3.3 基于动态权重分配的多模态信息平衡技术

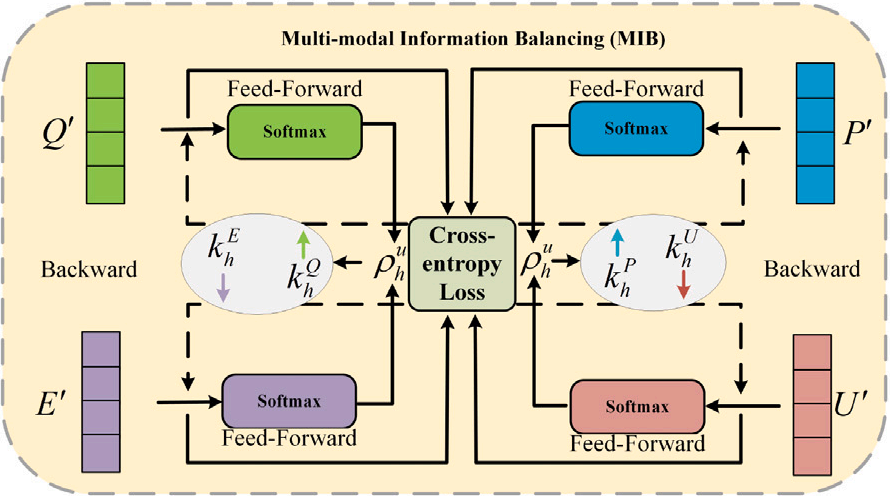


图1 多模态信息平衡模块架构图

模态不平衡问题是指在优化过程中，主导模态会抑制其他模态性能的现象。为解决这一问题，我们设计了多模态信息平衡（MIB）模块，在各模态的优化过程中动态地为它们分配权重，自适应地调节模型的优化过程，以实现模态平衡，如图1所示。

对于的每个通道，其参数为，梯度更新过程由下式表示：

其中是第次迭代更新后的参数。

我们通过更精细地监测信息，自适应地调整每个模态的梯度。这些信息可以表述为每个模态的特定信息，对数几率得分（logit score）能直接反映不同模态的激活程度。因此，在本文中，我们计算每个模态的模式视图和语义视图的对数几率得分，以探究每个模态的信息量。对数几率得分包含，其中和是的参数，表示中的第个特征，是第次迭代中的一个随机批次。我们定义来进一步量化不同模态视图的信息：

其中 ，是类别数。

我们设计差异率来衡量各模态对优化过程的影响：

越大，与其他模态视图相比，相应模态视图中的信息量就越大，从而导致模态不平衡。

信息量较大的模态在模型的优化过程中起着主导作用[1][2][3]，并抑制其他模态的优化过程。因此，我们设计平衡因子来平衡各模态的优化过程。

然后，我们将整合到公式 (1) 中。的更新过程如下：

通过使用，我们抑制了性能较好的模态视图的优化，而性能较差的模态视图则不受影响。

在获取多模态的融合特征后，我们将其输入到一个分类器中，该分类器包含一层多层感知器（MLP）和 激活函数，以获得预测标签：

其中是分类器，是分类器的参数。

为了增强模型的分类能力，我们最小化交叉熵损失：

为了增强特征的判别能力，我们采用跨模态监督对比损失。是中与属于同一类别的索引集，是中与类别不同的索引集。

其中，是的基数，是一个标量参数，根据经验设置为。因此，总损失可以表示为：

其中，是一个平衡因子。

我们使用 Adam 优化器在总损失下寻找最优的网络参数。对于测试样本集 ，我们使用最优参数、、、、、来获取相应的特征，然后将这些特征输入到分类器中以获得标签。通过我们的方法，可以调节每个模态的优化过程，有效缓解模态不平衡问题。

1. X. Peng, Y. Wei, A. Deng, D. Wang, D. Hu, Balanced multimodal learning via on-the-fly gradient modulation, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8238–8247.
2. Y. Wei, R. Feng, Z. Wang, D. Hu, Enhancing multimodal cooperation via samplelevel modality valuation, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 27338–27347.
3. R. Xu, R. Feng, S.-X. Zhang, D. Hu, Mmcosine: Multi-modal cosine loss towards balanced audio-visual fine-grained learning, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2023, pp. 1–5.

|  |
| --- |
| 算法2 基于动态权重分配的多模态信息平衡技术 |
| 输入数据：训练集 ，包含文本模态和图像模态以及标签矩阵 ；测试集 ，包含文本模态 和图像模态 。  for each epoch ：   1. 计算各模态特征 ； 2. 计算每个模态的对数几率得分 ； 3. 计算差异率 ； 4. 计算平衡因子 ； 5. 融合多模态特征 ； 6. 通过分类器 得到预测标签 ； 7. 反向传播更新参数；   optimizer.zero\_grad()  L\_total.backward()  更新模型参数；  输出：预测标签。 |