

Google 使用机器学习进行设计时的 7 个步骤 (上)

原创：Google 谷歌开发者 2017-09-14



原文标题：以人为中心的机器学习 (Human-Centered Machine Learning)

文 / Josh Lovejoy 和 Jess Holbrook

机器学习 (ML) 是帮助计算机揭示数据模式与关系（而不是依靠手工编程）的科学。机器学习是一个强大的工具，可用于打造个性化的动态体验。机器学习已经有力地推动了从 **Netflix** 建议到自动驾驶汽车等各项技术的发展。然而，随着使用机器学习打造的体验越来越多，显而易见的是，用户体验 (UX) 设计者仍然需要更多地了解如何让用户感觉自己能轻松驾驭科技（而不是反之）。

正如移动革命以及之前网络技术的兴起一样，机器学习将促使我们对构建的几乎所有体验进行反思、重构、置换和考量，从而开辟全新可能。在 Google 用户体验社区，我们启动了一个“以人为中心的机器学习”(HCML) 项目，旨在帮助聚焦和引导上述对话。从这个角度出发，我们审视了各个产品，以查看机器学习如何能够以人类需求为中心并以机器学习所独有的方式来满足这些需求。

在 Google，我们团队与公司内部用户体验设计者合作，探讨如何加速核心机器学习概念的开发，了解如何将机器学习集成到用户体验工具腰带中以及如何确保以包容的方式构建机器学习和人工智能 (AI)。

如果您刚刚开始接触机器学习，机器学习领域的复杂性以及无限的创新机遇可能让您感觉无所适从。不必恐慌，先停下来，给自己留点时间适应一下。您并不需要从头开始，也可以在团队中发挥重要作用。

为了帮助设计师适应设计机器学习驱动的产品这一全新领域，我们提出了七个观点。这些观点源自我们与 Google 用户体验和人工智能团队的合作（并进行了必要的试验，允许合理的错误数量），将帮助您遵循用户优先的原则，实现快速迭代，了解机器学习带来的独特机遇。

我们开始吧！

1. 不要指望机器学习去思考它要解决什么问题

现在，围绕机器学习和人工智能的宣传铺天盖地，天花乱坠。许多公司和产品团队不是首先关注他们要解决什么重大问题，而是一开始就直奔以机器学习作为解决方案的产品战略。



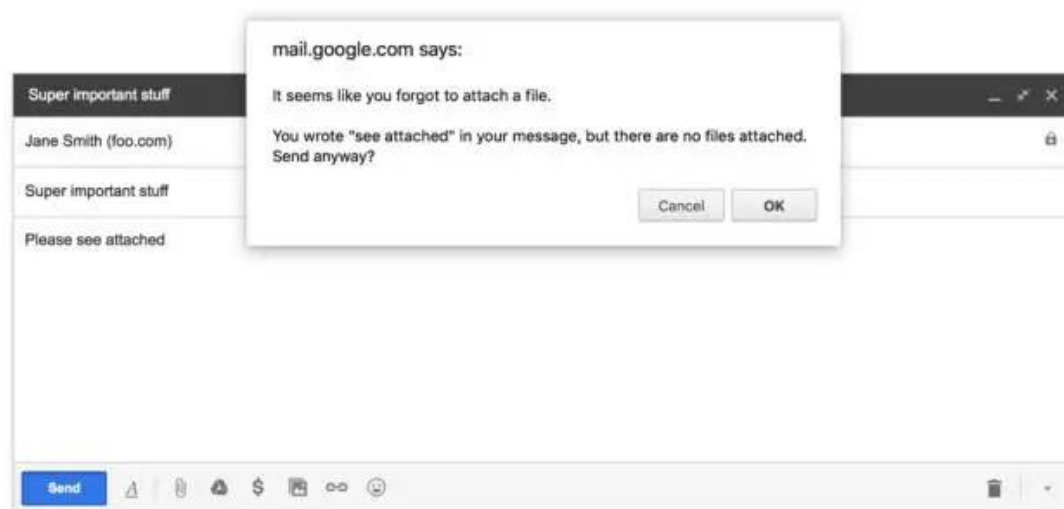
这种模式对纯粹探索或了解某项技术的作用很有用，并且常常会启迪对新产品的思索。然而，如果您并未瞄准人类需求，那么您只是在构建一套非常强大的系统来解决一个很小、也许根本不存在的问题。

因此，我们的第一个观点便是您仍然需要一如既往地努力发掘人类的需求。这涉及到人种学、背景查询、采访、深入研究、调研、查阅客户支持服务单据、日志分析并与人们接触，以弄清楚您是否正在解决某个问题或是满足某个未曾言明的人类需求。不要指望机器学习去思考它要解决什么问题。我们仍然需要定义问题。作为用户体验设计者，我们已经拥有指导我们团队的工具，无论主流的技术范式为何。

2. 询问自己机器学习是否能以独特方式解决问题

一旦您确认您需要满足哪些需求之后，您将需要评估机器学习是否能够以独特方式满足这些需求。有许多符合前述条件的问题并不需要机器学习解决方案。

目前，产品开发中的一大挑战是确定哪些体验需要机器学习，哪些体验可通过机器学习得到实质性增强，哪些体验不能通过机器学习得到改善，甚至会反受其害。有很多产品无需使用机器学习，也可以让人感觉很“智能”或“人性化”。不要想当然地认为只有机器学习才能解决问题。



▲ Gmail 会查找包含类似“attachment”和“attached”之类词语的短语，以便在您忘记插入附件时弹出提醒。启发式设计在这里就很有效。尽

管机器学习系统很可能能够捕获更多的潜在失误，但其构建成本却要高得多。

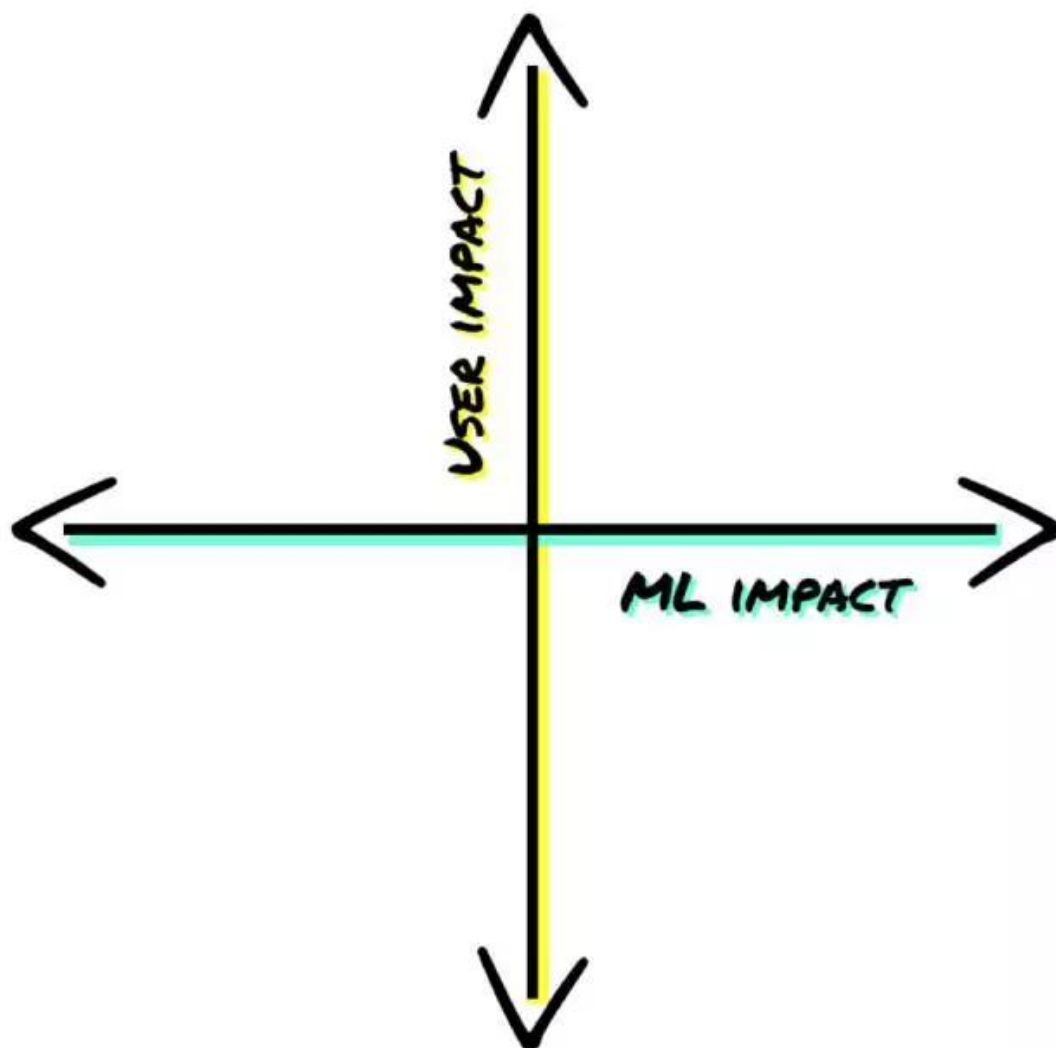
我们设计了一组练习来帮助团队理解机器学习对于其用例的价值。为此，这些练习深入探究人们在与机器学习系统交互时可能提出的具体心智模式和期望以及该系统需要哪些数据。

以下是我们让团队围绕他们要使用机器学习解决的用例进行研究和解答的三个示例练习：

- 描述理论上的人类“专家”目前执行此任务的方法。
- 如果您的人类专家要执行此任务，您如何为他们提供响应，以便他们在下一次能改进方法？对于**混淆矩阵**中的所有四个阶段均执行上述步骤。
- 如果由人类来执行此项任务，用户会希望他们做出哪些假设？

请花几分钟时间回答各个问题，了解人类能为机器学习驱动的产品提供的自动假设。这些假设如同产品团队在探讨产品时的提示一样重要，或者如同用户研究时的刺激一样有用。我们稍后讲解定义标签和训练模型时还会再讲到这方面的内容。

在完成这些练习以及特定产品和其他一些**草图**和**串连图板**后，我们将团队关于产品的各种想法分别写入一个简单的 **2x2** 网格中：



▲ 将想法写入这个 2x2 网格中。让团队投票选择哪些想法能够给用户带来最大的影响，哪些想法通过机器学习解决方案可以得到最大程度的加强。

这有助于我们区分作用显著的想法和作用不大的想法，区分能够通过机器学习加以增强的想法和机器学习对其无甚帮助或帮助不大的想法。在这些对话期间，您应该已经开始与工程团队合作，但如果您尚未开始合作，那现在正是参照机器学习的发展现状衡量一下这些想法

的好时机。您应当首先专注于机器学习能够以独特方式给用户带来最大影响（见上述矩阵右上角）的想法。

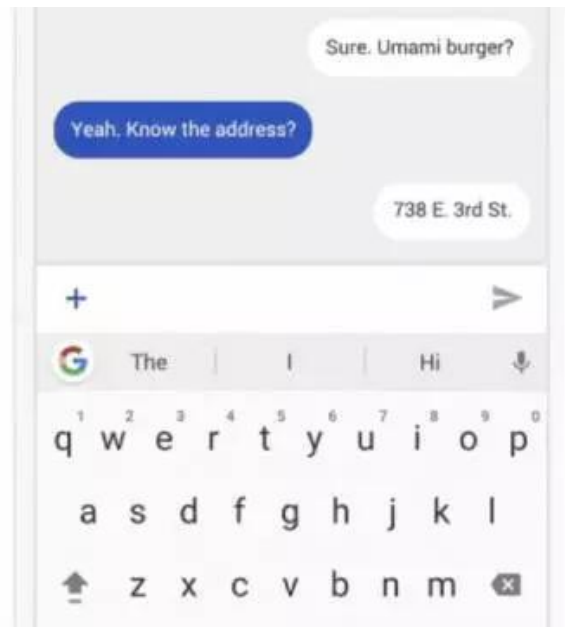
3. 利用个人数据和向导进行模仿

原型设计是机器学习系统面临的一大挑战。如果您的产品的所有价值在于使用独特的用户数据来打造专属定制体验，那就注定不能快速设计出贴近真实的原型。同样，如果您等待建成一套齐全的机器学习系统来测试设计，那在测试完毕后再进行任何有意义的更改很可能都为时已晚。但有两种用户研究方法可以帮助您：使用参与者的个人案例和 **Oz** 研究向导。

在利用早期模型进行用户研究时，让参与者为课程提供一些他们自己的数据，例如个人照片、联系人列表，或者他们收到的音乐或电影建议。请记住，务必明确完整地告知参与者在测试期间会如何使用这些数据以及何时删除这些数据。您甚至可以将这作为参与者的趣味“家庭作业”，要求他们在课前完成（毕竟大家都喜欢讨论自己喜欢的电影）。

您可以利用这些示例数据来模拟系统的正确或错误响应。例如，您可以模拟系统向用户推荐不合适的电影来观察其反应，了解用户对系统返回该结果的原因做出的假设。相比使用虚拟数据或概念性说明，这种方法能够帮助您更有效地评估这些可能性涉及的成本与收益。

第二种有效测试尚未成型的机器学习产品的方法是执行 Oz 研究向导。这种用户研究方法曾经十分流行，但在过去约 20 年间渐失风头。现在它们又卷土重来了。



▲ 适合使用 Oz 向导方法测试的最简单的体验之一便是聊天界面。只需安排团队成员在聊天的另一端输入“人工智能”提供的“回答”。

快速提醒：Oz 研究向导参与者认为自己是在与自主系统进行互动，但系统其实是由人控制的（通常由团队成员控制）。

让团队成员模仿机器学习系统的操作（例如聊天回复，建议参与者应该给谁拨打电话，或者电影建议），可以模拟与“智能”系统互动。这些互动对于指导设计思路至关重要，因为参与者真正认为自己是在与人工智能互动时，他们会自然而然地形成系统的心智模式，并根据这些模式调整自己的行为。观察他们的调整以及与系统的二阶互动非常有助于启迪设计灵感。

4. 权衡误报和漏报的成本

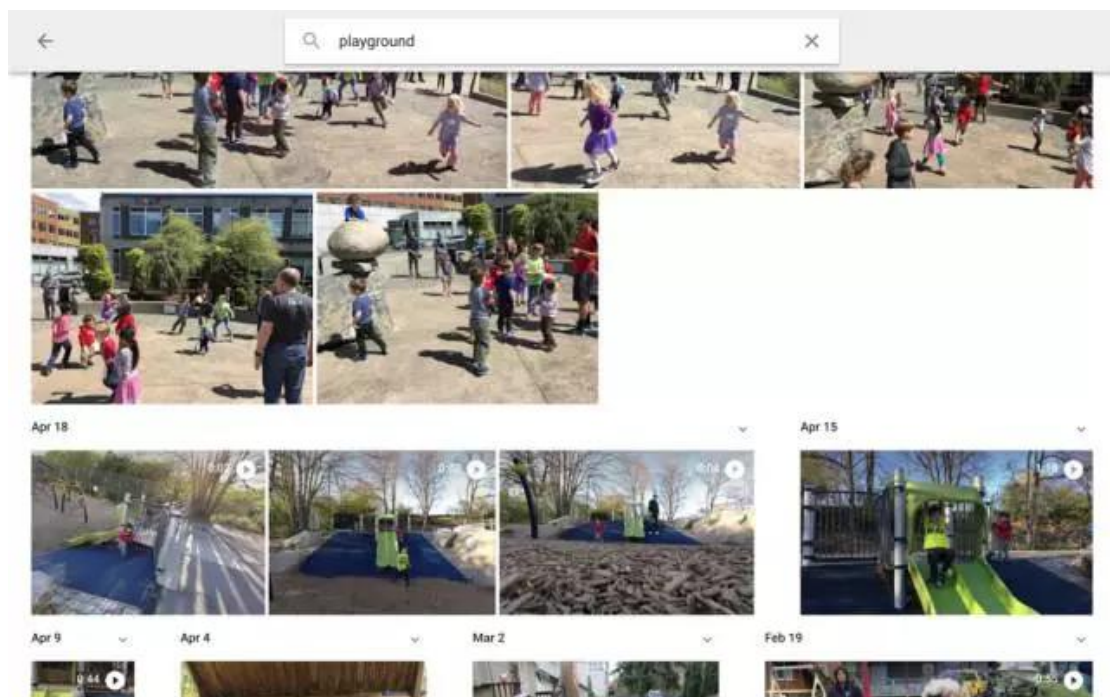
机器学习系统出错在所难免。了解这些错误的性质以及它们对产品用户体验的影响非常重要。我们在第 2 点的一个问题中提到了混淆矩阵这一概念。这是机器学习中的一个关键概念，描述了当机器学习系统判断正确或出错时的情况。

		Prediction	
		Positive	Negative
Reference	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

▲ 混淆矩阵的四个状态，以及可能对用户意味着什么。

所有错误对于机器学习系统而言没什么不同，而对于人则不是这样。例如，如果有一个“这是人类还是怪物？”的分类器，那么偶然地将一个人归类为怪物对系统来说只是一个错误。它不会认为所做的分类会冒犯用户或其文化背景。它也不会明白，与将怪物意外地标记为人相比，系统用户对于被意外标记为怪物会感到更不舒服。但这可能就是我们以人为中心的偏见：)

用机器学习术语来说，您需要有意识地在系统的准确率和召回率之间进行权衡。也就是说，您需要决定是包括所有正确答案重要，即使这意味着会加入较多错误答案（优化召回率），还是宁可遗漏一些正确答案也要最大程度减少错误答案（优化准确率）重要。例如，如果您在 Google 照片中搜索“游乐场”，您可能会看到如下结果：



这些结果包含几个儿童玩耍的场景，但不是在游乐场上。在此案例中，召回率优先于准确率。与仅包含游乐场照片但可能排除了您要寻找的照片相比，获取所有游乐场照片但包含几个相似但不完全正确的照片更重要。