

从游戏 logfile 预测数学成就与推理能力 ——机器学习的应用

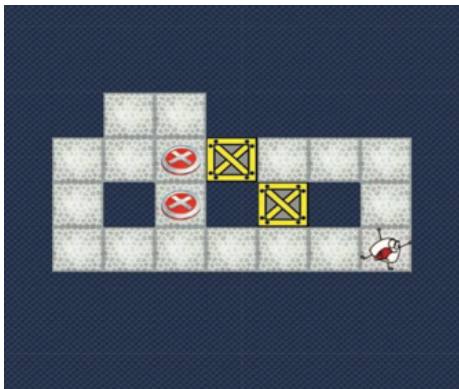
符植煜

北京师范大学心理学部

2017-4-20

引言：logfile 分析技术

- 通过追踪、分析被试在计算机上完成认知任务过程中的操作行为、操作时间等信息取测量指标 (Veenman, Bavelaar, De Wolf, & Van Haaren, 2014)。
- 典型的应用如推箱子游戏。



引言：传统技术的局限性

- 只提取少数指标，浪费了潜在可挖掘的信息；
- 使用线性模型，预测效果受到限制；
- 这些指标往往并不只是纯粹的包括一个元认知成分，而是多个元认知成分的混合 (Veenman, 2013)；

机器学习技术可以提供更好的解决方案。

引言：机器学习

- 给定一个数据集，能够同时观察到一系列 y （结果变量）以及 x （特征）；
- 在该数据集中我们训练得到模型；
- 获得的模型能够根据已知的 x 预测对应的 y 。
- 允许更复杂的模型，充分利用数据：
 - ▶ 随机森林；
 - ▶ 支持向量机；
 - ▶ 神经网络模型；
- 现有的在心理学中的应用：
 - ▶ 使用社交网络数据预测人格特质 (Wu, 2015)；
 - ▶ 使用可穿戴智能设备记录的生理指标预测情绪 (Zhang, 2016)。

引言：本文

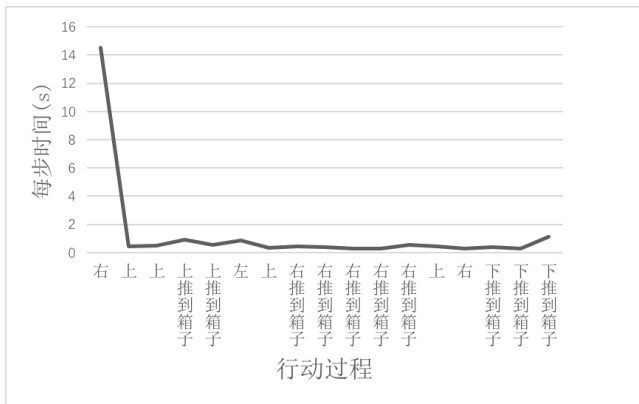
- 将机器学习技术引入 logfile 分析中；
- 使用推箱子的 logfile：
 - ▶ Li et al.(2015) 之前使用推箱子游戏测量被试的计划性 (Planning)；
 - ▶ 使用了第一步之前的时间与总时间之比作为计划性的指标，信效度良好。
- 以数学能力、推理能力作为结果变量；
 - ▶ 推箱子游戏涉及不止一种元认知、认知能力；
 - ▶ 数学能力、推理能力作为同样涉及了综合的认知、元认知能力，因此能充分利用数据。

方法：数据采集

- 被试：396 名初一、初二学生，最终有效观测值 360 个。
- 结果变量：
 - ▶ 瑞文标准推理测验 (Standard Progressive Matrices, SPM, Raven, 1989)：用于测量一般智力中的推理能力；
 - ▶ 数学成绩：3 次年级统测的数学成绩的平均值；
- 两个结果变量均分别取前 25% 得分的学生记为 1，后 25% 得分的学生记为 0，其余学生不予使用。最终使用被试数为 180 名。
- 推箱子：所有被试按照相同顺序完成 20 关。题目由 Li et al. (2015) 筛选获得。

方法：特征提取

程序记录了被试在每一关中的每个行动以及花费时间，见图：



图：一个典型的行动过程

方法：特征提取

1. 第一步所用时间占比及其衍生；
2. 执行间思考：
 - ▶ 除第一步外各步时间的标准差；
 - ▶ 超过（除第一步外）平均值一个标准差以上的步数（定义为思考步数）占总步数的比例；
3. 执行速度：剔除有思考的行动后（包括第一步）后的行动时间的平均值；
4. 冗余步数：被试行动路径中访问到曾访问过的状态的次数；
5. 与最优路径重合比例；
6. 与最优路径相差步数；
7. 各关是否成功、放弃；

方法：特征提取

1. 第一步所用时间占比及其衍生；
2. 执行间思考：
 - ▶ 除第一步外各步时间的标准差；
 - ▶ 超过（除第一步外）平均值一个标准差以上的步数（定义为思考步数）占总步数的比例；
3. 执行速度：剔除有思考的行动后（包括第一步）后的行动时间的平均值；
4. 冗余步数：被试行动路径中访问到曾访问过的状态的次数；
5. 与最优路径重合比例；
6. 与最优路径相差步数；
7. 各关是否成功、放弃；

方法：特征提取

1. 第一步所用时间占比及其衍生；
2. 执行间思考：
 - ▶ 除第一步外各步时间的标准差；
 - ▶ 超过（除第一步外）平均值一个标准差以上的步数（定义为思考步数）占总步数的比例；
3. 执行速度：剔除有思考的行动后（包括第一步）后的行动时间的平均值；
4. 冗余步数：被试行动路径中访问到曾访问过的状态的次数；
5. 与最优路径重合比例；
6. 与最优路径相差步数；
7. 各关是否成功、放弃；

方法：特征提取

1. 第一步所用时间占比及其衍生；
2. 执行间思考：
 - ▶ 除第一步外各步时间的标准差；
 - ▶ 超过（除第一步外）平均值一个标准差以上的步数（定义为思考步数）占总步数的比例；
3. 执行速度：剔除有思考的行动后（包括第一步）后的行动时间的平均值；
4. 冗余步数：被试行动路径中访问到曾访问过的状态的次数；
5. 与最优路径重合比例；
6. 与最优路径相差步数；
7. 各关是否成功、放弃；

方法：降维

- 各关是否成功、放弃，总共有 20×2 个特征，我们通过因素分析提取出 n 个因子，作为最终使用的特征；
- n 由模型拟合效果决定，最终选择为 2；
- 其他特征根据当前关的成功与否，有不同的分类：
 - ▶ 成功时，与最优路径相差步数越大，说明能力越弱；
 - ▶ 失败时，与最优路径相差步数越大，说明越坚持。
- 在个体内按照成功与否分别求平均，构造出 $x2$ 个特征。
- 最终使用了 23 个特征。

方法：模型训练

使用随机森林模型：

- 将样本中的 70% 用于训练：
 - ▶ 交叉验证；
 - ▶ 超参数搜索以获得某项评估指标最优的模型；
- 将获得的模型在剩下 30% 的样本中进行预测，评估该模型的表现。

方法：模型评估指标

	预测为阳性	预测为阴性
实际为阳性	TP	FN
实际为阴性	FP	TN

图：分类任务表现评估表

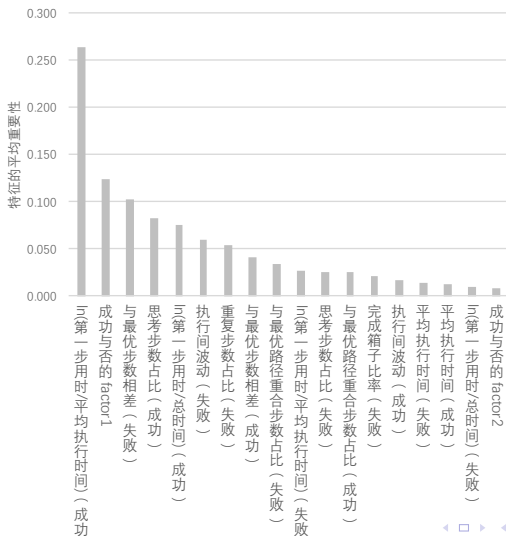
- 准确率 (Accuracy): 准确率是最为简单、直接的一个指标，为正确预测的观测值数量除以总观测值数量，即 $(TP+TN)/(FP+FN+TP+TN)$;
- 精确率 (Precision, P): 精确率述在预测为阳性的样本中，真实阳性的比例，即 $TP/(FP+TP)$;
- 召回率 (Recall, R): 召回率述在所有实际阳性样本中，预测为阳性的比例，即 $TP/(TP+FN)$;
- F1: F1 是召回率与精确率两者之间较为平衡的一个指标，公式为 $F1 = \frac{2RP}{R+P}$ 。

结果

最优化目标	F1	精确率	召回率	准确率
数学成绩				
F1 优先	71.14%	79.35%	71.11%	68.02%
精确率优先	75.57%	83.07%	73.33%	73.70%
召回率优先	73.09%	81.06%	71.78%	70.62%
准确率优先	71.65%	80.19%	69.67%	69.44%
推理能力				
F1 优先	63.83%	74.40%	61.19%	63.46%
精确率优先	63.72%	75.51%	59.17%	65.03%
召回率优先	65.01%	74.91%	63.10%	64.21%
准确率优先	64.22%	76.11%	59.05%	65.72%

图：模型预测结果

结果：特征的相对重要性



讨论

- 我们使用随机森林模型，从推箱子的 log file 中预测学生的数学成绩分类以及推理能力分类，预测结果尚可；
- 为 logfile 分析技术提供了新的思路；
- 需要承认我们只完成了对前后 25% 的预测，目前模型对于连续变量的预测表现不佳，需要进一步工作；
- 后续方向：
 - ▶ 增大样本量，提高被试间的异质性；
 - ▶ 使用与游戏过程数据关联更紧密的结果变量；
 - ▶ 提取更多更有效的特征；

谢谢!