

文章

Michael Lei · 二月 17, 2022 阅读大约需 23 分钟

InterSystems IRIS – 适用于实时人工智能/机器学习的多功能通用平台

实时人工智能/机器学习计算的挑战

我们将从我们在 InterSystems 数据科学实践中遇到的示例开始讲起：

- “高负载”客户门户与在线推荐系统集成。
计划是在整个零售网络层面重新配置促销活动（我们将假设使用“细分策略”矩阵而非“平面”促销活动母版）。推荐机制会有哪些变化？推荐机制内的数据馈送和数据更新会有哪些变化（输入数据量增加了 25000 倍）？推荐规则生成设置会有哪些变化（生成规则的总量和“分类”呈千倍增加，因此需要将推荐规则筛选阈值缩小千倍）？
- 设备健康监控系统使用“手动”方式馈送数据样本。现在，它将连接到每秒可传输数千个过程参数读数的 SCADA 系统。监控系统会有哪些变化（它能否应对以秒为单位的设备健康监控）？当输入数据接收到包含数百列最近在 SCADA 系统中实现的数据传感器读数的新块时，会发生什么（是否有必要关闭监控系统以将新的传感器数据整合到分析当中，以及要关闭多久）？
- 复杂的人工智能/机器学习机制（推荐、监控、预测）依赖于彼此的结果。
要调整这些人工智能/机器学习机制的功能以适应输入数据的变化，每月需要多少人工工时？人工智能/机器学习机制在支持制定业务决策方面的总体“延迟”是多少（支持信息的刷新频率对比新输入数据的馈送频率）？

综合考虑包括上述在内的示例后，我们总结了以下因过渡到使用实时机器学习和人工智能而出现的挑战：

- 我们对于自家公司的人工智能/机器学习机制的创建和适应速度（相对于形势变化的速度）是否满意？
- 我们的人工智能/机器学习解决方案在支持制定实时业务决策方面的表现如何？
- 我们的人工智能/机器学习解决方案能否自适应（即无需开发者介入即可继续工作）以应对数据漂移和由此产生的业务决策制定方法的变化？

本文综合概述了 InterSystems IRIS 平台在全面支持人工智能/机器学习机制部署、人工智能/机器学习解决方案装配（集成）和基于密集数据流的人工智能/机器学习解决方案训练（测试）方面的功能。我们将关注市场研究、人工智能/机器学习解决方案实例以及我们在本文中称为实时人工智能/机器学习平台的概念方面的内容。

调查内容：实时应用类型

由 Lightbend 在 2019 年面向约 800 名专业 IT 人士进行的调查，结果不言自明：

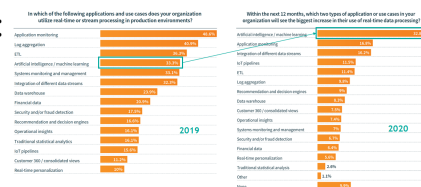


图 1：实时数据的主要使用者

我们将引用该调查结果报告中对我们最重要的片段：“… 流式传输数据流水线和基于容器的基础架构有着并驾齐驱的增长趋势，二者相结合可应对在更快、更高效且更敏捷地交付有影响力的结果方面存在的竞争压力。

与传统的批处理相比，流式传输能够更快地从数据中提取有用信息。

它还可以及时地集成高级分析数据，例如基于人工智能和机器学习 (AI/ML)

模型的推荐，所有功能都旨在通过提高客户满意度来实现差异化竞争优势。时间压力也会对 DevOps

团队构建和部署应用造成影响。诸如 Kubernetes 等基于容器的基础架构可帮助常要通过快速、重复构建和部署应用以应对变化的团队化解其面对的许多效率低下和设计问题。… 804 位 IT

专业人士提供了有关其组织内使用流式处理的应用的详细信息。受访者主要来自西方国家/地区（欧洲占比 41%，北美占比 37%），任职于大中小型组织的人数占比大致相等。… 人工智能并非投机炒作。

在已于人工智能/机器学习生产应用中使用流式处理的受访者中，有 58% 的人表示明年将出现一些最大幅增长。

- 受访者一致认为人工智能/机器学习用例将于明年出现一些最大幅增长。
- 随着实时数据处理得到更大规模的利用，不仅会拓宽在不同用例中采用的广度，还会加大在现有用例中采用的深度。
- 除了人工智能/机器学习之外，物联网流水线采用者的热情也十分高昂 — 有 48% 已经整合物联网数据的人员表示，此用例将于近期出现一些最大幅增长。 ... ”

这项非常有趣的调查表明，机器学习和人工智能场景是实时数据的主要使用者，这种看法已得到广泛认可。

另一个重要的收获是透过 DevOps 视角折射出的对于人工智能/机器学习的看法：我们现在已经可以断言，仍占主导地位的“基于完全已知数据集的一次性人工智能/机器学习”文化已发生转变。

实时人工智能/机器学习平台概念

实时人工智能/机器学习最典型的使用领域之一是工业中的制造过程管理。

让我们以这一领域为例并考虑上述所有想法，为实时人工智能/机器学习平台的概念给出确切的定义。

使用人工智能和机器学习来满足制造过程管理的需求具备多项与众不同的特性：

- 有关制造过程状况的数据会非常密集地生成：生成频率高、涉及参数广泛（SCADA 系统每秒可传输数以万计的参数值）
- 有关检测到的缺陷的数据（且不说不断演化的缺陷，它们反而是稀少且偶发的数据）已知存在缺陷分类不足和定位不及时的问题（通常以人工记录的形式在纸上找到）
- 从实际角度来看，模型训练和应用只有一个“观察窗口”，可反映以最近的过程参数读数作为结尾的合理移动间隔内的过程动态

这些特性使我们除了需要实时接收和基本处理来自制造过程的密集“宽带信号”之外，也需要以实时方式执行（并行）人工智能/机器学习模型应用、训练和准确率控制。我们的模型在移动观察窗口中“看到”的“框架”在不断变化 – 基于之前某一“框架”训练的人工智能/机器学习模型在准确率方面也会发生变化。如果人工智能/机器学习建模准确率下降（例如，“警报范数”分类误差的值超过了给定的容差边界），则应自动触发基于较新“框架”的重新训练 – 同时，在选择开始重新训练的時刻时，必须考虑到重新训练过程持续时间和当前模型版本的准确率下降速度（因为在获得“重新训练”版本的模型之前，在执行重新训练的过程期间会一直持续应用当前版本）。InterSystems IRIS 拥有关键的平台内功能，可充分支持用于制造过程管理的实时人工智能/机器学习解决方案。

这些功能可以分为三大类：

- 基于 InterSystems IRIS 平台实时运行的生产解决方案中新的或经修改的现有人工智能/机器学习机制的持续部署/交付 (CD)
- InterSystems IRIS 平台的单一生产解决方案中的入站过程数据流、人工智能/机器学习模型应用/训练/准确率控制队列、围绕与数学建模环境实时交互的数据/代码/编排的持续集成 (CI)
- 在使用传递自 InterSystems IRIS 平台的数据、代码和编排（“决策制定”）的数学建模环境中执行的人工智能/机器学习机制的持续训练 (CT)

将与机器学习和人工智能相关的平台功能归入上述类别并非随意而为。我们引用了由 Google

发表的方法论[文章](#)，文中为这种分组提供了概念基础：“... DevOps 是开发和运行大规模软件系统的一种常见做法。

这种做法具有诸多优势，例如缩短开发周期、提高部署速度、实现可靠的发布。

如需获得这些优势，您需要在软件系统开发中引入两个概念：

- 持续集成 (CI)
- 持续交付 (CD)

机器学习系统是一种软件系统，因此类似的做法有助于确保您能够可靠地大规模构建和运行机器学习系统。

但是，机器学习系统在以下方面与其他软件系统不同：

- 团队技能：在机器学习项目中，团队通常包括数据科学家或机器学习研究人员，他们主要负责进行探索性数据分析、模型开发和实验。这些成员可能不是经验丰富的、能够构建生产级服务的软件工程师。
- 开发：机器学习在本质上具有实验性。

您应该尝试不同的特征、算法、建模技术和参数配置，以便尽快找到问题的最佳解决方案。您所面临的挑战在于跟踪哪些方案有效、哪些方案无效，并在最大程度提高代码重复使用率的同时维持可重现性。

- 测试：测试机器学习系统比测试其他软件系统更复杂。除了典型的单元测试和集成测试之外，您还需要验证数据、评估经过训练的模型质量以及验证模型。
- 部署：在机器学习系统中，部署不是将离线训练的机器学习模型部署为预测服务那样简单。机器学习系统可能会要求您部署多步骤流水线以自动重新训练和部署模型。此流水线会增加复杂性，并要求您自动执行部署之前由数据科学家手动执行的步骤，以训练和验证新模型。
- 生产：机器学习模型的性能可能会下降，不仅是因为编码不理想，而且也因为数据资料在不断演变。换句话说，与传统的软件系统相比，模型可能会通过更多方式衰退，而您需要考虑这种降级现象。因此，您需要跟踪数据的摘要统计信息并监控模型的在线性能，以便系统在值与预期不符时发送通知或回滚。

机器学习和其他软件系统在源代码控制的持续集成、单元测试、集成测试以及软件模块或软件包的持续交付方面类似。但是，在机器学习中，有一些显著的差异：

- CI 不再仅仅测试和验证代码及组件，而且还会测试和验证数据、数据架构和模型。
- CD 不再针对单个软件包或服务，而会针对自动部署其他服务（模型预测服务）的系统（机器学习训练流水线）。
- CT 是机器学习系统特有的一个新属性，它主要涉及自动重新训练和提供模型。...

我们可以得出结论，基于实时数据的机器学习和人工智能需要更为广泛的工具和更加丰富的功能（从代码开发到数学建模环境编排）、所有功能和主题领域之间更加紧密的集成、更好的人力及机器资源管理。

实时场景：识别进料泵中出现的缺陷

继续以制造过程管理领域为例，我们将探讨一个已在开头引用过的实际案例：需要建立一种基于制造过程参数值流以及维护人员缺陷检测报告来实时识别进料泵中出现的缺陷的机制。

图 2：识别出现缺陷的案例解析

许多类似的实际案例中都有一个共同的特点，即在考虑规律且及时地馈送数据 (SCADA) 时，还需要同时考虑偶发且不规则地检测（和记录）各种缺陷类型。换句话说：SCADA 数据每秒馈送一次以供分析，但需要用纸笔记录缺陷并注明日期（例如：“1 月 12 日-第三轴承区域漏油渗入泵盖”）。因此，我们可以通过添加以下重要限制来对案例解析加以补充：我们只有一种具体缺陷类型的“指纹”（即具体缺陷类型由截至具体日期的 SCADA 数据表示 - 该特定缺陷类型没有其他示例）。这一限制立即使我们脱离了假定有大量可用“指纹”的传统机器学习范式（监督学习）。

图 3：细化缺陷识别案例解析

我们能否以某种方式“倍增”我们可用的“指纹”？是的，可以。泵的当前状况由其与已记录缺陷的相似度表征。即使不运用定量方法，仅通过观察从 SCADA 系统接收到的参数值的动态，也可以了解很多信息：

图 4：泵状况动态与具体缺陷类型“指纹”

然而，视觉感知（至少目前）在我们动态发展的场景中并不是最合适的机器学习“标签”生成器。

我们将使用统计检验来评估当前泵状况与已记录缺陷的相似度。

图 5：对传入数据与缺陷“指纹”应用统计检验

统计检验可评估一组包含制造过程参数值的记录（以“批次”形式取自 SCADA 系统）与具体缺陷“指纹”的记录相似的概率。使用统计检验（统计相似度指数）评估的概率随后会被转换为 0 或 1，成为我们评估相似度的每组记录中的机器学习“标签”。

即，使用统计检验处理获取到的一批泵状况记录后，我们就能够 (a)

将该批次添加到人工智能/机器学习模型的训练数据集以及 (b)

评估人工智能/机器学习模型当前版本应用于该批次时的准确率。

图 6：对传入数据与缺陷“指纹”应用机器学习模型

在之前的一期[在线讲座](#)中，我们展示并讲解了 InterSystems IRIS

平台如何将任何人工智能/机器学习机制实现为能够控制建模输出似然并调整模型参数的持续执行的业务流程。

我们实现泵场景依赖于在线讲座中介绍的完整 InterSystems IRIS 功能 - 这些功能在作为我们的解决方案一部分的 ANALYZER 流程中使用，通过自动管理训练数据集实现强化学习，取代了传统的监督学习。

我们会将在应用统计检验（相似度指数转换为 0 或

1) 和当前版本的模型之后呈现“检测一致性”（即统计检验和模型基于这些记录均输出

1) 的记录添加到训练数据集中。重新训练模型时，在其验证阶段（将新训练的模型应用于自身的训练数据集，在此之前需要对该数据集提前进行统计检验），应用统计检验后“未能保持”输出 1 的记录（由于训练数据集中永远存在属于原始缺陷“指纹”的记录）将从训练数据集中移除，并基于缺陷“指纹”和数据流中的“成功”记录训练新版本模型。

图 7：InterSystems IRIS 中人工智能/机器学习计算的机器人化

如果需要对通过 InterSystems IRIS

中的本地计算获得的检测准确率寻求“第二意见”，我们可以创建一个顾问流程，以使用云提供商（例如 Microsoft Azure、Amazon Web Services、Google Cloud Platform 等）基于控制数据集重做模型训练/应用）：

图 8：来自 Microsoft Azure 的“第二意见”，由 InterSystems IRIS 编排

我们场景的原型在 InterSystems IRIS 中实现为分析过程的代理系统，与设备（泵）、数学建模环境（Python、R 和 Julia）交互，并支持所有相关人工智能/机器学习机制的自我训练 – 基于实时数据流。

图 9：InterSystems IRIS 中实时人工智能/机器学习解决方案的核心功能

基于我们的原型获得的一些实际结果：

- 由模型检测到的缺陷“指纹”（1 月 12 日）：
- 模型检测到的原型已知“指纹”中未包含的新出现的缺陷（于 9 月 11 日检测到，而缺陷本身在两天后，即 9 月 13 日被维修队发现）：

对包含多次发生相同缺陷的真实数据进行的模拟表明，我们使用 InterSystems IRIS 平台实现的解决方案可以在维修队发现问题的几日之前检测到缺陷。

InterSystems IRIS – 适用于实时人工智能/机器学习计算的多功能通用平台

[InterSystems IRIS](#) 是一款完整、统一的平台，可简化实时富数据解决方案的开发、部署和维护。它提供了并发事务和分析处理能力，支持多个完全同步的数据模型（关系、分层、对象和文档），一个可集成不同数据孤岛和应用的完整的互操作性平台，以及支持批处理和实时用例的复杂结构化和非结构化分析功能。

该平台还提供了一个开放的分析环境，可将同类最佳的分析整合到 InterSystems IRIS

解决方案中，并提供灵活的部署功能以支持云和本地部署的任意组合。由 InterSystems IRIS 平台提供支持的应用目前已在各行各业中得到广泛使用，帮助公司在战略和战术执行中获得切实的经济利益，促进明智的决策制定并消除事件、分析和行动之间的“差距”。

图 10：实时人工智能/机器学习背景下的 InterSystems IRIS 架构

与上图相同，下图将新的“基础”（CD/CI/CT）与平台工作元素之间的信息流结合起来。可视化始于 CD 宏机制，并继续贯穿于 CI/CT 宏机制。

图 11：InterSystems IRIS 平台人工智能/机器学习工作元素之间的信息流图

InterSystems IRIS 中 CD 机制的基本要素：平台用户（人工智能/机器学习解决方案开发者）使用专门的人工智能/机器学习代码编辑器调整现有和/或创建新的人工智能/机器学习机制。上述代码编辑器为 Jupyter（全称：Jupyter Notebook；为简洁起见，在此编辑器中创建的文档也常被称为相同的名称）。在 Jupyter

中，开发者可以在传输（“部署”）到 InterSystems IRIS

之前编写、调试和测试（也使用可视化表示）具体的人工智能/机器学习机制。

显然，以这种方式开发的新机制只能进行基本的调试（特别是因为 Jupyter 不处理实时数据流） –

但我们对此没有意见，因为原则上，在 Jupyter

中开发代码的主要目标是验证单独的人工智能/机器学习机制的功能。类似地，已部署在平台中的人工智能/机器学习机制（请参阅其他宏机制）在调试前可能需要“回滚”到其“平台前”版本（从文件读取数据、通过 xDBC 而非本地表或 global（即 InterSystems IRIS 中的多维数据数组）访问数据等）。在 InterSystems IRIS 中实现 CD

具有一项重要特性：平台和 Jupyter

之间存在双向集成关系，支持在平台内部署（利用进一步的平台内处理）Python、R 和 Julia

内容（三种语言均为各自主要开源数学建模环境的编程语言）。

也就是说，人工智能/机器学习内容开发者获得了在平台中“持续部署”其内容的能力，同时能够使用其常用 Jupyter 编辑器以及 Python、R、Julia 所提供的常用函数库，在平台外部提供基本的调试功能（必要时）。继续聊聊

InterSystems IRIS 中的 CI 宏机制。

该图展示了“实时机器人化”的宏流程（一组数据结构、以数学环境语言编写的业务流程和代码段，以及以其编排的 InterSystems IRIS 原生开发语言 ObjectScript）。宏流程的目标是：支持人工智能/机器学习机制运行所需的数据处理队列（基于实时传输到平台的数据流），对人工智能/机器学习机制的排序和“分类”（又名“数学算法”、“模型”等 –

可根据实现细节和术语偏好采用多种不同名称) 做出决策, 对围绕人工智能/机器学习输出 (多维数据集、表格、多维数据数组等 – 生成报告、仪表板等) 的智能保持最新的分析结构。在 InterSystems IRIS 中实现 CI 具有一项重要特性: 平台和数学建模环境之间存在双向集成关系, 支持在平台内执行使用 Python、R 或 Julia 在各自的环境中编写的内容并接收执行结果。

这种集成在“终端模式”(即人工智能/机器学习内容被制定为对数学环境执行标注的 ObjectScript 代码) 和“业务流程模式”(即人工智能/机器学习内容被制定为使用可视化编辑器、Jupyter 或 IDE (IRIS Studio、Eclipse、Visual Studio Code) 的业务流程) 下均适用。需要使用 CI 层内的 IRIS 和 CD 层内的 Jupyter 之间的链接指定在 Jupyter 中编辑业务流程的可用性。

本文将进一步提供与数学建模环境相集成的更为详细的概述。我们认为此时完全有理由向您说明, 平台中提供了将人工智能/机器学习机制 (源自“持续部署”) “持续集成”到实时人工智能/机器学习解决方案中所需的所有工具。最后要聊到的是至关重要的宏机制: CT。没有它, 就不会存在人工智能/机器学习平台 (即便可以通过 CD/CI 实现“实时”)。CT 的本质是平台能够在数学建模环境的会话中直接操作机器学习和人工智能的“工件”: 模型、分布表、向量/矩阵、神经网络层等。在大多数情况下, 这种“互操作性”体现于在环境中创建上述工件 (例如, 对于模型, “创建”包括模型规范及其参数的后续评估, 即所谓的模型“训练”), 对工件进行应用 (对于模型: 借助目标变量的“建模”值进行计算 – 预测、类别指定、事件概率等), 以及对已经创建的和应用的工件进行改进

(例如, 通过根据模型的性能重新定义模型的输入变量以便提高预测准确率, 作为一种可行选项)。CT 角色的关键属性是它对 CD 和 CI 现实的“抽象”: CT

可以在具体环境中存在的限制下使用人工智能/机器学习解决方案的计算和数学细节来实现所有工件。将由 CD 和 CI 负责“提供输入数据”和“交付输出”。在 InterSystems IRIS 中实现 CT 具有一项重要特性: 使用上述与数学建模环境的集成时, 平台可以从其编排的数学环境内的会话中提取工件, 并且 (最重要的是) 将它们转换为平台内数据对象。

例如, 刚刚在 Python 会话中创建的分布表可以 (无需暂停 Python 会话) 传输到平台中作为 global (InterSystems IRIS

中的多维数据数组), 并进一步重用于在不同的人工智能/机器学习机制下的计算 (使用不同环境的语言实现, 如 R), 或者作为虚拟表。另一个示例: 与模型的“例程”功能并行 (在 Python 会话中), 它的输入数据集使用“自动机器学习”进行处理 – 自动搜索优化的输入变量和模型参数。

与“例程”训练一起, 生产模型可以实时接收“优化建议”, 以根据调整后的输入变量集、调整后的模型参数值 (不再是 Python 训练的结果, 而是作为其“替代”版本的训练结果, 例如使用 H2O

框架), 使整个人工智能/机器学习解决方案能够以自主方式处理输入数据和建模对象/流程中不可预见的漂移。

我们现在将以现有原型为例, 深入了解 InterSystems IRIS 的平台内人工智能/机器学习功能。

在下图中, 在图像的左侧部分, 我们看到了实现 Python 和 R 脚本执行的业务流程的片段。

在中央部分, 我们看到了执行这些脚本后的 Python 和 R 的相应可视化日志。

接下来是两种语言的内容示例, 在各自环境中执行。右侧为基于脚本输出的可视化。右上角的可视化使用 IRIS Analytics 开发 (数据从 Python 传输到 InterSystems IRIS

平台, 并使用平台功能呈现在仪表板中), 右下角所示为直接在 R 会话中获取并从中传输到图形文件的内容。重要说明: 所讨论的业务流程片段在此原型中负责基于从设备模拟器流程实时接收的数据进行模型训练 (设备状况分类)

, 该流程由监控分类模型性能的分类准确率监控流程触发。

文中会进一步讨论将人工智能/机器学习解决方案实现为一组交互式业务流程 (“代理”)。

图 12: 在 InterSystems IRIS 中与 Python、R 和 Julia 的交互

平台内流程 (又称“业务流程”、“分析过程”、“流水线”等, 具体取决于上下文) 可以编辑, 首先是使用平台中的可视化业务流程编辑器, 所用方式可同时创建流程图及其相应的人工智能/机器学习机制 (代码)。我们所说的“创建人工智能/机器学习机制”是指从一开始就进行混合 (在流程级别): 使用数学建模环境的语言编写的内容与使用 SQL (包括 [IntegratedML](#) 扩展程序)、InterSystems ObjectScript 以及其他支持的语言编写的内容相邻。此外, 平台内范式以嵌入片段集的形式 (如下图所示) 为“绘制”流程提供了广泛的功能, 有助于高效地构建有时相当复杂的内容, 避免视觉组合中的“随机失活” (“非可视化”方法/类/过程等)。

即, 在必要情况下 (可能在大多数项目中), 可以采用可视化的自文档格式实现整个人工智能/机器学习解决方案。请您注意下图中心部分, 该部分展示了“更高级的嵌入向量层”, 并表明除了这样的模型训练 (使用 Python 和 R 实现) 之外, 还有针对训练模型的所谓 ROC 曲线分析, 可对其训练质量进行视觉 (和计算) 评估 – 该分析使用 Julia 语言实现 (在其各自的 Julia 环境中执行)。

图 13: InterSystems IRIS 中的视觉人工智能/机器学习解决方案组合环境

如前所述, 已实现的平台内人工智能/机器学习机制的初始开发和 (在其他情况中) 调整将在平台外部的 Jupyter 编辑器中执行。在下图中, 我们可以找到一个编辑现有平台内流程的示例 (与上图中的流程相同) – 这就是其模型训练片段在 Jupyter 中的外观。Python 语言的内容支持在 Jupyter 中编辑、调试、查看内嵌计算图。更改 (如果需要) 可以立即复制到平台内流程, 包括其生产版本。

同样, 新开发的内容也可以复制到平台 (自动创建一个新的平台内流程)。

图 14 在 InterSystems IRIS 中使用 Jupyter Notebook 编辑平台内人工智能/机器学习机制

平台内流程的编辑不仅可以使用可视化或笔记本格式执行, 还可以使用“完整”的 IDE (集成开发环境) 格式执行。

IDE 为 IRIS Studio（原生 IRIS 开发工作室）、Visual Studio Code（VSCode 的 InterSystems IRIS 扩展程序）和 Eclipse（Atelier 插件）。在某些情况下，开发团队可以同时使用全部三种 IDE。下图展示了在 IRIS Studio、Visual Studio Code 和 Eclipse 中编辑完全相同流程的示例。

内容的任何部分都完全可以编辑：Python/R/Julia/SQL、ObjectScript 和业务流程元素。

图 15：在各种 IDE 中编辑 InterSystems IRIS 业务流程

在 InterSystems IRIS 中使用业务流程语言 (BPL) 组合和执行业务流程的方法值得特别提及。BPL 允许在业务流程中使用“预先配置的集成组件”（活动）– 确切地说，这让我们有权声明 IRIS 支持“持续集成”。预配置的业务流程组件（活动和活动之间的链接）对于装配人工智能/机器学习解决方案而言是极其强大的加速器。不仅用于装配：由于活动及其链接，在不同的人工智能/机器学习机制之上引入了“自主管理层”，能够根据具体情况做出实时决策。

图 16：InterSystems IRIS 平台中用于持续集成 (CI) 的预配置业务流程组件

代理系统（又名“多代理系统”）的概念在机器人化领域被广泛接受，InterSystems IRIS 平台通过其“生产/流程”结构为其提供有机支持。除了使用整体解决方案所需的功能“武装”每个流程的无限能力之外，“代理”作为平台内流程家族中的一员，能够为非常不稳定的建模现象（社会/生物系统的行为，实施部分观察的制造过程等）创建高效的解决方案。

图 17：在 InterSystems IRIS 中以业务流程代理系统形式运行的人工智能/机器学习解决方案

现在继续概述 InterSystems IRIS

平台，我们将向您展

示包含适用于所有实时场景类别的解决方案的应用领域（我们在之前的一期[在线讲座](#)中提供了对一些基于 InterSystems IRIS 的平台内人工智能/机器学习最佳做法的十分详细的发现）。

紧接上图，我们在下面提供了一幅更具说明性的代理系统图。

在该图中，完全相同的原型显示了其四个代理流程以及它们之间的交互：GENERATOR – 模拟设备传感器的生成数据；BUFFER – 管理数据处理队列；ANALYZER – 执行机器学习；MONITOR – 监控机器学习质量以及在有必要重新训练模型时发出信号。

图 18：在 InterSystems IRIS 中以业务流程代理系统形式构成的人工智能/机器学习解决方案

下图展示了不同机器人化原型（文本情感分析）在一段时间内的运行。上半部分 –

模型训练质量指标演变（质量提高），下半部分 – 模型应用质量指标和重新训练（红条）的动态。

正如所见，该解决方案显示出有效且自主的自我训练，同时继续以所需的质量水平运行（质量指标值保持在 80% 以上）。

图 19：基于 InterSystems IRIS 平台的持续（自）训练 (CT)

我们之前已经提到过“自动机器学习”，在下图中，我们将以另一个原型为例提供有关此功能的更多详细信息。

在业务流程片段图中，我们看到了在 H2O

框架中启动建模的活动，以及该建模的结果（与其他“手工”模型相比，获得的模型在 ROC

曲线方面具有明显优势，以及自动检测原始数据集中可用变量中“最具影响力的变量”）。这里的一个重要优势是“自动机器学习”可节省时间并提供专家资源：我们的平台内流程可以在半分钟内交付专家可能需要一周到一个月时间才能完成的工作（确定和证明最佳模型）。

图 20：在基于 InterSystems IRIS 平台的人工智能/机器学习解决方案中嵌入的“自动机器学习”

下图“直击高潮”，同时也很好地将有关各类实时场景的故事推向尾声：在此提醒您，尽管 InterSystems IRIS 可提供所有平台内功能，但在其编排下训练模型并非强制要求。

该平台可以接收来自在非平台编排工具中训练的模型的外部来源的所谓 PMML 规范，然后自其导入 [PMML 规范](#)

之时起继续实时应用该模型。请务必记住，尽管大多数最广泛使用的人工智能/机器学习工件都支持，但并非每个给定的人工智能/机器学习工件都可以解析为 PMML 规范。因此，InterSystems IRIS

是一种开放的平台，对其用户而言完全不存在“平台奴役”情况。

图 21：InterSystems IRIS 平台中基于 PMML 规范的模型应用

让我们聊聊 InterSystems IRIS 平台的其他优势（为了更好地说明，请参考制造过程管理），这些优势对于人工智能和机器学习的实时自动化非常重要：

- 强大的集成框架，可与任何数据源和数据使用者（SCADA、设备、MRO、ERP 等）互操作
- 内置多模型数据库管理系统，用于无限量制造过程数据的高性能混合事务和分析处理 (HTAP)
- 用于将人工智能/机器学习机制持续部署到基于 Python、R、Julia 的实时解决方案的开发环境
- 用于持续集成到人工智能/机器学习机制的实时解决方案和（自）训练的自适应业务流程

- 用于制造过程数据和人工智能/机器学习解决方案输出可视化的内置商业智能功能
- API 管理，可将人工智能/机器学习输出提供给 SCADA、数据集市/仓库、通知引擎等。

在 InterSystems IRIS 平台中实现的人工智能/机器学习解决方案可轻松适应现有的 IT 基础架构。得益于高可用性和灾难恢复配置支持，以及在虚拟环境、物理服务器、私有云和公共云、Docker 容器中的灵活部署能力，InterSystems IRIS 能够有效保障人工智能/机器学习解决方案的可靠性。也就是说，InterSystems IRIS 确实是实时人工智能/机器学习计算的多功能通用平台。我们平台的多功能特质在实际应用中得到了证明：对实现计算的复杂性实际上毫无限制；InterSystems IRIS 能够结合（实时）执行不同行业的场景；具有出色的适应性，能够提供满足用户具体需求的任何平台内功能和机制。

图 22：InterSystems IRIS — 适用于实时人工智能/机器学习计算的多功能通用平台

为了与对本文感兴趣的读者进行更加具体的对话，我们建议您继续与我们进行“实时”交流。我们将随时提供支持，制定与贵公司具体情况相吻合的实时人工智能/机器学习场景，运行基于 InterSystems IRIS 平台的协作原型设计，设计和执行路线图以在您的制造及管理过程中实现人工智能和机器学习。我们人工智能/机器学习专家团队的联系电子邮件：MLToolkit@intersystems.com。

[#Artificial Intelligence \(AI\)](#) [#IoT](#) [#云](#) [#平台](#) [#Machine Learning \(ML\)](#) [#融合分析](#) [#InterSystems IRIS](#)

源

URL:<https://cn.community.intersystems.com/post/intersystems-iris-%E2%80%93%E9%80%82%E7%94%A8%E4%BA%8E%E5%AE%9E%E6%97%B6%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0%E7%9A%84%E5%A4%9A%E5%8A%9F%E8%83%BD%E9%80%9A%E7%94%A8%E5%B9%B3%E5%8F%B0>