

## 第2章 RapidMiner Studio 简介

RapidMiner Studio 结合技术性和适用性，为最新的及已建立的人性化数据挖掘技术提供服务。通过推拽算子，设置参数及组合算子，在 RapidMiner Studio 中定义分析流程。

正如我们将在下面看到的，流程能从大量的随机的可嵌套的算子中产生，最终表示为所谓的流程图（流程设计）。流程结构由内部的 XML 来描述，通过图形用户界面来开发。在后台，RapidMiner Studio 不断地检查当前流程开发状态，确保语法一致，并在问题出现时，能自动推荐解决方案。以上功能是通过所谓的元数据转换实现的，即在流程设计阶段转换基础元数据，预知流程开发结果，并在出现不合适的算子组合时确定解决方案（快速修复）。此外，RapidMiner Studio 也能定义断点，因此能检查几乎所有的中间结果。成功组合的算子会被合并到构建模块中，因此在后期流程中它们还能被再次使用。

RapidMiner Studio 包含了 1500 多个专业数据分析流程：从数据划分到以市场为基础的分析，再到属性生成，它包含了所有您需要的数据挖掘工具。同时也包含了文本挖掘，网页挖掘，对来自网络论坛的信息的自动情感分析（情感分析，观点挖掘）及时间序列分析，并能作出预测。

RapidMiner Studio 为我们提供了强大的视觉化功能，例如三维图，散列矩阵和自组织地图等。在 RapidMiner Studio 里，您能够把您的数据转换为完全可定制的可输出的图表，并且支持缩放、移动及调整功能，以求最优化视觉效果。

### 2.1 安装及第一个资源库

在使用 RapidMiner Studio 前，先要下载和安装该软件。在 RapidMiner 官网 <http://www.rapidminer.com> 上的下载区域下载相应安装包。

根据您的操作系统下载相应的安装包，并根据官网上 RapidMiner Studio 说明书进行安装。如果您是第一次运行 RapidMiner Studio，您需要建立一个新的资源库（参见图 2.1）。开始我们会被限制在自己电脑的本地资源库中，之后您便可以在网络上定义资源库并把它分享给其他人。



图 2-1 在第一次开始使用 RapidMiner Studio 前，先在您的电脑上新建一个本地资源库

对于本地资源库，您只需为它命名（别名），然后把它存在您硬盘的任何目录下即可。（参见图 2.2）您可以直接点击右边的文件夹图标来选择目录，建议在文本对话框中选择一个方便的位置，在那里新建一个目录。然后您可以把这个新的目录作为您本地资源库的一个基础。

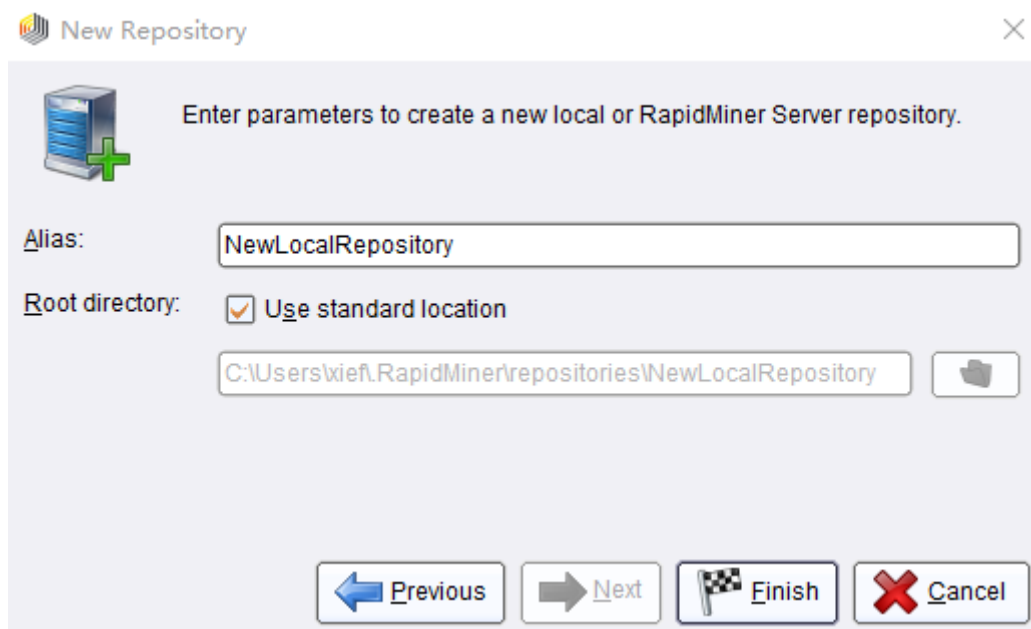


图 2.2：为存储数据和分析流程定义一个新的本地资源库。建议新建一个目录作为基础。

不论何种方式，RapidMiner Studio 将自动切换至设计透视图界面。

## 2.2 透视图及视图

在选择了资源库以后，您会进入到透视图的主页（图 2.3）。如果您处于联网状态，右侧部分会出现 RapidMiner 的近期新闻，中间的清单中列出了您以后会经常用到的操作。详细操作内容如下：

1.New Process（新流程）：打开设计透视图页面，新建一个分析流程。

2.打开：点击“Open（打开）”按钮，打开一个资源库浏览器。您可以在“Perspective（设计透视图）”中选择和打开一个已有的流程。如果您点击右侧的箭头按钮，将会出现一个近期被打开过流程的清单。您可以选择其中一个流程，该流程会在“设计透视图”界面中打开。

3.Application Wizard（应用向导）：利用应用向导，只需要三步您就能解决典型的数据挖掘问题。通过 Direct Maintenance Wizard（预测性维护向导）能够预测必要的维护需求。Churn Analysis Wizard（客户流失分析向导）帮助您识别哪些客户最有可能流失及其原因。通过 Sentiment Analysis（情感分析向导）分析社交媒体流，让您洞察客户的想法。（新版 RapidMiner Studio 更名为 Accelerator（加速器））

4.Tutorials（教程）：打开一个教程窗口，里面是一些教程：从创建一个分析流程到数据转换。每个教程都能在 RapidMiner Studio 中被打开使用，通过一些被选定的分析流程，这些教程为您介绍了一些数据挖掘的概念。

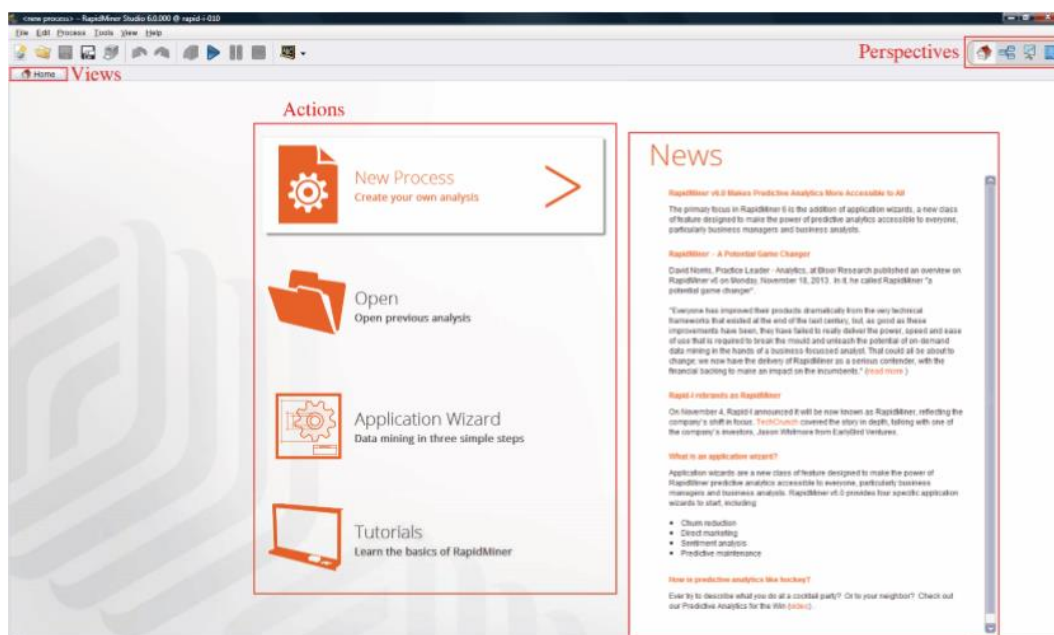


图 2.3：RapidMiner Studio 透视图主页

RapidMiner Studio 上部工具栏中,在右侧您可以发现 4 个按钮,这可以在各个 RapidMiner Studio 透视图之间切换。透视图由一组可自由配置的个人用户界面元素组成,即所谓的视图,它能根据您的喜好自主安排。

在透视图主页只有一个视图,至少一个预设视图,即您现在看见的主屏幕。进入“视图”菜单您能够打开其他的视图(图 2.4)。

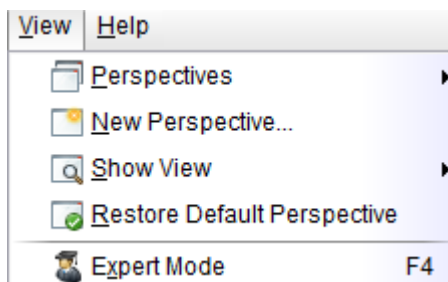


图 2.4: 视图菜单

在子菜单“打开视图”中,您能看到 RapidMiner Studio 的所有可用视图。我们在当前透视图看到一个视图,该视图被标记了一个勾。选择另一个视图,譬如名称为“LOG”的视图。现在,在图 2.5 中您能看到在透视图主页中增加了另外一个名称为“LOG”的视图。

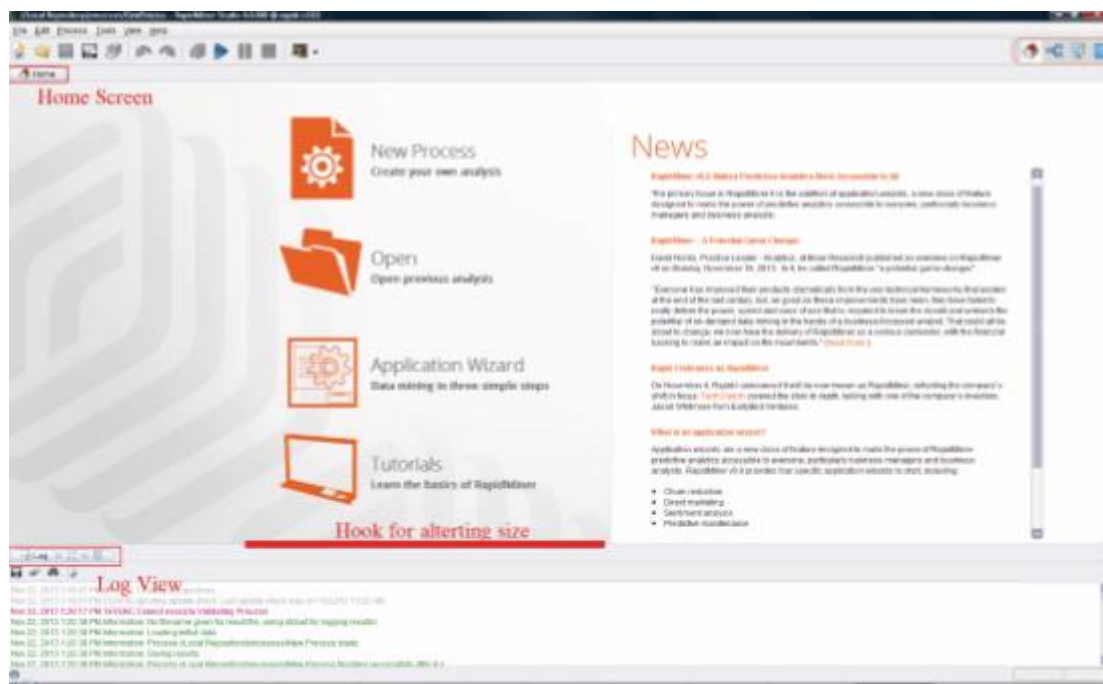


图 2.5: 视图间的大小变化

现在,您能看到之前的主页视图及在底部的“LOG”视图。如果您把鼠标移动到他呢不之间的区域,光标就会改变形状,这表示您现在能够通过拖动改变视图的大小。点击并按住鼠标键,移动光标,您可以试一下。

如之前提到过的,您也能根据自己的喜好改变视图的位置。只需把光标移动到视图标题

栏并把视图拖到其他位置，就能改变视图位置。在松开鼠标键后，将会安放该视图的位置会用透明的灰色来突出显示：

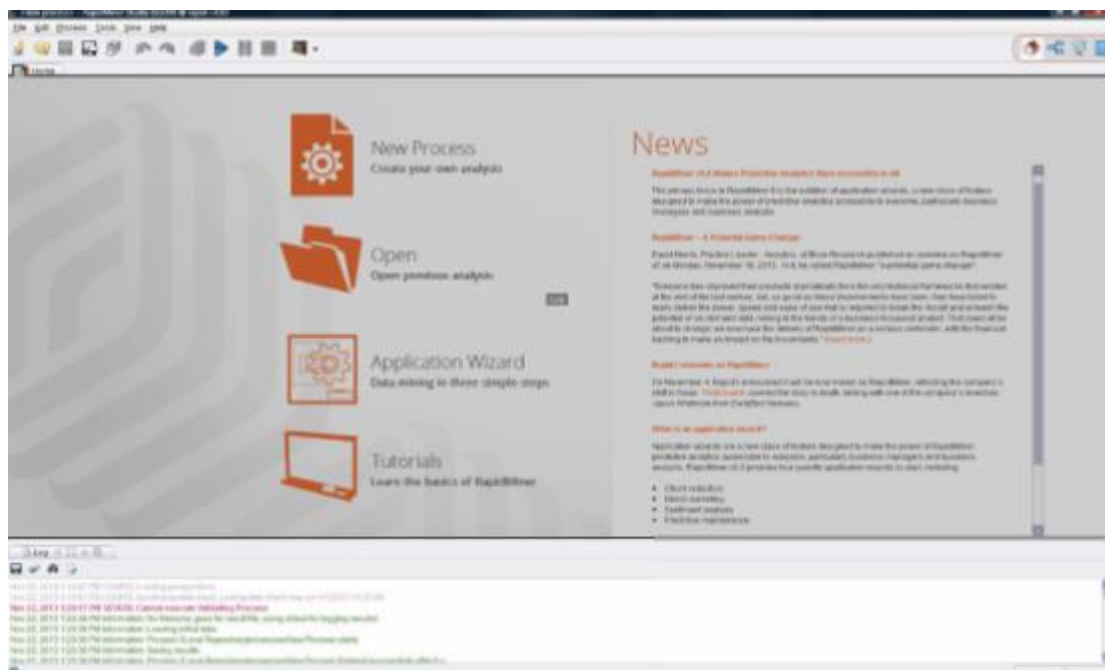


图 2.6：把底部的“LOG”视图拖到中间并突显新位置

您可以用这个方法组合视图，创建一些文件选项卡，即只有一个视图是可见的。或者，您也可以把“LOG”视图从下面拖动到右边区域，这样以后，区域划分就是纵向而不是横向的了。您甚至可以完全解除对一个视图的锁定并把它移动到 RapidMiner Studio 窗口之外。如果您想暂时全屏看一个视图，您可以最大化该视图并在之后再次最小化。您也可以通过右击视图标题栏来选择最大化该视图。每个视图带有“关闭、最大化、最小化及分离”操作选项，如图 2.7 所示。



图 2.7：视图操作选项

RapidMiner Studio 的所有视图还带有以下操作选项，其他操作选项都是自带说明的：

- 1.关闭：关闭当前透视图中的视图，通过菜单栏中的“视图”“打开视图”按钮，您可以在当前或者其他透视图重新打开该视图。
- 2.最大化：在当前透视图最大化视图。
- 3.最小化：在当前透视图最小化视图。视图被最小化到视图的左边，您可以再次最大化该视图或者在左边简要查看该视图。
- 4.分离：从当前透视图把视图分离出来单独显示，该单独显示的视图窗口能被移动到

任何地方。

现在我们稍微练习一下以不同的方式放置两个视图。有时候为了能够把视图准确地拖到想要的位置，我们需要进行一些练习。进行这些移动放置视图的练习是很有好处的，因为根据屏幕分辨率和个人喜好设定的界面可以提高您的工作效率。

有时候，您可能不小心删掉了一个视图或者无意间把一个透视图移动到了一个非常不好的位置。如果遇到这种情况，“视图”菜单能够帮您解决这个问题，因为除了能够通过“打开视图”按钮重新打开被关闭了的视图外，还能在任何时候通过视图菜单中的“Restore Default Perspective（恢复默认透视图）”来恢复初始状态。

此外，点击“新建透视图”按钮（图 2.4），您可以随意命名并保存自己的透视图。在“视图”菜单或者右边的工具栏，您可以切换已保存的和预先定义的透视图。

## 2.3 设计透视图

如一开始所提到的那样，您会在工具栏的右边发现对应每个（预先定义的）透视图都有一个图标：



图 2.8：工具栏中的透视图图标

这里展示的图标对应以下几个透视图：

**Home 主页透视图：**透视图的欢迎界面在上面已经提到过，在您启动流程后 RapidMiner 会打开欢迎界面。

**Design 设计透视图：**这是 RapidMiner Studio 透视图的中心，所有分析流程的创建、编辑和管理都在这里进行。

**Results 结果透视图：**在一个流程以数据、模型或者其它类似形式输出分析结果时，RapidMiner Studio 就会为您打开这个透视图。这里为您提供了统计、图表、高级图表及其它数据结果。

**Wizard 向导透视图：**这个透视图为您展示了用于典型数据挖掘问题的应用向导。

通过点击工具栏中的按钮或者在菜单中点击“视图” - “透视图”选项选择想要透视图，您可以切换到想要的透视图。如果切换到其他透视图是符合用户需求的话，RapidMiner Studio 最终会切换到其他透视图，譬如在完成分析流程以后切换到结果透视图。

现在，点击工具栏切换到设计透视图。在这章我们将详细介绍设计透视图。第四章的主题是结果透视图。现在请看图 2.9。

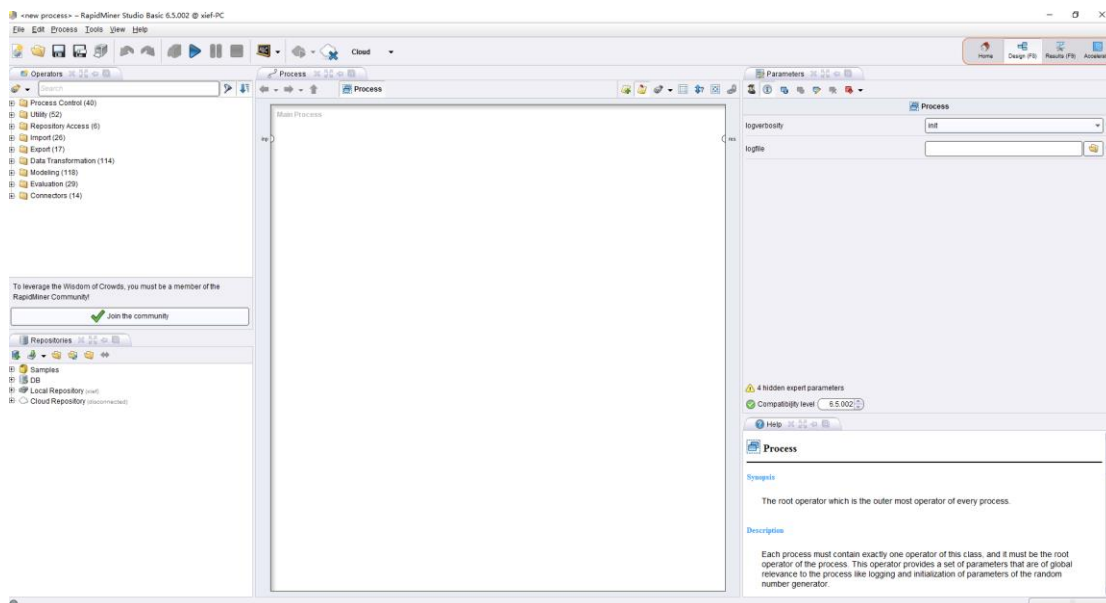


图 2.9: RapidMiner 的 Design（设计）透视图界面

因为 Design（设计）透视图是 RapidMiner Studio 的主要操作界面，下面我们将分别介绍设计透视图的所有部分及其关联视图的基本功能。

### 2.3.1 算子和资源库视图

在这个区域，至少在标准设置中，有两个非常重要的视图，详情如下。

#### Operators View 算子视图

RapidMiner Studio 中所有的操作步骤（算子）都以分组形式存储在这里，因此里面也能存储当前的流程。您可以以一个简单的方式操作不同组，浏览您想看的算子。如果 RapidMiner Studio 其中一个可用扩展被扩展了，那新增的算子也会被存储在这里。



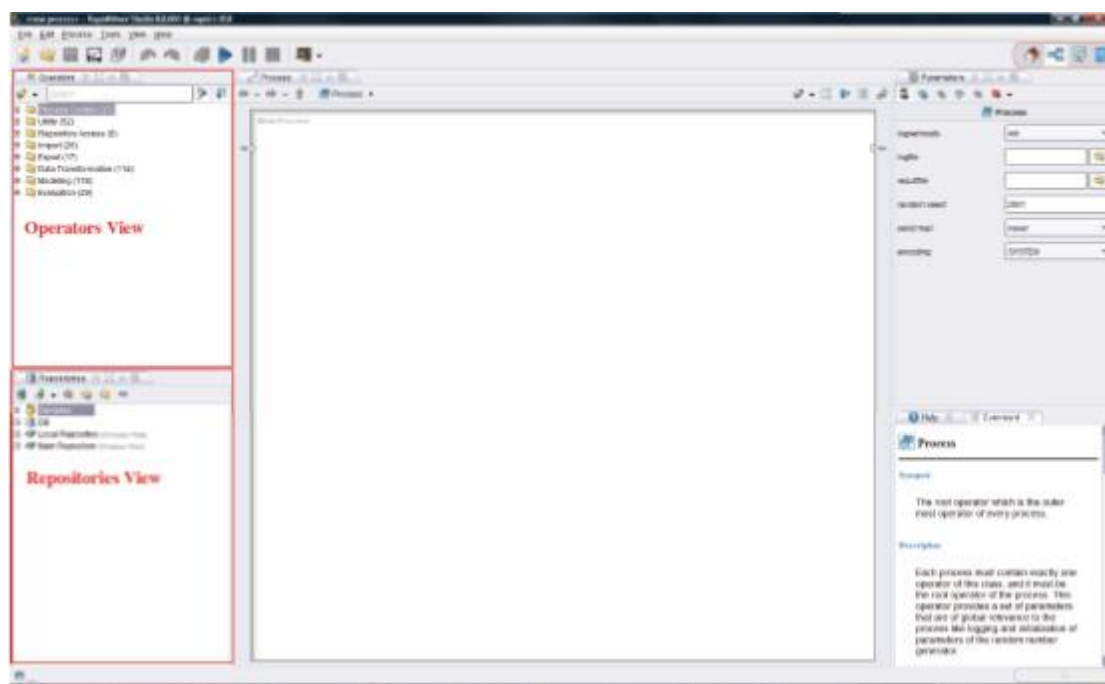


图 2.10: RapidMiner 的设计算子

在没有扩展的情况下，您至少会在树形结构中看到以下几个算子组。

**Process Control 流程控制：**循环或者条件分支等能控制流程的算子。

**Utility 辅助：**用于与“子流程”一起对子流程进行分组的辅助算子，也包含了重要的宏流程及日志流程。

**Repository Access 存储路径：**包含了资源库中的读写算子。

**Import 导入：**包含了大量的算子，用于读取文件、数据库等外部数据源中数据及对象。

**Export 导出：**包换了大量的算子，用于在文件、数据库等外部数据源中写入数据和对象。

**DataTransformation 数据转换：**就大小跟相关性而言，这可能是最重要的一组。所有的算子，包括数据跟元数据，都是在这里转换的。

**Modelloing 建模：**为了把生成的模型应用到新的数据集中，这个组包含了真正的数据挖掘流程，例如分类法，回归法、聚类、权重、关联规则法、相关性分析法和算子。

**Evaluation 评价：**这个群集中包含了可以计算模型质量的算子，如交叉验证、bootstrapping 等。

您可以在算子视图中选择算子并通过拖放把他们放置在理想位置。通过在流程中输出和输入端口画一条线来连接流程。在插入流程以后，您也可以选择让流程自动连接。在视图选择工具栏左边的插入符号（图 2.11）并定义自动发起的对外或者对内连接。



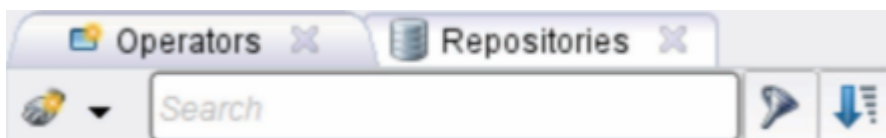


图 2.11：算子视图的操作及筛选

为了让操作尽量简便，算子视图也支持筛选功能，可能与搜索算子的部分名称或者全部名称。在筛选框内输入关键词，一旦所有搜索结果少于 10 个，树形结构就会被展开，显示全部搜索结果。这就意味着您无需每次都点开所有的文件夹。点击搜索后的红叉消除当前输入的关键词并关闭树形结构。

点击搜索框右边的图标能够筛选已被删除的算子并根据最常用的算子为它们排序。

提示：随着时间的推移，专业人员会越来越频繁的知道必要算子的名称。除了搜索（全部）名称，搜索框也支持首字母搜索（即所谓的驼峰式拼写法搜索）。试着输入“REx”来搜索“Read Excel”或者输入“DN”来搜索“Date to Nominal”和“Date to Numerical”-这大大加快了搜索速度。

## Repository View 资源库视图

资源库是 RapidMiner Studio 的一个核心组成部分，在第五章中我们会详细介绍。它被用于管理和使您项目中的分析流程结构化，同时它也是数据及相关元数据的一个来源。在接下来的一章我们会详细介绍如何使用资源库，现在让我们接着看下面的内容。

警告：因为 RapidMiner Studio 大多数流程支持使用元数据设计流程，我们强烈建议您使用 RapidMiner 资源库，否则（例如直接从文件或数据库中读取数据）元数据将无法使用，即 RapidMiner Studio 很多流程支持无法使用。

### 2.3.2 Process View 流程视图

流程视图（图 2.12）展示了分析流程及其相互连接的每一步。我们可以通过多种方法在现有的流程中新增步骤。我们能再次定义跟分离这些步骤间的连接。最终，我们甚至能在这个透视图中定义这些步骤的顺序。下文介绍了使用流程视图的方法。

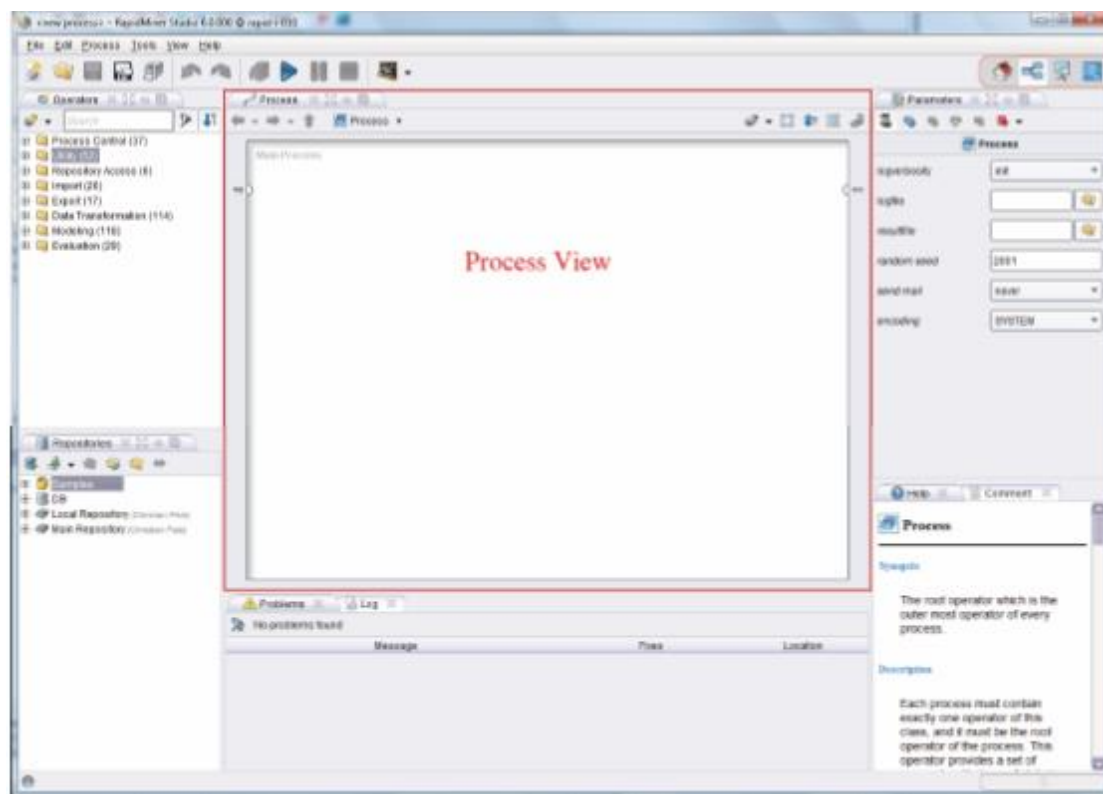


图 2.12：在流程视图中，连接 RapidMiner 的构成要素，即所谓的算子

### 2.3.3 Operators 和 Processes 算子和流程

从根本上说，RapidMiner Studio 的工作包括通过指明一系列的单个工作步骤定义分析流程。

在 RapidMiner Studio 中，这些流程构成要素被称为算子，一个算子是由很多东西进行定义的：

- 对预期输入的描述
- 对提供输出的描述
- 算子的输入动作，该动作最终会引起提供输出
- 若干能够控制动作的参数

算子的输入和输出都通过端口来产生或者消耗。端口需要特定种类的输入。在 RapidMiner Studio 中，我们会看到一个算子由如下一个模型被表示出来，其中输入端口在左边，输出端口在右边：

这样的算子能够从比如资源库、数据库或文件中输入数据。有时候，尽管可能会有一个至少指明数据位置的参数，但是这个算子可能会没有输入端口。其他的算子会转换输入数据并返回一个同样类型的对象。转换数据的算子属于这个组。其他算子会消耗输入数据并把它们转换成一个全新的对象：许多数据挖掘方法都属于这一个类别，比如会为给出的输入

数据提供一个模型。

端口的颜色指明了应该为该端口提供的的数据输入类型。举例来说，蓝色指需要一组例子。如果端口的上半部分及其标题是红色的，说明有问题。在图 2.13 的算子中，很容易看到这个问题：算子没有连接，输入端口需要与一个合适的来源连接起来。

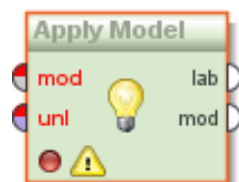


图 2.13：一个算子能通过（左边的）输入端口和（右边的）输出端口连接起来

如果结果不清楚或者在当前配置下无法提供结果，输出端口的颜色会是白色的。一旦所有的配置都齐了，即所有必须参数都定义完毕，所有必须输入端口都连接完毕，输出端口会根据输出的数据类型变成相应的颜色。

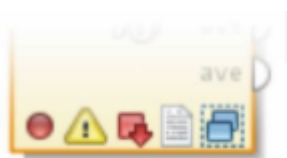


图 2.14：算子的状态指示符

不仅端口能通过不同的状态指示符表明他们的状态，整个算子（图 2.14）也能办到。这些指示符按照从左到右的顺序排列如下：

**状态指示灯：**红色指示灯说明有参数未被设置或输入端口未被连接等问题，黄色指示灯说明还未执行算子，不管配置是否基本齐全，绿色指示灯说明一切正常，已成功执行算子。

**三角警示牌：**用于表明是否有算子的状态信息出现。

**断点：**用于表明分析员是否在这个算子前面或是后面暂停了流程，以检测中间效果。

**注释：**如果这个算子中出现了注释，则会通过这个图标表示出来。

**子流程：**这是一个非常重要的指示，因为有些算子有一个或多个子流程。这个图标用来指示是否有子流程。您能双击算子，进入子流程。

如果一些算子被相互连接，我们称之为分析流程或者简称为流程。这一系列的步骤包括例如加载一组数据，转换数据，计算一个模型并把模型应用到其它数据组中等。这样一个流程在 RapidMiner Studio 中可能会像图 2.15 所示。

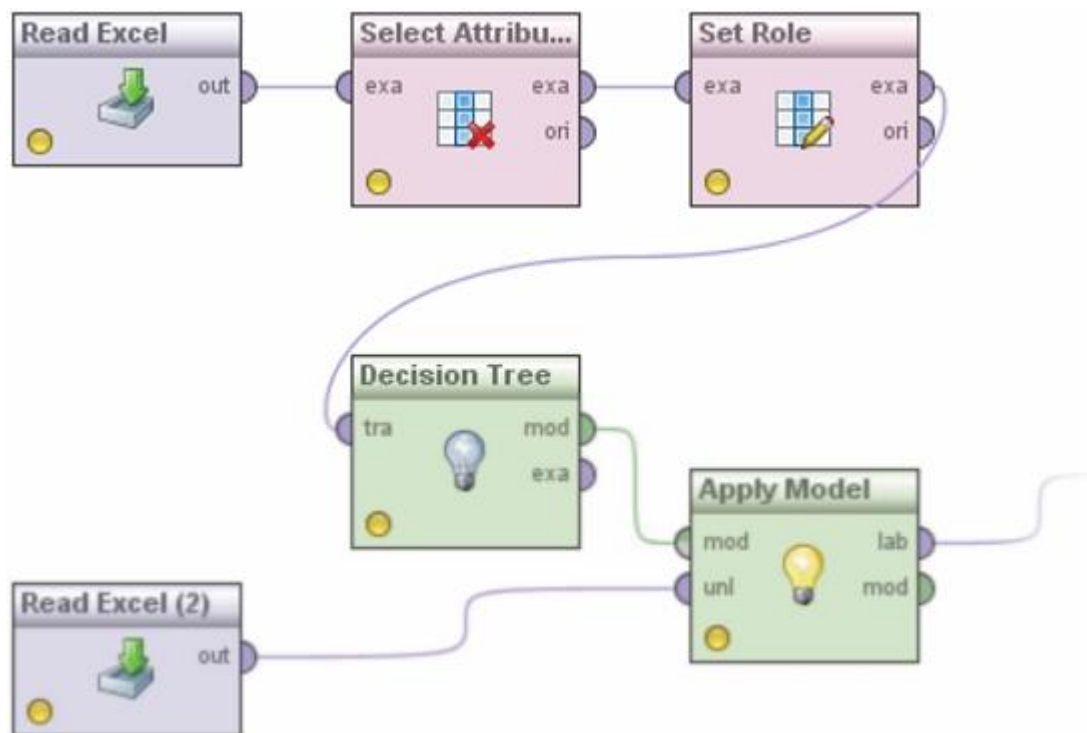


图 2.15：一个分析流程由几个算子组成。数据流的颜色指明了正在被传输的对象类型

在 RapidMiner Studio 中，这样一个流程很容易就能衍生出几百个算子，这些算子分布在几个不同层次的子流程中。流程检验在后台持续运行，同时流程导航帮助在下方显示以确保您不会迷失以及即使是在处理一些复杂任务时定义正确的流程。

## 插入算子

您可以用不同的方法在流程中插入新的算子。详细内容如下：

- 如上文所说，从算子视图中拖放
- 在算子视图中双击算子
- 通过菜单进入“编辑”“新建算子...”（或者直接按 Ctrl-I）打开对话框
- 在空白流程区域中的一个自由区域右击打开快捷菜单，点击子菜单“新建算子”新建算子

在不同情况下，新建的算子会根据算子视图的设置，自动连接至合适的算子，或需要用户手动进行或纠正算子连接。

## 连接算子

在您插入新算子以后，您可以相互连接插入的算子。有三种基本的连接方法，详细描述如下：

连接法 1：在插入时自动链接

在算子视图中，如果您选择了插入符号下的自动连接功能，在插入算子时，RapidMiner 会试着把算子连接到合适的输出端口。譬如，如果新的算子有一个输入端口，要求输入一组例子，那 RapidMiner 会试着找到一个已经能生成这样一组例子的算子，如果只有一个可选算子，RapidMiner 很容易选择和连接该算子。如果有多个可选算子，RapidMiner 会试着选择左边离鼠标现在的位置最近的一个算子。连接好的算子会被一个带有阴影的框给框起来，如图 2.16 所示。这样的话，您可以在插入算子期间，尽早的改正错误的算子连接。

提示：建议至少把输入端口的连接设置为自动连接。即使基于元数据的连接算法偶尔会创建一个错误的连接，但是正确的连接会被自动的识别出来，您仍然能后节省很多连接算子的时间。

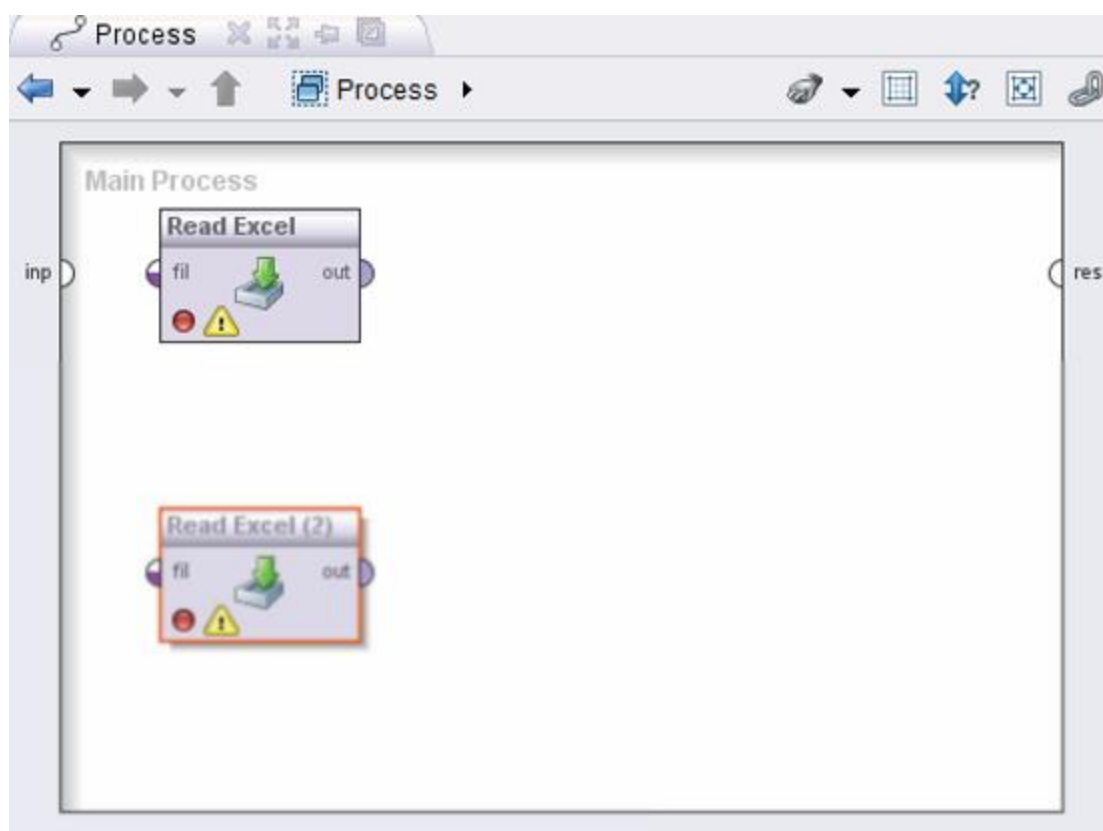


图 2.16：在拖动过程中，第二个算子被突出来（带框带阴影），如果这个时候有新算子被插入进来并需要一组例子，那第二个算子很有可能被连接到新算子上。

#### 连接法 2：手动

您还可以手动连接算子，对于比较复杂的流程来说，这很有必要。为此，先点击输出端口。然后画一条橙色的线，如图 2.17 所示。点击一个输入端口，把选择的输出端口连接到这个输入端口。如果要取消这一步，右击一下鼠标。橙色连接线就会消失，您能继续进行其它算法连接。

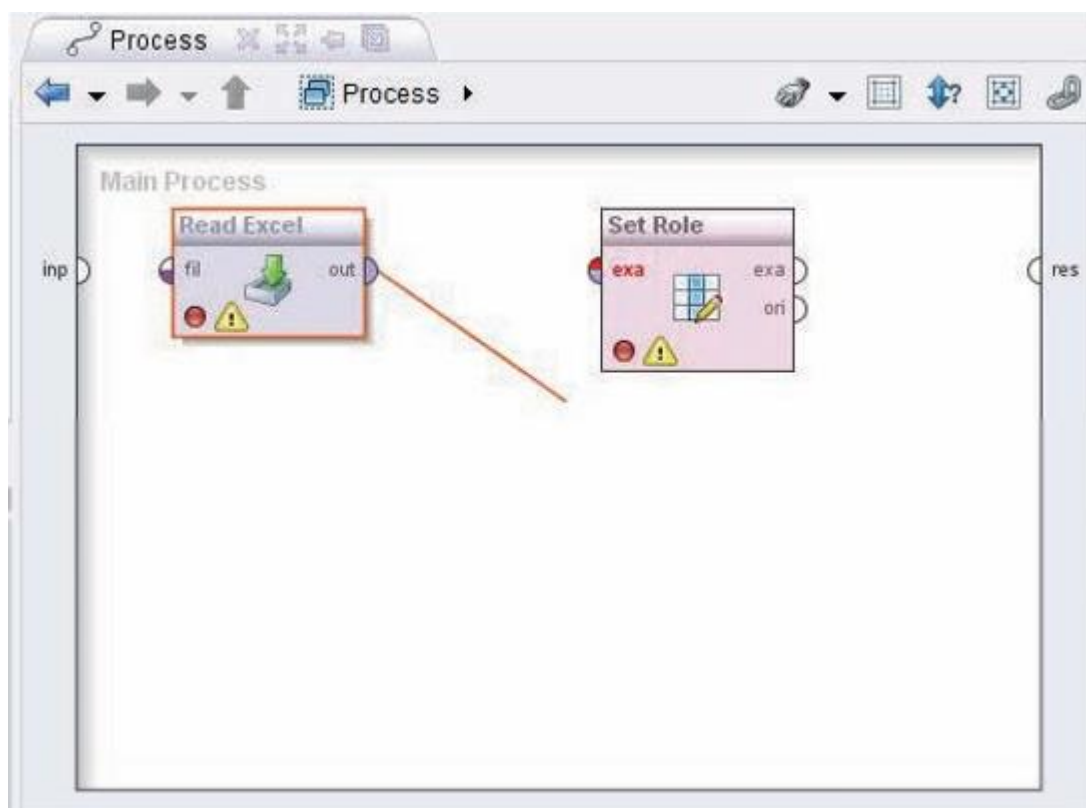


图 2.17：点击输出端口连接算子，右击取消连接。

### 连接法 3：全自动

有时候，大量的算子已经在流程（或子流程）中，但是还未连接起来。这种情况下，隐藏在流程视图插入符号下的“自动链接”跟“重新连接”功能会帮您的大忙。

如果采用一个相对有序的方法创建新的算子，所有算子被一个接一个的正确的排列好，即在插入新算子时前面一个算子始终都以一个带阴影的框标示出来，这个方法会更加有效。比较明智的做法是在所有自动连接完成以后，手动验证一下，因为可能会有意外的连接，尤其是在比较复杂的流程中。

## 选择算子

为了编辑参数，您必须选择一个算子。被您选中的算子会显示有橙色带阴影的边框。

如果您想同时对几个算子执行一个动作，譬如，移动或删除，青围绕这些算子拖出一个框以选择相关算子。

如果想多选或删除一个算子到当前已选的算子中，按住 **CTRL** 键并点击相关算子或者按住 **CTRL** 拖出另一个框。

## 移动算子

如上所述，选择一个或多个算子。现在把光标移到其中一个已选的算子上，按下并拖动

鼠标。

这时候，所有选中的算子都会移动到一个您鼠标所在的新位置。

如果在移动过程中，碰到了白色区域的边缘，白色区域会相应的自动放大。如果您碰到了可见区域的边缘，它们也会自动移动过去。

## 复制算子

如上所述，选择一个或多个算子。现在按下 **Ctrl+C** 键复制已选算子，按下 **Ctrl+V** 键复制它们。所有复制的算子会挨着原来的算子，您也可以移动它们。

## 删除算子

如上所述，选中一个或多个算子。现在您可以通过以下方法删除算子：

- 按“DELETE”键
- 右击所选算子的其中一个，在快捷菜单中选择“delete”
- 通过菜单中的“编辑”“删除”选项

## 删除连接

在按住 **ALT** 键的同时点击其中一个端口，就能删除连接。或者，您也可以通过相关端口的快捷菜单删除连接。

## 流程中的导航

我们看流程视图中的工具栏，我们会发现到目前为止，我们只使用了其中一项功能。在这部分，我们会介绍左边工具栏的另外四个功能：指向左边的箭头，指向右边的箭头，指向上面的箭头，导航栏（痕迹导航）。

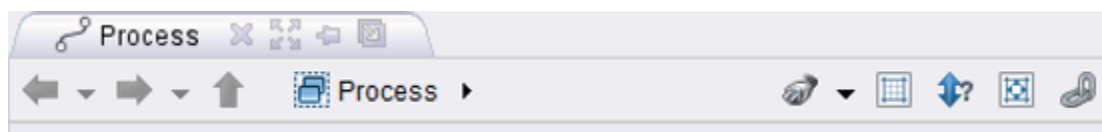


图 2.18：流程视图中的操作

每个操作的详细如下：

**指向左边的箭头：**返回编辑上一步，导航中操作很类似，这在因特网浏览器中很常见。通过快捷菜单也可以跳过其中几步。

**指向右边的箭头：**返回至最近一次历史编辑，导航中操作很类似，这在因特网浏览器中很常见。通过快捷菜单也可以跳过其中几步。

**指向上面的箭头：**离开当前子流程，回到上一级流程。



**导航栏:** 导航栏上显示了通过穿过所有层级从主流程到当前子流程的路径。一旦点击其中一个算子, 将会显示一个相关流程。您可以点击向右的小箭头进一步向下导航。

您可以双击一个右下方带有子流程图标的算子来进入子流程。如果想返回上一级，您可以使用箭头向上导航。导航栏（图 2.19）显示的当前路径也可以被用来导航至向上或向下一级。

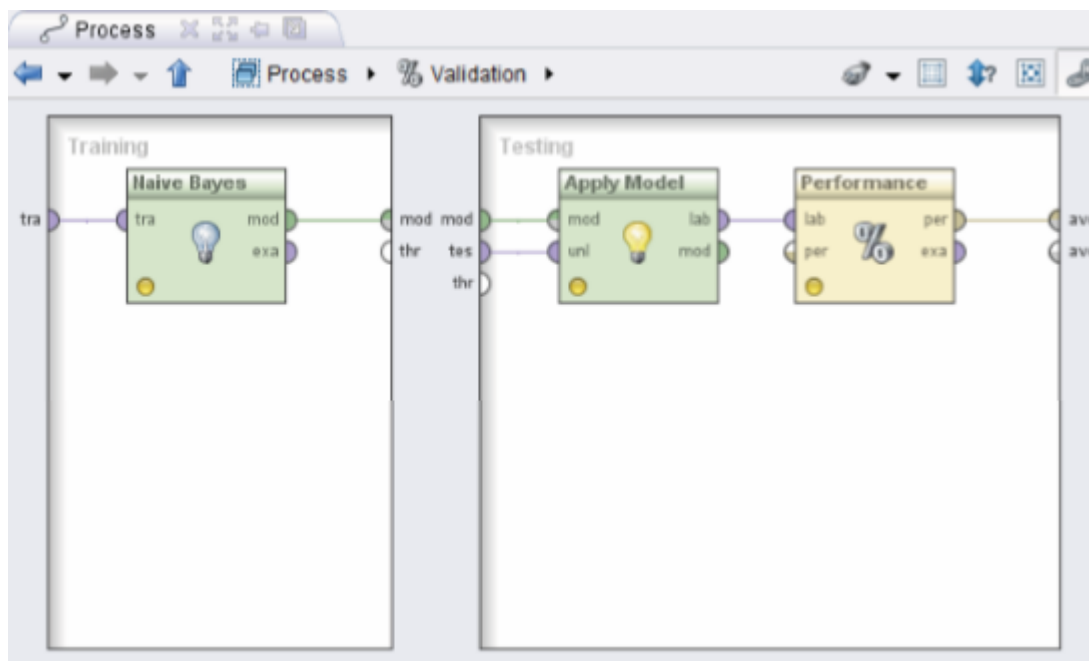


图 2.19: 一个标题为“有效期”的子流程, 可以通过点击向上的箭头或者导航栏离开该子流程

## 定义执行顺序

几乎在所有情况下，**RapidMiner** 都能自动成功确定正确的算子执行顺序。为了做到这一点，**RapidMiner** 会使用一个明显之前已被执行过的算子的连接信息和事实情况，该算子的执行结果会被用于另外的算子。

然而，有时候会出现算子顺序不能被自动定义的情况，譬如完全与子流程平行或者自动排列顺序不正确，原因可能为必须先计算一个宏，之后在把它当作一个参数用在之后的算子中。但是也有其他一些中重要原因，譬如该执行（报告，例子）需要一个更高效的数据处理或一个准确的顺序。

为了避免这种情况，RapidMiner 为显示算子程序，乃至便携地编辑执行顺序提供了一个绝妙的方法。请点击在流程视图的工具栏中的那个带有向上向下双箭头并带有问号的图标（图 2.18），这时，流程视图会显示算子顺序的定义。显示的不是每个算子的图标，而是算子的执行序号。橙色的透明线条以这种顺序连接算子，参见图 2.20。

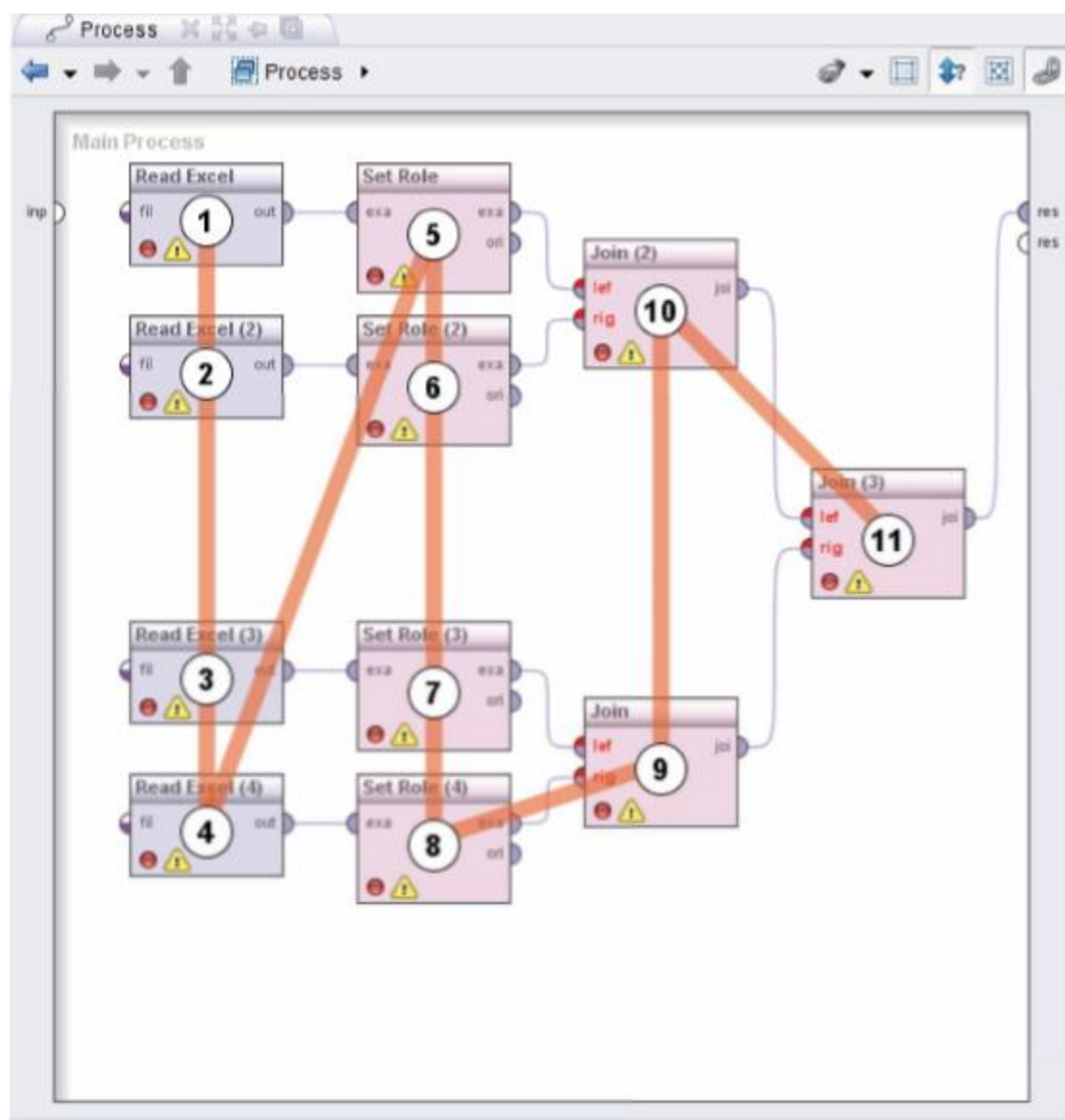


图 2.20: 执行顺序展示图。这个执行顺序不好，因为需要同时处理很多的数据组。

如果要改变这样一个执行顺序，点击一个算子的任何部位，以选中该算子。现在该算子的路径不能被改变，但是再次点击其他算子，就能改变这种一旦第一个算子被执行，第二个算子紧接着就会被执行的顺序。当您把鼠标移到剩下的算子上时，您会看到一直到这个算子，当前选择的算子都是橙色，从这个算子开始的其余算子为灰色。不能被选择的算子会以红色数字标示。右击可以取消当前的选择。您只需以这种方式点击几下，就可以改变上面所述的流程顺序（如图 2.21 所示）。

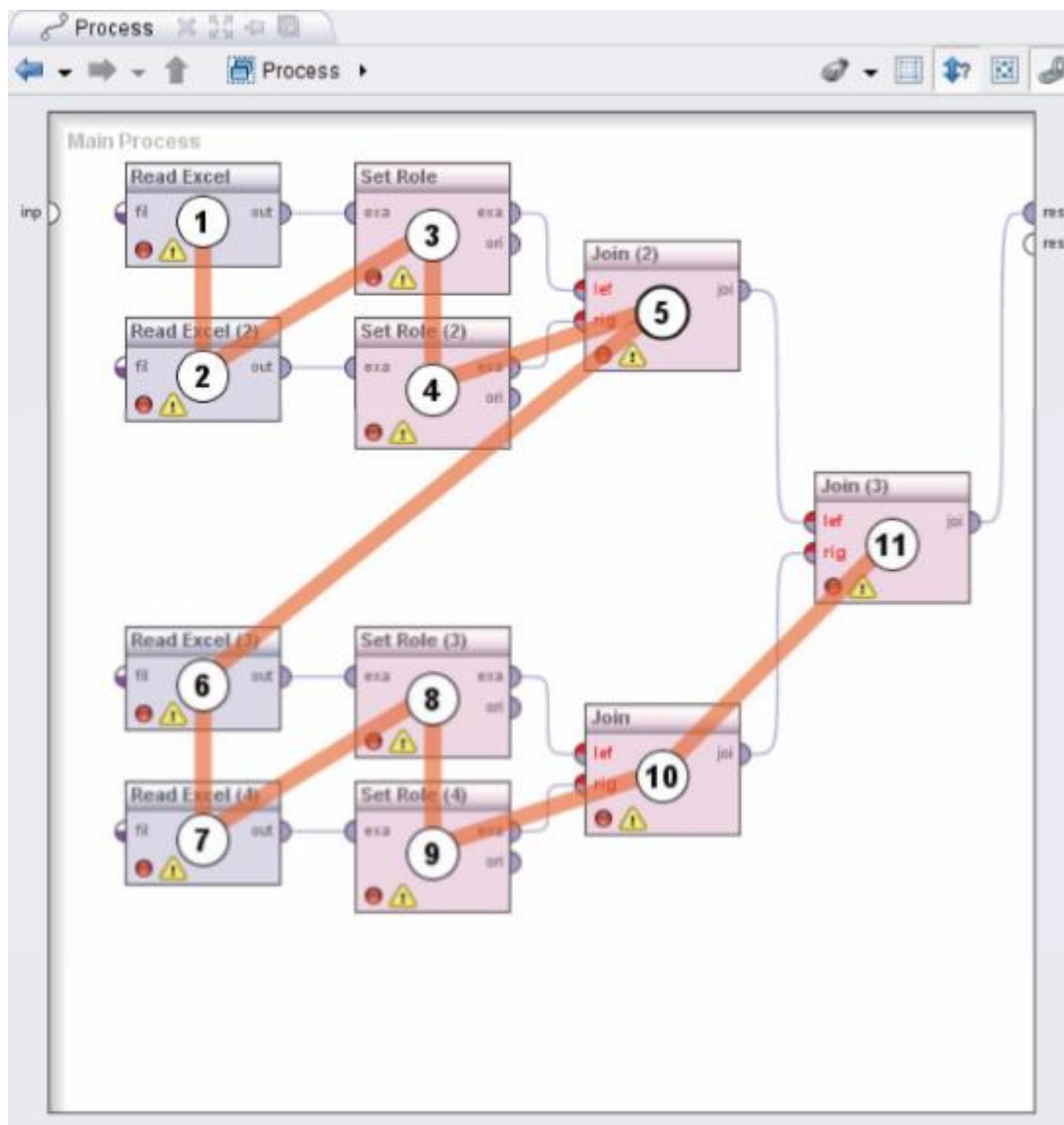


图 2.21: 做了一些改变之后的新顺序

### 2.3.4 流程视图的更多选项

在讨论了 RapidMiner 设计透视图的核心部分的几乎所有的选项以后，我们现在来介绍工具栏中剩余的选项，如图 2.18 所示，及如何在流程视图中实现这些选项。

流程视图工具栏右侧的 5 个图标有以下几个功能：

**自动连接及重新连接：** 使用插入符号可以自动连接及重新连接算子。

**自动排序：**根据连接及当前执行顺序重新排列所有算子。

**显示及改变执行顺序：**显示算子的执行顺序及改变顺序。

**自动调整大小:** 自动调整白色工作区域大小以使现有所有算子刚好有足够的空间。这个图标对于自动缩小（尺寸优化）特别实用。

**更新计划元数据：**如果点击该图标，在端口的计划元数据信息会更新以匹配算子执行后的实际数据。

另外，通过快捷菜单可以把流程另存为 PDF 或其他格式并把它打印出来。

### 2.3.5 Parameter View 参数视图

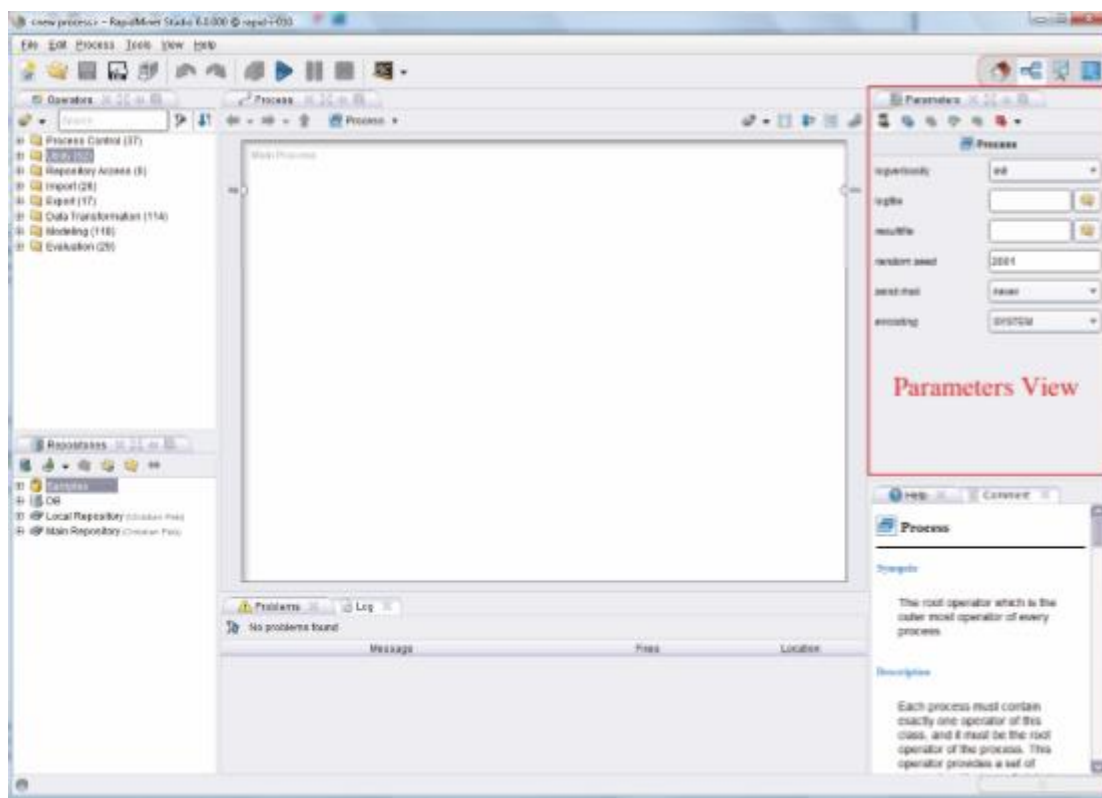


图 2.22：在参数视图中设置当前已选算子的参数

图 2.22 展示了 RapidMiner 的参数视图。大量算子为了发挥其正确的功能需要一些参数了指明。譬如，读取数据时需要指明文件路径。虽然通过定义特定参数值可以控制算子的执行，考虑到建模，也经常会优化参数值，但是，很多时候，参数不是绝对必须的。

在流程图中为算子选择了参数以后，它的参数会被参数视图中显示。像其他视图一样，参数视图也有自己的工具栏，详细描述如下。在工具栏下面，您会看到当前选择的算子的图标及其标题，跟在后面的是实际参数。粗体字指该参数必须被定义，它没有默认值。斜体字指该参数被划分为专业参数，一般情况下，数据分析初学者不要去更改它。图 2.23 展示了参数视图的详细内容。



图 2.23: 算子“Nominal to Date”的参数

请注意，一些参数只有在其他参数有了特定的值以后才会定义。例如，**Sample** 项选择 **absolute**，在选择 **absolute** 选项为抽样类型后，算子“**Sample（抽样）**”才能被指定一个绝对值。

正如参数一样，工具栏的动作指当前已选算子。

**专业模式：**左边的该图标用于切换专业者模式及初学者模式。只有专业者模式时，所有的参数才都会被显示出来，在初学者模式下，专业参数不会显示出来。

**算子信息：**展示了这个算子的一些基本信息，例如预期输入或描述等。通过流程视图中的快捷菜单选好算子以后按 **F1** 或者点击菜单栏上的“编辑”“显示算子信息...”也能跳出这个对话框。

**启用/禁用：**算子能被（暂时）禁止活动。算子的连接会被断开，然后它们就不会被执行了。被禁用的算子会以灰色显示。也可以通过流程视图中的快捷菜单或者点击菜单上“**Edit（编辑）**”“**Enable Operator（启用算子）**”来启用（或禁用）算子。

**重命名：**重命名算子的途径之一。其他途径如：选择算子以后按 **F2**；在流程视图中该算子的快捷菜单中选择“重命名”；或者通过菜单栏进入“**Edit（编辑）**”-“**Rename（重**

命名）”。

**删除：**删除算子的途径之一。其他途径如：选择算子以后按 DELETE；在流程图中该算子的快捷菜单中选择“删除”；或者通过菜单栏进入“Edit（编辑）”-“Delete（删除）”。

**断点开关：**在算子被执行以后，可以通过“Toggle Breakpoints（断点开关）”来设置断点，设置断点以后，执行流程被暂停，然后就能检测中间结果。也可以通过流程视图中该算子的快捷菜单或者菜单栏中的“Edit（编辑）”设置断点。按 F7 可以启用或者禁用断点。

## 2.3.6 Help & Comment View 帮助视图及注解视图

### Help View 帮助视图

每次在流程视图或算子视图中选择一个算子以后，帮助窗口中的帮助视图会显示关于这个算子的描述信息。

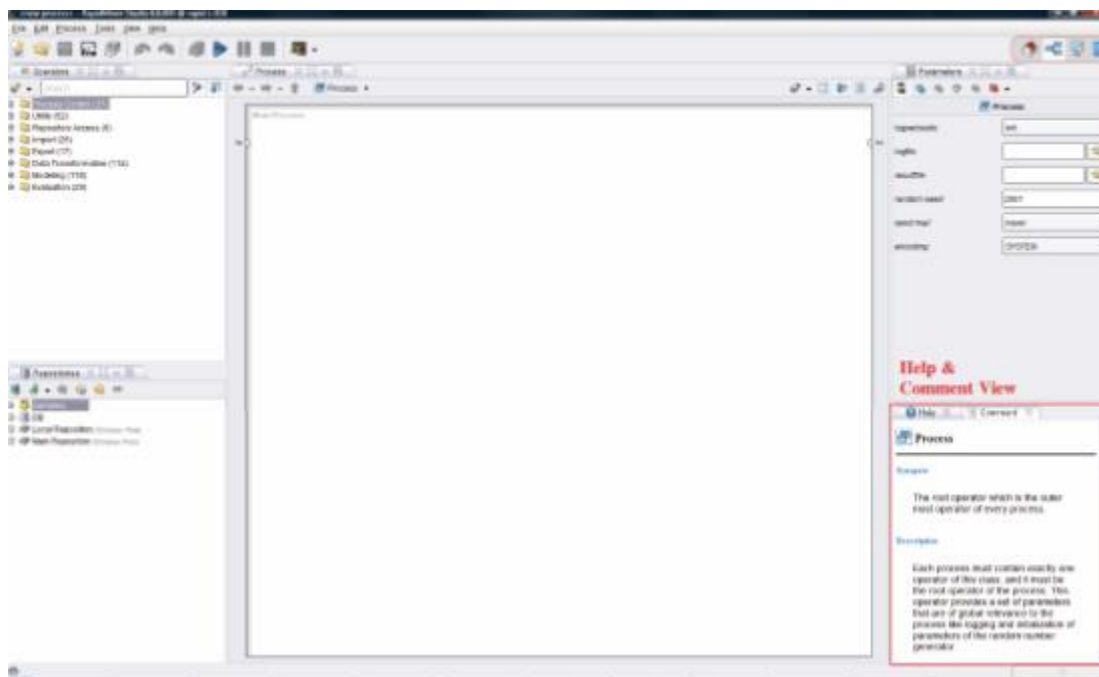


图 2.24：帮助文本会显示算子视图中或流程视图中当前已选算子的信息

描述信息包括：用一到两句话概括算子功能，算子功能的详细描述，包括参数的简短描述在内的参数清单，默认值（如果有的话），参数是否为专业的说明，参数相关性说明。

### Comment View 注解视图

与帮助视图不同，注解视图里的内容不是预先定义的描述，而是您个人对于流程的每一步的注释。仅仅是选择一个算子然后在注释区域写上任何内容。注释将跟流程定义一同保存起来，对于在后期设计时跟踪各个步骤很有帮助。如果某个算子有注释的话，在它的下边缘会有一个小小的文本图标。



### 2.3.7 Overview View 综述视图

特别是在扩展流程的时候，白色的工作区域会不够，然后我们会通过流程视图的快捷菜单或者按住 Ctrl 键然后向上下左右移动箭头或者简单地把一个算子拖到边缘来扩大工作区域。这个时候，我们同时看到整个工作区，流程里的导航会变得更困难。为了改善观感及方便导航，RapidMiner Studio 提供了一个综述视图（图 2.25），综述视图显示了整个工作区并把当前显示部分用一个小框突显出来。

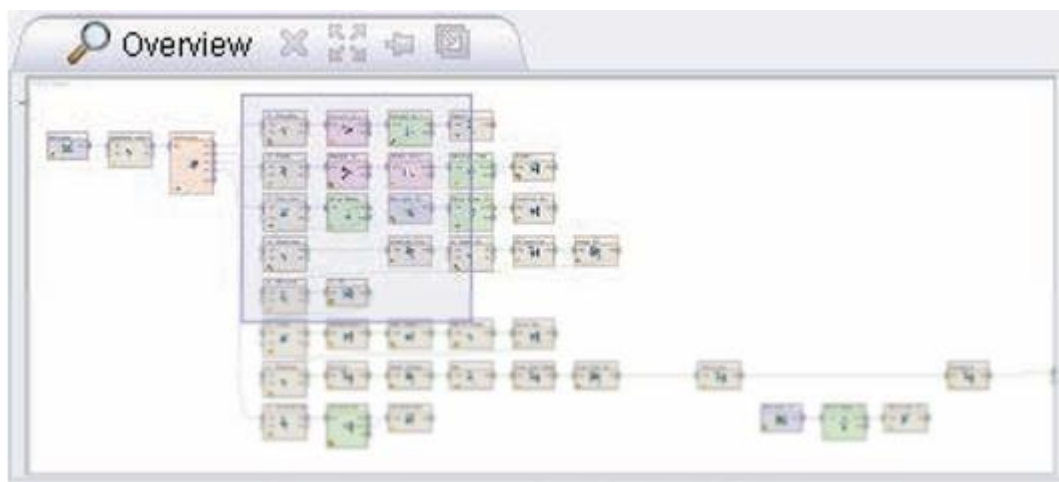


图 2.25：综述视图显示了整个流程并突出了可见部分

当在流程视图中转动滚动条或直接把一个算子拖到区域边缘的时候，您会看到显示部分会在流程视图中移动。或者您也可以直接拖动综述视图中的突显区域到想要的位置，流程视图会随之自动调整。

### 2.3.8 Problems & Log View 问题视图及日志视图

#### 问题视图

问题视图是一个更加核心的部分，能在设计分析流程时提供很有价值的帮助。任何警告及错误信息都会在这里被清楚地列明（图 2.26）。

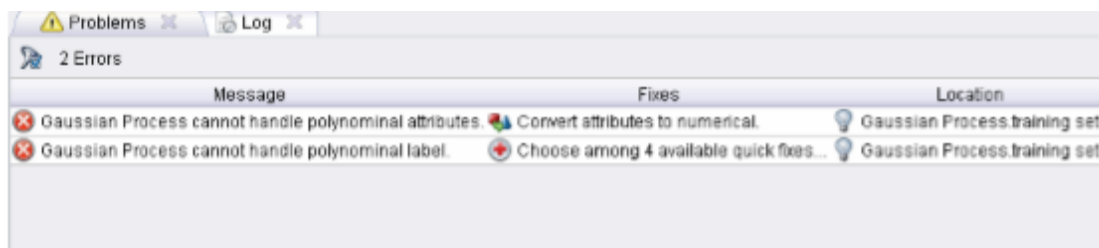


图 2.26：当前所有问题的描述

在标题为“信息”的第一列，您会看到问题的一个简短概括。这个时候，数据挖掘方法“高斯流程”不能处理多项式（多个值的分类）属性。最后一个标题为“Location（位置）”



的列以算子名称和相关输入端口名称的形式显示了问题出现的位置。问题视图工具栏左边的图标是一个过滤器，点击过滤器图标，问题视图中就只显示当前选择的算子的问题。这对于有很多问题的大型流程来说很实用。

问题视图也可能会有解决问题的建议，您可以执行该建议。这些解决方案被称为“Quick Fixes（快速修复）”。第二列简要的展示了几个可能的剷除方案，如果只有一个可能的解决方案时，会直接显示一个文本，如果有多个可能时，会指明有多少种不同的解决现有问题的可能。在上面的例子中，有两种处理第二个问题的可能。但是问什么这种解决问题的建议被叫做“快速修复”？这时，您双击一下表格中的相关的快速修复区域试试。第一行的解决方案会直接被执行，会自动配置和插入一个相关的算子，通过这样的方式，必要的预处理就完成了。双击第二行，会出现一个对话框，让您从多个解决方案中选择一个，之后会配置并插入一个或多个必要的算子，通过这样的方式，问题就被解决了。通过问题视图，您可以尽早发现问题，只需要稍微点几下就能在设计流程的过程中解决这些问题。

注意：RapidMiner Studio 的确定潜在问题并生成快速修复方案的功能是建立在有正确的元数据的基础上的。我们强烈建议您使用资源库，因为如果不使用资源库的话（例如直接从文件或数据库中读取数据），元数据将不能使用，这些功能也因此不能使用。

### 日志视图

在设计过程中，尤其是流程执行过程中，会同时写入大量信息并在出现错误时提供相关信息，例如如何修改流程设计以消除问题。



图 2.27：我们能在日志视图中看到补充信息，尤其是关于流程执行和发生错误时的

您可以跟往常一样在日志视图里复制文本，并把它运用到其他应用中。点击日志视图工具栏上的图标，您也可以保存或者删除所有这些内容或者在其中搜索相关内容。