

MoME: Mixture of Multimodal Experts for Cancer Survival Prediction

摘要

生存分析是一项具有挑战性的任务，需要整合全幻灯片图像 (WSIs) 和基因组数据以进行全面决策。这项任务有两个主要挑战：两种模态之间的显著异质性和复杂的模态间和模态内相互作用。以前的方法利用共同注意方法，在单独编码后仅将两种模态的特征融合一次。然而，由于模态之间的异质性，这些方法不足以对复杂任务进行建模。为了解决这些问题，论文提出了一种有偏渐进编码 (BPE) 范式，同时执行编码和融合。该范式在编码另一种模态时使用一种模态作为参考。它通过多次交替迭代实现模态的深度融合，逐步减少跨模态差异并促进互补相互作用。除了模态异质性之外，生存分析还涉及来自 WSI、基因组学及其组合的各种生物标志物。关键生物标志物可能在个体差异下以不同的方式存在，因此需要灵活地调整模型以适应特定场景。因此，进一步提出了多模态专家混合 (MoME) 层，以动态选择 BPE 范式每个阶段的定制专家。专家在不同程度上结合了来自另一种模态的参考信息，从而能够在编码过程中平衡或有偏向地关注不同的模态。大量实验结果表明，该方法在各种数据集 (包括 TCGA-BLCA、TCGA-UCEC 和 TCGA-LUAD) 上表现出色。

关键词：多模态学习；生存预测；计算病理学

1 引言

通过全切片图像 (WSIs) 和基因组数据进行生存分析对于全癌症预后至关重要，因为它可以评估死亡风险并为治疗计划提供重要参考。这项任务的关键是如何有效地利用来自两种模式的信息，例如，检测图像组学生物标志物以及探索组织病理学图像中的肿瘤微环境与基因组数据共表达之间的相互作用。近年来，研究的重点也从单模态预测转向利用多模态信息的更复杂的生存分析。这项任务的关键挑战之一是组织病理学图像和基因组数据之间存在显著的异质性，这源于它们固有的差异和不同的预处理方法。此外，模态间和模态内相互作用非常复杂，因为两种模态都拥有丰富的信息，但只有一小部分可以相互关联并用于生存预测。以前的方法试图通过使用基于跨模态注意（共同注意）的方法来应对这一挑战。然而，在整个过程中，特征融合只进行一次。考虑到任务的复杂性和两种模态之间的显著差异，这些方法可能被认为是肤浅的。为了解决这些问题，提出了一种偏置渐进编码 (BPE) 范式。与之前在融合之前分别编码模态的方法不同，该方法同时编码和融合特征。在这种方法中，一种模态被编码，同时利用另一种模态作为参考，这有助于提取更多相关信息。此外，两种模态的特征编码是交替进行的，从而逐步减少它们的特征空间之间的差异。这允许更深层次的融合

过程并有助于发现模态之间的相互作用。除了模态异质性之外，个体间差异还可能导致生存分析的关键特征出现在每个患者的不同模态中。这给模型结构的设计带来了新的挑战，因为它需要选择性地关注特定模态或两种模态之间的相互作用。为了实现这一点，提出了一个多模态专家混合 (MoME) 层，采用 BPE 范式。MoME 层由多个专业专家组成，能够对复杂的模态间和模态内相关性进行建模。此外，这些专家在不同程度上结合了来自另一种模态的参考信息，从而能够在编码过程中关注不同的模态。此外，可以灵活地选择每一层的专家，因为参考信息的功能可能不同。在网络的浅层，MoME 层可以使用参考信息作为过滤器，以消除与任务无关的特征并增强每个模态中的相关特征。相反，在网络的更深层，它可以用作寻找跨模态组合表示作为生物标记的指导。

2 相关工作

2.1 癌症生存预测中的多模态学习

多模态数据融合：在癌症生存预测中，很多研究已经开始使用多种数据源（如临床数据、基因数据、影像数据等）来改善预测准确性。例如：

- DeepSurv：使用深度学习方法预测生存时间，基于 Cox 回归模型，结合基因数据。 [1]
- Multi-modal deep learning：一些研究结合了基因组学数据和临床数据，使用深度学习模型进行联合建模，以提高预测精度。 [2]

2.2 Mixture of Experts (专家混合) 模型

专家混合模型 (Mixture of Experts, MoE) 是一种经典的机器学习方法，广泛应用于多个领域。MoE 模型通过为不同数据模式分配不同的专家来进行建模，适用于多模态问题。

- 在生物信息学中，MoE 方法被用来融合基因数据和临床数据，提升预测效果。例如，ExpertNet 是一个结合了多种生物数据源的混合专家模型，在药物响应预测中取得了较好的效果。 [3]
- MoE 模型有助于通过加权不同专家的输出，使得模型能够针对不同的输入模式动态调整，使其在多模态学习任务中更具适应性。 [4]

2.3 深度学习与癌症生存预测

深度神经网络 (DNN)：近年来，深度学习在医学领域的应用越来越广泛，尤其是在癌症预测和诊断方面。深度学习模型能够自动从数据中提取特征，从而避免了传统机器学习方法中手动特征工程的需求。

- SurvivalNet：使用深度神经网络模型进行生存预测，能够有效地处理生存时间数据。 [5]
- Deep Learning for Cancer Prognosis：一些研究使用深度学习模型，特别是卷积神经网络 (CNN) 和递归神经网络 (RNN)，来处理影像数据和基因组数据的结合，从而提高癌症预测的准确性。 [6]

3 本文方法

3.1 本文方法概述

论文的 BPE 范式概览如图 1 所示。

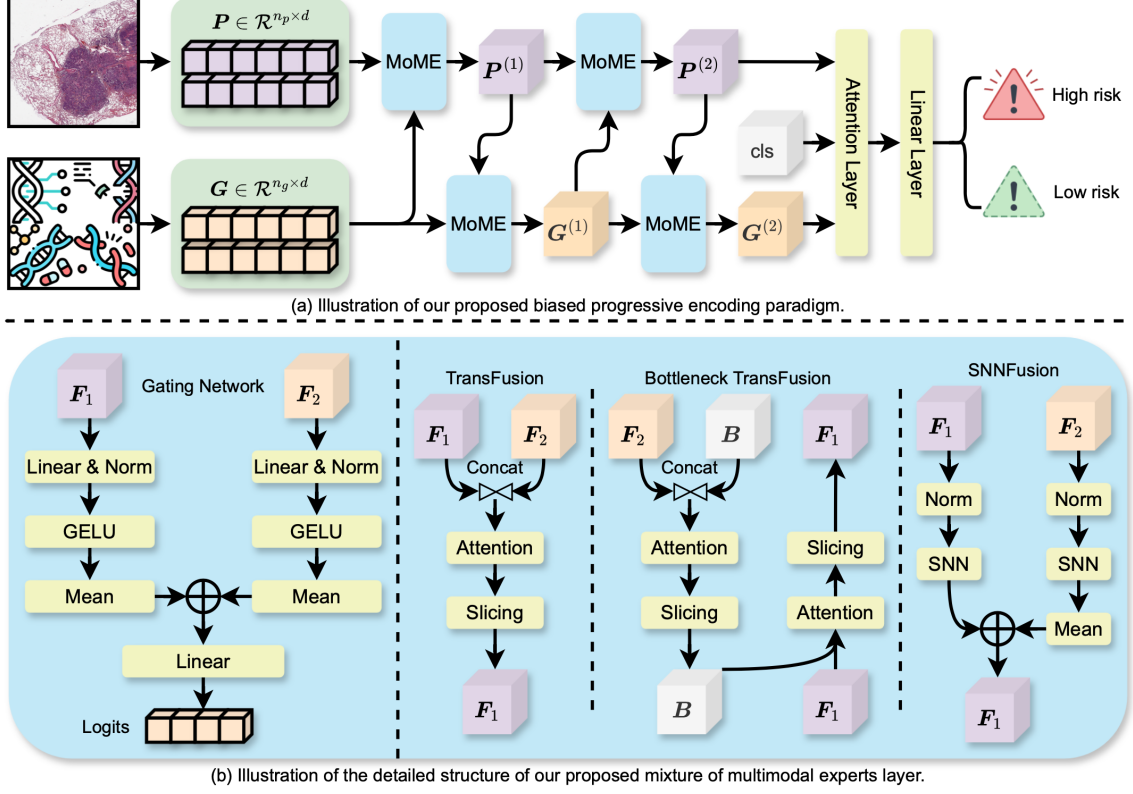


图 1. (a) 有偏渐进编码范式和 (b) 多模态专家混合结构的图示。(b) 的左侧部分代表门控网络，(b) 的右侧部分描述了提出的三个专家组件，它们设计用于不同程度的参考模态集成。

BPE 中的深度特征提取和渐进式学习策略可实现深度融合以减少模态间差异。完整的渐进式编码过程涉及对两种模态进行编码，如下所示：

$$F_1^{(i+1)} = BPE_{2i}(F_1^{(i)}, F_2^{(i)}) \quad (1)$$

$$F_2^{(i+1)} = BPE_{2i+1}(F_2^{(i)}, F_1^{(i+1)}) \quad (2)$$

其中 BPE 是 MoME。我们对所有数据集的两个特征进行两次编码，以促进两种模态的深度融合。编码后，两个特征与分类标记一起输入到注意块中，然后用于最终的生存预测。

3.2 特征提取模块

特征提取模块的主要设计包括以下内容：

1. Biased Progressive Encoding (BPE) Paradigm: BPE 是特征提取和模态融合的核心模块，通过以下策略实现多模态数据的编码和交互：

- **参考模态的交替编码：**在编码一种模态时，另一种模态作为参考，帮助提取更相关的信息。
- **逐步减小模态之间的特征空间差异：**通过交替编码两个模态的特征，逐渐实现更深层次的模态融合。
- **深度特征提取：**此过程通过深度学习模型捕获跨模态的复杂交互信息。

2. Mixture of Multi-modal Experts (MoME) Layer MoME 层由以下两个部分组成：

- **Gating Network：**轻量级的模块，用于选择合适的专家模型处理输入特征。包括线性层、GELU 激活函数、RMS 层归一化等，最终将多模态特征映射到一个新的特征空间。与传统的 Mixture of Experts 模型不同，MoME 强制选择一个最适合的专家，而非对多个专家的输出加权求和
- **Multimodal Expert Pool：**包括专注于单模态（全片图像或基因数据）和模态交互的专家。同时实现模态融合和编码。

3. Attention Mechanism

- MoME 模型选择自注意力机制 (Self-Attention, SA)，而非共注意力 (Co-Attention, CA)。
- 自注意力机制可以同时捕获单模态和模态间的复杂交互特征，而共注意力仅适合浅层交互。

3.3 损失函数定义

为了优化模型参数，使用离散生存模型的对数似然函数，给定二元审查状态 c_j

$$L = -c_j \cdot \log(f_{surv}(Y_j|h_{bag_j})) - (1 - c_j \cdot \log(f_{surv}(Y_j - 1|h_{bag_j}))) - (1 - c_j \cdot \log(f_{hazard}(Y_j|h_{bag_j}))) \quad (3)$$

在这个公式中，我们对在随访期结束后还活着的患者使用 $c_j = 1$ ，如果患者恰好在时间 $T_{j,cont}$ 去世， $c_j = 0$ 。在训练期间，可以通过最小化 L 和 $L_{uncensored}$

$$L_{surv} = -(1 - \beta \cdot L + \beta \cdot L_{uncensored}) \quad (4)$$

损失函数的第二项对应于未经审查的患者，其定义为：

$$L_{uncensored} = -(1 - c_j \cdot \log(f_{surv}(Y_j - 1|h_{bag_j}))) - (1 - c_j \cdot \log(f_{hazard}(Y_j|h_{bag_j}))) \quad (5)$$

4 复现细节

4.1 与已有开源代码对比

原始的 MoME 模型代码包含以下几个关键部分：

- 数据预处理：对多模态数据进行标准化、清洗和合并。
- 模型定义：为每个模态数据定义不同的专家模型。
- 专家选择与加权融合：通过最大似然估计（MLE）来训练每个专家模型的参数，并学习如何加权各专家的输出。
- 训练与评估：使用交叉验证来评估模型的表现，并进行损失函数优化。

我的工作：引入残差连接。在实现和改进 MoME 模型时，我在原始模型基础上引入了残差连接（Residual Connections），这一修改旨在缓解深度神经网络中的梯度消失问题，并提高模型的训练效率和性能。具体来说，残差连接能够有效地缓解层数增加时带来的训练难度，使得信息在网络中流动时不易丢失，尤其对于较深的网络结构来说，残差连接能够帮助优化模型的收敛速度和预测效果。

4.2 实验环境搭建

- **CPU:** Intel i7
- **GPU:** RTX 4060
- **内存:** 32GB
- **Python 环境:**
 - Name: torch Version: 1.13.0
 - Name: torchvision Version: 0.14.0
 - Name: scikit-survival Version: 0.22.2
 - Name: numpy Version: 1.24.3
 - Name: h5py Version: 3.11.0
 - Name: scipy Version: 1.10.1
 - Name: scikit-learn Version: 1.3.0
 - Name: pandas Version: 2.0.3
 - Name: nystrom-attention Version: 0.0.12
 - Name: admin-torch Version: 0.1.0

4.3 创新点

论文的基础上，我引入了残差连接，并通过在全连接神经网络和卷积神经网络中添加跳跃连接来改进模型结构。通过这一修改，模型的训练速度得到了加快，收敛性和精度均有显著提升，尤其是在处理复杂的多模态数据时，残差连接帮助保留了更多的信息流，进一步提高了癌症生存预测的准确性。

5 实验结果分析

Model	Dataset
	HNSC
Propoise	0.5836 ± 0.0235
Mome	0.5989 ± 0.0209

图 2. 复现结果，可以看到 mome 的效果要比 propoise 要好

	mome	去掉mome部分	加入残差	模态对比	
				单wsi	单gene
Cindex	0.6189 ± 0.0209	0.4714 ± 0.0238	0.6314 ± 0.0221	0.4757 ± 0.0215	0.5907 ± 0.0627
AUC	0.5625 ± 0.0490	0.3811 ± 0.0793	0.5636 ± 0.0507	0.3957 ± 0.0781	0.5202 ± 0.0707
Sensitivity	0.7159 ± 0.1596	0.1761 ± 0.6136	0.8580 ± 0.1619	0.3807 ± 0.8987	0.6761 ± 0.4671
Specificity	0.4625 ± 0.1848	0.8438 ± 0.6250	0.3000 ± 0.2475	0.6375 ± 0.9142	0.4625 ± 0.4873
Accuracy	0.5952 ± 0.0389	0.4940 ± 0.0238	0.5923 ± 0.0352	0.5030 ± 0.0357	0.5744 ± 0.0626
Precision	0.5958 ± 0.0462	0.8821 ± 0.4717	0.5767 ± 0.0418	0.7650 ± 0.5428	0.5866 ± 0.0559
Negative Predictive Value	0.6003 ± 0.0618	0.4824 ± 0.0019	0.6920 ± 0.1659	0.4879 ± 0.0170	0.6055 ± 0.1799
F1-score	0.6482 ± 0.0581	0.1777 ± 0.5329	0.6871 ± 0.0240	0.3238 ± 0.6564	0.6114 ± 0.1847
Youden ' s J statistic	0.1784 ± 0.0814	0.0199 ± 0.0114	0.1580 ± 0.0890	0.0182 ± 0.0182	0.1386 ± 0.1245

图 3. 消融实验

Model	Dataset
	HNSC
Propoise	0.5836 ± 0.0235
Mome	0.5989 ± 0.0209
Mome + ResNet	0.6169 ± 0.0318

图 4. 改进结果，模型效果提升显著

6 总结与展望

论文提出了一种新的方法 MoME，用于癌症生存预测。该方法结合了多模态数据并基于 Mixture of Experts (MoE) 模型进行生存预测。通过将不同模态的数据分别交给不同的“专家”模型进行处理，MoME 利用加权融合的方式来进行最终预测。在多模态数据中，缺失数据是一项常见的问题。未来的研究可以考虑引入缺失数据插补方法（如数据增强、深度学习插补）来解决数据缺失问题。通过设计针对性的数据预处理方案，使模型能够更好地应对缺失数据，进而提升预测效果。

参考文献

- [1] Katzman J. L. et al. DeepSurv: Personalized treatment recommender using a Cox proportional hazards deep neural network. *Proceedings of the 1st Machine Learning for Healthcare Conference*, 2018.
- [2] Zhou J. et al. Deep learning for cancer diagnosis and prognosis using multi-modal data. *International Conference on Machine Learning*, 2018.
- [3] Zhang L. et al. ExpertNet: A deep learning approach for predicting cancer drug response based on multi-modal data. *Bioinformatics*, 2020.
- [4] Chen H. et al. Moe-GAN: A generative model for cancer prognosis prediction. *Bioinformatics*, 2019.
- [5] Lee J. et al. SurvivalNet: A deep learning-based model for cancer survival prediction. *the IEEE International Conference on Bioinformatics and Biomedicine*, 2019.
- [6] Wu Q. et al. DeepPFS: A multi-modal deep learning framework for predicting cancer patient survival. *IEEE Transactions on Medical Imaging*, 2020.