如何将“机器学习”引入自动化？原来只需要这三步！

发布时间：2020-10-12 作者：www.cechina.cn

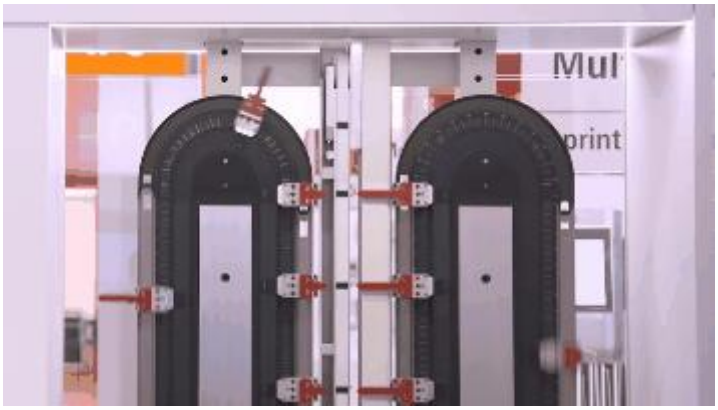
要说现在最热门的前沿技术，那非人工智能（AI）莫属。而人工智能的核心却是机器学习（ML）。可以说，掌握了机器学习，你也就掌握了人工智能技术。

那么，对于工业用户来说，如何将机器学习引入到自动化领域，突破传统自动化技术发展的天花板呢？面对人工智能、机器学习、深度学习、神经网络……这些深奥的概念，如何快速了解和掌握呢？

今天，给我 5 分钟，我告诉你答案！一文让您对机器学习的概念、关键技术、如何应用到工业自动化之中等全部都轻松掌握！

首 先

我们先来看一个，用机器学习进行优化的一个运动控制案例，以便你对机器学习有一个感性的认识。



这是两个相同的直线加圆弧的传输轨道，但我们可以看到，左边轨道上的工件输送十分平缓，而右边轨道上的工件传输很不平稳，加速很急，产品都快要被甩出去了。这不仅对轨道上的工件影响很大，而且轨道自身的磨损也很严重，右边轨道上工件运动曲线的设计显然不如左边的传输轨道。

那如何才能设计出左边这样的运动曲线呢？这里就需要用到机器学习，通过对工件多次的速度、加速度、位置等信息的记录，再经过建立数据模型，不断优化（训练）模型，最后得出一个最佳的运动曲线。为何要用机器学习来设计呢，这是因为这样的运动曲线设计并没有现存的曲线（如正圆、椭圆、渐开线等），也不能通过数学方程计算出来，所以只有借助机器学习的“算法模型加训练”来求解出来。

看完这个例子，你对机器学习的作用应该有了一个初步认识。下面我们再来理解几个常见概念。

人工智能(AI)

能够模仿人智力的智能，分为弱 AI 和强 AI，目前 AI 处于弱 AI 阶段。

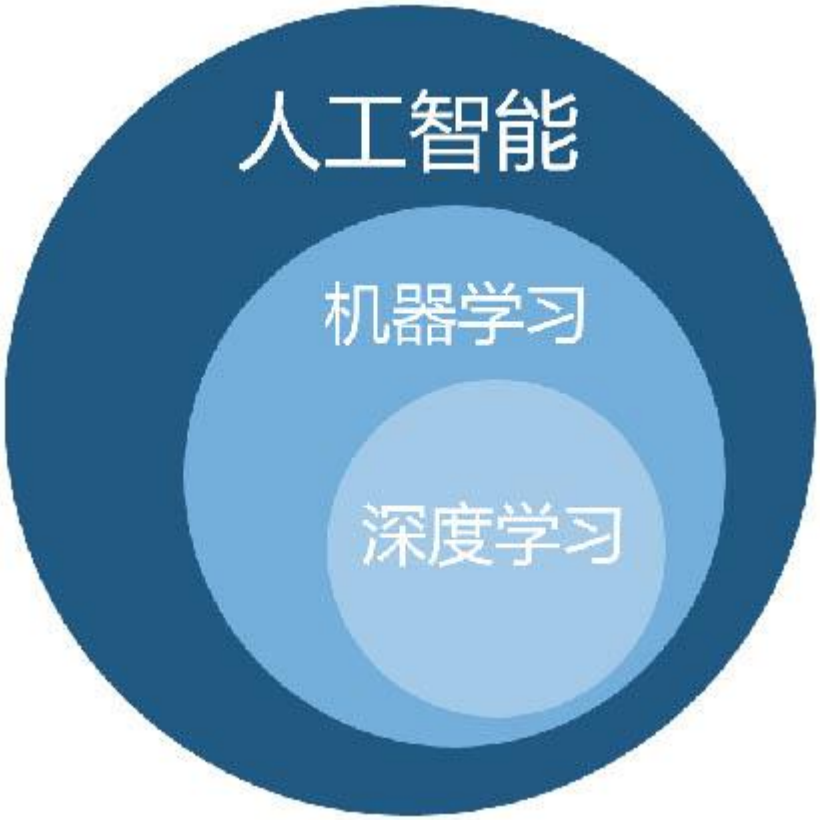
机器学习(ML)

达到弱 AI 的水平，基于可以通过“训练数据”学习特定任务的数学模型进行优化。

深度学习(DL)

专注于深度神经网络（DNN）作为模型，需要大量数据集进行训练的复杂模型，目前主要用于强大的视觉应用。

三者的关系是从属关系，如下图所示：



简单理解，机器学习就是通过根据各类算法建立数学模型，然后通过数据不断训练模型，提高模型准确性，最后将训练好的模型放到实际应用场景中运行做推理计算，解决用普通数学方法难以解决的实际问题。

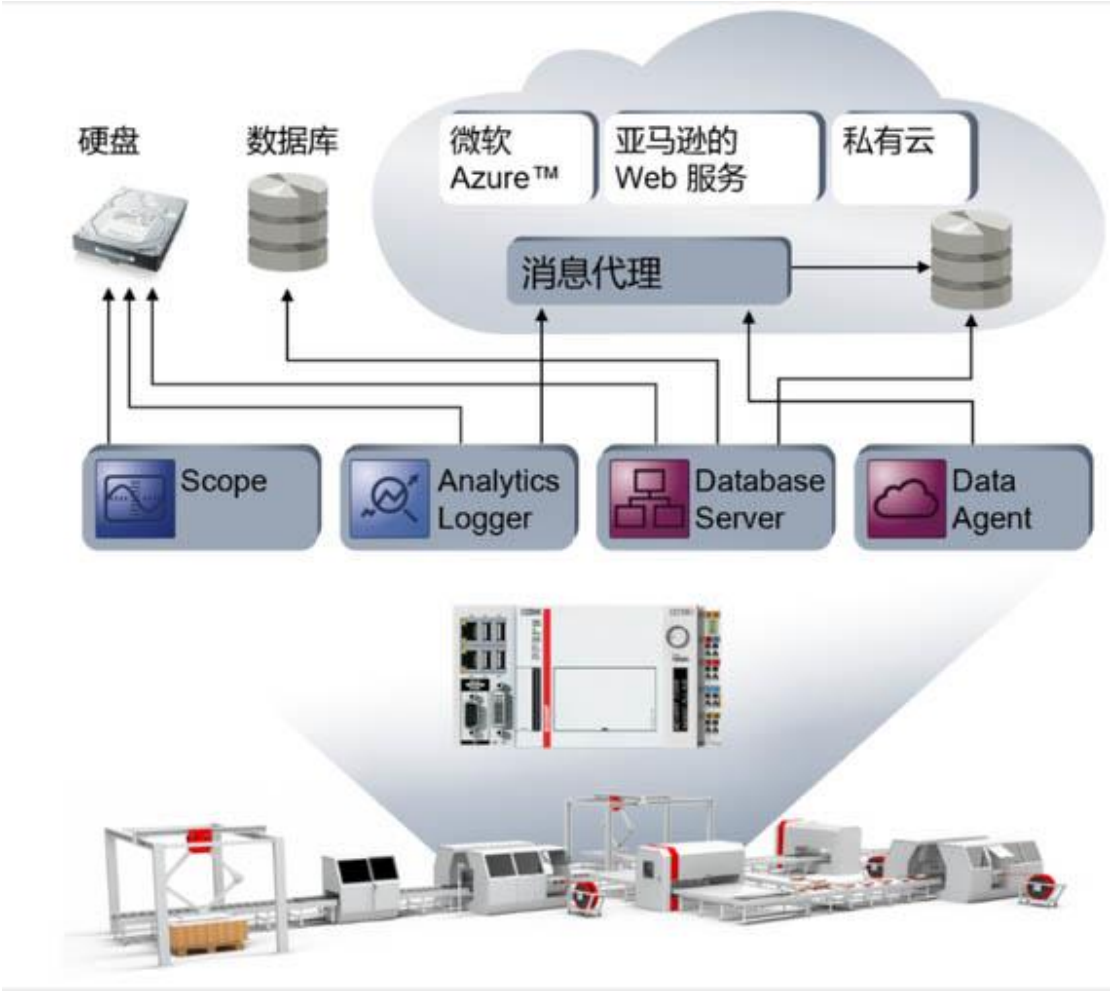
所以我们可以总结一下，将机器学习引入到工业自动化中，需要三步：收集工业现场数据、建立模型并训练模型、下载到实际应用中运行，如下图所示：

看上去是不是过程很简单？

当然了，实际使用过程并非如此简单，每个环节都会涉及到专业知识和工具，下面我们就来一一展开介绍一下，让你不仅入门，而且成为“专家”！

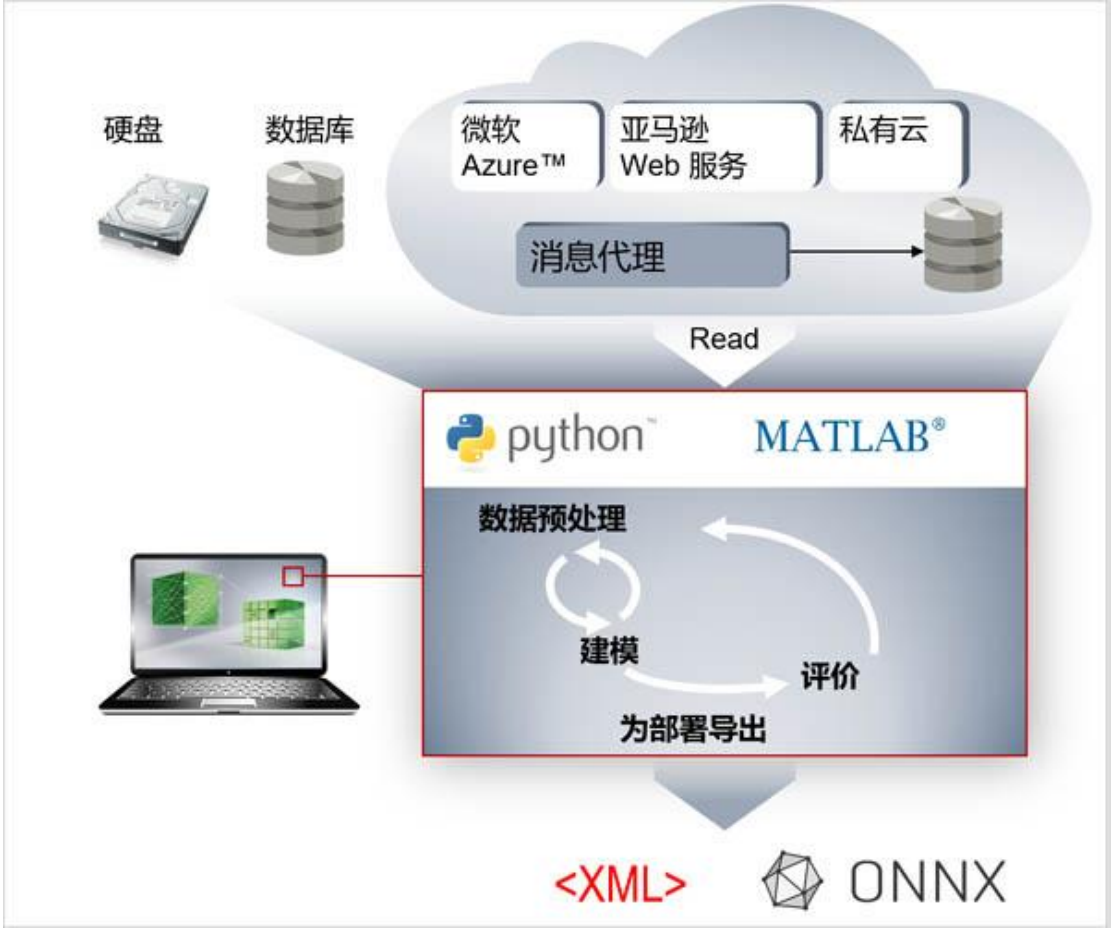
第 /一/ 步　收集工业现场数据

首先，在数据收集阶段，就是要通过各类传感器和测试测量工具来采集现场数据，这个环节就会用到我们自动化控制中的很多产品，比如像倍福的 TwinCAT3 Scope、TwinCAT3 Database Server、TwinCAT3 Data Agent 和 TwinCAT3 Analytics Logger 等工具，可以利用这些工具将数据采集到本地数据库或者云端存储、呈现，以便下一步来建模和训练。



第 /二/ 步　模型的搭建和训练

这一步是至关重要的一步，也是目前机器学习中最难、研究最多的一步。这一步里首先需要对上一步采集到的数据进行预处理，数据清洗除去异常值，数据转化或者数据集成等。然后，选择特征数据确定数学模型，进行学习微调，并进行未知数据的学习模型验证。模型训练后，生成导出一个可供 TwinCAT3 等模型运行环境的描述文件：XML 文件或者 ONNX 文件。这一步中特征数据的挖掘，也就是提取哪些数据来建模是整个机器学习能否成功的关键，往往需要精通行业知识经验的人才能做到。



在这一步中，搭建模型时往往需要用到第三方框架（平台工具），比如：Python SciKit、MATLAB Machine Learning Toolbox，以及深度学习框架 TensorFlow (谷歌)、Keras (frontend for TensorFlow, CNTK, ...)、PyTorch (脸书)、MxNet (亚马逊)、CNTK (微软)、MATLAB Deep Learning Toolbox (MathWorks)等，其中大多数是开源的和基于 Python 的。

当然，除了这些框架外，还有一个重要的事，数学模型的选择和建立。在数学上，可以把万事万物所有问题分为两大问题：回归问题和分类问题。回归问题通常是用来预测一个值，如预测房价、未来的天气情况等。分类问题是用于将事物打上一个标签，通常结果为离散值，如判断一幅图片上的动物是一只猫还是一只狗。解决这两类问题需要用到不同的数学模型，比如常见的有支持向量机（SVM）、神经网络、决策树和随机森林、线性回归、贝叶斯线性回归等，这些模型在框架中是现存的，可以直接使用。

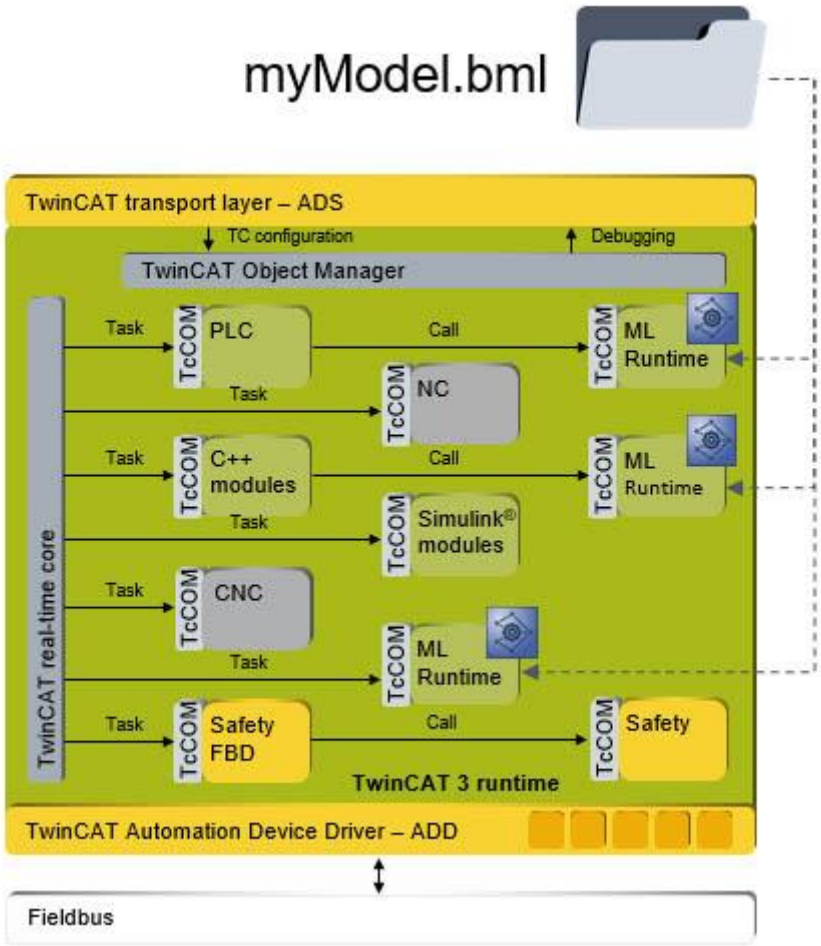
在这里，还需要提到一个知识点，那就是 ONNX 开放神经网络交换文件，这是一种针对机器学习所设计的开放式文件格式，用于存储训练好的模型。它使得不同的人工智能框架（如 Pytorch，MXNet）可以采用相同格式存储模型数据并交互。主要由微软，亚马逊 ，Facebook 和 IBM 等公司共同开发。

第 /三/ 步　加载模型到控制器里运行

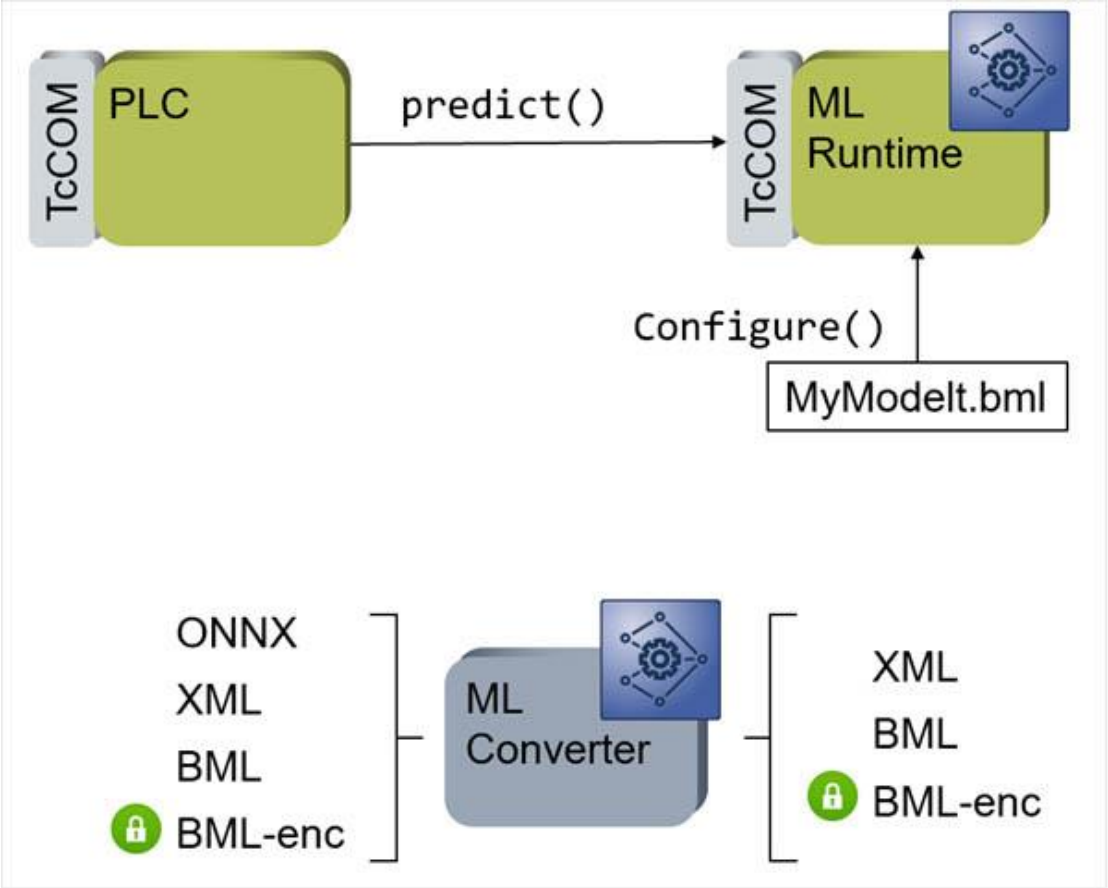
说完了模型搭建和模型训练后，最后一步就是将模型加载到工业电脑或者控制器中运行计算。由于模型描述文件并不能被工业控制器所识别，所以就需要用到像倍福 TwinCAT 3 这样的自动化控制软件平台作为引擎，将训练好的模型文件加载到控制器，才能在自动化中应用机器学习。

目前、TwinCAT3 以及无缝集成了机器学习引擎接口。可以通过机器学习文件转换器（ML Converter）把训练生成的模型文件 XML 或者 ONNX 转化

成 BML（倍福的机器学习文件）进行加密保护，经过 TwinCAT3 的 ML Runtime 进行加载，这样已训练好的模型就可以被 TwinCAT TcCOM 对象进行实时调用执行，同时可以被 PLC、C/C++ 封装的 TcCOM 的接口进行调用！如果神经网络较小，如权值大小为 10K 的多层感知器（MLP）可以在一个亚毫秒的任务周期中多次调用，以确保实时性！



同时，TwinCAT 3 本身所提供的支持多核技术也同样适用于机器学习应用，不同的任务程序可以访问同一个特定的 TwinCAT 3 推理引擎而不会相互限制。机器学习应用完全可以访问 TwinCAT 中所有可用的现场总线接口和数据，这将使其能够使用到大量数据。

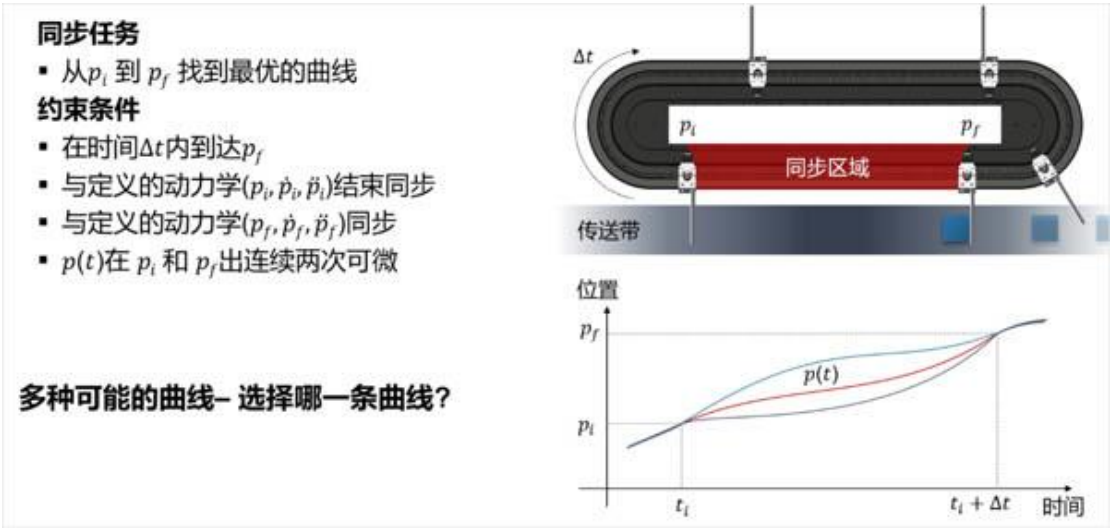


TwinCAT 3 现在有两个机器学习推理引擎，TF380x TC3 和 TF381x TC3，前者是经典机器学习模型的推理引擎，包括支持向量机(SVM)，主成分分析(PCA)，k 均值(k-means)等，后者是神经网络（NN）推理引擎，包括多层感知器（MPL）,卷积神经网络（CNN）,长短期记忆模型（LSTM）等。

举个例子　最优运动曲线是如何“机器学习”出来的

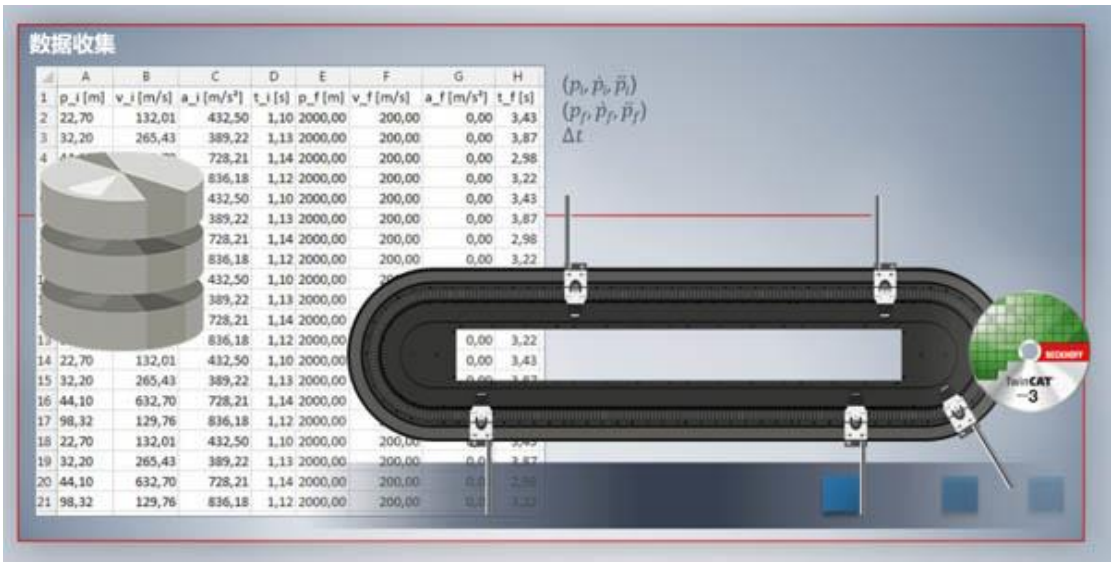
看完了上面的一般性方法介绍，下面我们再拿文章开头的那个传输轨道最优运动曲线是如何通过机器学习来进行优化的。

首先，将这个运动曲线优化的问题转化为数学问题，在一定的时间内，从 pi 顺时针到 pf 找到最优（最柔和）的运动曲线，运动过程中加速度要尽量小。



1 第一步

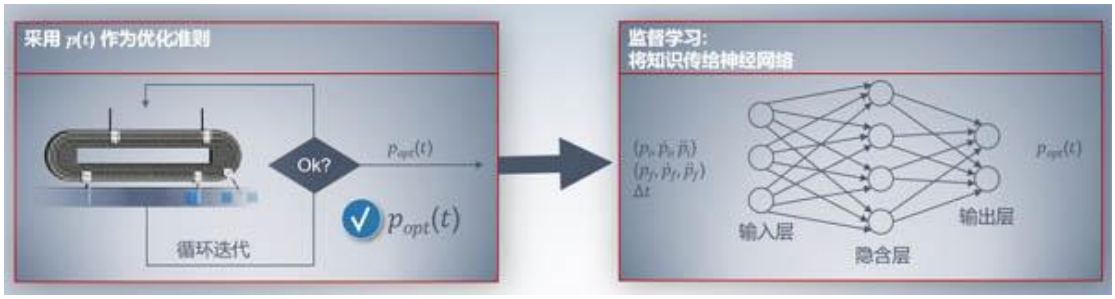
就是要采集数据，包括工件的位置、速度、加速度、时间等，把这些数据收集存储起来，以便下一步优化。



	A	B	C	D	E	F	G	H
1	p_i [m]	v_i [m/s]	a_i [m/s²]	t_i [s]	p_f [m]	v_f [m/s]	a_f [m/s²]	t_f [s]
2	22,70	132,01	432,50	1,10	2000,00	200,00	0,00	3,43
3	32,20	265,43	389,22	1,13	2000,00	200,00	0,00	3,87
4			728,21	1,14	2000,00	200,00	0,00	2,98
			836,18	1,12	2000,00	200,00	0,00	3,22
			432,50	1,10	2000,00	200,00	0,00	3,43
			389,22	1,13	2000,00	200,00	0,00	3,87
			728,21	1,14	2000,00	200,00	0,00	2,98
			836,18	1,12	2000,00	200,00	0,00	3,22
			432,50	1,10	2000,00			
			389,22	1,13	2000,00			
			728,21	1,14	2000,00			
13			836,18	1,12	2000,00			
14	22,70	132,01	432,50	1,10	2000,00		0,00	3,22
15	32,20	265,43	389,22	1,13	2000,00		0,00	3,43
16	44,10	632,70	728,21	1,14	2000,00			
17	98,32	129,76	836,18	1,12	2000,00			
18	22,70	132,01	432,50	1,10	2000,00	200,00		
19	32,20	265,43	389,22	1,13	2000,00	200,00		
20	44,10	632,70	728,21	1,14	2000,00	200,00		
21	98,32	129,76	836,18	1,12	2000,00	200,00		

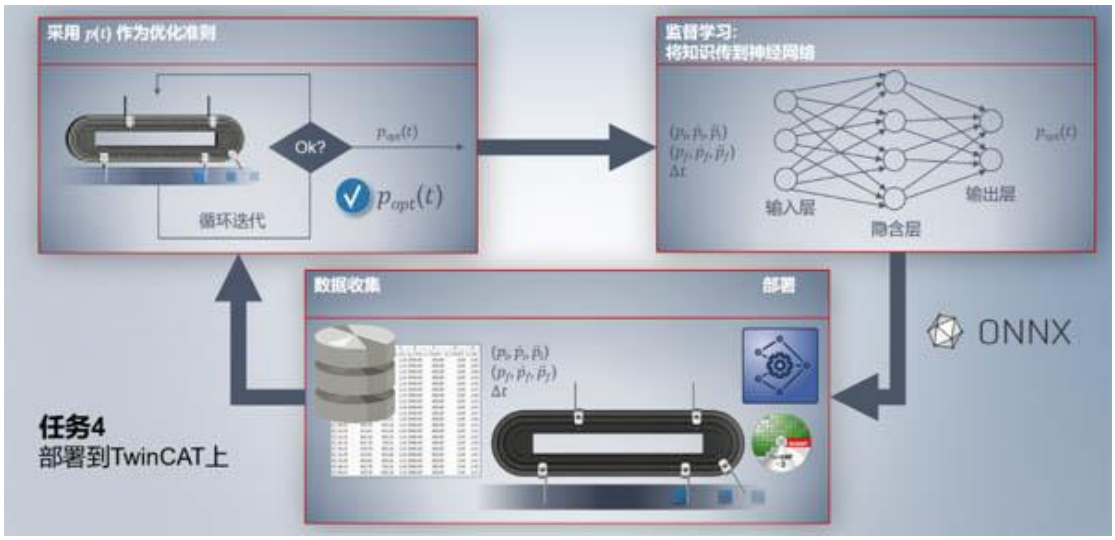
2 第二步

提取步骤一里面的特征数据建立模型，用 ??(??) 来作为优化准则，通过神经网络算法来循环迭代，监督学习训练模型。



3 第三步

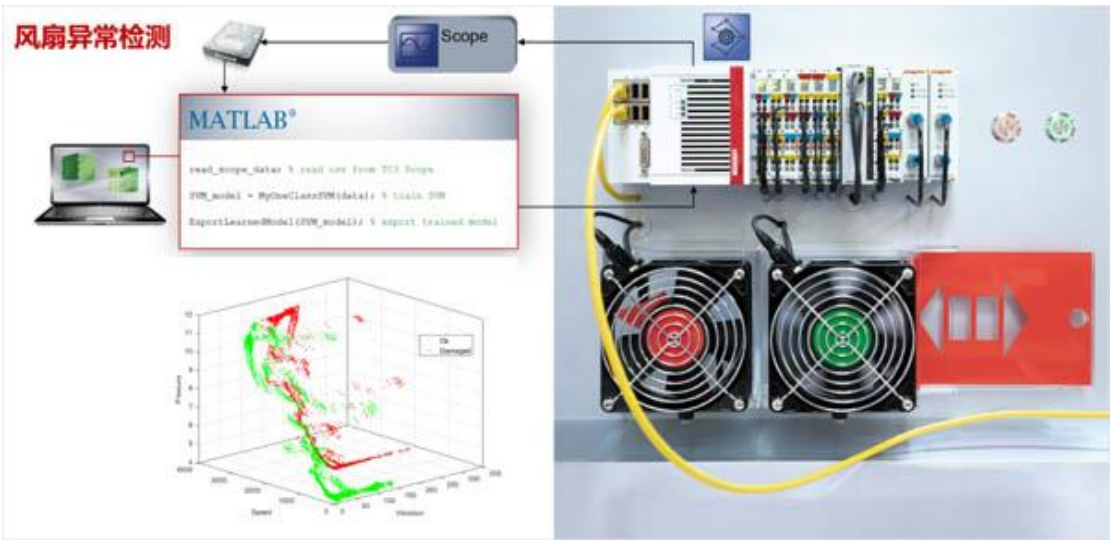
将训练好的模型通过 ONNX 文件部署到 TwinCAT 3 里，从而通过控制器实现最优的运动曲线。



举个例子　一个不需要故障数据的风扇异常检测

通常，我们通过大数据分析来做预测性维护，需要很多故障数据，可工业实际场景中往往没有太多故障数据，比如一个风机在开始几年并不太会有故障，只有到后期才有故障数据。可是，通过机器学习的方法，就可以解决这个问题，在无需故障数据的情况做到故障检测。

比如，要检测下图服务器工作站上的风扇是否有异常，就可以通过机器学习来做。



首先，通过 TwinCAT 3 Scope 来采集大量风扇正常情况下的压力、转速、振动等数据，然后用 MATLAB 来读取这些数据，使用 one-class SVM（一类支持向量机）模型来训练，等模型学习了大量正常数据后，就会自动产生一个正常数据的边界。最后将模型从 MATLAB 导出为一个 ONNX 文件，转换后加载到控制器的 TwinCAT 3 中。这样，当采集的数据超出边界时，控制器就会检测到风扇发生了异常状况。

这个应用看上去是十分简单的，但是这其中最难的一部分是特征数据的挖掘和提取，也就是常说的特征工程。至于数据的采集、模型的创建、训练，以及最后的控制器上的运行，已经有很多现存的工具和平台，比如 MATLAB 和倍福的 TwinCAT 3。站在这些“巨人”的肩膀上，你只要专注用工业现场知识和经验，就可以轻松将机器学习这一“高大上”的新技术引入到工业自动化之中。

标签：机器学习,控制器,人工智能,TwinCAT3,现场数据,工业自动化