**Representation learning for neural population activity with Neural Data**

**Transformers**

**Abstract**

神经集群活动是可以用状态空间模型（如RNN）来展示

循环卷积如RNN有缺点（会降低实时应用的效果）

所以提出了NDT 一个非循环卷积的替代

通过使用一个猴运动皮层的数据集，数据集包含一个reaching task，可以被RNN很好的衡量（即可以横向对比），两个同时建模

测试了其捕捉自治动力系统的能力，并表现出了更快的能力

**Introduction**

[1]神经集群活动理论上拥有一个随着时间驱动的潜在的动态模型

可以由线性、switching线性动态系统，或者非线性动态系统（RNN）表示

这种建模方法可以将独立的一个trial中的神经元活动和行为联系起来

例如一个LFADS的RNN方法表现得很好，并且可以到达毫秒级精度

1. RNN被应用于语言的建模，可以捕捉语言结构，但是他的性能不行，介入转向了Transformer。T是平行输入参数，而RNN拥有时序，这个让他更加快速。减少时间很重要。
2. 介绍NDT，用来给神经集群spike活动建模。基于修改后（因为spike和其他序列数据还是有区别的）的BERT编码。相比其他机器学习文章，神经数据集也小，需要调整训练策略。
3. 同时测试了人造和真是数据集。自己使用动态系统生成FR并从中采样，NDT和LFADS都可以从采用后的spike回复出未被观察到的部分。当使用真实数据集（后脑运动皮层），NDT比LFADS更快。
4. 结果表明循环卷积并不必须，同时打开了新的范式

**NDT model**

1. 同时使用NDT和LFADS处理按bin包装的spike数据序列，实时处理使用滑动窗，两个模型均假定是泊松的：预测出来的发放率将和观察到的进行泊松似然估计（用NLL来衡量参数）。然后介绍模型结构：LFADS会拥有一个序列模型解码器像是RNN，前一个状态会影响后一个状态，而NDT则是同时处理所有输入，它拥有一些非线性处理模块（图2a），其self-attention（SEC6.1）会联合其他其他输入的信息。
2. NDT在该研究中是一个6层编码器，输入会被映射到不同通道里，研究中只设定了两个通道，计算NLL之前，输入线形层和求幂，损失函数：交叉熵->泊松似然。
3. （2b）无监督的训练模型，使用了BERT的masked modeling建模方法:掩盖掉比如20%的连续序列要求模型进行重建，此时需要通过上下文进行重建（比如发放率平稳时，高发放率会引发高发放率）（类似于dropout）
4. 训练策略的调整：
5. BERT：使用特殊的掩盖方法，将被掩盖的部分直接归0（证明过有效）
6. 数据集规模小，采用 intensive regularization,dropout 0.2-0.6 (SEC3.4)

**Results**

比较了生成数据和M1 reaching activity的表现

介绍了两个编码器的超参数和方法

NDT：grid search；checkpoint: least validation NLL

LFADS：AutoLFADS模型； PBT(population-based training) : least smoothed validation NLL checkpoint

每一个都选择了20个模型,,每个模型都参与了三个研究的训练

AutoLFADS只是个baseline 没有特意调参

3.1 **The NDT achieves high-fidelity inference on synthetic autonomous dynamical systems** NDT在人工数据上具有高准确度

测试了两个合成数据集 the Lorenz system and the chaotic RNN

L数据集是根据洛伦兹方程生成的、高维投射的、发放速率数据，然后根据泊松分布抽样生成spikes。

C数据集通过权重由正态分布初始化的vanilla RNN生成。它比L更复杂，因此更有挑战性。

两个模型都有很好的表现，但是NDT会更不平滑一些，使用 判定系数 来衡量。但是，在HP search（超参数搜索中），NDT发生了AutoLFADS没有发生的结果：NDT（with high data likelihoods）可以很好的匹配潜在的系统。说明了likelihood（似然估计）是个很好的工具

3.2 **NDT infers motor cortical firing rates in autonomous settings with high fidelity NDT**可以高精度推断真实数据集

Monkey j Maze数据集，reaching task 202神经元 2296trials 108个不同条件下

动作开始前的250ms和动作开始后的450ms，通过10ms画bin,尝试过2-20ms的bin都有类似的结果。

计算PSTHS( peri-stimulus time histograms ),两个模型在相同条件下的不同trail基本都有稳定的表现。为了量化表现，计算了一个使用高斯核的30ms标准差的经验PSTHS，并计算两个模型之间的 判定系数。

也可以通过运动解码来评估。采用最优线性估计来映射发放率到手的速度上，同时合成轨迹，两个模型拥有类似的结果。

降低训练集大小，6-NDT依旧比baseline表现好，虽然在更小的时候（92）比另一个模型差,但是依旧展示了一个2-NDT消除了这个缺陷

**3.3 Efficiency Gains from Parallelism** 有效的信息获取，通过平行输入的方式

推断速度，平行输入，能推断出长度和推断速度之间的关系，大致在3.9ms水平，随着输入长度变长慢慢增长，而LFADS增长的非常迅速，在长度70bins（of 10ms）的时候，能到达6.7倍的速度差距，跟之前的工作比起来，比如使用20bins of 15ms 也能达到4倍的速度。理论上，这种对比是建立在非迭代过程的循环卷积上的，就是说采用类似卡尔曼滤波这样的方式去迭代，同时有逐步更新的Internal state。

**3.3.1 NDT结构简化**

NDT的结构可以缩短推理和训练的时间。但是六层结构相比于L方法确实会花费大量时间训练，简化结构在保持良好的效果的同时也能显著降低训练时间。

在有限制的数据集配置时（比如更小），更简化的模型可能会有更好的效果。

3.4 Ablative Analysis

验证训练策略：巨大影响的三个因素：log rate、 zero task、dropout

也验证了其他策略，但是收效不明显

4 Discussion

验证了NDT在神经数据集大小上的可行性

Limitation 不能表征时滞动力学，比如那些具有不可预测的扰动，比如不受监控的脑区给记录脑区发送了信号，而LFADS在这方面拥有很好的结果（在某篇研究中20%优于）但依旧相信她很好。

比如，RNN在长短时检测中，长时间特征很难被学习到，而T因为其平行输入的特点则可以学习到，映射到神经活动上，快特征和慢特征的学习，解释层次结构。同时它可以以用于多模态的模型建立，比如建立不同脑区之间的不同记录模态和行为测量。

T更受欢迎的原因是它对大规模数据的处理更优，。。。

6 Methods

**6.1 模型细节**

6.2 超参数

6.3 合成数据集

6.4 运动解码

6.5 时间测试

**Pyramidal Cell-Interneuron Circuit Architecture and Dynamics in Hippocampal Networks**

In vivo 体内的

Inhibitory neuron 抑制性神经元

Neural Syntax: Cell Assemblies, Synapsembles, and Readers

Transiently 短暂地

Underlie 构成...的基础

Numerous 许多的

Formation and disbanding 形成和分裂

Facilitate 使促进

Hierarchical 分等级的

Constituent 成分

Constellation 一群的

Segregated 隔离的

Paradox 矛盾

Inherent 内在的

Exemplify 是...的典范

Coherent 有条理的

Orchestration 编排

Distinct 不同的

Cognitive 认知的

Sufficient 充分的

Putative 推断的 假定的

Presumption 假定

In light of 根据，鉴于...的观点

Derive from 源于，继承

In turn 反过来，轮流

In correspondence with 与...相对应

From Parametric Representation to Dynamical System: Shifting Views of the Motor Cortex in Motor Control

Dogmas 教条

Anatomically 解剖结构学的

Primates 灵长类

Isometric 等距的

Foster 促进，培养

Reconcile 调和

Integration 融合

Doctrine 教条

Innervate 支配

Causal 因果关系的

Nonetheless 尽管如此

Elusive 难以理解的

Induce 引发，引诱

Intrinsic 内在的 固有的

Prominent 突出的

Substantial 大量的

Heterogeneity 异质性的

Conjunctive 连接的

Interpret 解释

Oscillating 震荡

Spinal cord 脊髓

Corticospinal tract 皮质脊髓束（CS）

Effector 感受器

In line with 与...一致

Have provider a deeper insight into 拥有更深的见解

初级运动皮层 M1