



09 | 实验设计

2018-06-23 王天一

机器学习40讲

[进入课程 >](#)



讲述：王天一

时长 12:36 大小 5.81M



和其他科学学科一样，机器学习也会借助实验获取关于目标的信息。宏观来看，实验的设计与分析正在逐渐脱离具体问题的限定，有成为一门独立学科的趋势。不管是物理学还是经济学，对实验的处理都存在着一些共性的准则。在本篇文章中，我就和你简单谈谈机器学习中有关实验设计与分析的一些原则性问题。

在讨论实验设计之前，先得知道实验设计到底是怎么回事。**实验设计** (experimental design)，或者叫**设计实验** (designed experiment)，指的是在实验之前制定详细的实验计划，确定实验目标并选择待研究的过程因子 (process factor)。精心挑选的实验设计可以在给定资源的条件下使实验获得的信息量最大化，让实验结果最大程度地接近真实结果。实验设计需要人为改变一个或多个过程因子，以观察这种变化对一个或多个因变量的影响，其目的是分析获得的数据以产生有效和客观的结论。

在现有的关于机器学习的文献中，对设计实验部分的讨论似乎并不多见，其原因在于这部分工作已经由他人代劳，而不需要放在机器学习的应用层面来解决。在各种各样的图像识别竞赛中，无论是训练集还是测试集都是预先给定的，其中的每张图片都有精确的标注。看起来，设计实验似乎是一项蓝领工作，它被处理高大上算法的白领工作给人为地屏蔽了。可真实情况是什么呢？通过人工数据训练出来的算法，在真实世界中的行为可能完全不同，从“人工智能”变成“人工智障”只是捅破一层窗户纸这么简单。

设计实验绝非简单的蓝领工作，它甚至比机器学习本身更加注重策略的作用。机器学习的目标是模拟数据的内在生成机制，取得较低的泛化误差和较强的预测精度。对数据生成机制的学习是建立在足量数据的基础上的，因此机器学习需要的数据多多益善，不会对数据提出这样那样的要求。即使是类别不平衡或者属性有缺失值，机器学习对这些非完美的数据也会来者不拒。这就像一个给啥吃啥的傻小子，只要能填饱肚子就无欲无求了。

相比之下，在设计实验中，数据都是经过精挑细选的，最重要的特性就是**平衡**（balance）。一般来说，“平衡”意味着对于数据来说，每个特征可能的取值个数都是相同的，而对于特征来说，每个取值上的数据比例也是要近似相同的。这种平衡保证了每个特征对结果都会产生同等程度的影响，从而回避了特征和因子之间的偏好关系。这就像个讲究生活品质的人，不仅要填饱肚子，还要通过口味和营养的搭配吃得美味、吃得健康。这在实验中就表现为精确衡量不同因子对实验结果的影响。

设计实验要完成的任务是对整个机器学习过程的优化。影响作为一个过程，而不单单是一个模型的机器学习的因子是多种多样的，其中最基本的因子就是**选用的模型形式**。当解决一个分类问题时，主成分分析（通常作为预处理方法出现）、 K 近邻和神经网络都是可用的模型。当模型确定后，**模型的超参数**（hyperparameter）就是下一个可控因素，主成分分析中主成分的个数 d 、 K 近邻算法中 K 的取值，神经网络中隐藏层的数目 n 都属于超参数。超参数会对模型的性能和复杂度产生直接的影响，在不同的超参数配置下，模型训练所得到的参数和性能也会有所区别。

实验中因子的设置取决于实验的目的。如果想要比较不同算法在同一个问题上的性能，算法的类别就是因子；如果要测试固定的算法在不同问题上的性能，算法的超参数就是因子；如果要用算法来区别数据集的数据质量，那么不同的数据集也是因子。设计实验时，要将发生影响的所有因子全部纳入考虑范围并加以调节，从而实现对学习过程的系统性优化。

当实验中的因子数目比较多时，如何唯一地确定单个因子的影响就需要一些技巧。你一定听说过**控制变量法**，这是最简单的实验设计技法之一，它通过将其余变量设置为固定值来观察

单一因素的影响效果，以及其背后的因果关系。在机器学习中，这种方法被称为**一次一因子** (one factor at a time)，它首先为所有因子都设定一个基线值，再在其他因子保持在基线水平的前提下令单个因子波动，观察它对学习性能的影响。

在应用控制变量变量法时需要注意的是，它暗含着一个较强的假设，就是不同因子之间相互独立，互不影响。这个假设在实际的学习任务中显然并不总是成立的，埃塞姆·阿帕丁 (Ethem Alpaydin) 在他的《机器学习导论》 (Introduction to Machine Learning, 3rd Edition) 中提到，在主成分分析的预处理与 K 近邻分类的级联算法中，主成分数目的每个选择都给后面的 K 近邻定义出一个全新的输入空间，这会使 K 近邻的最优超参数出现变化——在 10 维输入中计算出的最优近邻数目未必适用于 15 维的输入。

如果在每次实验中不是控制单个因子，而是让所有的因子一起变化来发现它们之间的协同关系，这就是**因子设计** (factorial design) 的方法。因子设计关注的是不同因子之间系统化的变化对学习效果的影响，它的一个特例是**全因子实验** (full factorial experiment)，也叫**完全交叉设计** (fully crossed design)。在全因子实验中，每个因子都有有限个离散的取值，实验则覆盖了所有因子所有取值的所有可能组合。如果总共有 3 个因子，每个因子的取值数目分别是 3/4/5，那么全因子实验的执行次数就是 $3 \times 4 \times 5 = 60$ 。

全因子实验不仅能够研究每个因子对因变量的影响，还能发现因子之间相互作用对因变量的影响，它可以进一步被看成是**连续实验** (sequential experimentation) 的一个阶段。在待分析的因子数目较多时，连续实验体现出来的就是“**粗调 + 微调**”的处理方式。它首先执行全因子实验，但只给每个因子赋予较少的可能取值。这种广撒网的做法能够评判所有因子的重要程度，确定哪些是对学习结果影响较大的活跃因子并保留下来，剩下的不活跃的因子就会被放弃。接下来，连续实验会聚焦活跃因子，通过增加取值数目并进行精细的微调来精确刻画它们之间的关系，以及对因变量的影响方式，进而优化学习的设置。

在对筛选出的少量因子进行微调时，可以使用**响应面方法** (response surface methodology) 来降低计算开销。微调的目的是找到最优的因子取值，在不可能对所有取值都计算出性能指标的情况下，通过**插值**的方法来拟合出因子和性能之间的响应面就是一种更容易操作的办法。在得到的响应面上寻找最值，找到的就是最优的因子取值。

响应面通常被设定为二次曲面，用来拟合初始曲面的数据通常取在因子的基线附近。在初始曲面上找到的因子最优值被应用到学习模型之中，得到的结果作为一个新样本被添加到初始曲面里面，然后继续对响应曲面进行拟合和优化，直到得到的最优结果没有明显的改进为止。

响应曲面其实可以看成一种替代模型。替代模型 (surrogate model) 是对真实模型的逼近，以数据驱动的自底向上的方法构建，目标是尽可能地模拟真实模型的行为。如果机器学习中的因子较多的话，它们之间的关系可能就是无法用解析式来描述的复杂关系，替代模型就是对这种复杂关系的拟合，就像机器学习训练模型来拟合输入特征和输出结果之间的关系那样。

今天我和你分享了机器学习中实验设计的一些基本原则，其要点如下：

实验设计的任务是观察一个或多个因子对实验结果的影响；

机器学习中，实验设计中的因子包括算法类型、超参数、数据集等；

连续实验可以用来评估多个因子对实验的影响；

响应面方法通过二次曲面的拟合寻找可变因子的最佳取值。

本篇中提到的实验设计类似于机器学习中的模型调参，只不过调参调的是模型内部的参数，而实验设计调的是模型外部的因子。在实际中通常没有得到应有的重视。那么在机器学习的实践中，你是否专门关注过这个过程，又或者掉进过实验设计的坑里，还是历练出什么独门绝技呢？

欢迎分享你的观点。

机器学习中的试验设计

影响因子

法类型、超参数、数据集

一次一因子、因子设计

连续实验，粗调 + 微调

响应面方法

实验设计的任务

对整个机器学习过程的优化

对观察一个或多个结果的影响因子



机器学习

40讲

— 帮你打通机器学习的任督二脉 —

王天一 工学博士，副教授



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金奖励**。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

[上一篇 08 | 模型的评估指标](#)

[下一篇 10 | 特征预处理](#)

精选留言 (5)

写留言



青刀快马

2018-06-24

1

老师，什么时候能具体说下如何创建一个模型？

展开 ▼

作者回复：先根据任务来选择分类模型或者回归模型，再来选择模型具体的形式，比如线性回归或者逻辑回归，模型的形式也可以根据你的数据来调整。确定模型的形式以后就要用数据训练出最优的参数，再对不同的模型进行比较，选出一个最合适的。



王宇直

1

2019-03-13

老师你好，如果把超参也一样用grid search来选择，会有什么不好的结果？

展开 ▼



Kevin.zh...

2018-12-26



作业：王老师是我的机器学习入门老师，所以还没有真正开始的实践，不过后续我会注意老师叮嘱的地方，期待进一步学习！



林彦

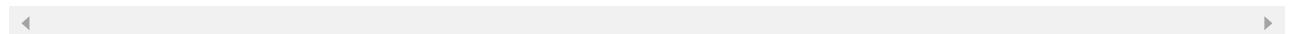
2018-07-02



现在投入的时间有限，实践少，模型的形式，超参数的选择，如何选择衡量的方法和参数都只会照搬，还不太理解后面的因果和细节。因此是否做错了或者掉进什么坑里已自己现在的认知都还未意识到，怎么探索的思路也不清晰。

展开 ▼

作者回复: 其实都是试错试出来的。



手持酱油, ...

2018-06-25



学习了

展开 ▼