

大模型原理与技术课程设计

姓名：程 挺 学号：SX2524015

一、选题理由

（一）领域重要性

从功能磁共振成像（fMRI）等非侵入性脑信号中解码自然语言文本，是脑机接口（BCI）研究领域的核心挑战，同时也是揭示语言系统神经基础的关键途径。该研究不仅能够深化人类对大脑语言处理机制的理解，还能为实用化脑机接口的开发提供重要支撑，在医疗康复、人机交互等领域具有广泛的应用前景。例如，可为语言功能障碍患者提供新的交流方式，实现大脑与外部设备的直接语言交互。

（二）现有技术局限性

现有框架通常将整个 fMRI 序列一次性处理，忽略了人类大脑在语言理解中采用的分段式和归纳式处理策略。随着输入序列长度增加，会出现内存过载和语义偏移问题，导致解码性能显著下降，难以满足长文本解码的需求。此外，fMRI 文本解码任务是从人类语言理解过程中产生的神经活动模式中重建文本内容，与传统机器翻译的双模态一对一映射范式存在本质区别。现有针对跨模态翻译的模型无法直接迁移应用，亟需设计契合该任务特性的解码框架。fMRI 数据具有复杂的时空结构且个体差异较大，现有基于掩码自编码器（MAE）的 fMRI 表征学习方法，未考虑不同时间帧对应的文本语义重要性差异，导致学习到的表征难以有效捕捉关键文本信息。fMRI-文本配对数据集规模有限，使得同时优化 fMRI 表征学习模块和解码模块面临巨大挑战，现有方法在数据利用效率和泛化能力上仍有不足。

（三）研究可行性

理论基础支撑：认知科学领域已证实，人类语言理解是增量处理与分段整合协同作用的过程，这为设计脑启发式解码框架提供了坚实的理论依据，可将人类语言处理机制融入模型设计。

技术发展赋能：大型语言模型（LLMs）的快速发展为开放词汇表的 fMRI 文本解码提供了强大的语言生成能力，而掩码自编码器等自监督学习范式则为解决数据稀缺问题、提升 fMRI 表征质量提供了有效技术路径。

数据集保障：现有 Human Connectome Project（HCP）、Narratives、Huth 等公开数据集，涵盖了大规模 fMRI 数据和配对文本数据，为模型的预训练、微调及性能评估提供了充足的数据支持。

二、相关工作对比

（一）与传统解码范式的对比

如 UniCoRN、BP-GPT 等主流方法，仍普遍采用“一步式全序列”机器翻译框架（即将完整 fMRI 映射为全文本，或以 fMRI 为 prompt 激活 LLM），这类方法未能充分考虑人脑认知分段、语义归纳机制，往往在长句子、长文本解码中表现显著劣化，且跨模态对齐易丢失细粒度语义。本方法的分段-递进-归纳架构彻底打破全局粗粒度映射，实时动态处理局部信息、分布式整合长距离依赖，保障了模型内存消耗与语义贯穿的双重优化。

（二）与对比学习、自监督 fMRI 特征学习方法的对比

传统 MAE 或对比学习仅关注空间或时序建模，掩码机制随机或均匀分配，未

针对与语言语义的强相关时刻或空间点做针对性优化，导致最终 fMRI 表示对文本还原能力有限。本研究提出“文本引导的掩码策略”，显著提高了 fMRI 嵌入语义丰富度，并兼容自监督+微调，从而突破小样本配对数据的瓶颈。

三、贡献与创新点

(一) 核心贡献

受人类语言处理认知机制的启发，提出了一种新颖的 fMRI 到文本解码框架，结合了增量处理和基于总结的语义整合策略，命名为 CogReader。如图 1 (b) 所示，首先将连续的 fMRI 时间序列划分为多个连续的片段。对于每一个片段，模型进行增量解码，实时逐词生成对应文本。本研究还设计了一个总结整合模块，将当前片段的解码结果总结为语义表示。然后，将该表示作为先验知识传递，用以指导下一个片段的解码，从而实现有效的跨片段信息流。此外，为了学习包含更多文本信息的 fMRI 特征，引入了文本引导的掩码策略，并将其整合到基于掩码自编码器(MAE)的 fMRI 表示学习框架中。主要贡献总结如下：

1. 设计了一种新的 fMRI 到文本解码框架，该框架整合了增量处理和总结语义整合。模型能够实现每个片段的实时解码，并逐步融合跨片段的知识，为从神经活动中解码长文本提供了有效的解决方案。
2. 提出了一种文本引导的掩码策略。通过利用关键语义词的注意力分布，本方法选择性地掩码相应的 fMRI 时间点，并结合 MAE，鼓励模型关注关键时间点的神经活动，从而学习包含更多关键信息的脑表征。
3. 大量实验表明，本方法在标准的 fMRI 到文本解码基准上显著优于现有的最先进方法。此外，随着句子长度的增加，性能优势愈发明显，这进一步验证了受认知启发的解码框架的可行性和有效性。

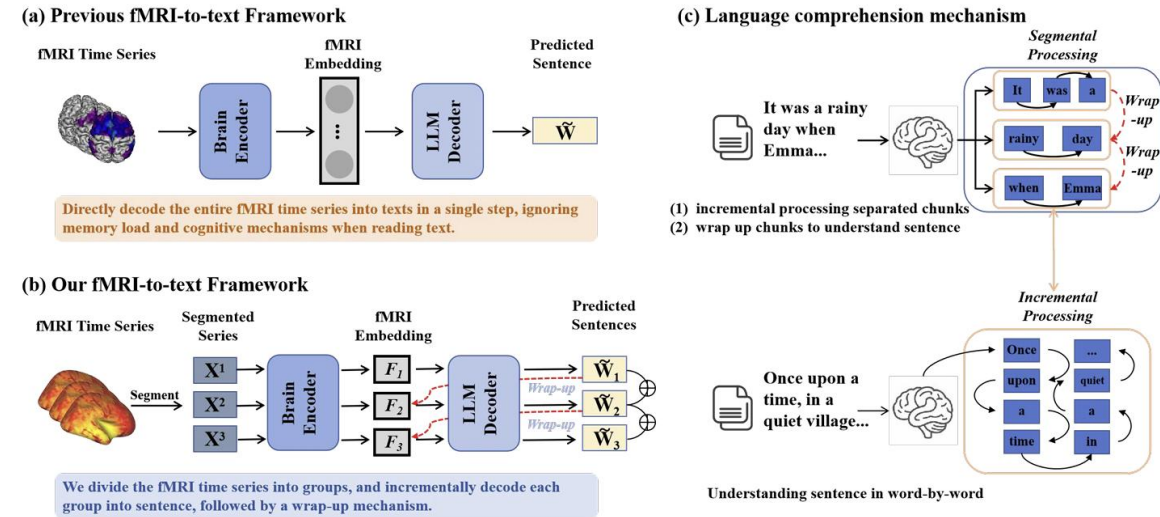


图 1: fMRI 到文本解码框架的比较。(a) 现有的框架在一步中直接解码对应于目标文本的整个 fMRI 序列。(b) 提出的基于分段的顺序解码框架。(c) 人类语言理解的认知机制，其中增量处理和分段总结并行进行。

(二) 创新点

解码框架的脑启发创新：首次将人类语言理解的“增量处理 + 分段整合”双机制融入 fMRI 文本解码框架。通过将长 fMRI 序列划分为适配人类语言处理的最优长度片段，逐段增量解码生成文本，并通过语义整合模块 (wrap-up) 将

当前片段语义总结作为先验知识传递给下一段，既缓解了内存压力，又保障了跨片段的语义连续性。

表征学习的语义引导创新：提出文本引导掩码策略，突破了现有 MAE 方法随机掩码的局限。利用预训练 BERT 模型计算文本词的语义重要性得分，将其映射到对应的 fMRI 时间帧，对高语义重要性帧采用更高掩码比例，引导模型重点学习语义关键区域的神活动特征，使 fMRI 表征与文本语义更紧密对齐。

训练与解码的协同优化创新：实现了表征学习与解码策略的深度协同。两阶段表征学习为解码提供了富含语义信息的 fMRI 嵌入，而分段式解码框架则充分利用了表征的时序特性和语义关联性，通过增量解码与语义整合的循环迭代，持续优化文本生成质量，形成“表征 - 解码 - 语义反馈”的闭环优化机制。

四、实验结果

将所提出的 fMRI 到文本的方法与四种先进的方法进行了比较：UniCoRN、EEG-Text、BP-GPT 和 PREDFT。

与本方法类似，UniCoRN 采用了一个两阶段解码框架，包括 fMRI 表征学习和随后 fMRI 到文本的解码。EEG-Text 最初是为 EEG 信号生成任务设计的。本方法使用 fMRI 数据重新训练了其编码器，以使其适用于 fMRI 解码。BP-GPT 是一种基于提示的解码方法，通过对比学习将 fMRI 表征与文本嵌入对齐，从而引导文本生成。PREDFT 是一种端到端的 fMRI 到文本解码模型，能够联合建模神经解码和脑预测编码。

在三种不同的文本解码长度下进行了比较，分别对应于 20TR、40TR 和 60TR 的 fMRI 时间序列。Narratives 数据集的比较结果见表 1 和表 2。

Length	Method	BLEU-N(%)				ROUGE-1(%)			BERTScore(%)		
		BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-F	ROUGE-P	ROUGE-R	BERTScore-F	BERTScore-P	BERTScore-R
20TR	UniCoRN	22.9	2.5	0.3	0	20.3	19.6	21	43.9	44.2	42.8
	EEG-Text	24.6	9.3	4.4	1.9	21.9	21.1	23.4	44.6	43.9	45.4
	BP-GPT	21.6	3.8	2.5	1.7	21.6	20.9	23.4	44.1	42.1	46.3
	PREDFT	24.3	4.2	0.7	0.1	20.1	22.3	18.3	45.9	45.5	46.7
	CogReader(ours)	25.4	10.5	4.7	2.6	23.4	22.6	24.6	46.3	45.7	46.9
40TR	UniCoRN	19.1	2.3	0.5	0.1	17.8	18.2	17.6	43.8	44.8	42
	EEG-Text	20.1	7.3	3	1.3	24.4	25.1	24.7	45.4	45.8	45.5
	BP-GPT	19.9	3.6	2.3	1.5	21.1	19.3	22.9	42.6	39.4	46.1
	PREDFT	25.9	4.8	1.4	0.4	21.1	24.8	18.6	46.3	46.2	46.8
	CogReader(ours)	31.2	15.3	10.3	8.2	29.6	28.7	30.4	50	49.3	51.1
60TR	UniCoRN	18	1.7	0.2	0.4	16.5	15.9	17	43.2	43.7	42.7
	EEG-Text	22.1	8.2	3.4	1.6	28.1	29.4	28.1	47.7	47.8	47.7
	BP-GPT	19.3	3.4	1.3	0.6	19.4	19.6	19.3	41.6	38.2	45.3
	PREDFT	26.4	6.1	1.9	0.6	28.1	25.5	20.5	48.1	47.7	48.5
	CogReader(ours)	36.2	20.4	14.7	12.1	36.2	35.6	37.2	53.5	52.6	54.5

表 1. 在 Narrative 数据集上，不同文本解码长度下本方法与其他方法的比较

表 2. 基于 Narrative 数据集的 fMRI 文本解码案例分析。目标句子与预测句子完全一致的部分以加粗显示，而语义相似性则以斜体显示

Results
Target: <i>was I gave the figure a name I called him Pie Man capital P capital M And I described him as a cape wearing masked avenger Though in fact he'd been capeless And said that as he fled the scene he clicked his heels in rakish glee And I gave him a catchphrase in Latin I said that he cried out Ego sum non an bestia which means I am not an animal Which makes no sense</i>
UniCoRN: <i>and me gray to the little Man and no taller than my knee, with a coat too big for his frame lingered by the gate. It's been nigh of years he to the then said a case here. The leather sat weathered on the step, its lock rusted shut, but I knew what was inside: the girl of the story, the one who scribbled tales of dragons in the margins of her notebooks, had tucked her most precious drawings there before she</i>
EEG-Text: <i>and this is where it gets really interesting I'm trying to think of a title for the book and I finally come up with an idea that sounds like something out of one of those old pulp stories by John Wilmot who lived in Waukegan Illinois but he was actually born in Westchester County New York which is about twenty five miles outside Chicago So I am going to describe my novel as The Man Who Walks into Times Square and stops before he can get too lost</i>
BP-GPT: <i>I'm a teacher and I teach people how to make things better by using my own talents And so one day I was walking down the street and this old lady came up to me and she said You know you sound like an intelligent black woman So I told her that's exactly what I'm here to do</i>
PREDFT: <i>the a said the girl a man I said the Man and use of best and she me and I man fl of Litgo be and the It's flirting and she guy of the eyes thelee raised non crate and she littleiving it then to of crate non a best owan is best me best to</i>
Ours: <i>I think that you realize what happened next Pie Man emerged from the late night library drop made his delivery and fled away crying Ego sum non an bestia Or that's what it said in my story in the newspaper next day which ran with photos of him leaving the scene cape flowing behind him doing this And I'm just like praying my life doesn't flash before my eyes and ruins</i>

量化比较如表 1 所示，本方法在所有解码长度和评估指标上始终优于所有其他的方法，展示了提出的类脑框架的整体有效性。从词汇级指标（BLEU-N 和 ROUGE-1）的角度来看，随着解码文本长度的增加，SOTA 方法的性能会下降，而本方法则呈上升趋势。在语义级评估（BERTScore）方面，SOTA 方法表现相对稳定，而本方法随着解码序列的增加仍在不断提升。在 60TR 时间窗口下，本方法实现了明显优于其他方法的文本解码性能。

解码文本比较如表 2 所示，本方法在语义信息提取和叙事一致性方面始终优于其他方法。例如，模型能够准确识别并生成关键短语，如“Pie Man”、“Ego sum non an bestia”、“scene”、“said”和“fled”，同时生成更连贯、更符合上下文的句子。相比之下，其他方法在目标内容上偏差更大。这些结果表明，随着解码长度的增加，本方法的优势愈加明显，突出了所采用的类脑顺序解码策略的有效性和可行性。