

# Exploring Cognitive Variables in the HCP Dataset : Predictive Modeling and Inter-Variable Relationship Analysis

马俊程, 屈泽凯, 魏昕原

## 1 研究背景

认知神经科学的一个基本目标是描述大脑解剖、功能和行为之间的关系，研究脑特征的个体差异与个体行为之间的关系是神经科学的基础。

个体化行为预测建模利用机器学习方法，基于大脑的神经影像数据对行为指标进行估计。由大脑神经影像数据得到的功能连接和结构连接均与认知功能相关，前者代表通过功能磁共振成像（fMRI）测量的区域血氧水平依赖性信号之间的时间依赖性模式，后者代表通过扩散 MRI 估计的区域间白质通路的完整性，二者被广泛用于正常人的认知行为预测任务 [3, 5, 6]。HCP 数据集 [7] 是研究健康人个体化预测使用最广泛的数据集之一，提供了高分辨率的神经影像数据和丰富的行为指标。

对人的行为预测在精准医学、青少年教育等多个领域都有潜在应用，如预测个体治疗的结果以指导治疗选择与剂量，对不同神经疾病进行临床亚组分类，以及在发育阶段预测未来的认知能力和心理健康等。同时，行为预测模型可以提供客观的生物标志物，让我们了解认知行为背后潜在的大脑机制。

## 2 核心问题

我们小组基于 HCP 数据集的功能数据和结构数据，为 13 个认知类变量建立预测模型，并探索认知任务之间在脑层面的潜在联系。首先，我们训练核岭回归模型评估每种输入特征是否可以很好地预测个体认知行为；其次，我们尝试结合不同的连接数据，观察是否可以改善预测；最后，我们选择最显著的预测模型，计算特征重要性，并可视化认知任务间的相似度以进行模型解释，分析认知任务间在脑层面的关系。

## 3 方法

### 3.1 数据

我们使用的是提供的 HCP 数据集，预测变量选择了“Cognition”类别的 13 个变量，输入特征分别用以下缩写表示

- FC1(753\*79800):rsfc\_atlas400\_753\_4.txt
- FC2(753\*79800):rsfc\_Yeo400\_753\_GSR.txt
- SC(753\*79800):scfp\_atlas400\_753.txt
- HC1(753\*159600): FC1+SC
- HC2(753\*159600): FC2+SC(+ 号表示连接)

## 3.2 模型

我们使用核岭回归 (KRR) 模型基于不同的输入特征训练行为预测模型, 因为它在几篇相关工作 [2-4] 中表现出了强大的行为预测能力。简单来说, KRR 模型假设输入特征相似的个体表现出更相似的行为, 利用核函数度量相似性, 并引入  $\lambda^2$  正则化项降低过拟合。

具体来说,实验使用 Pearson's correlation 作为核函数,输入特征分别为 FC1,FC2,SC,HC1 和 HC2,以 Pearson's correlation 和 Coefficient of determination(COD) 作为精度指标,以性别为回归协变量。每个回归模型都进行了 30 次 3-fold 嵌套交叉验证,外圈循环用来降低偶然性,内圈循环使用网格搜索确定超参数。

为了进一步探索结合不同类型的特征是否会带来更好的预测准确性，我们还使用了多核岭回归模型 [2, 4]，模型设置与单核岭回归类似，输入特征分别为 FC1+SC, FC2+SC, FC1+FC2, FC1+FC2+SC。

### 3.3 统计检验

为了评估显著性，我们对每个预测模型进行了置换检验。首先，将预测变量值随机打乱 100 次，分别训练和测试模型以得到预测模型性能的零分布；其次，将原始模型 30 次训练/测试中的预测精度与零分布进行比较；最后，统计预测精度小于或等于 100 个零模型性能的原始模型个数，计算比例记为  $p$  值。

为了控制多重比较，我们对所有统计测试进行了  $q < 0.05$  的错误发现率 (FDR) 校正 [1]，以评估模型的性能是否优于偶然性。

### 3.4 模型解释

我们选择了效果最佳的预测模型（基于 FC2）进行模型解释，通过 Haufe 变换计算每个认知任务的预测模型的特征重要性，这已经在 [2, 4] 中表现出很好的效果。简而言之，对于给定的预测模型，计算每个特征与每个受试者的预测行为得分之间的协方差，正的特征重要性得分表明脑特征与预测行为值之间较高的相关性。

上述过程得到了每个认知任务的 79800 维特征重要性向量，为了通过认知任务进一步分析脑连接层面的潜在关系，我们计算了所有认知任务的特征重要性之间的相关性，并将相关性矩阵可视化。

### 3.5 代码

我们使用 Matlab 进行实验，使用 python 画图，代码已随报告提交，也可见🔗，README.md 文件对代码复现进行了详细的说明。



图 1: README.md

## 4 结果与讨论

### 4.1 实验结果

我们优化了置换检验的代码，因此显著性结果与 *PPT* 中有一定区别

基于 **FC2(经过 GSR 处理)** 的 **KRR 模型**可以显著预测认知变量 使用 KRR 模型在不同输入特征 (FC1, FC2, SC, HC1, HC2) 上训练并测试，以 Pearson's correlation 为精度指标，对 13 个认知变量的预测结果如图2所示，标红圈表示该模型预测显著优于随机模型（错误发现率 FDR  $q < 0.05$ ）。

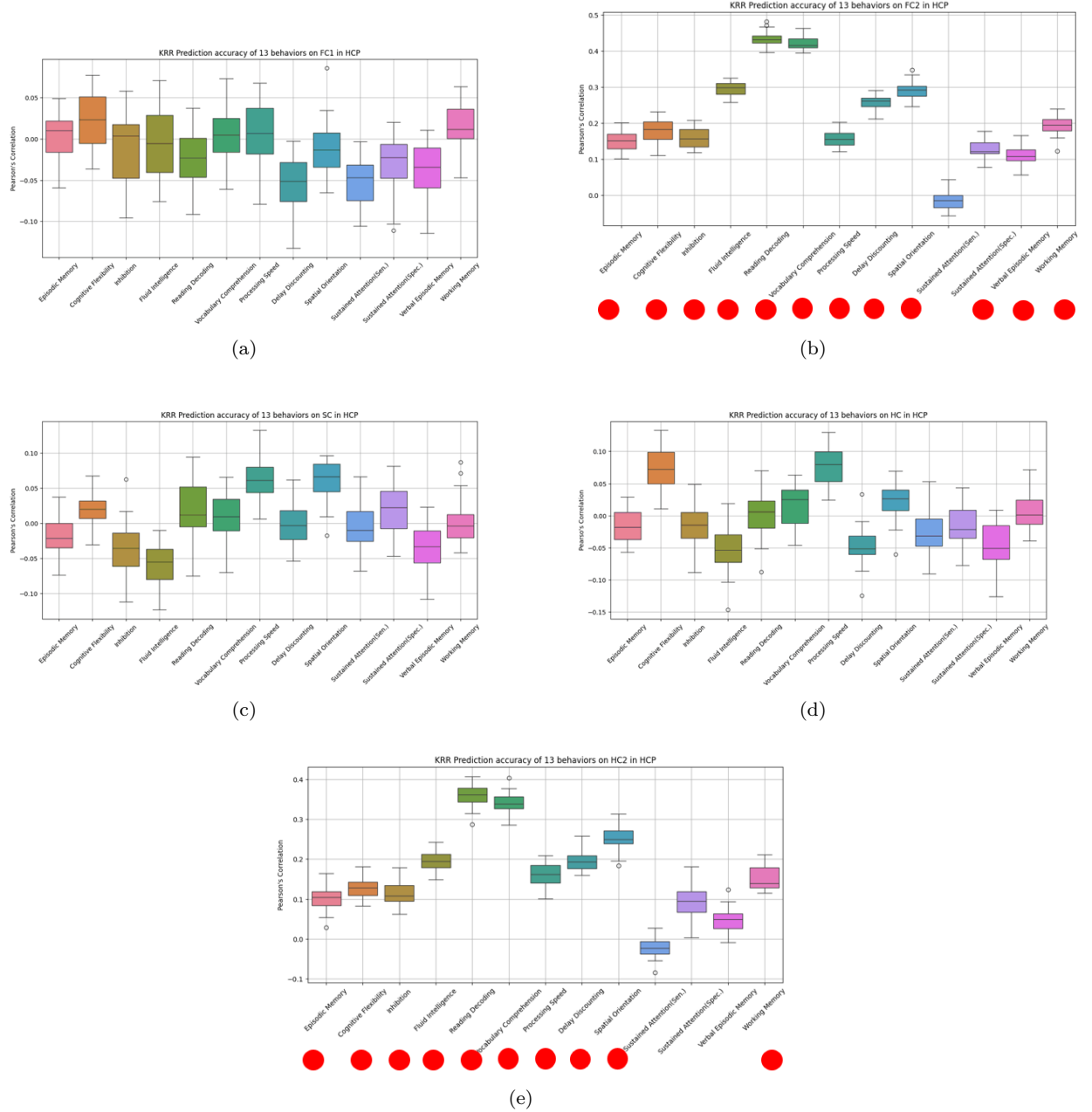


图 2: (a)FC1; (b)FC2; (c)SC (d)HC1; (e)HC2

可以发现，基于 FC2 的 KRR 预测模型在 12 个认知变量上实现显著预测，我们认为这是由于 GSR 处理有效降低扫描过程中由呼吸、心跳等生理因素引起的噪声，从而提高数据的信噪比，并且减少大脑不同区域间因

全球信号波动而产生的伪造相关性；Reading Decoding 和 Vocabulary Comprehension 两个认知任务的预测效果最好；输入 HC2 (FC2+SC) 特征时，在部分任务上实现显著预测，但是我们认为这主要归功于 FC2 特征，SC 特征的加入相较于单纯的 FC2 是负提升。

对 FC2 预测模型的置换检验结果如图3所示，可以看出对大部分认知任务有  $p < 0.005$  (FDR  $q < 0.05$ )，模型预测显著性很强。

	1	2	3	4
TASK	Episodic Memory	Cognitive Flexibility	Inhibition	Fluid Intelligence
P-Value	<b>0.002999</b>	<b>0.000664</b>	<b>0.001999</b>	<b>0.000333</b>
TASK	Reading Decoding	Vocabulary Comprehension	Processing Speed	Delay Discounting
P-Value	<b>0.000333</b>	<b>0.000333</b>	<b>0.002999</b>	<b>0.000333</b>
TASK	Spatial Orientation	Sustained Attention (Sen.)	Sustained Attention (Spec.)	Verbal Episodic Memory
P-Value	<b>0.000333</b>	0.623126	<b>0.011663</b>	<b>0.026991</b>
TASK	Working Memory			
P-Value	<b>0.000666</b>			

图 3: FC2 的单核 KRR 结果的置换检验结果

我们还以 Coefficient of determination(COD) 作为精度指标进行了一组实验，结果类似，见附录7.1。

**多核 KRR 模型略提升预测效果** 使用多核岭回归模型结合不同特征进行训练和测试的结果如图4所示。相较于部分 KRR 模型 (FC1,SC 和 HC1)，多核岭回归的性能有一定提升，在”Proceeding speed” 任务上均实现显著预测 (FDR  $q < 0.05$ )。

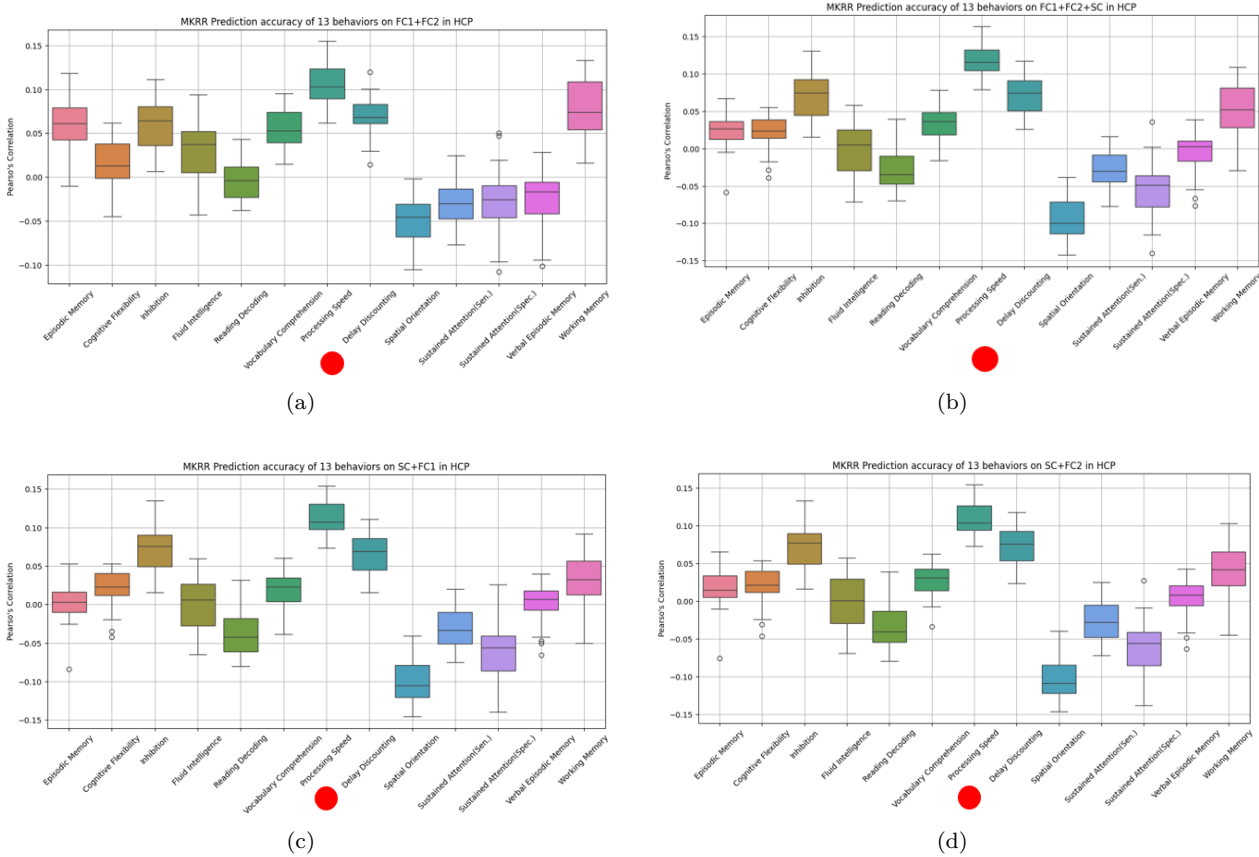


图 4: (a)FC1+FC2; (b)FC1+FC2+SC; (c)SC+FC1; (d)SC+FC2

将四种多核岭回归模型在所有认知任务上的精度做平均，性能对比如图5(a)所示，可以看出结合 FC1 和

FC2 的模型性能略佳。将效果最好的多核岭回归模型与 KRR 模型放在一起对比，精度均为所有认知任务上的平均值，结果如图5(b)所示，可以看出多核岭回归方法虽相较部分 KRR 模型略提升了效果，但与基于 FC2 和 HC2 的 KRR 预测模型精度差距仍很大。

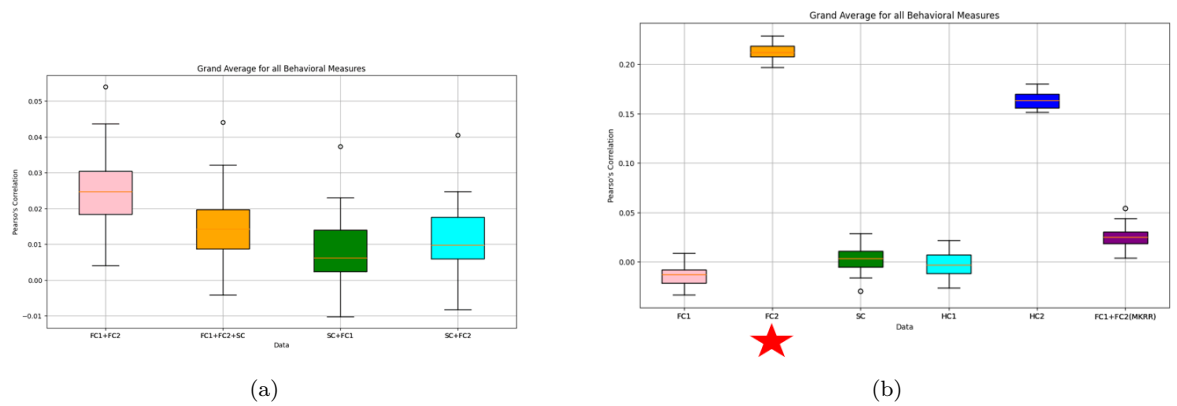


图 5: (a) 多核 KRR 性能; (b) 多核与单核对比

**预测变量间在脑连接层面存在潜在联系** 我们选择了预测显著性最强的预测模型（基于 FC2 的 KRR 模型），对每个认知任务进行 Haufe 变换，计算特征重要性 (Feature Importance, 后以 FI 表示)，得到 13\*79800 维向量，再计算行向量间的相关性以衡量响应相似度，得到了一个 13\*13 相似度矩阵，如图6所示。我们认为，两个认知任务间的 FI 相似度越高，对 79800 维特征中某些位置特征的响应越相似，这种响应相似性在某种程度上可以反映脑连接层面的潜在联系，例如激活共同的功能脑区。因此，根据相似度矩阵，我们可以从神经生物学角度去分析不同认知任务背后生理调控、神经通路的关联性，从而得出认知行为相关性的生理解释。值得注意的是，这种解释是由神经影像结果推测得来，基于“不同认知任务间的响应相似性可能源自它们在大脑中的共同神经基础”的假设，需通过解剖学等方法进一步验证。

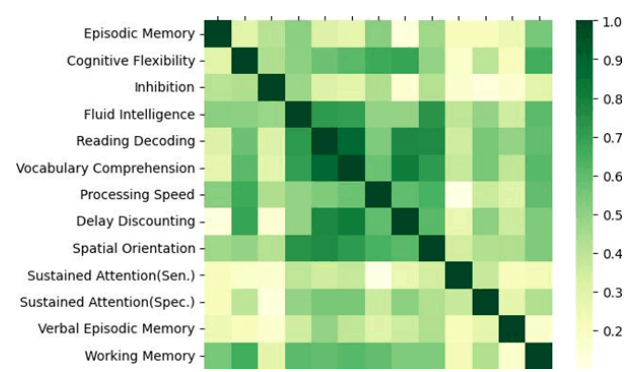


图 6: FI 相似度矩阵

如图7所示，“阅读解码”和“词汇理解”两种认知任务拥有最高的 FI 相似度，这两项任务都涉及语言处理与语义理解。我们认为，较高的 FI 表明大脑中存在负责文字解析和理解的脑区，在个体进行这两项任务时均处于活跃状态。根据已知的脑区功能，大脑的 Broca’s 区和 Wernicke’s 区主要负责语言的组织与理解能力，它们在左脑半球的颞叶和额叶中是特别活跃的，可能导致这两项任务在神经活动上会有较高的相似性。

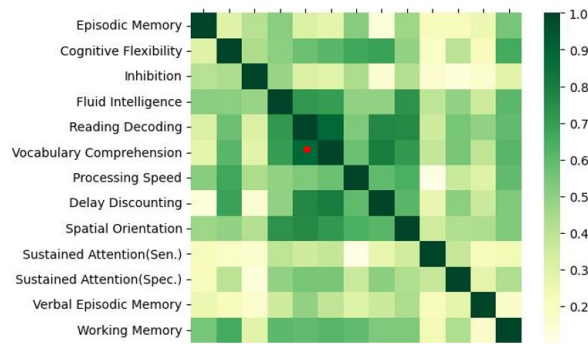


图 7: Reading Decoding vs Vocabulary Comprehension

如图8所示，“延迟折扣”和“词汇理解”、“阅读解码”任务分别具有第二、第三高 FI 相似度。虽然“延迟折扣”与其他两个任务在表面上看似不相关，一个涉及奖励价值的评估和决策，其余两个是语言理解的过程，但较高的 FI 表明它们可能都需要调用大脑中用于复杂认知过程的相同区域。我们认为，在进行“延迟折扣”任务时，解码文字和理解文字所传达的奖励信息都需要语义理解功能的参与，个体必须理解不同时间点潜在奖励的相对价值以做出决策，这需要对时间和价值的抽象概念有深刻理解，而这种理解与语言能力紧密相连。此外，决策过程中的抑制控制也与工作记忆有关，而工作记忆是“阅读解码”过程中不可或缺的，这也是两者 FI 相似度较高的神经学依据之一。

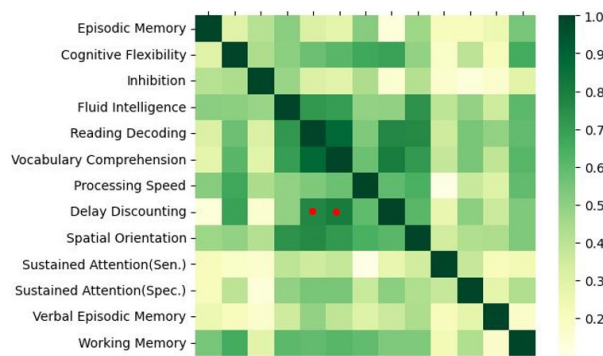


图 8: Reading Decoding vs Vocabulary Comprehension vs Delay Discounting

如图9所示，“情景记忆”和“延迟折扣”这一对任务的 FI 相似度最低，我们认为这反映了两个任务在功能区响应的差异。首先，“情景记忆”是关于回忆个人过去经历的认知过程，属于长期记忆，细分为陈述性记忆的情景记忆部分，它依赖于大脑中海马体以及相关的颞叶区域；而“延迟折扣”涉及对即时和未来奖励的评估，这是一个决策过程，更多地涉及前额叶，尤其是与奖励处理和决策制定相关的区域。其次，“情景记忆”强调对特定事件的细节和情境信息的回忆，主要与回忆和记忆重建相关，而“延迟折扣”则是一个涉及价值判断和自我控制的任務，更多涉及决策制定和自控力。这两个任务在神经通路上有较明显的差异，反映了大脑在执行记忆任务和决策任务时的功能专一性。



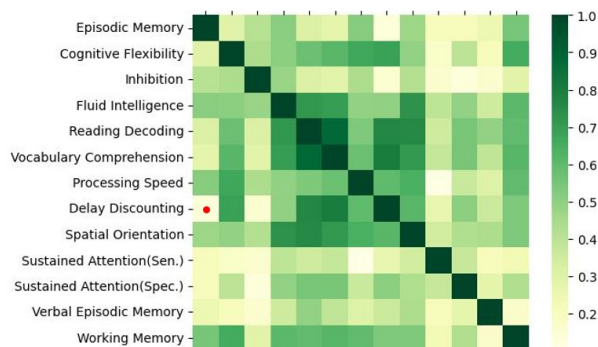


图 9: Delay Discounting vs Episodic Memory

## 4.2 讨论

在行为预测方面，[2] 在 ABCD 数据集上，利用静息 FC 和任务 FC 对行为进行预测，发现任务 FC 在认知行为预测上优于静息 FC，并且，当使用多核岭回归模型结合静息 FC 和任务 FC 时，效果得到明显提升；[4] 在 ABCD 和 HCP 数据集上，利用 fMRI 的多种数据形式进行行为预测，其中功能连接效果明显优于结构连接，也发现结合静息 FC 和任务 FC 明显提高预测性能；[3] 在 HCP 数据集上，利用静息 FC、结构连接和二者连接的 HC 对流动智力与固定智力进行预测，发现 HC 和 FC 的预测优于 SC。

我们发现经过 GSR 处理的静息 FC 可以很好的预测认知行为，这与 [2,4] 的结论一致，并且 FC 的效果大多优于 SC，与 [3] 的结论相符。但是结合 FC 和 SC 或两种静息 FC 未能像 [2,4] 结合静息 FC 和任务 FC 一样显著提升预测效果，我们认为可能有两方面原因，一方面，相比静息 FC 和任务 FC，静息 FC 和 SC 之间可能缺少对行为预测的互补性，无法实现效果提升；另一方面，相比我们目前使用的方法（多核岭回归，或直接连接），可能有更好的方法或模型结合静息 FC 与 SC 或两种静息 FC，实现协同预测。

在模型解释方面，由于我们不清楚 79800 维的 FC 数据如何恢复到  $400 \times 400$  上三角矩阵、亦不清楚特征进一步对应到脑图谱的哪一区域，我们没有像 [2-4] 等工作在脑图谱层面进行模型解释和分析，而是计算每个预测任务的特征重要性，进而分析脑层面的潜在关系。

## 5 未来展望

通过这次大作业，我们使用 KRR，多核岭回归等模型，基于大脑影像数据对认知行为进行建模，发现经过 GSR 处理的静息态功能连接可以显著预测 12 个认知行为，多核岭回归略提升预测性能，并且通过模型解释，发现一些认知任务具有较高的特征重要性相似度，可能在脑层面存在潜在联系，如涉及共同的功能脑区。

但是，我们的项目还有一定的不足，可以在以下方面在未来进一步探索

- 我们现在使用的多核岭回归模型结合多种输入特征未能明显实现效果提升，在未来，可以进一步探索是否有更合适的机器学习方法结合不同类型的特征，实现协同学习，提高预测精度。
- 我们的模型解释与分析由于缺少脑图谱信息，目前仅局限于行为层面，将特征重要性整体计算相似度，无法关注其中最相似特征对应的脑连接含义，进一步分析认知任务背后的神经机制。未来可以具体到脑图谱，将 79800 维特征重要性恢复到  $400 \times 400$  矩阵，再根据脑图谱进一步分析 13 个认知任务之间在人脑具体区域的响应相似度，以探索认知任务背后的脑连接机制。
- 站在机器学习的角度，基于 79800 维的功能/结构连接特征预测认知行为在小数据集的情况下是一个较病态的问题，数据在特征空间中极为稀疏。我们曾试图使用一些降维方法（如主成分分析，LDA 等），但这些经典机器学习方法会为模型解释带来困难。未来可以进一步探索认知神经学习领域的降维方法，在可解释的基础上，提取真正有判别性的特征，实现降维，相信可以大大提高认知行为预测的精确度。

## 6 小组分工

马俊程：方法调研，实现 Matlab 和 python 代码，跑出实验结果，制作 PPT 并汇报，完成部分实验报告并汇总。

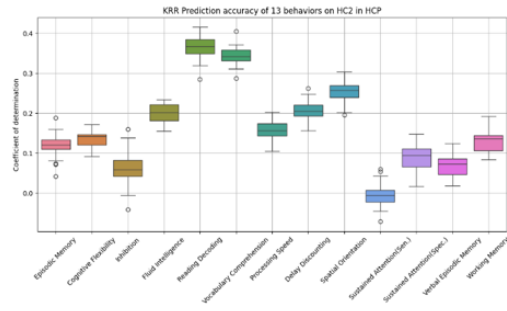
魏昕原：方法调研，处理实验数据并画图，制作部分 PPT，完成部分实验报告。

屈泽凯：方法调研，处理实验数据并画图，制作部分 PPT，完成部分实验报告。

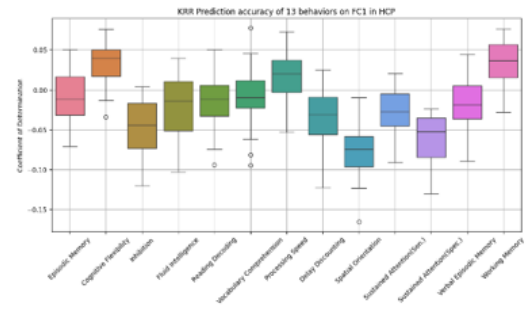
## 7 附录

### 7.1 补充实验结果

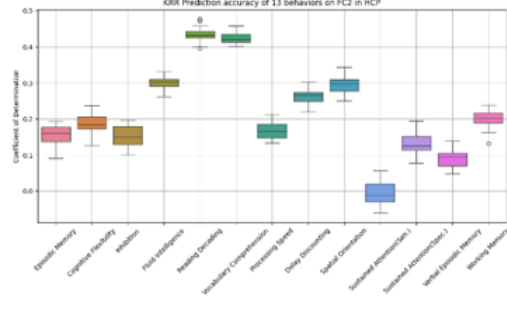
下面是以 COD 为精度指标的 KRR 模型结果，与报告中的结论一致。



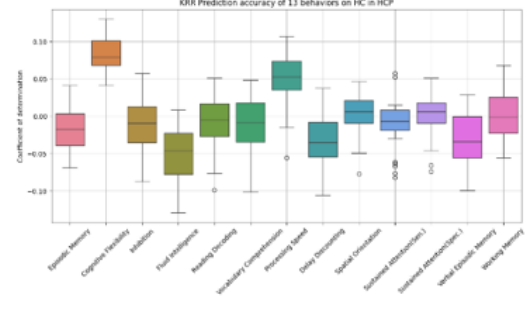
(a) KRR Prediction on HC2 in HCP



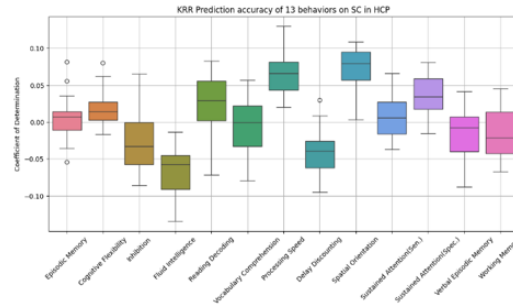
(b) KRR Prediction on FC1 in HCP



(c) KRR Prediction on FC2 in HCP



(d) KRR Prediction on HC in HCP



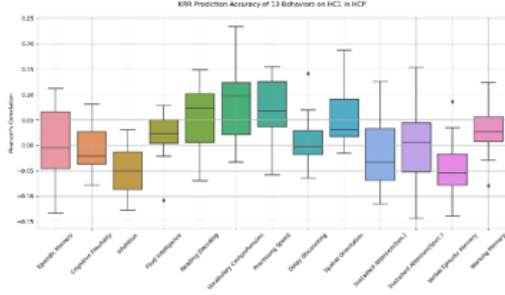
(e) KRR Prediction on SC in HCP

图 10: COD 为精度指标的结果

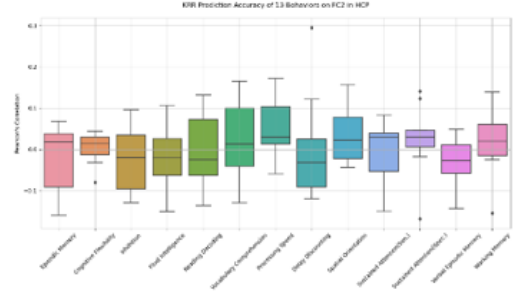


## 7.2 失败尝试

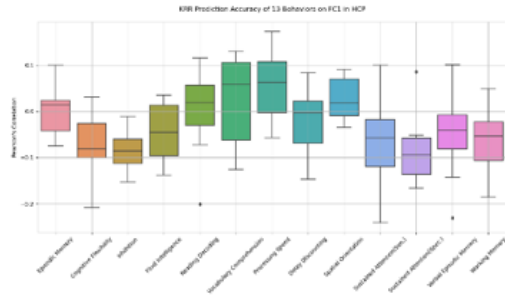
实验之初，我们使用 python，结合 sklearn 库，自主实现了整个预测流程，但是效果很不理想，过拟合严重，无论是改变超参数，还是更换模型（岭回归、KRR、SVR、KNN……），都无法改变这一状况，实验结果如下



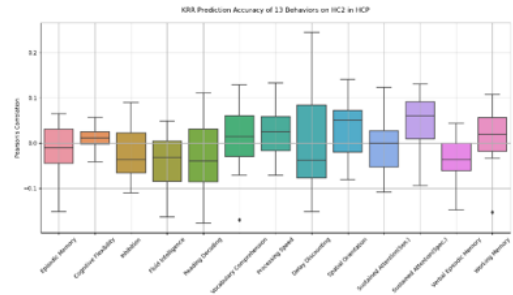
(a) KRR Prediction on HC1 in HCP



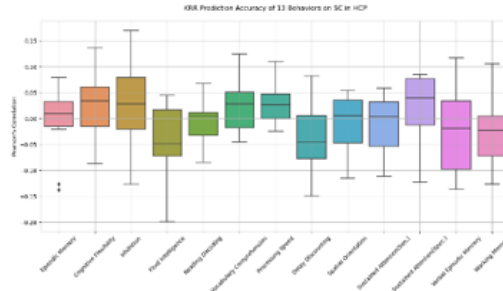
(b) KRR Prediction on FC2 in HCP



(c) KRR Prediction on FC1 in HCP

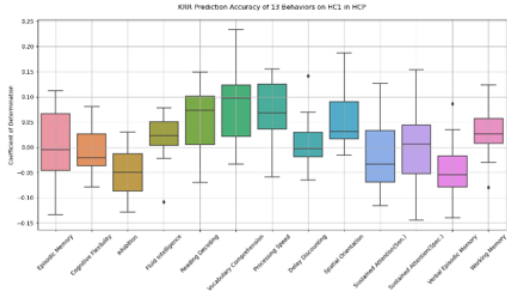


(d) KRR Prediction on HC2 in HCP

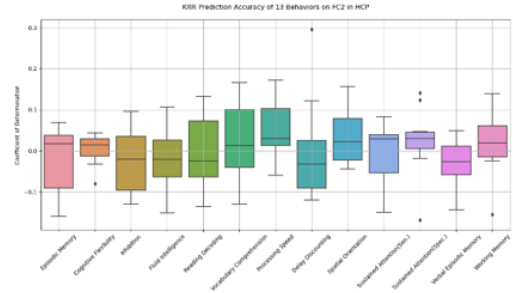


(e) KRR Prediction on SC in HCP

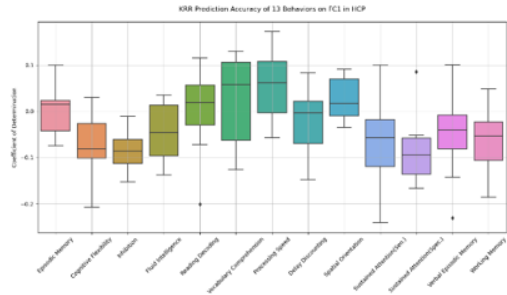
图 11: KRR 的 Pearson's correlation 结果



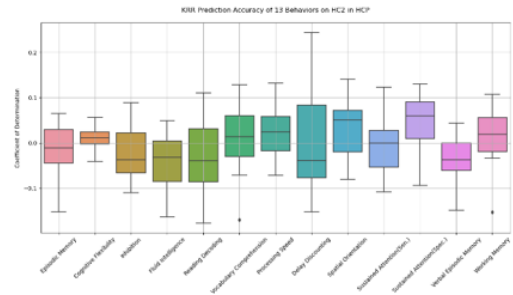
(a) KRR Prediction on FC1 in HCP



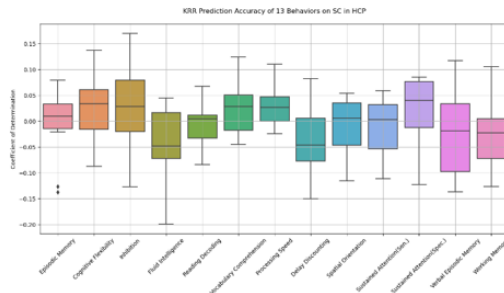
(b) KRR Prediction on FC2 in HCP



(c) KRR Prediction on FC1 in HCP

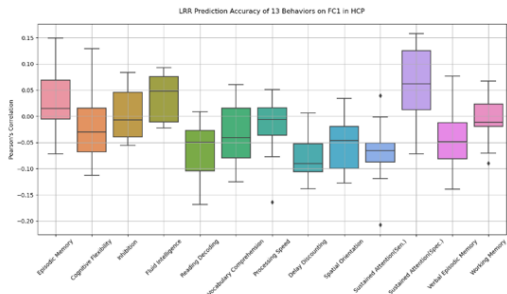


(d) KRR Prediction on FC2 in HCP

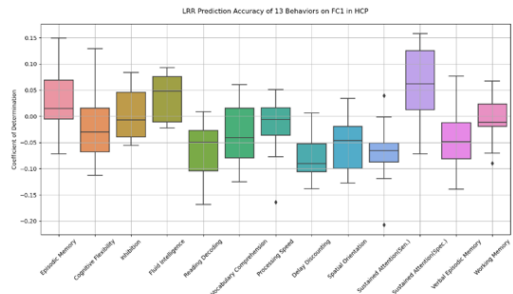


(e) KRR Prediction on SC in HCP

图 12: KRR 的 COD 结果



(a) LRR's Pearson's Correlation on FC1



(b) LRR's COD on FC1 in HCP

图 13: LRR 的结果

**经验总结** 经过我们分析代码，基于 python 的 sklearn 中的 KRR 模型，与基于 matlab 的 CBIG 库的 KRR 模型，主要有下面两个区别，可能导致效果的差异

- 后者对协变量（性别、年龄……）进行了回归，用统计学方法消除潜在的混杂因素（我们只回归了性别）。
- 后者使用的是相关性核函数，而 sklearn 中我们使用的是 RBF 核。

此外，我们使用 python 训练时，由于速度和时间限制，嵌套交叉验证对超参数的搜索范围较窄；而 CBIG 库的速度更快，得以搜索更广的范围，也可能是影响效果的原因之一。

## 参考文献

- [1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1):289–300, 1995.
- [2] Jianzhong Chen, Angela Tam, Valeria Kebets, Csaba Orban, Leon Qi Rong Ooi, Christopher L Asplund, Scott Marek, Nico UF Dosenbach, Simon B Eickhoff, Danilo Bzdok, et al. Shared and unique brain network features predict cognitive, personality, and mental health scores in the abcd study. Nature communications, 13(1):2217, 2022.
- [3] Elvisha Dhamala, Keith W Jamison, Abhishek Jaywant, Sarah Dennis, and Amy Kuceyeski. Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults. Human brain mapping, 42(10):3102–3118, 2021.
- [4] Leon Qi Rong Ooi, Jianzhong Chen, Shaoshi Zhang, Ru Kong, Angela Tam, Jingwei Li, Elvisha Dhamala, Juan Helen Zhou, Avram J Holmes, and BT Thomas Yeo. Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity mri. NeuroImage, 263:119636, 2022.
- [5] Javier Rasero, Amy Isabella Sentis, Fang-Cheng Yeh, and Timothy Verstynen. Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. PLoS computational biology, 17(3):e1008347, 2021.
- [6] Chandra Sripada, Saige Rutherford, Mike Angstadt, Wesley K Thompson, Monica Luciana, Alexander Weigard, Luke H Hyde, and Mary Heitzeg. Prediction of neurocognition in youth from resting state fmri. Molecular psychiatry, 25(12):3413–3421, 2020.
- [7] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. Neuroimage, 80:62–79, 2013.