



### SAS® Contextual Analysis 功能?

采用 SAS Contextual Analysis (SAS 语境分析), 可通过单一界面对文档分类、提取事实、了解文档情感, 迅速获得文本数据的分析结果。同时, 可以定义规则, 无需创建训练语料库。

### SAS® Contextual Analysis 重要性?

SAS Contextual Analysis 便于透过文本数据—用以更好地了解业务和客户的信息, 轻松识别出现的问题和今后发展趋势, 甚至可以帮助企业遵守行业法规。

### SAS® Contextual Analysis 目标用户?

企业很容易积累过多文件, 没有时间或资源手动查看。由于 SAS Contextual Analysis 便于安装运行—并且可以定制—因此是 IT 和业务用户的理想工具。

# SAS® Contextual Analysis

利用非结构化文本更全面地分析企业

文档、客户通信和其他类型的非结构化数据快速积累, 形成无法手动整理或查看的海量资料。而正确的技术可以把丰富的信息变为财富。

SAS Contextual Analysis (SAS 语境分析) 是一种先进的语言解决方案, 帮助您从非结构化文本中提取重要业务分析结果。它利用机器学习和相关专业知​​识, 揭示数据中隐藏的趋势和主题。否则, 难以发现这些有利于深入了解业务的洞察结果。

虽然 SAS Contextual Analysis 是一种先进的技术, 但使用并不复杂。您不必创建分类训练语料库 - 软件可以发现, 自动识别用作起点的主题。而且这种软件十分灵活。您可以细化机器生成的分类规则, 使其更加具体, 提取所需的分析结果, 进一步提高业务价值。

## 优点

- **单一集成系统便于管理分析。**采用 SAS Contextual Analysis, 可在单一集成环境下快速启动运行规则, 并在必要时进行定制。发现过程由起点自动执行。软件还允许利用自然语言处理技术自行设置定义。
- **引导式混合方法加强控制。**采用 SAS Contextual Analysis, 您可以通过混合方法定制文本分析, 为自动机器学习分析增加相关专业知​​识。这种技术依照语法结构识别数据的共同主题、分类规则和文本情感。您可以随时查看和修改机器学习结果, 满足您的特定需求。同时提高灵活性、处理能力和控制能力, 更加快速地获得结果。
- **准备数据加大覆盖范围。**SAS Contextual Analysis 不仅能够帮助获取非结构化数据的价值, 而且可为其他技术准备数据, 从而可以与结果直观交互。实施可视化分析或创建新的元数据以丰富其他应用。



## 产品概述

非结构化数据很难使用，但如果有正确的技术，便可以发挥其潜在价值。忽略非结构化信息，可能导致丢失重要的业务分析结果，如客户关注、新趋势，甚至潜在问题的早期迹象。为保持竞争力，企业需要这种情报。

但是，并非所有的企业都有时间或资源梳理无穷无尽的文档。即使能够这样做，也很容易产生人为错误。今天数据丰富的工作环境需要不仅功能强大，而且直观、有能力并且可以定制的技术。因此，您可以采用 **SAS Contextual Analysis**，这种技术专门为您提供重要分析结果，从而提高竞争力。

## 单一系统环境下完成文本模型开发和部署

单一环境下完成语言模型的开发、优化和部署，有助于简化工作量和文档分类。由于通过单一界面管理，因此，您可以更好地控制系统参数。

## 混合方法分类文档

机器学习依照语法结构发现共同主题、分类规则和文档情感，同时，可以利用相关专业定制分析。这两种方法相结合，可以更加全面地了解语境含义—快速获得更好的分析结果。

## 自然语言处理

手动查看文档耗费时间，而且容易出错。数据驱动的自然语言处理不需要繁琐的人工定义。确定并输入数据源之后，解决方案可自动识别词条频度、词类、同义词和词干。

## 发现词首分类形成的主题

如果采用人工的话，确定多个文档的中心主题十分艰难的工作。**SAS Contextual Analysis** 可以自动完成这个工作，利用机器学习和统计方法，确定文档组合中的核心主题。

## 生成可配置的规则改进语境

**SAS Contextual Analysis** 简化布尔规则定义和分类开发，利用申报专利的算法自动创建分类规则。确定规则后，可利用定义的概念进行语义增强和细化。

## 直接集成 SAS®

通过文本分析发现主要业务洞察结果本身具有重要价值，但还可以进一步深入分析。**SAS Contextual Analysis** 提供丰富的文本分析结果，为进一步分析准备数据，如 **SAS Visual Analytics** (SAS 可视化分析)。

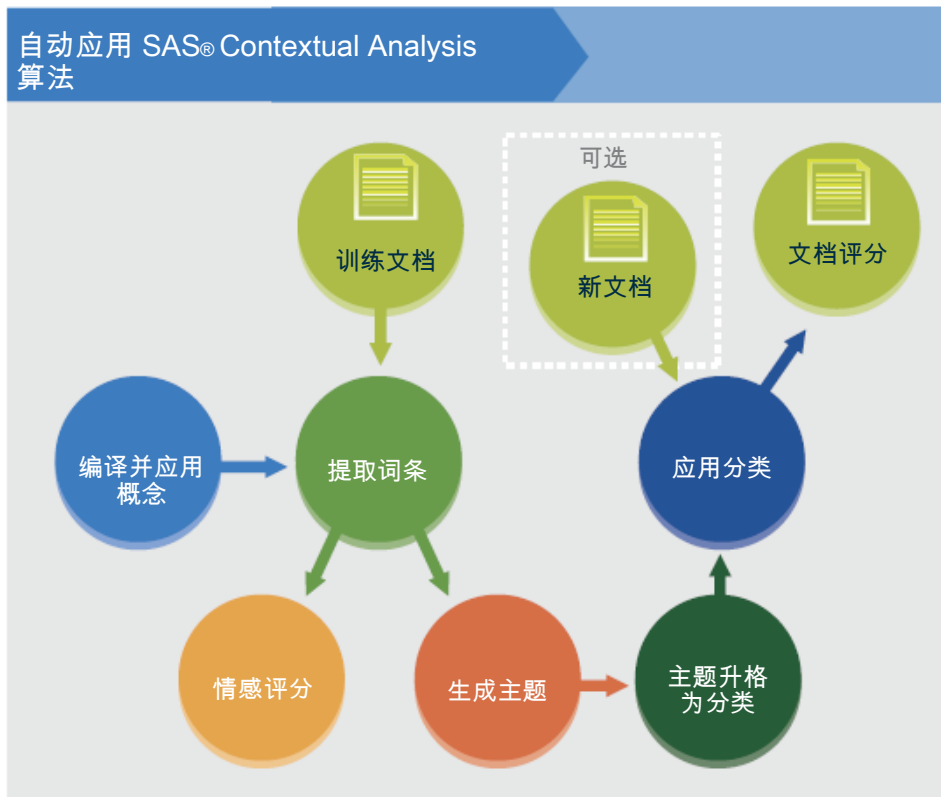


图 1: 分析师可进一步提炼机器生成的结果。

如想了解 **SAS Contextual Analysis** 系统要求的更多信息，  
下载白皮书，查看截屏图片，阅读其他资料，请访问  
[sas.com/contextual-analysis](https://sas.com/contextual-analysis)

## 主要特点

### 单一系统环境下完成文本模型开发和部署

- 单击 **GUI** 界面支持自然语言处理及：
  - 词条
  - 主题发现.
  - 分类定义
  - 概念规格
  - 文档级情感识别
- 帮助文本模型开发的研究、细化和文本模型部署。
- 项目启动时可导入现有 **SAS Content Categorization** 分类。
- 可利用同一交互界面中的验证检查创建、修改和启用 (或禁用) 定制概念，并测试语言规则定义。
- 预建项目创建向导，支持初始项目定义。
- 在线帮助指导分析师完成文本模型开发过程。
- 属性面板详细显示文本模型创建每一个步的状态、处理状态和消息对话，帮助诊断模型开发问题 (如收集生成主题的文档数量不足)。
- 项目管理选项允许同时开发和运行多个文本模型，附每种模型开发状态的相关说明。
- 集中所有项目属性的元数据管理。

### 混合方法分类文档

- 自动进行词条定义 (修改起止列表)、同义词检测、拼写检查和词干提取的自然语言处理。
- 拖放定制功能便于改变窗口之间保存或拖入的主词条。
- 基于 **SAS Text Miner** 的主题发现，支持机器生成的分类初始化。
- 可选择后用数据源 - 先建分类，然后使用文档语料库。
- 可直观修改主题中包含的词条，合并相似主题，以细化生成的结果。
- 主题可升格为分类，规则生成支持词首数据驱动的分类规则定义。
- 可利用丰富的预制语言运算符编辑系统派生的分类，进一步细化规则。
- 可使用 **17** 种普通实体预定义的可配置概念，建立定制概念以增加语境敏感性，包括通讯地址和车辆信息。
- 可启用或停用特定评估的概念。
- 验证部分或全部定义概念规则的规则语法。
- 生成的文档自动形成数据驱动文档级情感，作为主题发现或选择分类的一部分。

### 自然语言处理

- 软件自然语言处理包括自动解析、断词、词性标注、同义词检测、拼写检查和词干提取。
- 输出包括拖放/保存词条列表和词条频度。
- “保存”与“拖入”词条窗口之间的拖放 (仅主词条)。
- 附加文本处理包括利用起止列表 (加入/排除处理的词条)。包括一般不做分析的常用词条初步禁止列表 (如冠词和短连词)。分析师可编辑预定义禁止列表 (目前仅英文)。
- 软件可使用用户定义的起/止列表。
- 可通过定义输入 **SAS** 数据的相应列包含自定义分类。
- 定制概念面板中也可添加自定义概念，采用预定义运算符语法编写。
- 文本分析师可开发分类，不需要输入文档，然后根据语料运行分类。

## 特点 (续)

### 自动发现词首分类形成的主题

- 自动机器发现功能识别输入文档集中的核心主题及相关的关联度
- 主题中的词条关系可用语条云 (可配置阈值)、交互式词条图查询, 钻取评估相关性的主题并细化发现的主题。
- 直观描述主题相关文档的主题级情感。
- 通过拆分主题 (一个主题分为两个, 如有必要可重复), 以及相似主题合并为一个, 控制生成的主题数量。
- 所需主题升格为类别规则定义的分类, 进一步增强概念定义。

### 生成可配置的分类规则

- 基于用户细化生成的主题自动定义词首分类规则。
- 易于理解的布尔规则定义创建分类模型 (即分类)。
- 真阳例、假阳例和假阴例可视化诊断详细列出系统生成规则的强健性。
- 文档视图生成直观描述的分类情感。
- 规则可编辑、增强、删除或按定制分类重新定义。
- 丰富的预制规则运算符列表可用于详细规则模型分析。
- 分析可包括输入数据集或现有 **SAS Content Categorization Studio** 项目预定义字段预制的分类层级 (即分类)。
- 部署之前, 可基于输入数据集测试规则。

### 利用概念改进语境

- 概念提供细化模型特定分类语境, 提取更确切含义。
- 可选预定义概念 - 无需编写规则。这些概念用于普通实体定义日期、地点、时间等。
- 可导入 **SAS Content Categorization Studio** 项目既有概念。
- 可采用一套预定义运算符编写自定义概念。
- 预定义和既有概念规则可根据需要编辑细化。
- 点击按钮验证概念规则, 详细诊断识别出的任何语法问题的错误信息。

### 多种本地语言支持

- 支持中文 (简体和繁体)、荷兰文、英文、芬兰文、法文、德文、意大利文、日文、韩文、葡萄牙文、俄文、西班牙文和土耳其文文档处理。
- 提供英文、法文、荷兰文、中文和韩文本地化语言界面。

### 可扩展文本分析解决方案

- 采用单个服务器多核分布式处理, 扩展文档分类和情感发现。
- 自动创建分布式处理的分数代码, 扩展新文档评分范围。

### 直接集成 SAS®

- 集成 **SAS Metadata Server (SAS 元数据服务器)** (SAS 数据集现在可直接输入文本分析)。
- 直接集成 **SAS Visual Analytics Hub** 可进行交互式结果搜索。
- 模型评分代码下载功能适用于任何其他 **SAS** 应用的语境模型。包括查看下载情感和概念分数代码, 以及用于其他文档评分的分类分数代码。