Beijing Forest Studio

北京理工大学信息系统及安全对抗实验中心

缺失模态的情绪变化识别

硕士研究生 杨桢弘

2025年12月28日

问题回溯

- 总结反思
  - 实验讲解不清晰
  - 算法部分讲解不够深入
- 相关内容
  - 2024.11.10 杨桢弘《深度学习语音情绪识别技术》
  - 2023.06.11 李新帅《基于Transformer的时间序列分析》

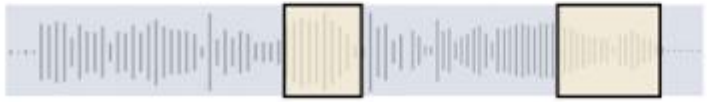
内容提要

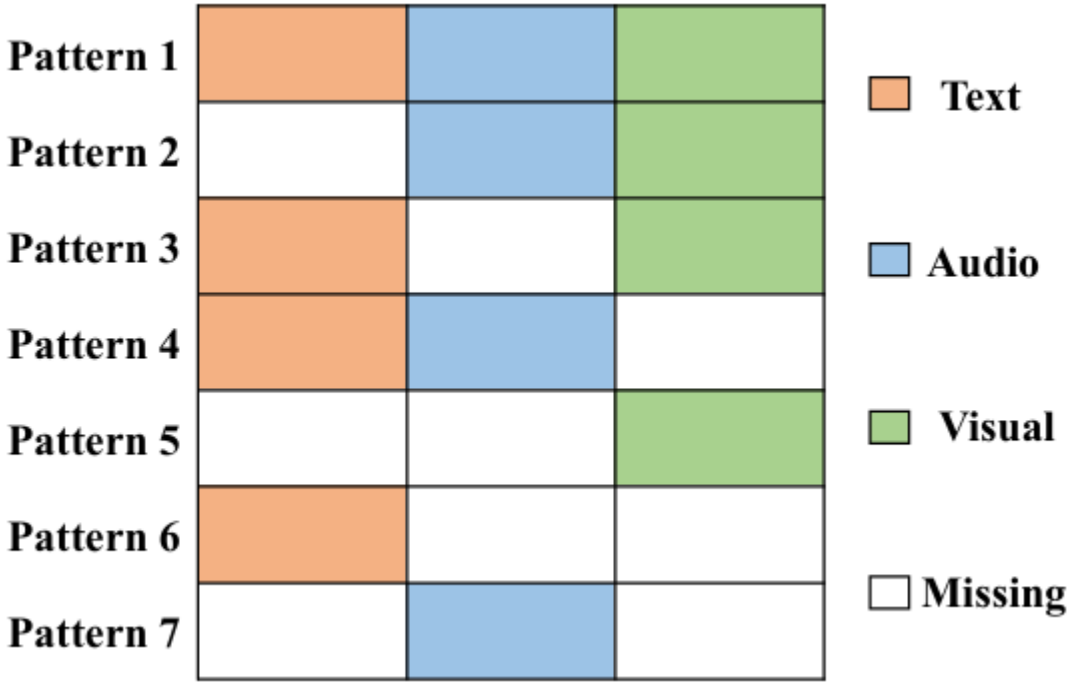
- 预期收获
- 内容引入
- 内涵解析与研究目标
- 研究背景与意义
- 研究历史与现状
- 知识基础
- 算法原理
  - MoMKE
  - HARDY-MER
- 特点总结与未来展望
- 参考文献

预期收获

- 预期收获
  - 掌握缺失模态情绪变化识别的研究现状与基本概念
  - 理解缺失模态情绪变化识别的基本模型及其原理
  - 了解缺失模态情绪变化识别未来发展方向

内容引入

Modality	Content	True Label	Predict
Visual		Positive	Positive ✓
Audio			
Text	I mean they had a great football season last year.		
Visual		Positive	Neutral ✗
Audio			
Text	I mean they had a great football season last year.		



缺失模态场景

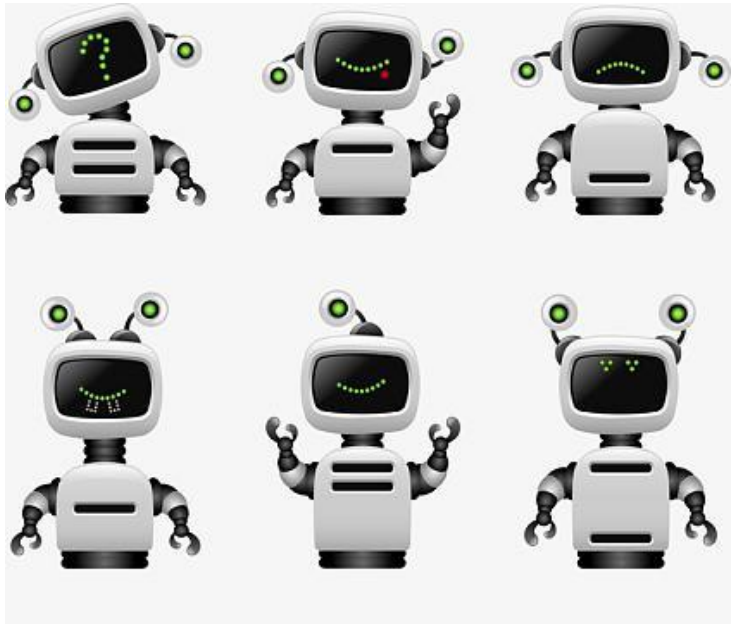
内涵解析与研究目标

- 内涵解析
  - 语音情绪识别
    - 语音→情绪类别/强度
  - 多模态情绪变化识别
    - 语音+文本+视频→情绪变化序列/情绪对
  - 缺失模态的情绪变化识别
    - 语音/文本/视频组合→情绪变化序列/情绪对

- 研究目标
  - 结合深度学习、图神经网络等技术
  - 在模态信息不完整、不对齐或质量受损的现实条件下，构建具有稳定判别能力、良好泛化性和训练鲁棒性的多模态情绪识别模型。

研究背景与意义

- 研究背景
  - 多模态情绪识别，相比单模态方法能够有效缓解噪声、歧义和模态偏置问题，因此在情感计算、人机交互、对话系统等场景中具有显著优势
  - 在真实场景中，不同模态的数据常常存在缺失、噪声或不同步等问题，缺失模态不仅影响单轮情绪识别，还会扰乱情绪变化的连续建模与上下文传播
- 研究意义
  - 在缺失模态学习与对话情绪变化识别等多个领域中具有广泛的应用
  - 能够在部分模态缺失时仍准确捕捉情绪变化趋势的模型，提升模型在真实场景中的可靠性



研究历史与现状

2020 Zhang等人提出了CPM-Nets，通过学习可重建各视图的统一潜在表示并引入结构化分类与对抗补全机制，实现对复杂视图缺失模式的鲁棒多视图学习

2021 Zhao等人提出了MMIN，通过跨模态想象机制联合级联残差自编码器与循环一致性学习，在不确定模态缺失条件下学习鲁棒的联合多模态表示，从而提升多模态情绪识别性能

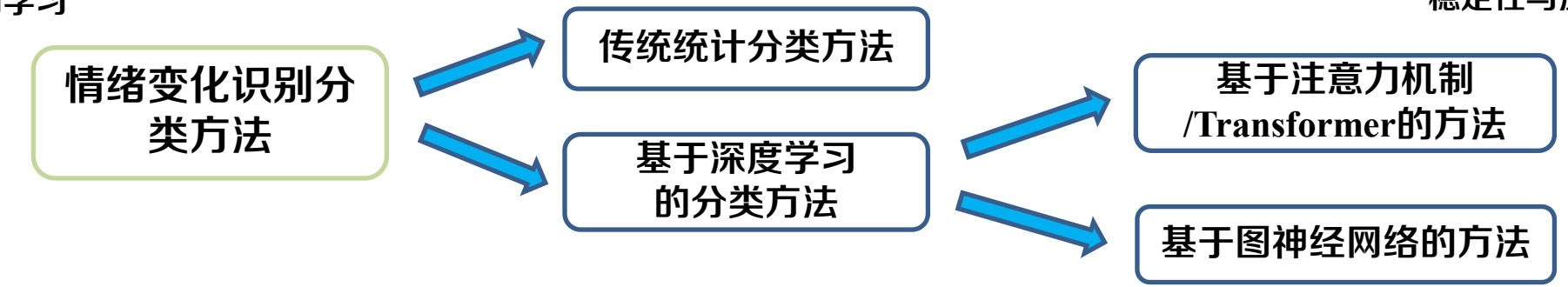
2023 Lian等人提出了GCNet，通过构建包含说话人关系与时间依赖的对话图结构，并联合分类与重建任务，提升多模态对话情绪识别性能

2024 Fu等人提出了SDR-GNN，通过构建情感交互图并在谱域融合高低频信息，实现缺失模态条件下对话情绪的有效重建与识别

2024 Zhao等人提出了MoMKE，通过构建单模态专家并引入动态专家混合与路由机制，在模态缺失条件下实现单模态知识与联合知识的自适应融合

2025 Shou等人提出了GSDNet，通过在图谱空间中引入扩散建模，仅对邻接矩阵特征值进行噪声建模，从而在保持图结构语义的同时实现高质量缺失模态恢复

2025 Li等人提出了HARDY-MER，引入多视角样本难度评估与检索驱动的动态课程学习策略，通过难度感知的样本组织与训练调度提升多模态情绪识别的稳定性与泛化能力



研究历史与现状

- 现有存在问题
  - 单模态知识与联合知识如何协同，尚缺乏统一有效建模方式
  - 训练策略普遍对样本难度、语义复杂性不敏感
  - 真实场景中复杂缺失模式难以覆盖
- 现有的常见方法
  - 基于Transfomer架构大模型的方法
    - 具备强大的全局注意力建模能力，能够捕捉长距离语义依赖
    - 适用于多轮长对话建模，能够刻画上下文级情绪演化信息
  - 基于图神经网络的方法
    - 利用图结构可建模跨轮、跨说话人的高阶情绪依赖关系
    - 支持多模态节点特征融合，能够灵活引入不同模态信息

知识基础

- 模式(Pattern)
  - 对客体（研究对象）特征的描述（定量的或结构的描述），是取自客观世界的某一样本的测量值的集合（或综合）。
  - 例如，对一只猫的文字描述：这只猫身上的毛黑白相间

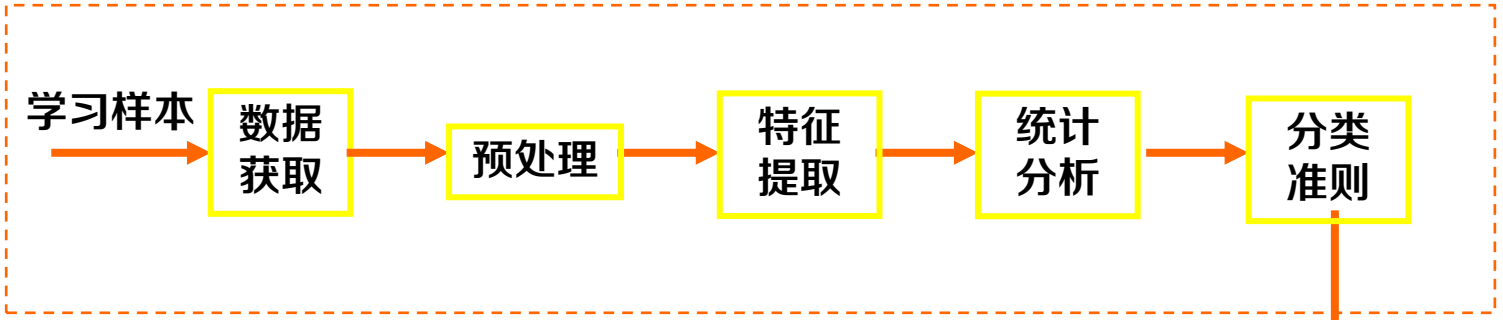
- 样本（Sample）
  - 一个具体的研究（客观）对象。例如，某人画的一幅图片等。
- 模式识别(Pattern Recognition)
  - 确定一个样本的模式类【具有共同特性的模式集合】过程，把某一样本归属于多个类型中的某个类型。
  - 例如，某种类型的动物、某种交通工具、某种移动设备

模，法也；式，法也

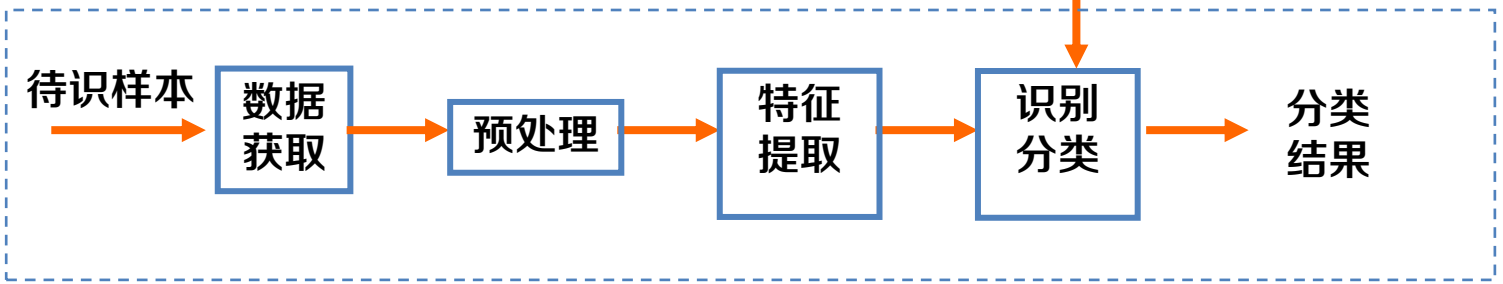
通过一系列的数学方法让机器来实现类似人的识别能力

- 模式识别的过程

学习过程：

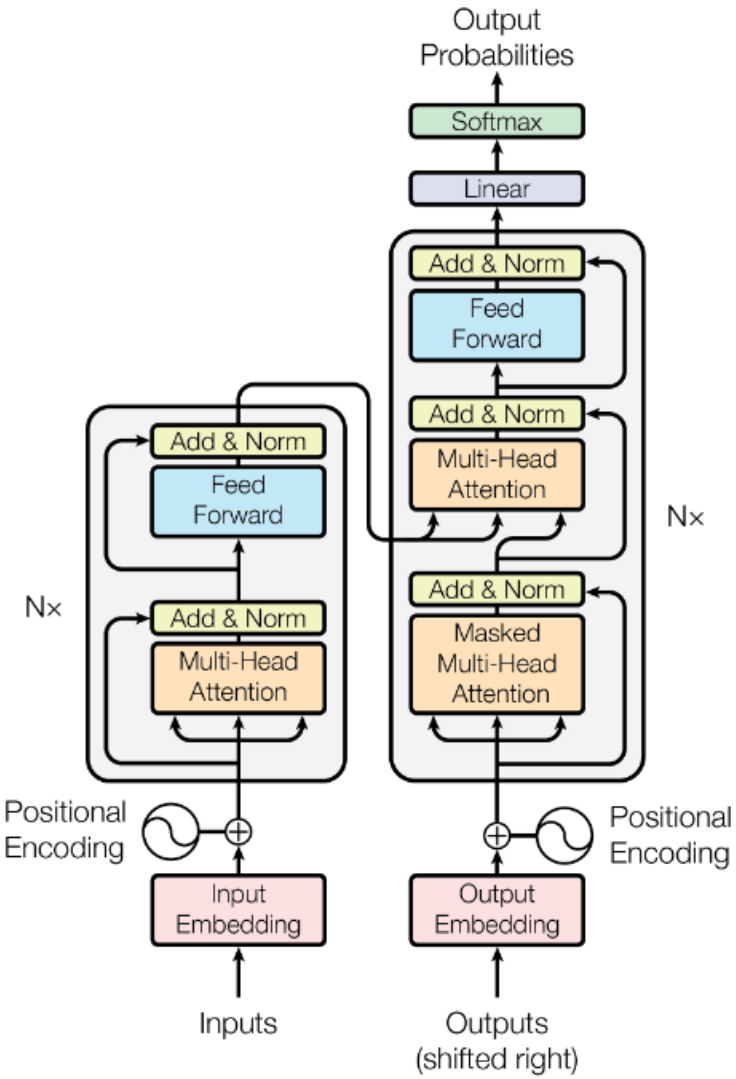


应用过程：



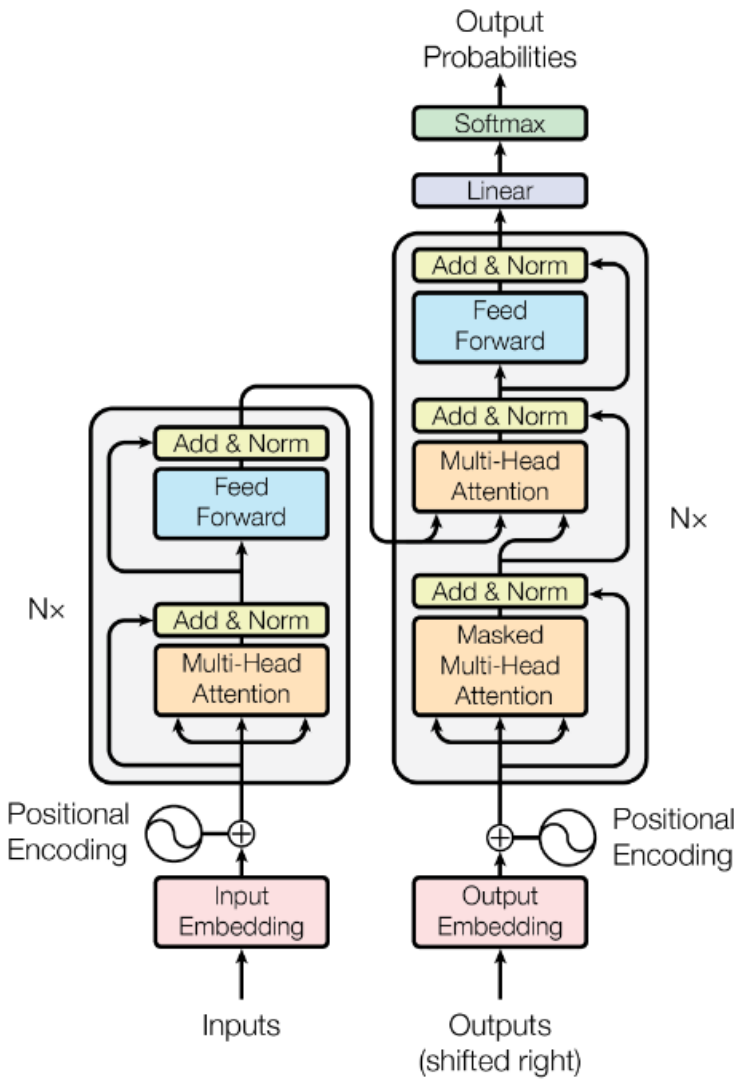
T	目标	获取样本分类结果
I	输入	学习样本与待识样本
P	处理	1. 数据预处理 2. 特征提取 3. 分类准则 4. 训练识别
O	输出	样本分类结果

- Transfomer架构
  - 输入、编码器、解码器、输出
- 重要组成部分和特点：
  - 自注意力机制
  - 多头注意力：自注意力机制被扩展为多个注意力头，每个头可以学习不同的注意权重，以更好地捕捉不同类型的关系
  - 堆叠层：有助于模型学习复杂的特征表示和语义
  - 位置编码
  - 残差连接和层归一化：有助于减轻训练过程中的梯度消失和爆炸问题，使模型更容易训练
  - 编码器、解码器



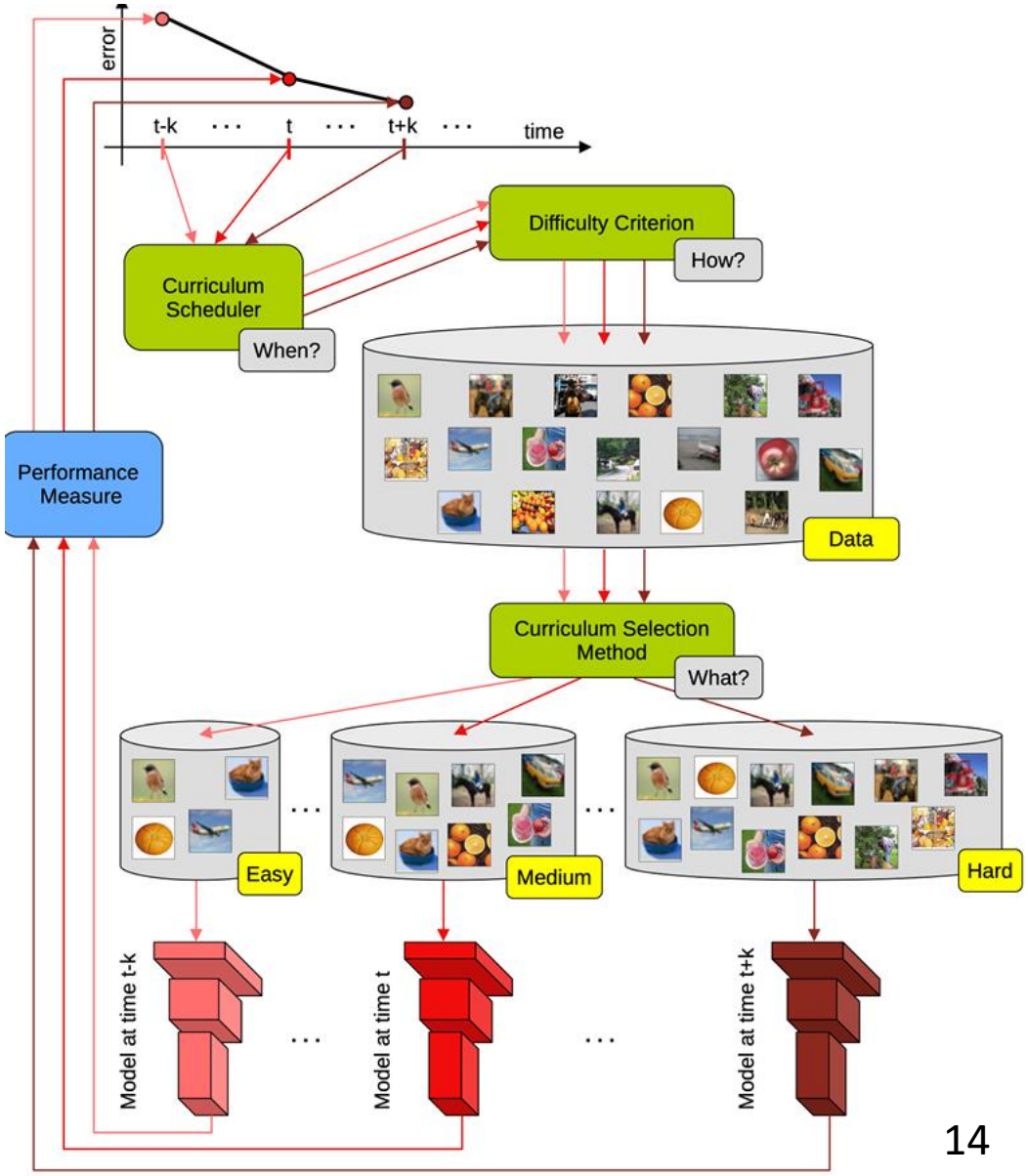
知识基础

- 编码器：对输入序列进行深度分析与理解，提取出富含上下文信息的特征表示
  - 子层1：多头自注意力机制
  - 子层2：前馈神经网络
  - 编码器的自注意力机制是双向的，序列中的每个词都可以看到并关注到输入序列中所有位置的词
- 解码器：基于编码器所提供的“深度理解结果”，并结合已经生成出来的目标序列，自回归地逐词生成出完整的目标序列
  - 子层1：掩码多头自注意力
  - 子层2：编码器-解码器注意力，又称交叉注意力
  - 子层3：前馈神经网络



知识基础

- 课程学习
  - 样本难度评估：根据一定标准对训练样本进行“难易”度量
    - 损失大小
    - 预测置信度
    - 重构误差
    - 语义一致性程度
  - 课程构建：根据样本难度对训练数据进行排序或分组，形成学习路径
    - 从易到难逐步引入
    - 分阶段开放更困难样本
  - 课程调度：控制不同难度样本在训练过程中的出现频率或权重，动态调整



- 缺失模态情绪变化识别的评价指标
  - 非加权准确率（UA）：

$$UA = \frac{\text{预测正确的样本数}}{\text{样本总数}}$$

  - 加权准确率（WA）：

$$WA = \frac{1}{C}\sum_{i=1}^{C}\frac{\text{第}i\text{类预测正确数}}{\text{第}i\text{类样本数}}$$

  - F1值：

$$F1 = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Leveraging Knowledge of Modality Experts for Incomplete Multimodal Learning

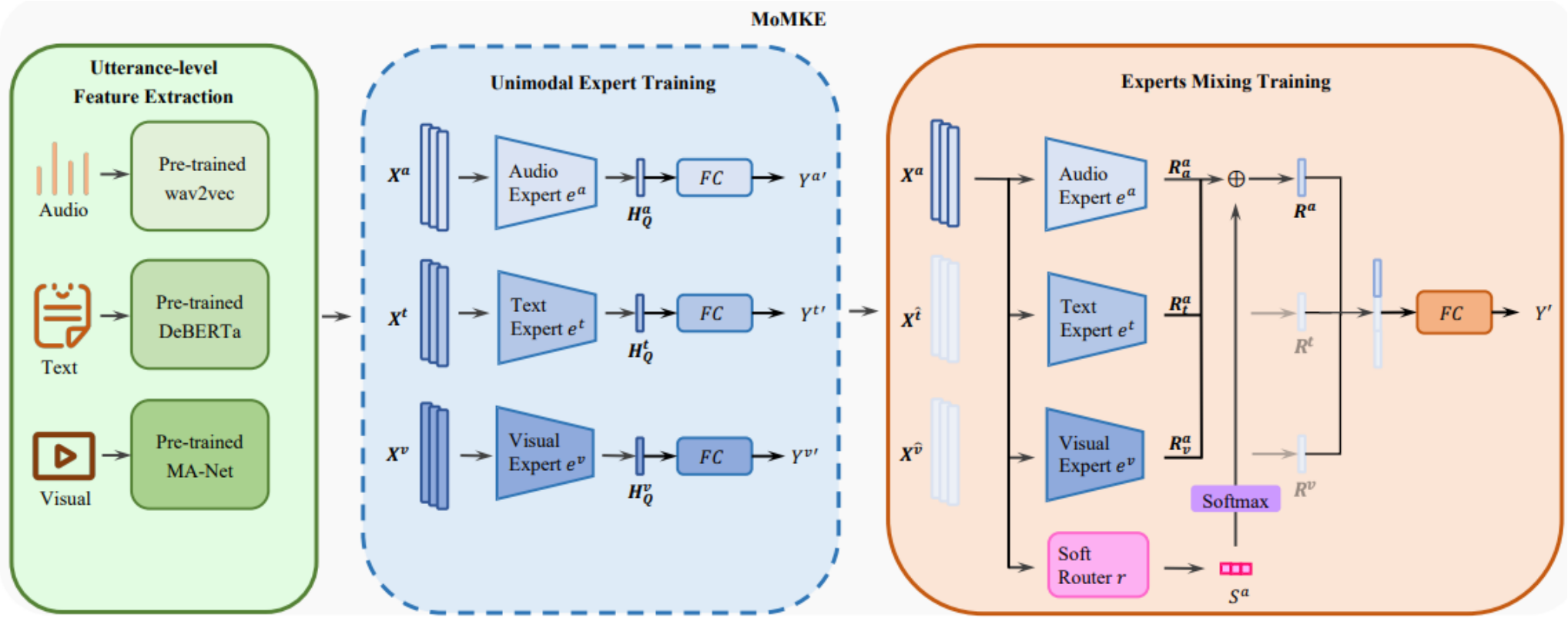
MoMKE TIPO

T	目标	提升严重缺失模态条件下的情绪识别性能，避免单模态判别信息的削弱
I	输入	3个数据集（IEMOCAP4类、CMUMOSEI、CMUMOSI数据集）
P	处理	1.单模态专家训练 2.专家混合训练 3.动态融合机制 4.预测情绪
O	输出	情绪类别（2/4类）

P	问题	1.现有方法过度依赖联合表示，忽视单模态本身的判别能力 2.现有方法在严重缺失条件下，模型性能显著退化
C	条件	需要预训练的特征提取模型
D	难点	1.如何在缺失模态条件下同时建模单模态与联合知识 2.如何根据不同输入场景动态调整不同模态知识的贡献
L	水平	2024 CCF A类

MoMKE 算法原理

- 算法原理图
  - 单模态专家训练
  - 专家混合训练
  - 动态融合机制



MoMKE 创新点分析

- 现有方法存在问题
  - 现有方法过度依赖联合表示，忽视单模态本身的判别能力

- MoMKE的解决方法
  - 为每个模态分别构建一个模态知识专家；每个专家通过独立的编码器和全连接层，学习该模态自身的情绪判别知识；获得高判别性的单模态表示，避免信息被联合空间过早稀释

- 解决方法 单模态专家训练
  - 构建模态知识专家
    - 对任意一个模态 $m \in \{a, t, v\}$中的话语在该模态下都有一个特征向量：

$$X_m = \{x_i^m\}_{i=1}^L \in R^{L*d_m}$$

    - 投影到固定维度：

$$Z_m = X_m W_m$$

    - 加位置编码：

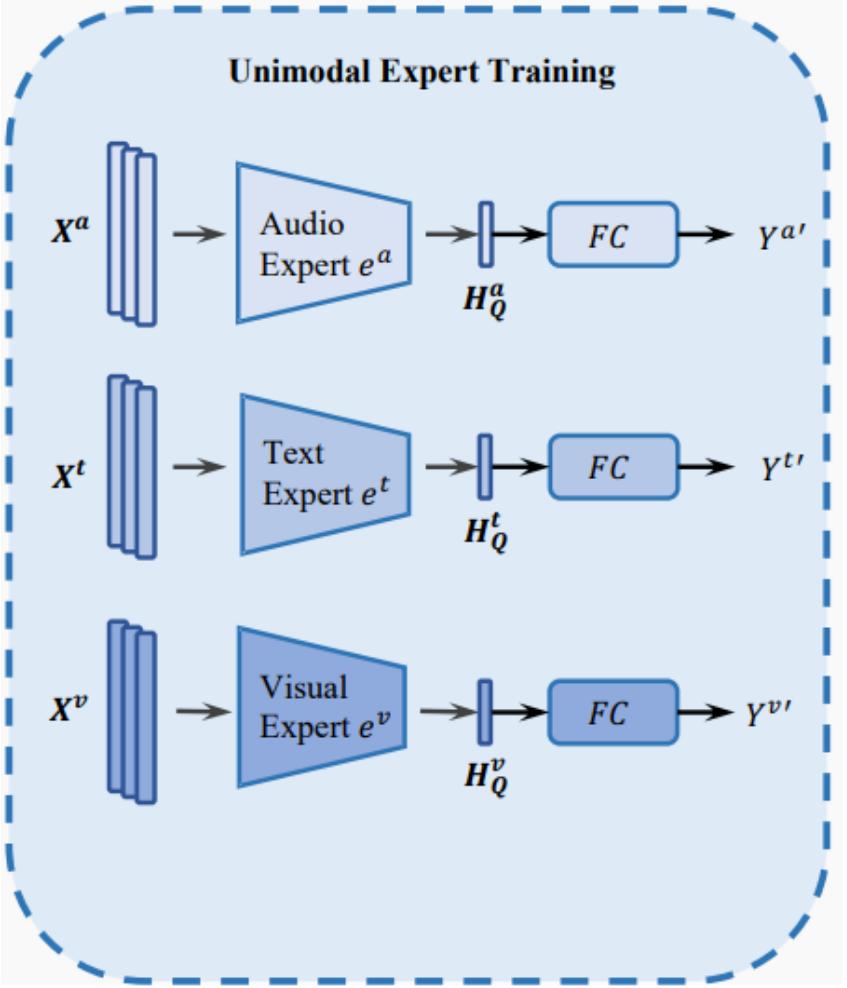
$$H_0^m = Z_m + Z_{pos}$$

    - 专家编码器编码：

$$H_j^{m\prime} = MSA(LN(H_{j-1}^m)) + H_{j-1}^m$$

$$H_j^m = FFN(LN(H_j^{m\prime})) + H_j^{m\prime}$$

其中，MSA是多头子注意力，FFN是前馈网络



- 解决方法 单模态专家训练
  - 构建模态知识专家
    - 输出预测结果：

$$H_Q^m = Transformer_{\theta_m}^T(X_m)$$

$$Y'_m = FC_{\theta_m}^{FC}(H_Q^m)$$

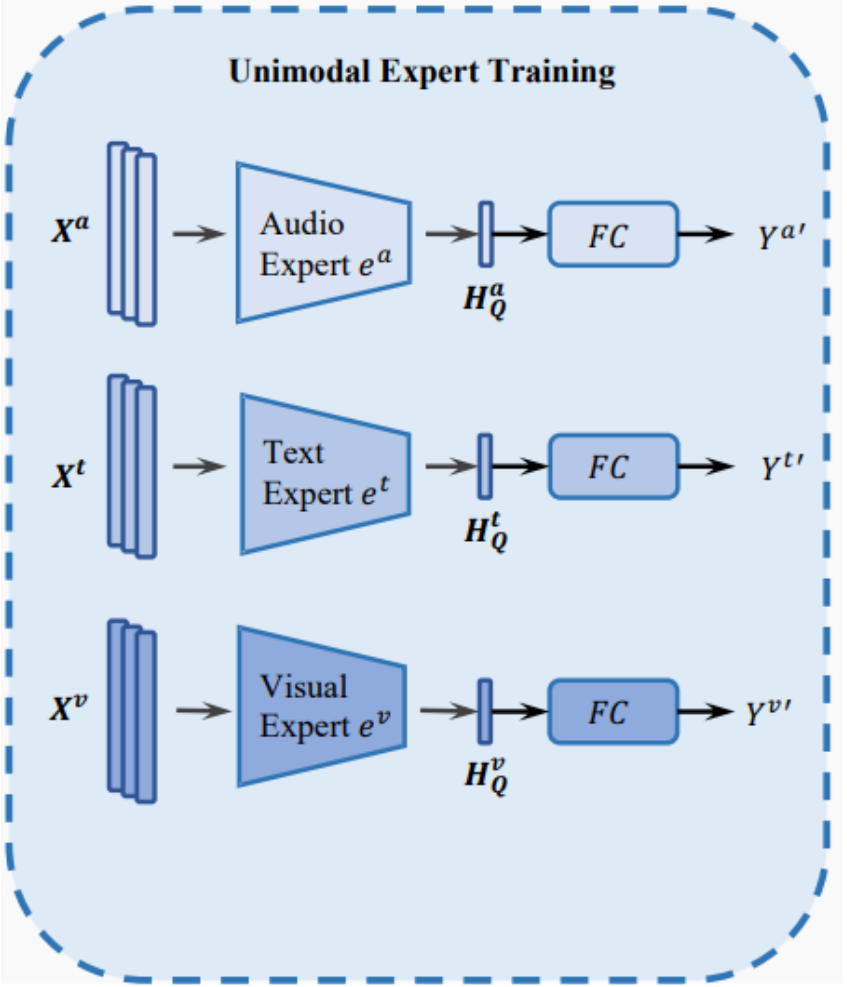
    - 损失函数：

$$l_{task} = CrossEntropy(Y, Y'_m)$$

$$l_{task} = MSE(Y, Y'_m)$$

      多分类用交叉熵，回归用MSE
    - 训练结果（专家）：

$$Transformer_{\theta_m}^T(\cdot) = e_m(\cdot)$$



MoMKE 创新点分析

- 现有方法存在问题
  - 现有方法在单模态推理上无法利用跨模态知识，联合表示强，但一旦缺失就崩溃
  - 现有方法使用固定的模态融合策略在不同缺失场景下无法适应，低质量或噪声模态干扰下导致模型预测出现偏差

- MoMKE的解决方法
  - 专家混合训练，在缺失模态条件下，将可用模态特征同时输入所有模态专家信息，既通过其对应的单模态专家获得单模态表示；也通过其他模态专家获得联合表示
  - 动态路由与表示结合，根据输入模态条件动态生成融合权重

MoMKE 创新点分析

- 解决方法 专家混合训练
  - 同一个输入X_a，分别过三个专家，得到三套表示

$$R_a^a = e_a(X_a)$$
$$R_a^t = e_t(X_a)$$
$$R_a^v = e_v(X_a)$$

  - 动态路由+表示融合
    - 输出每个专家的“重要性分数”：

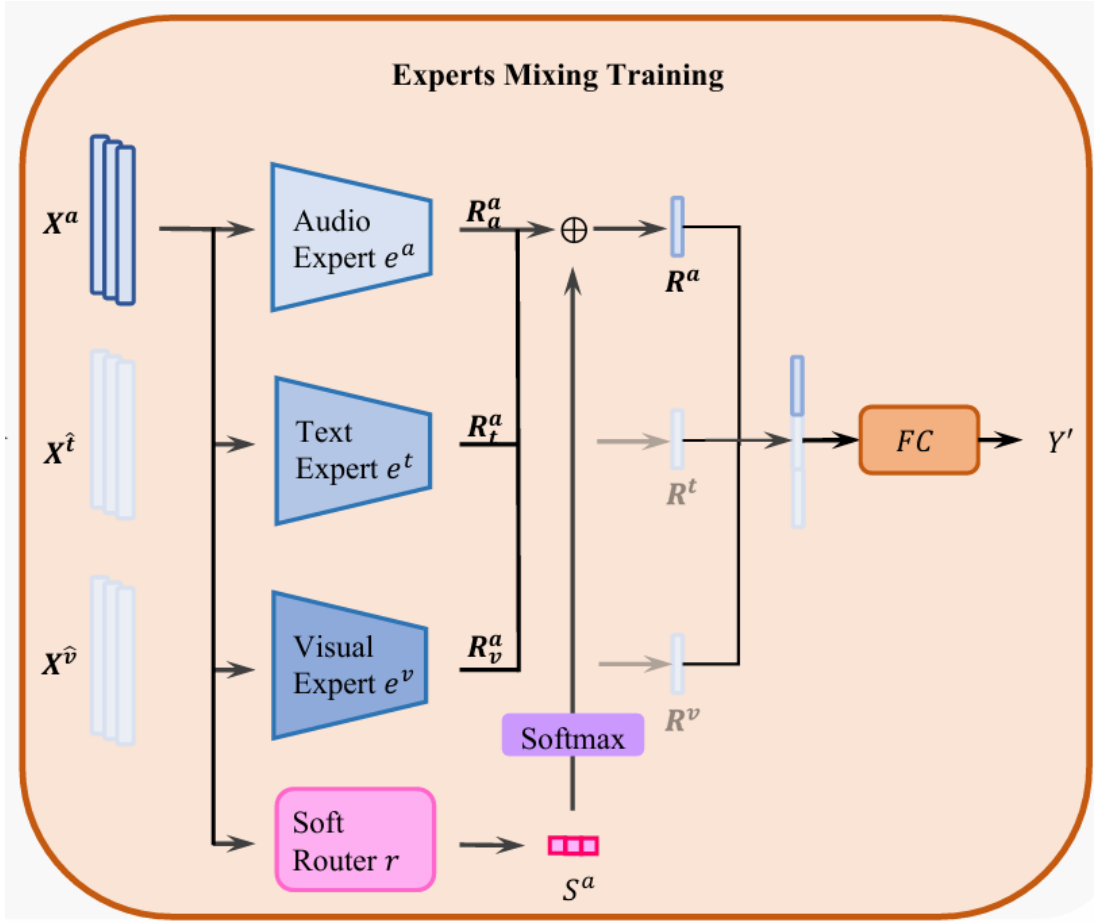
$$S^a = [s_a^a, s_a^t, s_a^v] = r(X^a) = MLP(X^a)$$

    - 归一化加权：

$$w_i^a = softmax(s_i^a)$$
$$R^a = \sum_{i\in\{a,t,v\}} w_i^a \cdot R_i^a$$

    - 最终表示：

$$Y' = FC'(R^m)$$



实验设计　数据资源

- 数据集
  - IEMOCAP（10位演员，5男5女，151个对话，4类情绪）

Neutral	Happy	Sad	Angry
1708	1636	1084	1103

  - CMUMOSI（2199个句子，英语，二分类）
  - CMUMOSEI（22856个句子，英语，二分类）
- 对比方法

  MCTN(2019)、MMIN (2021) 、IF-MMIN (2023)、MRAN(2023)
- 评价指标
  - UA：非加权平均召回率
  - WA：加权平均召回率
  - ACC:准确率
  - F1：F1值

实验设计　对比实验

- 评估MoMKE在数据集上的表现
  - 所有模态所有数据集上均有提升

Datasets	Models	Testing Condition							
		$\{a\}$	$\{t\}$	$\{v\}$	$\{a,v\}$	$\{a,t\}$	$\{t,v\}$	Average	$\{a,t,v\}$
		WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)
IEMOCAP	MCTN[21]	49.75/51.62	62.42/63.78	48.92/45.73	56.34/55.84	68.34/69.46	67.84/68.34	58.94/59.13	-
	MMIN[38]	**56.58/59.00**	66.57/ 68.02	52.52/**51.60**	63.99/65.43	72.94/75.14	72.67/73.61	64.10/65.24	-
	IF-MMIN[41]	55.03/53.20	**67.02/68.20**	51.97/50.41	**65.33/66.52**	**74.05/75.44**	**72.68/73.62**	**64.54/65.38**	-
	MRAN[17]	55.44/57.01	65.31/66.42	**53.23**/49.80	64.70/64.46	73.00/74.58	72.11/72.2	63.97/64.08	-
	MoMKE(ours)	**70.32/71.38**	**77.82/78.37**	**58.60/54.70**	**68.85/67.65**	**79.89/79.53**	**77.87/77.84**	**72.23/71.58**	**80.13/79.99**
	Δ_{SOTA}	↑13.74/↑12.38	↑10.80/↑10.17	↑5.37/↑3.10	↑3.52/↑1.13	↑5.84/↑4.09	↑5.19/↑4.22	↑7.69/↑6.20	-
		ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)
CMU-MOSI	MCTN^a[21]	56.10/54.50	79.10/79.20	55.00/54.40	57.50/57.40	81.00/81.00	81.10/81.20	68.30/67.95	81.40/81.50
	MMIN^a[38]	55.30/51.50	83.80/83.80	57.00/54.00	60.40/58.50	84.00/84.00	83.80/83.90	70.72/69.28	84.60/84.40
	GCNet^a[13]	56.10/54.50	83.70/83.60	56.10/55.70	62.00/61.90	84.50/84.40	84.30/84.20	71.12/70.72	85.20/85.10
	IMDer^a[31]	**62.00**/**62.20**	**84.80/84.70**	**61.30/60.80**	**63.60/63.40**	**85.40/85.30**	**85.50/85.40**	**73.77/73.63**	**85.70/85.60**
	MoMKE(ours)	**63.19**/**58.61**	**86.59/86.52**	**63.35/63.34**	**64.04/64.66**	**87.20/87.17**	**87.04/87.00**	**75.24/74.55**	**87.96/87.89**
	Δ_{SOTA}	↑1.19/↓3.59	↑1.79/↑1.82	↑2.05/↑2.54	↑0.44/↑1.26	↑1.80/↑1.87	↑1.54/↑1.60	↑1.47/↑0.92	↑2.26/↑2.39
CMU-MOSEI	MCTN^a[21]	62.70/54.50	82.60/82.80	62.60/57.10	63.70/62.70	83.50/83.30	83.20/83.20	73.05/70.60	84.20/84.20
	MMIN^a[38]	58.90/59.50	82.30/82.40	59.30/60.00	63.50/61.90	83.70/83.30	83.80/83.40	71.92/71.75	84.30/84.20
	GCNet^a[13]	60.20/60.30	83.00/83.20	61.90/61.60	64.10/57.20	84.30/84.40	84.30/84.40	73.10/72.80	**85.20/85.10**
	IMDer^a[31]	**63.80/60.60**	**84.50/84.50**	**63.90/63.60**	**64.90/63.50**	**85.10/85.10**	**85.00/85.00**	**76.00/75.30**	85.10/**85.10**
	MoMKE(ours)	**72.56/71.03**	**86.46/86.43**	**70.12/70.23**	**73.34/71.82**	**86.68/86.61**	**86.79/86.69**	**79.33/78.80**	**87.12/87.03**
	Δ_{SOTA}	↑8.76/↑10.43	↑1.96/↑1.93	↑6.22/↑6.63	↑8.44/↑8.32	↑1.58/↑1.51	↑1.79/↑1.69	↑4.80/↑5.08	↑2.02/↑1.93

- 评估MoMKE的不同模块在数据集上的表现
  - Without unimodal expert training：去除单模态专家训练
  - Without experts mixing training：去除混合专家训练
  - Without router：去除动态路由机制

Datasets	Modules	Testing Condition							
		{a}	{t}	{v}	{a,v}	{a,t}	{t,v}	Average	{a,t,v}
		WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)	WA(%)/UA(%)
IEMOCAP	Without unimodal expert training	67.89/68.56	75.30/76.54	56.97/53.09	66.58/65.91	77.36/77.93	76.99/76.81	70.02/69.64	77.62/77.47
	Without experts mixing training	67.88/67.91	75.74/75.34	56.87/53.37	66.65/65.46	77.25/78.33	76.99/76.93	70.06/69.39	77.17/77.80
	Without router	68.32/69.03	76.22/76.30	57.00/53.70	67.10/66.32	78.78/78.53	77.00/77.21	70.74/70.18	78.78/78.23
	MoMKE(ours)	**70.32/71.38**	**77.82/78.37**	**58.60/54.70**	**68.85/67.65**	**79.89/79.53**	**77.87/77.84**	**72.23/71.58**	**80.13/79.99**
		ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)	ACC(%)/F1(%)
CMU-MOSI	Without unimodal expert training	56.09/56.36	85.20/85.09	62.20/61.82	63.41/63.57	86.20/86.13	85.50/85.51	73.10/73.08	86.89/86.85
	Without experts mixing training	59.76/57.34	85.04/85.94	62.35/61.96	63.10/63.28	86.04/85.64	85.35/85.34	73.61/73.25	87.20/87.08
	Without router	61.31/58.56	85.30/86.10	62.20/61.82	63.58/63.20	86.20/86.43	85.78/85.71	74.06/73.64	87.00/86.70
	MoMKE(ours)	**63.19/58.61**	**86.59/86.52**	**63.35/63.34**	**64.04/64.66**	**87.20/87.17**	**87.04/87.00**	**75.24/74.55**	**87.96/87.89**
CMU-MOSEI	Without unimodal expert training	71.18/70.13	84.93/84.95	68.23/67.39	71.18/70.52	84.23/84.24	85.26/85.25	77.50/77.08	86.20/86.06
	Without experts mixing training	70.85/69.37	85.01/85.84	68.23/67.08	70.40/70.12	84.20/84.10	85.34/85.25	77.17/76.79	86.51/86.43
	Without router	71.59/70.46	85.45/85.48	68.99/68.00	72.00/71.18	84.88/85.12	85.34/85.55	78.04/77.63	86.89/86.43
	MoMKE(ours)	**72.56/71.03**	**86.46/86.43**	**70.12/70.23**	**73.34/71.82**	**86.68/86.61**	**86.79/86.69**	**79.33/78.80**	**87.12/87.03**

算法总结 MoMKE

- 算法贡献
  - 单模态专家训练：显式建模单模态知识，解决了传统方法在严重缺失条件下判别力严重下降的问题
  - 专家混合训练：不单是融合多模态输入，而是“用其他模态的知识来理解当前模态”
  - 动态路由机制：实现了对不同缺失场景的自适应表示建模

- 算法不足
  - 未考虑到不同样本难度不同，训练存在差异
  - 跨模态迁移 完全依赖训练数据分布一致性
  - 小数据集上容易过拟合

Hardness-Aware Dynamic Curriculum Learning for Robust Multimodal Emotion Recognition with Missing Modalities

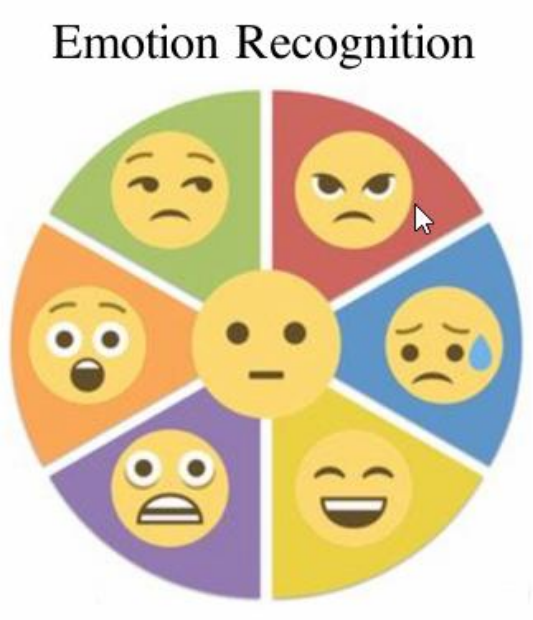
HARDY-MER TIPO

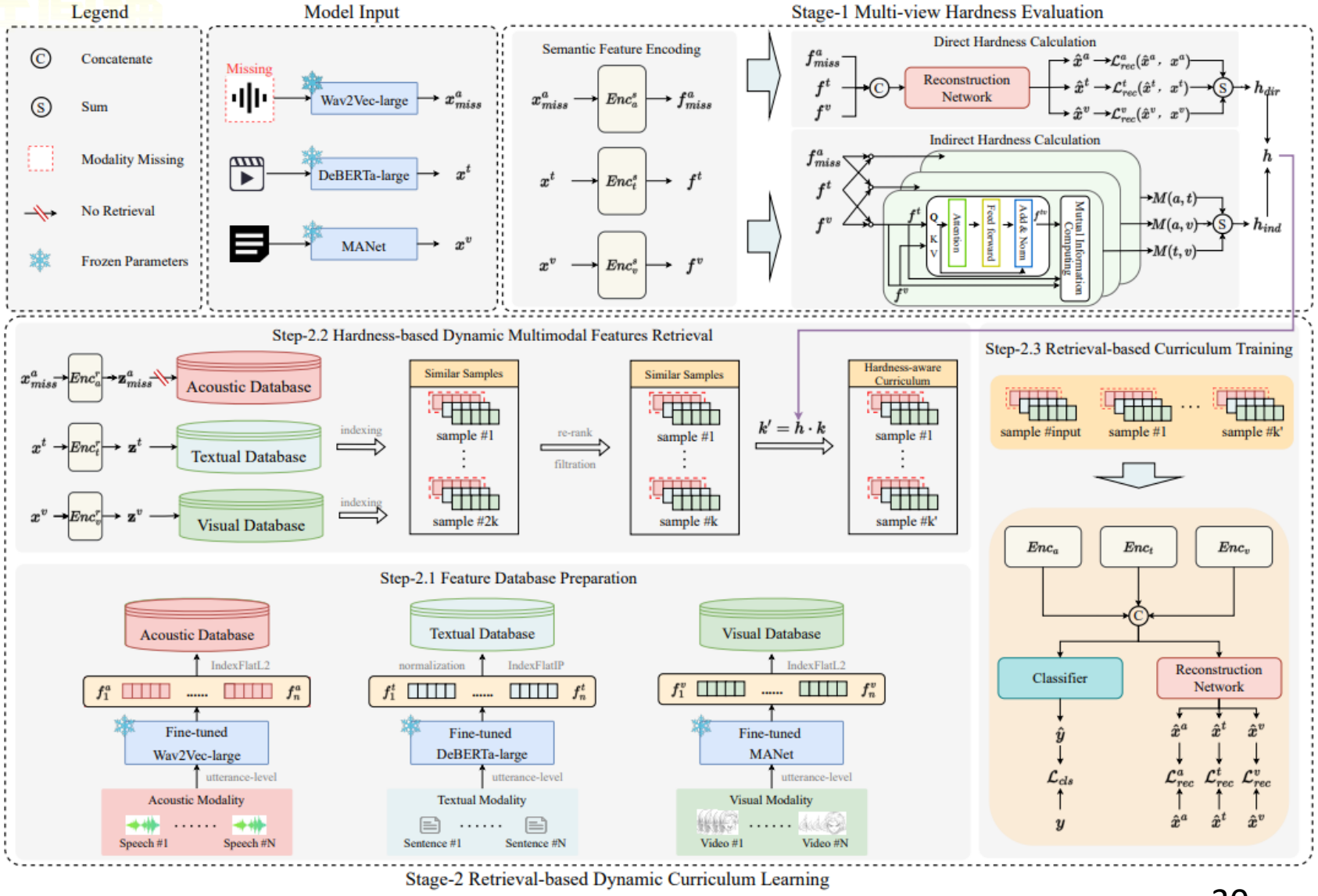
T	目标	实现缺失模态条件下的情绪识别，提升模型在复杂缺失模式下的泛化性能
I	输入	3个数据集（IEMOCAP4/6类、CMUMOSEI数据集）
P	处理	1.多视角难度评估 2.动态特征检索 3.难度感知课程学习 4.预测情绪
O	输出	情绪类别（2/4/6类）

P	问题	1.现有缺失模态方法在训练阶段对样本难度与语义复杂性不敏感 2.高难样本在训练早期主导梯度更新，导致模型不稳定、泛化能力下降
C	条件	需要完整模态样本用于训练初期的难度估计
D	难点	1.如何在缺失模态场景下合理量化样本学习难度 2.如何设计动态、自适应的课程学习策略而非固定训练顺序
L	水平	2025 CCF A类

HARDY-MER 算法原理

- 算法原理图
  - 多视角难度评估
  - 动态特征检索
  - 难度感知课程学习





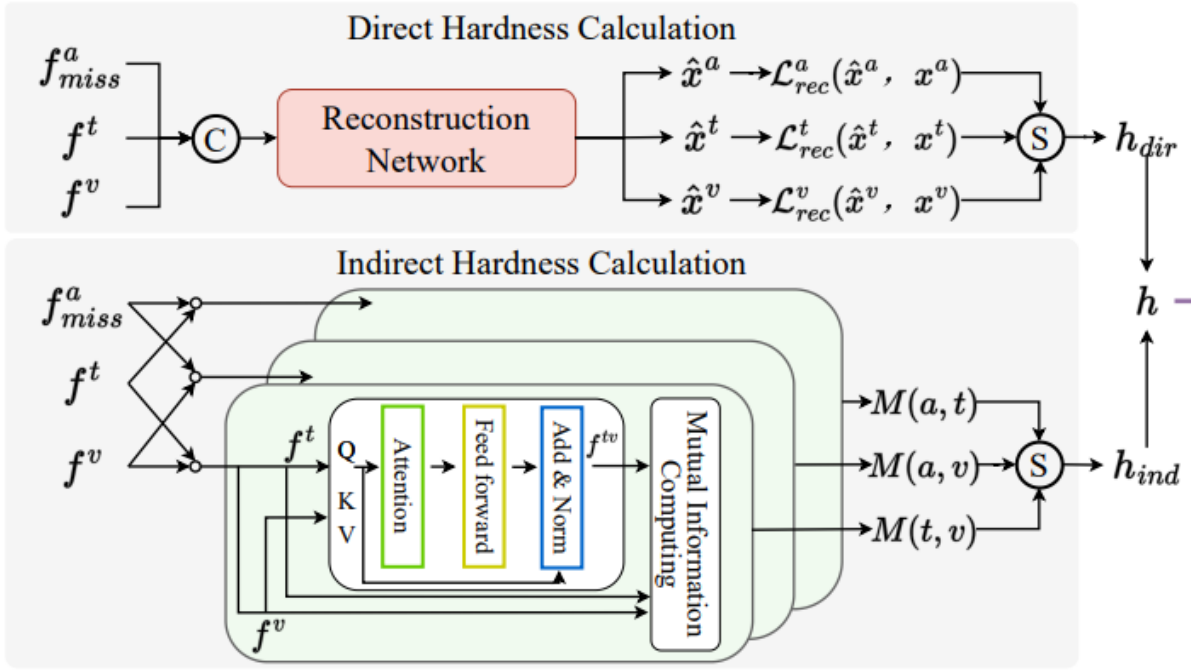
HARDY-MER 创新点分析

- 现有方法存在问题
  - 现有方法无法识别“存在但不可替代”的关键模态
  - 现有方法多模态样本的学习难度可能来源于多个互不等价的因素，单一指标只能反映其中一个侧面无法覆盖多模态交互带来的复杂性
- 解决方法
  - 多视角难度评估

$$h = h_{dir} + h_{ind}$$

其中 h_{dir} 直接难度表示信息是否可被重建

h_{ind} 间接难度表示跨模态语义是否一致

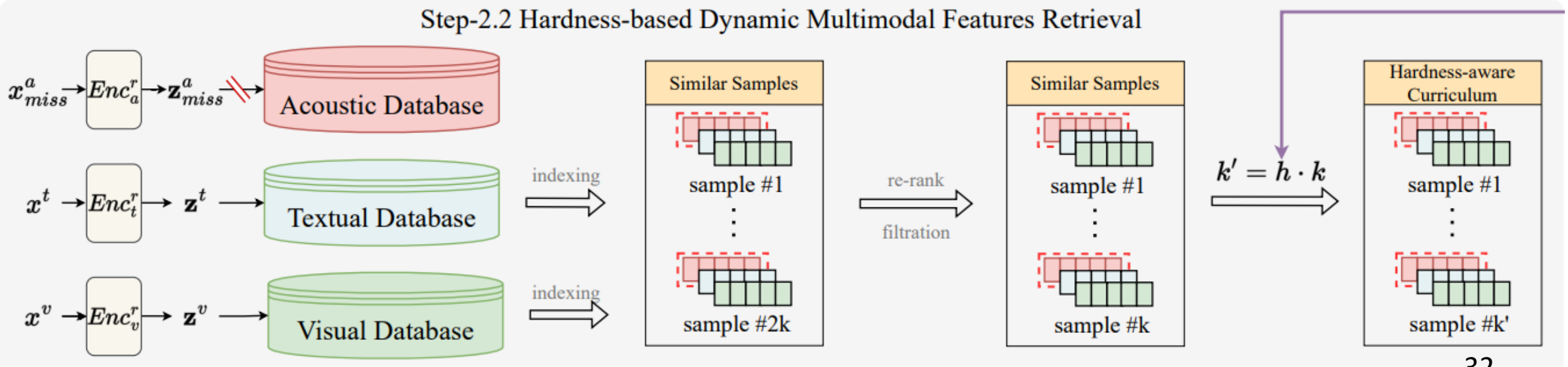


难度构建

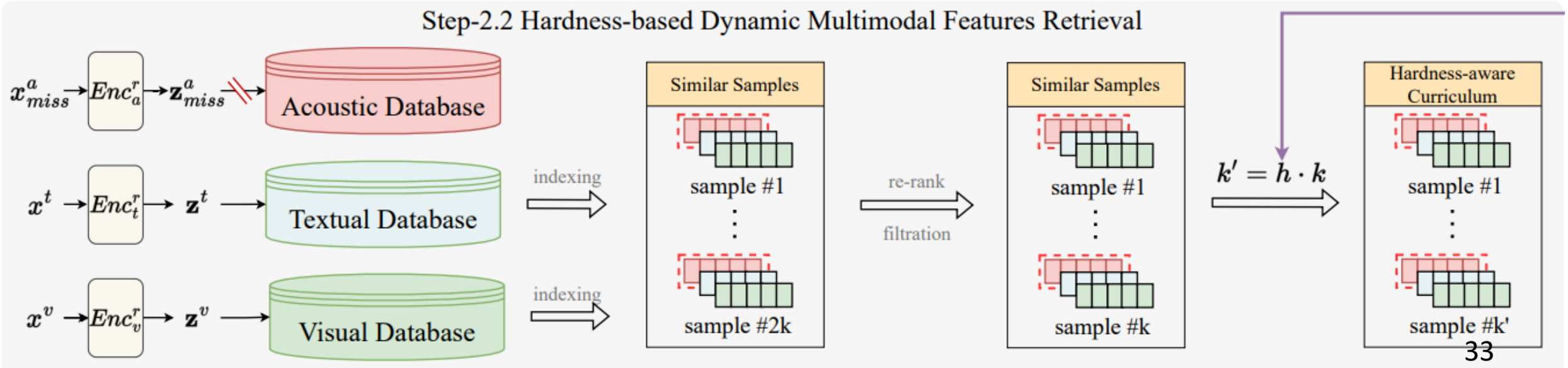
HARDY-MER 创新点分析

- 现有方法存在问题
  - 现有方法在缺失模态和高难样本条件下，单一样本信息不足、直接学习不稳定
- 解决方法
  - 动态特征检索
    - 检索编码器得到查询向量

$$z^m = Enc_r^m(x^m)$$



- 解决方法
  - 动态特征检索
    - 分模态检索top-k相似样本
    - 索引合并去重，选取候选样本
    - 计算综合相似度： $s(cand, x) = \frac{1}{|M|}\sum_{m\in M} ||f_{cand}^{m} - f_{x}^{m}||_2$
    - 最终样本数量： $k' = \lceil h \cdot k \rceil$



HARDY-MER 创新点分析

- 解决方法
  - 难度感知的课程学习
    - 情绪分类损失

$$L_{cls}(x_i) = l(\hat{y}_i, y_i)$$

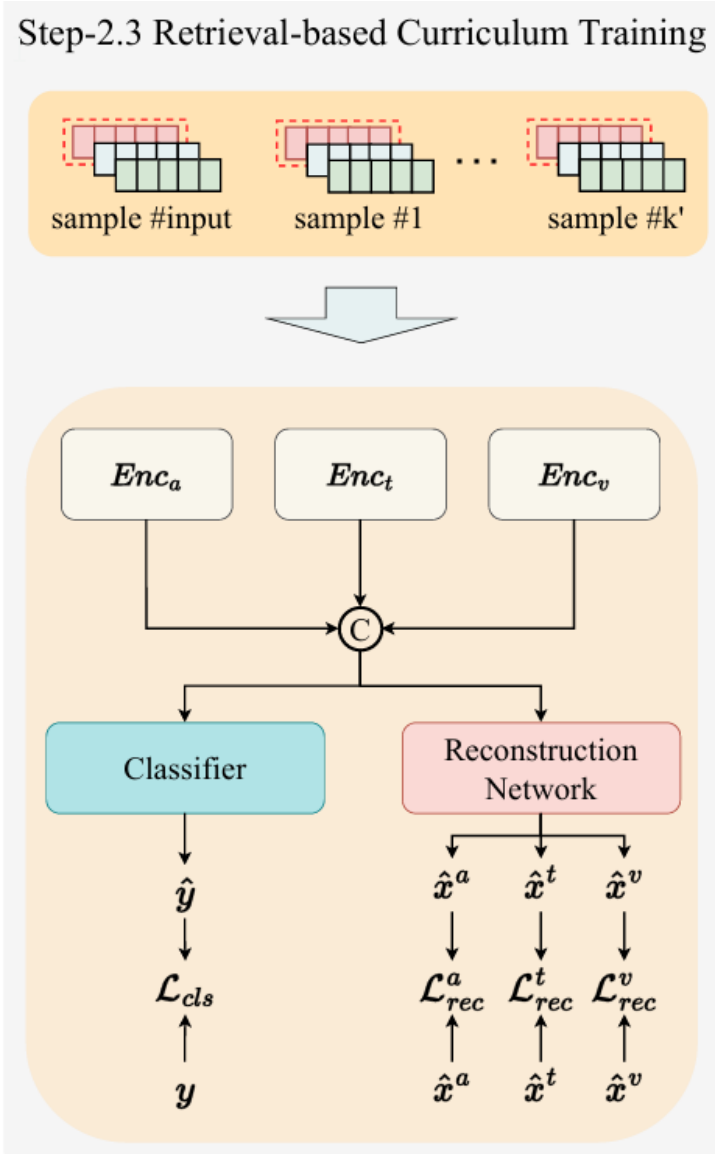
$$L_{sup}(x_i) = \frac{1}{k'}\sum_{x_{ij}\in S(x_i)} l(\widehat{y_{ij}}, y_{ij})$$

    - 重建损失

$$L_{rec}(x_i) = \sum_{m\in(a,t,v)} \delta_i^m \cdot ||x_i^m - \widehat{x_i^m}||$$

    - 总损失

$$L = L_{cls}(x_i) + \lambda_{sup} L_{sup}(x_i) + \lambda_{rec}\, L_{rec}(x_i)$$



实验设计　数据资源

- 数据集
  - IEMOCAP（10位演员，5男5女，151个对话，4类和6类情绪）

Neutral	Happy	Sad	Angry
1708	1636	1084	1103

Neutral	Happy	Sad	Angry	Excited	Frustrated
1708	595	1084	1103	1041	1849

  - CMUMOSEI（22856个句子，英语，二分类）
- 对比方法

  CPMNet(2020)、MMIN(2021)、GCNet(2023)、 CIF-MMIN (2024)、MoMKE (2024)
- 评价指标
  - UA：非加权平均召回率
  - WA：加权平均召回率
  - ACC:准确率
  - F1：F1值

实验设计　对比实验

- 评估HARDY-MER在数据集上的表现
  - 单模态下HARDY-MER在所有数据集均为最优
  - 在音频和文本的提升要大于视频的提升

Dataset	model	a		v		t	
		WA	UA	WA	UA	WA	UA
IEMOCAP four-class	CPMNet [53]	0.4685	0.5172	0.4495	0.4449	0.4563	0.4532
	MMIN [55]	0.5658	0.5900	0.5252	0.5060	0.6657	0.6802
	GCNet [18]	0.6558	0.6876	0.5796	0.5254	0.7233	0.7042
	CIF-MMIN [23]	0.5753	0.6006	0.5346	0.5156	0.6722	0.6899
	MoMKE [50]	0.6953	0.7021	0.5680	0.5203	0.7730	0.7766
	HARDY-MER (our)	**0.7265**	**0.7387**	**0.6319**	**0.6054**	**0.8249**	**0.8269**
	Δ_{Sota}	↑0.0312	↑0.0366	↑0.0523	↑0.0800	↑0.0519	↑0.0503
IEMOCAP six-class	CPMNet [53]	0.2947	0.2980	0.2620	0.2495	0.3244	0.3495
	MMIN [55]	0.4408	0.4296	0.3574	0.3065	0.4217	0.3855
	GCNet [18]	0.4995	0.4645	0.3978	0.3497	0.5648	0.5562
	CIF-MMIN [23]	0.4496	0.4356	0.3611	0.3135	0.4340	0.3971
	MoMKE [50]	0.5051	0.4738	0.3907	0.3451	0.6109	0.6019
	HARDY-MER (our)	**0.5158**	**0.4914**	**0.4302**	**0.3649**	**0.6589**	**0.6195**
	Δ_{Sota}	↑0.0107	↑0.0176	↑0.0324	↑0.0152	↑0.0480	↑0.0176
Dataset	model	a		v		t	
		ACC	F1	ACC	F1	ACC	F1
CMUMOSEI	CPMNet [53]	0.6571	0.6518	0.6123	0.6173	0.7287	0.7244
	MMIN [55]	0.5890	0.5950	0.5930	0.6001	0.8220	0.8240
	GCNet [18]	0.7204	0.7034	0.6808	0.6725	0.8426	0.8417
	CIF-MMIN [23]	0.6387	0.6460	0.6196	0.6266	0.8353	0.8304
	MoMKE [50]	0.7256	0.7103	0.6450	0.6346	0.8610	0.8603
	HARDY-MER (our)	**0.7482**	**0.7411**	**0.6935**	**0.6750**	**0.8720**	**0.8713**
	Δ_{Sota}	↑0.0226	↑0.0308	↑0.0127	↑0.0025	↑0.0110	↑0.0110

实验设计　对比实验

- 评估HARDY-MER在数据集上的表现
  - 双模态下HARDY-MER在IEMOCAP4类和6类上均为最优
  - 在模态互补性较强的组合（av）上优势明显
  - 在CMUMOSEI数据集上，av是提升的，其他也是次优

平均性能最优

Dataset	model	at		av		tv	
		WA	UA	WA	UA	WA	UA
IEMOCAP four-class	CPMNet [53]	0.3481	0.3623	0.4867	0.4933	0.4562	0.4657
	MMIN [55]	0.7294	0.7114	0.6399	0.6343	0.7167	0.6861
	GCNet [18]	0.7702	0.7687	0.6740	0.6564	0.7563	0.7362
	CIF-MMIN [23]	0.7419	0.7259	0.6499	0.6353	0.7240	0.6991
	MoMKE [50]	0.7903	0.7988	0.6857	0.6622	0.7555	0.7418
	HARDY-MER (our)	**0.8167**	**0.8243**	**0.7419**	**0.7450**	**0.7918**	**0.7851**
	Δ_{Sota}	↑0.0264	↑0.0255	↑0.0562	↑0.0828	↑0.0355	↑0.0433
IEMOCAP six-class	CPMNet [53]	0.3349	0.3394	0.2692	0.2546	0.3134	0.3043
	MMIN [55]	0.5195	0.4831	0.4192	0.3815	0.4749	0.4063
	GCNet [18]	0.5824	0.5725	0.4757	0.4331	0.5743	0.5466
	CIF-MMIN [23]	0.5243	0.4920	0.4254	0.3922	0.4888	0.4491
	MoMKE [50]	0.6318	0.6194	0.4865	0.4408	0.5992	0.5755
	HARDY-MER (our)	**0.6518**	**0.6298**	**0.5291**	**0.4745**	**0.6166**	**0.5786**
	Δ_{Sota}	↑0.0200	↑0.0104	↑0.0426	↑0.0337	↑0.0174	↑0.0031
Dataset	model	at		av		tv	
		ACC	F1	ACC	F1	ACC	F1
CMUMOSEI	CPMNet [53]	0.7265	0.7224	0.6156	0.6199	0.6629	0.6684
	MMIN [55]	0.8370	0.8330	0.6355	0.6191	0.8175	0.8142
	GCNet [18]	0.8510	0.8510	0.7149	0.6996	0.8474	0.8454
	CIF-MMIN [23]	0.8401	0.8347	0.6468	0.6208	0.8250	0.8194
	MoMKE [50]	**0.8632**	**0.8629**	0.7237	0.7207	**0.8690**	**0.8691**
	HARDY-MER (our)	0.8542	0.8501	**0.7482**	**0.7411**	0.8572	0.8539
	Δ_{Sota}	↓-0.0090	↓-0.0128	↑0.0245	↑0.0204	↓-0.0118	↓-0.0152

实验设计　消融实验

- 评估HARDY-MER的不同模块在数据集上的表现
  - -w/o h_{dir}：不做直接难度评估（重建难度）
  - -w/o h_{ind}：不做间接难度评估（跨模态一致性难度）
  - -w/o h：去掉完整难度建模
  - -w/o retrieval features:去掉检索特征
  - -w/o fine-tuning features:去掉特征微调

model	Testing Condition													
	a		v		t		at		av		tv		Average	
	WA	UA	WA	UA	WA	UA	WA	UA	WA	UA	WA	UA	WA	UA
HARDY-MER (our)	**0.7265**	**0.7387**	**0.6319**	**0.6054**	**0.8249**	**0.8269**	**0.8167**	**0.8243**	**0.7419**	**0.745**	**0.7918**	**0.7851**	**0.7556**	**0.7542**
w/o h_{dir}	0.7202	0.7246	0.6196	0.5933	0.8149	0.8184	0.8087	0.8164	0.7345	0.7307	0.7778	0.7754	0.7460	0.7431
w/o h_{ind}	0.7231	0.7281	0.6186	0.5945	0.8161	0.8161	0.8090	0.8129	0.7374	0.7374	0.7867	0.7775	0.7485	0.7444
w/o h	0.7215	0.7301	0.6136	0.5852	0.8142	0.8175	0.8124	0.8184	0.6889	0.6881	0.7719	0.7693	0.7371	0.7348
w/o retrieval features	0.7201	0.7217	0.6200	0.5806	0.8128	0.8137	0.8093	0.8132	0.7331	0.7282	0.7883	0.7719	0.7473	0.7382
w/o fine-tuning features	0.7126	0.7218	0.5916	0.5345	0.7299	0.7421	0.7496	0.7647	0.7364	0.7390	0.7387	0.7404	0.7098	0.7071

算法总结 HARDY-MER

- 算法贡献
  - 多视角难度评估：从直接难度与间接难度两个互补视角刻画样本学习难度，避免了单一指标难以全面反映多模态复杂性的不足
  - 动态特征检索：利用FAISS在模态特征空间中高效检索样本，缓解缺失模态导致的训练不稳定问题语义相似
  - 难度感知课程学习：根据样本难度动态决定检索支持样本的数量，困难样本获得更多语义相似的支持样本，简单样本几乎不受干预

- 算法不足
  - 课程学习效果依赖超参数设置
  - 检索索引需预先构建，难以完全在线更新

特点总结与未来展望

特点总结与未来展望

- 特点总结
  - MoMKE
    - 构建单模态专家，混合专家训练，通过动态路由机制调整权重
      - 解决了缺失模态条件下联合表示退化的问题
  - HARGY-MER
    - 多视角难度评估，动态检索相似样本，进行难度感知课程学习
      - 缓解了样本难度不均衡导致的训练偏置问题

- 未来发展
  - 如何将缺失模式和情绪相结合
  - 如何判断需要使用哪些模态

参考文献

[1] Xu, Wenxin, Hexin Jiang, and Xuefeng Liang. "Leveraging knowledge of modality experts for incomplete multimodal learning." Proceedings of the 32nd ACM International Conference on Multimedia. 2024.

[2] Liu, Rui, et al. "Hardness-Aware Dynamic Curriculum Learning for Robust Multimodal Emotion Recognition with Missing Modalities." Proceedings of the 33rd ACM International Conference on Multimedia. 2025.

知人者智，自知者明。胜人者有力，自胜者强。知足者富。强行者有志。不失其所者久。死而不亡者，寿。

谢谢！