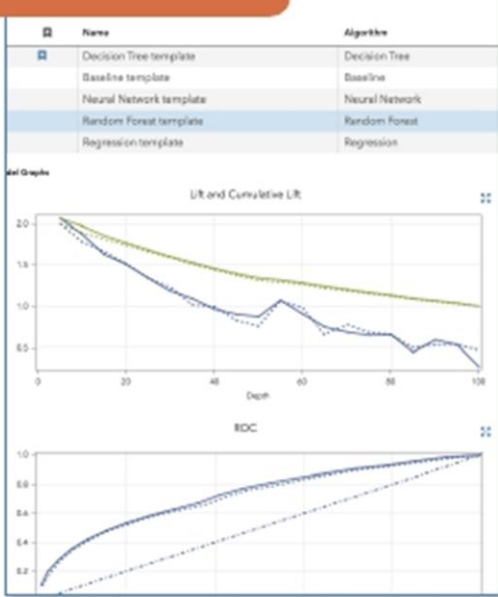


产品说明书



SAS模型工厂可以做什么？

SAS模型工厂（SAS[®] Factory Miner）提供一个基于Web自动化的可定制环境，用来在多个业务条线中大规模地建立、比较和再训练预测模型。仅仅只需几次点击操作，就可以在实战中测试您最喜爱的机器学习算法，并发现最佳模型。

SAS模型工厂为什么重要？

SAS模型工厂具有自动化、易于使用和可扩展特性，可以提升您的预测模型建模能力，提高分析人员的工作效率。他们可以快速共享模型、模板、结果和最佳实践。拥有更好更精确的模型，确保您的企业不会错过盈利机会或陷于未知的风险中。

SAS模型工厂为谁而设计？

主要为数据挖掘人员、统计人员、数据科学家、数据库营销人员、业务分析师等等而设计；同时为那些需要自动化大数据量处理，在短时间内来建立和评估众多模型的人而打造。

SAS[®]模型工厂

利用高级机器学习技术，快速方便且自动化地开发大量分析模型

随着企业开始应用分析技术来发展客户和业务，预测模型的开发也变得越来越精细。这就需要创建更多的模型，然而大多分析人员和数据科学家却没有这么多的时间分给这项工作。

SAS模型工厂提供一个基于Web的交互式环境，使得创建、修改和运行成千上万的模型变得极为简单快捷。

只需要几下点击操作，您就能够修改和转换您的数据，选择您需要使用的机器学习技术，在自动化建模环境中运行模型，快速识别每个分组中性能表现最佳的模型。如果某个分组中的模型表现不佳，您可以很方便地进行手动微调，来提高预测的准确性。

带来的好处

- 提高知识发现的效率通过对耗时的模型开发过程的自动化，获得巨大的模型生产能力。包括数据准备、变量转换、预测变量和算法选择等等。SAS模型工厂具有易于使用的Web界面，让您可以为每个客户群建立多个模型，并且自动化地识别最准确的模型。
- 自动化模型开发。选择最佳细分战略，解决您的业务问题。通过一套预置的建模模板，让您可以快速开始自定义预测建模，满足您的要求。与同事共享成功的模板，使用自动化的报告和文档化功能，在您的企业中应用模型设计和结果的最佳实践。
- 更快地探索新的想法。对大型复杂的数据应用机器学习和预测分析技术，快速获得结果。如果某个模型失败了，您可以快速使用不同的输入或思维来重新尝试。当变量改变或发现了新的变量时，您可以在不建立整个数据挖掘流程或挑战现有算法的情况下，测试这些变量。
- 快速将模型部署上线。可通过点击按钮部署冠军模型。SAS模型工厂自动生成完整的评分代码——包括所有必须的数据准备和转换步骤。而且，重训练模型也很容易，因为与模型开发和部署相关的一切工作都被集中管理起来。

产品概述

SAS模型工厂提供一个自动化的框架，用于快速便捷自动化地建立、比较和再训练成千上万的模型。

通过一个基于Web的拖拽操作界面，针对多个业务条线（例如产品、区域和渠道）或者为不同的客户群（低端或高端客户等等）建

立模型。软件使用统计和机器学习算法同时测试多个模型，基于预定义的性能指标对不同分组识别出性能最佳的模型。如果您对这种自动化的方式不甚满意，您可以对每一个模型进行手动微调。通过自动化预测模型开发流程（数据定义、数据清洗、特征选取、建模和模型评估），SAS模型工厂大大缩短了模型开发时间。

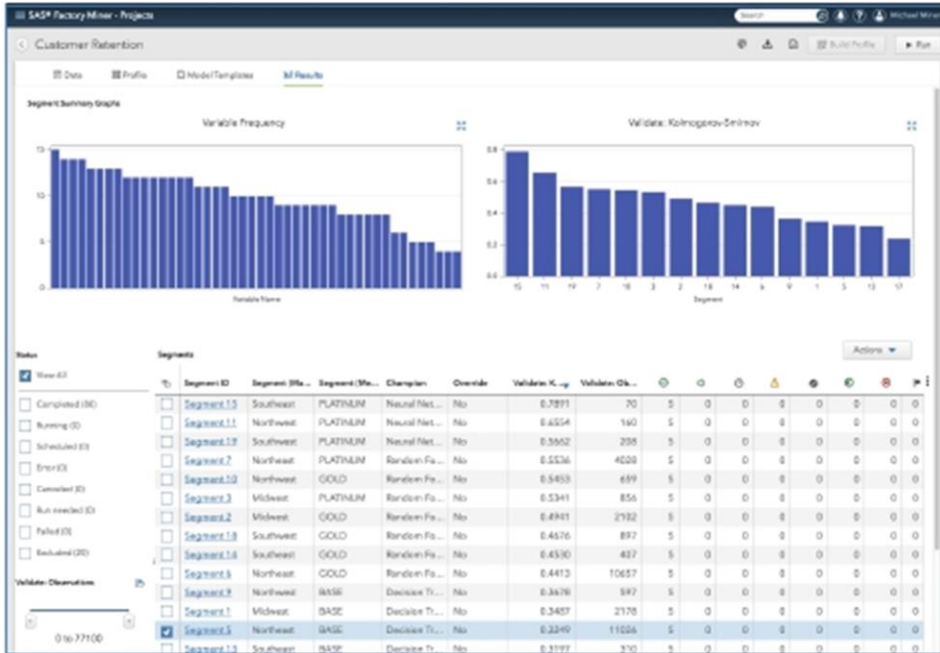


图1：客户化评估技术让您能够在数据中为每个分组生成冠军模型。

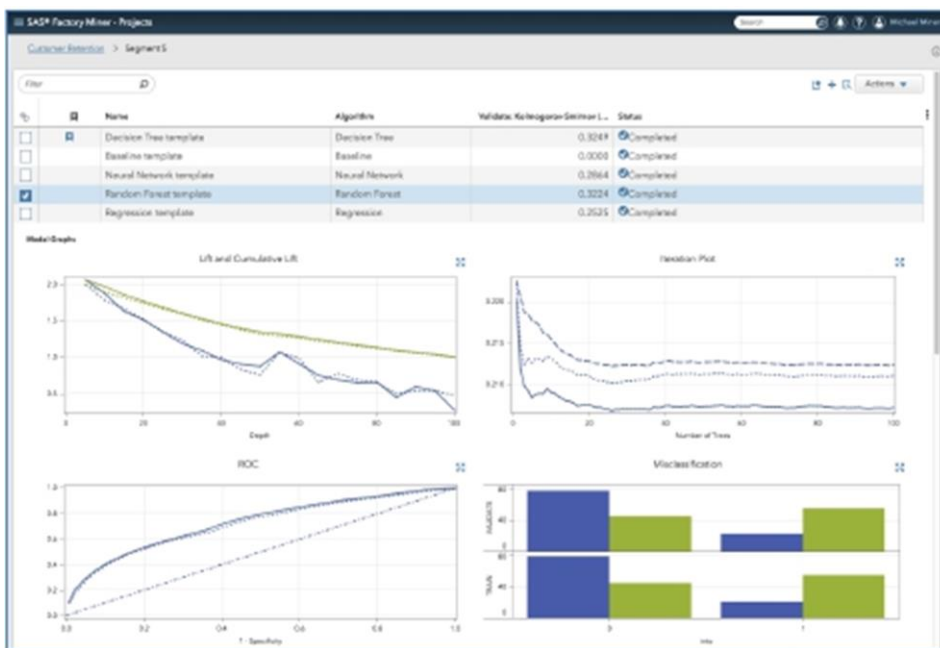


图2：您可以查看模型拟合统计量，对预测模型进行必要的改动。

经验丰富的分析师和新手都可以提高生产效率，专注于从精细模型中获得价值，得出更准确的洞察，制定更好的决策。

数据访问和数据准备

通过这个界面，用户可以轻松地访问SAS数据集，数据表和Hadoop文件。

下一步是准备数据，这通常是分析建模中最耗时的工作。使用SAS模型工厂，用户一次性指定变量的角色和水平，就可以其它人共享这些设置。这意味着每个用户都可以使用一致的基础定义，例如哪些属性是目标变量，哪些变量定义分组策略，以及哪些变量是输入特征。

自动刻画数据特征，提供对分析宽表的快速概览，能够识别有问题的数据，包括缺失值，异常值或偏斜的变量分布。交互式的数据准备工具，让数据转换变得十分简单，派生新的变量，进行智能化的特征选取，例如基于树和随机森林的变量选取。

客户自定义模型模板

易于使用的建模模板，利用尖端的机器学习技术，如开箱即用的随机森林、梯度推进、支持向量机和贝叶斯网络，以及决策树和回归模型等。所有模板都包括为模型开发的流程块和基于SAS最佳实践的缺省参数设置。

您可以在各个项目间定制、编辑和共享您的模型模板，和他人一起为企业建立一套建模流程最佳实践。向下钻取和定制模型，提高模型的性能。在模型模板中拖拽模型块，决定是否和如何应用冗余消减技术，例如变量选取或主成分、异常值过滤、补缺和变量转换等。

对每个建模项目来说，您可以从缺省模板、自己的模板或同事共享的模板中，选择想要用来参与模型竞赛的建模模板。

自助式机器学习

自助式机器学习和预测分析，让您能够在直观界面中测试新的想法。无需编写任何的程序代码或熟知机器学习技术如何工作，您就可以非常快速地创建一套基准模型。

对于性能不佳和未能达到企业标准的模型，分析专家可以轻松地深入挖掘引擎，自定义建模流程中的模型块，甚至是为每个模型块微调参数设置。可以保存微调的设置，并在自定义模板中共享。

SAS模型工厂应用客户化的评估技术，自动为每个分组选取冠军模型。您还能够推翻系统选取的模型，使用您想要的规则来手动选取冠军模型。

模型还可以跨组进行比较，发现性能之间的相对差异，以及每个模型的主要因子。

识别模型异常

为帮助提高模型性能，SAS模型工厂提供了容易理解的标准报表，可以发现模型问题和高置信度的最佳模型，甚至非技术用户也可以做到。然后您可以轻松地识别出性能差的模型，发现需要微调的建模细节。自动文档化使得在模型设计和模型结果的最佳实践，可以轻松地在企业内部相关人员之间共享。这还能够帮助您满足合规要求。

主要特点

数据访问和数据准备

- 访问注册在SAS元数据服务器中的数据源，包括SAS数据集、数据表和Hadoop文件。
- 交互式设定数据源的元数据，例如变量角色、水平和顺序，或者使用自动设定来在项目间共享变量设置。
- 定义数据分组，进行分层建模。
- 使用自动的数据特征刻画和交互式变量分布图形，评估数据问题。
- 使用下列技术过滤输入数据：缩尾，标准差，修剪和稀有值。
- 为了得到更好的模型，可对数据进行转换：**log**，**log 10**，平方根、倒数、平方、指数、中心化、标准化、区间、分箱、分位数、优化分箱和主成分分析。
- 使用统计和机器学习补缺方法来清洗数据：均值、最小值、最大值、中位数、常数、计数、基于分布。

定制有指导学习模板

- 交互式地建立自定义模板，包括模型和后续处理步骤：过滤、主成分、补缺、转换和有指导的和无指导的变量选取。
- 创建您自己的模型模板。
- 编辑数据准备或模型的参数，并保存为自定义模板。
- 跨项目和用户共享模型模板。

自助服务式机器学习技术

- 使用以下技术建立模型：
 - 贝叶斯网络
 - 决策树
 - 梯度推进
 - 神经网络
 - 随机森林
 - 支持向量机
 - 广义线性模型
 - 线性回归
 - 逻辑回归
- 交互式查看模型结果

冠军模型的识别

- 使用下列选取规则为每个分组自动选取冠军模型：
 - K-S统计量
 - 提升和累计提升
 - 增益和累积增益
 - 误分类率
 - 捕获事件的百分比
 - 平均捕获事件的百分比
 - 均方误差
- 推翻系统选取的模型，人工选取冠军模型
- 在一个和多个分组中交互式比较和评估模型

灵活的模型部署和管理

一旦模型竞赛确定了每个分组的冠军模型，就需要将冠军模型部署到生产环境中进行日常评分。这项工作可能会非常冗长乏味，比如人工重写或修改代码，延迟模型的部署，可能导致代价高昂的潜在错误。

SAS模型工厂为每个模型自动生成SAS评分代码——包括所有必须的数据准备和转换步骤。结合SAS评分加速器，SAS模型工厂的模型可以直接发布在多种数据库或Hadoop中。此外，所有模型开发和评分的过程都可以注册到SAS模型管理器中。通过集中化的Web用户界面，SAS模型管理器可以实现SAS或第三方模型的生命周期管理，管控您的模型资产。

模型重训练

预测模型也有有限的寿命。在开发阶段通过机器学习算法发现的数据关系，常常随时间的改变而改变，所以重训练模型是必需的。使用SAS模型工厂，只需要选定已经保存的模型，使用新的数据重新运行就可以了。

内置的模型评估结果历史记录，方便您轻松查看新的模型性能是否超出之前的一批。新的冠军模型随后可投产上线，代替性能不佳的现有模型。

主要特点(续)

识别模型异常

- 查看报告中的模型性能异常。
- 轻松识别性能差的模型并查看详细信息。
- 修改每个模型模板的缺省设置。

模型跟踪与报表

- 生成汇总报表，包括模型结果、重要变量和模型设置。
- 生成PDF和RTF格式的报告并共享。

模型重训练

- 在新的数据集上重新训练现有模型模板。
- 对重训练的模型生成模型评估统计量。
- 生成模型性能衰减报告。

灵活的模型管理和部署

- 为所有模型模板自动生成SAS评分代码。
- 注册模型到SAS模型管理器中，可进行集中化的模型部署和管理。（要求安装SAS模型管理器）
- 使用SAS评分加速器在数据库和Hadoop中部署模型。（要求安装SAS评分加速器）

可扩展的处理

- 在SAS服务器上使用多线程的过程步训练模型，利用多核服务器的处理能力。
- 通过SAS网络管理器使用异步处理训练模型，进行负载均衡和调度。（要求安装SAS网络管理器）
- 在数据库一体机（Oracle, Teradata, Greenplum和SAP HANA）或Hadoop上使用SAS高性能数据挖掘，在内存中训练模型。（要求安装SAS高性能数据挖掘）

可扩展的处理

SAS模型工厂可以在单机环境下运行分析模型，或利用SAS网络管理器的负载均衡和调度功能在网络环境下运行，或使用SAS高性能数据挖掘在内存中运行。也就是说您可以在最适合解决问题的环境中运行。而且，随着数据量增长、问题变得越来越复杂或您的分析能力越来越成熟，也很容易对SAS模型工厂进行性能扩展。

若要深入了解SAS模型工厂的系统需求、下载白皮书、查看屏幕截图或者浏览其它相关资料，请访问：
sas.com/factory-miner