

基于 R 语言的大数据挖掘平台

使用指南

作者：R&R&R（山东大学）

一、编写目的

本使用指南的编写目的是让熟悉或从事数据挖掘工作的人员能够快速掌握本软件的使用，减少在使用过程中的阻碍。

二、适用人员

熟悉或了解数据挖掘，有一定的数据挖掘专业知识。对于没有相关经验的用户，阅读本手册不能使你正确使用本软件，更不会让你掌握数据挖掘专业知识。但它可能会增加你对数据挖掘的感性认识。

三、使用说明

1.数据的导入

导入数据有两种途径。

- 1) 当你启动本软件时，你会在主面板中找到数据源子面板，其中提供了文件、数据库和 R 数据集三种数据来源。点击相应按钮后，会弹出对话框，按照提示操作即可。
- 2) 在任何时候，你都可以点击菜单栏中的文件按钮，从弹出的菜单项中选择“导入...”这一项，其它操作同上。

2.感受数据

在导入数据之后，可以看到数据面板中有“摘要”、“头部”、“缺失值”三个按钮。

“摘要”可以显示数据的常用统计信息；

“头部”，即 head，可以显示数据的前若干行，使用户对数据有直观的了解；

“缺失值”显示数据的缺失值信息，并给出处理缺失值的集中方法。

此外，您还可以切换到可视化面板，其中提供了直方图、散点图、QQ 图、箱图、分位箱图等可视化方法。

3.分析数据

本软件目前版本只支持分类、聚类和简单的关联性分析。它们分别在同名的面板中。打开之后，选择相应的分析方法，并设置参数，即可观察图像和文字结果。部分方法还可以对数据进行进一步操作。该部分对数据挖掘知识要求较高。

4.数据的展示

切换到导出面板后，可以看到有四部分。它们分别可以对数据生成文字摘要、图像打包、电

子表格和图文并茂的报告（测试功能）。本软件目前的版本暂不支持自定义导出文件的路径和文件名。请到指定位置寻找。

此外，您还可以通过文件>导出...导出相应的文件，但该位置只提供简化的功能，不能自定义导出的内容。建议使用导出面板。

四、注意事项

本软件仍不完善，虽然开发人员已经尽力保证程序的健壮性，但仍然有程序出错的风险。如果问题，请邮件联系 andong777@163.com。