



中国研究生创新实践系列大赛
“华为杯”第二十三届中国研究生
数学建模竞赛

学 校

长春师范大学

参赛队号

XXXXXX

队员姓名

1. 谢洪博

2. 崔曦月

3. 赵昱森

中国研究生创新实践系列大赛
“华为杯”第二十三届中国研究生
数学建模竞赛

题 目： 全国研究生数学建模竞赛论文标题

摘 要：

XXXX

关键词： XXX

目录

1	问题重述	5
1.1	引言	5
1.2	问题的提出	5
2	模型的假设	6
2.1	建模假设及适用范围	6
2.2	三层掩码与问题一的逐维观测规则	7
2.3	训练集限定的尺度处理与泄漏防护	8
3	符号说明	9
4	问题一：多视图时间锚定与可追溯时序对齐	10
4.1	输入输出与样本处理流程	10
4.2	子模型一：文本、语音和视觉的分层情感表征	11
4.2.1	上下文文本表征与词元归并	11
4.2.2	短时谱、倒谱与韵律的联合声学表征	11
4.2.3	面部动作、头部姿态和几何变化表征	12
4.3	子模型二：受转写约束的单调词级时间锚定	13
4.3.1	词边界后验与占据概率	13
4.4	子模型三：逐维质量加权的时间区间投影	14
4.4.1	计算素材范围与时间归属区间的分离	14
4.4.2	加权投影、局部方差与连续覆盖率	14
4.4.3	观测质量的主时间轴归集与两问接口	15
4.4.4	多层聚合的一致性与可重构条件	16
4.5	子模型四：固定秒级、词级与多尺度互补表示	16
4.5.1	固定秒级主时间视图	16
4.5.2	词级核验与多尺度上下文	17
4.5.3	保留 50 段相对进程辅助视图	17
4.6	子模型五：边界后验驱动的概率时间投影与低阶动态签名	17
4.6.1	概率时间投影：传播边界不确定性而不重排时间	17
4.6.2	低阶路径签名：补充均值与方差丢失的变化顺序	18

4.6.3	主结果保护与分支状态	19
4.7	子模型六：词查询驱动的双视图核验与受控内容增强	19
4.7.1	物理对应、内容匹配和情感贡献的分离	19
4.7.2	以原词为行、原生来源为列的物理高亮矩阵	19
4.7.3	不含额外位置通道的内容探针	20
4.7.4	多正例、时间保护带与共享内容学习	20
4.7.5	时间引导视图只作辅助，不作为独立内容证据	21
4.7.6	防止位置捷径、恒等循环与内容坍塌的核验	21
4.8	特征存储、全量汇总与读取规范	22
4.8.1	按样本存储与视图组织	22
4.8.2	变长序列、填充与来源关系	23
4.8.3	100 条样本与 300 条模态明细	23
4.9	核验指标、典型样本与模型对照	24
4.9.1	样本、时间和对齐准确性	24
4.9.2	典型样本核验卡	24
4.9.3	单因素对照而非复杂模块累加	25
4.10	参数锁定、运行流程与复现说明	25
4.10.1	全量可追溯特征生成流程	25
5	问题二：连续局部缺失下的概率补全与可靠度门控预测	28
5.1	问题接口、建模主线与适用范围	28
5.1.1	输入版本一致性与两问的职责	28
5.1.2	共享动态、轻量时序和双层门控的有机分工	28
5.2	连续区间缺失生成与可观测状态	29
5.2.1	同步与异步区间	29
5.2.2	三种缺失率及连续跨度	29
5.3	子模型一：质量感知的共享-私有动态概率补全	30
5.3.1	增广状态与观测噪声	30
5.3.2	后验均值、协方差及联合轨迹	31
5.3.3	补全器估计、冻结与不确定性核验	31
5.4	子模型二：连续缺口驱动的动态可靠度	32
5.4.1	观测质量与补全不确定性的统一但不混同	32
5.4.2	单调可靠度代理	32
5.5	子模型三：轻量缺失感知双向 GRU	32

5.5.1	输入状态与低可信更新控制	32
5.6	子模型四：GRU 后的来源门控跨模态注意力	33
5.6.1	有界内容得分与空来源	33
5.7	子模型五：可靠度与内容分离的融合门控	34
5.7.1	条件单调权重与先验专家	34
5.7.2	避免门控被归一化或旁路抵消	34
5.8	子模型六：可关闭低秩残差与时间汇聚	35
5.9	子模型七：中性点质量与有符号幅值分布	35
5.9.1	分类概率与条件幅值	35
5.9.2	混合预测、温度校准和一致解码	36
5.10	目标函数与分阶段训练	36
5.10.1	对混合分布的联合监督	36
5.10.2	完整可用视图教师与有限强度蒸馏	37
5.10.3	场景平衡与平滑最坏组风险	37
5.10.4	求解顺序与关键参数	38
5.11	验证设计、缺失规律与错误归因	38
5.11.1	分组验证、基础指标和不确定性	38
5.11.2	缺失类型、位置与持续跨度的受控实验	39
5.11.3	可推导性质与待检验规律	39
5.11.4	消融方案与错误归因	40
5.12	附件三全量推理、提交与复现记录	40
5.12.1	混合预测与低信息退化	40
5.12.2	正式预测与审计记录分离	41
5.12.3	训练与推理的数学求解流程	41
6	模型评价	42
	参考文献	43
	附录 A MATLAB 源程序	45
A.1	第 1 问程序	45
	附录 B Python 源程序	46
B.1	第 2 问程序	46

1 问题重述

1.1 引言

1.2 问题的提出

围绕 xxxx，本文依次提出如下问题：

问题一：

问题二：

问题三：

2 模型的假设

2.1 建模假设及适用范围

假设一：样本身份和标签可审计。原始主键由 `video_id` 与 `clip_id` 共同确定。若发现重复主键、文件缺失或标签冲突，须保留异常记录并按来源材料核对，不能按模型预测改写标签。极性与强度的逻辑关系为

$$c_i = \begin{cases} -1, & y_i < 0, \\ 0, & y_i = 0, \\ +1, & y_i > 0. \end{cases} \quad (2.1)$$

数值相等以来源文件的存储精度为准，不自行设置“近零即中性”的标签重编码带。文本、语音和视觉特征提取以及时间定位均不使用情感标签。

假设二：观测状态先于数值变换确定。填充位置、真实缺失、分量自然未定义和正常观测具有不同含义。只有来源说明明确约定“有效时段内全维零向量代表缺失”时，才能对相应预计算文件采用零向量判缺失规则。问题一自生成特征依据解码、检测、估计器状态和逐维定义域判断有效性，不因特征值为零而直接认定缺失。

假设三：不同模态具有部分共享但非完全相同的潜在情感变化。共享动态与私有动态的区分可供后续缺失补全和融合使用；问题一不要求估计共享情感潜变量，也不以强制提高模态间相似性作为对齐目标。讽刺、转折、表情掩饰等跨模态差异不被强制压平。

假设四：局部时间依赖有用，但不保证平稳。短区间缺失可在后续任务中借助邻近位置推断；长区间或多模态同步缺失的不确定性应扩大。线性高斯动态仅作为后续概率补全的可控先验，不能将推断值重新标记为问题一的真实观测。

假设五：解释是对指定模型与指定基线的决策归因。删除、遮蔽和特征置换不是自然世界中的随机干预，相关性归因不自动等于心理或生理因果解释。原始视频映射也不由矩阵形状保证，而由明确的时间区间与来源索引保证。

假设六：物理时间对应与边界不确定性传播严格区分。媒体呈现时间戳、原音轨偏移及可靠词边界构成物理时间依据。边界后验仅用于描述“词位于何处”的时间不确定性，不修改原始媒体时间戳，不跨越已确认的解码缺口或可靠锚点屏障。模态间情感表达不同不被直接解释为音画同步错误；概率时间视图是物理对齐的辅助表示，而不是用语义相似度重新排列时间。

假设七：词级时间对应不等于潜在槽位相似性。一个词允许覆盖多个声学帧和视频帧，面部行为也可跨多个词持续。物理对应由独立时间依据确定；可学习的固定数量槽位不是原词，也不自动等于物理时间格。三模态内容不要求逐词完全相似，内容探针与物理对齐分别核验。

假设八：质量记录与任务可靠度分层产生。问题一质量元数据只依据真实来源、检测与定位状态，不使用情感标签；问题二结合当前可见质量、连续缺口与补全后验估计可靠度，再学习内容相关的门控。质量未知不等于质量为零；低信息不等于中性。任何辅助内容不得绕过专项测试已施加的缺失。

2.2 三层掩码与问题一的逐维观测规则

先依据长度字段、填充规则和结构元数据得到 P_{it}^m 。仅对来源已明确采用全维零向量标记缺失、且一行各维度共同观测的预计算专项文件，使用

$$\begin{aligned} O_{it}^m &= P_{it}^m \mathbf{1}\{\mathbf{x}_{it,\text{raw}}^m \text{ 各分量有限且不全为零}\}, \\ M_{it}^m &= O_{it}^m B_{it}^m. \end{aligned} \quad (2.2)$$

若文件提供更精确的观测元数据，应优先使用元数据。训练完整视图取 $B \equiv 1$ ，人工缺失只作用于原本观测位置；不通过原始文本等辅助字段绕过专项集的既有缺失设置。对问题一自生成特征，逐维定义

$$O_{ird}^m = P_{ir}^m S_{ir}^m E_{ird}^m \mathbf{1}\{x_{ird,\text{raw}}^m \in \mathbb{R}\}. \quad (2.3)$$

这里 S_{ir}^m 表示原始来源可用， E_{ird}^m 表示相应估计成功且该分量有定义，集合 $\mathbb{R}$ 不含非有限数值。正常零值不改变 O_{ird}^m 。原生文本向量可有效而词时间尚不可定位，此时另设时间可定位标记，不删除原生文本，也不将其伪装成已有时间支撑。逐维位置区分 $P = 0$ 的填充、 $P = 1, O = 0$ 的未观测和 $P = 1, O = 1$ 的观测。未观测原因另行区分解码失败、检测失败、自然未定义和时间不可定位。概率边界派生视图有独立状态标记，不覆盖原始 O 。局部覆盖率由式 (4.24) 给出，不能由二值掩码代替。对未说明填充规则的零尾段，不用“最后一个非零位置”推断真实长度。若填充和末尾缺失两种机制可产生相同数组，则单凭数组不能唯一识别二者。此时设置 $U = 1$ ，保守保留为潜在有效但不可观测位置，待合法元数据核验后确定 P ，并单列歧义比例。对 `text_bert` 的使用仅限经确认不会泄露缺失内容的长度或填充元数据；`raw_text` 和词元编号不得进入问题二学生模型，也不得绕过专项集的文本置零直接重编码恢复。预计算特征与原始词元的关系不明时，不得擅自建立一一对应。

2.3 训练集限定的尺度处理与泄漏防护

记 Ω_{md}^{tr} 为训练集中模态 m 的第 d 维实际观测位置集合；对于整行共同观测的文件，由式 (2.2) 构造，对于逐维观测的文件，由式 (2.3) 构造。令

$$\begin{aligned}\mu_{md} &= \text{median}_{(i,t) \in \Omega_{md}^{\text{tr}}} x_{itd}^m, \\ r_{md} &= 1.4826 \text{ median}_{(i,t) \in \Omega_{md}^{\text{tr}}} |x_{itd}^m - \mu_{md}|, \\ s_{md} &= \begin{cases} r_{md}, & r_{md} > \varepsilon_s, \\ s_{md}^{\text{fb}}, & r_{md} \leq \varepsilon_s \text{ 且该维并非常量}, \\ 1, & \text{该维为常量}. \end{cases} \end{aligned} \quad (2.4)$$

其中 1.4826 为固定稳健尺度系数， s_{md}^{fb} 为预先声明的回退尺度：具有稳定物理范围的维度采用固定物理尺度，其余连续维度采用训练折标准差或预先固定的参考尺度。这样可避免“非恒定但 MAD 为零”的稀疏维度被 ε_s 人为放大后全部截断。训练中无有效观测的维度不估计任何尺度，单列状态。对有效维度应用

$$\tilde{x}_{itd}^m = \begin{cases} \text{clip}((x_{itd}^m - \mu_{md})/s_{md}, -b_m, b_m), & O_{itd}^m = 1, \\ 0, & O_{itd}^m = 0. \end{cases} \quad (2.5)$$

这里 $\text{clip}(z, -b, b) = \min\{b, \max\{-b, z\}\}$ 。若合法数据共同逐行观测，令 $O_{itd}^m = O_{it}^m$ 。回退尺度来源、常量维状态及使用次数均写入日志。标准化后的零可能是正常观测，不能再由零值重建掩码。真实填充在变换后置零，且不参与参数估计。原生未标准化特征必须保留。问题一若仅作描述性展示，可以另外保存明确标识为“全量描述性统计”的尺度，但该尺度不得用于声称无泄漏的预测评估。凡进行预测评估，尺度、降维和其他数据驱动参数均只在训练折拟合，验证与测试应用同一变换。按 `video_id` 核验同源片段相关性。已有官方划分时不擅自重划；官方划分若存在同源视频跨集，报告该限制。自行设置问题一的核验或后续实验划分时，以原始视频为分组单位；时间步数增加不等于独立样本数增加。

3 符号说明

表 3.1 符号约定

符号	含义
$i, j; m \in \mathcal{M} = \{T, A, V\}$	样本索引；文本、语音、视觉模态。
$D_i; r, t, k; \tau$	片段实际时长；原生行、目标区间、词索引；秒级时间。
$\mathbf{X}_i^m \in \mathbb{R}^{L_i^m \times d_m}$	模态特征矩阵；问题一主维度为 $d_T = 768, d_A = 74, d_V = 35$ 。
$P_{it}^m, O_{it}^m, U_{it}^m$	结构有效、整行观测和填充/缺失歧义标记；整行规则仅用于满足其条件的预计算文件。
O_{ird}^m, q_{ird}^m	问题一原生特征第 d 维的有效观测标记和质量权重。
$B_{it}^m; M_{it}^m$	后续训练的人工保留掩码；参与推断的观测掩码。
$\mathcal{R}_{ird}^m; J_{ir}^m$	特征计算使用的素材范围；原生行在时间轴上的归属区间。
$I_{it}; \Delta; h_s$	目标区间；固定秒级步长；多尺度上下文窗口长度。
$\bar{x}_{itd}^m, v_{itd}^m, c_{itd}^m$	物理对齐均值、局部方差和逐维真实时间覆盖率。
$S_{ik}, E_{ik}; \pi_{ik}(\tau)$	第 k 个词的随机起止边界；时刻 τ 属于该词的后验占据概率。
$\alpha_{itk}, \beta_{ikrd}^m$	目标区间-词的期望重叠质量；词-原生来源的后验重叠质量。
$\text{Sig}^{(1)}, \text{LogSig}^{(2)}$	低阶路径签名的一阶增量与二阶有向交互，用于辅助描述窗口内部变化顺序。
$\mathcal{D}_{tr}, \mathcal{D}_{va}, \mathcal{D}_{te}$	合法训练、验证和测试划分；已有官方划分时不擅自重划。
$\mathbf{w}_{it}, A, C_m, Q, R_m$	后续动态模型的增广状态、转移矩阵、观测矩阵及噪声协方差。
$\boldsymbol{\mu}_{it}, \Sigma_{i,tt}$	后续条件平滑后验均值和状态协方差块。
$\rho_{it}^m, \alpha_{it}^m$	可靠度和融合模态权重，不直接等于解释贡献。
p_{ic}, a_{ic}, b_{ic}	极性概率；非中性幅值的 Beta 分布形状参数。
$\hat{c}_i, \hat{y}_i$	最终报告的极性与情感强度。
$\mathcal{T}_i, \mathcal{L}_i$	原始素材映射和处理日志。
$\bar{q}_{itd}^m, q_{it}^{m, \text{obs}}, J_{it}^{m, Q}$	对齐后逐维质量、行级质量摘要及质量字段可用标记。
$H^{m, \text{time}}, H^{m, \text{cont}}, H^{m, \text{guide}}$	词查询的物理对应、无额外位置内容图和时间引导图。
$F_m, S_{ikr}^{m, \text{cont}}$	可选内容投影器及跨模态内容得分；不修改时间轴。
$K_{MC}; b$	问题二联合后验轨迹数及轨迹索引；与词索引 k 区分。
$u_{it}^m, g_{it}^m, b_{it}^m$	标准化补全方差、最近观测距离与连续缺口跨度比例。
$A_{tr}^{m \leftarrow n}, \alpha_{it}^{\emptyset}$	跨模态来源读取权重与先验专家融合权重。
G_{ic}, T_{cal}	条件幅值混合分布函数及最终类别概率校准温度。
ε	数值稳定常数；按变量量纲分别设定并记录。

4 问题一：多视图时间锚定与可追溯时序对齐

4.1 输入输出与样本处理流程

输入为附件一的 100 条原始视频 V_i 、英文转写 W_i 和样本主键。处理算子定义为

$$\mathcal{F}_1 : (V_i, W_i, \text{id}_i) \mapsto \{ \mathbf{X}_{i,\text{native}}^m, \mathbf{X}_{i,\text{word}}^m, \mathbf{X}_{i,\text{sec}}^m, \mathcal{X}_{i,\text{multi}}^m, \mathbf{X}_{i,50}^m, \mathbf{X}_{i,\text{post}}^m, \mathbf{X}_{i,\text{dyn}}^m, \\ P_i^m, O_i^m, Q_i^m, \mathcal{H}_i, \mathcal{T}_i, \mathcal{L}_i \}_{m \in \mathcal{M}}. \quad (4.6)$$

其中，固定秒级结果为主时间视图，词级结果用于语义核验，多尺度及 50 段相对进程结果作为辅助；词边界后验视图只在对齐器能够提供可靠路径后验时生成，低阶动态签名作为独立辅助分支保存。空的可选视图以未启用状态表示，不视为样本遗漏。新增 Q 为独立质量元数据， $\mathcal{H}$ 为词查询物理矩阵及可选内容核验图；二者不覆盖主特征。首先建立“标签表一行—原视频文件—输出记录”的双向清单，主键由 `video_id` 和 `clip_id` 共同组成，保留原始字符串及文件哈希。检查标签主键、素材主键与输出主键集合的一致性，分别登记重复、缺少、额外及冲突记录。标签只用于一致性核验，不参与特征定位、异常修正、边界后验或动态分支参数选择。采用 FFmpeg/ffprobe 类媒体工具获取音视频流和时间信息。设共享片段原点为 τ_{i0} ，呈现时间戳和时间基分别为 PTS_{ir}^m 与 TB_i^m ，则

$$\tau_{ir}^m = \text{PTS}_{ir}^m \text{TB}_i^m - \tau_{i0}. \quad (4.7)$$

若音频在共享时间轴上连续，则可写为 $\tau_i^A(n) = \delta_i^A + n/f_s$ 。若存在局部解码缺口，则按第 j 个连续有效音频段 $\mathcal{N}_{ij}$ 分段映射：

$$\tau_i^A(n) = \delta_{ij}^A + \frac{n - n_{ij}}{f_s}, \quad n \in \mathcal{N}_{ij}, \quad (4.8)$$

其中 n_{ij} 为该连续段的起始采样索引， δ_{ij}^A 为该段在共享时间轴上的真实起点。若实现采用保留缺口占位的全长数组，也可继续使用全局索引公式，但缺口位置必须由来源可用标记排除，不能参与声学估计。音频统一解码为单声道 $f_s = 16\,000$ Hz 浮点波形；起始偏移、各连续段边界、重采样映射与原始采样率同时存档。视频按真实时间戳解码，保留可变帧率信息，不以固定帧率公式替换未经核验的时间戳。不进行静音拼接、变速或改变强弱关系的强降噪。无音轨、局部解码缺口和无人脸分别标记，不删除样本，也不把缺口两侧拼接成连续内容。英文转写保留原文副本；工具输入的规范化、字符偏移、子词、原词与标点之间的映射独立保存。

4.2 子模型一：文本、语音和视觉的分层情感表征

4.2.1 上下文文本表征与词元归并

采用冻结的通用预训练 BERT 编码器 [1]，模型标识为 `bert-base-uncased`，不使用本题情感标签微调。设原词 w_{ik} 对应子词集合 $\mathcal{B}_{ik}$ ，第 ℓ 层子词表示为 $\mathbf{h}_{ib}^{(\ell)}$ ，定义

$$\mathbf{x}_{ik}^T = \frac{1}{4|\mathcal{B}_{ik}|} \sum_{\ell=L-3}^L \sum_{b \in \mathcal{B}_{ik}} \mathbf{h}_{ib}^{(\ell)} \in \mathbb{R}^{768}. \quad (4.9)$$

对于重复覆盖的子词，先按其距离窗口边界的权重归并多窗口表示，再按式 (4.9) 归并到原词。初始配置采用最多 510 个非特殊子词的窗口和 384 个子词的步长，特殊词元占用模型其余位置；实际分词器版本、最大长度和窗口起止索引均存档。保留否定词、程度词、转折词和重复表达，不机械删除停用词。标点保留在编码输入中，但不单独虚构发音时间；特殊词元不构成独立物理时间步。词时间只表示向量的定位锚，不表示上下文向量仅使用该词。句级汇总可另存，不能替代词级可定位序列。

4.2.2 短时谱、倒谱与韵律的联合声学表征

保留原方案的 74 维可解释主特征。短时频谱配置为窗长 $W = 400$ 点、步长 $H = 160$ 点和 Hann 窗，分别对应 25 ms 与 10 ms。将 FFT 长度单独设置为 $N_{\text{FFT}} = 512$ ，定义

$$\begin{aligned} S_{ir}[u] &= \sum_{n=0}^{W-1} s_i[rH + n]h[n] \exp\left(-\frac{2\pi i u n}{N_{\text{FFT}}}\right), \\ E_{irb} &= \log\left(\varepsilon_E + \sum_{u=0}^{N_{\text{FFT}}/2} H_b[u]|S_{ir}[u]|^2\right). \end{aligned} \quad (4.10)$$

这里 H_b 为第 b 个 Mel 滤波器。主配置不对分析窗作居中平移；末尾不足窗长时仅为计算作右侧零延拓，并保存真实采样范围和填充比例，不把延拓部分记作有效观测时长。Mel 频率映射、滤波器归一化、频率上下限和能量底数均在配置中固定。采用 40 维对数 Mel、13 维 MFCC、13 维 MFCC 一阶差分 and 8 维韵律/谱统计量：

$$\mathbf{x}_{ir}^A = [\text{LogMel}_{40}; \text{MFCC}_{13}; \Delta\text{MFCC}_{13}; \text{Prosody}_8] \in \mathbb{R}^{74}. \quad (4.11)$$

MFCC 采用对数 Mel 的正交归一化第二型离散余弦变换，保留 0 至 12 号系数。差分初始设置为左右各两步的局部线性斜率，内部点使用对称邻域，边界使用实际存在的单侧邻域；不足以估计时逐维标记无效。差分计算的邻域也属于其素材计算范围。8 项统计量依次为对数能量、对数基频、有声概率、谱质心、谱带宽、谱通量、过零率和谐噪比。基频与有声概率采用明确输出这两类量的概率估计器；若工具实际提供的是周期性强度或未校准分数，须按实际含义另命名，不冒称概率。谐噪比采用单独锁定的谐性估计器，并记录单位、基频范围和计算窗口。不同声学分量允许具有不同计算窗口。谱特征的 25 ms 窗不自

动适用于基频、有声概率或谐噪比。各分量先保留其原生时间锚与计算范围，再依据真实归属区间投影到 10 ms 声学网格，形成式 (4.11)。无声区间中的基频等分量可自然未定义，但不意味着整个音频模态缺失；局部损坏只影响相应局部。全程保留逐维掩码。频谱与倒谱存在冗余，但提供不同形式的谱描述；是否裁减可在合法训练数据上消融。**这里的自定义 74 维向量不是对附件二既有 74 维预计算特征的逐维解释。**冻结的深层语音表示可作为独立辅助分支另存，但不改变主特征维度，也不默认学习公共空间后重新安排时间对应。

4.2.3 面部动作、头部姿态和几何变化表征

使用 OpenFace 类通用面部分析工具提取动作单元、头部姿态、视线和关键点，并保留检测成功标记、置信度和轨迹记录 [3, 4]。定义

$$\mathbf{x}_{ir}^V = [\text{AU}_{17}; \text{Pose}_6; \text{Gaze}_6; \text{Geometry}_6] \in \mathbb{R}^{35}. \quad (4.12)$$

17 个 AU 强度按 AU01、02、04、05、06、07、09、10、12、14、15、17、20、23、25、26、45 的顺序保存。具体构建不支持的维度标为不可用，不临时换成其他物理量。**Pose** 为三维旋转与三维平移，保留工具原始单位再按式 (2.4) 处理；不以二维像素眼间距除以毫米平移。**Gaze** 为双眼各三维方向，视线估计失败时仅将相关维度标为无效。**Geometry** 保留六个无量纲量：左右眼开合度、嘴部开合度、嘴角间距、左右眉眼距离。为固定几何定义，设二维关键点按零基编号 $0, \dots, 67$ 记为 $\mathbf{p}_j$ ，距离为 $d_{ab} = \|\mathbf{p}_a - \mathbf{p}_b\|_2$ ，眼中心为

$$\begin{aligned} \mathbf{e}_1 &= \frac{1}{6} \sum_{j=36}^{41} \mathbf{p}_j, & \mathbf{e}_2 &= \frac{1}{6} \sum_{j=42}^{47} \mathbf{p}_j, & \ell &= \|\mathbf{e}_1 - \mathbf{e}_2\|_2, \\ g_1 &= \frac{d_{37,41} + d_{38,40}}{2\ell}, & g_2 &= \frac{d_{43,47} + d_{44,46}}{2\ell}, & g_3 &= \frac{d_{61,67} + d_{62,66} + d_{63,65}}{3\ell}, \\ g_4 &= \frac{d_{48,54}}{\ell}, & g_5 &= \frac{\|\frac{1}{5} \sum_{j=17}^{21} \mathbf{p}_j - \mathbf{e}_1\|_2}{\ell}, & g_6 &= \frac{\|\frac{1}{5} \sum_{j=22}^{26} \mathbf{p}_j - \mathbf{e}_2\|_2}{\ell}. \end{aligned} \quad (4.13)$$

只有关键点完整且 $\ell > \varepsilon_\ell$ 时才计算相应几何量。编号组的左右含义按工具映射锁定，不随画面镜像自动交换。单脸置信度初始门槛可取 $\theta_V = 0.8$ ，作为待核验设计值而非实测最优值。多人画面综合轨迹连续性、位置和口部运动与音频活动的一致性选择主说话者轨迹；选择不唯一时降低权重并保留候选，不逐帧任意切换。归一化差分可作为辅助动态统计，但不跨轨迹切换或缺失段计算。视觉量在后续时间聚合时按几何类型处理：**AU**、平移量和无量纲几何量可采用质量加权标量均值；头部旋转不直接对欧拉角做全局算术平均，而优先转换为旋转矩阵后在 $SO(3)$ 上求局部加权旋转均值，或在确认角度变化远离 $\pm\pi$ 分支点时使用局部近似；视线方向采用加权合向量并重新单位化，合向量范数过小时标记方向不稳定。这样避免 179° 与 -179° 被错误平均为 0° 。

表 4.2 问题一主特征与原始素材追溯字段

模态	维度	原生时间组织	质量依据	追溯字段
文本	768	原词／子词，另存词时间	词元映射、对齐分数及边界不确定性	字符、子词、词区间及编码窗口
语音	74	10 ms 网格；各分量计算范围独立	解码与逐维估计状态	原始采样范围、音轨偏移、分析窗口
视觉	35	原始帧时间戳与归属区间	检测、轨迹和逐维有效性	原始帧号、PTS、轨迹及关键点映射

4.3 子模型二：受转写约束的单调词级时间锚定

已有英文转写时采用强制对齐，不重新识别后覆盖原文。可使用 Montreal Forced Aligner 所代表的声学-音素对齐流程 [2]。将原转写转换为按顺序展开的状态图，允许自循环、向后续状态移动和合法非词状态，不允许逆序跳回。给定对齐器声学观测 $\mathbf{a}_r$ ，求解

$$z_{1:R} = \arg \max_{z_{1:R} \in \mathcal{G}(W_i)} \left\{ \log p(z_1) + \sum_{r=2}^R \log p(z_r | z_{r-1}) + \sum_{r=1}^R \log p(\mathbf{a}_r | z_r) \right\}. \quad (4.14)$$

这里 $\mathcal{G}(W_i)$ 为原转写允许的状态路径集合，对齐器内部声学特征不必等于式 (4.11)。Viterbi 递推为

$$V_r(k) = \log p(\mathbf{a}_r | k) + \max_{j \in \text{Pred}(k)} \{V_{r-1}(j) + \log p(k | j)\}. \quad (4.15)$$

回溯并合并同词音素，得到 $I_{ik}^{\text{word}} = [s_{ik}, e_{ik})$ 。可靠词区间须满足

$$0 \leq s_{i1} < e_{i1} \leq s_{i2} < e_{i2} \leq \dots \leq s_{iK_i} < e_{iK_i} \leq D_i. \quad (4.16)$$

零时长、逆序或越界结果标记为异常，不直接作为可靠区间使用。词间及首尾未覆盖部分单独保存，区分已确认静音、非词声音和未可靠对齐区域。没有对齐到单词不等于静音，也不能靠拉长相邻词吸收全部间隙。可靠词区间与补充区间共同形成覆盖 $[0, D_i)$ 的时间分区；区间类型不等于情感标签。工具若只提供路径分数，称为对齐质量分数，不称为校准概率。若可在固定状态图上获得前向-后向权重，可另外计算边界均值与分位区间，并与原始最优路径边界分别保存。低质量词保留原文及向量，时间映射标为不确定。完全失败时保留原生文本，时间视图标记不可定位，绝不按词数均分视频后宣称精确对齐。

4.3.1 词边界后验与占据概率

当对齐器能够在固定转写状态图上提供前向-后向后验时，不把单一 Viterbi 边界视为唯一确定事实，而将第 k 个词的起止边界记为随机变量 (S_{ik}, E_{ik}) ，其联合分布由合法单调路径诱导。定义词占据概率

$$\pi_{ik}(\tau) = \Pr(S_{ik} \leq \tau < E_{ik} | A_i, W_i), \quad k = 1, \dots, K_i. \quad (4.17)$$

另设 $\pi_{i0}(\tau)$ 表示状态图中合法静音或非词状态的后验质量；在已建模且媒体来源有效的时间位置满足

$$\sum_{k=0}^{K_i} \pi_{ik}(\tau) = 1. \quad (4.18)$$

解码缺口不属于“被模型观察到的静音”，因此不纳入式 (4.18) 的有效时间域。对固定目标区间 I_{it} ，定义第 k 个词在其中的期望占据时长

$$\alpha_{itk} = \mathbb{E}[|I_{it} \cap [S_{ik}, E_{ik})| \mid A_i, W_i] = \int_{I_{it}} \pi_{ik}(\tau) d\tau. \quad (4.19)$$

由此构造文本后验时间视图

$$\mathbf{x}_{it,\text{post}}^T = \frac{\sum_{k=1}^{K_i} \alpha_{itk} \mathbf{x}_{ik}^T}{\sum_{k=1}^{K_i} \alpha_{itk}}, \quad (4.20)$$

仅在分母大于零时定义。若边界后验退化为确定区间，则 $\pi_{ik}(\tau) = \mathbf{1}\{\tau \in [S_{ik}, E_{ik})\}$ ，式 (4.19) 自动退化为普通区间交长，因此概率视图与物理投影具有一致的极限情形。需要强调，式 (4.20) 是“以期望占据质量归一化”的表示，通常不等于逐条随机边界路径聚合结果的严格后验均值。若需要后者，应从联合合法路径后验采样，在每条路径上先执行完整聚合，再统计样本均值、方差与不可估计比例。边界后验的窄宽也不自动等于现实世界中的校准准确率，仍须通过人工审计边界验证其覆盖率。

4.4 子模型三：逐维质量加权的时间区间投影

4.4.1 计算素材范围与时间归属区间的分离

对每个原生特征同时保存 $\mathcal{R}_{ird}^m$ 与 J_{ir}^m 。前者是第 d 维计算实际使用的采样、帧或上下文范围，允许重叠且可以随维度变化；后者是该原生行在物理时间轴上的归属区间，用于时间质量分配。文本使用可靠词区间作为 J ；音频与视频在每个连续有效段内，按相邻时间锚的中点构造不重叠归属区间，再与该段真实边界相交。已确认的缺口保留为空，不通过中点扩展跨越。区间外推到首尾边界的规则固定并记录。由此，重叠分析窗不会被重复计成多份有效时长。语音分量异步输出时，其原生分支到联合声学网格的投影也使用同一规则，并保留两级来源关系。

4.4.2 加权投影、局部方差与连续覆盖率

给定目标区间 I_{it} ，令

$$\begin{aligned} w_{itrd}^m &= |I_{it} \cap J_{ir}^m| q_{ird}^m O_{ird}^m, \\ W_{itd}^m &= \sum_r w_{itrd}^m, \quad \bar{x}_{itd}^m = \frac{\sum_r w_{itrd}^m x_{ird}^m}{W_{itd}^m} \quad \text{当 } W_{itd}^m > 0. \end{aligned} \quad (4.21)$$

质量权重满足 $0 \leq q_{ird}^m \leq 1$ ，来源和计算公式须记录。无法可靠评估额外质量时，正常有效维可取单位权重，而不是构造未经支持的概率。 $W_{itd}^m = 0$ 时不计算分式，只输出占位值及

无观测标记。式 (4.21) 是一维加权最小二乘问题的闭式解：

$$\bar{x}_{itd}^m = \arg \min_{u \in \mathbb{R}} \sum_r w_{itrd}^m (u - x_{itrd}^m)^2, \quad W_{itd}^m > 0. \quad (4.22)$$

因此，三模态使用一致的时间投影算子，而不是分别采用无时间依据的帧号平均。原生不可定位的文本不参与该投影。对标量特征，式 (4.22) 直接给出聚合值；对视觉旋转和单位方向量，则沿用相同的时间质量权重，但在各自几何空间内完成聚合，因此“一致的投影算子”指一致的时间质量分配，而不是强迫所有物理量使用同一种欧氏算术平均。另外保存加权局部方差及观测数量：

$$v_{itd}^m = \frac{\sum_r w_{itrd}^m (x_{itrd}^m - \bar{x}_{itd}^m)^2}{W_{itd}^m}, \quad n_{itd}^m = \#\{r : w_{itrd}^m > 0\}. \quad (4.23)$$

上述方差是局部分散程度，不是均值估计的不确定性；仅一个来源时数值可为零，但须标记动态证据不足。均值、方差、掩码和计数分别保存，不改变 768/74/35 的主维度。定义参与估计的真实时间集合及连续覆盖率为

$$\mathcal{U}_{id}^m = \bigcup_{r: O_{itrd}^m=1, q_{itrd}^m>0} J_{ir}^m, \quad c_{itd}^m = \frac{|I_{it} \cap \mathcal{U}_{id}^m|}{|I_{it}|}. \quad (4.24)$$

$c_{itd}^m \in [0, 1]$ 衡量真实时间支撑，不等于检测置信度。一个 100 ms 区间只有 10 ms 有效来源时，即使均值可计算，其覆盖率也仅为 0.1。目标逐维观测标记取 $O_{itd, \text{sec}}^m = \mathbf{1}\{W_{itd}^m > 0\}$ ，与覆盖率分开保存。区间并集保证重叠来源不重复计时。另存解码可用率与分量定义率，避免将自然未定义误记为模态完全缺失。

4.4.3 观测质量的主时间轴归集与两问接口

原生质量 q_{itd}^m 描述提取与定位条件，不能根据情感标签、特征幅值或跨模态情感一致程度反推。视觉可记录检测置信度、轨迹稳定状态；声学记录解码、窗口完整性及分量估计状态；文本分别记录词元映射质量和边界定位质量。工具分数未经校准时仍称质量分数，不称“真实正确概率”。令 $a_{itrd}^m = |I_{it} \cap J_{ir}^m| O_{itrd}^m$ ，定义

$$\bar{q}_{itd}^m = \begin{cases} \frac{\sum_r a_{itrd}^m q_{itrd}^m}{\sum_r a_{itrd}^m}, & \sum_r a_{itrd}^m > 0, \\ 0, & \text{否则}. \end{cases} \quad (4.25)$$

这里不再次用含 q 的 w 来平均 q ，以免无解释地对高质量来源重复加权。结合式 (4.24) 的真实时间覆盖，定义行级辅助分数与有效维比例

$$q_{it}^{m, \text{obs}} = \frac{\sum_d O_{itd, \text{sec}}^m c_{itd}^m \bar{q}_{itd}^m}{\max\{1, \sum_d O_{itd, \text{sec}}^m\}}, \quad o_{it}^m = \frac{1}{d_m} \sum_d O_{itd, \text{sec}}^m. \quad (4.26)$$

q^{obs} 是可见维度的质量与时间支撑摘要, o 是维度可用程度, 二者分别保存。自然未定义的基频等不使所有声学维度失效。局部方差 v 反映真实变化, 不直接作为质量惩罚; 强烈情绪引起的高变化幅度不能因此被删除。另记质量字段可用标记 $J_{it}^{m,Q}$, 组成

$$\mathcal{Q}_i^m = \{\bar{q}_{itd}^m, q_{it}^{m,\text{obs}}, o_{it}^m, J_{it}^{m,Q}, \text{质量来源、规则及版本}\}. \quad (4.27)$$

这些记录独立存入 `quality/sec`, 不改变主特征维数, 也不以 $q\mathbf{x}$ 覆盖原特征。**问题一记录观测质量**, **问题二估计动态可靠度并学习融合门控**。前者不是后者的已训练答案。专项特征未提供合法质量字段时, 设置 $J^Q = 0$, 并以 $q^* = 1$ 表示“不施加额外质量折扣”, 不表示已核实为百分之百可靠。若部署端没有此字段, 默认训练端同样不使用此字段; 需要研究元数据缺失时, 另做明确的元数据遮蔽实验。被人工遮蔽位置的 q^{obs} 、局部方差、动态签名和原文不得成为学生的隐藏内容旁路。

4.4.4 多层聚合的一致性与可重构条件

若目标区间 I 被互不重叠的 $I_1, \dots, I_J$ 完整划分, 并且各子区间统计量均由同一套底层来源权重产生, 则均值满足

$$\bar{x}_d^m(I) = \frac{\sum_{j=1}^J W_d^m(I_j) \bar{x}_d^m(I_j)}{\sum_{j=1}^J W_d^m(I_j)}. \quad (4.28)$$

零权重子区间不参与求和。为使均值与方差能够无损合并, 底层同时保存加权充分统计量

$$S_0(I) = \sum_r w_r(I), \quad S_1(I) = \sum_r w_r(I) x_r, \quad S_2(I) = \sum_r w_r(I) x_r^2, \quad (4.29)$$

于是 $\bar{x} = S_1/S_0$, $v = S_2/S_0 - \bar{x}^2$ 。等价地, 若只保存子区间均值与方差, 则大区间方差按

$$v(I) = \frac{\sum_j W_j \left[v(I_j) + (\bar{x}(I_j) - \bar{x}(I))^2 \right]}{\sum_j W_j} \quad (4.30)$$

合并, 不能简单平均各子区间方差。上述可合并性只对“完整、互不重叠且与目标边界嵌套”的子区间成立。若新目标边界切穿某个已聚合子区间, 则必须回查其原生来源记录重新计算交长, 不能把中间格均值当成区间内部均匀分布后再次投影。观测数量 n 也按原生来源索引去重, 不能在重复视图之间直接相加。直接原生聚合与充分统计量合并所得结果应在预设数值容差内一致, 并作为实现审计条件。

4.5 子模型四: 固定秒级、词级与多尺度互补表示

4.5.1 固定秒级主时间视图

采用固定物理步长 $\Delta = 0.1 \text{ s}$, 定义

$$I_{it}^{\text{sec}} = [(t-1)\Delta, \min\{t\Delta, D_i\}), \quad t = 1, \dots, L_i, \quad L_i = \left\lceil \frac{D_i}{\Delta} \right\rceil. \quad (4.31)$$

由题面时长 2.648–34.567 s 可得 $27 \leq L_i \leq 346$ 。最后一个区间保留真实长度，不补出额外物理时长。主特征形状为

$$\mathbf{X}_{i,\text{sec}}^T \in \mathbb{R}^{L_i \times 768}, \quad \mathbf{X}_{i,\text{sec}}^A \in \mathbb{R}^{L_i \times 74}, \quad \mathbf{X}_{i,\text{sec}}^V \in \mathbb{R}^{L_i \times 35}. \quad (4.32)$$

同一样本三模态第 t 行共享同一个物理区间；跨样本的相同步号共享相同名义步长。某模态不具有有效来源时保留该时间位置及掩码，而不是删除一行破坏共同时间轴。

4.5.2 词级核验与多尺度上下文

词级视图直接以 I_{ik}^{word} 为目标，调用式 (4.21)，回答“某原词对应哪些声音采样、视频帧和特征行”。非词区间单独保存类型及来源；未可靠定位的原词仍保留在原生文本记录中。以固定秒级区间的中心 κ_{it} 为锚，构造上下文窗口

$$R_{it}^{(s)} = [\kappa_{it} - h_s/2, \kappa_{it} + h_s/2) \cap [0, D_i), \quad (4.33)$$

$$(h_1, h_2, h_3) = (0.1, 0.3, 0.7) \text{ s}, \quad \mathcal{X}_{i,\text{multi}}^m \in \mathbb{R}^{L_i \times 3 \times d_m}.$$

各尺度直接从原生特征或保存权重的底层统计聚合，并分别保存实际起止时间。边界窗口截断后可不再对称，且 0.1 s 上下文窗口不必与末尾不足步长的主区间相同。较大窗口表示邻域上下文，不宣称是当前 100 ms 内直接观察到的全部内容。各尺度分开存储，不默认加权压成一个新向量。

4.5.3 保留 50 段相对进程辅助视图

同时保存均分 50 段的相对进程辅助视图，定义为

$$I_{it}^{50} = \left[\frac{(t-1)D_i}{50}, \frac{tD_i}{50} \right), \quad t = 1, \dots, 50. \quad (4.34)$$

它表示片段相对进程，而非统一的物理分辨率。题面最短和最长视频的每段分别为 0.05296 s 和 0.69134 s，因此不将其作为唯一主时间轴。所有区间采用左闭右开记法；终止时间戳单独记录，不为包含末端单点而增加虚假时长。50 段视图覆盖全片，不截取前 50 个词；重复使用同一原生来源时保留来源编号，不声称新增了独立观测。式 (4.34) 仅定义问题一的相对进程视图，不解释附件二既有 **aligned** 特征的时间含义。

4.6 子模型五：边界后验驱动的概率时间投影与低阶动态签名

4.6.1 概率时间投影：传播边界不确定性而不重排时间

物理时间投影仍是默认主结果。仅当固定转写状态图能够提供可信的前向–后向后验时，额外生成概率时间视图。对模态 $m \in \{A, V\}$ 的第 r 个原生来源区间 J_{ir}^m ，定义词 k 对该来源的后验重叠质量

$$\beta_{ikrd}^m = q_{ird}^m O_{ird}^m \int_{J_{ir}^m} \pi_{ik}(\tau) d\tau. \quad (4.35)$$

于是词级后验特征为

$$\bar{x}_{ikd,\text{post}}^m = \frac{\sum_r \beta_{ikrd}^m x_{ird}^m}{\sum_r \beta_{ikrd}^m}, \quad \sum_r \beta_{ikrd}^m > 0. \quad (4.36)$$

对固定秒级网格仍使用式 (4.19) 将文本词向量映射至共同时间轴；语音和视觉的媒体时间戳本身不因文本边界不确定而改变。若研究需要把“某词附近的音频/视觉证据”作为词级核验视图，则采用式 (4.35)–(4.36)，而不是依据情感向量相似度移动来源。为审计边界不确定性，保存每个词的后验均值边界、预设分位区间、区间宽度以及人工参考边界是否落入后验区间。若对齐器只能输出单一路径和路径分数，则该概率分支标记为未启用，不以人为高斯扰动伪造后验。

4.6.2 低阶路径签名：补充均值与方差丢失的变化顺序

固定窗口内的均值与方差不能区分“先升后降”和“先降后升”等顺序结构，因此在不改变 768/74/35 主特征维度的前提下，增加一个独立的低维动态辅助分支。对连续有效区间 $I = [a, b]$ ，从少量预先指定且物理含义清楚的动态通道构造时间增广路径

$$\mathbf{z}(\tau) = [\tilde{\tau}, z_1(\tau), \dots, z_p(\tau)], \quad \tilde{\tau} = \frac{\tau - a}{b - a}. \quad (4.37)$$

主配置仅从可解释声学动态量、AU、嘴部开合等通道中选择，选择表和顺序在配置中固定；不在全体 BERT 维度上直接展开高阶交互。一阶签名为各通道总增量

$$S_j(I) = \int_a^b \mathrm{d}z_j(\tau), \quad (4.38)$$

二阶迭代积分为

$$S_{jk}(I) = \int_{a < u < v < b} \mathrm{d}z_j(u) \mathrm{d}z_k(v), \quad (4.39)$$

其反对称部分

$$A_{jk}(I) = \frac{1}{2}(S_{jk}(I) - S_{kj}(I)) \quad (4.40)$$

描述通道间有向变化关系。若含时间坐标共有 $d = p + 1$ 个通道，则二阶对数签名维度为

$$d_{\text{dyn}} = d + \frac{d(d-1)}{2}, \quad (4.41)$$

规模可控，并作为 $\mathbf{X}_{i,\text{dyn}}$ 独立保存。路径签名只在连续有效段内计算：解码缺口、轨迹切换或所选通道未定义时立即分段，不把缺失值填零后连接成虚假轨迹；若窗口内有效连续长度不足预设阈值，则动态分支标记为不可估计。插值仅用于在已观测连续段内定义分段线性路径，插值方法、原生节点和窗口边界全部存档。二阶有向项只作为顺序交互描述，不解释为因果效应。

4.6.3 主结果保护与分支状态

最终交付中， $\mathbf{X}_{i,\text{sec}}^m$ 及其来源链始终作为问题一主结果； $\mathbf{X}_{i,\text{post}}$ 和 $\mathbf{X}_{i,\text{dyn}}$ 均为可选辅助分支。任何概率或动态分支均不得覆盖原始时间戳、真实覆盖率和原生观测掩码。对齐器后验不可得、后验校准审计失败或动态窗口有效长度不足时，只记录“未启用/不可估计”，不使用人为填充值冒充观测。

4.7 子模型六：词查询驱动的双视图核验与受控内容增强

4.7.1 物理对应、内容匹配和情感贡献的分离

本节针对“某个词对应哪些音频、视觉位置”给出可直接核验的数学对象。区分三类对象：物理对应说明素材在何时发生；内容匹配说明指定表征能否检索到同时段内容；任务贡献说明内容如何影响给定情感预测。三者不共享同一个真实性解释，不能以注意力图的亮度代替时间准确性或因果贡献。注意力解释需要明确对象与对照实验，相关争论与诊断研究见文献 [5, 6]。

4.7.2 以原词为行、原生来源为列的物理高亮矩阵

对模态 $m \in \{A, V\}$ 预先指定核验通道组 $\mathcal{D}_m^{\text{audit}}$ ，令 χ_{ir}^m 表示该来源至少有核验所需的合法观测。它只用于显示，不能覆盖逐维掩码。以可靠词区间 $I_{ik}^{\text{word}} = [s_{ik}, e_{ik})$ 构造

$$d_{ikr}^m = |I_{ik}^{\text{word}} \cap J_{ir}^m| \chi_{ir}^m, \quad D_{ik}^{m,\text{src}} = \sum_r d_{ikr}^m. \quad (4.42)$$

当 $D_{ik}^{m,\text{src}} > 0$ 时，定义

$$H_{ikr}^{m,\text{time}} = \frac{d_{ikr}^m}{D_{ik}^{m,\text{src}}}, \quad \sum_r H_{ikr}^{m,\text{time}} = 1. \quad (4.43)$$

无来源时整行保存为空及失败标记，而不是归一化一个全零行。 $H^{m,\text{time}}$ 的形状为 $K_i \times N_i^m$ ，查询某词就是选中第 k 行，并把列号回查至真实音频采样或视频帧。一个词覆盖多帧时，应呈连续条带，不要求单像素尖峰。相邻词和持续表情可以共享部分分析窗口，但归属时间质量不重复计数。若有合法的边界后验，可另以 $d_{ikr}^{m,\text{post}} = \chi_{ir}^m \int_{J_{ir}^m} \pi_{ik}(\tau) d\tau$ 替代硬交长形成独立后验图；不能在无后验的情况下人为平滑后称为后验概率。质量加权图 $H^{m,Q} \propto H^{m,\text{time}} q_{ir}^{m,\text{audit}}$ 也单独命名，不替换纯时间图。音频与视觉帧数不同，行和为 one 的质量在更多列上被分摊，因而单格峰值不能直接跨模态比较。可另外显示密度

$$h_{ikr}^{m,\text{time}} = \frac{H_{ikr}^{m,\text{time}}}{|J_{ir}^m|}, \quad |J_{ir}^m| > 0, \quad (4.44)$$

并标注“每秒质量”；或者仅在显示端用共同真实时间区间归集质量。显示重采样不得回写训练特征，不能将未知区间内部分配假装为新的实测定位。

4.7.3 不含额外位置通道的内容探针

BERT 词向量、手工声学量和 AU/姿态量不在天然相同的坐标系内。将三者截成相同维数后直接比较余弦，不能证明存在或不存在跨模态语义。为研究内容可检索性，可另设小型投影器；它是**可选的训练折限定辅助分支**，默认不改变物理对齐。记 $\tilde{\mathbf{x}}_{ik}^T$ 为保留词身份的标准化文本向量， $\tilde{\mathbf{x}}_{ir}^m$ 为标准化原生来源向量。缺失维以独立掩码处理。定义

$$\begin{aligned} \mathbf{e}_{ik}^T &= \mathcal{N}_\varepsilon(F_T(\tilde{\mathbf{x}}_{ik}^T)), \\ \mathbf{e}_{ir}^m &= \mathcal{N}_\varepsilon(F_m(\tilde{\mathbf{x}}_{ir}^m)), \quad \mathcal{N}_\varepsilon(\mathbf{v}) = \frac{\mathbf{v}}{\max(\|\mathbf{v}\|_2, \varepsilon_e)}, \\ S_{ikr}^{m,\text{cont}} &= (\mathbf{e}_{ik}^T)^\top \mathbf{e}_{ir}^m / \tau_c. \end{aligned} \quad (4.45)$$

投影器可取线性层或一层小型感知机，公共维数初值 32，冻结底层提取器。 F_m 不额外接收时间戳、原词编号、槽位编号、视频 ID、位置编码或相对时间偏置。底层上下文编码器本身仍可能含隐式位置和整句信息，因此“不加显式位置”不等于完全消除了位置线索，仍须进行后述控制实验。对同一样本的全部合法来源，绘制

$$H_{ikr}^{m,\text{cont}} = \frac{\chi_{ir}^m \exp S_{ikr}^{m,\text{cont}}}{\sum_{r'} \chi_{ir'}^m \exp S_{ikr'}^{m,\text{cont}}}. \quad (4.46)$$

该图不乘词区间先验、不加近对角偏置。改变显示色阶或减小温度使图变尖，不增加原始证据；候选来源集与温度必须固定并记录。

4.7.4 多正例、时间保护带与共享内容学习

词对应一段声音及多帧画面，因此不能以单个帧号作为唯一正确类别。将真实重叠位置作为正例集合 $\mathcal{P}_{ik}^m$ ，将与该词时间保护带及正例实际计算范围不相交的合法来源作为候选负例集合 $\mathcal{N}_{ik}^m$ 。重复词、跨窗口复用及近邻上下文可能形成假负例，需排除或降权；同视频远段与不同视频样本分层抽取并分别评估，以防仅学到说话者身份。令 $\mathcal{C}_{ik}^m = \mathcal{P}_{ik}^m \cup \mathcal{N}_{ik}^m$ ，以 $T_{ikr}^m = H_{ikr}^{m,\text{time}}$ 为多正例软目标，定义

$$\begin{aligned} p_{ikr}^{m,\text{cand}} &= \frac{\exp S_{ikr}^{m,\text{cont}}}{\sum_{s \in \mathcal{C}_{ik}^m} \exp S_{iks}^{m,\text{cont}}}, \\ \mathcal{L}_{\text{cont}} &= - \sum_{m \in \{A, V\}} \lambda_m \frac{\sum_{i,k} \zeta_{ik}^m \sum_{r \in \mathcal{P}_{ik}^m} T_{ikr}^m \log p_{ikr}^{m,\text{cand}}}{\max(\varepsilon_\zeta, \sum_{i,k} \zeta_{ik}^m)} + \lambda_F \sum_m \|F_m\|_F^2. \end{aligned} \quad (4.47)$$

ζ 只来自可审计词边界与来源质量，不由当前相似度或情感标签自我确认。正例、负例不齐全时不计算相应项。式 (4.47) 借鉴对比表征学习的候选区分思想 [7]，其多正例目标由本题时间对应构造，并非词义与画面内容的一一语义标注。文本与语音可作为主要检索探针；文本与视觉不预设必须逐词匹配， $\lambda_V = 0$ 必须是允许的候选。面部动作描述表情、姿态和视线，不是英文词义的词典。若视觉内容探针在按原始视频分组的留出样本中未提供

稳定增益，不通过增加尖锐化、循环约束或一致性惩罚强行获得漂亮热图。主问题一不因上述投影器训练而重排时间。投影器只学习共享内容子空间，私有原特征仍完整保留。参数选择不用情感标签；若按情感任务选择该分支，则归入问题二，并遵循该问的嵌套验证。

4.7.5 时间引导视图只作辅助，不作为独立内容证据

确有必要在可靠词区间内细分读取权重时，可定义

$$H_{ikr}^{m,\text{guide}} = \frac{H_{ikr}^{m,\text{time}} \exp(\eta_c S_{ikr}^{m,\text{cont}})}{\sum_s H_{iks}^{m,\text{time}} \exp(\eta_c S_{iks}^{m,\text{cont}})}, \quad \eta_c \geq 0. \quad (4.48)$$

它只在已有真实时间支撑内重分配， $\eta_c = 0$ 退化为物理对应。图中时间先验造成的高亮必须明确标注，不能使用该图在原词区间内的集中程度证明“仅靠内容完成定位”。主特征继续使用原来的物理投影， H^{guide} 及相关聚合另存，并允许完全关闭。

4.7.6 防止位置捷径、恒等循环与内容坍塌的核验

若某跨模态图经 K_ℓ 个共享槽位线性分解为 $F_{A \rightarrow V} = R_A L_V$ ，则

$$\text{rank}(F_{A \rightarrow V} F_{V \rightarrow A}) \leq K_\ell. \quad (4.49)$$

当音频位置数 $N_A > K_\ell$ ，该循环矩阵不可能等于 I_{N_A} 。在没有额外满秩旁路、且确实满足上述线性乘积分解时，还满足

$$\|F_{A \rightarrow V} F_{V \rightarrow A} - I_{N_A}\|_F \geq \sqrt{N_A - K_\ell}. \quad (4.50)$$

因此，不把所有跨分辨率循环都压成单位阵作为主训练目标。声学高采样、词级重复和持续表情还会自然产生多对一关系。必要的循环检查应比较返回的真实时间邻域，而不是每个原生帧必须精确回到同一编号。在相同投影器、候选集合、温度和色阶下分别报告：纯物理图、无额外位置的内容图、时间引导图、内容置换图和仅位置基线。置换保持位置与掩码不变，独立置换内容；同时做同视频分块置换和跨视频匹配置换，说明各自破坏了什么。物理图在内容置换后保持不变是正常现象，不是失败。若带时间引导的图不变，也不能由此证明内容有效。注意力仅说明特定读取机制，不自动构成解释 [5, 6]。不循环绕回地平移目标真实时间窗，设 Δ_0 为预先固定的偏移集合，使用对全部偏移均合法的共同查询集 $\mathcal{K}_0$ ，定义

$$S_m(\delta) = \frac{1}{|\mathcal{K}_0|} \sum_{(i,k) \in \mathcal{K}_0} (e_{ik}^T)^\top e_{ik}^m(\delta), \quad (4.51)$$

$$\Delta S_m = S_m(0) - \frac{1}{|\Delta_0|} \sum_{\delta \in \Delta_0} S_m(\delta).$$

$e_{ik}^m(\delta)$ 由平移窗中真实来源聚合后投影，聚合规则固定。无秒级元数据则只报告索引偏移。按 `video_id` 重采样置信区间；同视频时间步、重复词和多个偏移不算独立样本。另报正

确窗与错位窗的配对检索 AUC、有效查询数、投影输出方差及有效秩；余弦绝对值较小不等于无效，须比较匹配与错位的差异。上述时间软目标与生成它的对齐器并不构成独立真值。物理定位质量必须回到人工审计词边界、媒体时间戳和原片段；内容检索性能必须在未参与投影器拟合的来源视频上评估。没有显著内容优势时如实保留“物理对齐成立、内容检索证据不足”的结论，不把它写成时间对齐失败。

4.8 特征存储、全量汇总与读取规范

4.8.1 按样本存储与视图组织

采用每个样本一个 HDF5 文件的组织方式，文件路径由无歧义主键清单定位；等价的 NPZ 加变长偏移表也可使用，但不得丢弃变长内容。元数据另存为 manifest_q1.jsonl，每个文件包含以下层次。读取顺序为：通过清单主键定位样本文件，选择视图，读取特征、

表 4.3 特征文件的逻辑分组与读取内容

分组	主要内容	读取与核验作用
meta	主键、素材路径、标签、哈希、时长、流偏移、配置标识	判断样本和文件是否一一对应
native/T,A,V	原生特征、 $\mathcal{R}$ 、 J 、逐维掩码、质量及来源索引	保留未聚合数据，允许重建任意视图
views/word	词文本、时间、三模态聚合及非词区间类型	核验原词、声音和视频帧对应
views/sec	固定秒级均值、方差、覆盖率、计数及真实长度	主时序输入；读特征时同时读掩码
views/multi	三尺度特征及每个窗口真实区间	区分局部观测和邻域上下文
views/phase50	相对进程特征及对应区间	固定长度辅助展示，不替代秒级主轴
views/dynamics	选定通道、连续有效段、低阶签名及状态	描述窗口内变化顺序；不覆盖主特征
alignment	硬边界、词边界后验、占据概率、分位区间及后验投影权重	回查确定边界与概率边界，不混淆观察与估计
quality/sec	对齐质量、覆盖、可用性和质量来源	为第二问提供独立且可关闭的质量入口
query/word	词-原生来源时间矩阵及列索引	选中一个词即可回查对应素材
probe/content	可选内容投影器、内容图及控制实验状态	与时间引导图分开；不作为默认预测输入
audit	工具版本、模型哈希、状态、错误码及处理日志	支持复现和异常追踪

时间区间、有效长度及掩码，再按来源索引回查原生行和原素材。时间戳使用足以保持精

度的浮点表示；原始采样与帧索引使用整数；质量权重、稀疏矩阵的行列含义、坐标顺序和端点规则均写入模式说明。

4.8.2 变长序列、填充与来源关系

磁盘上的原生和主时序采用真实长度存储；组成批次时才填充到批内最大长度 $L_{\max}$ ，定义

$$P_{it}^{\text{sec}} = \mathbf{1}\{1 \leq t \leq L_i\}, \quad 1 \leq t \leq L_{\max}. \quad (4.52)$$

填充值在标准化后设为零，填充区间时间戳使用无效哨兵并配合 P 读取；不能以 $[0, 0)$ 冒充真实片段。真实时间内的模态缺失仍满足 $P = 1$ ，以 O 及原因字段区分。零占位不表示中性情感。对每个目标区间保存参与聚合的原生行索引及逐维权重；语音分支到联合声学行再到目标区间的两级投影须保留完整链路。多尺度重复使用同一观测只增加视图，不增加独立观测数量。

4.8.3 100 条样本与 300 条模态明细

实际实施后输出 100 行样本级清单和 300 行模态级明细。即使某模态提取失败，也保留相应记录与原因；样本覆盖完整不等于所有模态成功。不得用虚构编号或设计维度表替代真实提取结果。

表 4.4 全量模态明细表的字段规范（非实测结果表）

字段	定义
sample_id, video_id, clip_id	原样保留的唯一主键与来源索引
modality, source_duration	模态类型与原片段时长 D_i
observed_duration	可靠来源归属区间并集长度，另存逐维有效时长
native_length, native_dim	原生行数与实际主维度
aligned_length, aligned_dim	主视图真实长度与维度；不能填批次填充长度
alignment_type, granularity	固定秒级主对齐及辅助视图类型、实际时间粒度
coverage, missing_reason	连续覆盖率、自然未定义及失败原因分列
quality_available, quality_path	质量字段可用性、计算规则与独立记录路径
query_map_path, probe_status	原词查询矩阵路径与可选内容探针状态
feature_path, source_hash	特征文件定位与素材校验值
status, config_hash	成功、部分成功或失败；完整配置的校验值

4.9 核验指标、典型样本与模型对照

4.9.1 样本、时间和对齐准确性

先要求标签、素材和输出主键集合相等且分别唯一，再计算覆盖指标。定义

$$C_{\text{id}} = \frac{\#\{\text{有且仅有一条输出记录的原始样本}\}}{100}, \quad (4.53)$$

$$C_{\text{valid},d}^m = \frac{\sum_{i,t} |I_{it}^{\text{sec}}| c_{itd}^m}{\sum_{i,t} |I_{it}^{\text{sec}}|}, \quad C_{\text{valid}}^m = \frac{1}{d_m} \sum_{d=1}^{d_m} C_{\text{valid},d}^m, \quad (4.54)$$

$$E_{\text{bnd}} = \frac{1}{2K_{\text{audit}}} \sum_{k \in \mathcal{K}_{\text{audit}}} (|\hat{s}_k - s_k^{\text{audit}}| + |\hat{e}_k - e_k^{\text{audit}}|). \quad (4.55)$$

覆盖率只在真实主网格求和，不含填充；逐维有效率的均值不替代各维状态明细。 $C_{\text{id}} = 1$ 仍需同时排查额外输出记录；不能仅凭这一比例断言主键集合完全一致。 E_{bnd} 仅在人工建立参考边界且 $K_{\text{audit}} > 0$ 时报告，否则注明未测，不输出虚构误差。所有有效区间须满足 $0 \leq s < e \leq D_i$ ，原生索引存在，主网格覆盖完整，相应归属区间在连续段内不重叠。强制对齐词序单调、概率边界只在合法状态图路径上取值、多尺度聚合满足式 (4.28)，均作为硬性审计条件。概率边界分支不以“更平滑”作为准确性证据，而在人工审计词集合 $\mathcal{K}_{\text{audit}}$ 上同时报告边界点误差与区间校准。设第 k 个词的 $(1 - \alpha)$ 后验边界区间分别为 $[L_k^s, U_k^s]$ 、 $[L_k^e, U_k^e]$ ，定义经验覆盖率

$$C_{\text{post}} = \frac{1}{2K_{\text{audit}}} \sum_{k \in \mathcal{K}_{\text{audit}}} [\mathbf{1}\{s_k^{\text{audit}} \in [L_k^s, U_k^s]\} + \mathbf{1}\{e_k^{\text{audit}} \in [L_k^e, U_k^e]\}]. \quad (4.56)$$

同时报告平均区间宽度

$$W_{\text{post}} = \frac{1}{2K_{\text{audit}}} \sum_{k \in \mathcal{K}_{\text{audit}}} [(U_k^s - L_k^s) + (U_k^e - L_k^e)]. \quad (4.57)$$

覆盖率和宽度必须联合呈现，避免用极宽区间换取表面上的高覆盖。若没有人工参考边界，则只报告模型内部后验宽度，不将其称为真实校准率。动态签名分支则通过与相同主特征基线的独立消融检验其增量价值，不以维度增加本身作为优势。

4.9.2 典型样本核验卡

按时长、对齐质量和缺失情况预先分层选样，至少展示一条真实样本；条件允许时再增加一条困难样本。每张核验卡应给出主键、英文原词、词边界、对应音频采样范围、原始帧时间戳与帧号、特征行编号、权重、覆盖率和异常原因，不只展示最易对齐的情况。本节给出的是模型和交付规范。原始素材未实际处理时，核验卡、100 行真实清单、实测边界误差及特征文件均不能宣称已经生成。

表 4.5 典型样本联合核验卡的填写规范（不含虚构样本）

核验项	必须填写的真实来源内容
样本与原文	双字段主键、原视频哈希、原转写片段与字符位置
文本时间	原词编号、 $[s_{ik}, e_{ik})$ 、边界质量和定位状态
音频对应	音轨偏移、原生采样率、采样点范围及分析窗口记录
视频对应	实际帧号和 PTS、目标轨迹编号，不以假定帧率换算
三模态特征	词级与秒级行索引、768/74/35 维定义及有效维度
对齐权重	物理重叠权重；启用后验分支时另列词占据概率、期望重叠质量与后验区间
词查询核验	原词为行、原生来源为列的物理图；标出真实秒数和对应区间
内容对照	内容图、时间引导图、仅位置和内容置换对照分别展示
核验材料	词边界叠加音频波形、代表帧及异常说明；人工边界另存

4.9.3 单因素对照而非复杂模块累加

采用相同原生特征、主时间网格和观测掩码，设置相互独立的对照：B0 为质量加权物理区间投影；B1 在 B0 上仅加入修正后的尺度处理与视觉几何一致聚合；B2 在 B1 上仅加入词边界后验概率投影；B3 在 B1 上仅加入低阶路径签名动态分支；B4 在 B1 上加入冻结深层语音辅助分支，用于区分“表征增强”与“时间模型增强”的贡献。各版本都比较主键完整性、时间边界误差、连续覆盖、来源可追溯性与人工核验；若后续用轻量预测器检验情感信息保留，则所有尺度、降维和调参只在训练折拟合，并按 `video_id` 分组。新增 B5 仅检查式 (4.43) 的词查询与人工素材对应，B6 在 B1 上独立加入式 (4.47) 的内容探针；B6 的验证不使用时间引导先验。质量记录本身作为元数据不改变主特征，是否改善预测由问题二消融检验。不预先宣称某一复杂版本预测性能最高。Soft-DTW、语义相似度驱动的全局重排和公共空间学习不进入默认问题一链路，以免将“表达差异”误当成“时间错误”。若另作研究，只能作为独立对照，不得覆盖问题一保存的物理特征及来源记录。

4.10 参数锁定、运行流程与复现说明

表 4.6 中未给出数值的后验分位水平、动态窗口等参数，应在实施前写入完整配置，说明候选范围和无标签的时间核验选择规则；不能将“待锁定”当作已经复现。工具确切版本与模型哈希应从真实运行环境写入，而不是在未运行时虚构。软件环境不同可能引入数值差异，须预先声明逐位一致或容差一致的复现目标。

4.10.1 全量可追溯特征生成流程

输入：原始视频、转写、标签表和已锁定的配置。

输出：原生层、固定秒级主视图、词级及辅助视图、掩码、来源映射、真实全量清单和日志。

表 4.6 关键配置与实际复现时必须锁定的信息

模块	本稿固定或建议配置	实际执行必须记录
媒体与时间	单声道 16 kHz; 原始 PTS 和流偏移	FFmpeg/ffprobe 构建、命令、解码及重采样映射
文本	冻结 BERT; 末四层; 510 子词窗口、384 步长	权重及词典修订号、哈希、分词器与推理依赖版本
谱与倒谱	400 点窗、160 点步、512 点 FFT; 40+13+13 维	窗函数、Mel 映射、DCT、差分、边界及幅度规则
韵律	8 项; 逐分量原生窗口与掩码	估计器名称版本、基频范围、阈值、单位和时间锚
视觉	17+6+6+6 维; 门槛初值 0.8	OpenFace 构建、列序、几何编号、置信度及轨迹规则
词级锚定	固定转写状态图、单调路径	MFA 版本、声学模型和发音词典哈希、异常词处理
时序视图	主步长 0.1 s; 尺度 0.1/0.3/0.7 s; 辅助 50 段	实际窗口端点、归属区间、完整来源权重及长度
边界后验	仅在固定转写状态图可得可靠前向-后向权重时启用	对齐器后验导出方法、分位水平、状态映射、采样数及人工校准结果
动态签名	时间增广、最高二阶; 仅连续有效段计算	通道表、插值规则、最小有效长度、窗口尺度、签名实现与维度
质量接口	式 (4.25)–(4.27)	字段可用性、质量定义、训练/测试兼容规则
词查询	真实词区间与原生来源交长	矩阵方向、质量/密度单位、原生列索引、显示色阶
可选内容探针	公共维数初值 32; 冻结底层编码器	温度、保护带、正负例规则、视频分组及置换对照
环境与输出	固定随机源; 原生层和派生层分离	操作系统、依赖锁、模型哈希、随机种子、硬件及输出校验值

1. 核验双字段主键、文件集合、原标签逻辑及哈希; 保留所有失败记录。
2. 对每个样本解析统一时间原点, 解码音轨、原始帧和真实时间戳, 登记偏移及缺口。
3. 按式 (4.9)、(4.11)、(4.12) 提取主特征, 保存逐维有效性、质量和计算素材范围。
4. 在固定转写状态图上求解式 (4.14), 记录可靠词边界、非词区间与不可定位词; 若可获得前向-后向后验, 则进一步计算式 (4.17)–(4.19), 不以标签修正。
5. 建立连续段内不重叠的时间归属区间, 按式 (4.21) 生成物理聚合, 同时计算方差、连续覆盖、观测数量以及式 (4.29) 的充分统计量。
6. 根据式 (4.31)、(4.33) 和(4.34) 构造主视图与辅助视图; 对非嵌套目标边界回查原生来

源，不对中间均值二次切分。

7. 若词边界后验可用，生成概率时间视图并保存后验区间与校准字段；在连续有效段上按式 (4.37)–(4.40) 生成低阶动态签名辅助分支。
8. 按式 (4.25)–(4.27) 归集质量；按式 (4.43) 生成以原词为行的物理来源矩阵。
9. 检查维度、数值、区间、索引、掩码、覆盖与来源链路；按合法训练规则生成可选标准化层，批处理时另行填充。
10. 如研究内容可检索性，另以视频分组训练式 (4.47) 的小型探针，分别报告内容、仅位置和置换对照；不回写主时间轴。
11. 写入特征文件、模式说明和日志；生成真实 100 行样本汇总与 300 行模态明细，并按质量分层完成典型样本联合回看。

最终方案以“**可靠时间锚定、逐维质量投影、词查询双视图核验、连续缺失接口**”为主线；边界后验、低阶顺序动态和内容探针为可关闭辅助分支。主要结果是忠实、完整、可回溯的三模态时序表示，而不是将每个位置强行补齐或强迫三模态产生相同情感。

5 问题二：连续局部缺失下的概率补全与可靠度门控预测

5.1 问题接口、建模主线与适用范围

5.1.1 输入版本一致性与两问的职责

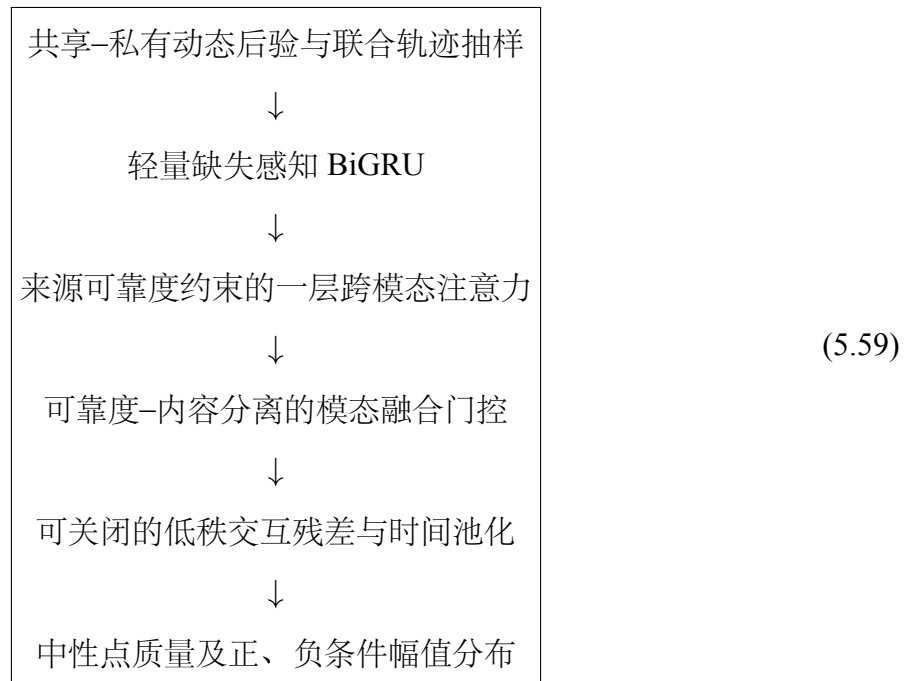
本问输入为合法训练资料与附件三**同一特征版本**的三模态数值序列、真实长度、结构掩码以及两端均可获得的质量元数据。自生成的 768/74/35 维特征不因维数一致便等同于专项集特征。对训练与测试核验

$$\mathcal{I} = (\text{编码器及修订号、列定义、时间轴、聚合方式、尺度、填充与缺失规则}). \quad (5.58)$$

接口一致或有合法训练资料支持的确定转换时才使用同一模型；未知接口不能靠随机投影或补零掩盖。本节的状态步要求三模态已有可核验的共同位置含义；原生采样率不同时，有合法时间戳才能按第一问规则投影至共同轴。只有各自未知索引的序列不能因为行号相同而共享同一状态步，应先解决接口定义，不擅自强制对齐。若有官方划分，遵循官方划分；若自行划分，以原始 `video_id` 分组。已知附件一有 100 条片段、37 个来源视频，但不由此臆测附件二的训练规模。以下采用紧凑结构作为起点，参数容量由实际合法训练量和内层验证决定。第一问的物理对齐不使用情感标签，也不通过本问的损失反向移动原始时间戳。第一问可选内容探针、词查询矩阵和多尺度辅助特征**不是专项测试默认可用输入**。只有接口匹配、缺失已同步传播且不含被隐藏内容时，才可另行评估其增量价值。主方案不依赖这些可选分支，保证附件三只提供基础特征时仍有完整定义。

5.1.2 共享动态、轻量时序和双层门控的有机分工

构建可靠度约束的概率补全-时序交互融合模型，记为 C_{RG} ：



概率补全负责推断缺失而非伪造观测；BiGRU 负责模态内非线性变化；来源门控限制低可信信息被其他模态读取；融合门控决定最终模态权重；低秩残差提供受控交互；联合分布保证极性与强度解释一致。教师蒸馏与分组鲁棒风险各保留一套，不把同义损失重复叠加。XGBoost 统计分支不列入默认主模型，只可作为合法折外对照。设 X_{obs} 、 X_{mis} 分别为可见和缺失数值，完整预测定义为

$$p(y | X_{\text{obs}}, M, P) = \int p_{\theta}(y | X_{\text{obs}}, X_{\text{mis}}, M, P) p_{\omega}(X_{\text{mis}} | X_{\text{obs}}, M, P) dX_{\text{mis}}. \quad (5.60)$$

后验均值输入是该框架的简化对照，不与对非线性网络输出进行后验平均混为一谈。

5.2 连续区间缺失生成与可观测状态

5.2.1 同步与异步区间

主公式针对专项文件中“有效行的全部特征维共同缺失”的情况。先在原始数值上依据来源规则识别 O 与 P ，再标准化，不能对标准化后的零重新判缺失。令 K_i^m 为模态 m 的遮蔽区间个数，定义

$$\begin{aligned} \mathcal{B}_i^m &= \bigcup_{j=1}^{K_i^m} [a_{ij}^m, a_{ij}^m + \ell_{ij}^m - 1], \\ B_{it}^m &= 1 - \mathbf{1}\{t \in \mathcal{B}_i^m\}, \quad M_{it}^m = O_{it}^m B_{it}^m. \end{aligned} \quad (5.61)$$

$K_i^m = 0$ 时 $B^m \equiv 1$ 。不同模态起止相同为同步缺失，起止不同为异步或交错缺失。覆盖单模态、双模态以及三模态局部区间；对原本有观测且被人工遮蔽的模态要求 $\sum_t M_{it}^m > 0$ ，不把主任务改为整模态消失。过短且不能构造局部缺失的样本保留原视图。训练采用与标签独立的场景抽样，同时保留不额外遮蔽的视图。默认先在题目提供的特征层遮蔽，再重建全部学生侧统计。上下文特征本身可能已编码邻域信息，因此这里模拟的是赛题定义的特征级缺失，不等价于“原词从未进入编码器”。不能通过未遮蔽原文、多尺度窗口、词查询图或动态签名恢复题目隐藏的内容。

5.2.2 三种缺失率及连续跨度

令 w_{it} 为有效位置的真实时长；只有索引时统一取 1。在共同有效时间轴上定义

$$\begin{aligned} r_{i,\text{nat}}^m &= 1 - \frac{\sum_t w_{it} O_{it}^m}{\sum_t w_{it} P_{it}^m}, \\ r_{i,\text{add}}^m &= \frac{\sum_t w_{it} O_{it}^m (1 - B_{it}^m)}{\sum_t w_{it} O_{it}^m}, \\ r_{i,\text{total}}^m &= 1 - \frac{\sum_t w_{it} M_{it}^m}{\sum_t w_{it} P_{it}^m} = r_{i,\text{nat}}^m + (1 - r_{i,\text{nat}}^m) r_{i,\text{add}}^m. \end{aligned} \quad (5.62)$$

分母为零时记不可估计。主总体口径为三模态等权平均 $r_i^{\text{all}} = \frac{1}{3} \sum_m r_{i,\text{total}}^m$ ，不以特征维数加权使文本主导缺失率。另记同步无观测比例

$$r_i^{\text{sync}} = \frac{\sum_t w_{it} P_{it} \mathbf{1}\{M_{it}^T + M_{it}^A + M_{it}^V = 0\}}{\sum_t w_{it} P_{it}}. \quad (5.63)$$

附件三只从现有文件可得总缺失状态，通常不能再分辨原有与新增缺失，不能虚构 O 的完整真值。缺失位置按起点、中心相对进程报告。无秒级元数据时 ℓ/L 只是序列跨度比例；有可靠时间戳时才额外报告秒级跨度。

5.3 子模型一：质量感知的共享-私有动态概率补全

5.3.1 增广状态与观测噪声

为简化记号，训练折标准化后的观测记为 $\mathbf{x}_{it}^m$ 。令

$$\begin{aligned} \mathbf{w}_{it} &= [\mathbf{z}_{it}; \mathbf{r}_{it}^T; \mathbf{r}_{it}^A; \mathbf{r}_{it}^V], \\ \mathbf{w}_{it} &= A\mathbf{w}_{i,t-1} + \boldsymbol{\eta}_{it}, \quad \boldsymbol{\eta}_{it} \sim \mathcal{N}(0, Q), \\ \mathbf{x}_{it}^m &= \mathbf{b}_m + C_m \mathbf{w}_{it} + \boldsymbol{\epsilon}_{it}^m. \end{aligned} \quad (5.64)$$

共享状态维数初值 $r_s = 8$ ，每模态私有状态维数 $r_p = 4$ ； A 按共享和私有块组织， C_m 对其他模态私有块为零。保留观测偏置 $\mathbf{b}_m$ ，因为按中位数中心化不保证数据均值为零。初始状态 $\mathbf{w}_{i1} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Pi_0)$ 。只对当前可见位置读取合法质量：

$$q_{it}^{m,*} = \begin{cases} q_{it}^{m,\text{obs}}, & M_{it}^m = 1, J_{it}^{m,Q} = 1, \\ 1, & \text{其他情况}. \end{cases} \quad (5.65)$$

学生中被遮蔽位置的质量字段可用标记置零，不传递其原始隐藏质量。采用可消融的质量-噪声假设

$$R_{it}^{m,\text{eff}} = \frac{R_m}{\max\{q_{it}^{m,*}, q_{\min}\}}, \quad M_{it}^m = 1, \quad (5.66)$$

其中 R_m 对角正定、 $q_{\min} > 0$ 。低质量观测获得较大观测噪声，从而较弱地约束共享状态。式 (5.66) 是设计假设，不是检测分数的天然概率解释；未知质量时退化为 R_m 。缺失位置的预测残差噪声使用训练得到的 R_m ，不利用被隐藏质量调整。 Q, Π_0, R_m 设置正特征值下界，并对 A 施加预先固定的范数约束。离散状态步对应当前合法时间轴；不把不同来源的未知索引轴混成物理时间。潜在坐标存在旋转不唯一性，不把某一状态维度命名为真实心理量。该模型是线性高斯动态模型的结构化应用，估计思想可参考文献 [8]。

5.3.2 后验均值、协方差及联合轨迹

省略样本下标，定义 $\|\mathbf{v}\|_G^2 = \mathbf{v}^\top G \mathbf{v}$ 。仅对有效观测建立条件能量

$$\begin{aligned} \mathcal{J}(\mathbf{w}_{1:L}) = & \frac{1}{2} \|\mathbf{w}_1 - \boldsymbol{\mu}_0\|_{\Pi_0^{-1}}^2 + \frac{1}{2} \sum_{t=2}^L \|\mathbf{w}_t - A\mathbf{w}_{t-1}\|_{Q^{-1}}^2 \\ & + \frac{1}{2} \sum_{t,m:M_t^m=1} \|\mathbf{x}_t^m - \mathbf{b}_m - C_m \mathbf{w}_t\|_{(R_t^{m,\text{eff}})^{-1}}^2. \end{aligned} \quad (5.67)$$

固定参数时，写为 $\mathcal{J} = \frac{1}{2} \mathbf{w}^\top \Lambda \mathbf{w} - \mathbf{h}^\top \mathbf{w} + \text{const}$ ，则

$$\boldsymbol{\mu} = \Lambda^{-1} \mathbf{h}, \quad \Sigma = \Lambda^{-1}, \quad p_\omega(\mathbf{w}_{1:L} \mid X_{\text{obs}}, M) = \mathcal{N}(\boldsymbol{\mu}, \Sigma). \quad (5.68)$$

填充不形成状态步，真实缺失形成状态步但不形成观测项。正定初始与过程噪声保证条件二次能量有唯一解。可用 **Kalman** 滤波和平滑求解，不要求实际计算稠密逆矩阵。若处理问题一逐维观测，使用行选择矩阵仅抽取已定义且可观测维度形成似然；自然无定义的物理量不能补成“真实可测值”。缺失行的预测分布为

$$\hat{\mathbf{x}}_{it}^m = \mathbf{b}_m + C_m \boldsymbol{\mu}_{it}, \quad V_{it}^m = C_m \Sigma_{i,tt} C_m^\top + R_m. \quad (5.69)$$

保留跨时间与跨模态协方差。第 b 条联合轨迹先抽取 $\mathbf{w}_{i,1:L}^{(b)}$ ，再在缺失位置抽取观测噪声，形成

$$\mathbf{x}_{it}^{m,(b)} = M_{it}^m \mathbf{x}_{it}^m + (P_{it}^m - M_{it}^m) \mathbf{x}_{it,\text{mis}}^{m,(b)}, \quad b = 1, \dots, K_{\text{MC}}. \quad (5.70)$$

不能独立抽每一行后声称保留了动态联合后验；也不能用后验均值覆盖真实观测。离线任务可利用已提供的双侧上下文，实时版本则需另改为单向滤波。

5.3.3 补全器估计、冻结与不确定性核验

只用当前训练折拟合

$$\omega^* = \arg \min_{\omega} \left\{ - \sum_{i \in \mathcal{D}_{\text{tr}}} \log p_\omega(X_{i,\text{obs}} \mid O_i, P_i, Q_{i,\text{legal}}) + \lambda_C \sum_m \|C_m\|_F^2 + \lambda_A \|A\|_F^2 \right\}. \quad (5.71)$$

训练似然包括协方差归一化项，不能只最小化式 (5.67) 的残差部分来估计噪声。可采用 EM 的后验矩与带正则加权估计，再实施稳定性约束；非凸参数估计不承诺全局最优。对训练或内层验证中原本可见、后来人工遮蔽的位置，分别检验补全 MAE、区间覆盖率和区间宽度。必要时用该合法校准部分确定固定方差修正因子，随后冻结。主方案冻结补全器再训练情感网络，避免网络通过任意放大方差逃避困难证据。自然缺失与附件三缺失没有真值，不计算伪造的补全误差。

5.4 子模型二：连续缺口驱动的动态可靠度

5.4.1 观测质量与补全不确定性的统一但不混同

取训练折确定的正定对角参考尺度 D_{ref}^m ，在标准化坐标上可固定为单位阵，定义缺失位置的标准化方差

$$u_{it}^m = \frac{1}{d_m} \text{tr} [(D_{\text{ref}}^m)^{-1} V_{it}^m]. \quad (5.72)$$

使用固定参考尺度而非 $R_m^{-1}V$ ，使指标表达可比的总预测方差；后者是相对观测噪声的比值，不能无条件视为总不确定性。真实观测位置的补全不确定性通道置零。令 v_{it} 为合法归一化时间或索引位置， $\delta_{it}^{m,-}$ 、 $\delta_{it}^{m,+}$ 分别为距最近前向、后向同模态观测的距离。对应方向没有观测时置为预先规定的最大距离 1，并保存无前向/后向观测标记。定义

$$g_{it}^m = \min\{\delta_{it}^{m,-}, \delta_{it}^{m,+}\}, \quad b_{it}^m = \frac{\text{包含当前位置的连续缺口跨度}}{\text{样本有效跨度}}. \quad (5.73)$$

有效观测处取 $g = b = 0$ ；填充不构成缺口。距离没有物理单位时只能解释为序列跨度。长缺口中央远离两侧证据，端部缺口缺少一侧上下文，二者由距离、边界标记与动态后验共同描述。

5.4.2 单调可靠度代理

定义

$$\rho_{it}^m = \begin{cases} q_{it}^{m,*}, & P_{it}^m = 1, M_{it}^m = 1, \\ \rho_{\text{imp}} \exp[-\lambda_u u_{it}^m - \lambda_g g_{it}^m - \lambda_b b_{it}^m], & P_{it}^m = 1, M_{it}^m = 0, \\ 0, & P_{it}^m = 0, \end{cases} \quad (5.74)$$

其中 $0 < \rho_{\text{imp}} < 1$ ， $\lambda_u, \lambda_g, \lambda_b \geq 0$ 。系数经内层验证确定后固定，允许冗余距离惩罚系数为零。补全可靠度有受限上界，但不声称任意真实观测都优于任意补全；真实质量很低时也会被降低权重。对缺失位置，在其他量固定时有

$$\frac{\partial \rho}{\partial u} = -\lambda_u \rho \leq 0, \quad \frac{\partial \rho}{\partial g} = -\lambda_g \rho \leq 0, \quad \frac{\partial \rho}{\partial b} = -\lambda_b \rho \leq 0. \quad (5.75)$$

ρ 是可靠度代理而非校准正确概率；其作用是约束信息使用。不能依据模态之间情感相反便降低质量，否则会把讽刺、掩饰或转折抹平。区间长度与方差可能相关，因此距离项是否保留需消融，而非惩罚越多越好。

5.5 子模型三：轻量缺失感知双向 GRU

5.5.1 输入状态与低可信更新控制

对第 b 条联合补全轨迹，构造

$$\mathbf{e}_{it}^{m,(b)} = F_m^{\text{in}} \left[\mathbf{x}_{it}^{m,(b)}; M_{it}^m; q_{it}^{m,*}; J_{it}^{m,Q}; \log(1 + u_{it}^m); \delta_{it}^{m,-}; \delta_{it}^{m,+}; b_{it}^m; \mathbf{e}_{it}^{\text{edge}}; \text{PE}(v_{it}) \right]. \quad (5.76)$$

这里是**预测网络**，允许读取合法时间位置；与第一问“无额外位置内容探针”不同。后者的独立性要求不限制预测网络使用时间信息。 F_m^{in} 取小型感知机，输出初值 32 维。以前向为例，令

$$\mathbf{h}_{it}^{m,-} = \exp[-\text{softplus}(\mathbf{a}_m)\delta_{it}^{m,-}] \odot \vec{\mathbf{h}}_{i,t-1}^m. \quad (5.77)$$

标准 GRU 更新门和重置门为

$$\begin{aligned} \mathbf{z}_{it}^m &= \sigma(W_z \mathbf{e}_{it}^m + U_z \mathbf{h}_{it}^{m,-} + \mathbf{b}_z), \\ \mathbf{r}_{it}^m &= \sigma(W_r \mathbf{e}_{it}^m + U_r \mathbf{h}_{it}^{m,-} + \mathbf{b}_r), \\ \mathbf{h}_{it}^{m,\text{cand}} &= \tanh[W_h \mathbf{e}_{it}^m + U_h(\mathbf{r}_{it}^m \odot \mathbf{h}_{it}^{m,-}) + \mathbf{b}_h]. \end{aligned} \quad (5.78)$$

为降低不可靠补全对后续记忆的持续污染，在已有更新门上施加可靠度系数，不另增大网络：

$$\tilde{\mathbf{z}}_{it}^m = \rho_{it}^m \mathbf{z}_{it}^m, \quad \vec{\mathbf{h}}_{it}^m = (1 - \tilde{\mathbf{z}}_{it}^m) \odot \mathbf{h}_{it}^{m,-} + \tilde{\mathbf{z}}_{it}^m \odot \mathbf{h}_{it}^{m,\text{cand}}. \quad (5.79)$$

反向同理并使用后向距离，最终拼接两方向为 $\mathbf{h}_{it}^{m,(b)}$ 。每方向隐维初值 16、仅一层。式 (5.79) 可独立消融，避免与后续门控重复压制过强。时间间隔和掩码进入循环网络的思想借鉴 GRU-D[9]，这里改写为离线双向、多模态与后验补全版本。GRU 内部门控制时间记忆，后续门控控制模态间读取和融合，两者不相互替代。填充不更新隐藏状态，真实时间内的缺失位置保留。该限制减少某类污染通道，但不构成对整个非线性网络误差的全局单调保证。

5.6 子模型四：GRU 后的来源门控跨模态注意力

5.6.1 有界内容得分与空来源

跨模态注意力用于读取其他模态的不同时间位置，借鉴 MulT 的方向性交互 [10]，仅保留一层。省略样本和轨迹下标，定义有界值表示 $\mathbf{v}_r^n = \tanh(W_V \mathbf{h}_r^n)$ ，以及

$$e_{tr}^{m \leftarrow n} = \kappa_A \tanh \left[\frac{(W_Q \mathbf{h}_t^m)^\top (W_K \mathbf{h}_r^n)}{\sqrt{d_h}} + b_{tr}^{mn} \right], \quad n \neq m. \quad (5.80)$$

b_{tr}^{mn} 基于合法相对时间或索引差，不强制跨模态所有情感事件同时发生。参数尽量在方向之间共享，另设小型模态嵌入。内容分数被限制在 $[-\kappa_A, \kappa_A]$ ，以免任意大的得分完全抵消可靠度折扣。在所有其他模态有效来源中定义与内容无关的基准质量 $\pi_r^{n|m}$ ，使 $\sum_{n \neq m, r} \pi_r^{n|m} = 1$ 。默认先对可用来源模态等权，再在每个模态内按真实区间长度或均匀索引分配。这样高采样率模态不因行数多而天然获得更大总权重。无候选时所有 $\pi = 0$ 。定义来源门控权重

$$A_{tr}^{m \leftarrow n} = \frac{\pi_r^{n|m} \rho_r^n \exp(e_{tr}^{m \leftarrow n})}{\lambda_\emptyset^A + \sum_{j \neq m, s} \pi_s^{j|m} \rho_s^j \exp(e_{ts}^{m \leftarrow j})}, \quad \lambda_\emptyset^A > 0. \quad (5.81)$$

空来源取零值向量，其权重为分母中 $\lambda_\emptyset^A$ 所占比例。若所有实际来源可靠度趋零，读取结果趋零而非重新归一化为一整份不可靠信息。更新

$$\tilde{\mathbf{h}}_t^m = \tanh \left[W_0 \mathbf{h}_t^m + \eta_A W_O \sum_{n \neq m, r} A_{tr}^{m \leftarrow n} \mathbf{v}_r^n \right], \quad 0 \leq \eta_A \leq 1. \quad (5.82)$$

第一层门控放在 GRU 后、跨模态读取时。来自其他模态的表示不能把原来的 $M_t^m = 0$ 改为 1；借到音频证据不等于重新观察到了缺失单词。

5.7 子模型五：可靠度与内容分离的融合门控

5.7.1 条件单调权重与先验专家

对交互后的表示计算有界内容相关性

$$s_t^m = \kappa_F \tanh \left[\mathbf{v}^\top \tanh(W_s \tilde{\mathbf{h}}_t^m + \mathbf{b}_s) \right]. \quad (5.83)$$

采用乘性可靠度融合，加入先验专家 $\emptyset$ ：

$$\begin{aligned} Z_t &= \lambda_\emptyset^F + \sum_{n \in \mathcal{M}} \rho_t^n \exp(s_t^n), \\ \alpha_t^m &= \frac{\rho_t^m \exp(s_t^m)}{Z_t}, \quad \alpha_t^\emptyset = \frac{\lambda_\emptyset^F}{Z_t}, \\ \mathbf{g}_t &= \alpha_t^\emptyset \mathbf{h}_\emptyset + \sum_{m \in \mathcal{M}} \alpha_t^m \tilde{\mathbf{h}}_t^m, \quad \lambda_\emptyset^F > 0. \end{aligned} \quad (5.84)$$

质量与内容分工明确： ρ 描述证据可靠程度， s 描述当前任务相关性。乘性门控融合可参考 GMU 思想 [11]，本稿额外固定了可靠度入口和先验分支。空专家不恒等于中性，而是在低信息条件下的保守表示。固定内容分数、其他模态可靠度和先验质量，得到

$$\frac{\partial \alpha_t^m}{\partial \rho_t^m} = \frac{e^{s_t^m} [\lambda_\emptyset^F + \sum_{n \neq m} \rho_t^n e^{s_t^n}]}{Z_t^2} > 0. \quad (5.85)$$

并有

$$\alpha_t^m \leq \frac{\rho_t^m e^{\kappa_F}}{\lambda_\emptyset^F + \rho_t^m e^{\kappa_F}} \rightarrow 0 \quad (\rho_t^m \rightarrow 0). \quad (5.86)$$

这给出“其他条件固定时，可信度越高，门控值越高”的精确定义，不要求任意两模态在内容不同的情况下仍按质量严格排序。该偏导是门控层的条件性质，不是整个网络输入变化时的总导数结论。

5.7.2 避免门控被归一化或旁路抵消

归一化原则上作用于乘性门控之前；不把 $\text{LN}(\rho \mathbf{h})$ 作为保留幅度可靠度的保证，因为忽略数值稳定项时正标量可能被层归一化抵消。门控后的融合及残差不再立即进行逐位置尺度归一化，权重矩阵实施正则和必要范数上界。最终预测不直接拼接未经门控的原特征、完整文本或共享后验均值。缺失结构统计是合法状态信息，但不含隐藏原文和被遮蔽数值。主通路中每份内容须经过可靠度控制；仍通过对照检查历史记忆和非线性交互造成的残余敏感性，不宣称消除了所有不可靠信息影响。 α 是信息使用权重，不是自动成立的因果贡献。

5.8 子模型六：可关闭低秩残差与时间汇聚

令 $\mathbf{z}_t^m = \alpha_t^m \tilde{\mathbf{h}}_t^m$ 。采用低秩多模态融合 [12] 的受控残差形式

$$\mathcal{T}_R(\mathbf{z}^T, \mathbf{z}^A, \mathbf{z}^V) = \sum_{r=1}^R \bigodot_{m \in \mathcal{M}} U_{mr}^\top [1; \mathbf{z}^m]. \quad (5.87)$$

令

$$\begin{aligned} \mathbf{l}_t &= \mathcal{T}_R(\mathbf{z}_t^T, \mathbf{z}_t^A, \mathbf{z}_t^V) - \mathcal{T}_R(\mathbf{0}, \mathbf{0}, \mathbf{0}), \\ \mathbf{f}_t &= \mathbf{g}_t + \eta_L W_L \tanh(\mathbf{l}_t), \quad 0 \leq \eta_L \leq 1. \end{aligned} \quad (5.88)$$

增广常数保留单、双、三模态交互；减去全零输入的常数项，避免所有内容趋零时低秩分支仍制造额外内容残差。初始秩 $R = 4$ ，允许 $\eta_L = 0$ 作为正式候选。网络复杂不等于一定有效，保留该分支需要验证增益。对每条后验轨迹进行时间汇聚：

$$\begin{aligned} a_{it}^{\text{pool}} &= \kappa_P \tanh[\mathbf{v}_P^\top \tanh(W_P \mathbf{f}_{it})], \\ \beta_{it} &= \frac{P_{it} w_{it} \exp(a_{it}^{\text{pool}})}{\sum_s P_{is} w_{is} \exp(a_{is}^{\text{pool}})}, \\ \mathbf{z}_i &= \left[\sum_t \beta_{it} \mathbf{f}_{it}; \text{Stats}(M_i, P_i, \rho_i) \right]. \end{aligned} \quad (5.89)$$

w 的定义与缺失率一致。统计项仅含缺失比例、最大缺口、边界标记和合法可靠度摘要。注意力热图用于展示读取位置，不替代第一问的物理词边界。

5.9 子模型七：中性点质量与有符号幅值分布

5.9.1 分类概率与条件幅值

对第 b 条轨迹输出

$$p_{ic}^{(b)} = \text{softmax}(W_c \mathbf{z}_i^{(b)} + \mathbf{b}_c)_c, \quad c \in \{-1, 0, +1\}. \quad (5.90)$$

为控制小样本异方差估计复杂度，先采用正、负类别内共享集中程度

$$\begin{aligned} m_{ic}^{(b)} &= \sigma(\mathbf{w}_c^\top \mathbf{z}_i^{(b)} + b_c), \quad \kappa_c = \varepsilon_\kappa + \text{softplus}(\xi_c), \\ a_{ic}^{(b)} &= m_{ic}^{(b)} \kappa_c, \quad b_{ic}^{(b)} = (1 - m_{ic}^{(b)}) \kappa_c, \quad c \in \{-, +\}. \end{aligned} \quad (5.91)$$

$a, b > 0$ ；数值上对 m 采用预定内部边界，不改变标签。在训练量充足且验证支持时，才将 κ_c 扩展为带收缩正则的样本相关参数。定义统一随机变量

$$Y_i^{(b)} = \begin{cases} -3U_{i,-}^{(b)}, & C_i = -1, \quad U_{i,-}^{(b)} \sim \text{Beta}(a_{i,-}^{(b)}, b_{i,-}^{(b)}), \\ 0, & C_i = 0, \\ +3U_{i,+}^{(b)}, & C_i = +1, \quad U_{i,+}^{(b)} \sim \text{Beta}(a_{i,+}^{(b)}, b_{i,+}^{(b)}). \end{cases} \quad (5.92)$$

因此 $\Pr(Y = 0) = p_0$ ，零只属于中性。单分量均值及二阶矩为

$$\begin{aligned}\mu_i^{(b)} &= 3p_{i,+}^{(b)}m_{i,+}^{(b)} - 3p_{i,-}^{(b)}m_{i,-}^{(b)}, \\ \mathbb{E}[(Y_i^{(b)})^2] &= 9 \sum_{c \in \{-,+\}} p_{ic}^{(b)} \frac{a_{ic}^{(b)}(a_{ic}^{(b)} + 1)}{(a_{ic}^{(b)} + b_{ic}^{(b)})(a_{ic}^{(b)} + b_{ic}^{(b)} + 1)}.\end{aligned}\quad (5.93)$$

5.9.2 混合预测、温度校准和一致解码

令 $I_u(a, b)$ 为正则化不完全 Beta 函数。联合轨迹平均后

$$\bar{p}_{ic} = \frac{1}{K_{\text{MC}}} \sum_b p_{ic}^{(b)}, \quad G_{ic}(u) = \frac{\sum_b p_{ic}^{(b)} I_u(a_{ic}^{(b)}, b_{ic}^{(b)})}{\sum_b p_{ic}^{(b)}}, \quad c \in \{-, +\}.\quad (5.94)$$

分母正性由有限 softmax 保证，计算使用稳定形式。 G 是条件混合分布，不能把 Beta 参数平均后冒称等价。分类校准作用于最终实际混合概率：

$$p_{ic}^{\text{cal}} = \frac{\bar{p}_{ic}^{1/T_{\text{cal}}}}{\sum_{c'} \bar{p}_{ic'}^{1/T_{\text{cal}}}}, \quad T_{\text{cal}} > 0.\quad (5.95)$$

只用预留校准集或严格折外校准预测估计 T_{cal} ，温度缩放的基本思路见文献 [15]。式 (5.95) 修改类别边际，不重新定义各类别条件混合 G ；由此得到一致的校准联合分布。不能在单次预测上校准后另行平均而不检查最终输出。最终报告

$$\hat{c}_i = \arg \max_c p_{ic}^{\text{cal}}, \quad \hat{y}_i = \begin{cases} -3G_{i,-}^{-1}(1/2), & \hat{c}_i = -1, \\ 0, & \hat{c}_i = 0, \\ +3G_{i,+}^{-1}(1/2), & \hat{c}_i = +1. \end{cases}\quad (5.96)$$

并列时执行预先锁定的中性优先、再固定顺序规则。条件中位数对应条件绝对损失最小化，但先分类再取中位数不等于全混合分布无条件 MAE 最优，这是保证极性与强度一致的明确权衡。验证和正式结果都按式 (5.96) 计算，不只报告通常更平滑的混合均值误差。

5.10 目标函数与分阶段训练

5.10.1 对混合分布的联合监督

对非中性样本取 $u_i = |y_i|/3$ ，以合法标签精度及内层验证确定观测分辨率 $\Delta_u > 0$ ，定义

$$l_i = \max\{0, u_i - \Delta_u\}, \quad h_i = \min\{1, u_i + \Delta_u\}.\quad (5.97)$$

这不是新的中性阈值，而是处理 $y = \pm 3$ 端点及标签观测精度的区间似然。第 b 条轨迹赋予真实标签的概率为

$$Q_i^{(b)} = \begin{cases} p_{i0}^{(b)}, & y_i = 0, \\ p_{i,c_i}^{(b)} [I_{h_i}(a_{i,c_i}^{(b)}, b_{i,c_i}^{(b)}) - I_{l_i}(a_{i,c_i}^{(b)}, b_{i,c_i}^{(b)})], & y_i \neq 0. \end{cases}\quad (5.98)$$

定义 $H_\delta(e) = \frac{1}{2}e^2$ 当 $|e| \leq \delta$, 否则 $H_\delta(e) = \delta(|e| - \delta/2)$ 。主监督为

$$\ell_{\text{sup},i} = -\log \left[\frac{1}{K_{\text{MC}}} \sum_b Q_i^{(b)} \right] + \lambda_y H_\delta \left(\frac{y_i - \bar{\mu}_i}{3} \right), \quad \bar{\mu}_i = \frac{1}{K_{\text{MC}}} \sum_b \mu_i^{(b)}. \quad (5.99)$$

负对数作用在混合概率外, 与式 (5.60) 的推理对象一致。由 Jensen 不等式, 该项不大于逐轨迹负对数的平均; 后者会更强地要求每条可能补全都给出正确判断。有限抽样的对数目标是近似, 并非无偏的精确边缘对数似然。类别重加权不是默认必选, 避免极小中性样本权重破坏概率校准。

5.10.2 完整可用视图教师与有限强度蒸馏

教师使用相同合法训练集未额外遮蔽的视图, 先训练后冻结; “完整”不表示补出了原本不存在的观测。学生只能读取当前缺失视图。借鉴知识蒸馏 [13], 对教师和学生的混合类别概率采用同样的温度变换, 定义

$$\begin{aligned} \ell_{\text{KD},i} &= \omega_i \left\{ T_d^2 \text{KL}(p_i^{\text{tea}, T_d} \| \bar{p}_i^{\text{stu}, T_d}) + \lambda_\mu H_\delta \left(\frac{\mu_i^{\text{tea}} - \bar{\mu}_i^{\text{stu}}}{3} \right) \right\}, \\ \omega_i &= \left[1 - \frac{H(p_i^{\text{tea}})}{\log 3} \right] (r_{i,\text{retain}})^\gamma. \end{aligned} \quad (5.100)$$

$H(p) = -\sum_c p_c \log p_c$; $r_{i,\text{retain}}$ 是各原有可观测模态保留比例的等权平均。教师分支停止梯度; T_d 不等于校准温度。教师不确定或学生证据极少时, 蒸馏减弱, 但该权重不是教师真实正确率。对 $\mathcal{K}_{im} = \{t : O_{it}^m = 1, B_{it}^m = 0\}$, 可用轻量解码器 D_m 对学生隐藏表示进行辅助重建:

$$\ell_{\text{rec},i} = \frac{1}{|\mathcal{M}_i^+|} \sum_{m \in \mathcal{M}_i^+} \frac{1}{|\mathcal{K}_{im}| d_m K_{\text{MC}}} \sum_{t \in \mathcal{K}_{im,b}} \|D_m(\mathbf{h}_{it}^{m,(b)}) - \mathbf{x}_{it}^m\|_1, \quad (5.101)$$

其中 $\mathcal{M}_i^+ = \{m : |\mathcal{K}_{im}| > 0\}$; 集合为空时整项取零。按模态、目标数和维数归一化, 避免文本维数主导。真值仅进入损失目标, 不输入补全器或学生。重建仅辅助情感任务, 不要恢复不可辨识的完整语义。

5.10.3 场景平衡与平滑最坏组风险

以缺失模态组合和粗缺失率构造有限场景组 g , 位置与同步性在组内均衡抽样。避免在小样本上把所有因素交叉成大量空组。定义

$$L_g = \mathbb{E}_{i,B|g} [\ell_{\text{sup},i} + \lambda_d \ell_{\text{KD},i} + \lambda_r \ell_{\text{rec},i}], \quad (5.102)$$

$$\mathcal{L} = (1 - \lambda_G) \sum_g \pi_g L_g + \lambda_G \tau_G \log \left[\sum_g \pi_g \exp(L_g / \tau_G) \right] + \lambda_W \|\theta\|_2^2. \quad (5.103)$$

$\pi_g > 0, \sum_g \pi_g = 1$, 并保留原视图组。第二项等价于

$$\max_{\mathbf{q} \in \Delta_G} \left\{ \sum_g q_g L_g - \tau_G \text{KL}(\mathbf{q} \| \boldsymbol{\pi}) \right\}, \quad q_g^* \propto \pi_g \exp(L_g / \tau_G). \quad (5.104)$$

两种写法是同一目标，不能作为两个模块重复相加。最坏组优化仍需要正则与早停 [14]，不保证每个重缺失样本都预测正确。

5.10.4 求解顺序与关键参数

训练按“分组划分-训练折尺度-动态补全器-未额外缺失教师-缺失学生-校准”的顺序进行。每折独立拟合所有数据驱动参数，不能先用全体样本训练内容投影器或补全器再声称验证无泄漏。补全器和教师在学生训练时冻结，可靠度系数由内层验证固定。人工缺失可由短到长增加，但验证缺失情景从开始就锁定。表中是可复现的候选起点，不是未经实

表 5.7 问题二起始配置与必须验证的选择（非最优参数表）

模块	起始配置	选择和限制
动态潜变量	共享 8、私有各 4	正定噪声与稳定范数约束，训练折估计
模态时序	投影 32，双向各 16	一层 BiGRU，底层通用编码器冻结
跨模态交互	一层，低维共享参数	κ_A, κ_F 初值 2；空来源质量正值
可靠度	ρ_{imp} 初值 0.5	$\lambda_u, \lambda_g, \lambda_b \geq 0$ ，允许关闭冗余项
低秩残差	$R = 4$	$\eta_L \in [0, 1]$ ，零值是正式候选
Beta 幅值	类别内共享集中程度	不默认拟合每样本异方差
后验抽样	训练 4，推理 16	增大抽样数检验预测与分位数稳定性
人工遮蔽	比例 0.1、0.3、0.5、0.7	约 20% 原视图；不抹去整个模态
优化与正则	学习率初值 3×10^{-4}	Dropout 0.2，内层早停与权重衰减
损失权重	$\lambda_y = 1, \lambda_d = 0.1$	$\lambda_r = 0.05, \lambda_G = 0.1$ 仅为起点
损失尺度	$\delta = 0.25, T_d = 2$	强度误差先除以 3； τ_G 内层选定

验确认的最优组合。实际参数、训练范围、随机种子、停止准则和失败情况必须记录。随机补全次数和人工遮蔽次数增加不等于独立有标签样本增加。

5.11 验证设计、缺失规律与错误归因

5.11.1 分组验证、基础指标和不确定性

有官方划分则遵循并核验同源限制；自行设计时按原始视频分组进行外层评价和内层选择。数据足够时采用 5 折外层、3 折内层；类别在来源视频上过度集中时减少折数并报告限制，不把中性重新编码为其他类。校准使用独立保留集或严格折外预测，不与结构选择重复消费同一评估结果。基于最终一致解码报告

$$\begin{aligned}
 \text{MAE} &= \frac{1}{N} \sum_i |\hat{y}_i - y_i|, & \text{RMSE} &= \sqrt{\frac{1}{N} \sum_i (\hat{y}_i - y_i)^2}, \\
 \text{Acc}_3 &= \frac{1}{N} \sum_i \mathbf{1}\{\hat{c}_i = c_i\}, & \text{MacroF1} &= \frac{1}{3} \sum_{c \in \{-1, 0, +1\}} \frac{2P_c R_c}{P_c + R_c}.
 \end{aligned} \tag{5.105}$$

另外报告 Pearson 相关系数、中性召回率、各类支持数、类别 Brier 分数及校准图。零分母与无方差情况按事先约定处理；缺类折不能用任意平滑掩盖支持不足，可优先汇总外层折外预测再报告全体指标。未校准的轨迹混合满足精确的模型内全方差分解

$$\text{Var}(Y \mid X_{\text{obs}}) = \mathbb{E}_b[\text{Var}(Y \mid b, X_{\text{obs}})] + \text{Var}_b[\mu_i^{(b)}]. \quad (5.106)$$

第二项反映当前补全分布下的预测变化，不自动等于全部参数不确定性。校准改变类别边界后，应按校准联合分布重新求矩与分位数，不照搬未校准数值。模型预测区间在验证集检查覆盖率与宽度，不承诺未知测试缺失机制下的无条件覆盖。

5.11.2 缺失类型、位置与持续跨度的受控实验

对所有模型使用同一验证样本与同一遮蔽实例，记录主键、模态、区间、随机种子及当前缺失率口径。至少比较 单缺口中“缺口时长 = 缺失率乘片段时长”，因此仅加长一个缺

表 5.8 局部缺失影响的控制变量设计

因素	比较水平	控制要求
缺失模态	T、A、V 及其组合	相同基础样本与可比较的总缺失量
新增缺失率	0, 0.1, 0.3, 0.5, 0.7	同时报最终总缺失率，短样本记录实际比例
缺失位置	开头、中段、末尾	固定相同跨度，仅平移区间
连续跨度	一长段、多个短段	固定总缺失率，不把两个因素混为一谈
同步程度	同步、部分重叠、交错	固定各模态总缺失率
语义位置	固定规则选定的转折/否定附近	只在合法有原文的验证样本上分析，不反向改训练标签

口不能独立识别缺失率和连续跨度的作用。需在总量相同条件下改变缺口分段结构。片段长度不同还应分层展示，不用单一绝对时长直接覆盖全部样本。定义性能退化及缺失曲线面积

$$\Delta \text{MAE}_g = \text{MAE}_g - \text{MAE}_{\text{original}}, \quad \text{AURC}_{\text{MAE}} = \frac{1}{r_{\max}} \int_0^{r_{\max}} \text{MAE}(r) \, dr. \quad (5.107)$$

按视频成组 bootstrap，并对模型差异使用相同重采样组。多次遮蔽先在样本内汇总，不把每次遮蔽当作新独立受试样本。

5.11.3 可推导性质与待检验规律

固定动态模型参数、噪声尺度和既有质量时，删除观测会从后验精度矩阵移除半正定信息项。因此若 $M' \leq M$ ，则

$$\Lambda(M') \preceq \Lambda(M), \quad \Sigma(M') \succeq \Sigma(M). \quad (5.108)$$

这是固定模型内不确定性的性质，不能推出有限样本训练模型每个样本误差随缺失率严格增加。长区间可能更多依赖先验，同步缺失可能更难跨模态补偿，首尾缺失缺少一侧上下文；但哪个模态、哪个位置损害最大必须实测。中段可能包含关键转折，完整噪声也可能误导模型，不预先规定排序。门控检查分别记录 q^{obs} 、 ρ 、来源权重 A 、融合权重 α 和先验权重。验证式 (5.85) 时固定内容表示做受控扫描；实际缺失实验中表示随缺失变化，不把这两类曲线混为一谈。门控趋势符合预期仍需与任务误差共同解释。

5.11.4 消融方案与错误归因

表 5.9 在同一验证框架下进行的消融

编号	结构	检验目标
C0	观测统计量与掩码的简单分类/回归基线	复杂时序模型的实际增益
C1	掩码 BiGRU，不做概率补全和显式质量门控	时序与掩码基础效果
C2	加概率补全，仅后验均值输入	动态补全作用
C3	联合轨迹积分，只有最终融合门控	不确定性传播和后门控作用
C4	加来源门控与有界得分	提前限制跨模态读取的增益
C5	加可靠度调制 GRU 更新	历史记忆污染控制是否有益
C6	加可关闭低秩残差	明确高阶交互是否提供增量
C7	加蒸馏与分组风险，分别单项对照	两种训练机制各自贡献

另对最终主干分别关闭质量-噪声映射、距离项和辅助重建；把连续遮蔽换成独立点遮蔽，以检验缺失机制匹配。只改变一个因素，参数规模变化单列说明。没有稳定改善的分支允许删去，不能因为称为“融合模型”就必须全部启用。可视化包括缺失率-误差曲线、模态类型与位置热图、混淆矩阵、强度散点图、区间覆盖图和逐位置门控记录。典型误差需回查同一样本的原始证据、遮蔽区间、补全方差、读取权重与最终输出。可分为关键语义不可见、补全过度自信、跨模态冲突、中性混淆及接口/时间映射异常；分类须由实际证据支持，不由单张注意力图推断。

5.12 附件三全量推理、提交与复现记录

5.12.1 混合预测与低信息退化

锁定全部参数后，附件三不再制造人工缺失、不重新拟合尺度、不更新补全器和校准温度。若整条样本合法但没有可信观测，退回训练标签的平滑先验

$$p_c^{\text{prior}} = \frac{N_c + \alpha_0}{N + 3\alpha_0}, \quad \alpha_0 > 0, \quad (5.109)$$

非中性幅值使用训练类别内拟合的先验分布；某类别无训练支持时报告该限制并采用事前固定的弱先验，不从测试标签补足。输出低信息标记，不固定把无信息判为中性。部分局

部无信息则通过动态后验与空专家处理，不直接删除区间。未知维度、版本冲突或无效长度是接口错误，不应伪装为合法低信息样本。保留全部 ID 和错误状态，核验或修复接口后再出正式结果；不能一边阻止错误推理，一边宣称已完成全量数值预测。

5.12.2 正式预测与审计记录分离

官方提交列名及精度以第四条要求为准，不臆定模板。正式文件仅含规定主键、极性和强度；独立审计记录包括总缺失率、连续最大缺口、同步缺失、质量字段可用性、先验权重、类别概率及区间摘要。核验

$$N_{\text{pred}} = N_{\text{test}}, \quad \text{ID}_{\text{pred}} = \text{ID}_{\text{test}}, \quad (5.110)$$

并逐 ID 检查唯一性和状态。非中性强度不能因输出小数位不足变成零；保存足够精度，强度边界与极性逻辑同时检查。附件三无标签，只能展示预测分布、缺失结构和模型不确定性，不能计算其真实 MAE、F1 或准确率。

5.12.3 训练与推理的数学求解流程

输入：合法训练与验证划分、同版本专项特征、固定场景及参数范围。

输出：每条专项样本的一致极性、强度及审计状态。

1. 检查主键、特征接口、原始零标记、有效长度和质量元数据；建立结构掩码，不读取隐藏原文。
2. 在当前训练折计算尺度；估计并核验共享-私有补全器，冻结其参数；训练并冻结未额外缺失教师。
3. 按式 (5.61) 生成连续缺失，同步清理全部辅助视图；只将当前可见值送入后验推断。
4. 按式 (5.68)–(5.70) 联合抽样，按式 (5.74) 计算可靠度。
5. 经 BiGRU、来源门控、融合门控和可选低秩残差得到分布，按式 (5.103) 更新学生。
6. 在固定内层情景上选择结构和停止点；在独立或严格折外部分校准最终混合概率。
7. 用同一组外层缺失实例报告最终解码性能、门控消融与成组不确定性，不改动困难样本定义。
8. 对附件三逐 ID 应用固定变换和联合推理，按式 (5.96) 解码；核验正式文件与审计文件一致性。

6 模型评价

这里是模型评价

参考文献

- [1] Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT, 2019: 4171–4186. DOI: 10.18653/v1/N19-1423.
- [2] McAuliffe M, Socolof M, Mihuc S, Wagner M, Sonderegger M. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. Interspeech, 2017. DOI: 10.21437/Interspeech.2017-1386.
- [3] Baltrušaitis T, Zadeh A, Lim Y C, Morency L P. OpenFace 2.0: Facial Behavior Analysis Toolkit. IEEE International Conference on Automatic Face and Gesture Recognition, 2018.
- [4] OpenFace 项目维护者. Output Format. 官方项目文档: <https://github.com/TadasBaltrušaitis/OpenFace/wiki/Output-Format>. 执行时须保存所用文档修订与工具构建版本。
- [5] Jain S, Wallace B C. Attention is not Explanation. NAACL-HLT, 2019: 3543–3556. DOI: 10.18653/v1/N19-1357.
- [6] Wiegrefe S, Pinter Y. Attention is not not Explanation. EMNLP-IJCNLP, 2019. arXiv:1908.04626.
- [7] van den Oord A, Li Y, Vinyals O. Representation Learning with Contrastive Predictive Coding. 2018. arXiv:1807.03748.
- [8] Ghahramani Z, Hinton G E. Parameter Estimation for Linear Dynamical Systems. Technical Report CRG-TR-96-2, University of Toronto, 1996.
- [9] Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. 2016. arXiv:1606.01865.
- [10] Tsai Y H H, Bai S, Liang P P, Kolter J Z, Morency L P, Salakhutdinov R. Multimodal Transformer for Unaligned Multimodal Language Sequences. ACL, 2019. arXiv:1906.00295.
- [11] Arevalo J, Solorio T, Montes-y-Gómez M, González F A. Gated Multimodal Units for Information Fusion. 2017. arXiv:1702.01992.
- [12] Liu Z, Shen Y, Lakshminarasimhan V B, Liang P P, Zadeh A, Morency L P. Efficient Low-rank Multimodal Fusion with Modality-Specific Factors. ACL, 2018. arXiv:1806.00064.
- [13] Hinton G, Vinyals O, Dean J. Distilling the Knowledge in a Neural Network. 2015. arXiv:1503.02531.
- [14] Sagawa S, Koh P W, Hashimoto T B, Liang P. Distributionally Robust Neural Networks for Group Shifts: On the Importance of Regularization for Worst-Case Generalization. ICLR, 2020. arXiv:1911.08731.

- [15] Guo C, Pleiss G, Sun Y, Weinberger K Q. On Calibration of Modern Neural Networks. ICML, 2017. arXiv:1706.04599.

附录 A MATLAB 源程序

A.1 第 1 问程序

code.m

附录 B Python 源程序

B.1 第 2 问程序

mip1.py
