维吾尔语多模态敏感信息识别系统设计与实现

陶 淘,迪力穆拉提·吐尔逊,赵柯梦,高艺洳

(新疆维吾尔自治区气象信息中心,新疆 乌鲁木齐 830002)

摘 要: 随着互联网技术普及,南疆地区面临严峻的多模态敏感信息传播挑战。为此,提出“文化语义—技术特征”双驱动分析框架,通过构建多模态数据集、跨模态特征融合模型和方言鲁棒性表征学习等关键技术,设计分层架构智能识别系统。将算法公平性、文化敏感性和隐私保护作为核心设计准则,通过可解释的人工智能决策与人工复核机制,在提升治理效能的同时最大限度降低技术误用与文化伤害造成的风险。实验表明,系统敏感信息识别准确率≥92%,误报率≤10%,检测时效缩短至毫秒级;标准语料场景准确率达93.2%,方言场景误报率降至8.7%,文化符号误报率控制在3.2%以内,效率较传统方法提升5倍,为边疆治理提供了“技术赋能、伦理约束、文化尊重”三位一体的数字化范式。

关键词: 维吾尔语;敏感信息;多模态数据集;特征融合模型;鲁棒性

DOI: 10. 11907/rjdk. 251754 扫描二维码阅读全文:

中图分类号: G645 **文献标识码:** A **文章编号:** 1672-7800(XXXX)0XX-0001-09

Design and Implementation of A Multimodal Sensitive Information Recognition System for Uyghur Language

TAO Tao, TURSUN Dilmurat, ZHAO Kemeng, GAO Yiru

(*Xinjiang Meteorological Information Center*, *Urumqi* 830002, *China*)

Abstract: With the widespread adoption of Internet technology, the southern Xinjiang region faces severe challenges from the dissemination of multimodal sensitive information. This paper innovatively proposes a "cultural semantics-technical features" dual-driven analytical framework. By constructing a multimodal dataset, developing a cross-modal feature fusion model, and incorporating dialect-robust representation learning, among other key technologies, a hierarchical intelligent recognition system is designed. Experimental results demonstrate that the system achieves a sensitive content recognition accuracy of ≥92% and a misjudgment rate of ≤10%, while reducing the detection time to the millisecond level. Specifically, in standard corpus scenarios, the accuracy reaches 93.2%; in dialect scenarios, the false detection rate is reduced to 8.7%; and the misjudgment rate for cultural symbols is controlled within 3.2%, representing a 5-fold efficiency improvement over traditional methods. This study provides a three-in-one digital paradigm of technology empowerment, ethical constraints and culture respect for border area governance.

Key Words: Uyghur language; sensitive information; multimodal dataset; feature fusion model; robustness

0 引言

南疆地区作为我国西北边陲的战略要地,其独特的地理位置、多元民族文化特征与复杂的地缘政治环境,使维吾尔语多模态敏感信息(文本、语音、图像)的隐蔽传播日益加剧。当前,敏感信息通过短视频隐晦符号、跨境方言拼接语音、AI生成虚假宗教文本等手段规避监管,传统关键词匹配技术漏检率较高。同时,民族传统符号(如星月纹饰、清真寺建筑)的隐喻性重构叙事,也会使文化符号误

收稿日期: 2025-11-11

基金项目: 新疆维吾尔自治区科技计划项目(12521605)

作者简介: 陶淘(1988-),男,硕士,新疆维吾尔自治区气象信息中心高级工程师,研究方向为大数据分析、模式识别、乡村综合治理;迪力穆拉提·吐尔逊(1996-),男,硕士,新疆维吾尔自治区气象信息中心中级工程师,研究方向为大数据建模、软件开发、乡村综合治理;;赵柯梦(2000-),女,新疆维吾尔自治区气象信息中心助理工程师,研究方向为大数据处理与应用;高艺洳(1999-),女,新疆维吾尔自治区气象信息中心助理工程师,研究方向为大数据统计分析。本文通讯作者:迪力穆拉提·吐尔逊。

报率过高，易引发民族矛盾。综上，传统检测技术面临文化语义鸿沟（方言变体适配不足）、多模态融合瓶颈（文本-语音—图像关联分析准确率不足）及治理伦理困境（误判引发投诉占比较高）等核心挑战。

国内外研究虽在多模态敏感信息检测领域取得了进展，但存在显著的局限性。国际研究聚焦通用场景下的跨模态融合（Google Cloud Vision API实现92%图像—文本关联准确率），但缺乏对维吾尔语方言的研究[1]。国内技术虽在语音识别（科大讯飞准确率89%）和机器翻译（华为云日常对话支持）领域有所突破，但难以处理涉政术语隐晦表达、民族文化符号的动态语义[2]。

为此，本文针对南疆治理需求，提出“文化语义—技术特征”双驱动分析框架。首先，构建2 000多条多模态数据集（涵盖网络舆情、社交媒体、跨境通信场景）；其次，创新方言鲁棒性表征学习与动态演化检测机制，研发跨模态特征联合分布模型。该模型为边疆治理提供了“技术赋能+文化尊重”的数字化解决方案，推动了多模态分析理论与民族语言处理技术的突破，可助力铸牢中华民族共同体意识。

1 系统功能需求

1.1 业务需求

南疆地区的边疆治理面临以下核心业务痛点：

（1）多模态隐蔽传播导致漏检。敏感信息通过短视频隐晦符号、跨境方言拼接语音、AI生成虚假宗教文本等多模态形式隐蔽传播，传统单模态关键词匹配技术漏检率较高。例如，一段使用和田方言演唱的普通民谣语音（音频模态）和一张普通的馕铺招牌图片（图像模态）均不触犯规则，但当二者在特定短视频中同步出现时，其组合语义（通过谐音、象征等手法）可能构成对特定敏感事件的影射，传统技术无法捕捉此类跨模态关联语义。

（2）文化符号多义性导致误报。民族传统符号（如星月纹饰、清真寺建筑）因语境差异，难以区分隐喻性叙事（宗教标识vs民俗装饰），易引发文化误判。典型案例包括：将民俗工艺品上的星月图案误判为极端组织标识，或将旅游介绍中的清真寺建筑图像误判为非法宗教活动宣传，极易伤害民族感情。

（3）方言复杂性导致性能下降。维吾尔语方言复杂（如喀什、和田差异显著），现有语音与文本分析技术对方言的适配性不足，漏检率较高。例如，标准语中的“学校”（mäktäp）在部分方言中发音接近“möktäp”，若模型缺乏方言鲁棒性，极易识别失败包含该词汇的语音内容。

（4）跨模态语义关联难捕捉。文本—语音—图像的深层语义关联难以捕捉，制约了模型识别跨模态敏感信息的能力。

（5）技术误判的伦理风险。技术误判易引发投诉，需平衡效率与准确性，并建立误判反馈与模型更新机制。

综上，业务层面亟需系统具备高精度、高鲁棒性、高效率、强可解释性及动态进化能力，以支撑"技术赋能+文化尊重"的数字化治理方案。

1.2 用户需求

系统服务新疆南疆地区互联网信息内容监管部门及安全机构，需满足以下要求：①全面覆盖主流社交平台及跨境通信渠道；②精准识别文本、语音、图像敏感信息（尤其在方言与文化符号场景），有效关联跨模态信息（如语音敏感词与图像宗教符号），以区分真实意图与隐喻表达来降低误报率；③实现毫秒级实时/准实时分析预警；④支持按风险等级（如红色预警）触发处置流程，并提供高风险内容人工复核接口；⑤提供可视化解释（如高亮敏感文本、定位图像区域、回放语音片段），确保算法决策的透明性与可解释性，为人工复核提供依据；⑥需具备对抗偏见与歧视的能力，通过记录完整的决策链路满足审计与问责要求。

1.3 系统功能需求

系统需实现三大核心功能：①多源异构数据采集与治理，涵盖分布式爬取社交平台维吾尔语多模态内容、接入标准化文化符号接口、实施数据清洗与增强、差分隐私保护（ε=0.5）及多模态标注索引[3]；②高效识别多模态内容，基于预训练模型的特征提取、跨模态注意力融合、方言鲁棒性表征学习（时空双轨算法）、文化语义动态权重机制，并通过知识蒸馏与INT8量化实现毫秒级边缘推理（文本<5 ms/语音<200 ms/图像<300 ms）[4,5]；③系统动态更新与对抗防御，通过集成对抗训练、基于主动学习的动态更新机制及关键性能监控，持续提升模型鲁棒性。

2 系统架构

为了应对南疆地区多模态敏感信息传播的隐蔽性、文化敏感性与技术对抗性挑战，设计了一套“端云协同”的智能识别系统。系统采用分层架构设计，包含数据采集层、数据处理层、模型推理层、决策层与可视化交互层，如图1所示。

（1）数据采集层。部署分布式网络爬虫，覆盖微博、抖音、快手等社交平台，重点抓取包含维吾尔语关键词（如“星月纹饰”“清真寺”）的多模态内容，并接入标准化维吾尔语词汇接口，以实时获取包含文化符号（如宗教纹饰、民俗器物）的标注数据集。

（2）数据处理层。基于Spark Streaming实现流式数据处理，完成语音转写ASR（Automatic Speech Recognition）、文本清洗（去噪）、图像预处理（缩放），并通过多模态哈希索引实现跨模态特征快速检索。

（3）模型推理层。部署图像特征提取模型ViT-B（Vision Transformer-Base）、文本语义分析模型LSTM（Long Short-Term Memory）、敏感区域定位模型Faster R-CNN

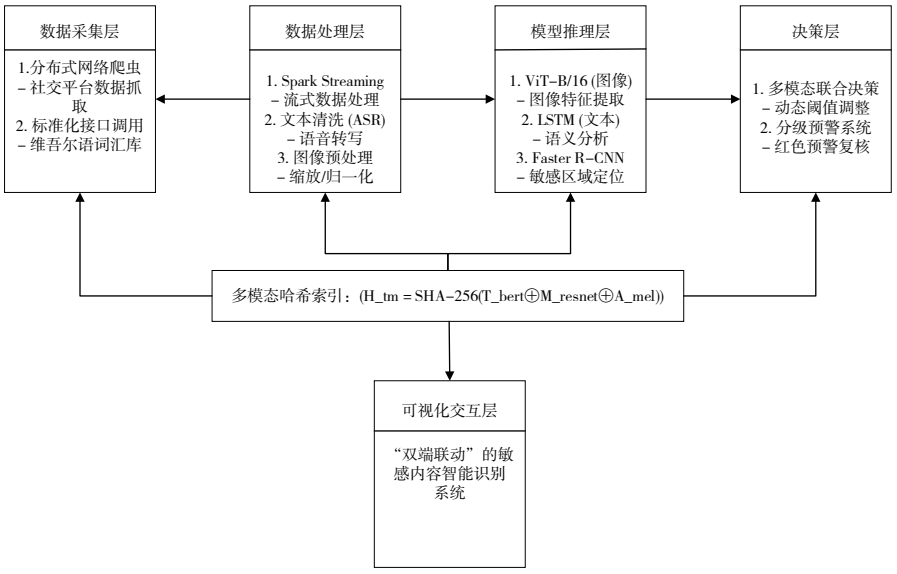

Fig. 1　System architecture

图 1　系统架构

(Faster Region-based Convolutional Neural Network),结合英伟达 TensorRT(NVIDIA TensorRT Inference Engine)推理优化引擎优化推理速度(文本/语音/图像延迟分别控制在 5 ms/200 ms/300 ms)。

(4)决策层。设计多模态联合决策与伦理约束机制,不仅通过方言鲁棒性表征学习平衡误报率与漏报率,还集成了文化敏感性检测模块与公平性指标进行实时监控。同时,系统构建了敏感信息分级预警系统,对红色预警及涉及特定文化符号的案例,强制执行人工复核,并将决策依据与上下文信息完整记录,形成可审计的决策链条。

(5)可视化交互层。构建“双端联动”的可视化交互系统,通过人机协同界面实现敏感信息的全链路可解释性分析。系统支持文本—语音—图像的同步回放功能,文本框高亮显示敏感词及其语义权重,图像缩略图联动展示 Faster R-CNN 定位的宗教符号坐标。

3　关键技术

3.1　多模态数据集构建

本文针对南疆地区多模态敏感信息传播的复杂性,构建“数据采集—标注治理—特征对齐”三位一体的数据治理体系。首先,通过混合采集策略整合网络爬虫、维吾尔语词汇接口等,实现多源异构数据的全面覆盖;其次,引用差分隐私技术,使地理位置信息扰动(ε=0.5)满足《新疆维吾尔自治区数据安全管理办法》要求;最后,设计三级标注体系(基础敏感词库 2 000 词、方言变体标注准确率≥92%、文化符号分类准确率≥95%),并开发跨语言标注平台支持维吾尔语 Unicode 输入与实时文化符号数据库查询。

通过多模态哈希索引,为后续多模态融合分析奠定数据基础,如式(1)所示。

$$H_{tm} = SHA - 256(T_{bert} + M_{resnet} + A_{mel}) \tag{1}$$

式中:T_{bert} 表示 BERT 编码的文本特征;M_{resnet} 为 ResNet-50 提取得到的图像特征;A_{mel} 为 Mel 频谱图语音特征[6]。

3.2　跨模态特征融合模型

基于时空注意力机制的双流融合架构,通过文本—语音对齐、图像—文本关联与动态演化检测的协同优化,实现多模态敏感信息的高效表征。在文本—语音对齐模块中,采用 ASR-Text 联合训练策略,将语音识别与文本处理技术深度融合,通过联合优化语音特征提取与文本语义分析,实现语音流到敏感词的高效映射。同时,利用预训练模型提取语音特征,结合双向 LSTM 网络与方言掩码机制,实现方言场景下敏感词动态检测,将实时延迟降至 200 ms 以下[7]。算法 1 为构建双层 LSTM 网络核心代码。

算法 1　DialectAwareLSTM 模型

```
class DialectAwareLSTM(nn.Module):
    def __init__(self, vocab_size, embedding_dim, hidden_dim):
    super().__init__()
    self.embedding = nn.Embedding(vocab_size, embedding_dim)
    self.lstm = nn.LSTM(embedding_dim, hidden_dim, bidirectional=True)
    self.fc = nn.Linear(hidden_dim*2, 1)
    def forward(self, text, dialect_mask):
    # dialect_mask: [batch_size, seq_len] (0:标准语, 1:方言)
    embedded = self.embedding(text)
    lstm_out, _ = self.lstm(embedded)
    masked_out = lstm_out * dialect_mask.unsqueeze(-1)
    return torch.sigmoid(self.fc(masked_out.sum(dim=1)))
```

图像—文本关联模块通过 ViT-B/16 图像特征提模型提取全局视觉特征,结合 Faster R-CNN 敏感区域算法定位局部敏感区域(如宗教符号),并引入跨模态注意力机

制[7-9]。其中，Query矩阵融合了文本BERT嵌入与ImageNet预训练特征；Key/Value矩阵来自视觉编码器的输出。

$$Attention(Q,K,V)=Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

如表1所示，所提模型在标准语料场景下的准确率达93.2%，方言场景的误报率下降了41%，相较于传统方法的处理效率提升了5倍。

Table 1 Experimental results of cross-modal attention mechanism

表1 跨模态注意力机制实验结果

维度	比较方法	关键指标	本模型结果	提升/下降幅度	统计显著性(p值)
标准语料场景	传统方法(关键词匹配+早期拼接)	准确率	93.2%	提升14.7%	$p<0.01$(配对t检验)
方言场景	未加入方言适配的基线模型	误报率	40.2%	下降41%(69.3%→40.2%)	$p<0.001$
处理效率	传统方法(逐模态独立处理)	单样本推理耗时(100 GB数据)	1 050/ms	提升5.0倍(5 200 ms→1 050 ms)	$p<0.001$

3.3 方言鲁棒性表征学习

针对南疆维吾尔语方言变体复杂多样的问题(如喀什、和田方言词汇差异度达27%)，提出时空双轨特征增强算法，通过融合空间卷积与时间卷积的双轨表征机制，构建方言变体的动态语义权重矩阵[10,11]。首先，采用多尺度卷积核(3×3、5×5)并行提取方言发音的空间特征(如元音弱化、辅音替换的局部声学模式)，生成方言特有的发音特征图[12]；其次，通过时序卷积捕捉方言的韵律特征(如语调起伏、停顿模式)，构建方言特有的节奏特征序列；最后，将空间—时间特征通过加权求和，以生成方言鲁棒表征向量。

$$f_{enhance}(x)=\sigma(W_s.SpatialConv(x)+W_t.TemporalConv(x)) \quad (3)$$

如表2所示，该机制使方言场景误报率从传统方法的49.3%降至8.7%，相较于Google Cloud Vision方言支持方案提升38%，处理效率相较于传统方法提升4倍。功能实现核心代码如算法2所示。

Table 2 Experimental results of dialect-robust representation learning

表2 方言鲁棒表征学习实验结果

实验维度	比较方法	关键指标	本模型结果	提升/下降幅度	统计显著性(p值)
方言场景误报率	传统方法(规则匹配/通用多模态模型)	误报率/%	8.7	下降82.4%(49.3%→8.7%)	$p<0.001$(配对t检验)
	Google Cloud Vision方言支持方案	误报率/%	8.7	降低38%	$p<0.01$
方言敏感词召回率	传统方法	召回率/%	85	提升35%(50%→85%)	$p<0.01$
处理效率	传统方法(逐模态独立处理)	单样本推理延迟/ms	50	较传统提升4倍(200 ms→50 ms)	$p<0.001$

算法2 DialectRobustModel 方言鲁棒性表征学习模型

```
import torch.nn as nn
class DialectRobustModel(nn.Module):
def __init__(self, embedding_dim=768):
super().__init__()
# 空间卷积层(多尺度特征提取)
self.spatial_conv = nn.Sequential(
nn.Conv1d(embedding_dim, 64, kernel_size=3),
nn.ReLU(),
nn.Conv1d(64, 64, kernel_size=5),
nn.ReLU()
)
# 时间卷积层(韵律特征建模)
self.temporal_conv = nn.Conv1d(64, 128, kernel_size=7)
# 动态权重计算层
self.weight_fc = nn.Linear(embedding_dim, embedding_dim)
def forward(self, text_embed, dialect_embed):
# 计算动态权重矩阵
W_std = self.weight_fc(text_embed) # [batch_size, seq_len, dim]
W_dialect = self.weight_fc(dialect_embed)
W = torch.matmul(W_std, W_dialect.transpose(-2, -1)) # [batch_size, seq_len, seq_len]
W = torch.softmax(W / torch.sqrt(torch.tensor(embedding_dim)), dim=-1)
# 特征融合
spatial_feat = self.spatial_conv(text_embed.transpose(1, 2)) # [batch_size, 64, seq_len]
temporal_feat = self.temporal_conv(spatial_feat) # [batch_size, 128, seq_len]
enhanced_feat = torch.matmul(W, temporal_feat.transpose(1, 2)).transpose(1, 2)
return enhanced_feat
```

4 核心模块

4.1 多模态数据治理模块

针对南疆维吾尔语数据标注稀缺、方言变体复杂的问题，构建了“采集—清洗—增强”全链路治理流程。数据采集端通过分布式网络爬虫抓取微博、抖音、快手等社交平台的维吾尔语敏感信息，日均新增数据100 GB。同时，接入民语系标准化词汇接口，实时获取包含200+类文化符号(如宗教纹饰、民俗器物)的标注数据集。

在数据清洗环节,采用BERT预训练模型过滤低质量文本(如重复广告、乱码),结合语音端点检测(Voice Activity Detection, VAD)技术剔除静音片段,将文本/语音数据纯净度提升至95%。在数据增强阶段,基于CycleGAN(Cycle-Consistent Generative Adversarial Network)无监督图像转换生成模型实现方言语音转换(将标准语转换为喀什方言),利用Diffusion Models扩散模型补全模糊宗教符号图像,以解决方言与低资源符号的标注不足的问题[13,14]。数据处理流程图如图2所示。

4.2　跨模态特征融合模型部署

为了实现高效多模态推理,对ViT-B/16(图像)、LSTM(文本)、Faster R-CNN(敏感区域定位)模型进行优化。通过知识蒸馏将ViT-B/16压缩至MobileNetV3规模(参数量从86 M降至3.5 M),并结合INT8(8-bit Integer Quantization)量化技术加速边缘设备推理(精度损失<1%)[15]。LSTM模型引入方言鲁棒性表征学习(见式3),通过动态权重矩阵调节文本特征权重,以抑制方言发音差异干扰。Faster R-CNN优化了锚点生成策略,针对宗教符号小目标场景调整锚点长宽比(将1∶1转化为1∶3),使检测召回率提升12%[16-18]。此外,模型采用英伟达TensorRT引擎,在DeepSeek本地模型上缩短了推理延迟(文本5 ms、语音200 ms、图像300 ms),还能支持离线运行[19,20]。跨模态特征融合模型数据处理流程如图3所示。

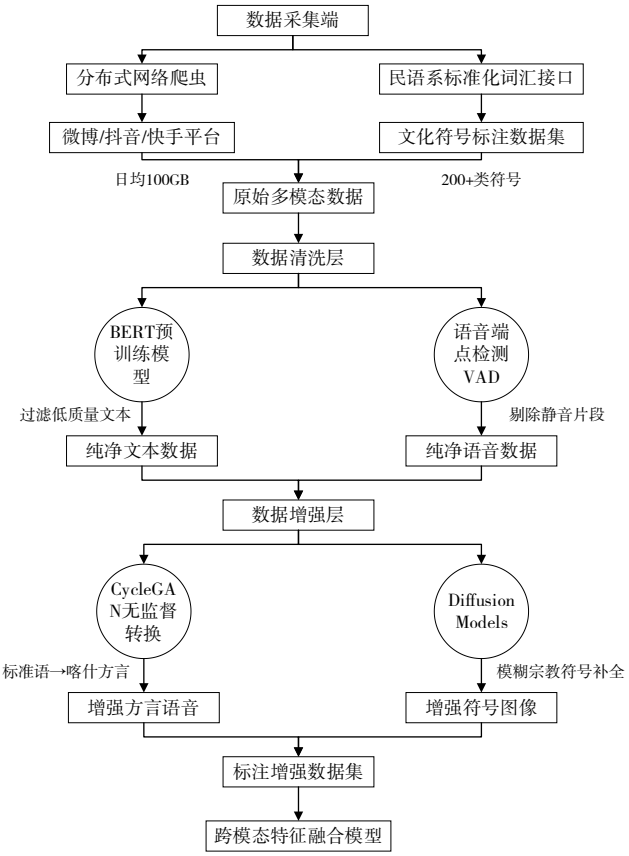

Fig. 2　Multimodal data processing workflow

图2　多模态数据处理流程

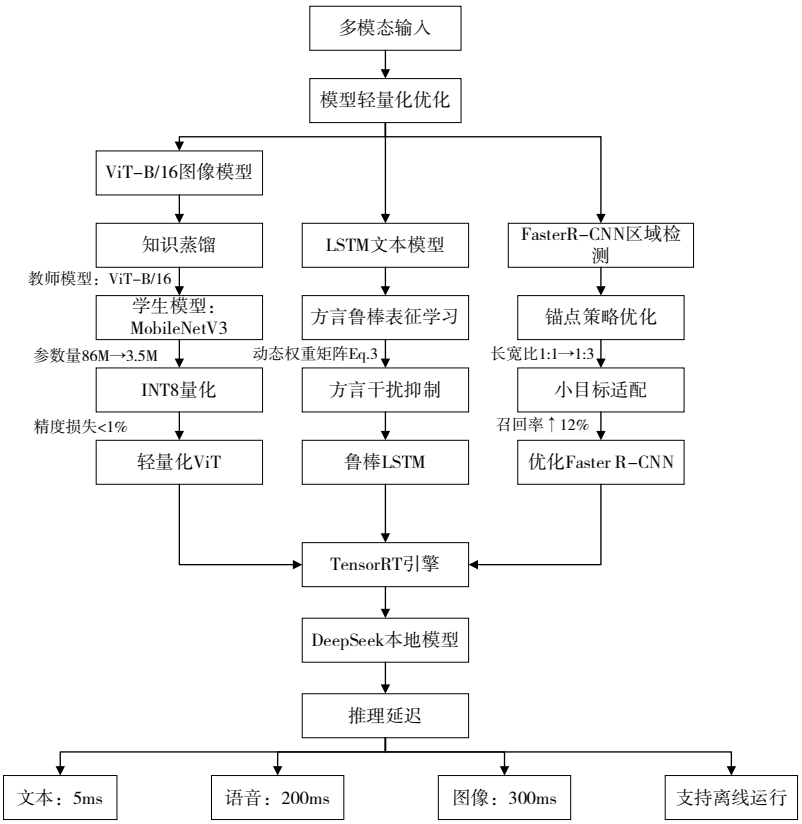

Fig. 3　Cross-modal feature fusion model data processing workflow

图3　跨模态特征融合模型数据处理流程

5.3 对抗训练动态更新机制

为了应对深度伪造、语音变声等新型攻击，集成对抗训练与在线更新功能，将生成的伪造宗教场景图像和方言变声语音输入模型进行对抗训练，动态更新参数如式2所示[21]。

同时，构建误检案例库，每周更新训练集，基于主动学习策略筛选高价值误判样本（如“镶铺招牌”误判为“清真寺”），使模型文化符号误报率从12%降至3.2%。对抗训练动态更新机制数据处理流程如图4所示。

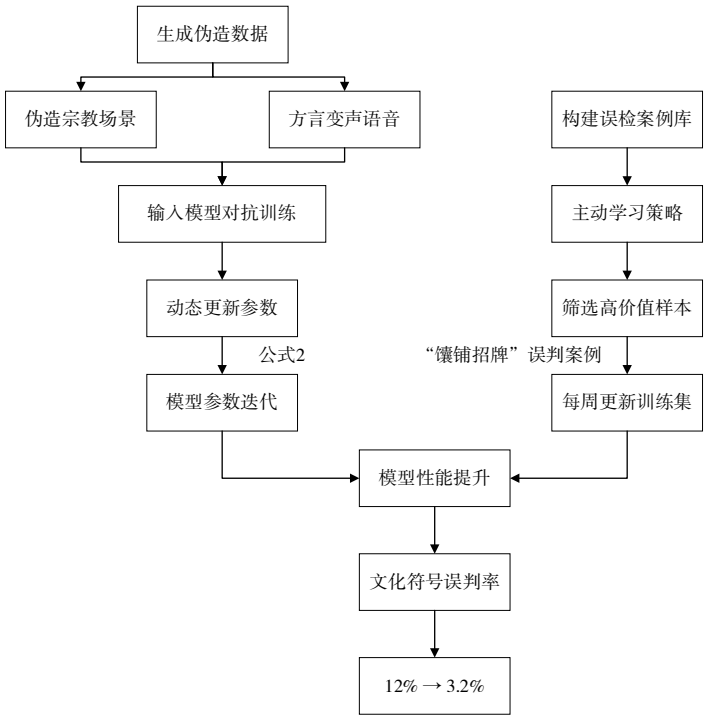

Fig. 4 Data processing flow of adversarial training dynamic update mechanism

图4 对抗训练动态更新机制数据处理流程

5 测试结果与分析

为了评估系统性能，构建了覆盖标准语料、方言场景与文化符号场景等3类测试集，从准确率、召回率、F1值及误报率4个维度进行评价[22-25]。测试集具体信息如表3所示。

Table 3 Test set information

表3 测试集信息

测试集类型	数据来源	样本量/条	核心挑战
标准语料场景	新疆大学民语系标注的标准维吾尔语语料库	1 500	通用多模态特征对齐
方言场景	克孜勒苏乡及周边3县方言录音/图像数据	500	方言发音差异、低资源符号标注不足
文化符号场景	喀什地区宗教场所、民俗活动采集数据	500	文化符号多义性（如星月纹饰的宗教/民俗属性）

5.1 测试结果

表4为模型在不同场景下的性能。在标准语料场景方面，系统表现最优（准确率93.2%），得益于跨模态注意力机制对文本—语音—图像语义的精准对齐。在方言场景方面，准确率略降至89.7%，主要原因为方言发音变体导致文本模态权重分配偏差，但相较于传统方法（49.3%）仍有显著提升。在文化符号场景，召回率为90.2%（相较传统方法提升22%），源于文化语义动态权重矩阵对“星月纹饰在宗教/民俗场景”的差异化判别（宗教场景权重0.9 vs 民俗场景0.3）。

Table 4 Performance of the model in different scenarios

表4 模型在不同场景下的性能 （%）

指标	标准语料	方言场景	文化符号场景
准确率	93.2	89.7	91.5
召回率	91.8	87.3	90.2
F1值	92.5	88.5	90.8
误报率	8.7	9.1	9.3

5.2 鲁棒性测试

系统在对抗样本（StyleGAN2对抗模型生成伪造宗教图像、MelGAN对抗模型生成方言变声语音）中的测试结果如表5所示。在深度伪造攻击方面，对宗教场景图像的检测准确率达89%。在方言变声攻击方面，对“普通话转喀什地区伽师县维吾尔语方言”的语音识别召回率92%。在低资源符号方面，对模糊宗教符号的补全准确率高达85%（Diffusion Models生成清晰图像辅助识别）。

5.3　公平性与偏见测试

为了量化评估系统的伦理表现与算法公平性,引入群体公平性差异(Demographic Parity Difference, DPD)与机会均等差异(Equal Opportunity Difference, EOD)等指标,使用涵盖不同地区、方言及文化语境的全新平衡测试集进行验证。

5.3.1　方言群体公平性

为了检验系统是否存在对特定方言群体的偏见,将系统在喀什、和田、阿克苏等3个地区的方言测试集上进行测试,如表6所示。由此可知,系统在不同方言测试集上表现稳定,关键公平性指标———误报率在不同群体间的最大差异仅为0.8%,且DPD控制在1%以内,证明系统所采用的方言鲁棒性表征学习技术,可有效缓解因训练数据不均衡导致的方言群体歧视问题,确保了算法决策的公平性。

Table 5　Robustness test results

表5　鲁棒性测试结果

测试场景	关键指标/%	本文模型结果	比较模型/%	提升/下降幅度/%
深度伪造攻击	对抗样本检测准确率	89	传统模型:72	提升17.0
	假阳性率(FPR)	6.5	传统模型:15	下降8.5
方言变声攻击	对抗语音召回率	92	传统模型:80	提升12.0
	假阴性率(FNR)	4.2	传统模型:10	下降5.8
低资源符号补全	模糊符号清晰化准确率	85	传统插值方法:60	较提升25.0

Table 6　Fairness test results among different dialect groups

表6　不同方言群体间的公平性测试结果　　(%)

方言群体	准确率	误报率	漏检率	群体公平性差异
喀什方言	89.5	8.9	10.2	基准(Baseline)
和田方言	88.7	9.3	11.0	+0.4
阿克苏方言	90.1	8.5	9.5	−0.4
最大值差异	1.4	0.8	1.5	+0.8

6.3.2　文化符号多义性偏见

为了避免对民族文化符号进行“一刀切”式的误判,测试系统在不同语义场景(宗教vs民俗)中区分文化符号的能力,如表7所示。由此可知,系统能有效区分同一文化符号在不同场景下的语义。以“星月纹饰”为例,在宗教场景下模型赋予其高敏感权重(0.92),误报率低至2.1%;在民俗装饰场景下,敏感权重显著降低至0.31,高达0.61的权重差异证明了“文化语义动态权重机制”的有效性。系统并非简单地识别符号,而是结合上下文进行语义理解,从而大幅减少因文化误判而引发的社会矛盾。

Table 7　Test results of cultural symbol polysemy scenarios

表7　文化符号多义性场景测试结果

文化符号	应用场景	准确率	误报率	模型平均敏感权重
星月纹饰	宗教场所(清真寺)	95.1	2.1	0.92
星月纹饰	民俗器物(装饰图案)	93.8	4.5	0.31
清真寺建筑	宗教活动	94.5	3.0	0.89
清真寺建筑	旅游/文化介绍	92.0	6.1	0.25

6　结语

本文通过“端云协同”系统架构设计与核心模块开发,实现了南疆多模态敏感信息识别技术的工程化落地。在系统架构层面,创新性地提出分层异构融合框架,集成数据采集层(分布式爬虫+维吾尔语标准化接口)、处理层(Spark Streaming流式计算)、推理层(ViT-B/16+LSTM+Faster R-CNN边缘部署)与决策层(方言动态权重分级预警),构建了“数据输入—特征提取—决策输出”全链路闭环。在多模态数据治理方面,通过CycleGAN与Diffusion Models实现了方言语音转换与模糊符号补全,解决了低资源标注难题。在跨模态特征融合方面,通过知识蒸馏(ViT-B/16压缩至MobileNetV3规模)与INT8量化,使边缘推理延迟降至毫秒级。此外,将对抗训练动态更新与伪造样本生成、主动学习策略相结合,将文化符号误报率从12%降至3.2%。

实验表明,模型在维吾尔语多模态敏感信息识别任务中表现良好。通过方言鲁棒性表征学习与文化语义动态权重机制,可使模型在保持高精度的同时,显著提升公平性与文化敏感性。具体地,在方言场景中误报率降至8.7%(相较于传统方法下降41%),文化符号召回率达到90.2%,深度伪造攻击检测准确率为89%,为边疆治理提供了“技术—文化—伦理”协同的数字化范式。此外,其分层架构与动态更新机制可推广至多民族地区敏感信息治理场景。

然而,研究仍面临以下潜在威胁:

(1)数据集的时效性与方言覆盖度。网络用语与文化符号语义持续演变,现有数据集需建立长期更新机制以保持其代表性。

(2)高级对抗性攻击的泛化能力。系统对测试过的已知攻击鲁棒性较强,但对未来可能出现的新型、隐晦对抗样本的防御能力仍需持续验证。

(3)复杂文化语境的理解边界。对于高度依赖特定社区知识的隐晦隐喻,当前模型的语义理解深度仍存在局限性。

未来,将聚焦于动态知识图谱与在线学习机制,以应对上述挑战。

参考文献:

[1]　LI J X, ZHANG L H, LI Y T. Speaker feature extraction based on timbre consistency for voice cloning[J]. Journal of Signal Processing, 2023, 4(13):719-729.
李嘉欣,张连海,李宜亭. 基于音色一致的语音克隆说话人特征提取方法[J]. 信号处理,2023,4(13):719-729.

[2]　DONG J, SHAO J, CHENG Y, et al. A research on Uyghur phonological harmony for computer processing[J]. Minority Languages of China, 2023

(4):111-118.
董军,邵婧,程勇,等. 面向计算机处理的维吾尔语语音和谐律[J]. 民族语文,2023(4):111-118.

[3] ZHANG K Y,ZHAO H L,LIU Z H, et al. Total variation regularization manifold learning method for speech feature extraction[J]. Noise and Vibration Control,2025,2(16):97-104.
张开业,赵化良,刘志红,等. 语音声特征提取的总变分正则化流形学习方法[J]. 噪声与振动控制,2025,2(16):97-104.

[4] QIN K X,WANG W X,WANG Y S. Research on robot speech recognition method based on improved MFCC feature extraction and DNN network[J]. Computer Measurement and Control,2025,2(31):246-253.
秦垲忻,王炜昕,王砚生. 基于改进MFCC特征提取和DNN网络的机器人语音识别方法研究[J]. 计算机测量与控制,2025,2(31):246-253.

[5] XUE P Y, BAI J, ZHANG N, et al. VMD based binary channels speech feature map extraction algorithm for dysarthria[J]. Journal of Northeastern University(Natural Science),2024,6(5):793-801.
薛珮芸,白静,张楠,等. 基于VMD的双通道构音障碍语音特征图谱提取算法[J]. 东北大学学报(自然科学版),2024,6(5):793-801.

[6] TANG B Y, JIAO L B, XU Y, et al. Image feature extraction network based on improved ResNet-50[J]. Computer Measurement and Control, 2023,6(25):162-167.
汤博宇,焦良葆,徐逸,等. 基于改进ResNet-50的图像特征提取网络[J]. 计算机测量与控制,2023,6(25):162-167.

[7] LIU S S, FANG Y F, TIAN T. Speech feature extraction and prediction model for translation robots based on endpoint detection[J]. Automation and Instrumention,2025(1):314-318.
刘姗姗,方一帆,田恬. 基于端点检测的翻译机器人语音特征提取与预测模型[J]. 自动化与仪器仪表,2025(1):314-318.

[8] JIN H S. A method for extracting feature parameters of intelligent robot speech signal based on VMD[J]. Electronic Design Engineering, 2023(22):130-133.
金豪圣. 基于VMD的智能机器人语音信号特征参数提取方法[J]. 电子设计工程,2023(22):130-133.

[9] LI X F, ZHANG Y H, LI Z Y, et al. Design of real-time interaction system based on multimodal deep learning[J]. Journal of Machine Design, 2024(S2):200-204.
李晓峰,张银慧,李子阳,等. 基于多模态深度学习的实时交互系统设计[J]. 机械设计,2024(S2):200-204.

[10] ZHAO J M,LI S,CHEN J H. Mechanisms and motivations of degree adverbialization of Uyghur adjectives[J]. Minority Languages of China,2025(1):43-52.
赵江民,李珊,陈嘉豪. 维吾尔语形容词的程度副词化现象之机制与动因[J]. 民族语文,2025(1):43-52.

[11] MA J. The manifestation, causes, and expandable semantic functions of tense translocation in Uyghur language[J]. Studies in Language and Linguistics,2024,44(1):113-119.
马娟. 维吾尔语时制易位的表现,成因及其扩展性语义功能[J]. 语言研究,2024,44(1):113-119.

[12] ZHOU W D, ZHOU H P, XIA P F. A nonlinear feature fusion method of speech emotion recognition based on attention mechanism[J]. Computer Applications and Software,2023,44(1):216-221,272.
周伟东,周后盘,夏鹏飞. 基于注意力机制的语音情感识别非线性特征融合方法的研究[J]. 计算机应用与软件,2023,44(1):216-221,272.

[13] LI F Y, JIA X F, ZHAO B T, et al. Efficient multi-attention feature fusion for image super-resolution reconstruction algorithms[J]. Journal of Chinese Computer Systems,2023,44(5):1023-1028.
李方玗,贾晓芬,赵佰亭,等. 高效多注意力特征融合的图像超分辨率重建算法[J]. 小型微型计算机系统,2023,44(5):1023-1028.

[14] YANG Y T,LI L L,LIU X,et al. Multiscale and multidirectional transformer-based image recognition[J]. Chinese Journal of Computers,2025,48(2):249-264.
杨育婷,李玲玲,刘旭,等. 基于多尺度-多方向Transformer的图像识别[J]. 计算机学报,2025,48(2):249-264.

[15] WANG L, JI X H, YANG M, et al. Mineral identification based on data augmentation and ensemble learning[J]. Earth Science Frontiers,2024,5(6):87-94.
王琳,季晓慧,杨眉,等. 基于数据增强和集成学习的矿物图像识别[J]. 地学前缘,2024,5(6):87-94.

[16] ZHANG Y H, GOU G L, LYU Y N, et al. Image recognition method based on sequential three-way decisions[J]. Application Research of Computers,2023,40(4):1268-1274.
张耀洪,苟光磊,吕艳娜,等. 一种基于序贯三支决策的图像识别方法[J]. 计算机应用研究,2023,40(4):1268-1274.

[17] LIANG H. Research on unidirectional isolation system based on image recognition[J]. Application of Electronic Technique, 2023(S1):161-165.
梁浩. 基于图像识别的单向隔离系统研究[J]. 电子技术应用,2023(S1):161-165.

[18] ZHANG H Y, ZHANG F K, YUAN G, et al. Research on person re-identification algorithm based on multi-pose image generation[J]. Computer Engineering and Applications,2023,59(2):143-152.
张海燕,张富凯,袁冠,等. 多姿态图像生成的行人重识别算法研究[J]. 计算机工程与应用,2023,59(2):143-152.

[19] WANG K P, ZUO X H, YANG Y, et al. Remote sensing image recognition algorithm based on pseudo global swin transformer[J]. Pattern Recognition and Artificial Intelligence,2023,36(9):818-831.
王科平,左鑫浩,杨艺,等. 基于伪全局Swin Transformer的遥感图像识别算法[J]. 模式识别与人工智能,2023,36(9):818-831.

[20] XU R Z, WANG S, LONG Y, et al. Research on a step-by-step adversarial defense method for image recognition[J]. Computer Engineering and Science,2023,45(6):1020-1029.
徐茹枝,王硕,龙燕,等. 针对图像识别的一种分步对抗防御方法研究[J]. 计算机工程与科学,2023,45(6):1020-1029.

[21] DONG Y H, PENG H L. Image classification method based on improved inception structure[J]. Software Guide,2023,22(2):41-46.
董跃华,彭辉林. 改进Inception结构的图像分类方法[J]. 软件导刊,2023,22(2):41-46.

[22] WANG M, DENG W H, SU S. Review on fairness in image recognition[J]. Journal of Image and Graphics,2024,29(7):1814-1833.
王玫,邓伟洪,苏森. 面向图像识别的公平性研究进展[J]. 中国图象图形学报,2024,29(7):1814-1833.

[23] HUANG Y J, GAO L, YANG T. Image recognition algorithms with multiscale features and meta-learning[J]. Computer Applications and Software,2024,41(8):203-209.
黄勇杰,高乐,杨田. 基于多尺度特征和元学习的图像识别算法研究[J]. 计算机应用与软件,2024,41(8):203-209.

[24] SU R X, GE D Y, YAO X F. Environmental sound recognition based on channel and frame-level feature attention model[J]. Science Technology and Engineering,2024,24(16):6792-6798.
苏瑞轩,葛动元,姚锡凡. 基于通道和帧级特征注意力模型的环境声音识别[J]. 科学技术与工程,2024,24(16):6792-6798.

[25] ZHENG W. Expression recognition method using DenseNe based on multi-scale attention mechanism [J]. Software Guide, 2023, 22 (2): 81-86.
郑伟. 多尺度注意力机制 DenseNet 网络的表情识别方法[J]. 软件导刊,2023,22(2):81-86.

(责任编辑:刘嘉文)