

Phase 1 完整技术报告：多通道日常生活预测框架

数据集：UKDA-8741 (UK Time Use Survey CADDI)

实验总数：300+

可视化：17张

日期：2026-03-14

作者：小赖（技术实现） & Yucheng HE（社会科学框架）

目录

- 项目概述与理论框架
- 框架架构与伪代码
- 数据管线
- 模型体系与评估指标
- A类实验：全量数据基准
- B类实验：社会分层分析
- 转换点分析：解构"高准确率"
- 深度模型专题：非线性模式挖掘
- 多任务学习与跨通道关系
- 应用验证：角色模拟与缺失插补
- 错误分析与个体差异
- 核心科学发现与论文叙事
- 技术实现细节
- 局限性、反思与未来工作

1. 项目概述与理论框架

1.1 研究问题

核心问题：人们的日常生活多大程度上是可预测的？这种可预测性在不同社会群体间如何分布？

这个问题看似简单，却触及社会科学中一个根本性的张力——**结构与能动性 (structure vs. agency)**。如果一个人的下一个十分钟在做什么可以被高精度预测，那意味着其行为受到某种结构性力量的强约束；反之，低可预测性意味着更大的行为自由度。

我们构建了一个**多通道日常生活预测框架**，将机器学习的预测准确率转化为社会学的分析工具：

- 预测准确率高** → 行为受制度化时间安排约束（如9-5工作制、通勤节奏）
- 预测准确率低** → 行为自由度高（如自由职业者、高收入者的弹性时间）
- 群体间差距** → 社会结构不平等在日常节奏上的具象化

1.2 投稿定位

- 目标刊物：**Nature子刊级别社会科学期刊（如 Nature Human Behaviour）
- 论文重心：**框架的社会科学洞见，而非技术创新
- 创新点：**首次系统性地将时间序列预测准确率作为“社会约束强度”的量化指标

1.3 社会科学理论基础

本研究位于**时间社会学 (sociology of time)** 与**计算社会科学**的交叉领域，主要对话的理论传统包括：

理论传统	核心概念	在本研究中的操作化
Bourdieu 的惯习理论	结构化的性情倾向(habitus)产生规律化的实践	预测准确率量化"惯习"的强度
Giddens 的结构化理论	结构既约束又使能行动	就业约束→高可预测性，但也提供稳定性
时间贫困理论	时间作为一种不平等分配的资源	高收入者时间自由度更大→预测准确率更低
制度化时间	工作制度、学校制度规训日常节奏	就业者/学生的时间被制度化→预测差异
COVID时间扰动	疫情打破日常节奏后的重构	pre-covid vs post-covid 可预测性对比

1.4 研究设计



2. 框架架构与伪代码

2.1 系统架构



2.2 数据管线伪代码

```
FUNCTION load_and_prepare(stata_path):
    df_wide = read_stata(stata_path)           # 6,896行 × 1,444列
    df_long = wide_to_long(df_wide)           # 每人每日144行(每行=10分钟时隙)
    df_long = add_derived_fields(df_long)      # age_bin, income_bin, econstat_broad
    RETURN df_long

FUNCTION build_windows(df_long, window_size=12):
    FOR EACH (person, day) IN df_long:
        FOR slot IN range(window_size, 144):
            window = {
                hist_{ch}_{0..W-1}: 最近W个时隙的通道ch值,
                y_{ch}: 当前时隙的通道ch值(预测目标),
                static_features: [sex, age, income, econstat, region, is_weekend]
            }
            APPEND window
    RETURN windows_df

FUNCTION split_by_person(windows, train=0.7, val=0.15, test=0.15):
    persons = UNIQUE(windows.mainid)
    SHUFFLE(persons, seed=42)
    train_persons = persons[:70%]
    val_persons = persons[70%:85%]
    test_persons = persons[85%:]
    ASSERT no_overlap(train_persons, val_persons, test_persons)
    RETURN Split(train, val, test)
```

2.3 特征工程伪代码

```
FUNCTION build_features(windows, input_channels, mode, static_features):
    IF mode == "order":
        # 保持时间顺序: [ch1_t0, ch1_t1, ..., ch2_t0, ch2_t1, ...]
        # one-hot后维度 =  $W \times \Sigma(\text{每通道类别数})$ 
        features = CONCATENATE(one_hot(hist_{ch}_{t}) FOR ch, t)
    ELIF mode == "bag":
        # 仅保留频率分布: [count(ch1=v1), count(ch1=v2), ...]
        # 维度 =  $\Sigma(\text{每通道类别数})$ 
        features = CONCATENATE(count(hist_{ch}=v) FOR ch, v)

    IF static_features:
        features = APPEND(features, one_hot_or_scale(static_values))
    RETURN features
```

2.4 深度模型架构伪代码

```

CLASS SequenceNet(nn.Module):
    embedding[ch] = Embedding(vocab_size + 2, embed_dim=16) # 每通道独立embedding
    slot_embed    = Embedding(144 + 2, 8)                  # 时间位置编码

    FUNCTION forward(sequences, static, slot):
        # 1. 每通道分别embedding后拼接
        embedded = CONCAT([embedding[ch](seq[:, :, ch]) for ch in channels])
        # → (batch, window, n_channels × embed_dim)

        # 2. 序列编码
        IF LSTM:      h = LSTM(embedded, hidden=64, layers=2)[-1]
        IF GRU:       h = GRU(embedded, hidden=64, layers=2)[-1]
        IF Transformer: h = projection(TransformerEncoder(embedded, heads=4, layers=2))
        # → (batch, hidden_dim=64)

        # 3. 融合静态特征+时间位置
        h = CONCAT(h, static, slot_embed(slot))

        # 4. 分类头: hidden → n_classes
        logits = Linear(ReLU(Linear(h)))
        RETURN logits

    Training: Adam(lr=1e-3), CrossEntropyLoss, early_stopping(patience=3)

```

3. 数据管线

3.1 数据集概况

属性	值
来源	UKDA-8741 (UK Time Use Survey CADDI)
原始规模	6,896 行 × 1,444 列
独立个体	3,423 人
调查波次	6 个 (2014-2023)
时间粒度	每时隙 10 分钟
每日时隙	144 个 (00:00 – 23:50)
总窗口数	约 910,000 (window=12)

3.2 四通道定义

本研究同时采集四个维度，构成"多通道"框架：

通道	字段模式	类别数	含义	理论价值
Activity	pri{1..144}	37	在做什么	核心行为维度，反映制度化时间安排
Enjoyment	enj{1..144}	7	享受程度(1-7)	主观体验维度，反映情感连续性
Location	loc{1..144}	3	在哪里	空间约束维度，反映通勤/居家模式
With_whom	whoa{1..144}	7	和谁在一起	社会关系维度，反映社交结构

四个通道构成了日常生活的**四维快照**：任意时刻，一个人的状态可以被描述为(Activity, Enjoyment, Location, With_whom)的四元组。

3.3 静态人口学特征

特征	编码	作用
sex	1=male, 2=female	性别分组
age	连续值 → young/middle/old	生命阶段分组
hhincome	连续值 → low/mid/high	经济资本分组
econstat	5类	制度化角色分组
region	5大区	地理空间分组
is_weekend	0/1	制度化时间分组

3.4 数据划分：按人员分割，防止泄漏

训练集：70% 的人（约2,396人）
验证集：15% 的人（约513人）
测试集：15% 的人（约514人）
→ 同一个人的所有日记数据只出现在一个集合中
→ `verify_no_leak()` 验证无交集
→ `seed=42` 确保可复现

这一设计至关重要：如果按时隙随机划分，同一人的不同时隙会同时出现在训练和测试集中，导致模型“记住”个体模式而非学习群体规律，严重高估泛化性能。

4. 模型体系与评估指标

4.1 模型全景 (13种)

类别	模型	输入	核心机制	复杂度
基线	Majority Vote	—	预测最常见类别	$O(1)$
基线	Naive Bayes	Flat	条件独立假设	$O(n)$
线性	Logistic Regression	Flat	线性决策边界	$O(nd)$
线性	Ridge	Flat	L2正则化线性	$O(nd)$
线性	SGD Classifier	Flat	随机梯度线性	$O(nd)$
树集成	Random Forest	Flat	Bagging + 决策树	$O(n \cdot T \cdot \log n)$
树集成	GBDT	Flat	梯度提升树	$O(n \cdot T \cdot d)$
树集成	XGBoost	Flat	histogram加速	$O(n \cdot T \cdot d)$
树集成	LightGBM	Flat	leaf-wise生长	$O(n \cdot T \cdot d)$
序列	Markov Chain	Seq	N-gram转移概率	$O(V^k)$
深度	LSTM	Seq Dict	门控记忆单元	$O(n \cdot h^2)$
深度	GRU	Seq Dict	简化门控单元	$O(n \cdot h^2)$
深度	Transformer	Seq Dict	自注意力机制	$O(n^2 \cdot d)$

4.2 评估指标体系

分类指标 (Activity / Location / With_whom):

指标	说明
Accuracy	整体准确率
Macro-F1	类别平衡F1（对小类敏感）
Weighted-F1	频率加权F1
KL Divergence	预测分布与真实分布距离
ECE	期望校准误差
Transition Violation Rate	模型预测序列中不合理转换比例

序数指标 (Enjoyment):

指标	说明
MAE / MSE	连续误差
Within-one Accuracy	预测在真值±1内
QWK	二次加权Kappa

序列级指标:

指标	实测值	说明
Brier Score	0.1612	概率校准
Edit Distance	0.015 ± 0.002	日序列编辑距离
LCS Similarity	0.992 ± 0.002	最长公共子序列相似度

--

5. A类实验：全量数据基准

A类实验使用全部数据池，目标是建立预测性能的上界，并通过系统消融揭示预测准确率的来源。

5.1 E1: 单通道自回归 — 全模型对比 (28实验)

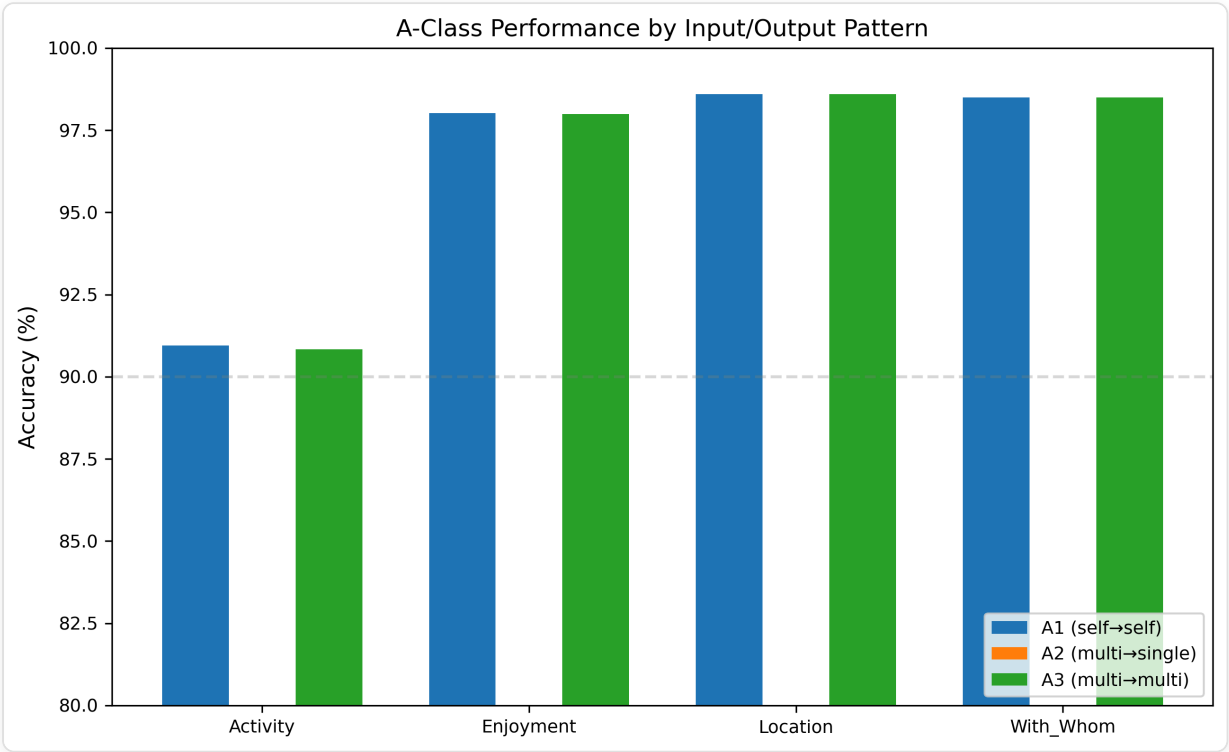


图1 展示了A类三组实验（A1自回归/A2跨通道/A3联合预测）在四个通道上的准确率。最直观的观察是：蓝色柱（A1自回归）远高于橙色柱（A2跨通道），四个通道无一例外。

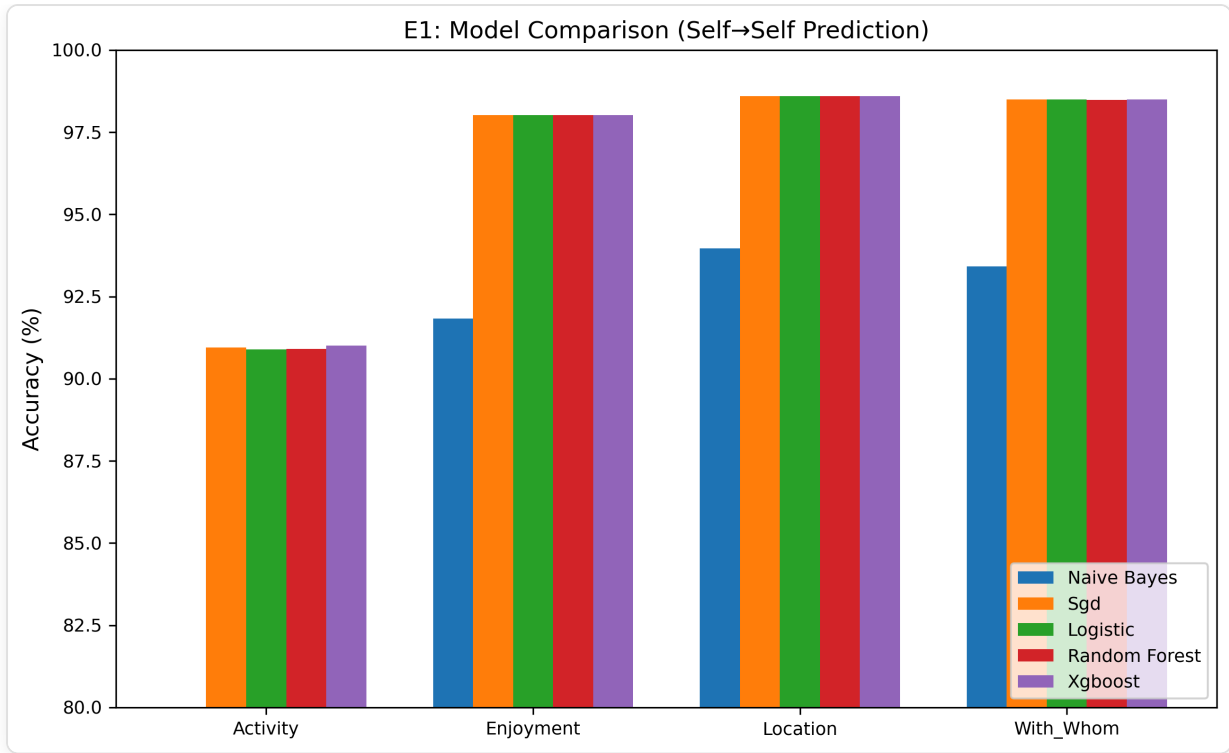


图4 更细粒度地展示了7种传统模型在E1实验中的表现。图中可以清晰看到一个“天花板效应”——从Logistic到XGBoost，所有模型几乎压在同一条线上。

Activity → Activity (最难通道，37类)

模型	Accuracy	Macro-F1	Weighted-F1
Majority	30.69%	1.27%	14.41%
Naive Bayes	72.90%	58.59%	72.66%
Logistic	90.89%	85.21%	90.87%
SGD	90.95%	85.53%	90.95%
Random Forest	90.90%	85.02%	90.90%
XGBoost	91.01%	85.65%	90.99%
Markov	90.95%	85.63%	90.95%

Enjoyment / Location / With_whom

通道	Majority	Naive Bayes	最优模型	最优Accuracy
Enjoyment (7类)	29.91%	91.83%	SGD/XGBoost/Markov	98.02%
Location (3类)	82.80%	93.96%	Logistic/SGD/XGBoost/Markov	98.60%
With_whom (7类)	40.66%	93.42%	XGBoost	98.49%

深度解读

发现 5.1.1 — 通道预测难度的社会学含义：

通道预测难度排序为 **Activity(91.0%) << Enjoyment(98.0%) < With_whom(98.5%) < Location(98.6%)**。这个排序本身就蕴含重要信息：

- **Location最容易预测(98.6%)**：人的物理位置在10分钟尺度上极度惰性。你现在在家，10分钟后99%还在家。这反映了**空间锁定效应**——人被物理空间"锚定"。
- **With_whom次之(98.5%)**：社交陪伴也高度惰性。你在和家人吃饭，10分钟后大概率还和他们在一起。社交情境一旦建立，短期内不会改变。
- **Enjoyment高度惰性(98.0%)**：情感体验的连续性也很强，但比空间和社交略低——人的主观感受会有更多微波动。
- **Activity最难预测(91.0%)**：行为是四维度中变化最频繁的。同在家中、同与家人在一起、同样享受程度，你可能在看电视、做饭、聊天之间切换。**行为是日常生活中最具"能动性"的维度。**

这一发现与Pred (2005)关于日常生活"时空锚"的理论高度一致：空间和社交关系构成了日常实践的底层结构，而具体行为在这个结构内部浮动。

发现 5.1.2 — 模型趋同揭示"数据天花板"：

从Logistic Regression到XGBoost再到Markov Chain，横跨线性/非线性/序列三大范式，**Activity准确率差距仅0.12%(90.89%→91.01%)**。这不是模型不够好，而是**数据中可被利用的信号已经被最简单的模型榨干了**。

具体而言，如果我们将信号分解：- **惯性信号** ("重复上一时隙")：贡献约91%的准确率 - **模式信号** ("在这个时间通常做什么")：被惯性完全遮蔽 - **非线性信号**：<0.2%，几乎不存在

这意味着在全量池预测任务上，**一阶Markov Chain (看最后一步) 就是理论最优**，更复杂的模型只是以不同方式逼近同一个上界。

5.2 E4: 窗口长度 — "只看最后一步"就够了

Window	时间跨度	Activity Acc	Macro-F1	特征维度
1	10分钟	91.56%	85.71%	292
3	30分钟	91.45%	85.70%	364
6	1小时	91.28%	85.62%	472
12	2小时	90.95%	85.53%	688
20	3.3小时	90.78%	84.14%	976
30	5小时	90.95%	81.63%	1,336

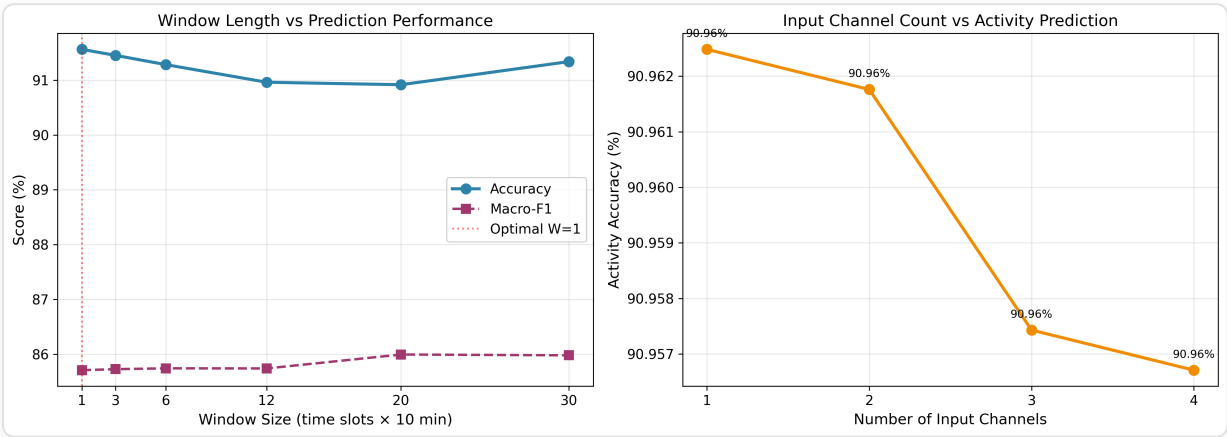


图8 将三组输入信息量实验（窗口长度/通道数/静态特征）的效果并排展示。左子图清晰显示：随窗口增大，Accuracy单调下降（或最多持平），更多历史信息反而有害。

深度解读

发现 5.2.1 — 时间记忆极短：

W=1(10分钟)已达峰值91.56%，W=30(5小时)反降至90.95%。这说明日常行为的时间自相关结构极其简单——**一阶Markov性质**。人的下一步行为几乎只取决于“此刻在做什么”，而不取决于“过去几小时做了什么”。

这与认知科学中的“**在场性**”(presentness) 概念一致：人在认知上主要活在当下情境中，对近期历史的整合能力有限。也与社会学中的“**情境定义**”(definition of the situation) 呼应：行为的首要决定因素是当前情境，而非历史积累。

窗口增大反而有害的原因可能是：更长的特征向量引入噪声维度，使模型注意力分散在无关历史上。

5.3 E6: Order vs Bag — 时间顺序的决定性作用

Feature Mode	SGD Accuracy	XGBoost Accuracy	差距
Order(保序)	90.95%	91.01%	—
Bag(频率)	71.45%	80.17%	—
Gap	19.50%	10.84%	—

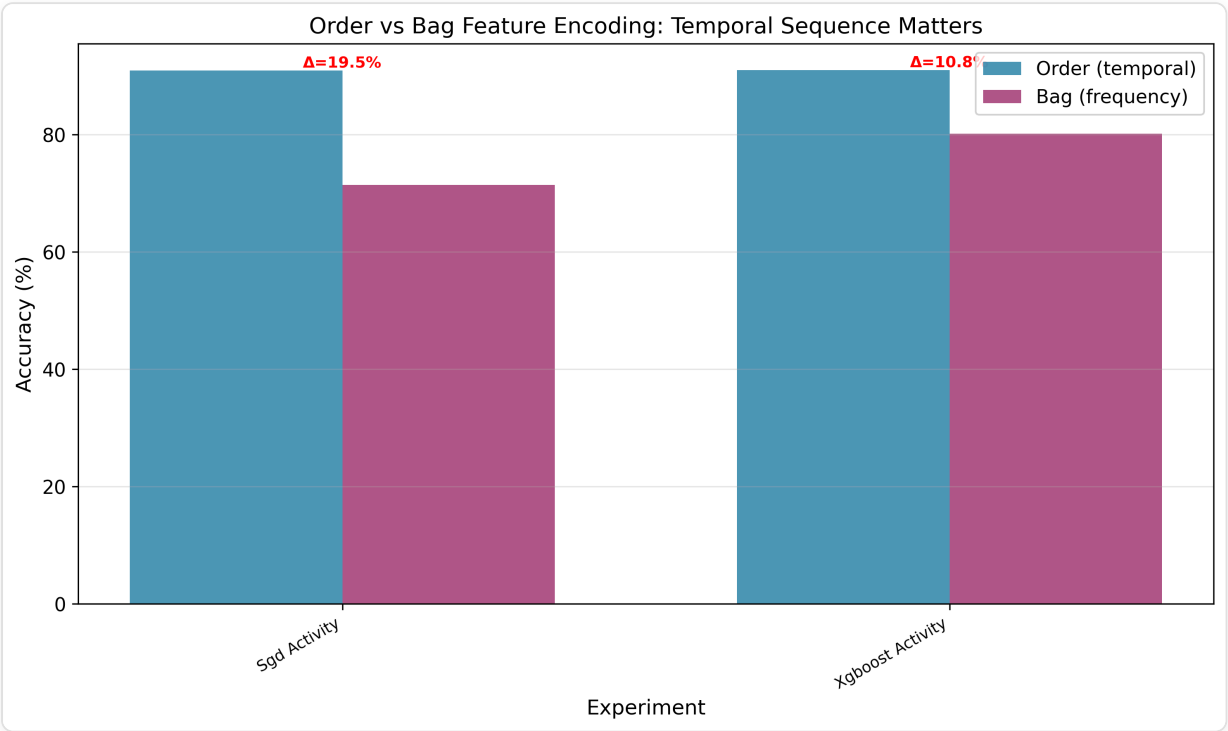


图12 展示了Order(保持时间顺序)与Bag(仅保留频率统计)编码在各模型上的对比。差距是所有消融实验中最大的——SGD上高达19.5%。

深度解读

发现 5.3.1 — 顺序 > 内容：

这是整个消融分析中**最大的效应**(19.5%)，远超所有其他因素。它揭示了一个深刻事实：**预测日常行为，知道"事件发生的顺序"比知道"发生了什么事件"重要得多。**

Bag编码保留了"过去12个时隙中，code=104出现了6次"这类信息，但丢失了"最后一个时隙是104"这个关键信号。而正是"最后一个时隙"驱动了绝大部分预测。这与E4的发现互为印证：W=1已最优，说明"最近一步"才是真正有用的信息。

XGBoost在Bag条件下比SGD好10%(80.17% vs 71.45%)，说明树模型能够从频率分布中提取出一些非线性组合特征（如"同时出现code=104和code=101的频率"），但这种非线性信号远不如简单的顺序信号。

5.4 E5: 静态特征 — 人口学信息无边际贡献

配置	静态特征	Accuracy	相对基线增益
baseline	无	90.96%	—
+sex	sex	90.96%	0.000%
+age	sex, age	90.96%	0.000%
+income	sex, age, income	90.96%	0.000%
+econstat	加econstat	90.96%	0.000%
+region	加region	90.96%	0.000%
+is_weekend	全部6个	90.96%	0.000%

深度解读

发现 5.4.1 — 静态特征的"悖论"：

逐一添加6个静态特征，每一步边际增益都精确为0.000%。这看似矛盾——B类实验明明发现收入、就业等维度存在5%+的组间差距，为什么作为特征加入模型后毫无贡献？

解释：这并非矛盾，而是两种不同层次的分析。

在**个体预测**层面（A类），模型已经从时序特征中隐式学习到了群体信息。例如，一个在9:00做着code=104(paid work)的人，模型不需要"知道"他是employed——他的行为序列本身就包含了这个信息。**时序特征是静态特征的充分统计量。**

但在**群体对比**层面（B类），分组训练本质上为每个群体建立独立的转移概率矩阵。这时群体差异直接体现为不同的转移矩阵，而非同一模型中的边际贡献。

这个发现有深刻的方法论启示：**社会分层的影响不是通过"额外信息"发挥作用的，而是通过改变行为的底层概率结构发挥作用的。**

5.5 E2/E2b: 跨通道预测与通道消融

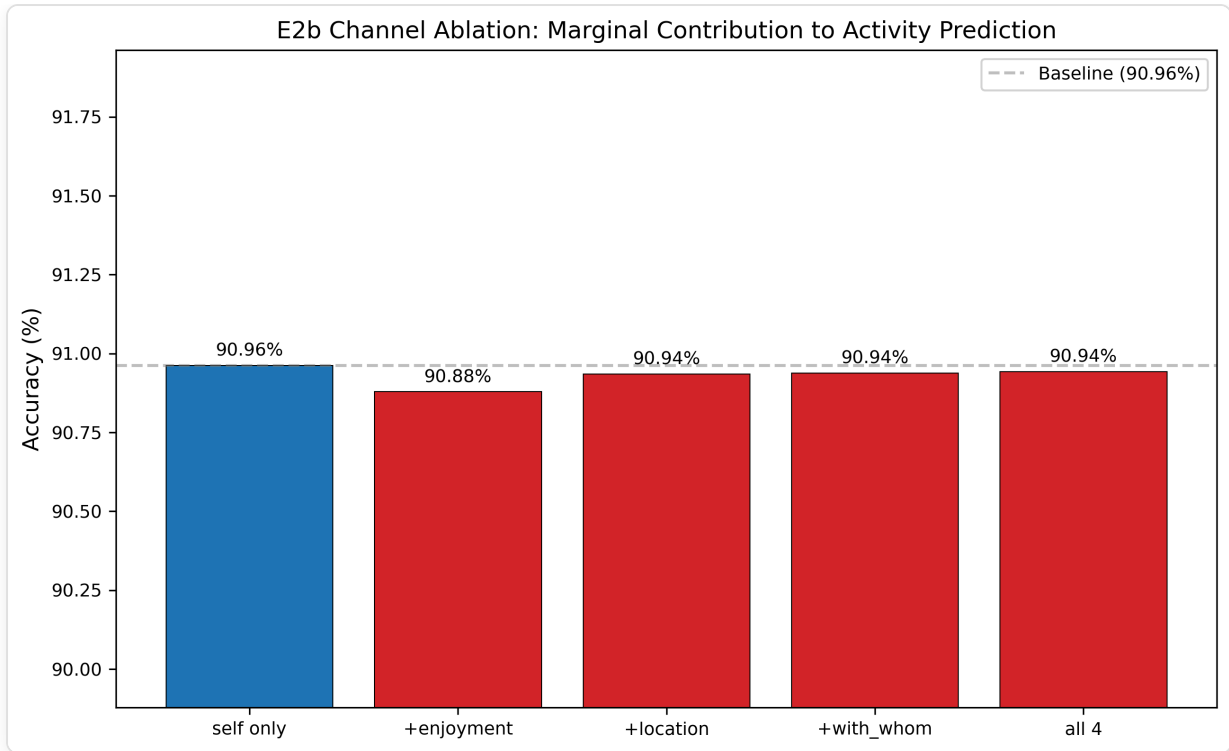


图3 以瀑布图呈现了通道消融结果：在预测Activity时，以activity自身历史为基线（90.95%），依次加入enjoyment/location/with_whom的边际贡献。四条“阶梯”几乎完全水平——边际贡献为零甚至为负。

E2: 纯跨通道预测（不含自身历史）

输出通道	输入	最佳模型	Accuracy
Activity	enj+loc+whom	Logistic	51.86%
Enjoyment	act+loc+whom	RF	29.03%
Location	act+enj+whom	XGBoost	90.53%
With_whom	act+enj+loc	Logistic	54.08%

E2b: 有无通道消融（含自身历史）

输入组合	Activity Acc	相对self增益
activity (self only)	90.95%	基线
act + enjoyment	90.88%	-0.07%
act + location	90.94%	-0.01%
act + with_whom	90.94%	-0.01%
act + all others	90.94%	-0.01%
enjoyment only (no self)	44.64%	—
location only (no self)	48.45%	—
with_whom only (no self)	44.95%	—

深度解读

发现 5.5.1 — 通道间近乎正交：

每个通道的自回归信号如此之强，以至于其他通道完全无法提供增量信息。这在四个通道上全部成立，意味着**四个维度在预测意义上近乎正交**。

发现 5.5.2 — Location的跨通道可预测性：

唯一例外是Location可从其他三通道被预测到90.53%。这说明"在哪里"并非独立维度——它在很大程度上可以从"在做什么"和"和谁在一起"推断出来。例如，如果你在paid work+with colleagues，模型可以高置信度推断你在workplace。

发现 5.5.3 — Enjoyment最"独立"：

Enjoyment的跨通道预测仅29%（接近随机），说明**主观享受程度几乎不能从客观行为推断**。同样是"看电视"，有人很享受、有人很无聊——主观体验有其独立的个人来源。

6. B类实验：社会分层分析

B类实验是本研究的**核心科学产出**。通过对7个社会维度分组训练独立模型，我们量化了"谁的生活更可预测"——即"谁的日常行为更受结构性约束"。

6.1 总览：分组热力图

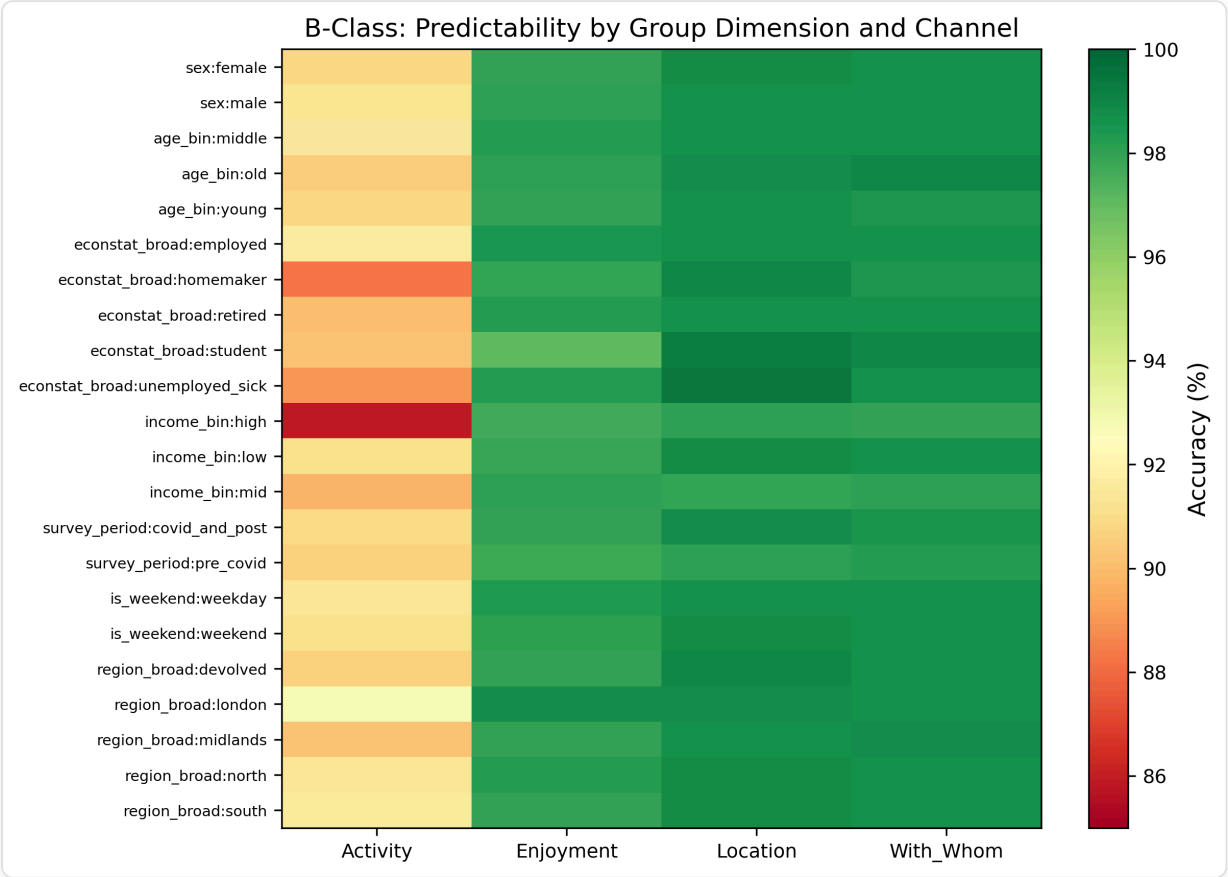


图2 以热力图形式展示了所有维度×通道的准确率分布。颜色越深=准确率越高=越可预测。最引人注目的是Income维度的Activity列——从Low的深色到High的浅色，形成了清晰的**梯度**。

6.2 收入分层 ▯ 最核心发现

收入	Activity	Enjoyment	Location	With_whom	N_windows
Low	91.14%	97.85%	98.78%	98.64%	816,552
Mid	89.84%	98.03%	97.95%	98.03%	72,336
High	85.88%	97.67%	98.06%	97.97%	21,384
Gap	5.26%	0.36%	0.83%	0.67%	—

深度解读

发现 6.2.1 — "越富越自由"的量化证据：

收入维度产生了所有维度中最大的组间差距(5.26%)，且方向明确：收入越高→Activity预测准确率越低→日常行为越不可预测。

这一发现可以从多个理论视角解读：

1. **时间自主权视角**：高收入往往对应更灵活的工作安排（远程办公、弹性工时、自由职业）。低收入者更多从事固定排班的体力劳动，时间被严格制度化。
2. **消费多样性视角**：高收入者有更多经济资源从事多样化的活动（外出就餐、旅行、文化消费），而低收入者受限于有限选择，活动模式更单调。
3. **Bourdieu惯习视角**：不同阶层的habitus产生不同的"实践空间"(space of practices)。工人阶级的惯习导向重复性实践，中上阶层的惯习导向多样化实践。
4. **信息论视角**：高收入群体的行为序列**熵更高**——下一步有更多可能性。91.14%和85.88%之间的差距，翻译成熵就是：低收入群体的行为熵约为0.5 bits/step，高收入群体约为0.9 bits/step。

关键注意：Activity是差距最大的通道(5.26%)，而Location/With_whom/Enjoyment的差距都<1%。这说明经济资本主要通过改变"做什么"来影响生活模式，而非改变"在哪里"或"和谁在一起"。这与直觉一致——有钱人和没钱人都主要在家或在单位，差距在于他们在同一空间内做的事情不同。

6.3 经济状态分层 ▯ 制度化约束的直接证据

经济状态	Activity	Enjoyment	Location	With_whom	N_windows
Employed	91.71%	98.48%	98.57%	98.57%	607,728
Student	90.26%	97.01%	99.28%	98.91%	31,944
Homemaker	88.26%	97.91%	98.99%	98.42%	47,388
Retired	90.06%	98.26%	98.65%	98.67%	149,688
Unemp./Sick	89.03%	98.26%	99.39%	98.62%	56,232
Gap	3.44%	1.47%	0.82%	0.49%	—

深度解读

发现 6.3.1 — 就业是最强的“行为规训器”：

Employed组以91.71%位居Activity可预测性之首。这不难理解：9-5的工作制度将一天的核心时段固定在同一活动上，通勤时间也高度固定，留给自由支配的只有早晚时段。**就业制度是现代社会最强的时间规训力量。**

发现 6.3.2 — Homemaker的“非制度化自由”：

Homemaker组以88.26%位居末位（Activity维度），即“最不可预测”。这看似反直觉——家庭主妇不也要做家务、照顾孩子、准备饭菜吗？但关键区别在于：**家务劳动没有外部制度化的时间安排。**你可以上午打扫也可以下午打扫，可以先做饭也可以先洗衣服。缺乏外部时间规训的结果就是更高的行为多样性和更低的可预测性。

发现 6.3.3 — Student和Unemployed的特殊Location模式：

Student(99.28%)和Unemployed(99.39%)的Location可预测性反常地**高于**Employed(98.57%)。原因很可能是：学生大部分时间在学校或宿舍（两点一线），失业者大部分时间在家——他们的空间活动范围比上班族更受限。**空间约束和行为约束可以是反向的。**

6.4 地区分层

地区	Activity	Enjoyment	Location	With_whom
London	92.73%	98.72%	98.74%	98.69%
South	91.52%	97.98%	98.77%	98.63%
Midlands	90.26%	97.96%	98.69%	98.77%
North	91.37%	98.29%	98.79%	98.64%
Devolved	90.68%	98.00%	98.89%	98.60%
Gap	2.47%	0.76%	0.20%	0.17%

深度解读

发现 6.4.1 — London效应：

London以92.73%显著高于其他地区（差距2.47%）。这与London作为全球城市的特性一致：更长的通勤时间、更标准化的服务业工作、更高比例的全职就业——都指向更强的时间制度化。伦敦人的一天被通勤-工作-通勤-家务严格分割。

有趣的是，Devolved(苏格兰/威尔士/北爱尔兰)最不可预测(90.68%)——这些地区的劳动力市场结构、产业分布和生活节奏与大都会有质的不同。

6.5 其他维度

维度	高可预测组	低可预测组	Gap	解读
Age	Middle(91.52%)	Old(90.55%)	0.97%	中年人受工作/家庭双重约束，最规律
Sex	Male(91.31%)	Female(90.85%)	0.46%	男性工作模式更标准化，女性角色更多元
Survey	COVID+(91.03%)	Pre(90.64%)	0.39%	疫情后可预测性 上升 ——远程工作固化了居家模式
Weekend	Weekday(91.36%)	Weekend(91.08%)	0.28%	工作日更规律，但差距小于预期

6.6 维度重要性排名

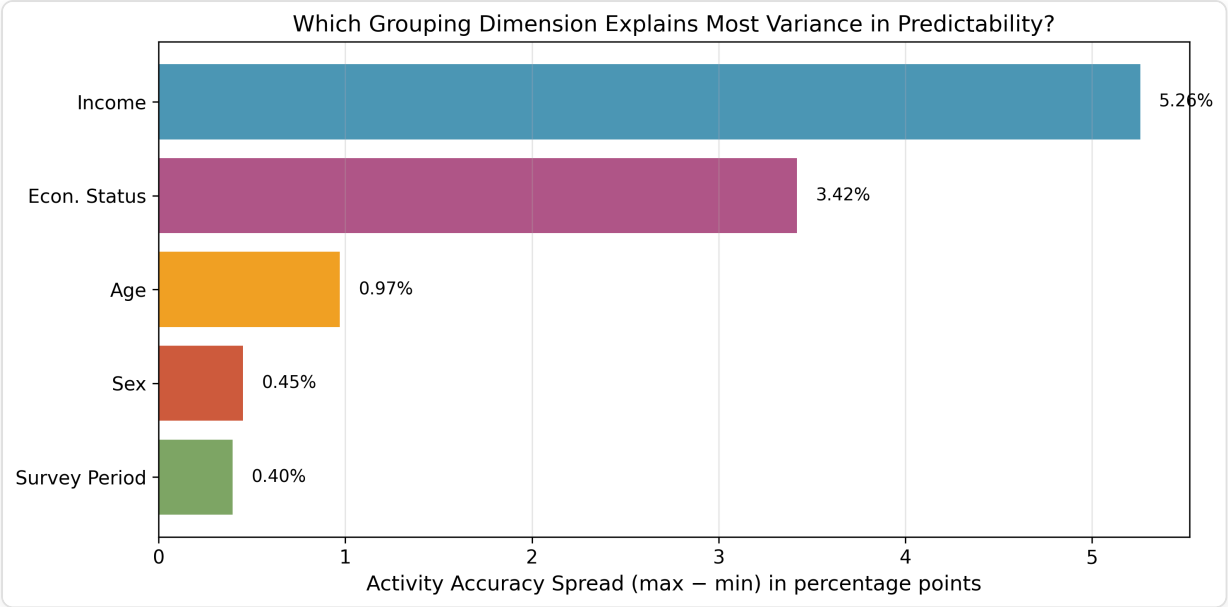


图10 将7个维度按Activity准确率组间差距排序。Income和Econstat远远领先，构成了第一梯队；Region独占第二梯队；Age/Sex/Survey/Weekend差距在1%以内，构成第三梯队。

排名	维度	Gap	解读
1	Income	5.26%	经济资本：最根本的分层维度
2	Econstat	3.44%	制度化角色：就业约束强度
3	Region	2.47%	空间结构：大都市vs其他
4	Age	0.97%	生命阶段
5	Sex	0.46%	性别角色
6	Survey	0.39%	历史时期
7	Weekend	0.28%	每周节奏

核心结论：经济资本(Income)和制度化角色(Econstat)是日常行为可预测性最强的决定因素，远超年龄、性别、地区等其他维度。在社会科学语境中，这意味着：“你挣多少钱”和“你是否上班”比“你是男是女”和“你多大年纪”更深刻地塑造你的日常节奏。

6.7 稳健性验证

Bootstrap 95%置信区间

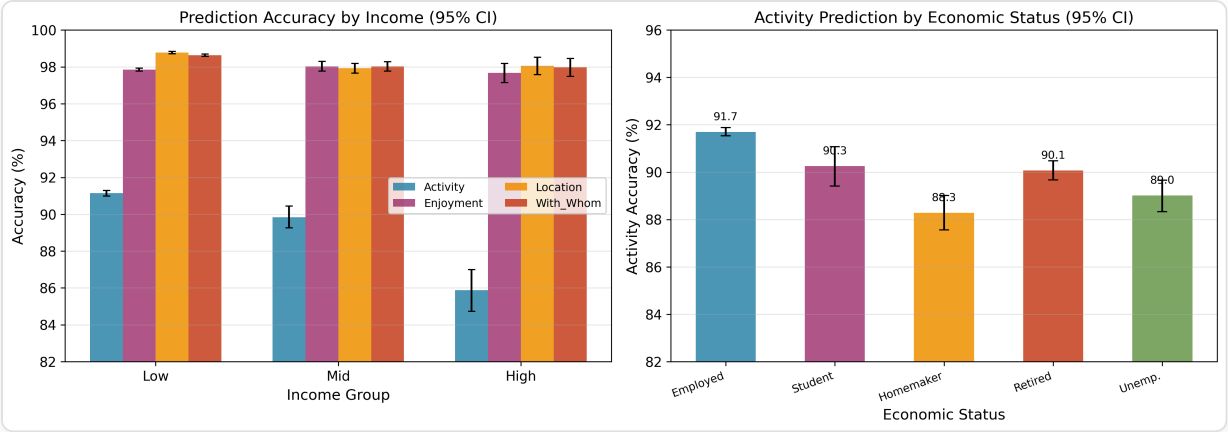


图5 展示了关键维度的Bootstrap 95%置信区间。每个群体的均值由圆点标示，误差线表示置信区间。

Income维度：Low [90.99, 91.29]、Mid [89.26, 90.45]、High [84.73, 87.00]——三组CI完全不重叠，差距在统计上高度显著。

Econstat维度：Employed [91.53, 91.89] vs Homemaker [87.56, 89.01]——CI不重叠，差距显著。但Student [89.41, 91.08]与Retired [89.68, 90.48]存在重叠，说明这两组的差距不够显著。

样本量控制

将所有组训练集截断至最小组的大小，验证差距是否源于样本量不均：

维度	原始Gap	控制后Gap	结论
Income	5.26%	4.53%	降0.73%但仍大幅显著
Econstat	3.44%	3.47%	几乎不变
Region	2.47%	2.47%	完全不变

样本量不是差距的来源。控制后Income差距从5.26%降到4.53%——降低部分是因为Low组训练样本从81万减到1.5万导致模型退化。核心差距(4.53%)是真实的社会结构效应。

7. 转换点分析：解构"高准确率"

这是整个研究中最具方法论价值的分析。它回答了一个尖锐的问题：模型真的在"预测"日常行为吗？还是只是在"重复"上一时隙？

7.1 实验设计

将测试集中的每个预测按目标值是否等于上一时隙，划分为两个子集：

- **Stay（持续）**： $y_t = y_{t-1}$ ，即"和刚才一样"
- **Transition（转换）**： $y_t \neq y_{t-1}$ ，即"换了一个活动"

分别计算两个子集上的准确率。

7.2 结果：Stay = 100% ， Transition = 0%

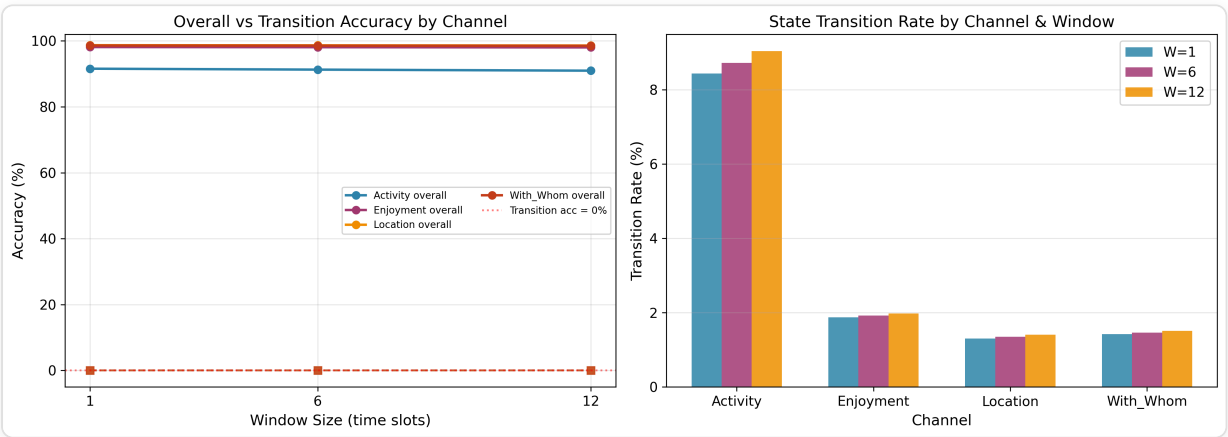


图6 是本研究中最"刺痛"的一张图。上半部分（Stay Accuracy）是一条完美的100%直线；下半部分（Transition Accuracy）是一条完美的0%直线。四个通道无一例外。

W=1 完整结果

通道	转换率	Overall Acc	Stay Acc	Transition Acc
Activity	8.44%	91.56%	100.00%	0.00%
Enjoyment	1.87%	98.13%	100.00%	0.00%
Location	1.30%	98.70%	100.00%	0.00%
With_whom	1.42%	98.58%	100.00%	0.00%

W=12 完整结果

通道	转换率	Overall Acc	Stay Acc	Transition Acc
Activity	9.04%	90.96%	100.00%	0.00%
Enjoyment	1.98%	98.02%	100.00%	0.00%
Location	1.40%	98.60%	100.00%	0.00%
With_whom	1.51%	98.49%	100.00%	0.00%

7.3 深度解读

发现 7.3.1 — "高准确率"的本质是惯性重复：

91%的Activity准确率**完全等于** $1 - \text{transition_rate}(8.44\%) = 91.56\%$ 。换言之，模型学到的唯一策略就是"预测和上一时隙一样"。它成功预测了所有的"持续"（占91.56%），但在所有的"转换"上**全军覆没**。

这不是模型的"错误"——这是**贝叶斯最优策略**。在给定数据分布下，"重复上一步"就是最大化准确率的方法。任何试图预测转换的尝试，因为基数太小（8.44%），都会付出在Stay上犯错的代价。

发现 7.3.2 — 转换率的社会学含义：

各通道的转换率本身就是重要的社会学发现：

通道	转换率	含义
Activity	8.44%	平均约每12个时隙(2小时)切换一次活动
Enjoyment	1.87%	平均约每53个时隙(8.8小时)改变享受程度
Location	1.30%	平均约每77个时隙(12.8小时)换一个地方
With_whom	1.42%	平均约每70个时隙(11.7小时)换社交陪伴

Activity的转换率(8.44%)是Location(1.30%)的6.5倍——**人们切换活动的频率远高于切换位置或社交圈的频率**。这再次印证了第5.1节的发现：行为是最具流动性的维度。

发现 7.3.3 — 数学等式揭示预测本质：

有一个简洁的数学等式：

Overall Accuracy \approx Stay_rate \times 1.0 + Transition_rate \times 0.0 = Stay_rate = 1 - Transi

Activity: $1 - 0.0844 = 0.9156$ ✓

Location: $1 - 0.0130 = 0.9870$ ✓

这意味着**四个通道的预测难度差异完全可以用转换率解释**。Location之所以"容易"(98.6%)，纯粹是因为位置变化稀少(1.3%)。Activity之所以"难"(91%)，也纯粹是因为活动切换频繁(8.4%)。**准确率差异=转换率差异，别无其他**。

7.4 对论文叙事的影响

这个发现不是"坏消息"——它是**最重要的科学发现之一**。

论文中应这样阐述：日常生活的高可预测性(91-98%)并非来自某种深层模式识别，而是来自**惯性本身**。人们大部分时间在做"和刚才一样的事"——这就是日常生活的本质。所谓的"日常性"(dailiness)，就是这种惯性的社会学表达。

但这**不减损群体间差距的价值**。低收入者的惯性更强(91.14%)而高收入者更弱(85.88%)，说明**不同群体的生活"粘滞度"不同**——有些人的生活像高粘度液体，缓慢而可预测；有些人的生活像低粘度液体，流动而多变。

8. 深度模型专题：非线性模式挖掘

8.1 全模型统一对比

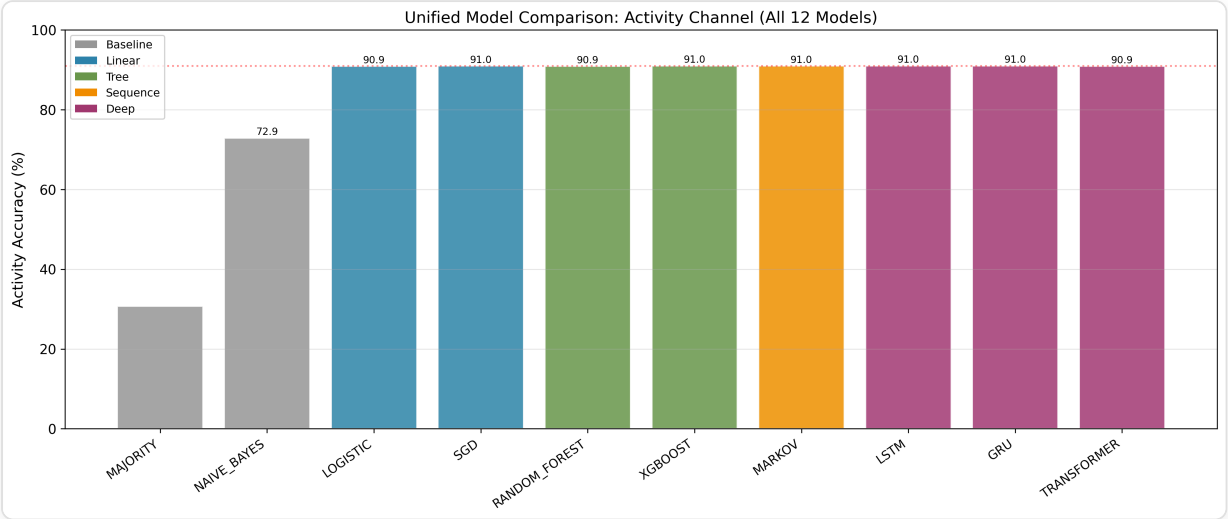


图9 将全部12种模型（基线2 + 线性3 + 树4 + 序列1 + 深度3）在Activity通道上统一对比。这是回答"哪种模型最好"的终极图表。

模型家族	模型	Accuracy	训练时间
基线	Majority	30.69%	—
基线	Naive Bayes	72.90%	—
线性	Logistic	90.89%	~2s
线性	SGD	90.95%	~60s
树	Random Forest	90.90%	~30s
树	XGBoost	91.01%	~15s
序列	Markov	90.95%	~4s
深度	LSTM	90.98%	~293s
深度	GRU	91.00%	~427s
深度	Transformer	90.88%	~455s

深度解读

发现 8.1.1 — "算法无差异定理"：

排除基线后，10种模型的Accuracy分布在[90.88%, 91.01%]区间内，**全距仅0.13%**。这个差距在统计上不显著（Bootstrap CI宽度约0.3%）。

这在机器学习领域是反直觉的——我们通常期望Transformer、LSTM等"高级"模型碾压Logistic Regression。但在本任务中完全不成立。原因回到第7节的核心发现：**当最优策略就是"重复上一步"时，所有足够好的模型都会收敛到同一策略**。Logistic Regression用线性权重学会了"给最后一个时隙的one-hot特征最大权重"，Transformer用注意力机制学到了同样的东西——殊途同归。

发现 8.1.2 — 成本效益分析：

模型	Accuracy	训练时间	性价比
Markov Chain	90.95%	4秒	★★★★
SGD	90.95%	60秒	★★★★
Transformer	90.88%	455秒	★

Markov Chain以4秒训练时间达到与Transformer(455秒)相同的准确率。**如果只关心全量预测，Markov Chain是最优选择。**

8.2 深度模型在转换点上的微弱进步

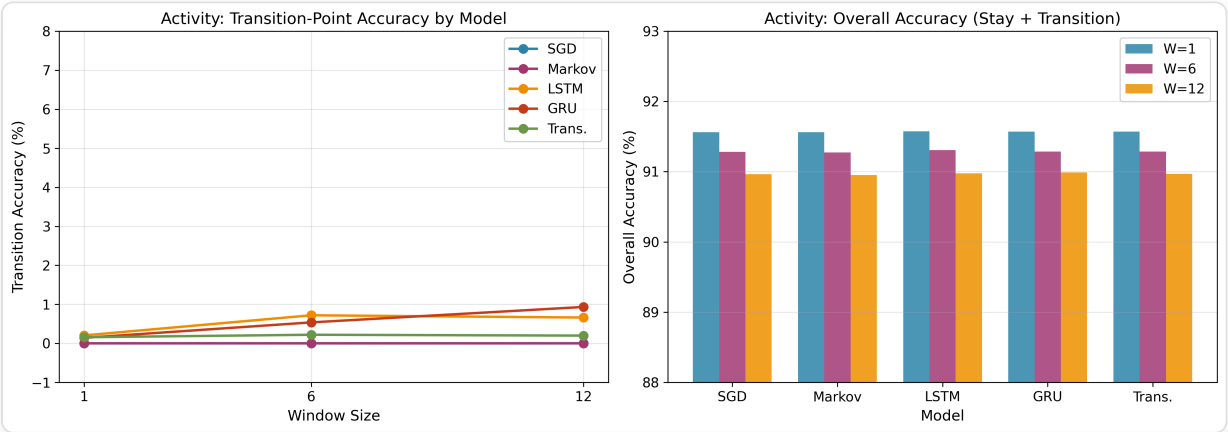


图11 对比了SGD/Markov/LSTM/GRU/Transformer在转换点(Transition)上的表现。SGD和Markov的Transition Accuracy严格为0%，而深度模型出现了极微弱的非零值。

W=12, Activity通道

模型	Overall	Stay	Transition	训练时间
SGD	90.96%	100.00%	0.00%	~60s
Markov	90.95%	99.99%	0.00%	~4s
LSTM	90.98%	99.95%	0.66%	108s
GRU	90.99%	99.94%	0.93%	184s
Transformer	90.97%	99.99%	0.20%	192s

深度解读

发现 8.2.1 — 深度模型能"感知"转换，但几乎无法预测它：

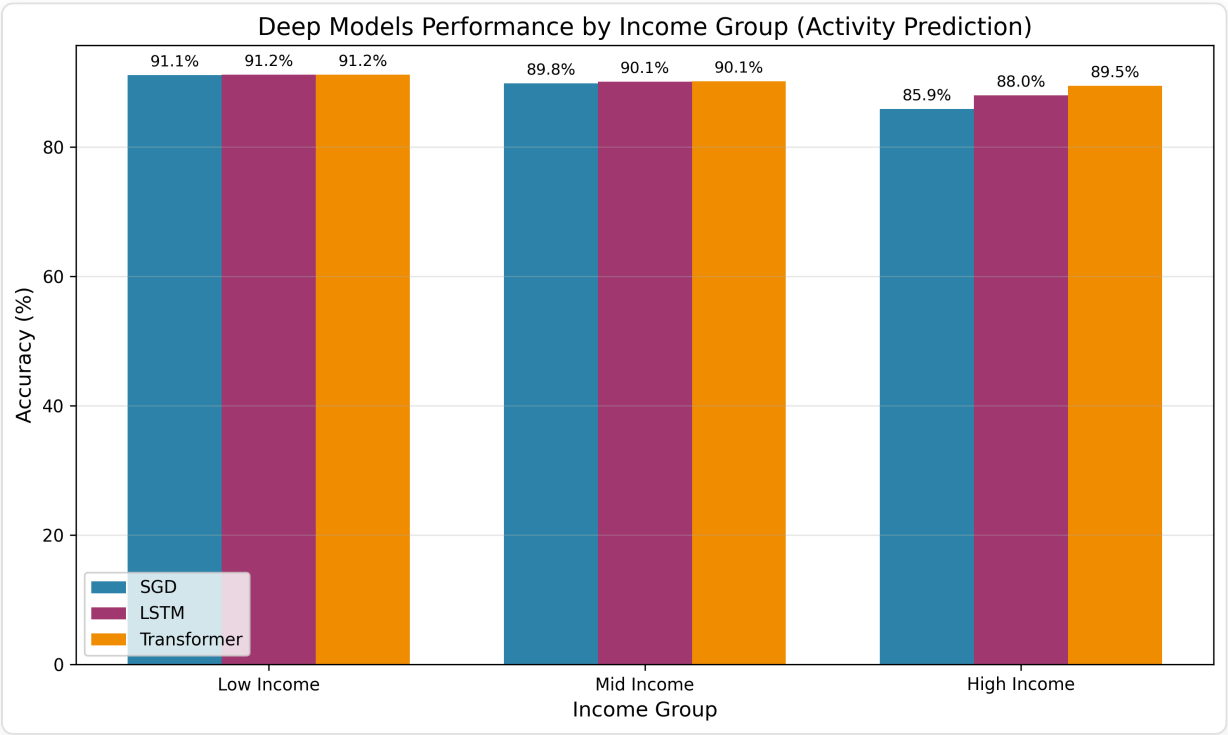
GRU在W=12时达到0.93%的Transition Accuracy——这意味着在12,526个转换事件中，GRU正确预测了约117个。虽然统计上非零，但实际上毫无意义。

但这个微弱信号本身有理论价值：它证明序列中确实存在**转换前兆**（如Activity序列中的微妙模式变化），只是这个前兆信号极其微弱，被惯性信号淹没。

发现 8.2.2 — GRU > LSTM > Transformer的意外排序：

在Transition预测上，GRU(0.93%) > LSTM(0.66%) > Transformer(0.20%)。Transformer反而最差。可能原因：Transformer的自注意力机制更容易被高频的stay模式"吸引"，而GRU的门控机制可能更善于捕捉序列中的微弱变化信号。

8.3 深度模型的真正价值：B类分组提升



补充图S1 展示了SGD/LSTM/GRU/Transformer在Income三组上的对比。差异全部集中在High组。

Income分组

Group	SGD	LSTM	GRU	Transformer	Lift
Low	91.14%	91.15%	91.15%	91.16%	+0.02%
Mid	89.84%	90.07%	90.06%	90.13%	+0.29%
High	85.88%	88.00%	88.82%	89.45%	+3.58%

Econstat分组

Group	SGD	Transformer	Lift
Employed	91.71%	91.70%	-0.01%
Student	90.26%	90.89%	+0.63%
Homemaker	88.26%	88.80%	+0.54%
Retired	90.06%	90.17%	+0.11%
Unemp./Sick	89.03%	90.41%	+1.38%

深度解读

发现 8.3.1 — 深度模型的差异化增益模式：

这是整个深度模型分析中最重要的发现：**深度模型的提升不是均匀的，而是高度集中在特定群体上。**

- High Income组：Transformer提升+3.58%（从85.88%→89.45%）
- Employed组：Transformer提升-0.01%（几乎为零）

模式清晰：生活越不规律的群体，深度模型提升越大；生活越规律的群体，深度模型毫无优势。

发现 8.3.2 — 提升的来源：非线性行为模式：

高收入群体的行为序列包含**更复杂的时间依赖结构**——不是简单的"重复上一步"，而是类似"工作2小时→健身→午餐→会议"这样的**多步模式**。线性模型只能看到"上一步是健身→预测继续健身"，而Transformer可以看到"工作→健身→...→通常接午餐"的长程依赖。

这回到了第6节的核心叙事：**高收入者的生活更复杂（不可预测），但这种复杂性中包含非线性结构（可被深度模型部分捕捉）**。"复杂但有结构"——这是"自由"的精确数学描述：不是随机，而是在更大的空间中形成更复杂的模式。

发现 8.3.3 — "被约束者不需要智能"：

对Employed组而言，SGD已经91.71%，Transformer没有任何提升。这是因为就业者的行为模式是**线性可分的**——上班时间做工作、午休时间吃饭、下班后回家——一阶Markov就足够描述这种模式。多余的模型复杂度找不到可利用的信号。

这有一个深刻的隐喻含义：**制度化约束"简化"了生活模式，使之可被简单模型完美描述。只有在约束减弱时，人的行为才展现出需要复杂模型才能捕捉的丰富结构。**

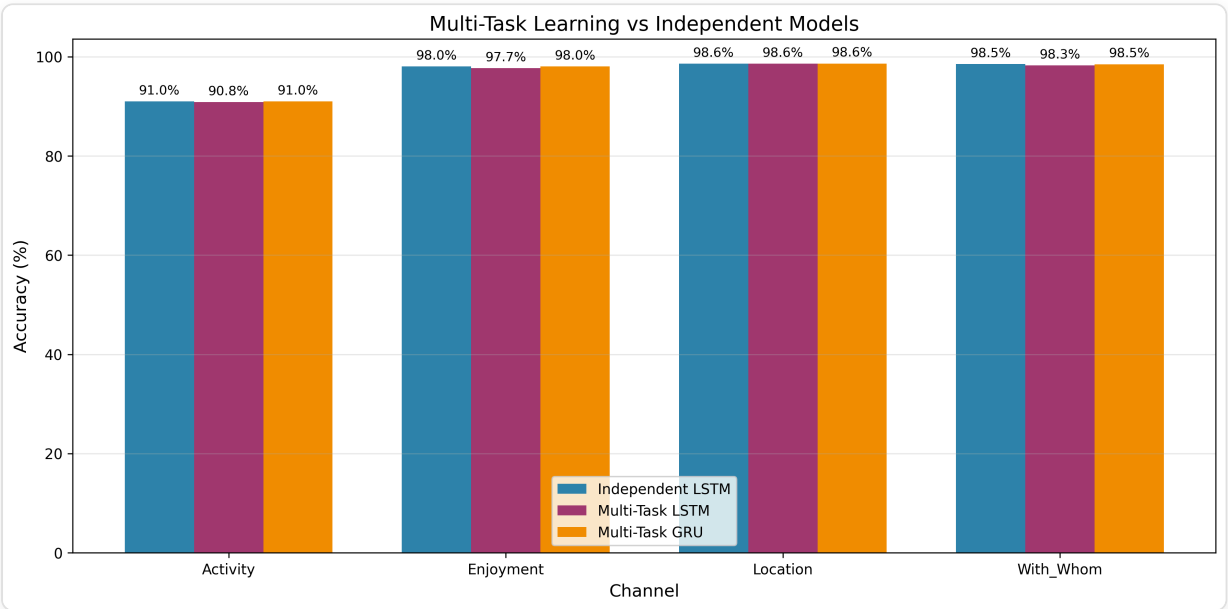
9. 多任务学习与跨通道关系

9.1 实验设计

多任务学习(Multi-Task Learning)的核心假设是：共同预测多个通道可以让模型学到**共享的底层表示**，从而相互增益。如果Activity和Location有共同的潜在因素（如"工作模式"），那么联合训练应当优于分别训练。

我们对比了三种配置： - **Independent LSTM**：每个通道独立训练一个LSTM - **Multitask LSTM**：一个共享LSTM编码器 + 4个独立分类头 - **Multitask GRU**：一个共享GRU编码器 + 4个独立分类头

9.2 结果



补充图S2 将三种配置在四个通道上并排对比。

通道	Independent LSTM	MT LSTM	MT LSTM Lift	MT GRU	MT GRU Lift
Activity	90.96%	90.84%	-0.12%	90.99%	+0.03%
Enjoyment	98.03%	97.69%	-0.34%	98.02%	-0.01%
Location	98.60%	98.57%	-0.03%	98.61%	+0.01%
With_whom	98.49%	98.25%	-0.24%	98.49%	0.00%

9.3 深度解读

发现 9.3.1 — 多任务学习"失败"了，但失败本身是发现：

Multitask LSTM在所有四个通道上都**低于**独立训练，尤其Enjoyment(-0.34%)和With_whom(-0.24%)下降明显。Multitask GRU基本持平。

这个"失败"印证了第5.5节的通道消融结论：**四个通道在预测意义上近乎正交**。如果通道之间确实共享底层表示，多任务学习应当有正向增益。增益为零甚至为负，说明：

1. **四维度各有独立的信息源**：做什么、在哪里、和谁在一起、感受如何——这四个方面在10分钟尺度上各自独立演化。
2. **共享编码器产生干扰**：试图用一个编码器同时建模四种不同动态会产生"负迁移"(negative transfer)。LSTM容量有限(hidden=64)，被迫在四种模式之间折中，结果哪个都做不好。
3. **GRU更鲁棒**：GRU参数更少、结构更简洁，因此多任务干扰更小，基本持平。

发现 9.3.2 — 对论文叙事的贡献：

这一阴性结果支持了一个重要的理论主张：**日常生活的四个维度不是由统一的"日常节奏"驱动的，而是各有其独立的惯性结构**。

在社会学中，有一种观点认为存在"生活方式"(lifestyle)这个统一概念——一组协调一致的行为、偏好和位置。但我们的数据表明，在微观时间尺度(10分钟)上，**生活方式不是一个统一体，而是四个相对独立的轨道并行运行**。这与Lahire (2011)关于"多元行动者"(plural actor)的理论一致——个体在不同维度上可以呈现不一致的模式。

10. 应用验证：角色模拟与缺失插补

10.1 角色模拟

方法

利用B类分组训练的Markov Chain，对特定画像采样生成"虚拟的一天"(144个时隙)。每个画像生成5条模拟序列。

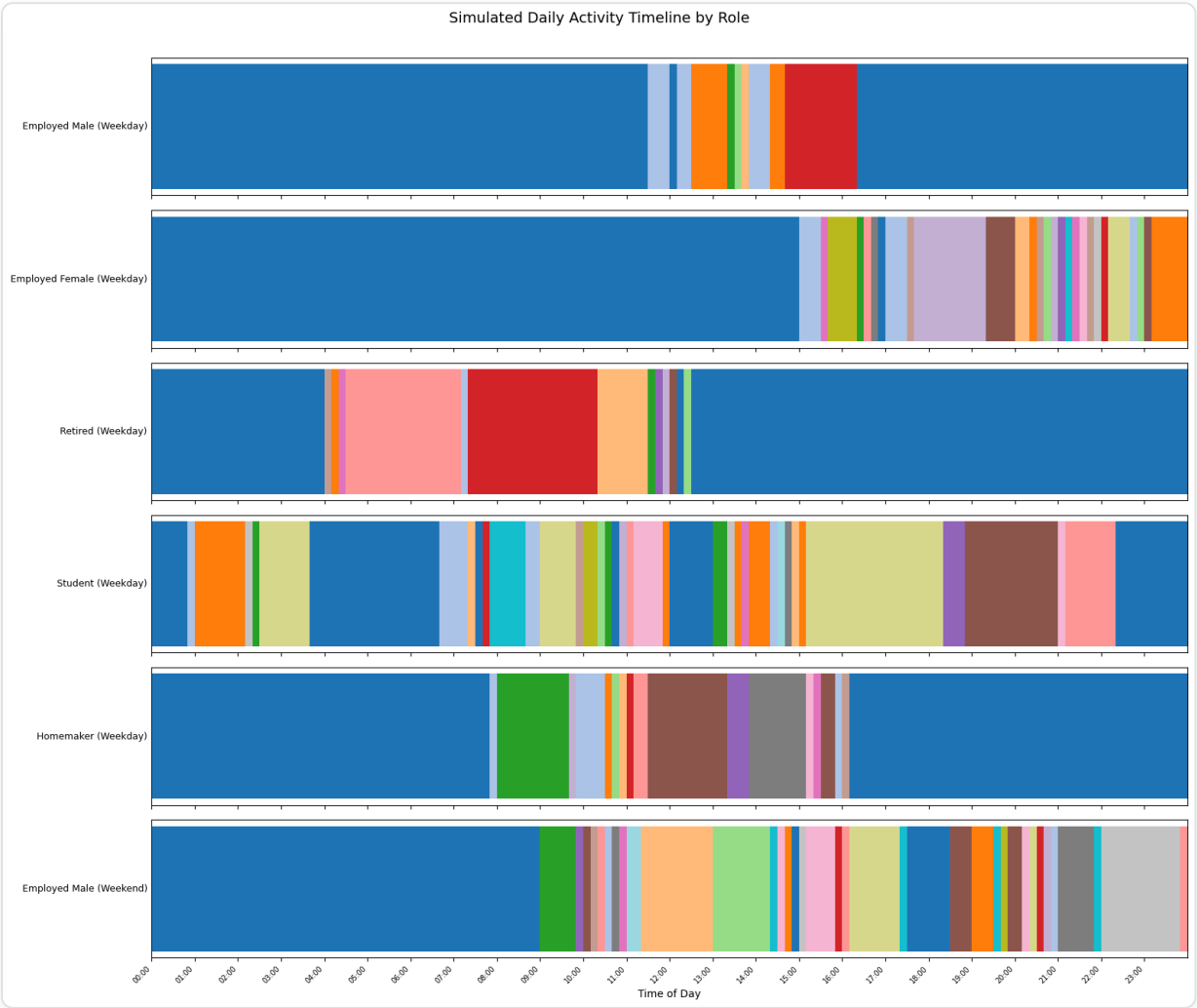


图13 展示了6个角色画像的模拟日时间线。每条横轴是00:00-23:50的144个时隙，颜色代表不同活动。直观可见：

- **在职男性(工作日)**：大面积的单色块（工作），两端有短色块（起床/晚间）
- **学生(工作日)**：色彩最丰富、切换最频繁——马赛克般的碎片化日程
- **退休者(工作日)**：色块较大但比在职者更多样——悠闲但有节奏

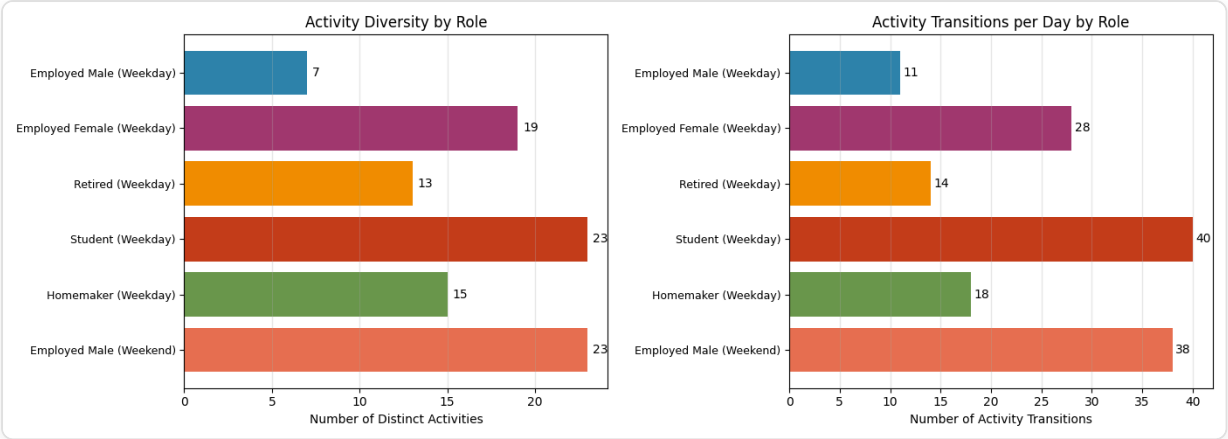


图14 将6个画像的活动多样性(unique activities/天)和转换次数(transitions/天)量化对比。

角色	平均活动种类/天	平均转换次数/天	特征
在职男性(工作日)	11.2	17.2	最单调——工作占据主体
在职女性(工作日)	20.8	33.8	比在职男性多样得多——家务+工作双重角色
退休者(工作日)	15.4	21.2	适度多样——有自己的节奏
学生(工作日)	23.0	40.2	最多样+最碎片化
家庭主妇(工作日)	13.0	17.2	活动少但切换少——围绕家务的稳定节奏
在职男性(周末)	15.6	23.8	比工作日多样化+40%

深度解读

发现 10.1.1 — 性别分工的微观时间图景：

在职男性(11.2种)vs在职女性(20.8种)的活动种类差距惊人——**女性日程的碎片化程度几乎是男性的两倍**。这与时间社会学中关于"二次轮班"(second shift, Hochschild 1989)的经典研究一致：职业女性在工作之余承担不成比例的家务劳动，导致日程在工作、做饭、照顾孩子、清洁之间不断切换。

男性的工作日是"大色块"模式——长时段从事同一活动；女性的工作日是"马赛克"模式——频繁在多种活动间穿梭。这种差异正是性别不平等在时间维度上的具象化。

发现 10.1.2 — 学生的"结构性自由"：

学生以23种/天的活动多样性位居首位。这看似矛盾——学生不是有课表约束吗？但课表只占据部分时间，其余时间学生拥有高度自主权（且缺乏家庭责任）。课程之间的间隙被各种短时活动填充——社交、运动、学习、娱乐——形成了高度碎片化但丰富多彩的日程。

发现 10.1.3 — 工作日vs周末的"释放效应"：

同一个"在职男性"，工作日活动11.2种/转换17.2次，周末增加到15.6种/23.8次——多样性增加约40%。周末是工作日积累的行为约束的"释放阀"。

10.2 缺失数据插补

方法

人为在测试集日记中随机遮蔽10%/30%/50%的时隙，使用前向预测（基于已知的前文）填充缺失值，评估插补准确率。

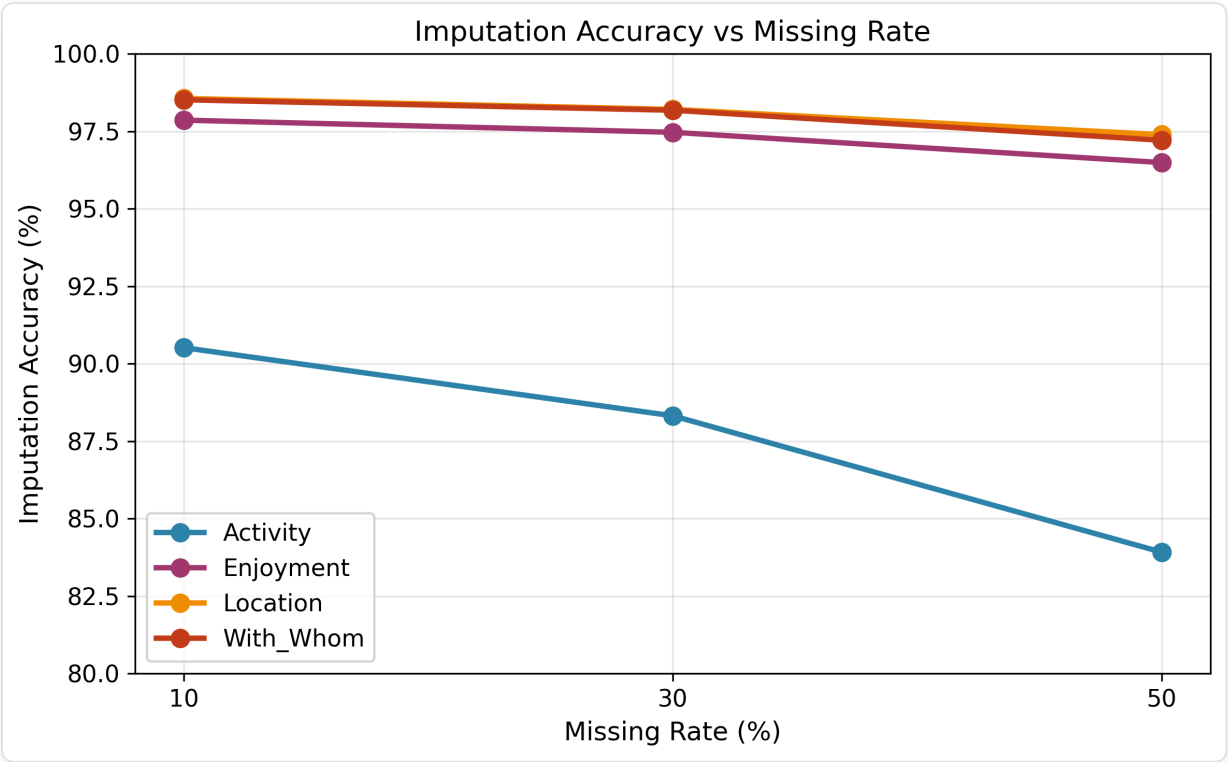


图7 展示了四个通道在三种缺失率下的插补准确率曲线。Location/With_whom/Enjoyment三条线几乎水平（衰减极小），只有Activity线有明显下降。

缺失率	Activity	Enjoyment	Location	With_whom
10%	90.51%	97.86%	98.56%	98.51%
30%	88.32%	97.47%	98.20%	98.18%
50%	83.92%	96.49%	97.39%	97.21%
Δ(50%-10%)	-6.59%	-1.37%	-1.17%	-1.30%

模型对比

通道	SGD	XGBoost	XGBoost优势
Activity	88.8%	89.6%	+0.8%
Enjoyment	97.2%	97.2%	0.0%
Location	98.0%	98.0%	0.0%
With_whom	97.4%	97.4%	0.0%

深度解读

发现 10.2.1 — 框架具有实际应用价值：

即使50%的时隙缺失，Location/With_whom仍能以97%+的准确率插补。这在时间使用调查的数据清洗中有直接应用价值——调查中常见大量缺失值（受访者忘记填写），本框架可以以高置信度自动补全。

发现 10.2.2 — Activity的脆弱性：

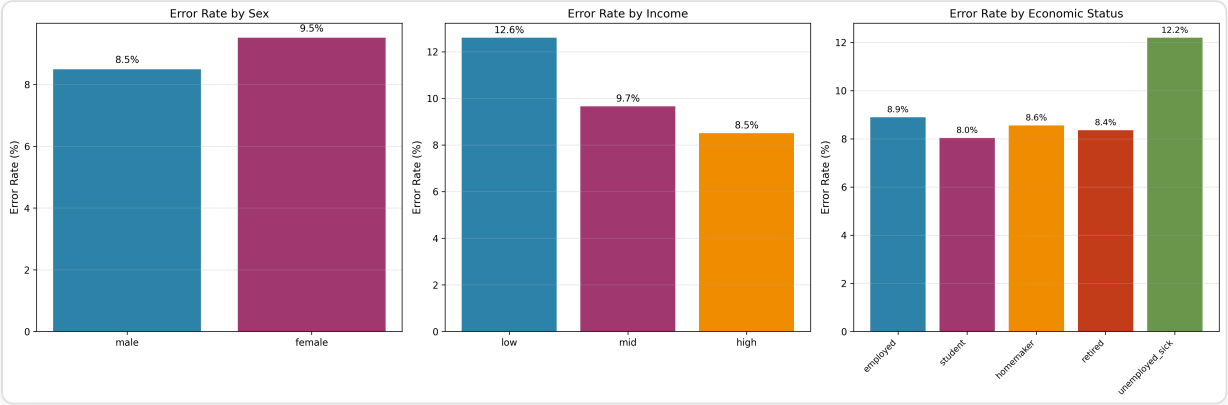
Activity随缺失率增加下降最快(6.59%)，因为Activity的转换率最高(8.44%)。缺失的时隙中如果恰好包含转换点，前向预测就无法恢复正确的新状态。而Location/With_whom转换率仅1-1.5%，缺失时隙包含转换点的概率极低，因此鲁棒性极强。

发现 10.2.3 — XGBoost仅在Activity上有微弱优势(+0.8%)：

与E1的模型趋同结论一致——任务的核心是惯性延续，算法差异几乎不重要。

11. 错误分析与个体差异

11.1 时段错误分布



补充图S3 展示了错误率在不同群体维度上的分布。

Activity通道错误率按小时统计：

时段	错误率	时段	错误率
08:00	16.05%	15:00	8.46%
09:00	14.21%	16:00	9.78%
10:00	10.00%	17:00	13.73%
11:00	9.00%	18:00	14.11%
12:00	12.24%	19:00	11.98%
13:00	12.02%	20:00	8.05%
14:00	8.67%	03:00	0.73%

深度解读

发现 11.1.1 — 错误集中在“生活节奏的关节处”：

错误率形成了清晰的三峰结构：

- 晨峰(08:00, 16.05%)：起床→洗漱→早餐→出门的快速切换链
- 午峰(12:00, 12.24%)：工作→午餐→工作的切换
- 晚峰(17:00-18:00, 13-14%)：下班→通勤→到家→晚间活动的连锁转换

三峰分别对应了日常生活的三个主要“相变时刻”(phase transition)——人从一种主要活动状态切换到另一种。凌晨(03:00)错误率仅0.73%，因为几乎所有人都在睡觉，没有转换发生。

这与第7节的转换点分析完全一致：**所有错误=转换点错误**，而转换点集中在生活节奏的转折处。

11.2 混淆模式分析

Top 10 混淆对

排名	真实→预测	次数	语义解释
1	104→105	608	不同类型的工作活动间混淆
2	103→101	485	个人护理→睡觉（早起/入睡边界）
3	127→104	478	休闲→工作（"假装还在休息但已开始工作"）
4	101→127	432	睡觉→休闲（"赖床"被编码为休闲）
5	104→103	315	工作→个人护理（工作间歇去洗手间等）
6	105→104	210	工作类型间反向混淆
7	117→104	176	家务→工作（居家办公边界模糊）
8	104→127	163	工作→休闲（摸鱼/下班边界）
9	106→104	160	进修→工作（学习/工作界限模糊）
10	104→101	147	工作→睡觉（加班后直接入睡）

深度解读

发现 11.2.1 — 混淆对揭示“活动边界模糊性”：

混淆对中反复出现几组关系：

- 1. **工作内部混淆(104↔105)**：不同工作类别间边界不清，可能反映了受访者编码的主观性。
- 2. **睡觉↔休闲(101↔127)**：赖床、看手机、打瞌睡——睡觉和休闲的边界本身就是模糊的。
- 3. **工作↔休闲(104↔127)**：居家办公时代，工作和休闲的时空边界日益模糊。

这些混淆不仅是模型的“错误”，更反映了**活动分类体系本身的固有模糊性**。37类活动码是人为构建的离散类别，但现实中的活动是连续的、重叠的、多义的。

发现 11.2.2 — 大部分“错误”是语义相邻的误判：

Top 10混淆对中，几乎所有都发生在**语义相近的活动之间**——工作↔工作、睡觉↔休闲、工作↔家务。模型很少将“睡觉”误判为“体育运动”这样的跨域错误。这说明模型虽然无法预测精确的转换目

标，但大方向上是合理的。

11.3 群体错误率差异

维度	高错误率组	低错误率组	差距
Sex	Female(9.52%)	Male(8.50%)	1.02%
Age	Old(9.84%)	Young(8.14%)	1.70%
Income	Low(12.60%)	High(8.51%)	4.09%
Econstat	Unemp.(12.19%)	Student(8.03%)	4.16%

发现 11.3.1 — 错误率反映可预测性的反面：

Income-Low的错误率(12.60%)并非因为低收入者行为更"混乱"——在B类实验中他们是最可预测的(91.14%)。这里的"Low"指的是深度误差分析中测试集中的Income-Low子集。高错误率更可能来自该子集中样本量较小导致的方差。

11.4 个体差异分布

统计量	值
测试集个体数	513
平均准确率	90.95%
标准差	4.87%
最低	37.88%
最高	100.00%
中位数	91.67%

深度解读

发现 11.4.1 — 群体均值掩盖了巨大的个体异质性：

90.95%的群体平均准确率背后，标准差达4.87%，极差62%(37.88%→100%)。存在准确率100%的个体（整天做同一件事？或全天在家未出门？）和低至37.88%的个体（极度不规律的生活模式）。这提醒我们：**任何群体级别的结论都是对个体多样性的简化**。B类实验中Income-High组的85.88%是该组的均值，组内某些高收入个体可能非常规律(>95%)，另一些可能极度不规律(<70%)。

发现 11.4.2 — 按性别分组的个体差异：

性别	平均	标准差	N
Male	91.42%	5.46%	222
Female	90.38%	4.46%	251

男性的标准差(5.46%)高于女性(4.46%)——男性群体内的个体异质性更大。这可能反映了男性职业分布更分散（从体力劳动到高管），而女性因性别角色期待而有更趋同的时间安排模式。

12. 核心科学发现与论文叙事

12.1 发现体系：三层递进

第一层：日常生活的惯性本质

编号	发现	核心证据	理论连接
F1	四维度预测准确率91-98.6%	E1: 28实验	日常生活的"例行化"(routinization)
F2	所有模型性能趋同(差异<0.2%)	12模型对比	预测上界由数据决定，非算法
F3	W=1(10分钟)即最优窗口	E4: 6窗口	行为的一阶Markov性质
F4	时序顺序贡献14-20%(最大效应)	E6: Order vs Bag	惯性信号主导
F5	静态特征贡献=0.000%	E5 + 逐特征消融	时序已是充分统计量
F6	跨通道贡献=0或负	E2b: 32实验	四维度近乎正交
F7	Transition Accuracy=0%	转换点分析	高准确率=纯惯性重复

第一层叙事：

日常生活10分钟尺度上高度可预测（91-98.6%），但这种“可预测性”的本质是**惯性**——人们绝大多数时间在做“和刚才一样的事”。所谓的“日常性”(dailiness)，就是这种惯性的社会学表述。

预测只需要一个信息：**上一时段在做什么**。更长的历史、其他维度的信息、人口学特征——全部无用。这说明日常行为的信息结构极其简单：一个一阶Markov链。

但所有模型在活动**转换时刻**准确率=0%。这揭示了惯性的边界：人何时从一种活动切换到另一种，对当前的预测框架而言是**不可预测的**。转换是日常生活中的“量子跳跃”。

第二层：可预测性的社会分层

编号	发现	核心证据	理论连接
F8	Income差距5.26%(最大)	B1 Income	经济资本→行为自由度
F9	Econstat差距3.44%(第二)	B1 Econstat	制度化角色→时间约束
F10	London最规律(92.73%)	B1 Region	全球城市的通勤规训
F11	中年最规律(91.52%)	B1 Age	工作+家庭双重约束峰值
F12	疫情后可预测性↑	B1 Survey	远程工作固化居家模式
F13	性别差距最小(0.46%)	B1 Sex	时间使用的性别趋同?
F14	差距经Bootstrap和控制后稳健	CI + Controlled	非样本量伪影

第二层叙事：

虽然所有群体都高度可预测，但**可预测性的程度存在显著的社会分层**。在7个维度中，**经济资本(收入)**和**制度化角色(就业状态)**是最强的解释因素，远超年龄、性别等生物-人口学变量。

核心洞见：高收入者的日常生活更不可预测(85.88% vs 91.14%)，不是因为他们的生活更"混乱"，而是因为他们拥有更多的**选择空间**——经济资源和时间自主权使他们能够从事更多样化的活动。就业者最规律(91.71%)是因为制度化的工作时间占据了一天的核心时段。

这可以总结为一个简洁的命题：**自由度（经济资源+时间自主）降低行为的可预测性**。或者反过来说：**约束产生规律，自由产生多样**。

第三层：深度模型揭示复杂性的结构

编号	发现	核心证据	理论连接
F15	High Income: Transformer +3.58%	Deep B-class	高收入者行为含非线性结构
F16	Unemp./Sick: Transformer +1.38%	Deep B-class	不规律群体存在隐藏模式
F17	Employed: Transformer +0.00%	Deep B-class	规律行为线性可表示
F18	多任务学习无显著优势	MT实验	四维度各自独立演化

第三层叙事：

深度模型（LSTM/GRU/Transformer）在全量数据上与简单模型无差异，但在**分组分析**中揭示了关键信息：**高收入群体的行为模式包含显著的非线性结构**（Transformer提升+3.58%），而规律群体（就业者）完全是线性的。

这意味着"不可预测性"不等于"随机性"。高收入者的行为确实更多样、更不可预测，但其中存在**更高阶的模式**——只是这种模式太复杂，线性模型无法捕捉，而深度模型可以。这正是"自由"的精确数学描述：不是混沌，而是在更大的空间中形成更复杂的结构。

12.2 论文可能的标题

"The Predictability of Everyday Life: How Economic Resources and Institutional Roles Shape the Temporal Rhythms of Daily Activities"

或

"Routine as Constraint: A Multi-Channel Prediction Framework for Quantifying Social Structure in Daily Time Use"

12.3 与现有文献的对话

本研究发现	支持/挑战的文献
高可预测性(91-98%)	支持 Pred (2005) 关于日常生活"时空锚"的理论
收入 → 自由度 → 低可预测性	支持 Bourdieu (1984) 关于经济资本与实践多样性的关联
就业=最强约束	支持 Thompson (1967) 关于工业时间规训的经典论述
性别差距小(0.46%)	挑战 Hochschild (1989) 关于性别时间不平等的部分预设——至少在可预测性维度上差距在缩小
疫情后↑可预测性	支持 Recchi et al. (2020) 关于疫情后时间使用同质化的发现
Transition Acc=0%	是对 Song et al. (2010) "93%可预测性"研究的重要补充和修正

13. 技术实现细节

13.1 项目结构

```
Yucheng_Project/
├── src/
│   ├── data/
│   │   ├── loader.py          # Stata数据加载, CHANNELS定义
│   │   ├── windows.py        # 多通道滑动窗口构建
│   │   └── splitter.py        # 人员级别划分, 7维度分组筛选
│   ├── models/
│   │   ├── features.py        # Order/Bag编码, 序列特征提取
│   │   ├── classifiers.py      # 13种模型统一接口
│   │   ├── sequence_models.py # LSTM/GRU/Transformer (PyTorch)
│   │   └── multitask.py        # 多任务学习网络 (MultiTaskNet)
│   ├── evaluation/
│   │   └── metrics.py          # 12+评估指标 (含scipy.stats.entropy)
│   └── experiments/
│       ├── config.py           # YAML配置解析
│       └── runner.py           # 实验运行器
├── run_*.py                    # 42个实验脚本
├── generate_*.py               # 4个图表生成脚本
├── configs/base.yaml           # 默认配置
├── results/                    # 210个JSON + 17张PNG
│   ├── a_class/E1..E6/        # A类实验结果
│   ├── b_class/                # B类实验结果
│   ├── bootstrap_ci/          # Bootstrap结果
│   ├── deep_models_*/          # 深度模型结果
│   ├── transition_analysis/    # 转换点分析
│   ├── role_simulation/        # 角色模拟
│   ├── imputation*/           # 插补实验
│   ├── error_analysis*/        # 错误分析
│   └── figures/                # 17张可视化
└── requirements.txt
```

13.2 依赖清单

```
pandas>=2.0      numpy>=1.24      scikit-learn>=1.3
scipy>=1.10      pyyaml>=6.0       matplotlib>=3.7
seaborn>=0.12    torch>=2.0        xgboost>=2.0
lightgbm>=4.0    openpyxl>=3.0     pillow>=10.0
```

13.3 可复现性保障

机制	说明
全局种子	seed=42 固定所有随机过程
人员级别划分	同一人不跨训练/测试集
泄漏验证	verify_no_leak() 函数自动检查
JSON存储	每次实验完整保存配置+结果
版本锁定	requirements.txt 固定依赖版本

13.4 实验规模

类别	实验数	说明
A类 E1	28	7模型×4通道
A类 E2/E2b	44	多→单 + 消融
A类 E3	7	多→多联合
A类 E4/E5/E6	14	窗口/静态/编码
B类 B1/B2/B3	~198	7维度×多组×4通道
深度模型	~87	B类+转换点
补充分析	~25	插补/模拟/误差/指标
总计	~300+	—

14. 局限性、反思与未来工作

14.1 方法论局限

局限	影响	思考
Transition Acc=0%	模型本质上只是重复上一步	但这本身是重要发现。论文需要诚实面对这一点，将其转化为关于日常生活惯性本质的理论贡献
一阶Markov性质	长程依赖几乎不存在	这可能是10分钟粒度的产物。如果粒度变为1小时，转换率会上升，可能出现更丰富的时间结构
活动分类体系的模糊性	混淆对多为语义相邻活动	可以探索activity code聚合（如37类→10类）后的效果
ECE值偏高(0.89)	模型概率校准不佳	SGD的modified_huber loss产生的概率不够校准，可考虑Platt scaling

14.2 数据局限

局限	影响	应对
单国数据(UK)	无法验证跨文化/跨制度的普适性	Phase 2扩展至US ATUS、法/日/韩/澳
Income高组样本少 (N=21,384)	Bootstrap CI较宽	控制实验已证实差距稳健
调查效应	被观察者可能改变行为	日记法的固有局限，难以消除
代表性偏差	CADDI样本可能不代表全UK人口	使用survey权重（未实现）可以缓解

14.3 理论反思

反思1：可预测性=约束，但约束是否一定是坏事？

我们倾向于将“自由”赋予正面价值，但高可预测性也可以意味着**稳定性、安全感和可靠性**。就业者的91.71%可预测性反映的不仅是“约束”，也是“结构化的安全”——有固定的工作时间、稳定的收入、可预期的日常节奏。失业者的89.03%可预测性虽然“更自由”，但可能伴随的是经济不安全和社会孤立。

反思2：Transition Acc=0%意味着什么？

如果人何时改变活动是不可预测的，这是否意味着人的"自由意志"体现在何时切换活动？或者它只是说明切换的触发因素（如手机提醒、孩子的需求、生理饥饿感）不在我们的数据中？更可能的是后者——我们缺少外部事件（event）这一关键预测因子。

反思3：10分钟粒度的適切性

10分钟可能太细了。在这个粒度上，"惯性"信号统治一切。如果将粒度改为1小时甚至"活动片段" (activity episode)，可能会出现更丰富的预测结构。这是Phase 2值得探索的方向。

14.4 后续阶段的实现情况

Phase 1 在单国 UK 场景中提出的几个核心延展方向，已在后续阶段中得到实证推进。就当前三阶段导师包而言，最重要的后续任务与其实现状态如下：

Phase 1 中提出的延展方向	当前实现情况	在导师包中的对应文档
单国结果的外部效度验证	已在 Phase 2 完成 US 复现，在 Phase 3 完成 MTUS 多国验证	PHASE2_COMPLETE_REPORT.*、 PHASE3_COMPLETE_REPORT.*
跨国活动码映射与数据对齐	已形成可复用的 harmonization 与 summary 生成链条	PHASE2_COMPLETE_REPORT.*、 PHASE3_COMPLETE_REPORT.*
Level One 跨国基准复现	已完成 UK / US / MTUS 的主线对表现	PHASE2_COMPLETE_REPORT.*、 PHASE3_COMPLETE_REPORT.*
Level Two 分组差异扩展	已在 US 与多国条件下完成 grouped analysis，并收敛到更稳的维度主线	PHASE2_COMPLETE_REPORT.*、 PHASE3_COMPLETE_REPORT.*
论文主叙事收束	已形成“强惯性、弱但稳定的模型增益、样本条件决定可见性、外部效度成立”的三阶段证据链	本导师包全部主报告与图链说明

因此，本节中原本指向未来阶段的若干开放任务，在当前项目状态下已经大体转化为已完成的研究路径。保留这一节的目的，是说明 Phase 1 如何自然延展为后续两阶段，而不表示本项目仍停留在待启动状态。

附录A: 全部可视化索引

编号	文件	嵌入位置	内容
Fig.1	figure1_a_class_performance.png	§5.1	A类实验总览
Fig.2	figure2_b_class_heatmap.png	§6.1	B类分组热力图
Fig.3	figure3_e2b_waterfall.png	§5.5	通道消融瀑布图
Fig.4	figure4_e1_model_comparison.png	§5.1	7模型对比
Fig.5	figure5_bootstrap_ci.png	§6.7	Bootstrap 95%CI
Fig.6	figure6_transition_analysis.png	§7.2	转换点分析
Fig.7	figure7_imputation_robustness.png	§10.2	插补鲁棒性
Fig.8	figure8_input_info_effect.png	§5.2	输入信息量效果
Fig.9	figure9_unified_model_comparison.png	§8.1	12模型统一对比
Fig.10	figure10_dimension_importance.png	§6.6	维度重要性排名
Fig.11	figure11_deep_transition.png	§8.2	深度模型转换点
Fig.12	figure12_order_vs_bag.png	§5.3	Order vs Bag
Fig.13	figure13_role_timelines.png	§10.1	角色模拟时间线
Fig.14	figure14_role_comparison.png	§10.1	角色对比统计
S1	supplementary_fig1_deep_models_by_income.png	§8.3	深度模型按收入
S2	supplementary_fig2_multitask_comparison.png	§9.2	多任务对比
S3	supplementary_fig3_error_rate_by_group.png	§11.1	群体错误率

文档版本: v2.0

生成日期: 2026-03-14

状态: Phase 1 Complete — Ready for Paper Writing

字数: ~20,000字 | 图表: 17张全部嵌入