

信也科技杯

学术指导单位

中国计算机学会自然语言处理专业委员会(CCF-NLP)

合作院校

复旦大学自然语言处理实验室(FudanNLP)

赛事合作

第15届CCF自然语言处理与中文计算国际会议(NLPCC2026)

第十一届信也科技杯 全球AI算法大赛

||||| LISTEN BEYOND 循声知意

2026 FINVOLUTION
GLOBAL DATA SCIENCE COMPETITION
TURN-TAKING MODELING IN CONVERSATION CHALLENGE



CONTENT

01

赛题背景

02

难点分析

03

数据与训练

04

模型优化

05

工程处理

06

总结感谢



赛题背景：五类轮次行为预测

第十一届信也科技杯
全球AI算法大赛

TASK SETTING

5-30 s

输入：双声道音频 + 文本 + 30 秒历史标签

01

输入窗口

过去 5-30 秒
双人对话片段
音频 / 文本 / 历史

02

预测目标

未来 2 秒
C / NA / I / BC / T
多标签独立预测

03

评测指标

Macro-F1
稀有类等权
排序与工作点并重

任务本质

根据说话人关系、历史状态与临界时序
预测未来 2 秒内的下一轮交互行为

数据包含普通话与多种方言；I / BC / T 稀有且边界短促，是主要误差来源。



Baseline 已具备 Whisper + Qwen3 + 历史标签 CNN + 19 维统计；结合赛题场景我们进行以下分析。

BASELINE 代码	导致的问题	我们的修改
固定 375 chunks start = end - 375	测试片段为 5-30 s 训练只见过 30 s，长度分布错位	长度对齐 随机采样 62-375 chunks 左补零、真实内容右对齐
range(..., stride=5) 均匀滑窗	I / BC / T 边界窗口稀少 大量相邻背景样本淹没临界点	边界学习 进入 / 退出边界多偏移采样 每个事件最多追加 4 个窗口
wave.mean(dim=1) 五类共享 MLP	声道角色被平均 稀有类与 C / NA 争夺共享表征	关系建模 ch0 / ch1 / 差值 + BiGRU 逐类 base head + 交互 refine

后续改进围绕数据采样、时序关系建模与部署决策展开。

① 变长上下文训练

Baseline 固定 375 chunks; 我们随机采样 62-375 chunks

5s



2s

10s



2s

20s



2s

30s



2s

0.6241 → 0.6293

单模型公榜 +0.0052

② 边界对比采样

Baseline 仅均匀滑窗; 我们围绕进入 / 退出边界追加偏移窗口



偏移: ± 1 、 ± 2 、5、10、20、24 ...

352,269

追加窗口

≈ 21%

占 exp25 训练样本

将训练关注点从海量背景转向边界邻域

只在训练阶段扰动 Audio / Text / Context; 推理完全确定, Handcrafted 默认不加噪

WITHOUT
NOISE

高度重叠滑窗

冻结骨干输出稳定, 相邻窗口高度相似

融合头反复看到近重复样本

容易记忆稀有类局部模式

验证工作点对个别样本敏感

作用机制

在冻结骨干不被破坏的前提下, 扩张融合头看到的局部邻域, 降低滑窗记忆化。

FEATURE
NOISE

训练时 $\sigma = 0.05$

audio [B,512] + ϵa

text [B,1024] + ϵt

context [B,128] + ϵc

concat $\rightarrow$ Interact Head; eval 时 $\epsilon = 0$

短周期信号

epoch 0: 离线 F1 +0.0015, BC AUC
+0.0013

MagicHub 方言对话只提供稀有类候选；三模型一致后，以低权重进入真实训练分布

DATA
BOTTLENECK

稀有类正例不足

BC / I / T 在真实数据中占比低

均匀增加背景不能补排序信号

单模型伪标签易放大自身偏差

直接等权混入会污染真实监督

PSEUDO
PIPELINE

3-vote × weight 0.5 × repeat 8

多模型独立推理外部方言对话

仅保留三票一致的高置信正例

只注入 BC / I / T，局部标签 mask

低权重抑噪，重复采样保证进入梯度

作用机制 增加稀有类真实声学 and 方言变化，而不是继续复制官方滑窗背景。

逐类排序证据 BC AUC 0.7551→0.7604 (+0.0053)

每个 batch 内只重加权模型最容易误判为 BC / I 的负样本，不改变真实标签

STANDARD
LOSS

未显式强化难负样本

背景负样本数量远多于稀有正例

大量易负样本的信息量较低

高分假阳性在批次损失中被稀释

BC/I决策边界缺少针对性修正

作用机制 让优化器持续修正决策边界附近的假阳性，适配 macro-F1 下的稀有类。

ONLINE
HNM

$p \geq 0.55 \cdot \text{top } 20\% \cdot \text{weight } 2.0$

筛选真实标签为 0 的 BC / I 项

按当前预测分数过滤模型仍给出高分的负样本

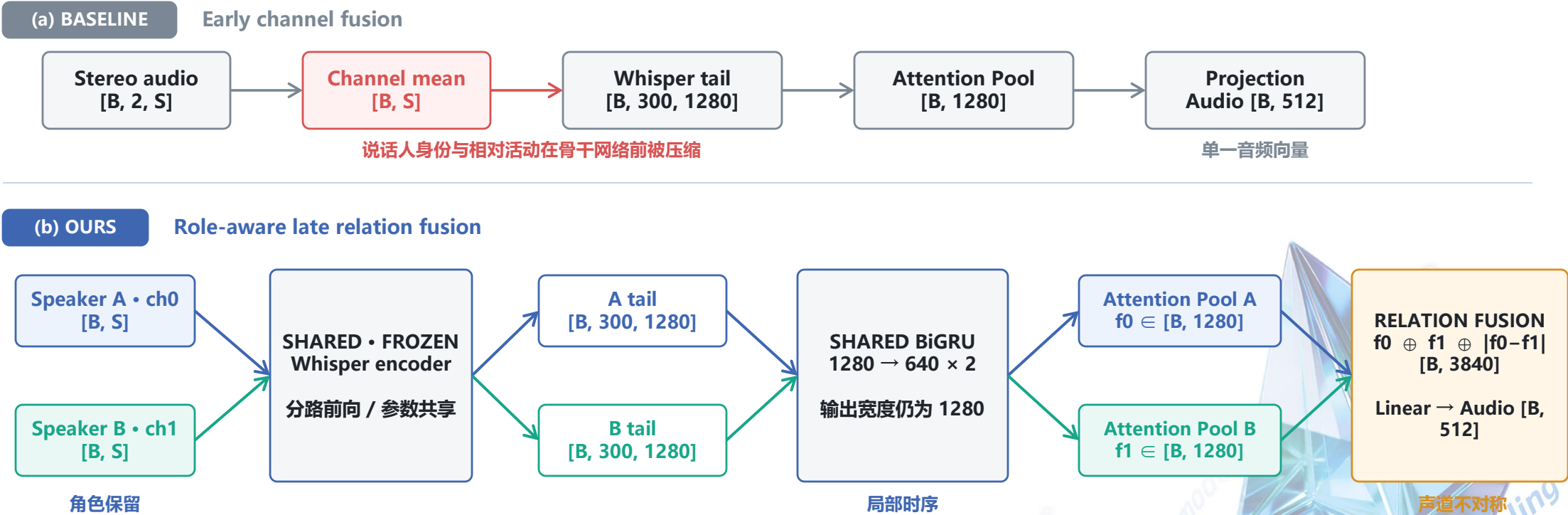
每类取 top 20% 难负样本

选中项损失 $\times 2$ ，其余保持原权重

消融实验 exp25→exp28: F1 +0.0013;
BC AUC +0.0029; I AUC +0.0010

模型优化一：角色感知的双声道时序交互

从声道早期平均，改为角色保留、共享时序编码与显式关系融合



机制 两路保留说话人角色；共享 BiGRU 提取停顿与话轮结束趋势； $|f_0 - f_1|$ 编码声道不对称性。

整体配方证据

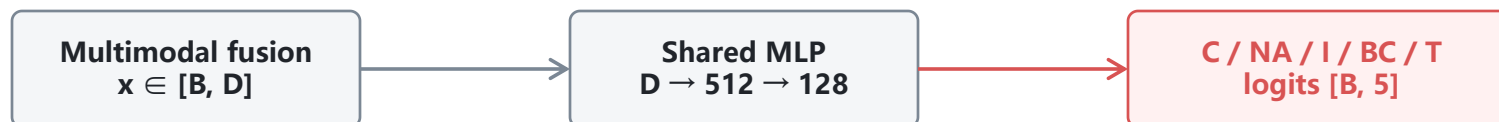
I / BC / T AUC: .826 / .755 / .871
公榜 .6608 → .6679 (+.0071) *

模型优化二：逐类决策头与标签交互修正

先学习类别专属边界，再用轻量 refine 建模多标签关系；保留多标签独立性

(a) SHARED HEAD

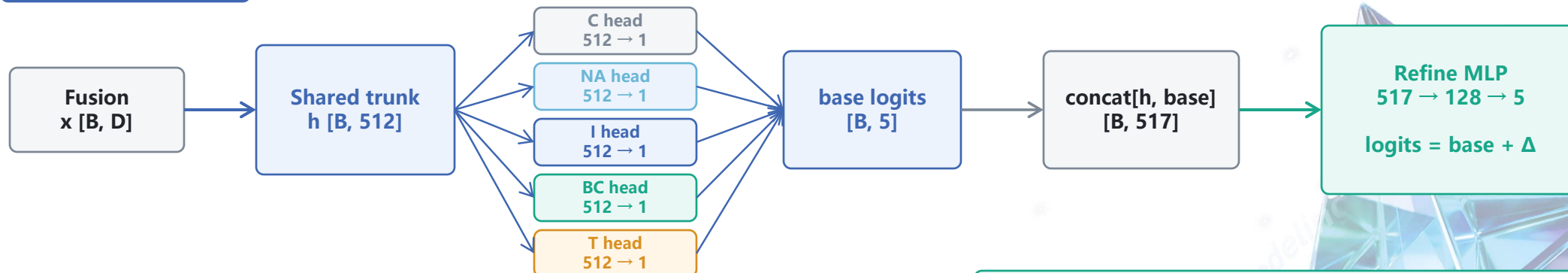
五类共享最后决策空间



C / NA 高频类别容易主导共享梯度，稀有类边界被折中

(b) INTERACT HEAD

Class-specific base logits + shared relational correction



机制 逐类头隔离决策边界；Interact refine 只学习标签间残差关系，不把 BC / I / T 强制互斥。

逐类 Head: 0.6137 → 0.6169 (+0.0032)
Interact: 0.6227 → 0.6246 (+0.0019)

离线结果：逐类 Head +0.0032, Interact +0.0019

Shape: h [B,512], base [B,5], concat [B,517], 最终输出仍为独立五标签 logits [B,5]。

模型优化三：冻结骨干与稀有类稳定训练

大骨干负责稳定表征，小规模时序层与分类头专注边界；强调稀有类但限制梯度失控

01 冻结骨干

Whisper + Qwen3

仅训练 BiGRU 与融合头
减少小数据过拟合与显存开销

02 Focal 加权 BCE

$$L = (1 - p_t)^{\gamma} \cdot BCEw$$

BCE 保持多标签独立监督
Focal 聚焦难例与高分误报

03 pos_weight cap

$$wc = \min(N - /N+, 6)$$

补偿类别不平衡
限制极稀有类梯度放大

04 Model EMA

$$\beta = 0.9998$$

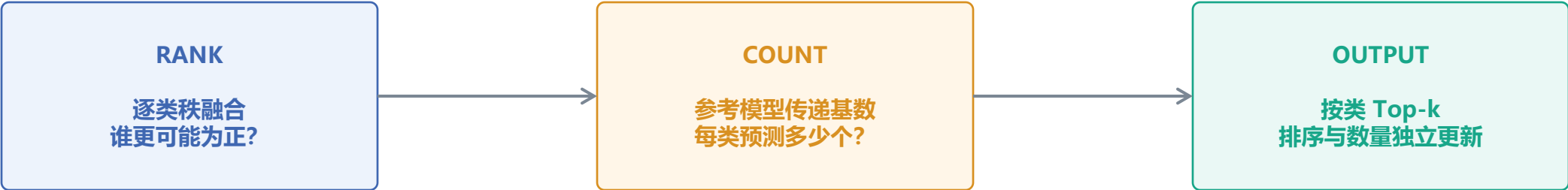
平滑参数更新
降低 checkpoint 与阈值波动

稳定表征 → 难例聚焦 → 类别梯度限幅 → 参数平滑

代表性实验	0.6233 解冻顶层 负向对照	≈ 0.6296 冻结 + EMA 稳定基线	0.6413 exp25 强配方 日志 tuned F1	0.8357 exp25 日志 macro-AUC
-------	------------------------	------------------------------	------------------------------------	---------------------------------

冻结骨干抑制小数据过拟合；Focal、cap=6 与 EMA 分别聚焦难例、限制稀有类梯度并平滑参数更新。

将类别内排序与预测数量分开建模，并按类别选择互补成员完成最终决策



$$\hat{k}_c = \sum_i 1[f_{ref,c}(x_i) \geq \tau_c] \qquad \hat{y}_c = \text{Top-k}(sc, \hat{k}_c)$$

逐类异构路由

- C / NA Whisper 主排序器
- I 主排序器 + 方言成员
- BC 主排序器 + FireRed / rare 专家
- T 主排序器 + 方言 / horizon 专家

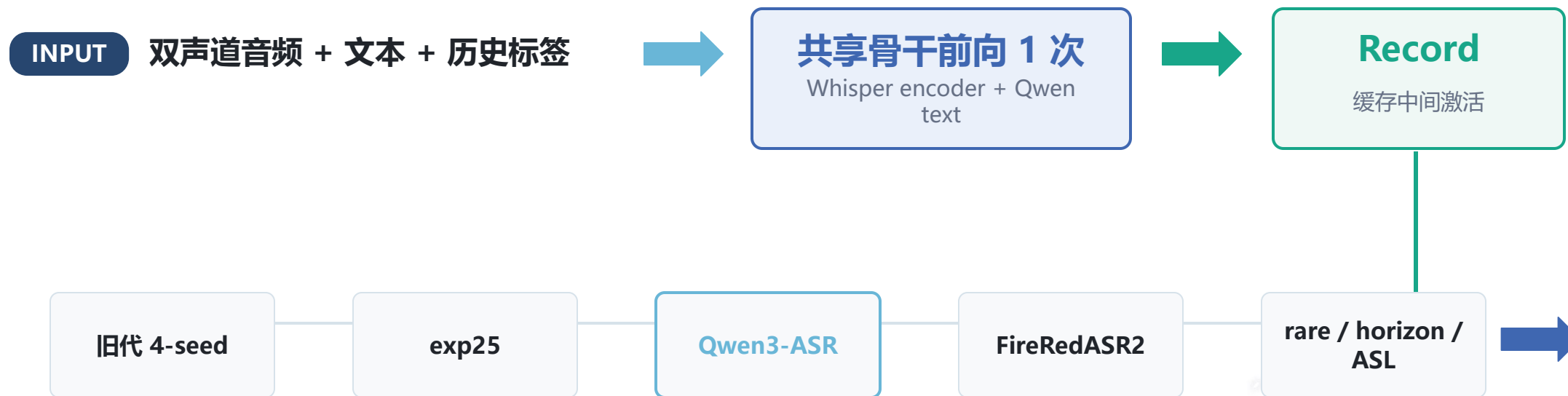
成员进入条件
逐类 AUC × 排序残差互补 × 跨折稳定

公开榜增益

Rate Matching	+0.0052
方言异构成员	+0.0017
逐类路由	+0.0010

工程落地：共享骨干与多头重放，60 分钟内完成推理

第十一届信也科技杯
全球AI算法大赛



12 heads → 9 ranking routes

2.24B

完整加载参数 < 8B

94.1M

12 个头部总参数

10.8M

单成员可训练参数

< 60 min

官方环境完成全量推理

run.sh 固化配置 | 无需外部参数 | 无在线依赖 | slim 权重约 43 MB / 成员

总结：从数据采样到模型与部署的完整方案

第十一届信也科技杯
全球AI算法大赛

训练阶段提升稀有类排序能力，部署阶段在资源约束内完成逐类决策

官方 baseline

0.593773



最终系统

0.674916

绝对提升
+0.081143

主要结果
强化边界样本学习
保留双声道角色信息
将排序增益转化为最终 F1

数据与训练

变长上下文
边界对比采样
Feature Noise / 伪标签
在线难负样本挖掘

模型结构

双声道共享 BiGRU
逐类 Head + Interact
冻结骨干
Focal-BCE + cap=6 + EMA

部署与推理

排序与数量解耦
逐类异构路由
共享骨干多头重放
官方环境 < 60 min

致谢主办方：让算法创新走向真实业务

比赛同时考察离线指标与 Docker 部署，使模型优化在统一环境中接受真实约束检验

算法优化

模型

×

Docker 实战

部署

贴近业务
AI 创新

算法探索

数据构造
时序建模
稀有类优化
多模型协同

部署实战

Docker 镜像
环境适配
效率优化
稳定交付

业务价值

统一环境验证模型效果
控制参数量与推理耗时
完成可复现的端到端交付
积累实际部署经验

信也科技杯

THANKS