SAS Enterprise Miner 7.1
企业级数据挖掘集成环境

简化数据挖掘流程，在海量数据上建立高精度的描述模型和预测模型。

SAS Enterprise Miner 可以做什么？

SAS Enterprise Miner 简化了数据挖掘流程，在海量数据上建立高精度的描述模型和预测模型。目前，富于前瞻性理念的企业都在使用 SAS 预测分析和数据挖掘软件，来检测欺诈行为，最大限度地降低风险，预测资源需求，减少资产的故障停机时间，提高营销的响应率，遏制客户流失。

SAS Enterprise Miner 为什么重要？

SAS 提供了业界最全面的预测分析和数据挖掘功能，帮助用户对复杂数据展开发掘和分析，获得有价值而且可信的洞察，并基于事实做出决策。

SAS Enterprise Miner 为谁而设计？

那些需要了解和分析海量数据，解决关键的业务问题或研究问题，并制定有效决策的人员，都可以使用 SAS Enterprise Miner。包括数据挖掘者，统计师，营销分析师，数据库营销人员，风险分析师，欺诈调查员，工程师，科学家和业务分析人员。

对于大多数企业来说，要把越来越多的原始数据转化为有用的信息，无疑是一个艰巨的任务。这是因为那些能够帮您识别重要商业机遇的内在联系和答案，常常被深埋在浩如烟海的数据之中。哪些客户会在什么时候购买什么产品？哪些客户正在流失，要想留住他们该做些什么？保险费率应该如何制定才可以确保盈利？维护计划和操作结果会对设备的无故障运行时间造成什么影响？

为了得到这些问题的答案，在当今竞争激烈的市场中保持领先，企业就需要功能强大的数据挖掘解决方案。从数据中发现以前未知的模式和规律，从而帮助整个企业的决策者建立有效的发展战略。那些选择了数据挖掘并将其落实到业务流程的企业，才可能在当今快速发展竞争激烈的市场中生存。

不过，对大量数据的挖掘还需要正确的流程和合适的工具，否则可能会弄巧成拙、效率低下。数据挖掘的新方法和新技术正在不断涌现，迅速发展。即便是相同的分析方法，也可能在某个数据集上表现良好，放到新的数据源或者新的业务应用上却变得无效。因此，拥有一套功能强大、用途广泛的分析工具就变得至关重要。不同的工具会产生不同的模型，只有当你逐个比较这些模型时，你才可能从中发现最适合分析问题的挖掘方法。如果你使用的是分析能力有限的工具（比如说，只支持回归或者决策树算法），得到的预测模型就可能缺乏业务价值。

各个行业对有价值分析信息的需求变得越来越强烈，由此也给数据挖掘人员和业务分析师带来了越来越大的压力，他们需要在更短的时间内开发出更多、更好的分析模型。安全可靠而且可扩展的数据分析，可以帮助企业解决很多业务问题。这需要在整个企业内分工合作，同时需要一个强大的多功能数据挖掘解决方案，该方案可以按照不同需求进行定制。

SAS Enterprise Miner 提供了一个优化的架构，使得数据挖掘人员和业务分析师有了更多的时间来建立高度准确的预测模型和描述模型。分析结果可以在整个企业内共享，并将分析模型嵌入业务流程。SAS 分析功能是 SAS 商业分析框架的重要组成部分。SAS 商业分析框架可以扩展您现有系统的价值，确保为整个企业提供清晰而一致的统一视图，并帮助您精确预测，减少运营中的不确定性，取得市场优势。

主要优点

- **了解数据中的重要联系，直观快速地开发模型。** SAS Enterprise Miner 的图形用户界面让它很容易使用，方便分析人员在整个建模过程中进行操作。分析人员和企业管理人员可以共享一个公共的、易于解释的、可视化的数据挖掘过程，彼此紧密合作，解决最严峻的业务挑战。一个互动的点击式操作环境，已经针对常见的分析需要进行优化，支持数据表和图形之间的联动分析，方便您的知识发现和可视化探索，更容易地理解数据中的联系。

- **功能全面的数据挖掘集成环境，让您可以更有效地开发出更好的模型。** SAS Enterprise Miner 提供了数据挖掘流程图环境，该环境允许您以流程图方式来设计整个数据挖掘过程，支持用户交互并自动记录参数设置，极大地缩短了商业分析师和数据挖掘人员的模型开发时间。它能有效地将您设计的数据挖掘过程映射为 SAS 的后台处理，从而产生最佳的分析结果。

- **自给自足而且自动化的分析方式，快速轻松地产生洞察，辅助您做出更好的决策。**对于缺乏深厚统计分析基础的商业分析师和业务专家来说，SAS 的快速预测建模工具非常适合他们，能够针对常见的业务场景自动生成预测模型，快速给出行动建议。分析结果中包含了易于理解的图表，帮助用户理解和决策。

- **提高了预测的准确性，确保您做出正确决策，并持续采取最佳行动。** 支持不断发展的创新算法，带来更精准高效的模型，同时结合行业分析应用经验，提高预测模型的稳定性和准确性，并可轻松通过可视化的模型评估和验证指标进行验证。可以采用多种算法建立预测模型，并将模型结果和评估统计量并排展示，方便对多个模型进行比较。挖掘流程图可用作带有自我记录功能的模板，它们可以方便地更新，或者重用以处理新的业务问题，无需从头开始进行分析。模型结果还可以帮助您深入理解自变量和目标变量之间的关系以及影响程度。

- **轻松进行模型部署和进行评分，快速获得评分结果。** SAS Enterprise Miner 简化了繁琐的模型部署和评分过程，自动生成 SAS、C、Java 和 PMML 等评分代码，支持以批处理方式对新数据进行评分。评分代码可以部署到实时评分或批处理环境中，包括 SAS 服务器、网站、嵌入到业务流程中，或直接放到关系数据库内运行（库内处理）。这些功能可以节省您的时间，帮助您获得准确的结果，并作出正确决定，为您带来最大价值。

产品概览

SAS Enterprise Miner 是一个帮助企业展开数据挖掘的先进的、分布式的客户机/服务器系统。为了强化数据挖掘过程，该软件进行了精心设计，让之可以和其它 SAS 技术无缝集成，包括数据整合、统计分析和报表等 SAS 技术。

完备的企业数据统一视图

当我们在一个集成的信息交付体系中展开数据挖掘时，数据挖掘就可以发挥出最大的威力。集成的信息交付体系会从各种不同的数据源收集信息，包括网站、呼叫中心、调查表、客户反馈意见、时间序列和来自业务系统的交易数据等等。辅助以 SAS 文本挖掘解决方案，您还可以将结构化数据和非结构化数据放在一起进行分析挖掘，从而形成一个综合的预测建模解决方案，包括全方位的数据分析和知识发现。

易于使用的图形用户界面

SAS Enterprise Miner 数据挖掘集成环境提供了一个易于
使用的鼠标拖拽操作界面，适合数据分析人员使用。它提供了先进的分析挖掘算法，并按照核心的分析任务进行组织，这些分析任务是保证数据挖掘成功的基础。在 SAS 数据挖掘方法论中包括 5 个主要步骤：抽样(Sampling)，探索(Exploration)，变换(Modifcation)，建模(Modeling)和评估(Assessment)。称为 SAS SEMMA 数据挖掘方法论。在每一个步骤中，您可以执行一系列分析动作，来完成数据挖掘项目的开发。通过使用 SEMMA 工具栏上的功能节点，您可以运用先进的统计方法，找出最显著的变量，借助表达式生成器来完成数据变换，开发模型来预测结果，验证模型的准确性，对新数据进行预测评分，并将评分结果发布到您的业务系统。

快速简便自助的模型开发方式。

SAS 快速预测建模(Rapid Predictive Modeler)，能够自动快速地生成预测模型，用来解决大量常见的业务分析问题。它会自动尝试一系列数据挖掘任务(例如，数据变换，变量选择，尝试多种常用算法，并对模型结果进行评估)，从而自动化地完成模型开发。

SAS 快速预测建模，是 SAS 企业向导(SAS Enterprise Guide)和 SAS 的 Microsoft Office 插件中的一个功能模块，在该模块中预置了一系列 SAS Enterprise Miner 的挖掘建模步骤。这是一种协作式的方法，业务人员可以首先使用 SAS 快速预测建模工具来快速开发出预测模型，然后由专业分析建模人员使用 SAS Enterprise Miner，来对模型进一步精炼和定制。

经典先进的分析建模技术集合

SAS Enterprise Miner 所提供的分析建模方法，在深度和广度上是其它软件无与伦比的。它包含了大量经典的和先进的预测性和描述性建模算法，包括决策树，袋装与推进，时间序列挖掘，支持向量机，神经网络，内存推理，层次聚类，线性回归和逻辑回归，关联分析，序列分析和 Web 路径分析，以及更多数据挖掘算法。SAS Enterprise Miner 同样也支持那些最先进的分析算法，包括梯度推进，LARS 算法和偏最小二乘回归方法。

全面的数据准备、汇总和探索工具

对于数据挖掘来说，数据准备是费时费力但又非常必要的一项工作。SAS Enterprise Miner 提供了功能强大的交互式数据准备工具，包括填补缺失值，过滤异常值，以及定义数据分组规则。核心的数据准备工具包括文件导入，数据追加，合并和变量的删除。广泛的汇总描述功能和交互式探索工具，让新用户也可以通过动态链接的多维图形，很方便地检视大量的数据。这将有助于产生高质量的、针对特定业务问题量身定制的数据挖掘结果。

基于业务的模型比较、报告和管理

模型比较的评估功能包括 Lift 曲线和投资回报率等指标，这些评估指标让数据挖掘人员可以更容易与业务领域专家讨论模型的结果，并展开协作。借助于高度可视化的评估界面，不同算法生成的模型可以进行一致的评估。创新的截止(Cutoff)节点检查后验概率分布，定义出解决现有业务问题的最佳行动。

传统评分和库内评分过程更快地生成结果

评分是一个定期将模型应用到新数据上的过程，从而完成模型在业务环境中的部署。这个过程可能是很乏味的，尤其是当它需要手动改写或转换代码的时候。这种手动工作会导致模型部署进展缓慢，同时也很容易导致一些潜在的、需要花费大量精力才可以修正的错误。评分代码必须完成产生预测模型之前的所有处理任务，包括每一个数据预处理步骤。SAS Enterprise Miner 能够自动生成 SAS，C，Java 和 PMML 的评分代码。评分代码可以部署到实时评分或批处理环境中，包括 SAS 服务器、网站、嵌入到业务流程中，或直接放到关系数据库内运行(库内处理)。借助于 SAS 的评分加速器(已经支持 Teradata，IBM DB2 和 Netezza 数据库)，SAS Enterprise Miner 的模型评分功能可以发布为数据库
专用的评分函数，直接在数据库中进行评分。SAS Enterprise Miner 的分析结果可以传递给其他的 SAS 解决方案，比如 SAS 营销自动化(Marketing Automation)，SAS 模型管理(Model Manager)和 SAS 实时决策管理(Real-Time Decision Manager)等，从而将数据挖掘成果部署到实时的运营环境中。

开放而且可扩展的设计提供了更多的灵活性

SAS Enterprise Miner 是一个可定制的软件环境，用户可以添加新的工具，导入个性化的 SAS 代码。在 SAS Enterprise Miner 集成环境之外开发的 SAS 模型，可以方便地集成到 Enterprise Miner 的分析过程流中，并且能够完全控制其中的每一条 SAS 代码。扩展节点中提供了交互式编辑器，方便您开发和调试模型训练代码和评分代码。用户可以交互式地编辑和提交代码，同时查看运行日志和输出结果。您还可以使用 SAS 代码或 XML 逻辑来开发自定义的功能，从而实现功能扩展，这项特色也向数据挖掘人员敞开了 SAS 的大门，您可以尽情利用 SAS 全部的分析能力。

支持网格计算的高性能平台

创新的 Java 客户端和 SAS 服务器架构提供了前所未有的灵活性，您可以根据需要对 SAS 来进行配置安装，从单机系统到超大型企业应用，您都可以得到最有效的系统部署。无论用户是下班回家抑或去了其它地方，都无需担心您提交到服务器上的挖掘任务会终止，功能强大的服务器会在后台完成这些计算过程。大量的处理密集型任务，例如数据排序、汇总、变量选择和回归建模，都可以实现高性能计算。而这种高性能计算是以多线程，并行处理、负载均衡、网格计算、批处理调度等技术来予以支撑。

先进的分布式数据挖掘系统，适合企业级应用

SAS Enterprise Miner 可以通过门户网站以瘦客户端的方式进行分发部署，从而将客户端的维护工作量减至最小。当然，您也可以将整个数据挖掘系统配置安装在单独的一台电脑上(即单机版)。SAS Enterprise Miner 支持多种主流操作系统，从 Windows 服务器到 UNIX 平台，从而成为企业级大规模数据挖掘项目的首选软件。借助于 SAS Enterprise Miner 的批处理功能，您可以合理调度大量的模型训练和评分过程，把它们安排在服务器负载的低峰时段来执行。在整个分析过程中，您都可以轻松创建相关报表，用于内部记录或者是作为外部报告。模型的结果可以打包，并注册到 SAS 元数据服务器。这样数据挖掘模型就可以与其它 SAS 软件进行整合，包括 SAS 模型管理、SAS 数据整合工作室(SAS 企业数据整合服务器和 SAS 数据整合服务器的一个组成部分)、SAS 企业向导等。借助于 SAS 商业分析平台，数据挖掘人员、业务经理和数据管理人员可以有效地管理企业内的大量挖掘模型。

主要特点

直观的界面

- 易于使用的图形用户界面，可以创建分析流程图:
  - 更快地建立更多更好的模型。
  - 支持浏览器/服务器架构。
  - 可以访问 SAS 编程环境。
  - 可以通过 XML 格式来交换分析流程图。
  - 其它项目或用户可以把流程图作为模板来重用。
- 批处理:
  - 封装了图形用户界面的所有功能。
  - 基于 SAS 宏来进行处理。
  - 可以将模型训练和评分过程嵌入客户化的应用程序。

可扩展的处理

- 基于服务器的处理。
- 内存处理。
- 异步模型训练。
- 以干脆的方式停止处理过程，不留系统垃圾。
- 网格计算:
  - 通过服务器集群，实现挖掘流程的分布式处理。
  - 调度执行模型训练和评分任务。
  - 负载均衡和资源分配。
- 并行处理 - 同时运行多个功能节点和分析流程。
- 多线程预测算法。
- 在服务器上存储所有数据。

数据的访问和管理

- 支持对结构化数据和非结构化数据的访问与整合，包括时间序列数据，交易数据，网站访问路径数据和调研数据作为候选人预测，并可将这些数据作为备选的预测自变量。
- 文件导入节点让您可以很方便地导入 Microsoft Excel 文件、CSV 文件、SAS 数据文件，JMP 数据文件和其他常见的文件格式。

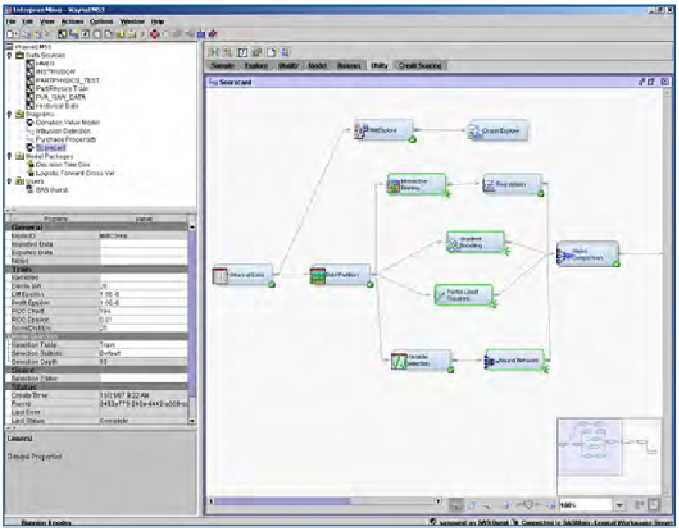

图 1: 在 SAS Enterprise Miner 的图形用户界面中，分析流程图是一种自我记录的模板，可以很容易地更新，应用到新的问题上，与其他建模人员或分析师共享。

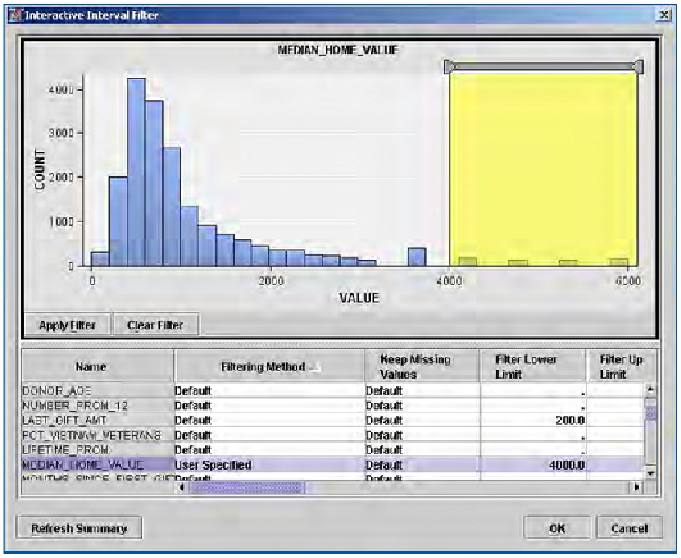

图 2: 使用“过滤器”节点，交互式过滤极端值。阴影区域定义了要保留的变量取值范围。

抽样

- 简单随机抽样。
- 分层抽样。
- 按权重抽样。
- 聚类抽样。
- 系统抽样。
- 对前 N 条记录抽样。
- 稀有事件抽样。
- 在 Teradata 13 中的分层事件水平抽样。

- 可以和 R 整合，扩展更多的建模算法并进行比较:
  - 在 JMP Pro 中可以直接调用 R。
- 允许变量中包含特殊字符。
- 增强的资源管理器窗口，快速查找并浏览表中数据，或使用交互式图形组件来进行图形化分析。
- SAS 逻辑库管理器和逻辑库分配向导。
- 变量删除节点。
- 合并数据节点。
- 添加数据节点。
- 异常值过滤:
  - 通过设置各种分布的阈值来除去极端的数值。
  - 把出现次数少于 n 次的类别值合并起来。
  - 交互式过滤类别值和数值。
- 元数据节点，用于修改变量的元数据，例如变量角色、测度水平和顺序。
- 通过 SAS 元数据服务器，与 SAS 数据整合工作室、SAS 企业向导、SAS Office 插件等软件进行整合:
  - 建立挖掘所需的训练表。
  - 部署评分代码。

数据分区

- 创建训练、验证和测试数据集。
- 通过使用数据分区，保证模型良好的泛化能力。
- 类目标的默认分层。
- 任意类变量的均匀分区。
- 输出 SAS 表或视图。

数据替换

- 集中性测量。
- 基于分布。
- 通过替代切分规则来进行树补缺。
- 两端边界值(滤除两端异常值后)的平均。
- 稳健的 M-估计。
- 缺省常数。
- 数据替换编辑器:
  - 为类变量指定新取值。

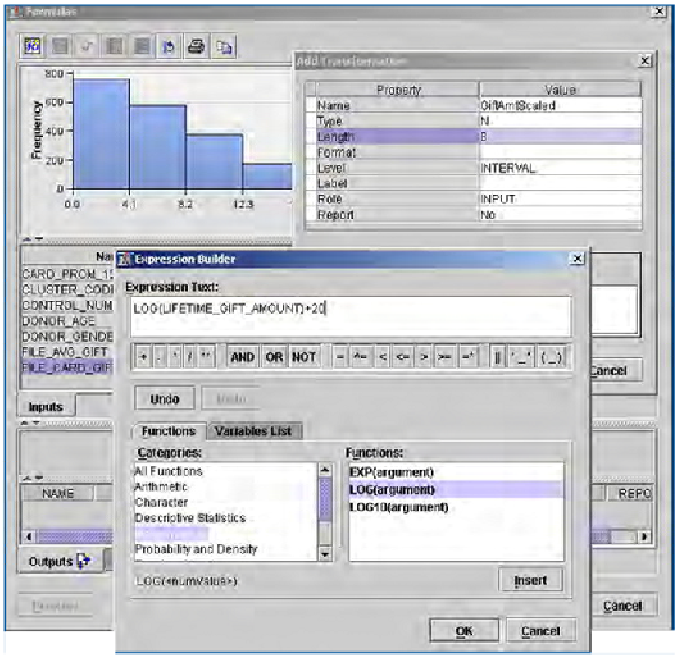

图 3: 交互式使用变量变换节点中的表达式生成器，自定义数据变换。

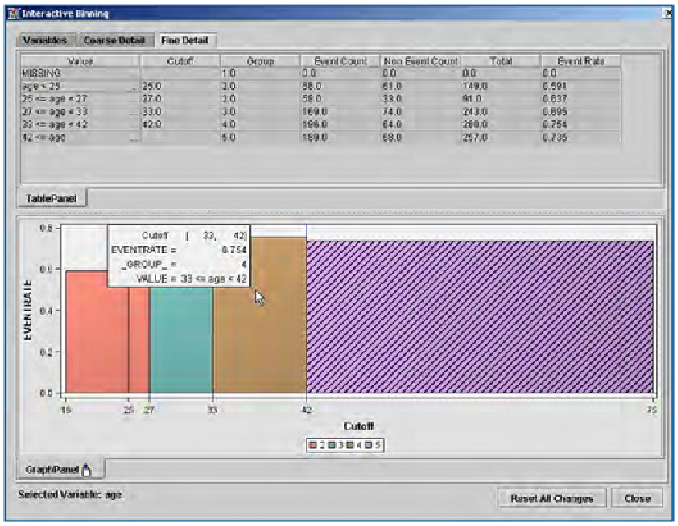

图 4: 交互式的变量分箱，以保证最大化变量与目标的关系，或者符合业务策略要求。数据分箱的交互式拆分与合并，并保存分组定义以备重用。

交互式变量分箱

- 分位数或分桶。
- Gini 变量选取。
- 将缺失值处理为单独分组。
- 精细分类和粗分类。
- 按目标对分箱进行特征刻画。
- 交互式修改分组。
- 保存分箱定义。

- 为未知值指定替换值。
- 交互式设定替换阈值，捕获区间变量的异常值。

描述性统计

- 单变量统计和绘图：
  - 区间变量：N，平均值，中位数，最小值，最大值，标准差，按比例调整的偏差和缺失百分比。
  - 类变量：类的数量，计数，众数、众数百分比、缺失百分比。
  - 分布图。
  - 类目标变量各个水平值的明细统计。
- 双变量统计和绘图：
  - 皮尔森相关和斯皮尔曼相关图。
  - 卡方分布图，支持对连续输入变量的 n 分箱。
  - 变异系数绘图。
- 基于 logworth 的变量选择。
- 其它交互图形：
  - 变量价值图 – 基于自变量对目标变量的价值进行排序。
  - 按目标变量或聚类变量的取值，来统计类变量的分布。
- 按比例调整的均值偏差图。

聚类和自组织映射图

- 聚类：
  - 用户自定义或自动选择最佳的聚类数目。
  - 类变量的多种重编码策略，以便纳入后续的聚类分析。
  - 缺失值处理。
  - 段剖面图中显示了各个聚类中输入变量和其它因子(未纳入聚类算法的变量)的数据分布，便于刻画聚类特征。
  - 段剖面是通过决策树算法来生成的，反映了输入变量和聚类之间的关系。
  - PMML 评分代码。
- 自组织映射图：
  - 支持批量处理的 SOM，支持

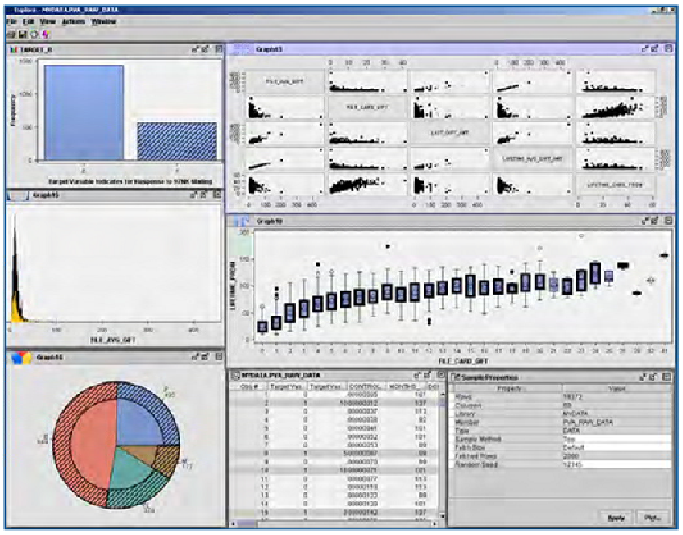

图 5：通过各种统计图形，对您的数据展开探索。在图形探索节点中，同时自动保存探索结果和交互式图形。

图形/可视化

- 批量和交互式绘图：散点图，矩阵，箱形图，星座，等高线图，针状图，网格，密度和多维图；三维图，饼图和带状图和直方图。
- 段剖面图形：
  - 对聚类节点和其它建模节点产生的聚类，交互式进行聚类的特征刻画(段剖面)。
  - 轻松识别反映聚类特征及类间差异的重要变量。
- 易于使用的图形探索向导和图形探索节点：
  - 创建标题和脚注。
  - 应用 WHERE 子句进行筛选。
  - 多种可选的配色方案。
  - 坐标轴值域可轻松缩放。
  - 在 SAS Enterprise Miner 结果窗口中探索底层数据，开发客户化的图形。
- 图形和数据表之间的交互式联动，支持选择、画笔、放大、平移等操作。
- 数据和图形都可以很容易地复制和粘贴到其他应用程序，或保存为 BMP 文件。
- 交互式图形可以自动保存在节点的结果窗口中。
- 图形生成器：
  - 可以将变量拖拽到模板中，交互式创建分析图形。当你将变量拖动到图形中时，可以获取即时反馈。
  - 在一个图形中，对变量进行可视化的分组比较。
  - 统计结果与图形的动态链接，让您轻松发现数据间的关系。

数据变换

- 简单变换：自然对数，常用对数，平方根，倒数，平方，指数和标准化。
- 分箱：桶处理，分位数和，基于目标关系的最优分箱。
- 最优指数变换：最大化正态分布，目标关联最大化，按目标变量水平均匀散布。

Nadaraya-Watson 或局部线性平滑方法。

- Kohonen 网络。
- 在图上覆盖其它变量的分布。
- 缺失值处理。

购物篮分析

- 关联和序列发现:
  - 根据置信度排列的规则网格图。
  - 对规则的提升、置信度、期望置信度和支持度，绘制统计线图。
  - 针对给定的支持度和置信度范围，绘制频次计数的直方图。
  - 期望置信度与置信度的对比散点图。
  - 规则描述表。
  - 规则的网络图。
- 基于提升、置信度、支持度、规则中的项数等条件，对规则进行筛选。
- 无缝整合规则和其它输入变量，强化预测模型的效果。
- 层次关联:
  - 可以生成多层次的关联规则。
  - 指定多维输入表中的父子关系。

降维

- 变量选取:
  - 基于卡方或 R 方选取准则，删除那些与目标无关的变量。
  - 删除层次结构中的变量。
  - 删除缺失值太多的变量。
  - 对水平数过多的类变量进行缩减。
  - 对连续型输入变量进行分箱，从而发现非线性关系。
  - 检测交叉效应。
- 基于最小角回归（LARS）进行变量选择:
  - AIC，SBC，Mallows C(P)，交叉验证和其他筛选准则。
  - 包括以下图形：参数估计，系

- 交互式编辑器：定义多项式和 n 阶交互效应。
- 交互式定义数据变换:
  - 使用表达式生成器或 SAS 代码编辑器，自定义数据变换。
  - 对比新变量与原始变量的数据分布。

预定义全局转换代码以备重用。

规则生成器节点

- 即席创建数据驱动的规则和策略。
- 交互式定义结果变量的取值和导出路径。

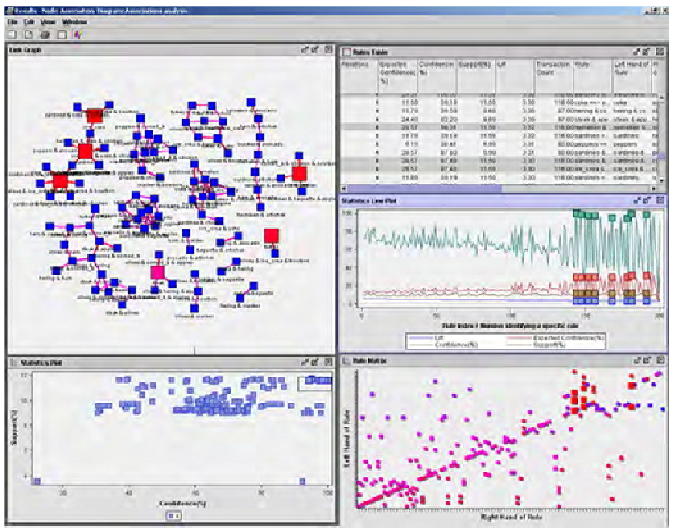

图 6: 查看购物篮特征。基于提升、置信度、支持度、规则中的项数等条件，对规则进行筛选。

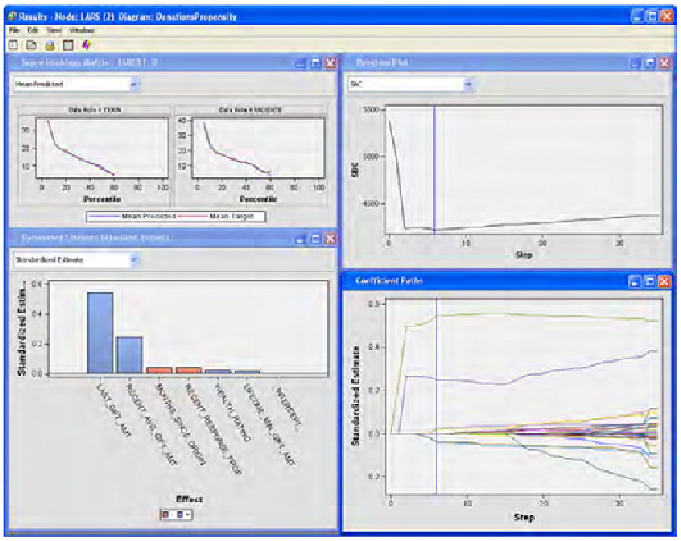

图 7: 使用 LARS 来预筛选重要的预测变量。

网站路径分析

- 在点击流数据中，可扩展高效率地挖掘最常见的导航路径。
- 在任何类型的序列数据中，挖掘出常见的连续序列。

数路径，迭代图，得分排名等等。

- 推广支持 LASSO(least absolute shrinkage and selection operator)。
- 支持类输入变量和目标变量，以及连续变量。
- 可生成评分代码。

- 主成分：
  - 计算相关性和协方差矩阵的特征值和特征向量。
  - 包括以下图形：主成分系数，主成分矩阵，特征值，对数特征值和累积比例特征值。
  - 交互式选择要保留的成分个数。
  - 使用预测建模技术来对选中的的主成分进一步挖掘。
- 变量聚类：
  - 将多个变量划分为不相交或分层的聚类。
  - 特征值或主成分分析。
  - 支持类变量。
  - 聚类的树状图展示。
  - 选定的变量表，含聚类和相关统计。
  - 聚类网络和 R 方图。
  - 交互式用户控制选定的变量。
- 时间序列挖掘：
  - 使用累计方法和变换，缩减时间序列中包含的交易数据。
  - 多种分析方法，包括周期性，趋势，时域，周期性分解。
  - 使用聚类和预测建模技术，来对缩减后的时间序列进一步挖掘。

决策树

- 支持以下决策树算法：
  - CHAID，分类回归树，袋装和推进，梯度推进和 bootstrap 森林。
  - 基于利润或提升来进行树的选择和修剪。
  - K 折交叉验证。
- 拆分规则：Prob 卡方检验，Prob F

SAS 代码节点

- 编写 SAS 代码，支持从简单到复杂的数据准备和数据变换任务。
- 可嵌入其它 SAS 模块中支持的过程步。
- 开发自定义模型。
- 创建 SAS Enterprise Miner 的扩展节点。
- 对评分代码予以增强。
- 易于使用的编程开发界面：
  - 宏变量列表，让您方便引用数据源和系统变量。
  - 交互式代码编辑器，并可提交执行。
  - 训练、评分和报告代码的分别管理。
  - SAS 代码的输出结果和运行日志。
- 创建图形。

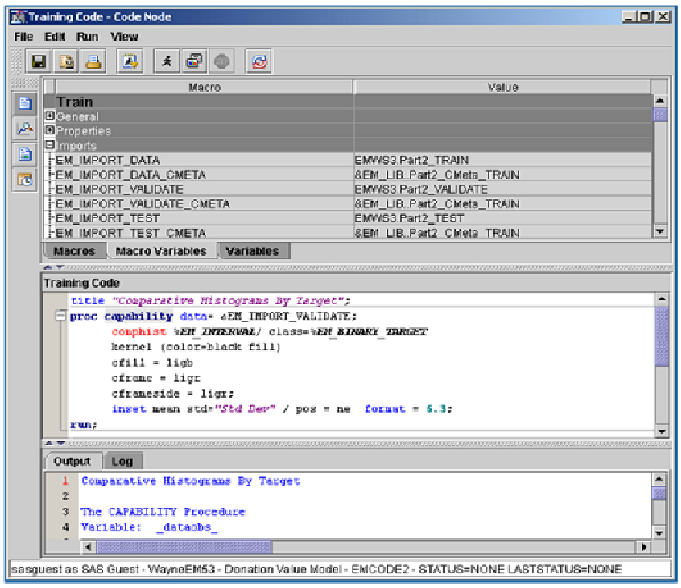

图 8：自定义的 SAS 代码，可用来创建变量变换、调用其它 SAS 过程步、开发新的功能节点、增强评分代码、生成复杂报告等等。

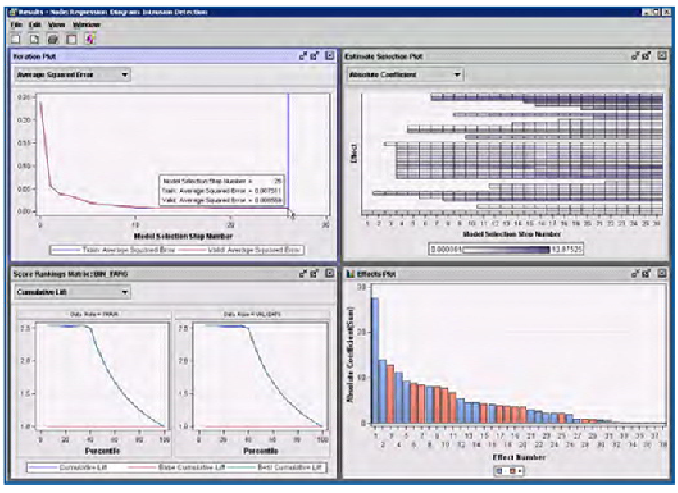

图 9：使用逐步选择方法以及其它模型选择诊断方法，来开发线性回归和逻辑回归模型。

检验，Gini，熵和方差缩减。

- 目标切换用于多目标分段策略。
- 自动输出叶节点 ID，用作后续建模和分组处理的输入变量。
- 显示英文规则。
- 计算变量的重要性，用于变量的预备筛选和模型解释。
- 独特的决策树统一树图表示。
- 交互式生成决策树功能:
  - 交互式地生长/修剪决策树；树节点的扩展与折叠。
  - 引入验证数据来评估树的稳定性。
  - 自定义拆分点，支持二叉或多叉拆分。
  - 可对任意候选变量进行拆分。
  - 拷贝拆分。
  - 将数据表和图形动态链接，让用户可以更好地评估树的性能。
  - 易于打印的树图，可打印在单页或多页上。
- 交互式的子树选择。
- 基于 ARBORETUM 快速算法过程。
- PMML 评分代码。

神经网络

- 神经网络节点:
  - 灵活的网络结构，配有丰富的组合和激活函数。
  - 10 种模型训练技术。
  - 预优化。
  - 对输入变量自动标准化。
  - 支持直接连接。
- 自动神经网络节点:
  - 自动搜寻多层感知器的最优配置。
  - 从四种不同的网络结构中选择合适的类型和激活函数。
- PMML 评分代码。
- DM 神经网络节点:
  - 在建模中支持维度归约和函数选择功能。
  - 快速训练模型；线性和非线性估计。

图 10: 交互式或按批处理方式开发决策树。提供大量评估图形来帮助用户全面衡量决策树的稳定性。

一致的建模功能

- 基于训练、验证(缺省)或测试数据集，使用利润或损失、AIC、SBC、均方误差、误分类率、ROC、Gini 或 KS 统计量等选择条件，来选择最优模型。
- 在模型开发过程中纳入先验概率。
- 支持二值型、名义型、序数型和区间型的自变量和目标变量。
- 轻松访问评分代码和分区的数据源。
- 在一个窗口中显示多个结果，帮助您更好地评估模型的性能。
- 在决策节点中可以设定目标事件，并定义先验概率和利润/损失矩阵。

回归

- 线性回归和逻辑回归。
- 支持逐步、向前和向后等选择方法。
- 公式编辑器：支持多项式、交叉效应和效应层次。
- 交叉验证。
- 效应层次规则。
- 提供以下优化技术：共轭梯度、双折线、Newton-Raphson 线性搜索或岭搜索，Quasi-Newton 和信赖域。
- Dmine 回归节点。
  - 快速向前最小二乘回归。
  - 可选的变量分箱，用于发现非线性关系。
  - 可选的类变量缩减。
- 支持交叉效应。
- 支持 Teradata v13 数据库中的库内建模。
- PMML 评分代码。

规则归纳

- 递归式的预测建模技术。
- 特别适用于对稀有事件进行建模。

基于内存的推理

- 采用 k 近邻算法，对观测进行分类或预测。
- 获得专利的降维树结构和扫描技术。

支持向量机（实验性的）

- 它是一种最大间隔分类器，适用于解决包含众多变量的问题。
- 支持二值目标变量。
- 估计方法包括完全密度二次式、分解的二次式、拉格朗日函数和最小二乘法。
- 线性、多项式、径向基函数和 sigmoid 核参数。
- 可选的抽样及交叉验证方法。
- 基于过程步的 SAS 评分功能。

时间序列挖掘（实验性的）

- 时间序列的数据准备:
  - 对事务型数据进行聚集、变换和汇总。
  - 自动对时间序列进行转置，以支持相似度分析、聚类和预测建模。
- 相似度分析:
  - 适用于新产品销量预报、模式识别和短期预报。
  - 可以计算目标序列和输入序列之间的相似度，也可以计算输入序列中的相似度。
  - 用于各种序列组合的相似度矩阵。
  - 在层次聚类中，使用相似度矩阵，并生成系统树图。
  - 用于评估聚类结果的星座图。
- 指数平滑:
  - 使用一个或多个平滑参数来控制权重衰减。
  - 自动选择最佳拟合平滑方法。
- 使用描述建模技术，来对缩减后的时间序列进一步挖掘。

生存分析

- 对离散事件进行回归，附加逻辑回归。
- 使用三次样条插值，计算时间效应上的事件概率。
- 用户可以自定义时间间隔，决定该如何收集和分析数据。
- 通过可选的抽样方法来自动扩展数据。

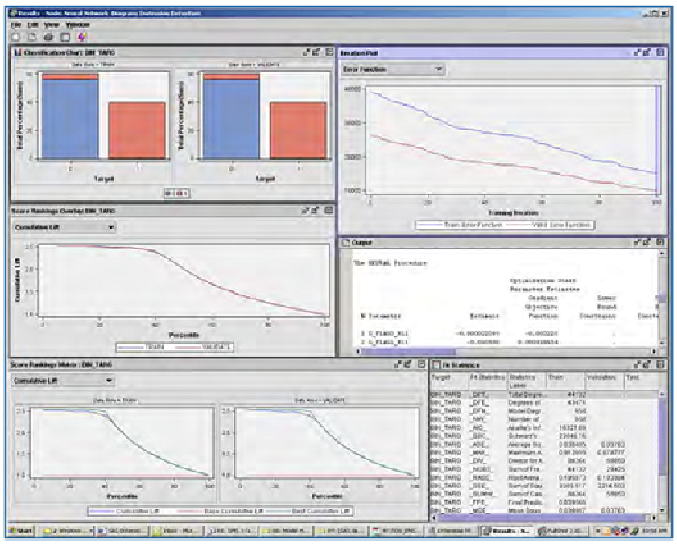

图 11: 通过使用神经网络节点，可以拟合高度复杂的非线性关系。在架构中包括了广义线性模型、多层感知器、径向基函数，以及大量组合、激活和误差函数。

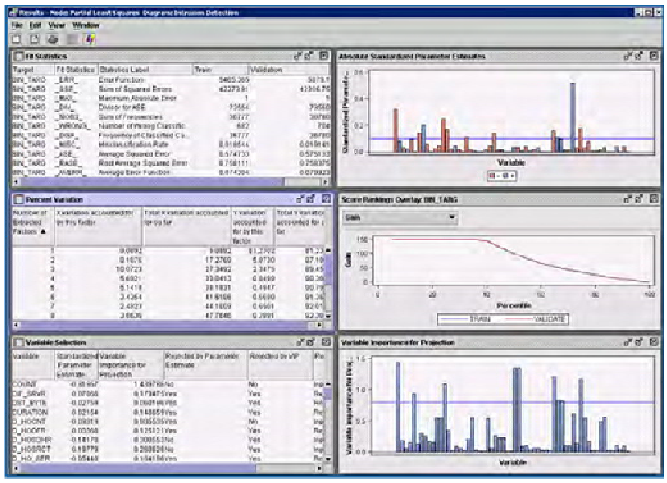

图 12: 使用偏最小二乘节点，抽取非关联的潜在因子或成分的候选集。生成 SAS Enterprise Miner 标准的模型拟合统计量，方便与其它挑战者模型算法进行比较。

偏最小二乘节点

- 特别适用于从大量潜在相关的变量中抽取重要因子。
- 还可以执行主成分回归和减秩回归。
- 用户可以人工指定或者自动选择因子个数。
- 五种可选的交叉验证策略。
- 支持变量选取。

两阶段建模

- 同时对类别和区间目标进行预测，采取顺序式和并发式的建模方法。
- 在每个阶段分别使用一种决策树、回归或神经网络模型。
- 类别目标的预测结果可以影响到对区间目标的预测，这种影响是可控的。
- 准确估计客户价值。

Ÿ 支持不随时间变化的协变量。

Ÿ 采用 holdout 验证来计算生存函数。

Ÿ 生成竞争风险或子风险函数。

Ÿ 基于生命平均残差计算的评分卡。

费率制定（用于保险业）

Ÿ 对索赔频率、损失程度和纯保费进行建模

Ÿ 自动选择合适的数据分布和链接函数，同时支持用户自行指定。

  Ÿ 支持如下数据分布：泊松分布、负二项分布、gamma 分布、正态分布、逆 guassian 分布、tweedie 和零堆积泊松分布。

  Ÿ 支持如下链接函数：对数函数、逆函数、power squared 函数、power cubed 函数和 log-log 函数。

Ÿ 支持效应预分箱的快速内存建模。

Ÿ Tweedie 模型优化方法，支持自动、鞍点、likelihood 和 final likelihood。

Ÿ 交叉效应生成器。

Ÿ 用户可以决定如何定义 ZIP 模型，来处理过度散布问题。

Ÿ 标准的拟合统计量包括 AIC、SBC、皮尔森卡方和离差。

Ÿ 对于每个模型效应，计算 relativities plot。

Ÿ 生成 SAS 评分代码。

报表生成器节点

Ÿ 采用 SAS ODS（输出交付系统）技术，可以创建 PDF 或 RTF 格式的分析流程报表。

Ÿ 有助于形成分析过程的文档，促进分析结果的共享。

Ÿ 文档可以另存，也可以包含在 SAS Enterprise Miner 结果包中。

Ÿ 包括分析流程图的图片。

Ÿ 用户输入的备注信息。

评分

Ÿ 在 SAS Enterprise Miner 图形用户

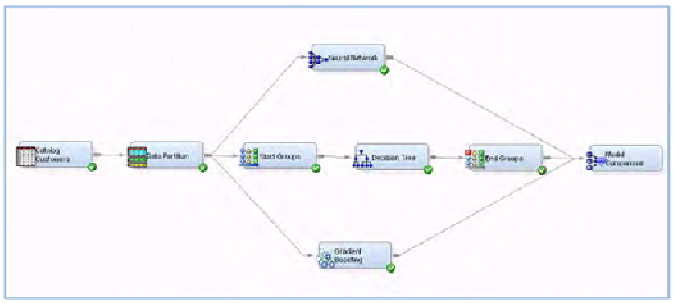

图 13：使用开始组和终止组节点以及决策树节点，来创建袋装与推进模型。

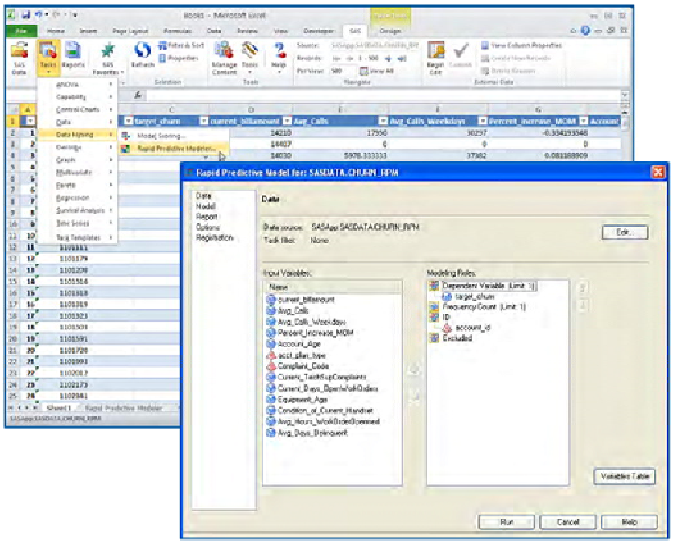

图 14：在 SAS Enterprise Guide 或 SAS 的 Microsoft Office 插件（目前仅限于 Microsoft Excel）中，可以使用 SAS 快速预测建模器，来自动建立预测模型，解决多种业务问题。

模型集装

Ÿ 通过合并多个模型的预测结果，来获得更健壮的模型。

Ÿ 提供以下集装方法：取平均值、投票和取最大值。

分组处理包括开始组和结束组节点

Ÿ 对一段分析处理流程重复执行

Ÿ 适合解决以下问题：分层建模、袋装与推进、多重目标、交叉验证。

SAS Enterprise Guide 或 SAS 的 Microsoft Office 插件（目前仅限于 Microsoft Excel）中的 SAS 快速预测建模器

Ÿ 自动建立预测模型，解决多种业务问题。

Ÿ 可以在 SAS Enterprise Miner 中打开、修改和增强这些自动建立的模型。

Ÿ 产生简要报告，包括变量重要性、提升图、ROC 图和模型评分卡，使之易于使用和检视。

Ÿ 能够对训练数据评分，同时允许您保存评分数据集。

模型导入节点

Ÿ 可以注册 SAS Enterprise Miner 模型，在其它流程图或项目中重用。

Ÿ 导入和评估外部模型。

界面中，包括了评分节点，用于交互式评分。

- 能够自动生成优化的评分代码，去除未用到的变量。
- 可以自动生成多种语言的评分代码，包括 SAS、C、Java 和 PMML（3.1 版本）。
- 在 SAS、C 和 Java 评分代码中，包含了模型代码、聚类代码、数据转换代码和缺失值的补缺代码。
- 通过 SAS Scoring Accelerator，可以在 Teradata、IBM DB2 和 Netezza 数据库中，直接使用 SAS Enterprise Miner 的模型进行库内评分。
- 支持将模型部署在多种环节中。

模型注册与管理

- 可以将 SAS Enterprise Miner 模型注册到 SAS 元数据服务器上。
- 支持与 SAS 模型管理器的整合，实现评分代码的版本控制、模型生命周期管理（从模型开发、投产到模型监控）。
- 支持与 SAS Enterprise Guide、SAS 的 Microsoft Office 插件、SAS 数据整合工作室等软件的整合，在这些软件中使用模型进行评分。

软件要求

若要深入了解 SAS Enterprise Miner 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：www.sas.com/enterpriseminer。

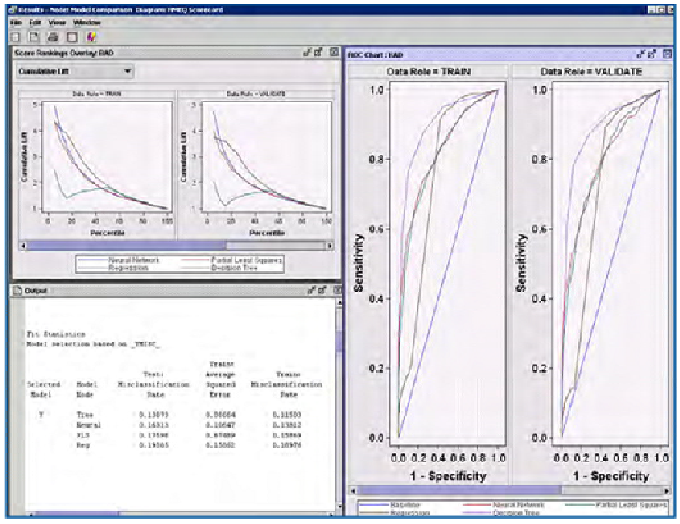

图 15：使用模型比较节点，可以在一个易于理解的框架下，对多个模型同时进行模型评估。

模型评估

- 模型比较节点，可用于在统一框架内对所有保留数据进行多个模型的比较。
- 基于用户指定的模型评估准则，自动选择最佳模型。
- 允许用户覆盖自动选择的结果。
- 丰富全面的拟合诊断统计量。
- 提升图、ROC 曲线。
- 基于决策数据的收益损失图、混淆矩阵。
- 分类概率评分分布图、评分排名矩阵图。
- 区间型变量的评分排名和分布。
- 截止值节点，可用于设定二值型目标变量的概率评分截止值。
- 允许用户覆盖缺省的选择。
- 最大的 KS 统计量。
- 最小的误分类率。
- 最大化累积特征。
- 最大的真阳性率。
- 最大的事件精度（基于训练集的先验概率）。
- 相同精度下的事件召回率。
- 通过一个交互式的图形化界面来刻画模型的特征：
  - 比较对照多个模型，这些模型可以使用同样算法或不同算法。
  - 评估变量的重要性，以及对目标变量的影响。
  - 基于要评估的结果，识别出最佳的函数设置，包括模拟。

THE POWER TO KNOW.

SAS 公司，免费咨询电话：400 818 1081

若要联系您当地的 SAS 分公司，请访问：www.sas.com/china

SAS 和所有其它 SAS 公司的产品或服务名称，是 SAS 软件有限公司在美国和其他国家的注册商标。® 表示美国注册商标。其他品牌和产品名称均为相关公司的注册商标。

版权所有，2011 年，SAS 公司保留所有权利。