


华院分析数据挖掘工具(Inside Miner)产品介绍

一、 华院数据挖掘工具的定位

目前市场上主流的数据挖掘软件较多，各自都有独特的优势。

公司	产品	优势	劣势	价格
	SAS/AP+EM	灵活、速度快、功能强大	图形化用户界面友好型差	非常贵
	Clementine	图形化用户界面友好	不灵活、速度慢、数据处理功能少	比较贵
	Kxen	自动化	不灵活、数据处理功能少，不支持工作流拖拉模式	一般
	Insightful Miner	图形化用户界面友好	数据处理功能少，模型方法少	一般

华院分析的数据挖掘工具选择 SAS/AP（基础模块）作为平台进行二次开发，吸取了众多数据挖掘软件之长处，并结合华院分析多年数据挖掘模型开发经验，定制化的开发了众多独特而又有价值的功能模块。

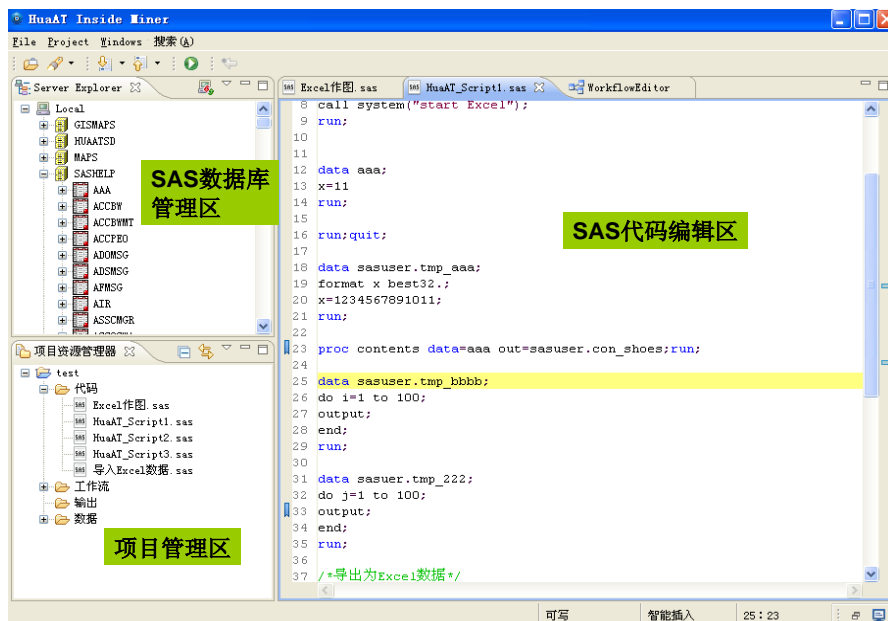
Inside Miner 的定位：



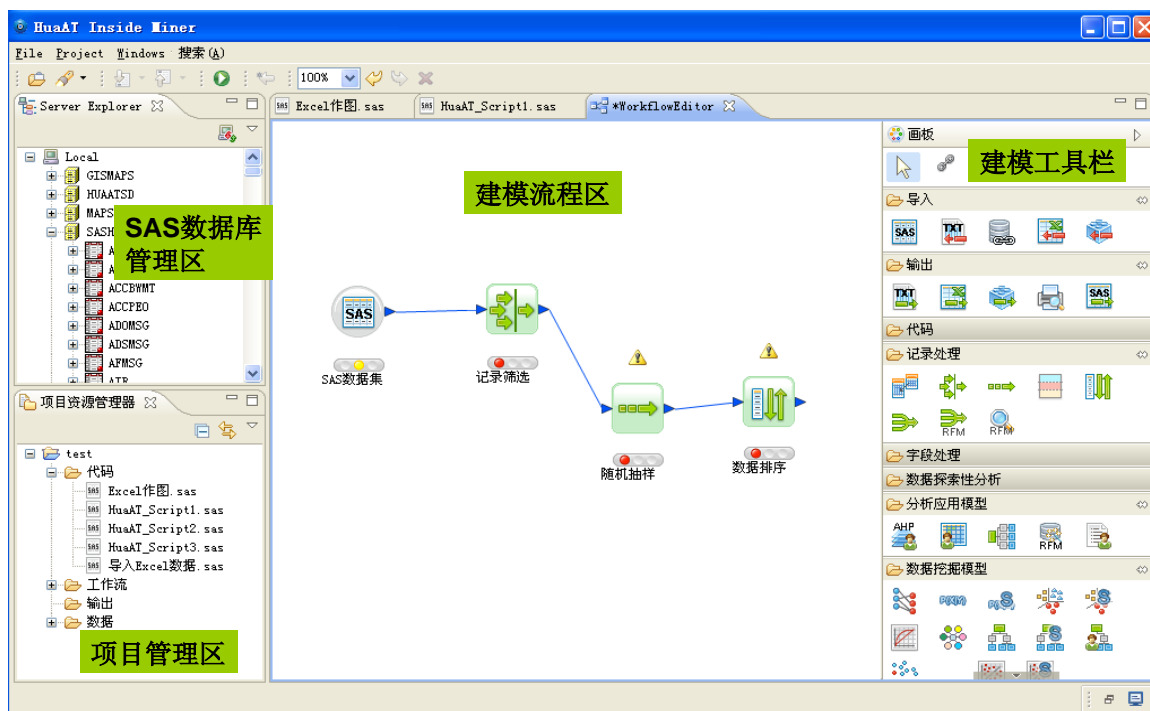
二、 华院数据挖掘工具功能

该工具使用 Eclipse RCP 技术开发，主要的 GUI 界面如下：

SAS IDE 界面（在该界面下编写 SAS 代码并直接提交给 SAS 运行）：



数据挖掘 workflow 界面（在该界面下拖拉节点开发数据挖掘模型）：



主要功能节点包括：

- 数据导入
 - SAS 数据导入
 - TXT 数据导入
 - EXCEL 数据导入
 - 数据库连接
 - PMML 模型导入
- 数据导出
 - 导出为 SAS 数据
 - 导出为 TXT 数据
 - 导出为 EXCEL 数据
 - 导出至数据库
 - 导出 PMML 模型
- 代码
 - SAS 代码
 - SQL 代码
- 记录处理功能
 - 记录筛选
 - 高级筛选
 - 抽样
 - 划分
 - 排序
 - 数据汇总
 - RFM 汇总
- 字段处理功能
 - 类型设置
 - 字段过滤
 - 新建字段
 - 数据表合并
 - 变量变换
 - 哑变量变换
 - 极值处理
 - 聚类筛选变量
 - 预测力筛选变量
 - 变量分箱
 - 重新分类
 - 缺失值处理
 - 标准化
 - 转置
- 数据探索分析
 - 描述统计
 - 交叉报表
- 分析应用模型
 - AHP 模型
 - 策略表
 - 策略树
 - RFM 分析
 - 评分卡
- 数据挖掘模型
 - Logistic 回归
 - 线性回归
 - 决策树
 - 人工决策树
 - 朴素贝叶斯
 - K-means 聚类
 - 因子分析
 - KNN
 - 关联规则

三、 华院数据挖掘工具特点

- 基于工程的项目管理（一个项目一个文件夹，包含代码、 workflow、数据等内容），可以实现团队知识共享(CVS)和项目文档的统一管理
- 以 SAS 作为后台进行二次开发，充分利用了 SAS 在数据统计领域的强大功能！
- 基于 workflow 的数据分析/挖掘拖拉式操作， workflow 增加“状态灯”、“输入输出端口”等辅助功能，比 Clementine 更胜一筹。
- 开放的平台，功能完全可以定制，任何人写的好的 SAS 代码都可以固化成一个可以复用的功能节点。