际研发协同、外派工程等等诸多的业务层面。系统建设不能再局限于以“普通话+英文”为主的那种服务框架了, 而应当朝着一个多元语种体系去转变, 这个体系要涵盖汉语少数民族语种, 像藏语、维吾尔语、壮语等等, 也要涵盖主要的国际语种, 比如英语、日语、韩语、俄语, 还得涵盖主要的国内方言, 像是粤语、吴语、闽南语、赣语等。系统不光得支持文本界面的翻译工作, 而且还要在语音输入识别、语义理解、语境交互、情感判断以及本地化回复等诸多层面都能有所覆盖。从用户的角度来讲, 要达成“无感切换”以及“个性定制”的效果, 也就是说系统要能够依据用户的语言偏好或者语种识别的结果, 自动地去调整交互语言, 并且要在不影响整体流程的基础之上, 完成对话翻译以及指令的执行操作。像这样以需求作为驱动来进行的体系重构, 不但体现出了服务的精细程度以及包容特性, 同时也为在技术层面实现语言平等提供了一个可以落脚的地方。

1.2 自然语言处理技术在多语环境下的关键应用

NLP作为多语言服务系统的根基性引擎, 其涵盖了分词、词性标注、命名实体识别、依存句法分析、语义匹配以及对话生成等诸多子任务。在多语种的实际环境

当中, 传统的NLP模型暴露出训练语料不均衡、针对低资源语种的建模能力欠佳等一系列问题。要想强化其对非主流语言的处理效能, 就必须依靠跨语言迁移学习、像mBERT、XLM-R这类多语言预训练模型以及语义对齐机制等相关技术路径。并且, 像BERT这样基于上下文的动态词向量技术, 也为实现多语言语义的一致性给出了可行的建模办法。与此同时, 端到端语音识别与语音合成技术能够对语音交互的实时处理能力予以支持。当下, 通过把大语言模型(LLMs)和小模型部署(TinyML)相结合所形成的混合式架构, 可以促使多语言处理在本地边缘设备以及云端处理当中达成协同运行的状态, 进而为园区赋予高性能、低延迟且契合场景的语言服务能力。

1.3 场景化智能语言服务的模块化建设策略

要想让智慧园区多语服务系统高效落地实施, 那就得采用这样一种体系架构, 即把“语种适配”“场景驱动”以及“功能模块化”这几方面结合起来。先说语种适配层, 它的主要职责就是能够自动去检测出用户在语言方面的偏好, 然后据此调取与之相对应的语种模型。再看场景驱动层, 它是通过接口来调用不同的业务模块的, 像咨询答疑呀、设施导航呀、投诉建议呀、政策解读等等这些业务模块都在其调用范围内。而功能模块层呢, 则会依据具体的任务需求, 把文本生成、语音识别、情感分析等这类自然语言处理的子模块给嵌入进去。举个例子来说吧, 在园区的访客导航系统当中, 语音助手是可以识别粤语发音指令的, 并且还能结合方位图来进行本地化的语音播报呢。再比如在园区企业服务终端这

里，系统能够依据企业登记所在的国家，自动去设定默认语言，而且还可以实现合同条款的双语对照生成这样的功能。通过构建起这种模块化的、能够被复用的语言服务组件，园区就能够比较灵活地去扩展语种的范围，也能快速地去适配各种各样的业务需求，进而构建起一个可以持续演化发展的智能语言服务生态环境。

2 园区方言保护的技术路径与融合机遇

2.1 方言资源濒危现状与数字化保护的紧迫性

中国乃是世界上语言资源极为丰富的诸多国家之一，拥有着数量多达数百种的地方方言。不过，因为普通话大力推广、城市化不断推进以及代际之间出现断层等因素的影响，方言的使用人群以及使用场景正以较快的速度在不断萎缩。有大量属于非通用语种的方言，其语音资料都还未曾被记录下来，而且这些方言的语法结构、词汇系统以及表达习惯也近乎处在被人遗忘的边缘境地。园区作为区域文化聚集的重要场所和传播的关键平台，如果能够把语言技术引入到方言资源的采集环节、归档工作以及活化处理当中去，那么毫无疑问将会变成推动地方文化再次焕发生机的一股重要力量。特别是在园区开展文化展示活动、进行对外交流以及针对青年展开教育等诸多场景之中，对方言加以灵活运用并且实现可视化的呈现，这不仅仅是展现文化自信的一种方式，而且还是构建具有差异化的园区品牌认同的一种行之有效的途径。所以说，数字化以及智能化的语言技术应当担负起保护方言的全新使命，实现从‘存量保存’到‘增量再生’的转变。

2.2 基于 NLP 的方言语料采集与建模技术路径

若要达成方言的智能识别以及交互式服务这一目标，首要之事便是去构建起一个高质量的语料资源库。能够凭借智慧园区其自身所具备的社区资源，还有职工呈现出来的多样性以及当地开展的各类在地文化活动，来构建起一套‘众包式’的方言采集机制：借助 APP 端发布语音任务、进行方言录音打卡以及开展互动故事接龙等多种方式，去激励园区的用户上传那些带有代表性的口音样本以及文本语料。在完成采集之后，经由音频自动转写、针对同音词展开纠错、就声学特征来开展建模等一系列的步骤，进而实现初步的语音识别模型的构建工作；随后再把多任务学习与领域自适应算法相结合起来，以此来提升针对发音变异以及句法结构所具有的特殊性的容错能力。就建模这个层面来讲，可以专门针对高频词汇、那些固定的短语以及情境对话去设计较为

轻量化的语义模型，让其即便在资源受限的设备之上也能够正常运行起来，从而推动方言识别系统在园区终端、智能机器人、语音导览等诸多场景当中实现低门槛的部署操作。

2.3 方言文化场景嵌入园区智能交互系统实践

在有技术给予支持的情形之下，园区能够把方言元素很自然地融入到日常服务以及文化传播里面去，进而形成一条关于方言再生的链条，其呈现出‘服务—认知—认同’这样的模式。就好比说在智慧屏当中引入‘本地方言问候语’这一内容，如此一来，每一位进入园区的访客都能够有机会去选择体验一回十分地道的方言问候。又比如在企业展厅之中配置‘方言故事播报’这样的一个模块，借助虚拟讲解人来讲述园区所在地的那些民间传说呀、地名的由来呀、饮食方面的风俗等等情况。再比如在青年文化沙龙里面，结合语音识别还有翻译技术，去实现‘普通话—方言’双语互译的游戏活动，以此来激发年轻人对于本土文化所抱有的兴趣。除此之外呢，园区还可以凭借数字人技术去开发出‘方言数字 IP 形象’，像是‘上海话讲解员’‘客家话调解官’这类的，通过那种拟人化的语言交互以及文化表达的方式，让用户对方言能够增添一份亲近之感。通过这一连串的应用举措，园区在为大家提供具有实用性的语言服务的同时，也肩负起了‘语言活化场’所赋予的文化使命。

3 多语言服务与方言保护的融合创新路径

3.1 构建多语—方言协同的语言服务中台

当面对智慧园区那愈发繁杂的语言服务需求状况时，去构建起一个能够让多语和方言协同起来的统一语言服务中台，这可不单单是技术集成方面的一种现实可行的选择，实际上更是达成语言平等、达成资源共享以及实现跨语种高效处理的极为重要的路径呀。这个中台必须得突破以往存在的那种‘多语言处理孤岛化’的架构方面所面临的困境，要把主流语种和边缘语种之间存在的信息壁垒给彻底打破掉，要将普通话、国际语种、民族语言以及地方方言全部都纳入到同一个调度与处理的框架里面去呢。在数据接入这个层面上，中台理应具备对多源异构数据的接入能力，要能够支持像文本、语音、语义标签、语料对齐等多种不同类型的语言资源的接入操作以及标准化的处理流程，从而构建起一个有着动态学习能力的语言知识图谱；而在服务调度层这里呢，就需要凭借语种识别以及语境判别这类技术，来实

现智能路由的机制，要依据用户的身份、所处的位置、语言方面的偏好等等因素，自动地去调用相应的模型来展开处理以及给出响应；到了模型管理层的话，那么它就应当具备模型加载以及热切换的机制，以便让主流语种的大模型、方言的小模型都可以依据具体的场景来实现灵活的调配以及资源的复用操作啦。

与此同时，语言服务中台还应当内置“语种效能评估系统”，凭借“语言资源评价指数”来对不同语种模型于响应速度、用户满意程度、识别准确程度等诸多维度的运行状况加以监控，进而达成闭环式的迭代优化。借助中台针对语言服务展开的集中统筹以及灵活调用操作，园区在确保服务质量和效率的情形下，能够推动像方言这类非标准语种在技术处理层面达成与主流语言相同的接入条件以及公平的待遇，从而构建起一个极具包容性、多种语言共同生存的语言服务生态体系。

3.2 智能语音设备中的语言个性化适配机制

伴随智能语音设备在园区场景里得以广泛布设，像导览机器人、自助问询终端以及语音控制面板等等这类设备，语言交互所呈现出的质量状况以及展现出的亲和程度，已然变成了左右用户体验的极为关键的指标。在多种语言以及多方言共同存在的这样一种环境之下，对于系统而言，怎样去精准识别并且妥善适配用户的语言特征，这无疑是达成精准服务以及给予文化关怀的在技术层面的关键要点。

当访客服务机器人察觉到用户是用粤语发出指令之时，系统能够自动地切换至粤语语音播报模式，并且同时启用粤语文本提示的界面，如此一来，便达成了内容理解以及反馈在语言方面的一致性。在政务咨询或者是投诉处理这类场景当中，系统可以融合情感识别方面的技术，以此来判定用户的语气以及情绪状态，再借助语言模型去匹配具备相同方言能力的数字客服，进而完成相应的回应操作，使得交流过程中的理解度以及情绪包容性都得以提升。个性化的语言适配不单单是对服务交互进行了优化，而且还在不知不觉间传递出系统对于使用者文化身份的尊重与理解之意，推动着园区的语言系统从“选择式多语”这种状态朝着“感知式语态协同”的方向不断演进，进而构建起一个以人为核心、让语言充满温度的智慧语音生态环境。

3.3 技术赋能语言文化共同体建设路径探索

最终要达成的目标可不单单是提供服务，而是要凭

借语言技术赋予园区能量，从而构建起语言文化共同体，塑造出那种让人能‘听得懂’、涵盖多种语种且充满温度的人文生态。园区能够依据语言行为方面的数据来开展语言地图的可视化分析工作，以此去揭示不同语言的使用分布状况以及社群聚集的相关特征。园区还会定期举办诸如‘语言开放日’还有‘方言节’这类活动，使得有着不同语言背景的人员可以借助AI翻译来实现文化方面的相互沟通以及情感层面的交流。并且在园区的媒体平台上开设‘语言共创实验室’，以此来鼓励用户积极参与到语料贡献、译文优化以及方言数字化共建这些事情当中。这样一个语言文化共同体一旦形成，不但有助于提升园区内部的凝聚力，而且还能为中国多语种城市的建设给出范式方面的示范，切实达成技术与文化之间相互支撑以及共生发展的局面。

4 结语

语言是连接人类认知、社会行为与文化记忆的核心媒介。在智慧园区发展不断加速的今天，基于NLP的多语言服务系统不仅应解决“能听懂”的问题，更要思考如何“听得准”“说得亲”“用得活”。将语言技术与文化保护深度融合，不仅能提升园区服务质量与国际形象，更可为地方语言的活化与传承开辟新路径。未来，随着大模型、语音智能与语义网的持续演进，多语与方言协同系统将在政务、教育、旅游、社区等更多场景中落地生根，绘制出一幅“语言多样性与技术普惠性共荣共生”的数字园区新图景。

参考文献

- [1]高雅奇.基于大语言模型和RAG技术的高校知识库智能问答系统构建与评价[J].电脑知识与技术,2024,20(29):18-20+38.
- [2]帕拉什·戈雅尔,苏米特·潘迪,卡兰·贾恩.面向自然语言处理的深度学习[M].机械工业出版社:202207.320.
- [3]王楠楠.面向垂直领域的知识图谱构建和智能问答系统设计与实现[D].北京邮电大学,2022.
- [4]王一成,马宁,万福成.面向知识图谱的多语言电子商务智能系统研究[J].西北民族大学学报(自然科学版),2019,40(03):64-69.
- [5]周建设,张凯,罗茵,等.语言智能评测理论研究与技术应用——以英语作文智能评测系统为例[J].语言战略研究,2017,2(05):12-19.