用于急诊科结局预测的多模态感知器语言模型

Sabri Boughorbel
卡塔尔计算研究所，哈马德·本·哈利法大学，卡塔尔

Fethi Jarray
LIMTIC 实验室，突尼斯大学，突尼斯，突尼斯；梅德宁高等计算机科学学院，加贝斯大学，突尼斯

Abdulaziz Al Homaid
卡塔尔计算研究所，哈马德·本·哈利法大学，卡塔尔

Rashid Niaz
医学信息学部 Sidra Medicine，
卡塔尔

Khalid Alyafei
Sidra Medicine 医学信息学部，
卡塔尔

arXiv:2304.01233v1 [cs.CL] 3 Apr 2023

摘要

语言建模在生成引人入胜、准确率高且语义连贯性强的文本方面取得了令人瞩目的进展。一个有趣的研究方向是利用上下文信息，增强这些强大的模型，使其适用于特定应用。在本研究中，我们探索了面向医疗保健应用的多模态语言建模。我们关注基于分诊时记录的主诉文本信息和生命体征，对医院急诊科中的结果进行预测并开展患者分诊。我们对 Perceiver 进行了改造；这是一种模态无关的基于 Transformer 的模型，已在多个应用中展现出良好前景。由于生命体征模态采用表格格式表示，我们修改了 Perceiver 的位置编码，以确保排列不变性。我们使用包含 12 万次就诊的 MIMIC-IV ED 数据集，评估了多模态语言模型在诊断代码预测任务中的表现。实验分析表明，与仅基于文本或生命体征训练的模型相比，多模态能够提升预测性能。我们确定了多模态能够改善模型性能的疾病类别，并证明生命体征对这些类别具有额外的预测能力。通过分析交叉注意力层，我们展示了多模态对模型预测的贡献。本研究为医疗保健应用中多模态语言模型的开发提供了有趣的见解。

1 引言

大语言模型近期在面向通用用途的人工智能应用方面取得了显著进展 Ouyang 等人（2022）；Adiwardana 等人（2020）。模型架构、大规模文本数据集以及精心设计的模型训练方案，是推动这一领域快速发展的关键因素。一项重要的活跃研究方向是将语言模型扩展至其他模态。近期，文本生成图像方面已取得令人瞩目的成果

Ramesh 等人（2022）。语言、图像、视频和音频信息的结合，也是多模态语言模型的热门研究领域 Baevski 等人（2022）。在医疗保健领域，电子健康记录（EHR）收录了多模态且互补的患者信息 Acosta 等人（2022）。电子健康记录收集自由文本临床笔记、数值型实验室检查结果，以及医学影像、结构化诊断编码、可穿戴设备数据或基因组诊断信息等其他类型的模态。将多模态方法应用于医疗保健领域，极有可能加速人工智能工具在临床实践中的应用。在下文中，我们将简要概述多模态方法，并说明本研究所选择方向的依据。

多模态方法可分为早期（输入）、中期（特征）和后期（决策）融合 Ramachandram 和 Taylor（2017）；Xu 等人（2022）。早期融合会将输入数据合并为单一表示，再输入深度网络。这类方法的主要缺点是，后续分类层无法区分不同模态。中期融合会先分别处理每种模态以提取特征，然后将不同模态的特征合并为中间表示。可通过拼接、求和和注意力机制等多种方式实现。Baevski 等人提出了一种基于 Transformer 的框架 Data2vec，用于从语音、视觉和语言数据中学习 Baevski 等人（2022）。每种模态都会转换为输入序列。图像采用 ViT 策略编码 Dosovitskiy 等人（2020），语音使用多层一维卷积神经网络编码，文本则通过字节对编码进行分词。模型采用自监督方式训练，并遵循与 BERT 类似的掩码策略 Devlin 等人（2018）。Data2vec 的缺点是计算成本高。后期融合方法会分别处理不同模态，然后在最后几层合并输出。最终决策可通过多数投票或平均等简单策略得出 Deng 等人（2020）。当不同模态相对独立时，这种方法很有用，也因此降低了信息丢失的风险。

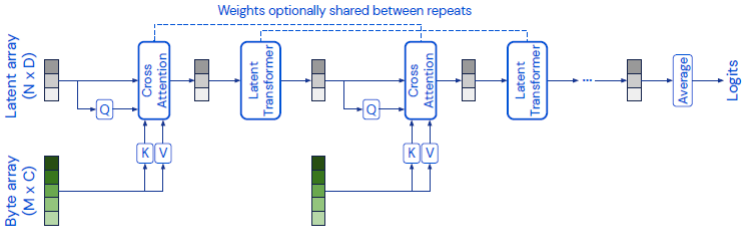
Perceiver 被提出作为一种模态无关 Transformer，用于解决自注意力中的计算瓶颈 Jaegle 等人（2021）。如图1所示，Perceiver 引入了一个固定大小的潜在向量 N。Perceiver 中的模态融合采用混合方式。输入数据在网络的不同阶段进行融合。因此，模型在网络深度方向上有多次机会，从各种模态中提取额外信息。交叉注意力层利用潜在数组对大小为 M 的输入数据进行注意力计算。该操作计算开销较低，并使模型深度与输入数据大小相互解耦。本研究旨在将 Perceiver 应用于急诊科（ED）就诊患者的诊断编码预测任务，同时利用主诉记录中的文本信息以及患者状态和生命体征的数值信息。在急诊科就诊过程中尽早预测患者结局，有助于更好地规划临床医护人员配置，并改善急诊科医疗需求的管理。

2 方法

我们使用 Perceiver 作为核心模型，将文本数据和生命体征整合起来，用于分类任务。生命体征以数值数组表示，文本数据则以在 Perceiver 训练过程中学习得到的嵌入表示。对于每次就诊，我们将这两种表示拼接起来，并输入 Perceiver 的不同模块，如图1所示。下文将介绍我们使用的数据集，以及如何在 Perceiver 中组合多模态。通过不同的实验和消融研究，我们展示了两种模态（文本、生命体征）在预测任务中的作用。

2.1 数据集

我们使用 MIMIC-IV ED 数据集开展实验 Johnson et al. (2021)。我们纳入了包含 ICD-10 编码的就诊记录。由于分诊信息非常有限，我们预测不带小数位的高级别ICD编码。我们专注于预测主要诊断，并舍弃次要诊断。为使数据能够代表模型训练所需的数据分布，我们将分析范围限定在数据集中最常见的前50位编码 Kaur et al. (2021)。急诊科 ICD 编码预测所用的文本数据包括主诉和入院时记录的生命体征。主诉是患者陈述的就诊原因，以简短的自由文本形式呈现。数据经过预处理，包括删除标点符号、将字母转换为小写以及分词。表格数值数据包括 6 项生命体征（体温、心率、呼吸频率、氧气



Fig图 1：Perceiver 架构。绿色的输入数据在不同深度通过交叉 -
atte注意力块进行融合（图片源自 Jaegle 等人（2021））。

血氧饱和度、收缩压和舒张压），以及护士评估的两个变量：疼痛程度，按 1 至 10 分计分；病情严重程度，按 1 至 5 分计分，其中 1 表示最严重，5 表示最轻。

2.2 模型设置与训练

由于我们应用中的文本较短，因此选择了大小为 16 的紧凑嵌入。此外，我们希望保持表格数据和文本数据的表示大小平衡。生命体征以一个 8 维数组表示，且不经过额外处理。文本和生命体征的表示拼接后，作为 Perceiver 模型的输入。模型深度设为 4，因此交叉注意力块和潜在变换器重复四次。潜在数组的大小设为 16。不同模态在每个交叉注意力块中融合。在实验中，我们分析了交叉注意力层，以考察不同模态在预测中的作用。我们将包含 12 万次就诊的数据集划分为训练集和测试集。Perceiver 模型训练 50 个训练轮次，直至交叉熵损失趋于稳定。表中呈现的所有结果均在测试集上获得。模型训练使用训练集和测试集的随机划分，重复进行 5 次。表中报告了不同数据随机化下各项指标的均值和标准差。

3 结果与讨论

我们研究了位置编码在 Perceiver 表格模态中的作用。表 1 对比了在表格模态（生命体征）中使用和不使用位置编码训练的 Perceiver。后者确保了排列不变性，并带来了更好的结果。排列不变的变换器在表格数据上表现良好（Dash 等人，2022）。

模型	Prec	Rec	F1	Acc	AUC
Perceiver-WPE	**74±0.35**	**50±0.28**	**60±0.13**	**43±0.13**	**96±0.04**
Perceiver-PE	73±0.21	48±0.78	57±0.37	39±0.26	94±0.31

表 1：使用傅里叶特征的 Perceiver，在表格模态（生命体征）上不使用位置编码（WPE）与使用位置编码（PE）的对比。数值为微平均结果。

在后续实验中，我们对生命体征采用不使用位置嵌入的 Perceiver。表 2 总结了使用不同模态预测主要诊断代码的结果。结果显示，与仅使用文本特征或生命体征相比，结合主诉记录和生命体征后，所有指标（精确率、召回率、F1分数、准确率、曲线下面积）的预测性能均有所提升。

为更好地理解各模态的作用，我们按疾病类别（ICD 编码）对性能结果进行了分层。表3展示了多模态模型与单模态（文本）模型之间的性能差异。脓毒症、心房颤动或脑梗死等疾病的性能提升幅度最大。这可能是因为生命体征和文本中记录的信息具有互补性，有助于诊断这些疾病。既往研究量化了生命体征症状作为这些疾病诊断特征的作用 Di Somma等 (2010);

模态	Prec	Rec	宏平均 F1	Acc	AUC
文本+生命体征	**66±0.27**	**38±0.36**	**48±0.28**	**32±0.34**	**94±0.10**
Text	60±1.17	29±0.84	39±0.97	25±0.76	92±0.43
生命体征	39±1.72	16±1.67	21±1.54	12±1.30	80±0.28
模态	Prec	Rec	微观 F1	Acc	AUC
文本+生命体征	**74±0.35**	**50±0.28**	**60±0.13**	**43±0.13**	**96±0.04**
Text	76±0.14	43±0.61	55±0.51	37±0.48	95±0.23
生命体征	46±0.21	27±0.42	32±1.24	21±1.31	71±0.4

表 2：单模态与多模态的性能比较。该任务是对 ICD 编码进行多分类。指标包括精确率、召回率、F1 分数、准确率和 ROC 曲线下面积。指标采用宏平均和微平均。

	AUC 差异	ICD-10	描述
主要编码	5.6	A41	脓毒症
	5.4	E87	体液和酸碱平衡紊乱
	4.0	R20	皮肤感觉障碍
	3.9	I48	心房颤动和扑动
	3.1	I63	脑梗死
底部编码	0	D64	其他贫血
	0	J45	哮喘
	0	T78	不良反应，未在其他地方分类
	0	R45	涉及情绪状态的症状和体征
	0	S72	股骨骨折

表3：多模态Perceiver（文本+生命体征）与纯文本模型在宏观AUC方面的性能差异。结果按疾病类别分层。

Hunt (2022); Rienstra等 (2012); Greenlund等 (2003)。表3底部的编码对应贫血、哮喘、骨折或不良反应等疾病。除哮喘外，这些疾病通常不会依据测得的生命体征进行诊断。对于哮喘，我们预期血氧饱和度应具有更高的预测贡献。这或许可以用入院时症状和生命体征的变异性来解释。图2以可视化方式展示了负责融合两种模态的交叉注意力层。图2-a显示，整体而言，文本模态的激活程度高于生命体征。图2-b揭示了体温生命体征对Perceiver预测发热相关急诊入院的贡献。

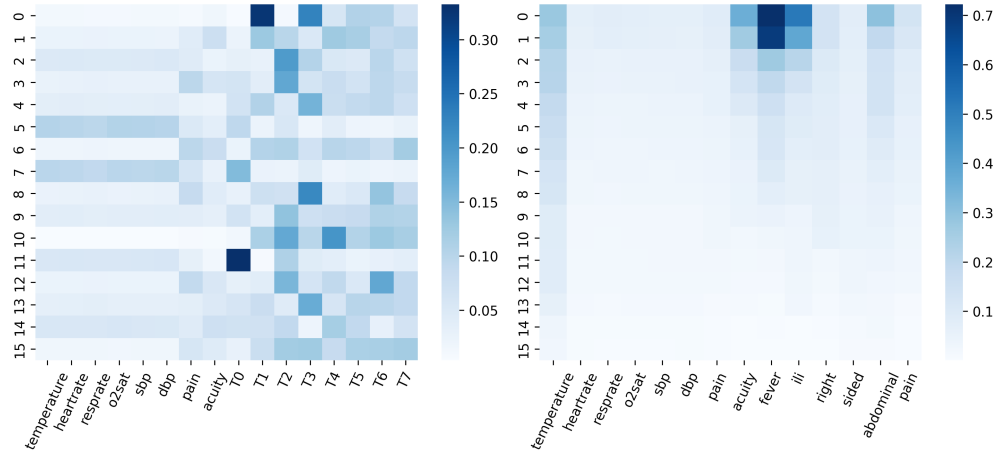
4 结论

我们研究了一种多模态方法，用于从文本和表格数据中学习，以预测医院急诊科的健康结局。该方法基于 Perceiever 模型，可实现模态的早期融合和中期融合。实验结果表明，与单模态方法相比，多模态方法具有有效性。未来，我们计划使用更大的数据集（如MIMIC-III重症监护病房）来扩展语言模型的规模。我们还计划探索微调预训练大语言模型的方向，并为其扩展其他模态，如表格数据或图像数据。

参考文献

Juli´an N Acosta、Guido J Falcone、Pranav Rajpurkar 和 Eric J Topol。多模态生物医学人工智能。《自然医学》，28(9):1773–1784，2022。

Daniel Adiwardana、Minh-Thang Luong、David R So、Jamie Hall、Noah Fiedel、Romal Thoppilan、Zi Yang、Apoorv Kulshreshtha、Gaurav Nemade、Yifeng Lu 等。《迈向类人开放域聊天机器人》。*arXiv* 预印本 *arXiv:2001.09977*，2020。



(a) 第 4 个（最后一个）
Perceiver 模块的平均交叉注意力矩阵。T0-T7 表示词元位置。

(b) 一次包含以下内容的急诊就诊数据的交叉注意力矩阵：
ICD 编码 R50（发热）。就诊文本记录为“发热、流感样疾病、右侧腹痛”。ILI 表示（流感样疾病）

图2：（a）在生命体征、疼痛和病情严重程度中，生命体征的贡献最大。文本模态具有较强的交叉注意力，但词元位置没有明显模式。（b）此次就诊中伴有发热症状的主诉。

Alexei Baevski、Wei-Ning Hsu、Qiantong Xu、Arun Babu、Jiatao Gu 和 Michael Auli。《Data2vec：语音、视觉和语言自监督学习的通用框架》。载于国际机器学习大会，第 1298–1312 页。PMLR，2022。

SarthakDash、SugatoBagchi、Nandana Mihindukulasooriya 和 AlfioGliozzo。采用变换器编码器进行表格理解的置换不变策略。载于计算语言学协会：*NAACL 2022*论文集，第788–800页，2022年。

Yifan Deng、Xinran Xu、Yang Qiu、Jingbo Xia、Wen Zhang 和 Shichao Liu。用于预测药物相互作用事件的多模态深度学习框架。生物信息学，36(15):4316–4322，2020年。

Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。BERT：用于语言理解的深度双向变换器预训练。*arXiv*预印本 *arXiv:1810.04805*，2018年。

Salvatore Di Somma、Chiara Serena Gori、Tommaso Grandi、Marcello GiuseppeRisicato 和 Emil- iano Salvatori。急诊科的液体评估与管理。液体超负荷，164:227–236，2010。

Alexey Dosovitskiy、LucasBeyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly 等。图像胜过千言万语：用于大规模图像识别的变换器。*arXiv* 预印本 *arXiv:2010.11929*，2020。

Kurt J Greenlund、Linda J Neff、Zhi-Jie Zheng、Nora L Keenan、Wayne H Giles、Carma A Ayala、Janet B Croft 和 George A Mensah。公众对中风主要症状的认知度较低。《美国预防医学杂志》，25(4):315–319，2003。

Anne Hunt。脓毒症：体征、症状、诊断、治疗和病理生理学概述。《急诊护士》，30(5)，2022。

Andrew Jaegle、Felix Gimeno、Andy Brock、Oriol Vinyals、Andrew Zisserman 和 Joao Carreira。《Perceiver：通过迭代注意力实现通用感知》。发表于国际机器学习大会，第 4651–4664 页。PMLR，2021。

Alistair Johnson、Lucas Bulgarelli、Tom Pollard、Leo Anthony Celi、Roger Mark 和 S Horng IV。Mimic-iv-ed。*PhysioNet*，2021。

RajvirKaur、Jeewani Anupama Ginige 和 OliverObst。基于出院摘要的自动 ICD 编码与分类系统：系统性文献综述。*arXiv preprintarXiv:2107.10652*，2021。

Long Ouyang、Jeff Wu、Xu Jiang、Diogo Almeida、Carroll L Wainwright、Pamela Mishkin、Chong Zhang、Sandhini Agarwal、Katarina Slama、Alex Ray 等。通过人类反馈训练语言模型以遵循指令。*arXivpreprint arXiv:2203.02155*，2022。

Dhanesh Ramachandram 和 Graham W Taylor。深度多模态学习：近期进展与趋势综述。*IEEE* 信号处理杂志，34(6):96–108，2017。

Aditya Ramesh、Prafulla Dhariwal、Alex Nichol、Casey Chu 和 Mark Chen。利用 CLIP 潜变量进行分层文本条件图像生成。*arXivpreprint arXiv:2204.06125*，2022。

Michiel Rienstra、Steven A Lubitz、Saagar Mahida、Jared W Magnani、Jo˜ao D Fontes、Moritz F Sinner、Isabelle C Van Gelder、Patrick T Ellinor 和 Emelia J Benjamin。心房颤动患者的症状与功能状态：现状及未来研究机遇。《循环》，125(23):2933–2943，2012。

Peng Xu、Xiatian Zhu 和 David A Clifton。基于变换器的多模态学习：综述。*arXiv* 预印本 *arXiv:2206.06488*，2022。