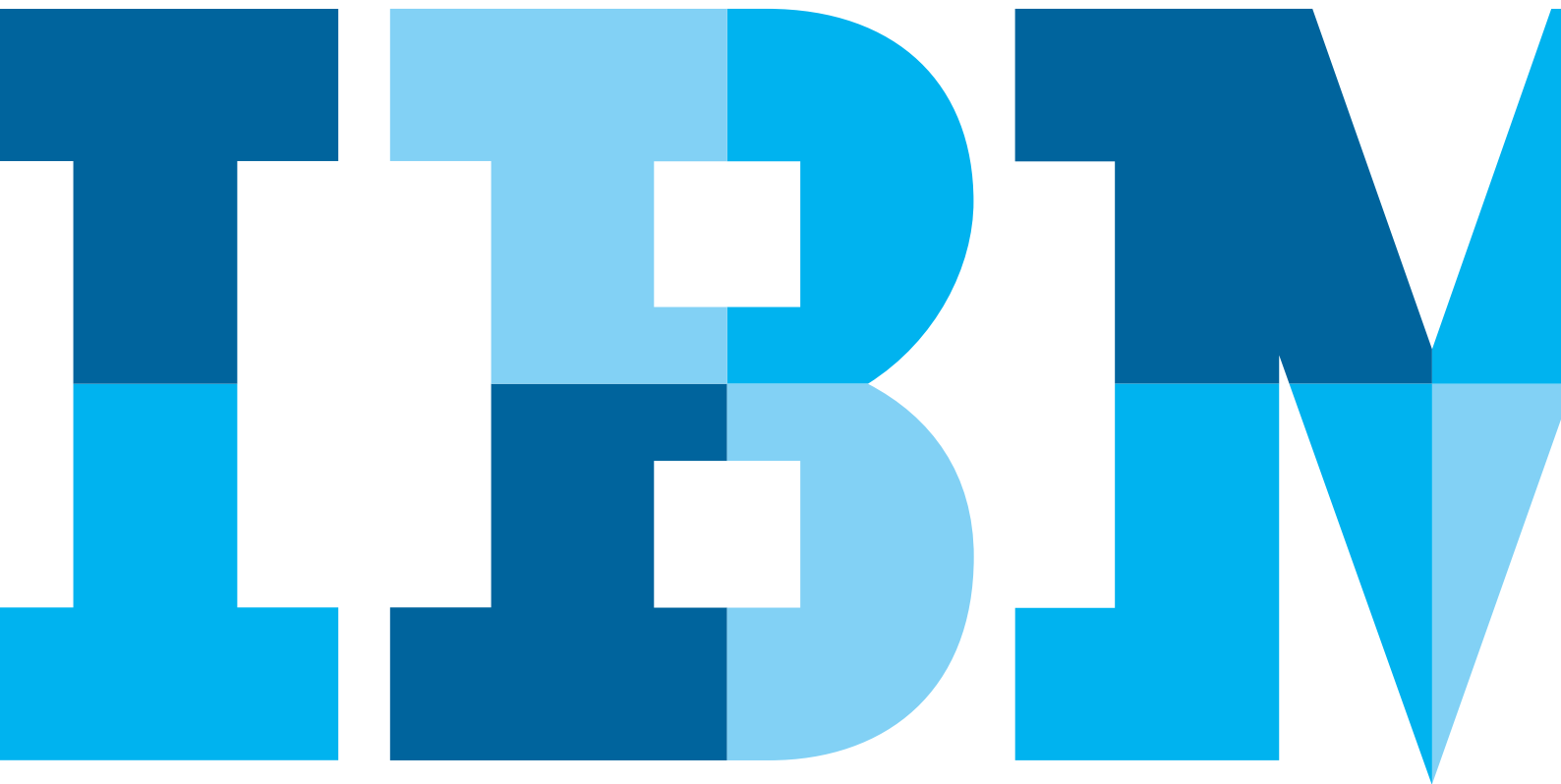


R 和 SPSS 软件:共赢策略

*Bore serum rernatem exerspis molorio tota pro eum
facearibea solorerum sequatqui sitiust*



目录

- 2 概述
- 3 R 的简介
- 3 R 的局限
- 4 R 和 SPSS 软件组合的优势
- 7 结论
- 7 关于 Business Analytics
- 7 咨询联系

概述

本论文旨在证明:与单纯使用 R 编程语言相比,R 与 IBM® SPSS® Statistics 和 Modeler 软件搭配使用时更具优势。R 用户通过 SPSS 软件可以实现高级数据管理,访问点击式界面,获取演示级别的输出信息和更高的可扩展性。SPSS 软件的用户可以获得丰富的不断扩展的统计分析和图形库,这些可以帮助他们获得对数据更深的见解。

搭配使用 SPSS 软件和 R 可以让二者都能发挥最大性能。SPSS 软件可以在您的 SPSS 界面上运行 R 语法。您可以通过自定义 R 编码来强化 SPSS 软件中现有的强大的数据操作、统计分析和预测算法功能,以提升效果和灵活性。您可以进行自定义分析,创建并处理输出信息,并在分析计划中融合新见解。此外,您还可以通过 SPSS Custom Dialog Builder 将 R 编码分享给可以从新的分析方案中获益、但选择不用编程进行分析的人员,以实现 R 编码的再利用。开发人员可以专注于写代码,而用户不用了解 R 就能专注于分析和使用基于 R 的功能。

R 的简介

R 是统计计算和制图中使用的一种开源编程语言和软件环境 (www.r-project.org)。统计师和数据挖掘员经常使用 R 语言来开发统计软件。此外, R 语言在高级数据分析中也被广泛使用。R 中包括众多统计和图解技术, 例如线性和非线性建模、经典统计检验、时间序列分析、分类和聚类。R 的使用须遵守 [免费软件基金会](#) 和 [GNU 通用公共许可证](#) 的条款。

R 中包含来自多个资源库的 4800 多个包, 涵盖经济计量学、数据挖掘、空间分析、生物信息学等多个专业领域。R 已经通过内部权威统计师和计算机学家的审核。但是由于 R 是开源编程语言, 因此其发布和程序包不含编码化进程。

R 的局限

因为 R 可免费获得, 所以人们普遍认为可以用它取代商业统计和建模软件, 以此为机构节约大量成本。但这一想法是有些偏差的。尽管 R 有很多优点, 但它尚不能满足您从数据分析中获取最大业务价值的所有需求。以下局限性是 R 用户的主要担忧:

- 部署。通过 R 将预测性输出信息整合至操作环境中比较困难。
- 界面。R 没有现代化图形用户界面, R 程序员之外的人员很难使用。
- 学习曲线。R 不是所有人都能轻松学会的, 因为并非所有人都是程序员。
- 数据。R 很难连接至本机数据库。
- 输出信息。很难生成可即时发布的输出信息。
- 性能。R 可以快速消耗所有可用内存。
- 协作。分析小组通过 R 分享工作内容比较困难, 尤其是在小组成员的 R 知识水平不一致的情况下。
- 企业安全。您下载的程序包没有安全保障。

此外, 用户必须确保拥有正确的 R 程序包和编码, 才能将程序包和输出信息串起来。这些局限往往会增加使用 R 的成本。

R 和 SPSS 软件组合的优势

通过 SPSS 软件使用 R 是可以解决 R 的局限性问题的。对于专家人数较少、又想从 R 中获益的组织, R 和 SPSS 的强强联合可以满足其需求,同时又不会带来陡峭的学习曲线。SPSS 软件是一个很便捷的平台, R 用户可利用它来处理大型数据集合,获得优质图形和其他格式的输出信息。其他优势还包括:SPSS 软件使用方便,能够将集成的 R 程序包分发给对 R 不熟悉的众多用户。更重要的是,通过 SPSS 软件使用 R 可以解决 R 的局限性。

部署

SPSS 软件能以经济节约的方式应对 R 的部署局限。例如,用户可通过 IBM SPSS Modeler Gold 来优化业务规则和预测模型对销售、客户服务、维护等其他方面的影响方式。SPSS Modeler Gold 可兼容集成 R 的模型。SPSS Modeler 还可通过集成 IBM Cognos® TM1® 提高商业智能的使用效率。IBM 能够向用户提供多种使用预测智能的方式。用户无需创建或购买其他软件,便可将预测洞察、历史和实时数据视图整合至其互动的移动仪表板中。

界面

SPSS Modeler 和 SPSS Statistics 可提供简单的图形用户界面,支持各种数据准备、统计分析和预测建模算法。R 编码可以与其他现有的全部功能和特性在同一界面运行。用户还可以为 R 添加 GUI 前端。程序员之外的用户可通过该前端输入客户值(图 1),如变量名等。用户无需了解背后的编码,便可使用高级分析功能。

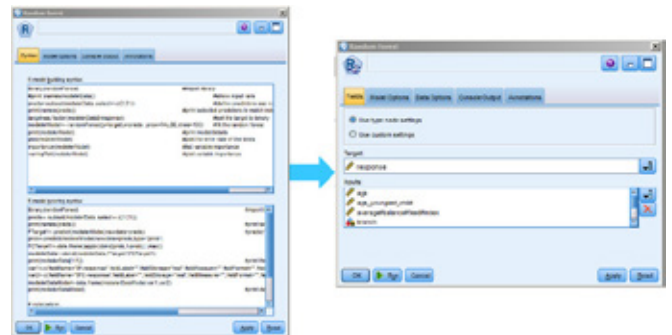


图 1:R 编码可用于创建客户对话框,可以让程序员之外的用户在分析中更轻松地使用编码。

学习曲线

R 属于编程语言, 因此是必须通过学习来掌握。但是 R 并不容易学。而且在花时间学习 R 的过程中, 您无法充分利用其功能。但是, 将 R 集成到 SPSS 软件就能让您立即充分体验 R 的功能。您可以重点学习那些真正独特的 R 程序, 并使用 SPSS 软件中已经包含的数据、统计分析和建模。

数据

通过 R 来访问分析所需的数据需要花费大量时间和精力。您必须编写大量代码, 执行程序包, 甚至使用 Java。SPSS Statistics 和 SPSS Modeler 可以解决这一数据存取问题。二者均能与 SQL、Oracle、SAP、IBM Netezza®、DB2® 商业数据库等共用。SPSS Statistics 和 SPSS Modeler 可以读取文本输入、电子表格、SAS 文件和其他格式的文件。SPSS Modeler 可以直接从 IBM Cognos Business Intelligence 和 Cognos TM11 中读取信息。Wizard 程序带有预制的存取数据接头的, 节省了分析之间提取、转化和操作数据的大量时间。SPSS Statistics 和 SPSS Modeler 可在点击式界面上提供强大的数据操作技术。用户可以对数据进行调换、检查和重定义格式。SPSS Modeler 还提供自动准备数据功能, 这可以一键优化预测建模的数据。

输出信息

SPSS Statistics 和 Modeler 包含多种制作现成演示表格和图形的方法。R 程序员可以通过 SPSS Statistics 软件将 R 功能导入 SPSS 软件语法中。这样您就可以生成演示级别的图形、数据透视表和其他格式的输出信息(图 2)。用户还可将结果发布为 PDF、Word、PowerPoint、Excel 等其他常用格式。此外, 您还可以将 R 与实体分析、社交媒体分析和文本分析等高级 SPSS Modeler 功能进行集成。

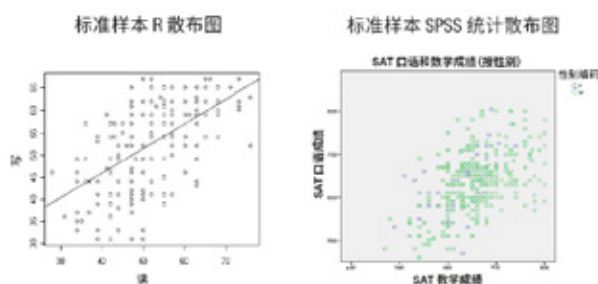


图 2: SPSS 可生成演示级别的图表和图形, 为企业利益相关者节约大量用于准备和调整分析输出信息格式的时间。

性能

R 指令需要在内存工作区运行。

R 指令在设计时并未考虑内存,在单个 R 会话中,所有对象都存储在临时的工作内存中。这样会导致现有内存被快速占用。通过组合 R 和 SPSS Modeler,您可以对传输至 R 的数据进行分割和取样(图 3)。此外,SPSS Modeler Server 是一种内存开发技术,可以将分析转到磁盘来进行,从而空出内存。您可以使用 R 指令和创建 R 对象,同时又不会对数据挖掘和建模的整体性能造成重大影响。此外,SPSS Modeler 还能用于 IBM Netezza 和 SAP Hana 等其他环境测量数据库中的 R。而且利用 SPSS Analytic Server 技术还可以衡量 Hadoop 中的 R。

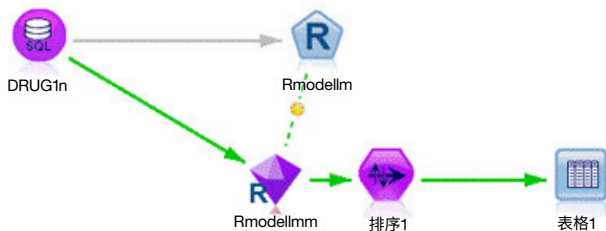


图 3: IBM SPSS Modeler 被用于在 SQL 数据库中存取本机数据,同时 R 模型被用于获取记录。紫色表示该分析步骤正在使用 SQL Pushback,可通过在数据库中执行内存密集型任务来提升性能。

合作

R 是一种针对个人设计的语言。但多数分析工作由团队合作完成,会有多人参与建模和统计分析。将 SPSS 软件和 R 联合使用可以摒弃 R 语言的个体局限,并且获得 IBM 世界级的协作功能。SPSS 软件提供了一套完整框架,可以对 SPSS Statistics 和 SPSS Modeler 开发的分析资产进行集中化、保护性和自动化操作。这样可以确保模型和统计分析得到安全分享,遵守公司法规。

企业安全

R 缺乏正式发布流程,因此不包括正式的质量保证环节。如果没有质量保证,您就需要自行承担使用 R 的风险。您无法确保某个发布或包是否能提供其所述的功能。此外,R 程序包从网页或用户社区下载而来,因此包的安全性没有保证。事实上,您可能会无意中下载恶意软件、特洛伊木马病毒、数据利用软件等。而作为 IBM 软件质量认证流程的一部分,SPSS Modeler 和 Statistics 已经通过严格测试。使用 IBM 的软件后,您不会再因风险操作而威胁工作环境的安全。

结语

SPSS 和 R 都有自己的优势, 不仅经得起时间考验, 而且在分析界得到了强烈认可。此外, 二者结合可以实现优势互补, 从而带来更强大的功能和特性, 可以令整个分析界受益。R 用户可以访问高级数据管理功, 从而处理更庞大的数据集。而 SPSS 软件则向 R 用户提供了更丰富的图形和透视数据表输出方案, 从而提升用户体验。最后, SPSS 软件可以将 R 程序包发布至众多用户, 是一个理想的部署工具。

SPSS 软件的用户也能访问更多的统计功能。因此他们无需学习复杂的编程语言(如 R 语言), 便能执行复杂的分析。R 和 SPSS 软件的组合优势绝对值得考虑。

关于 Business Analytics

IBM Business Analytics 软件可提供数据驱动型洞察力, 帮助各个机构以更加智能的方式运营, 并且超越竞争对手。这款综合性产品组合为商业智能、预测分析和决策管理、绩效管理和风险管理等领域提供了解决方案。Business Analytics 解决方案能让企业确定并直观显示客户分析等领域的趋势和模式, 从而对业务绩效产生深远的影响。公司可借其比较各种场景; 预测潜在的威胁和机会; 改进计划、预算和预测资源; 根据预期的回报平衡风险并致力于满足法规要求。通过进行广泛的分析, 各个组织可让策略和战略决策协调一致, 最终达成业务目标。更多信息请参见:

ibm.com/business-analytics。

咨询联系

如欲咨询或提问, 请访问

ibm.com/business-analytics/contactus。IBM 代表将在两个工作日内答复您的问题。



© IBM 公司版权所有 2014

IBM Corporation
Software Group
Route 100
Somers, NY 10589

美国印制
2014 年 3 月

IBM、IBM 徽标、ibm.com、Cognos、DB2 以及 SPSS 是 International Business Machines Corp. 在全球多个管辖区注册的商标。其他产品和服务名称可能是 IBM 或其他公司的商标。当前的 IBM 商标列表请见网站的“版权和商标信息”版块：www.ibm.com/legal/copytrade.shtml。

Netezza 是 IBM 旗下公司 IBM International Group B.V. 的注册商标。

Linux 是 Linus Torvalds 在美国和/或其他国家/地区的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家/地区的商标。

Java 和所有基于 Java 的商标和徽标为 Oracle 和/或其子公司的商标或注册商标。

UNIX 是 The Open Group 在美国和/或其他国家/地区的注册商标。

本文档包含截至发布之日的最新信息，IBM 可能随时进行更改。并非所有产品或服务在 IBM 开展业务的所有国家/地区均有提供。

本文所载信息按“原样”提供，不做任何明示或暗示的担保，包括对适销性、特定目的的适用性的任何担保，以及针对非侵权的任何担保或条件。

IBM 根据产品交付协议中规定的条款和条件为产品提供担保。



请回收利用