

文章

[Michael Lei](#) · 四月 14, 2021 阅读大约需 8 分钟

通过 InterSystems IRIS 实现分布式人工智能

什么是分布式人工智能 (DAI) ?

试图找到一个“无懈可击”的定义是徒劳的：这个术语似乎有些“超前”。但是，我们仍然可以从语义上分析该术语本身，推导出分布式人工智能也是人工

智能（请参见我们为提出一个“实用”定义所做[努力](#)）

），只是它分布在多台没有聚合在一起（既不在数据方面，也不通过应用程序聚合，原则上不提供对特定计算机的访问）的计算机上。即，在理想情况下，分布式人工智能的安排方式是：参与该“分布”的任何计算机都不能直接访问其他计算机的数据和应用程序，唯一的替代方案是通过“透明的”消息传递来传输数据样本和可执行脚本。

与该理想情况的任何偏差都会导致出现“部分分布式人工智能” - 一个示例是通过中央应用程序服务器分发数据，或者其反向操作。不管怎样，我们都会得到一组“联合”模型（即，在各自数据源上训练的模型，或者按自己的算法训练的模型，或者同时以这两种方式训练的模型）。

“面向大众”的分布式人工智能方案

我们不会讨论边缘计算、机密数据操作员、分散的移动搜索，或者类似的引人入胜但又不是最有意识和最广泛应用（目前不是）的方案。

我们将更“贴近于生活”，例如，如果考虑以下方案（其详细演示应该可以[在此处查看](#)）：一家公司运行一个生产级 AI/ML 解决方案，其运行质量正在由一名外部数据科学家（即，不是该公司员工的专家）系统地进行检查。由于种种原因，该公司无法授予数据科学家访问该解决方案的权限，但可以按照时间表或在特定事件（例如，解决方案终止一个或多个模型的训练会话）后向其发送所需表中的记录样本。我们据此假定，该数据科学家拥有某个版本的 AI/ML 机制，且这些机制已经集成在公司正在运行的生产级解决方案中，而且数据科学家本人很可能正在开发、改进和调整这些机制，以适应该具体公司的具体用例。将这些机制部署到正在运行的解决方案中、监控机制的运行以及其他生命周期方面的工作由一名数据工程师（公司员工）负责。

我们在这篇文章中提供了一个在 InterSystems IRIS 平台上部署生产级 AI/ML

解决方案的示例，该解决方案可自主处理来自设备的数据流。

上一段提供的链接下的演示中也运行了相同的解决方案。您可以使用我们的仓库 [Convergent Analytics](#) 中的内容（免费且无时间限制，请访问 [Links to Required Downloads](#) 和 [Root Resources](#) 部分）在 InterSystems IRIS 上构建您自己的解决方案原型。

通过这样的方案，我们可获得“分布程度”如何的人工智能？在我们看来，此方案相当接近理想情况，因为数据科学家与公司的数据和算法均保持“切割”（只是传输了有限的样本，但在某个时间点前很重要；而且数据科学家自己的“样本”永远不会与作为实时生产级解决方案的一部分部署和运行的“活跃”机制 100% 同步），他完全不能访问公司的 IT 基础架构。因此，数据科学家的作用变为在他的本地计算资源上部分重放该公司生产级 AI/ML 解决方案的运行片段，获得可接受置信级别的运行质量评估，并向公司返回反馈（在我们的具体方案中，以“审核”结果表示，可能还加上该公司解决方案涉及的 AI/ML 机制的改进版本）。

图 1 分布式人工智能方案表示

我们知道，在人类执行人工智能项目交换的过程中，不一定需要表示和传输反馈，这源于有关现代方法的出版物以及已有的分布式人工智能实现经验。但是，InterSystems IRIS 平台的优势在于，它允许同样高效地开发和启动“混合”（人类和机器串联）且完全自动化的人工智能用例，因此，我们将继续根据上述“混合”示例进行分析，同时为读者留下自行阐述其完全自动化的可能性。

如何在 InterSystems IRIS 平台上运行具体的分布式人工智能方案

本文上一节提到的带有方案演示的视频[介绍](#)对作为实时 AI/ML 平台的 InterSystems IRIS

进行了总体概述，并解释了其对 DevOps 宏机制的支持。在演示中，没有明确覆盖负责定期将训练数据集传输给外部数据科学家的“公司侧”业务流程，因此，我们将从该业务流程及其步骤的简介开始。

发送方业务流程的一个主要“引擎”是 while 循环（使用 InterSystems IRIS 可视业务流程编辑器实现，该编辑器基于平台解释的 [BPL](#) 表示法），负责将训练数据集系统地发送给外部数据科学家。该“引擎”内部执行以下操作（参见下图，跳过数据一致性操作）：

图 2 “发送方”业务流程的主要部分

(a) 负载分析器 – 将训练数据集表中的当前记录集加载到业务流程中，并基于它在 Python 会话中形成一个数据框架。调用操作会触发对 InterSystems IRIS DBMS 的 SQL 查询，并触发对 Python 接口的调用以将 SQL 结果传输给它，以便形成数据框架；

(b) Azure 分析器 – 另一个调用操作，触发对 Python 接口的调用，以向其传输一组 Azure ML SDK for Python 指令，从而在 Azure 中构建所需的基础架构，并在该基础架构上部署前一个操作中形成的数据框架数据；

作为执行上述业务流程操作的结果，我们在 Azure 中获得一个存储对象（一个 .csv 文件），其中包含公司的生产级解决方案用于模型训练的最近数据集的导出：

图 3 训练数据集“到达”Azure ML

这样，发送方业务流程的主要部分已经结束，但是我们还需要再执行一个操作，同时请记住，我们在 Azure ML 中创建的任何计算资源都是可计费的（参见下图，跳过数据一致性操作）：

图 4 “发送方”业务流程的最后部分

(c) 资源清理 – 触发对 Python 接口的调用，以向其传输一组 Azure ML SDK for Python 指令，从 Azure 中删除上一个操作中构建的计算基础架构。

数据科学家所需的数据已经传输完毕（数据集现在在 Azure 中），因此我们可以继续启动将访问数据集的“外部”业务流程，运行至少一次替代模型训练（从算法上讲，替代模型不同于作为生产级解决方案一部分运行的模型），并向数据科学家返回得到的模型质量指标及可视化，从而表示有关公司生产级解决方案运行效率的“审核结果”。

我们现在看一下接收方业务流程：与发送方业务流程（在包含公司自主 AI/ML 解决方案的其他业务流程中运行）不同，它不需要 while 循环，而是包含与在 Azure ML 和 IntegratedML（InterSystems IRIS 中用于自动 ML 框架的加速器）中训练替代模型有关的一系列操作，并将训练结果提取到 InterSystems IRIS 中（该平台也被认为在数据科学家处本地安装）：

图 5 “接收方”业务流程

(a) 导入 Python 模块 – 触发对 Python 接口的调用，以向其传输一组指令，导入进一步操作所需的 Python 模块；

(b) 设置 AUDITOR 参数 – 触发对 Python 接口的调用，以向其传输一组指令，为进一步操作所需的变量指定默认值；

(c) Azure ML 审核 –（我们将跳过任何对 Python 接口触发的进一步引用）将“审核任务”提交到 Azure ML；

(d) 解释 Azure ML – 将发送方业务流程传输到 Azure ML 的数据与 Azure ML 的“审核”结果一起获取到本地 Python 会话中（此外，在 Python 会话中创建“审核”结果的可视化）；

(e) 流式传输到 IRIS – 将发送方业务流程传输到 Azure ML 的数据与 Azure ML 的“审核”结果一起从本地 Python 会话中提取到 IRIS 中的业务流程变量；

(f) 填充 IRIS – 将发送方业务流程传输到 Azure ML 的数据与 Azure ML 的“审核”结果一起从 IRIS 中的业务流程变量写入 IRIS 中的表；

(g) IntegratedML 审核 – 使用 IntegratedML 加速器“审核”从 Azure ML 接收的数据以及上一个操作中写入 IRIS 的

Azure ML “审核”结果（在此特定情况下，该加速器处理 H2O auto-ML 框架）；

(h) 对 Python 进行查询 – 将数据和 IntegratedML “审核”结果传输到 Python 会话中；

(i) 解释 IntegratedML – 在 Python 会话中，创建 IntegratedML “审核”结果的可视化；

(j) 资源清理 – 从 Azure 中删除先前的操作中创建的计算基础架构。

图 6 Azure ML “审核”结果的可视化

图 7 IntegratedML “审核”结果的可视化

分布式人工智能一般如何在 InterSystems IRIS 平台上实现

InterSystems IRIS 平台实现分布式人工智能有三种基本方法：

- 根据用户定义的规则和算法，直接交换人工智能项目，并对其进行本地和中央处理
- 人工智能项目处理委托给专门的框架（例如：TensorFlow、PyTorch），交换的编排和各个准备步骤由用户在 InterSystems IRIS 的本地和中央实例上配置
- 人工智能项目的交换和处理都通过云提供商（Azure、AWS、GCP）来完成，本地和中央实例只向云提供商发送输入数据并接收最终结果

图 8 在 InterSystems IRIS 平台上实现分布式人工智能的基本方法

这些基本方法可以修改/组合使用：尤其是，在本文的上一节（“审核”）所描述的具体方案中，使用了第三种方法“以云为中心”，将“审核员”部分划分到云端，而在数据科学家一侧执行本地部分（作为“中央实例”）。

目前，在我们生活的现实中，“分布式人工智能”学科的理论和应用要素正在不断积累，但还没有形成“规范形式”，这使得创新实现具有巨大潜力。我们的专家团队密切关注分布式人工智能作为一门学科的发展，并为其在 InterSystems IRIS 平台上的实现设计加速器。

我们乐于分享我们的内容，并帮助每一个认为这里讨论的领域有用的人开始分布式人工智能机制的原型设计。

您可以使用以下电子邮件地址联系我们的 AI/ML 专家团队 – MLToolkit@intersystems.com。

[#AI](#) [#云](#) [#分布式数据管理](#) [#机器学习](#) [#融合分析](#) [#InterSystems IRIS](#)

源

URL: <https://cn.community.intersystems.com/post/%E9%80%9A%E8%BF%87-intersystems-iris-%E5%AE%9E%E7%8E%B0%E5%88%86%E5%B8%83%E5%BC%8F%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD>