面向模糊情感预测的大型音频语言模型推理解构

Xiaofeng Yu[1,*], *Jiaheng Dong*[2,*], *Jean Honorio*[2], *Abhirup Ghosh*[3], *Hong Jia*[1], *Ting Dang*[2]

[1] 新西兰奥克兰大学 [2] 澳大利亚墨尔本大学 [3] 英国伯明翰大学 xyu724@aucklanduni.ac.nz, jiahengd@student.unimelb.edu.au

arXiv:2603.08230v1 [cs.SD] 9 Mar 2026

摘要

语音情绪识别在各种应用中发挥着重要作用。然而，现有的大多数方法都只预测单一情感标签，过度简化了人类情感表达本身固有的模糊性。近期的大型音频语言模型展现出生成更丰富输出的潜力，但其对模糊情感进行理解的推理能力仍然有限。在这项工作中，我们将模糊情感识别重新表述为一个分布式推理问题，并首次系统研究了大型音频语言模型中的歧义感知推理。我们的框架包含两个互补部分：一个使预测与人类感知分布相一致的歧义感知目标，以及一种结构化的歧义感知思维链监督，用于引导模型围绕情感线索进行推理。在 IEMOCAP 和 CREMA-D 上进行的实验表明，采用监督微调、直接偏好优化和 GRPO 等训练策略均能持续提升效果。

索引词：语音识别、人机交互、计算副语言学

1. 引言

语音情绪识别（SER）在 [1, 2, 3], 对话智能体 [4],和心理健康应用 [5]中的人机交互方面发挥着重要作用。尽管近期取得了进展，大多数 SER 系统仍然经过训练，只能预测单一的离散情绪类别，这过度简化了人类情感表达的复杂性。在实际情境中，情绪往往具有歧义且相互交织，而非归属于单一类别。计算建模与人类情绪感知之间的这种差异，促使我们需要更具表达力且能感知不确定性的情绪理解范式。

近期的大型音频语言模型（LALMs）展现出令人期待的能力，能够生成超越单标签分类的、更丰富的文本输出 [6, 7]。先前研究表明，LALMs 能够捕捉语音中的多种情感线索，并生成分布式情绪预测 [6]。然而，在高度模糊的情绪情境下，它们的推理能力仍然有限。人类能够自然地在情绪歧义下进行推理，权衡多种线索并形成概率判断 [8, 9], ; 而当前的 LALMs 却难以做到这一点。

提升大型音频语言模型的推理能力日益受到关注，尤其是通过思维链（CoT）提示 [10] 以及基于强化学习（RL）的后训练 [11, 12]。例如，采用思维链监督 [13], 进行监督微调（SFT），或使用偏好优化方法，例如直接偏好优化（DPO） [14] 和强化学习方法，如组相对策略优化（GRPO） [15]，展现出巨大潜力。然而，这些研究主要聚焦于确定性语音理解任务，例如音频问答（AudioQA），此类任务只需选出一个正确答案。相比之下，模糊情绪推理本质上具有分布特性：多种情绪解读可能同时成立，并可表示为软标签或情绪类别上的概率分布（例如，40% 开心、60% 惊讶）。因此，挑战不仅在于提升推理质量，还在于存在不确定性时避免过早收敛到单一的确定性情绪解读。

为弥补这一空白，我们将模糊情感识别重新表述为一个分布式推理问题，并首次对大型音频语言模型中的歧义感知推理进行系统研究。有效的模型必须（i）在决策层面保留情感不确定性；（ii）围绕情绪歧义开展结构化推理，连贯地整合细微且异质的情绪证据。为此，我们的框架提出了两个关键组件：（i）歧义感知目标，通过前向 KL 正则化使预测的情绪分布与人类感知分布对齐，从而防止情感坍塌；（ii）结构化的歧义感知思维链监督，在预测前引导模型整合体现情绪歧义的证据。该框架即插即用，兼容不同的后训练策略，并在监督微调、直接偏好优化和 GRPO 下进行了评估。我们的贡献总结如下：

- 我们首次对大型音频语言模型中的歧义感知推理进行系统研究
- 我们设计了两个互补目标以支持歧义感知学习：歧义感知目标和结构化的歧义感知思维链监督。
- 我们在 IEMOCAP 和 CREMA-D 数据集上，针对多种后训练策略评估了所提出的范式，证明了这两个目标的有效性。

通过将决策层面不确定性建模与推理增强解耦，我们的工作为大型音频语言模型中具备歧义意识的情绪理解提供了新的见解。

2. 相关工作

模糊语音情感识别。早期关于模糊情感识别的研究包括：使用软标签而非类别硬标签来建模情绪歧义 [16], ; 通过多个分类器模拟多位标注者来捕捉歧义 [17], ; 或将模糊情绪视为分布外情况来处理 [18]。随着大型音频语言模型的兴起，出现了多项歧义感知情感识别研究。例如， [6] 通过分析词元级预测分布，考察大型音频语言模型是否隐式编码了情绪歧义。

[*] 这些作者对本文贡献相同。

另一项研究 [7]提出，利用大型音频语言模型生成的合成标注来扩充多标注者标签。然而，这些方法都没有致力于改进大型音频语言模型处理模糊情绪感知的推理过程。**大型音频语言模型的推理能力**。近期关于提升大型音频语言模型推理能力的研究主要分为两类：基于思维链（CoT）的方法和基于强化学习（RL）的方法。基于思维链的方法旨在教会模型逐步推理。例如，Audio-CoT [10] 探索基于提示的思维链诱导，而Audio-Reasoner [13] 则进一步训练模型，使其在生成答案前遵循结构化推理流程。然而，这些方法通常采用为确定性任务（如AudioQA）设计的思维链模式；此类任务旨在得出唯一正确答案，并未显式建模情绪歧义。

基于强化学习的方法则通过奖励驱动的学习来优化推理行为。SARI [11] 和 Sound-Mind [12] 分别使用 GRPO 和 REINFORCE++ [19] 来提升推理能力。然而，这些推理框架仍针对特定任务设计，其目标是识别唯一正确答案。对于歧义情绪推理，哪些后训练目标和策略至关重要，目前仍不清楚。

3. 方法

3.1. 问题表述

我们将情绪理解表述为一项歧义感知推理任务。对于每个语音话语 n，输入是一对多模态数据 $x_n = \{A_n, T_n\}$，其中 A_n 表示声学信号， T_n 表示其转录文本。为捕捉感知的主观性，每个样本都标注有：(i) 各类别上的真实情绪分布 $\mathbf{p}_n^{GT} \in \mathbb{R}^C$ C，以及 (ii) 一条推理轨迹$\mathbf{Z}_n^{GT}$：

$$p_{n,i}^{GT} = \frac{m_{n,i}}{\sum_{j=1}^{C} m_{n,j}}, \quad \mathbf{Z}_n^{GT} = (z_n^{(1)}, \ldots, z_n^{(L)}) \qquad (1)$$

其中 $m_{n,i}$ 表示选择类别i的标注者人数， $z_n^{(\ell)}$表示一个中间推理步骤。与单标签分类不同，$\mathbf{p}_n^{GT}$将群体情绪歧义编码为软分布（例如，70% 开心、30% 惊讶）。轨迹$\mathbf{Z}_n^{GT}$是一种结构化思维链，用于说明如何基于细微的声学和语言学线索进行推理，从而得出存在歧义的情绪分布；它在多模态证据与最终预测之间起到可解释桥梁的作用。

目标是训练一个 LALM f_θ，将输入映射到由预测推理轨迹 $\mathbf{Z}_n$ 和预测情绪分布 $\hat{}\mathbf{p}_n$组成的联合输出空间：$(\hat{}\mathbf{Z}_n,\ \hat{}\mathbf{p}_n) = f_\theta(A_n,\ T_n)$。模型通过两个互补目标进行优化：一方面最小化预测分布与真实分布之间的散度，另一方面鼓励生成的推理步骤遵循真实推理轨迹$\mathbf{Z}_n$。

3.2. 通过歧义感知提示进行思维链整理

为了使 LALM 能够针对歧义情感线索进行推理，我们首先整理一个结构化思维链数据集，为歧义情绪提供明确的推理监督信号。对于每个输入 $x_n = \{A_n, T_n\}$ 及其对应的真实分布 $\mathbf{p}_n^{GT}$，我们使用高性能闭源大型音频语言模型（例如 GPT-4o）合成推理轨迹 $\mathbf{Z}_n^{GT}$。

生成过程遵循结构化协议，以确保轨迹始终有依据且逻辑一致，

表1： 歧义感知思维链整理的综合。

输入： $x_n = \{A_n, T_n\}$，$\mathbf{p}_n^{GT}$
推理步骤关键规则：关键词提取 显著的声学和语言学线索。
步骤 1：文本分析。严格分析文本的语义 含义和语境。识别潜在的歧义。
步骤 2：音频分析。使用专业 术语（音量、语速、音高、音调）描述韵律。指出 支持多数标签和少数标签（若有）的具体线索。 少数标签（如果有）。
步骤 3：综合。综合证据以消除 歧义。　　推理必须足够有力，使得 读者只需阅读你的分析，就能推断出真实标签 ing 你的分析。

如表1所示。随后，使用同一个闭源模型自动验证生成的轨迹，检查解释是否与给定的目标分布一致。

3.3. 歧义感知学习目标

基于结构化推理轨迹$\mathbf{Z}_n^{GT}$和感知分布$\mathbf{p}_n^{GT}$，我们探索如何优化 LALM，以实现歧义感知情绪理解。具体而言，我们提出一种歧义感知目标，旨在使模型输出与人类情绪不确定性保持一致。我们这一方案的一项关键优势在于其“即插即用”的兼容性；它可以无缝集成到各种训练框架中，包括监督微调、直接偏好优化和 GRPO。

3.3.1. 歧义感知目标

目标是使模型预测的情绪分布 $\hat{}\mathbf{p}_n$ 与人类感知分布 $\mathbf{p}_n^{GT}$保持一致。我们采用 Kullback-Leibler（KL）散度作为对齐准则：

$$\mathcal{L}_{dist} = D_{KL}(\hat{\mathbf{p}}_n \parallel \mathbf{p}_n^{GT}) = \sum_{i=1}^{C} p_i^{GT} \log \frac{p_i^{GT}}{\hat{p}_i}, \qquad (2)$$

我们不依赖标准文本生成来获取 $\hat{}\mathbf{p}_n$，而是在最后一步读取情绪类别名称对应的词元级 logits，并对这些 logits 应用 Softmax，以捕捉更细致的情绪状态分布，遵循 [6]。这种基于逻辑值的形式使模型能够表达情绪类别上的分级不确定性，与多标注者的软标签相呼应。

3.3.2. 即插即用学习范式

我们的分布感知目标与具体框架无关，可以无缝集成到不同的训练策略中。

监督微调。采用监督微调进行学习时，模型经过训练，同时最大化推理词元的似然并最小化分布散度。总损失 $\mathcal{L}_{SFT}$是加权组合：

$$\mathcal{L}_{SFT} = \mathcal{L}_{CE}(\hat{\mathbf{Z}}_n, \mathbf{Z}_n^{GT}) + \lambda \mathcal{L}_{dist}(\hat{\mathbf{p}}_n, \mathbf{p}_n^{GT}), \qquad (3)$$

其中， $\mathcal{L}_{CE}$ 是用于自回归生成思维链轨迹的标准交叉熵损失。该目标对推理轨迹和分布进行联合约束。

直接偏好优化。对于直接偏好优化，我们在标准框架中引入歧义感知目标。我们不再依赖静态、离线的

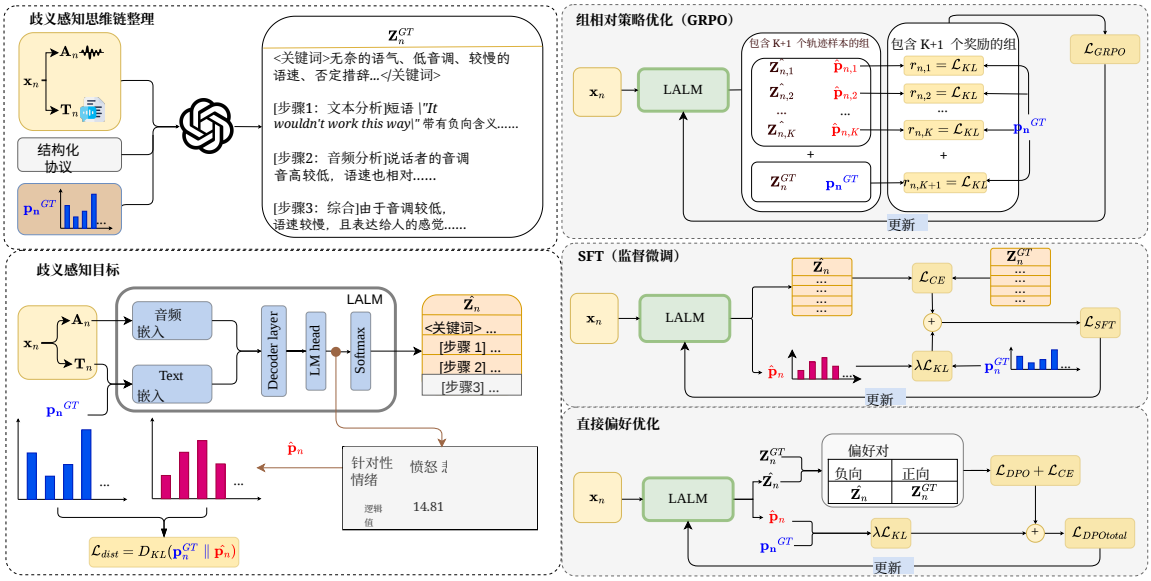

图 1： 所提出的歧义感知目标与即插即用学习范式概览。

偏好对，而是采用同策略方案，由当前策略生成动态 思维链轨迹样本$\mathbf{Z}_n$ 及其情绪分布 $\hat{\mathbf{p}}_n$。预测分布与真实分布偏差较大的轨迹样本被视为负样本，而精选真实思维链则作为正样本。我们通过詹森-香农（JS）散度 $JS(\hat{\mathbf{p}}_n, \mathbf{p}_n^{GT})$ 对此进行衡量；该散度是一种有界的对称度量，能更稳定地识别“较差”的推理路径。随后，直接偏好优化损失会促使模型偏好既遵循黄金推理、又更好地匹配人类感知分布的轨迹，并结合基于 KL 散度的分布损失和思维链交叉熵，形式如下：

$$\mathcal{L}_{\text{DPO-total}} = \mathcal{L}_{\text{DPO}} + \lambda_1 \mathcal{L}_{dist}(\hat{\mathbf{p}}_n, \mathbf{p}_n^{GT}) + \lambda_2 \mathcal{L}_{CE}(\hat{\mathbf{Z}}_n, \mathbf{Z}_n^{GT}), \quad (4)$$

$$\mathcal{L}_{\text{直接偏好优化}} = -\mathbb{E}\Big[\log \sigma\Big(\beta\big(\log \tfrac{\pi_\theta(y_{pos}|x_n)}{\pi_{\text{参考}}(y_{pos}|x_n)} - \log \tfrac{\pi_\theta(y_{neg}|x_n)}{\pi_{\text{参考}}(y_{neg}|x_n)}\big)\Big)\Big], \quad (5)$$

其中 π_θ 表示当前的 LALM，π_{ref} 表示未经任何更新的原始 LALM。

GRPO。 我们还研究了基于强化学习的推理优化方法，采用 GRPO，并同样通过加入歧义感知来增强奖励。

在旧策略 $\pi_{\theta_{\text{old}}}$的每次更新中，我们采样 K 条推理轨迹 $\{\hat{Z}_{n,k}\}_{k=1}^{K} \sim \pi_{\theta_{\text{old}}}(\cdot|x_n)$，并获得相应的情绪分布 $\{\hat{\mathbf{p}}_{n,k}\}_{k=1}^{K}$。每条轨迹都会被赋予一个奖励 $r_{n,k}$：

$$r_{n,k} = -\mathcal{L}_{dist}(\mathbf{p}_n^{GT}, \hat{\mathbf{p}}_{n,k}) + \lambda_3 r_{\text{format}}(\hat{Z}_{n,k}), \quad (6)$$

其中，第一项用于奖励与人类感知分布的准确匹配，第二项则确保遵循规定的思维链格式（定义方式与 [12]类似）。随后，我们将奖励归一化为优势值，以稳定训练：$\hat{A}_{n,k} = \frac{r_{n,k} - \mu_n}{\sigma_n}$，其中 μ_n 和 σ_n 分别表示针对话语 n采样的 K 条轨迹的优势值均值和标准差。

随后使用 GRPO 式目标优化策略：

$$\mathcal{J}_{\text{GRPO}} = \mathbb{E}\Bigg[\frac{1}{K}\sum_{k=1}^{K}\frac{1}{|Z_{n,k}|}\sum_{t}\min\Big(\rho_{n,k,t}(\theta)\hat{A}_{n,k}, \quad (7)$$

$$\text{clip}(\rho_{n,k,t}(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_{n,k}\Big) - \beta \text{KL}(\pi_\theta||\pi_{\text{ref}})\Bigg].$$

其中 $\rho_{n,k,t}(\theta)$ 表示词元级策略概率比率。该目标会提高高质量推理轨迹的可能性，同时抑制低质量推理轨迹。

由于 GRPO 通常依赖采样得到的轨迹，而不直接参考真实推理路径，我们进一步通过在计算优势时将真实轨迹$\mathbf{Z}_n^{GT}$作为额外的参考样本，来增强优势估计，这种方法称为 GRPO$_z$。这确保真实推理轨迹始终获得最高奖励，从而使优化倾向于在存在歧义时生成更忠实的推理路径。

4. 实验设置

数据集。 我们在 IEMOCAP [20]和 CREMA-D [21]上评估训练范式。IEMOCAP 包含四种情绪类别：愤怒、快乐、悲伤和中性，其中“兴奋”被合并到快乐类别中，并采用严格的 5折留一会话交叉验证进行评估。CREMA-D [21] 包含六种类别，新增厌恶和恐惧。我们汇总标注者投票（即 IEMOCAP 中的 3 位标注者和 CREMA-D 中的 4–12 位标注者），并对其进行归一化，以构建软真实标签。

实现细节。 我们使用 GPT-4o 构建结构化的思维链轨迹，因为它在生成连贯的推理链方面能力出众，能够将输入观察与目标输出联系起来 [13]。使用 Qwen2-Audio-7B-Instruct[1] 作为基础 LALM 进行后训练实验。对于所有学习范式，我们均对注意力模块和前馈模块应用 LoRA ($r = 8$, $\alpha_{\text{lora}} = 16$)。所有可训练参数均使用 AdamW 进行优化，学习率分别设为 1e-4（监督微调）、5e-6（直接偏好优化）和 2e-5（GRPO）。训练过程中采用 3% 的线性预热和余弦衰减。直接偏好优化期间用于正则化的相应损失权重为 $\lambda = 1$， $\lambda_1 = 0.1, \lambda_2 = 0.1$和 $\lambda_3 = 0.1$。我们将模型与 1）基础模型和 2）Audio-Reasoner [13]进行比较。

评估指标。 为了评估预测的情绪分布与真实分布的匹配程度，我们报告 Jensen–Shannon散度（JS ↓）和巴塔查里亚

[1]https://huggingface.co/Qwen/Qwen2-Audio-7B-Instruct

表 2： 模糊情感识别性能比较。最佳结果以粗体显示。*GRPO*$_z$ 表示在 *GRPO* 中注入真实值思维链轨迹。

数据集	方法	JS↓	布雷尔-柯蒂斯距离 ↑ R^2↑		布里尔分数↓
IEMOCAP	基础模型	0.40	0.64	0.51	0.15
	Audio-Reasoner	0.36	0.67	0.52	0.15
	SFT	0.29	0.73	0.60	0.12
	DPO	0.23	0.79	0.65	0.09
	GRPO	0.36	0.67	0.55	0.13
	GRPO$_z$	**0.20**	**0.82**	**0.67**	**0.07**
CREMA-D	基础模型	0.25	0.78	0.54	0.05
	Audio-Reasoner	0.37	0.69	0.51	0.09
	SFT	0.21	0.82	0.57	0.04
	DPO	**0.17**	**0.86**	0.67	**0.03**
	GRPO	0.25	0.77	0.46	0.04
	GRPO$_z$	0.20	0.82	**0.68**	0.04

系数（布雷尔-柯蒂斯距离 ↑） [22], , 这些指标用于衡量分布差异和概率质量重叠，以评估模型是否保留了不确定性，而非产生过度自信的预测。

5. 结果

5.1. 后训练策略的性能

我们首先在表 2 中评估所提出的歧义感知训练范式在不同后训练策略下的有效性。在两个数据集上，与基础模型相比，采用所提出的目标函数均能在 SFT、直接偏好优化和 GRPO$_Z$ 下稳定提升性能。

其中，SFT 的表现始终不及 GRPO$_Z$ 和直接偏好优化，这表明，歧义感知情感建模得益于在多条推理轨迹上进行学习，而非依赖单一的监督推理路径。值得注意的是，GRPO$_Z$ 在 IEMOCAP 的所有指标上均取得最佳表现，而直接偏好优化在 CREMA-D 上表现最佳。这一差异表明，后训练策略的效果取决于情感分布空间的结构复杂性。随着 CREMA-D 中情绪类别数量的增加，预测分布对中间推理步骤中的细微变化变得更加敏感。在这类高维歧义场景中，GRPO$_Z$ 使用的结果级奖励信号主要依赖最终决策阶段预测情绪分布与目标情感分布之间的散度，因此可能无法精确指导细粒度推理。相比之下，直接偏好优化等基于偏好的方法通过在词元级直接对比正负推理链，提供了更密集的监督，从而为朝向目标分布的中间轨迹提供更强的监督。

5.2. KL散度的影响

我们通过比较仅使用交叉熵（CE）训练的变体，分析基于 KL散度的分布监督的影响
图2中那些采用KL散度进行情绪分布监督的方法。

在两个数据集上，引入KL监督相较于仅交叉熵训练均能持续提升分布级指标。在IEMOCAP上，加入KL 正则化后，监督微调与直接偏好优化两种变体相对于各自的仅交叉熵对应模型，JS散度均有所降低，布雷尔-柯蒂斯距离均有所提高。类似的改进

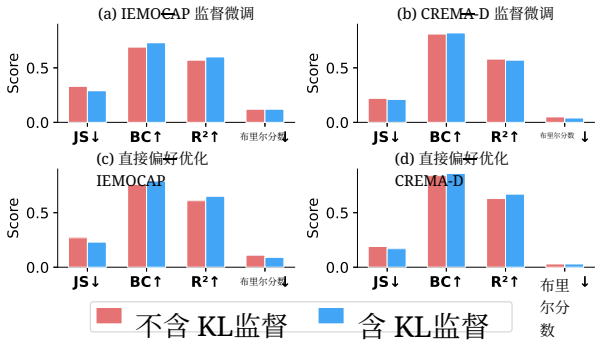

图2： 仅交叉熵（不使用*KL*监督）与本文提出的额外*KL*监督方法之间的比较。*(a)* 在*IEMOCAP*上进行监督微调，*(b)* 在*CREMA-D*上进行监督微调，*(c)* 在*IEMOCAP*上进行直接偏好优化，以及*(d)* 在*CREMA-D*上进行直接偏好优化。

表 3： 监督微调训练目标的消融实验。模型在*CREMA-D*上训练，并在域内（*CREMA-D*）和跨域（*IEMOCAP*）设置下进行评估。

设置	CoT	KL	JS↓	BC↑	R^2 ↑	布里尔分数↓
域内	✓	✓	0.21	0.82	0.57	0.04
域内	✗	✓	0.22	0.81	0.57	0.04
跨域	✓	✓	0.38	0.65	0.50	0.15
跨域	✗	✓	0.52	0.52	0.31	0.22

在CREMA-D上同样可以观察到。这些一致的提升表明，尽管基于交叉熵的训练会促使模型复现经过整理的推理轨迹和最终预测词元，但它在显式约束预测概率质量以匹配软标签情绪分布方面仅提供了较弱的监督。相比之下，KL散度直接惩罚分布不匹配，从而提高了与歧义的一致性。

5.3 歧义感知思维链监督的影响

我们在表 3 中分析了歧义感知思维链监督的影响。比较采用思维链监督与未采用思维链监督进行监督微调（SFT）后的表现，可以发现模型在域内和跨域评估中呈现出不同的行为。在 CREMA-D 上训练和测试时，加入思维链监督相比仅使用 KL 正则化训练，性能仅略有提升。然而，在 CREMA-D 上训练、在 IEMOCAP 上测试时，采用思维链监督训练的模型显著优于未采用该监督训练的模型。这一观察表明，如果不使用思维链，仅通过 KL 正则化优化最终的情绪分布，很容易过拟合数据集特有的分布模式。相比之下，歧义感知思维链监督促使模型基于多模态情绪线索进行推理，从而提升泛化能力。

6. 结论

本研究首次系统探讨了大型音频语言模型在存在歧义的语音情绪识别任务中的歧义感知推理。我们将存在歧义的情绪理解重新表述为一个分布式推理问题，并提出一种即插即用框架，将歧义感知目标与结构化思维链监督相结合。采用我们训练范式的不同后训练策略（如监督微调、直接偏好优化和 GRPO）在 IEMOCAP 和 CREMA-D 上均持续提升了性能。

7. 生成式人工智能使用披露

生成式人工智能工具仅用于少量语言编辑和润色，以提升稿件的清晰度和可读性。这些工具未用于生成科学内容、分析实验结果、总结相关工作、制定方法论或提出研究想法。本文所述的所有概念性贡献、实验设计、分析和结论均由作者提出并验证。作者对本稿件的内容承担全部责任。

8. 参考文献

[1] L. Cai、J. Dong 和 M. Wei，《基于深度学习的语音与面部表情多模态情感识别》，载于*2020*年中国自动化大会（*CAC*）。IEEE，2020年，页码 5726–5729。[2] G. Wadley、V. Kostakos、P. Koval、W. Smith、S. Webber、A. Cox、J. J. Gross、K. H¨o¨ok、R. Mandryk 和 P. Slov´ak，《人机交互中的情感之未来》，载于*CHI* 人机交互系统中的人为因素会议扩展摘要，2022年，页码 1–6。[3] C. Singla、S. Singh、P. S harma、N. Mittal 和 F. Gared，《利用高层描述符实现人机交互中的情感识别》，《*Scientific Reports*》，第 14 卷，第 1 期，第 12122 页，2024年。[4] E. K. Zadeh 和 M. Alaeifard，《结合混合深度学习模型与上下文感知的实时语音情感分析，实现自适应虚拟助手交互》，《*International Journal of Advanced Human Computer Interaction*》，第 1 卷，15，2023 1，1第。页码 –。[5] A. Sharma、K. Sharma 和 A. Kumar，《利用微调迁移网络与多模态融合进行实时情绪健康检测》，《神经计算与应用》，第 35 卷，第 31 期，页码 22 935–22 948，2023年。[6] J. V. Halim、S. Wang、H. Jia 和 T. Dang，《词元级 logits 至关重要：深入考察用于模糊情感识别的语音基础模型》，*INTERSPEECH*，2025年。[7] W. Zhang、H. Jin、S. Wang、Z. Wei 和 T. Dang，《扩展歧义：利用音频语言模型增强语音情绪识别中的人工标注》，*arXiv* 预印本 *arXiv:2601 14620*，2026年。。[8] K. R. Scherer，《情感的声音交流：研究范式综述》，《*Speech Communication*》，第 40 卷，第 1-2 期，页码 227–256，2003年。[9] A. Schirmer 和 R. Adolphs，《通过面部、声音和触觉感知情感：比较与趋同》，《认知科学趋势》，第 21 卷，第 3 期，页码 216–228，2017年。[10] Z. Ma、Z. Chen、Y. Wang、E. S. Chng 和 X. Chen，《Audio-CoT：探索大型音频语言模型中的思维链推理》，*arXiv* 预印本 *arXiv:2501.07246*，2025年。[11] C. Wen、T. Guo、S. Zhao、W. Zou 和 X. Li，《SARI：通过课程引导强化学习实现结构化音频推理》，*arXiv* 预印本 *arXiv:2504.15900*，2025年。[12] X. Diao、C. Zhang、K. Kong、W. Wu、C. Ma、Z. Ouyang、P. Qing、S. Vosoughi 和 J. Gui，《SoundMind：面向音频语言模型的强化学习激励逻辑推理》，载于《*2025*年自然语言处理经验方法会议论文集》，2025年，页码 528–540。[13] X. Zhifei、M. Lin、Z. Liu、P. Wu、S. Yan 和 C. Miao，《Audio-Reasoner：提升大型音频语言模型的推理能力》，载于《*2025*年自然语言处理经验方法会议论文集》，2025年，页码 23 840– 23 862。[14] R. Rafailov、A. Sharma、E. Mitchell、C. D. Manning、S. Ermon 和 C. Finn，《直接偏好优化：你的语言模型其实暗中充当着奖励模型》，《神经信息处理系统进展》，第 36 卷，页码 53 728–53 741，2023年。[15] Z. Shao、P. Wang、Q. Zhu、R. Xu、J. Song、X. Bi、H. Zhang、M. Zhang、Y. Li、Y. Wu 等，《DeepSeekMath：推动开放语言模型数学推理能力的极限，2024》，网址 *https://arxiv. org/abs/2402.03300*，第 2 卷，第 3 期，第 5 页，2024年。[16] E. Mower、A. Metallinou、C.-C. Lee、A. Kazemzadeh、C. Busso、S. Lee 和 S. Narayanan，《解读歧义情感表达》，载于*2009*年第三届情感计算与智能交互国际会议及研讨会。IEEE，2009年，页码 1–8。[17] Y. Zhou、X. Liang、Y. Gu、Y. Yin 和 L. Yao，《面向模糊语音情感识别的多分类器交互学习》，*IEEE/ACM* 音频、语音与语言处理汇刊，第 30 卷，页码 695–705，2022年。[18] W. Wu、B. Li、C. Zhang、C.-C. Chiu、Q. Li、J. Bai、T. Sainath 和 P. Woodland，《处理情感歧义：从域外检测到分布估计》，载于计算语言学协会第 *62* 届年会论文集（第一卷：长篇论文），2024年，页码 2078–2093。[19] J. Hu、J. K. Liu、H. Xu 和 W. Shen，《Reinforce++：利用全局优势归一化稳定无评论家策略优化》，*arXiv* 预印本 *arXiv:2501.03262*，2025年。[20] C. Busso、M. Bulut、C.-C. Lee、A. Kazemzadeh、E. Mower、S. Kim、J. N. Chang、S. Lee 和 S. S. Narayanan，《IEMOCAP：交互式情感双人动作捕捉数据库》，语言资源与评价，第 42 卷，页码 335–359，2008年。[21] H. Cao、D. G. Cooper、M. K. Keutmann、R. C. Gur、A. Nenkova 和 R. Verma，《CREMA-D：众包情感多模态演员数据集》，*IEEE* 情感计算汇刊，第 5 卷，第 4 期，页码 377–390，2014年。[22] X. Hong、Y. Gong、V. Sethu 和 T. Dang，《AER-LLM：利用大语言模型实现感知歧义的情感识别》，载于 *ICASSP 2025-2025 IEEE* 声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2025年，页码 1–5。