

IBM SPSS Modeler 16
Python 脚本编制和自动化指南

IBM

注释

在使用本资料及其支持的产品之前，请阅读第 235 页的『声明』中的信息。

产品信息

本版本适用于 IBM SPSS Modeler V16R0M0 以及所有后续发行版和修订版，直到在新版本中另有声明为止。

目录

第 1 章 脚本编制	1	导入、替换和删除节点	28
脚本编写概述	1	遍历流中的节点	28
脚本类型	1	获取节点的相关信息	29
流脚本	1	第 4 章 脚本编制 API	31
独立脚本	3	脚本编制 API 简介	31
超节点脚本	3	示例: 使用定制过滤器搜索节点	31
流中的循环和条件执行	3	元数据: 关于数据的信息	31
流中的循环	4	访问已生成的对象	33
流中的条件执行	7	处理错误	35
执行和中断脚本	7	流、会话和超节点参数	35
查找和替换	8	全局值	39
第 2 章 脚本语言	11	使用多个流 - 独立脚本	39
脚本编写语言概述	11	第 5 章 脚本编写技巧	41
Python 和 Jython	11	修改流执行	41
Python 脚本编制	11	使用模型	41
运算	12	生成加密密码	41
列表	12	脚本检查	41
字符串	13	从命令行编写脚本	42
备注	14	指定文件路径	42
语句语法	14	与早期版本的兼容性	42
标识	15	第 6 章 命令行自变量	43
代码块	15	调用软件	43
将参数传递给脚本	15	命令行自变量的使用	43
示例	16	系统自变量	44
数学方法	16	参数自变量	45
使用非 ASCII 字符	18	服务器连接自变量	45
面向对象的程序设计	19	IBM SPSS Collaboration and Deployment Services Repository 连接自变量	46
定义类	19	组合多个参数	47
创建类实例	19	第 7 章 属性参考信息	49
向类实例添加属性	19	属性参考信息概述	49
定义类属性和方法	20	缩写	49
隐藏变量	20	节点和流属性示例	49
继承	20	节点属性概述	50
第 3 章 在 IBM SPSS Modeler 中进行		公共节点属性	50
脚本编制	23	第 8 章 流属性	51
脚本类型	23	第 9 章 源节点属性	55
流、超节点流和图	23	源节点公共属性	55
流	23	asimport 节点属性	56
超节点流	23	cognosimport 节点属性	57
图	23	database 节点属性	58
执行流	23	datacollectionimport 节点属性	59
脚本编制上下文	24	excelimport 节点属性	60
引用现有节点	24	evimport 节点属性	61
查找节点	25	fixedfile 节点属性	61
设置属性	25		
创建节点以及修改流	26		
创建节点	26		
链接和取消链接节点	27		

sasimport 节点属性	64
simgen 节点属性	64
statisticsimport 节点属性	66
userinput 节点属性	66
variablefile 节点属性	67
xmlimport 节点属性	70

第 10 章 记录操作节点属性 71

append 节点属性	71
aggregate 节点属性	71
balance 节点属性	72
derive_stb 节点属性	72
distinct 节点属性	74
merge 节点属性	74
rfmaggregate 节点属性	75
Rprocess 节点属性	76
sample 节点属性	76
select 节点属性	78
sort 节点属性	78
streamingts 节点属性	79

第 11 章 字段操作节点属性 83

anonymize 节点属性	83
autodataprep 节点属性	83
binning 节点属性	86
derive 节点属性	88
ensemble 节点属性	89
filler 节点属性	90
filter 节点属性	90
history 节点属性	91
partition 节点属性	91
reclassify 节点属性	92
reorder 节点属性	93
restructure 节点属性	93
rfmanalysis 节点属性	93
settoflag 节点属性	94
statistictransform 节点属性	95
timeintervals 节点属性	95
transpose 节点属性	99
type 节点属性	100

第 12 章 图形节点属性 105

图形节点公共属性	105
collection 节点属性	105
distribution 节点属性	106
evaluation 节点属性	107
graphboard 节点属性	108
histogram 节点属性	110
multiplot 节点属性	111
plot 节点属性	112
timeplot 节点属性	113
web 节点属性	114

第 13 章 建模节点属性 117

公共建模节点属性	117
anomalydetection 节点属性	117

apriori 节点属性	118
autoclassifier 节点属性	119
设置算法属性	121
autocluster 节点属性	121
autonumeric 节点属性	122
bayesnet 节点属性	123
buildr 节点属性	124
c50 节点属性	125
carma 节点属性	126
cart 节点属性	126
chaid 节点属性	128
coxreg 节点属性	129
decisionlist 节点属性	131
discriminant 节点属性	132
factor 节点属性	133
featureselection 节点属性	134
genlin 节点属性	136
glm 节点属性	139
kmeans 节点属性	142
knn 节点属性	143
kohonen 节点属性	144
linear 节点属性	145
logreg 节点属性	146
neuralnet 节点属性	149
neuralnetwork 节点属性	150
quest 节点属性	152
regression 节点属性	153
sequence 节点属性	154
slrm 节点属性	155
statisticsmodel 节点属性	156
svm 节点属性	156
timeseries 节点属性	157
twostep 节点属性	159

第 14 章 模型块节点属性 161

applyanomalydetection 节点属性	161
applyapriori 节点属性	161
applyautoclassifier 节点属性	162
applyautocluster 节点属性	162
applyautonumeric 节点属性	162
applybayesnet 节点属性	162
applyc50 节点属性	163
applycarma 节点属性	163
applycart 节点属性	163
applychaid 节点属性	164
applycoxreg 节点属性	164
applydecisionlist 节点属性	164
applydiscriminant 节点属性	165
applyfactor 节点属性	165
applyfeatureselection 节点属性	165
applygeneralizedlinear 节点属性	165
applyglm 节点属性	166
applykmeans 节点属性	166
applyknn 节点属性	166
applykohonen 节点属性	166
applylinear 节点属性	167

applylogreg 节点属性	167
applyneuralnet 节点属性	167
applyneuralnetwork 节点属性	168
applyquest 节点属性	168
applyregression 节点属性	168
applyr 节点属性	168
applyselflearning 节点属性	169
applysequence 节点属性	169
applysvm 节点属性	169
applytimeseries 节点属性	170
appltwostep 节点属性	170

第 15 章 数据库建模节点属性 171

Microsoft 建模的节点属性	171
Microsoft 建模节点属性	171
Microsoft 模型块属性	173
Oracle 建模的节点属性	174
Oracle 建模节点属性	174
Oracle 模型块属性	179
IBM DB2 建模节点属性	180
IBM DB2 建模节点属性	180
IBM DB2 模型块属性	185
IBM Netezza Analytics 建模节点属性	186
Netezza 建模节点属性	186
Netezza 模型块属性	194

第 16 章 输出节点属性 195

analysis 节点属性	195
dataaudit 节点属性	196
matrix 节点属性	197
means 节点属性	198
report 节点属性	199
Routput 节点属性	200
setglobals 节点属性	200
simeval 节点属性	201
simfit 节点属性	201
statistics 节点属性	202
statisticsoutput 节点属性	203
table 节点属性	203
transform 节点属性	205

第 17 章 导出节点属性 207

公共导出节点属性	207
asexport 节点属性	207
cognosexport 节点属性	208
databaseexport 节点属性	208
datacollectionexport 节点属性	211
excelexport 节点属性	211
outputfile 节点属性	212

sasexport 节点属性	213
statisticsexport 节点属性	213
xmlexport 节点属性	213

第 18 章 IBM SPSS Statistics 节点属性 215

statisticsimport 节点属性	215
statisticstransform 节点属性	215
statisticsmodel 节点属性	216
statisticsoutput 节点属性	216
statisticsexport 节点属性	216

第 19 章 超节点属性 219

附录 A. 节点名引用 221

模型块名称	221
避免重复的模型名称	223
输出类型名称	223

附录 B. 从旧脚本编制迁移到 Python 脚本编制 225

旧脚本迁移概述	225
一般差异	225
脚本编制上下文	225
命令与函数	225
文字和注释	226
运算符	226
条件语句和循环	227
变量	228
节点、输出和模型类型	228
属性名	228
节点引用	228
获取并设置属性	229
编辑流	229
节点操作	230
循环	230
执行流	231
通过文件系统和存储库访问对象	232
流操作	232
模型操作	233
文档输出操作	233
旧脚本编制与 Python 脚本编制之间的其他差异	233

声明 235

商标	236
--------------	-----

索引 237

第 1 章 脚本编制

脚本编写概述

IBM® SPSS® Modeler 中的脚本编写是用于在用户界面上实现过程自动化的强大工具。您使用鼠标或键盘进行的操作,借助脚本同样可以完成,而且使用脚本可以自动化那些手动执行将造成大量重复操作且高耗时的任务。

脚本的作用包括:

- 根据是否满足执行条件,限制在流中执行节点的顺序以及以连续方式执行节点。
- 创建用于在流中重复执行节点的循环。
- 指定通常包含用户交互的操作的自动执行顺序,例如您可以构建一个模型,然后对其进行测试。
- 设置需要实际用户交互的复杂过程,例如需要重复模型生成和测试的交叉验证步骤。
- 设置流操纵过程 - 例如,您可以提取一个模型训练流,运行它,然后自动生成相应的模型测试流。

本章提供流级脚本、独立脚本以及 IBM SPSS Modeler 用户界面超节点内脚本的高级说明和示例。有关脚本编写语言、语法和命令的更多信息,请参阅随后的各节。¹

注意: 您无法导入和运行在 IBM SPSS Modeler 中的 IBM SPSS Statistics 中创建脚本。

脚本类型

IBM SPSS Modeler 使用三种类型的脚本:

- **流脚本**存储为流属性然后和指定流一起保存和装入。例如,可以编写自动化训练和应用模型块流程的流脚本。您还可以指定何时执行特定流,脚本应代替流画布内容运行。
- **独立脚本**不与保存在外部文本文件中的所有特定流关联。例如,可以使用独立脚本同时操作多个流。
- **超节点脚本**存储为超节点流属性。超节点只在终端超节点中可用。您可以使用超节点脚本控制超节点内容的执行序列。对于非终端(源或过程)超节点,可以为超节点定义属性或定义这种超节点直接在流脚本中包含的节点。

流脚本

脚本可用于定制特定流中的操作并与该流一起保存。流脚本可用于指定某个流中终端节点的特定执行顺序。可以使用“流脚本”对话框来编辑与当前流一起保存的脚本。

从流属性对话框访问流“脚本”选项卡:

1. 从“工具”菜单中,选择:

 流属性 > 执行

2. 单击**执行**选项卡以处理当前流的脚本。

1. IBM SPSS Modeler 旧脚本语言仍然可与 IBM SPSS Modeler 16 配合使用。请参阅文档《IBM SPSS Modeler 16 脚本编制和自动化指南》以获取更多信息。请参阅第 225 页的附录 B,『从旧脚本编制迁移到 Python 脚本编制』以获取将现有 IBM SPSS Modeler 旧脚本映射到 Python 脚本的指导信息。

3. 选择“执行”方式：缺省值（可选脚本）。

使用“流脚本”对话框顶部的工具栏图标可以执行下列操作：

- 将先前存在的独立脚本的内容导入窗口中。
- 将脚本保存为文本文件。
- 打印脚本。
- 追加缺省脚本。
- 编辑脚本（撤销、剪切、复制、粘贴及其他常见的编辑功能）。
- 执行整个当前脚本。
- 执行某个脚本中的选定行。
- 在执行期间停止脚本。（只有在脚本处于运行状态的情况下，才会启用此图标。）
- 检查脚本的语法，如果发现任何错误，就将其显示在对话框的下部面板中复查。

另外，也可以指定当执行流时是否应运行此脚本。每当按照脚本的执行顺序执行流时，您可以选择**运行该脚本**来运行脚本。此设置为快速构建模型提供流一级的自动化。但是，缺省设置为在执行流的过程中忽略此脚本。即使选择选项**忽略此脚本**，也可以直接从此对话框运行脚本。

您也可以选择将脚本编制类型从 Python 脚本编制更改为旧脚本编制。

脚本编辑器提供了下列功能，这些功能有助于脚本编写：

- 语法突出显示；将突出显示关键字、文字值（例如字符串和数字）以及注释。
- 行编号。
- 块匹配；当光标处于程序块的开始位置时，还将突出显示相应的结束块。
- 建议的自动补全。

可以使用 IBM SPSS Modeler 显示首选项来定制语法突出显示器使用的颜色和文本样式。通过选择**工具 > 选项 > 用户选项**，然后单击**语法**选项卡，您可以访问显示首选项。

通过从上下文菜单中选择**自动建议**或者按 **Ctrl + Space**，可以访问建议语法补全的列表。使用光标键在列表中上下移动，然后按 **Enter** 键可插入所选文本。按 **Esc** 可退出自动建议方式而不修改现有文本。

调试选项卡显示调试消息，并且可以用于在执行脚本后立即对脚本状态进行评估。**调试**选项卡包含一个只读文本区域和单行输入文本字段。文本区域显示由脚本发送到标准输出（例如，通过 `Python print` 命令）或标准错误（例如，通过错误消息文本）的文本。输入文本字段将接收来自用户的输入。然后，将在对话框内最近执行的脚本上下文（称为**脚本编制上下文**）中对此输入进行评估。文本区域包含命令和生成的输出，以便用户能够查看命令跟踪。输入文本字段始终包含命令提示符（对于 Python 脚本编制，此命令提示符为 `>>>`）。

在下列情况下，将创建新的脚本编制上下文：

- 脚本是使用“运行此脚本”按钮或“运行所选行”按钮执行的。
- 脚本语言会发生更改。

如果创建了新的脚本编制上下文，那么将清除文本区域。

注：在脚本面板外部执行流将不会修改此脚本面板的脚本上下文。在脚本对话框中，将无法查看该执行过程中创建的任何变量的值。

独立脚本

“独立脚本”对话框用于创建或编辑保存为文本文件的脚本。它显示了文件名称，提供了用于装入、保存、导入和执行脚本的实用程序。

要访问“独立脚本”对话框，请执行以下操作：

在主菜单中，选择：

工具 > 独立脚本

对流脚本可用的工具栏和脚本语法检查选项对独立脚本同样适用。有关更多信息，请参阅第 1 页的『流脚本』主题。

超节点脚本

通过使用 IBM SPSS Modeler 脚本语言，可以创建和保存所有终端超节点中的脚本。这些脚本只在终端超节点中可用，并且常在创建模板流或用于强制超节点内容以特定顺序执行时使用。使用超节点脚本，您也可以在流中运行多个脚本。

例如，假设需要指定一个复杂流的执行顺序，并且超节点包含若干个包括“设置全局值”节点的节点，而执行“设置全局值”节点又需要在派生用于散点图节点的新字段之前进行。这种情况下，可以创建一个首先执行“设置全局值”节点的超节点脚本。由“设置全局值”节点计算出的值，例如平均差或标准差，可在散点图节点的执行过程中使用。

在超节点脚本中也可以指定节点属性，操作方法与在其他脚本中的进行的操作一样。另外，为所有超节点或直接来自流脚本的超节点的封装节点更改和定义属性。有关更多信息，请参阅第 219 页的第 19 章，『超节点属性』主题。此方法适用于源和过程超节点以及终端超节点。

注意：因为只有终端超节点能够执行自身脚本，所以“超节点”对话框的“脚本”选项卡只在用于终端超节点时可用。

从主画布打开“超节点脚本”对话框：

从流画布选择终端超节点，然后从“超节点”菜单选择：

超节点脚本...

从放大超节点画布打开“超节点脚本”对话框：

右键单击超节点画布，然后从上下文菜单中选择：

超节点脚本...

流中的循环和条件执行

从 V16.0 开始，通过 SPSS Modeler，您可以选择各个对话框中的值在流中创建一些基本脚本，而无需使用脚本编制语言直接编写指令。可通过此方式创建的两种主要类型的脚本是简单循环以及在满足条件时执行节点的方式。

可以组合流中的循环规则和条件执行规则。例如，您可能具有来自世界各地制造商的汽车销售相关数据。您可以在流中设置一个用于处理数据的循环，从而按制造国家或地区标识详细信息，并将数据输出到各个显示了详

细信息（例如，按型号排列的销售量，按制造商和引擎大小排列的排放级别等）的图形。如果您希望仅分析欧洲信息，那么还可以向循环添加条件，以阻止针对总部设在美国和亚洲的制造商创建图形。

注：由于循环和条件执行均以后台脚本为基础，因此它们仅适用于运行的整个流。

- **循环** 使用循环可自动化重复任务。例如，这可能意味着向流添加给定数目的节点，并且每次更改一个节点参数。另外，您可以将流或分支的运行控制为反复运行给定数目的次数，如下示例所示：
 - 运行流给定数目的次数，并且每次都对源进行更改。
 - 运行流给定数目的次数，并且每次都对变量的值进行更改。
 - 运行流给定数目的次数，并且在每次执行时都输入一个额外的字段。
 - 构建模型给定数目的次数，并且每次都对模型设置进行更改。
- **条件执行** 您可以使用它根据预定义的条件来控制终端节点的运行方式，可能的示例如下：
 - 根据给定值是 `true` 还是 `false`，控制是否将运行节点。
 - 定义节点循环将以并行方式运行还是按顺序运行。

循环和条件执行都是在“流属性”对话框中的“执行”选项卡中设置的。任何在条件或循环要求中使用的节点都随附加到这些节点的附加符号一起显示在流画布上，此符号用于指示这些节点将参与循环和条件执行。

您可以通过下列三种方式中的其中一种来访问“执行”选项卡：

- 使用主对话框顶部的菜单：
 1. 从“工具”菜单中，选择：
流属性 > 执行
 2. 单击“执行”选项卡以处理当前流的脚本。
- 从流中：
 1. 右键单击节点，然后选择**循环/条件执行**。
 2. 选择相关子菜单项。
- 从主对话框顶部的图形工具栏中，单击流属性图标。

如果这是您第一次设置循环或条件执行详细信息，请在“执行”选项卡上选择**循环/条件执行**执行方式，然后选择**条件**或**循环**子选项卡。

流中的循环

通过循环，您可以自动化流中的重复任务；可能的示例如下：

- 运行流给定数目的次数，并且每次都对源进行更改。
- 运行流给定数目的次数，并且每次都对变量的值进行更改。
- 运行流给定数目的次数，并且在每次执行时都输入一个额外的字段。
- 构建模型给定数目的次数，并且每次都对模型设置进行更改。

可以在流“执行”选项卡的**循环**子选项卡上设置要满足的条件。要显示该子选项卡，请选择**循环/条件执行**执行方式。

如果设置了**循环/条件执行**执行方式，那么在您运行流时，您定义的所有循环要求都将生效。（可选）您可以针对您的循环要求生成脚本代码，并通过单击“循环”子选项卡右下角的**粘贴...**将此代码粘贴到脚本编辑器中；主要“执行”选项卡将显示此更改以显示**缺省（可选脚本）**执行方式，并将脚本显示在此选项卡的顶部。这意味着，您可以先使用多个循环对话框选项来定义循环结构，然后再生成可在脚本编辑器中进行进一步定制脚本。请注意，当您单击**粘贴...**时，您定义的所有条件执行要求也会显示在生成的脚本中。

要设置循环，请完成下列步骤：

1. 创建迭代关键字以定义将在流中执行的主要循环结构。有关更多信息，请参阅创建迭代关键字。
2. 在需要时，定义一个或多个迭代变量。有关更多信息，请参阅创建迭代变量。
3. 您创建的迭代和所有变量都将显示在该子选项卡的主要部分中。缺省情况下，将按显示顺序执行迭代；要在列表中上下移动迭代，请单击迭代以将其选中，然后使用该子选项卡右侧的向上或向下箭头更改顺序。

创建用于流中的循环的迭代关键字

使用迭代关键字可以定义将在流中执行的主要循环结构。例如，如果要对汽车销售进行分析，那么可以创建流参数**制造国家或地区**，并将其用作迭代关键字；在运行流时，此关键字将设置为各个迭代过程中您的数据中的各个不同的国家或地区值。使用“定义迭代关键字”对话框可以设置关键字。

要打开此对话框，请选择“循环”子选项卡左下角的**迭代关键字...**按钮，或者右键单击流中的任何节点，然后选择**循环/条件执行 > 定义迭代关键字（字段）**或**循环/条件执行 > 定义迭代关键字（值）**。如果是从流中打开此对话框，那么系统可能会自动为您填写一些字段，例如，节点的名称。

要设置迭代关键字，请填写下列字段：

迭代依据。您可以选择下列其中一个选项：

- **流参数 - 字段。**使用此选项可创建一个循环，用于将现有流参数的值依次设置为各个指定字段。
- **流参数 - 值。**使用此选项可创建一个循环，用于将现有流参数的值依次设置为各个指定值。
- **节点属性 - 字段。**使用此选项可创建一个循环，用于将节点属性的值依次设置为各个指定字段。
- **节点属性 - 值。**使用此选项可创建一个循环，用于将节点属性的值依次设置为各个指定值。

设置内容。选择将在每次执行循环时设置其值的项。您可以选择下列其中一个选项：

- **参数。**仅当您选择**流参数 - 字段**或**流参数 - 值**时才可用。从可用列表中选择所需参数。
- **节点。**仅当您选择**节点属性 - 字段**或**节点属性 - 值**时才可用。选择要对其设置循环的节点。单击浏览按钮以打开“选择节点”对话框并选择所需节点；如果列出的节点过多，那么可以通过选择下列其中一个类别对显示结果进行过滤，以仅显示特定类型的节点：“源”、“进程”、“图形”、“建模”、“输出”、“导出”或“应用模型”节点。
- **属性。**仅当您选择**节点属性 - 字段**或**节点属性 - 值**时才可用。从可用列表中选择节点的属性。

要使用的字段。仅当您选择**流参数 - 字段**或**节点属性 - 字段**时才可用。选择节点中要用于提供迭代值的字段。您可以选择下列其中一个选项：

- **节点。**仅当您选择**流参数 - 字段**时才可用。选择要对其设置循环且包含详细信息的节点。单击浏览按钮以打开“选择节点”对话框并选择所需节点；如果列出的节点过多，那么可以通过选择下列其中一个类别对显示结果进行过滤，以仅显示特定类型的节点：“源”、“进程”、“图形”、“建模”、“输出”、“导出”或“应用模型”节点。
- **字段列表。**单击右边列中的列表按钮可显示“选择字段”对话框，您可以在该对话框中选择节点中要用于提供迭代数据的字段。请参阅第 6 页的『选择用于迭代的字段』以获取更多信息。

要使用的值。仅当您选择**流参数 - 值**或**节点属性 - 值**时才可用。选择所选字段内要用作迭代值的值。您可以选择下列其中一个选项：

- **节点。**仅当您选择**流参数 - 值**时才可用。选择要对其设置循环且包含详细信息的节点。单击浏览按钮以打开“选择节点”对话框并选择所需节点；如果列出的节点过多，那么可以通过选择下列其中一个类别对显示结果进行过滤，以仅显示特定类型的节点：“源”、“进程”、“图形”、“建模”、“输出”、“导出”或“应用模型”节点。

- **字段列表。** 选择节点中用于提供迭代数据的字段。
- **值列表。** 单击右边列中的列表按钮可显示“选择值”对话框，您可以在此对话框中选择节点中要用于提供迭代数据的字段。

创建用于流中的循环的迭代变量

您可以使用迭代变量在每次执行循环时更改流中的流参数值或选定节点的属性值。例如，如果流循环将对汽车销售数据进行分析并使用**制造国家或地区**作为迭代关键字，那么您可能会具有一个按型号显示销售额的图形输出，以及另一显示了废气排放信息的图形输出。在这些情况下，您可以创建迭代变量，这些变量将为生成的图形创建新标题，例如**瑞典汽车排放和按型号排列的日本汽车销售额**。使用“定义迭代变量”对话框可以设置任何您需要的变量。

要打开此对话框，请选择“循环”子选项卡左下角的**迭代变量...** 按钮，或者右键单击流中的任何节点，然后选择：**循环/条件执行 > 定义迭代变量。**

要设置迭代变量，请填写下列字段：

更改。 选择要修改的属性的类型。可以从**流参数**或**节点属性**中进行选择。

- 如果选择**流参数**，请选择所需参数，然后通过循环的各个迭代，使用下列其中一个选项（如果在流中可用）定义应该将该参数设置为的值。
 - **全局变量。** 选择应该将流参数设置为的全局变量。
 - **表输出单元。** 要将流参数设置为表输出单元中的值，请从列表中选择表，然后输入要使用的行和列。
 - **手动输入。** 如果要手动为此参数输入将在各个迭代中采用的值，请选择此选项。返回到“循环”子选项卡时，将创建一个可在其中输入所需文本的新列。
- 如果选择**节点属性**，请选择所需节点以及该节点的其中一个属性，然后选择要用于该属性的值。通过使用下列其中一个选项，可以设置新属性值：
 - **单独。** 属性值将使用迭代关键字值。请参阅第 5 页的『创建用于流中的循环的迭代关键字』以获取更多信息。
 - **作为资源前缀。** 使用迭代关键字值作为在**资源**字段中输入的内容的前缀。
 - **作为资源后缀。** 使用迭代关键字值作为在**资源**字段中输入的内容的后缀。

如果选择前缀或后缀选项，那么系统将提示您向**资源**字段添加附加文本。例如，如果迭代关键字值为**制造国家或地区**并且您选择**作为资源前缀**，那么可以在此字段中输入 - **按型号排列的销售额**。

选择用于迭代的字段

创建迭代时，您可以使用“选择字段”对话框选择一个或多个字段。

排序方式。 可以通过选择下列其中一个选项可用于查看的字段进行排序：

- **自然。** 数据流向下遍历数据时，当前节点接收字段的顺序即为字段的查看顺序。
- **名称。** 使用字母顺序对要查看的字段进行排序。
- **类型。** 查看按其测量级别进行排序的字段。此选项在选择具有特定测量级别的字段时非常有用。

一次从列表中选择一个字段，或采用按住 **Shift** 并单击和按住 **Ctrl** 并单击的方法选择多个字段。此外，也可以使用列表下面的按钮根据测量级别选择多组字段，或选择或取消选择表中所有字段。

请注意，可供选择的字段将进行过滤，以仅显示适用于您使用的流参数或节点属性的字段。例如，如果您使用的是存储类型为字符串的流参数，那么将仅显示存储类型为字符串的字段。

流中的条件执行


通过条件执行，您可以根据与您所定义的条件相匹配的流内容来控制终端节点的运行方式；可能的示例如下：

- 根据给定值是 true 还是 false，控制是否将运行节点。
- 定义节点循环将以并行方式运行还是按顺序运行。

可以在流“执行”选项卡的**条件**子选项卡上设置要满足的条件。要显示该子选项卡，请选择**循环/条件执行**执行方式。

如果设置了**循环/条件执行**执行方式，那么在您运行流时，您定义的所有条件执行要求都将生效。（可选）您可以针对您的条件执行要求生成脚本代码，并通过单击“条件”子选项卡右下角的**粘贴...**将此代码粘贴到脚本编辑器中；主要“执行”选项卡将显示此更改以显示**缺省（可选脚本）**执行方式，并将脚本显示在此选项卡的顶部。这意味着，您可以先使用多个循环对话框选项来定义条件，然后再生成可在脚本编辑器中进行进一步定制的脚本。请注意，当您单击**粘贴...**时，您定义的所有循环要求也会显示在生成的脚本中。

要设置条件，请完成下列步骤：

1. 在“条件”子选项卡的右侧，单击“添加执行语句”按钮 ，以打开“条件执行语句”对话框。在此对话框中，可以指定执行节点所必须满足的条件。
2. 在“条件执行语句”对话框中，指定以下内容：
 - a. **节点**。选择要对其设置条件执行的节点。单击浏览按钮以打开“选择节点”对话框并选择所需节点；如果列出的节点过多，那么可以对显示结果进行过滤，以按下列其中一个类别显示节点：“导出”、“图形”、“建模”或“输出”节点。
 - b. **作为依据的条件**。指定执行节点所必须满足的条件。您可以从下列四个选项中选择其中一个：**流参数**、**全局变量**、**表输出单元**或**始终满足**。在对话框下半部分中输入的详细信息由您选择的条件控制。
 - **流参数**从提供的列表中选择参数，然后选择该参数的**运算符**；例如，运算符可以是大于、等于、小于和介于之间等等。然后，输入**值**或最小值和最大值，具体取决于运算符。
 - **全局变量**。从提供的列表中选择变量；例如，这可能包括平均值、总和、最小值、最大值或标准差。然后，选择**运算符**及所需值。
 - **表输出单元**。从可用列表中选择表节点，然后选择表中的**行和列**。然后，选择**运算符**及所需值。
 - **始终满足**。如果必须始终执行节点，请选择此选项。选择此选项后，将无需选择其他参数。
3. 重复步骤 1 和 2 所需次数，直到您设置了所有需要的条件。所选节点和执行该节点前所必须满足的条件将分别显示在该子选项卡主要部分中的**执行节点**和**如果满足此条件**列中。
4. 缺省情况下，将按显示顺序执行节点和条件；要在列表中上下移动节点和条件，请单击节点或条件以将其选中，然后使用该子选项卡右侧的向上或向下箭头更改顺序。

另外，您可以在“条件”子选项卡的底部设置下列选项：

- **按顺序对所有条件进行求值**。选择此选项可按条件在该子选项卡上的显示顺序对各个条件进行求值。对所有条件进行求值后，将立即执行那些条件求值为“true”的节点。
- **一次执行一个节点**。只有选中**按顺序对所有条件进行求值**时才可用。选中此选项表示，如果某个条件求值为“true”，那么将先执行与该条件关联的节点，然后再对下一个条件进行求值。
- **在首次命中之前进行求值**。选中此选项表示，将仅运行第一个根据您的条件返回“true”求值的节点。

执行和中断脚本

可以通过多种方法来执行脚本。例如，在流脚本或独立脚本对话框中，“运行此脚本”按钮将执行整个脚本：



图 1. “运行此脚本”按钮

“运行选定的行”按钮用于执行您在脚本中选择的单一行或者相邻行所组成的块:



图 2. “运行选定的行”按钮

可以使用以下方式执行脚本:

- 在流脚本或独立脚本对话框中，单击“运行此脚本”或“运行选定的行”按钮。
- 在**运行此脚本**设置为缺省执行方式的情况下运行流。
- 启动后以交互模式使用 `-execute` 标志。有关更多信息，请参阅第 43 页的『命令行自变量的使用』主题。

注意: 如果在“超节点脚本”对话框中选择**运行此脚本**，那么将在执行超节点时执行超节点脚本。

中断脚本执行

“流脚本”对话框中的红色“停止”按钮将在脚本执行过程中被激活。使用此按钮可以放弃脚本和任何当前流的执行。

查找和替换

可在编辑脚本或表达式文本的位置（包括脚本编辑器）或定义“报告”节点中的模板时使用“查找/替换”对话框。在上述任一区域编辑文本时，按 `Ctrl+F` 可访问此对话框，并确保光标的焦点位于文本区域中。例如，处理填充节点时，可以通过“设置”选项卡的任一文本区域或表达式构建器中的文本字段访问此对话框。

1. 使用文本区域中的光标，按 `Ctrl+F` 可访问“查找/替换”对话框。
2. 输入要搜索的文本，或从最近搜索项下拉列表中选择。
3. 输入替换文本（如果有的话）。
4. 单击**查找下一个**开始搜索。
5. 单击**替换**替换当前选定的内容，或单击**全部替换**更新所有项或选定的实例。
6. 每次操作完成后，此对话框将关闭。从任一文本区域中按 `F3` 键，可重复上一次查找操作，或按 `Ctrl+F`，可再次访问该对话框。

搜索选项

区分大小写。指定查找操作是否区分大小写；例如 `myvar` 是否与 `myVar` 匹配。无论怎样设置，替换文本始终完全按照输入插入。

仅限于整个单词。指定查找操作是否匹配单词中内嵌的文本。如果选中，`spider` 的搜索结果将不会包括 `spiderman` 或 `spider-man`。

正则表达式。指定是否使用正则表达式语法（请参阅下一节）。如果选中，**仅限于整个单词**选项将禁用并且会忽略其值。

仅限于选择文本。控制使用**全部替换**选项时的搜索范围。

正则表达式语法

使用正则表达式，您可以搜索特殊字符（如选项卡或换行符）、字符的类或范围（如 *a* 到 *d*）、任何数字或非数字以及边界（如行首或行尾）。正则表达式模式描述了该表达式将尝试在输入字符串中查找的字符串的结构。受支持的正则表达式构造的类型如下。

表 1. 字符匹配

字符	匹配
x	字符 x
\\	反斜杠字符
\\On	含八进制值的字符 On (0 <= n <= 7)
\\Onn	含八进制值的字符 Onn (0 <= n <= 7)
\\Omnn	含八进制值的字符 Omnn (0 <= m <= 3, 0 <= n <= 7)
\\xhh	含十六进制值的字符 Oxhh
\\uhhhh	含十六进制值的字符 Oxhhhh
\\t	制表符 ('\\u0009')
\\n	换行符 ('\\u000A')
\\r	回车符 ('\\u000D')
\\f	换页符 ('\\u000C')
\\a	警报（蜂鸣）符 ('\\u0007')
\\e	转义符 ('\\u001B')
\\cx	x 对应的控制字符

表 2. 匹配字符类

字符类	匹配
[abc]	a、b、或 c（简单类）
[^abc]	除 a、b、或 c 之外的所有字符（相减）
[a-zA-Z]	a 到 z 或 A 到 Z，包含（范围）
[a-d[m-p]]	a 到 d 或 m 到 p（合并）。也可指定为 [a-dm-p]
[a-z&&[def]]	a 到 z 和 d、e、或 f（交集）
[a-z&&[^bc]]	a 到 z，除 b 和 c 外（相减）。也可指定为 [ad-z]
[a-z&&[^m-p]]	a 到 z，而非 m 到 p（相减）。也可指定为 [a-lq-z]

表 3. 预定义字符类

预定义字符类	匹配
.	任意字符（可能或不可能与行终止符匹配）
\\d	任意数字: [0-9]
\\D	非数字: [^0-9]
\\s	空格字符: [\\t\\n\\x0B\\f\\r]
\\S	非空格字符: [^\\s]
\\w	单词字符: [a-zA-Z_0-9]
\\W	非单词字符: [^\\w]

表 4. 边界匹配

边界匹配符	匹配
^	行首
\$	行尾
\b	单词边界
\B	非单词边界
\A	输入的开头
\Z	除最后终止符外（如果有），输入的结尾
\z	输入的结尾

有关使用正则表达式的更多信息以及一些示例，请参阅 <http://www.ibm.com/developerworks/java/tutorials/j-introtojava2/section9.html>。

示例

以下代码将在字符串开头搜索并匹配 3 个数字：

```
^[0-9]{3}
```

以下代码将在字符串结尾搜索并匹配 3 个数字：

```
[0-9]{3}$
```

第 2 章 脚本语言

脚本编写语言概述

通过 IBM SPSS Modeler 的脚本编制工具，您可以创建一些脚本，这些脚本可以在 SPSS Modeler 用户界面上运行、处理输出对象并运行命令语法。您可以在 SPSS Modeler 中直接运行这些脚本。

IBM SPSS Modeler 中的脚本以脚本语言 Python 编写。IBM SPSS Modeler 所使用的基于 Java 的 Python 实现称为 Jython。脚本语言包含下列功能部件：

- 用于引用节点、流、工程、输出和其他 IBM SPSS Modeler 对象的格式。
- 可用于处理这些对象的一组脚本编制语句或命令。
- 用于设置变量、参数和其他对象的值的脚本编制表达式语言。
- 注释、连接符和文字文本块的支持。

以下各节描述了 Python 脚本语言、Python 的 Jython 实现以及在 IBM SPSS Modeler 内进行脚本编制的入门基本语法。具体属性和命令的有关信息则在随后的章节中提供。

Python 和 Jython

Jython 是 Python 脚本语言的实现，它以 Java 语言编写并与 Java 平台进行集成。Python 是一种面向对象的功能强大的脚本语言。Jython 具有成熟脚本语言的生产力特征，而且与 Python 不同的是，Jython 可以在任何支持 Java 虚拟机 (JVM) 的环境中运行。这意味着您在编写程序时可以使用 JVM 上的 Java 库。通过 Jython，您可以利用此差异并使用 Python 语言的语法和大部分功能

作为一种脚本语言，Python（及其 Jython 实现）易于学习并能够高效地进行编码，而且具备创建运行程序所需的最小结构。可以在交互方式下输入代码，即一次输入一行。Python 是一种解释性脚本语言；它没有 Java 中的预编译步骤。Python 程序仅仅是文本文件，系统将在输入这些文件时对其进行解释（在解析语法错误后）。简单表达式（例如已定义的值）以及更加复杂的操作（例如函数定义）将立即执行并可供使用。可以快速测试任何对代码进行的更改。但是，脚本解释确实存在一些缺点。例如，由于使用未定义的变量不是编译器错误，因此只有在执行使用了该变量的语句的情况下，才会检测到此错误。在这种情况下，可以编辑并运行程序以调试错误。

Python 将所有内容（包括所有数据和代码）视为对象。因此，您可以使用多行代码来处理这些对象。某些选择类型（例如数字和字符串）将被更加方便地视为值而不是对象；Python 支持此行为。有一个受支持的 Null 值。此 Null 值具有保留名称 None。

有关 Python 和 Jython 脚本编制的更深入介绍以及一些示例脚本，请参阅 www.ibm.com/developerworks/java/tutorials/j-jython1 和 www.ibm.com/developerworks/java/tutorials/j-jython2。

Python 脚本编制

本 Python 脚本语言指南介绍了在 IBM SPSS Modeler 中编制脚本时最可能使用的组件，其中包括概念和编程基础。这将为您的提供足够的知识来开发自己的 Python 脚本，以便在 IBM SPSS Modeler 中使用。

运算

赋值通过使用等号 (=) 来完成。例如，要将值“3”赋值给名为“x”的变量，您可以使用以下语句：

```
x = 3
```

等号还可用于将字符串类型数据赋值给变量。例如，要将值“a string value”赋值给变量“y”，您可以使用以下语句：

```
y = "a string value"
```

下表列出了一些常用的比较运算和数值运算及其描述。

表 5. 常用的比较运算和数值运算

运算	描述
$x < y$	x 是否小于 y?
$x > y$	x 是否大于 y?
$x \leq y$	x 是否小于或等于 y?
$x \geq y$	x 是否大于或等于 y?
$x == y$	x 是否等于 y?
$x != y$	x 是否不等于 y?
$x <> y$	x 是否不等于 y?
$x + y$	将 y 与 x 相加
$x - y$	从 x 中减去 y
$x * y$	将 x 乘以 y
x / y	将 x 除以 y
$x ** y$	求 x 的 y 次幂

列表

列表是元素序列。列表可以包含任意数目的元素，而列表的元素可以是任何类型的对象。也可以将列表视为阵列。随着添加、除去或替换元素，列表中元素的数目可能会增加或减少。

示例

<code>[]</code>	任何空列表。
<code>[1]</code>	包含单个元素（整数）的列表。
<code>["Mike", 10, "Don", 20]</code>	包含 4 个元素（两个字符串元素和两个整数元素）的列表。
<code>[[], [7], [8, 9]]</code>	列表的列表。每个子列表都是一个空列表或整数元素列表。
<code>x = 7; y = 2; z = 3;</code> <code>[1, x, y, x + y]</code>	整数列表。此示例说明了变量和表达式的使用。

您可以向变量分配列表，例如：

```
mylist1 = ["one", "two", "three"]
```

然后，可以访问列表的特定元素，例如：

```
mylist[0]
```

这将生成以下输出：

one

方括号 ([]) 中的数字称为索引, 它指向列表中的某个特定元素。将从 0 开始对列表中的元素编制索引。

您也可以选择列表中的一系列元素; 这称为切割。例如, `x[1:3]` 将选择 `x` 的第 2 个和第 3 个元素。结尾索引是所选内容后面的一个索引。

字符串

字符串是一个被视为值的不可变字符序列。字符串支持所有生成新字符串的不可变序列函数和运算符。例如, `"abcdef"[1:4]` 将生成输出 `"bcd"`。

在 Python 中, 字符由长度为 1 的字符串表示。

字符串字面值通过使用单重引用或三重引用来定义。使用单引号定义的字符串不能跨行, 而使用三重引号定义的字符串可以跨行。可以将字符串括在单引号 (') 或双引号 (") 中。引用字符可以包含其他未转义的引用字符或已转义 (即, 前面带有反斜杠 (\) 字符) 的引用字符。

示例

```
"This is a string"
'This is also a string'
'It's a string"
'This book is called "Python Scripting and Automation Guide".'
"This is an escape quote (\") in a quoted string"
```

Python 解析器将自动合并多个以空格分隔的字符串。这样您可以更轻松地输入长字符串, 并且更容易混合单个字符串中的引号类型, 例如:

```
"This string uses ' and " 'that string uses ".'
```

这将生成以下输出:

```
This string uses ' and that string uses ".
```

字符串支持一些有用的方法。下表列出了其中一些方法。

表 6. 字符串方法

方法	用法
<code>s.capitalize()</code>	对 <code>s</code> 执行首字母大写
<code>s.count(ss {,start {,end}})</code>	计算 <code>ss</code> 在 <code>s[start:end]</code> 中的出现次数
<code>s.startswith(str {, start {, end}})</code> <code>s.endswith(str {, start {, end}})</code>	测试以查看 <code>s</code> 是否以 <code>str</code> 开头 测试以查看 <code>s</code> 是否以 <code>str</code> 结尾
<code>s.expandtabs({size})</code>	将制表符替换为空格, 缺省 <code>size</code> 为 8
<code>s.find(str {, start {, end}})</code> <code>s.rfind(str {, start {, end}})</code>	在 <code>s</code> 中查找 <code>str</code> 的第一个索引; 如果找不到, 那么结果为 -1。 <code>rfind</code> 从右到左进行搜索。
<code>s.index(str {, start {, end}})</code> <code>s.rindex(str {, start {, end}})</code>	在 <code>s</code> 中查找 <code>str</code> 的第一个索引; 如果找不到, 那么将引起 <code>ValueError</code> 。 <code>rindex</code> 从右到左进行搜索。
<code>s.isalnum</code>	测试以查看字符串是否为字母数字字符串
<code>s.isalpha</code>	测试以查看字符串是否为字母字符串
<code>s.isnum</code>	测试以查看字符串是否为数字字符串
<code>s.isupper</code>	测试以查看字符串是否为全部大写
<code>s.islower</code>	测试以查看字符串是否为全部小写

表 6. 字符串方法 (续)

方法	用法
<code>s.isspace</code>	测试以查看字符串是否全是空格
<code>s.istitle</code>	测试以查看字符串是否为首字母大写的字母数字字符串序列
<code>s.lower()</code> <code>s.upper()</code> <code>s.swapcase()</code> <code>s.title()</code>	转换为全部小写 转换为全部大写 转换为大小写颠倒 转换为全标题形式
<code>s.join(seq)</code>	将 <code>seq</code> 中的字符串连接起来, 以 <code>s</code> 作为分隔符
<code>s.splitlines({keep})</code>	将 <code>s</code> 分割为多行, 如果 <code>keep</code> 为 <code>true</code> , 那么将使用换行
<code>s.split({sep [, max]})</code>	使用 <code>sep</code> (缺省 <code>sep</code> 为空格) 将 <code>s</code> 分割为“字”, 最多分割 <code>max</code> 次数
<code>s.ljust(width)</code> <code>s.rjust(width)</code> <code>s.center(width)</code> <code>s.zfill(width)</code>	在宽度为 <code>width</code> 的字段中, 将字符串左对齐 在宽度为 <code>width</code> 的字段中, 将字符串右对齐 在宽度为 <code>width</code> 的字段中, 将字符串居中对齐 用 0 进行填充。
<code>s.lstrip()</code> <code>s.rstrip()</code> <code>s.strip()</code>	除去前导空格 除去尾部空格 除去前导和尾部空格
<code>s.translate(str {, delc})</code>	除去 <code>delc</code> 中的所有字符后, 使用表转换 <code>s</code> . <code>str</code> 应该是长度为 <code>= 256</code> 的字符串。
<code>s.replace(old, new [, max])</code>	使用字符串 <code>new</code> 全部替换或按照 <code>max</code> 出现次数替换字符串 <code>old</code>

备注

备注是由井号 (或散列符号) (#) 引入的注释。同一行上位于井号后面的所有文本都将被视为备注的组成部分, 并且将被忽略。备注可以开始于任何列。以下示例说明了备注的使用:

```
#The HelloWorld application is one of the most simple
print 'Hello World' # print the Hello World line
```

语句语法

Python 的语句语法非常简单。通常, 每个源代码行都是单一语句。除 `expression` 和 `assignment` 语句外, 每个语句都由一个关键字名称 (例如 `if` 或 `for`) 引入。可以在代码中任何语句之间的任意位置插入空白行或备注行。如果某一行中有多个语句, 那么必须使用分号 (;) 来分隔每个语句。

超长语句可以分为多行。在这种情况下, 要分到下一行的语句必须以反斜杠 (\) 结尾, 例如:

```
x = "A loooooooooooooooooooooooooong string" + \
    "another loooooooooooooooooooooooooong string"
```

如果某个结构括在圆括号 (())、方括号 ([]) 或花括号 ({}) 内, 那么语句可以在任何逗号后面分为新行, 而不必插入反斜杠, 例如:

```
x = (1, 2, 3, "hello",
    "goodbye", 4, 5, 6)
```

标识

标识用于对变量、函数、类和关键字进行命名。标识的长度任意，但必须以大写或小写的字母字符或下划线字符 (`_`) 开头。以下划线开头的名称将通常保留作为内部名称或专用名称。在第一个字符后面，标识可以包含任意数目的字母字符、0 到 9 的数字以及下划线字符，并且这些字符和数字可以任意组合。

Jython 中的一些保留字不可用于对变量、函数或类进行命名。这些保留字分为下列类别：

- **语句引导词:** `assert`、`break`、`class`、`continue`、`def`、`del`、`elif`、`else`、`except`、`exec`、`finally`、`for`、`from`、`global`、`if`、`import`、`pass`、`print`、`raise`、`return`、`try` 和 `while`
- **参数引导词:** `as`、`import` 和 `in`
- **运算符:** `and`、`in`、`is`、`lambda`、`not` 和 `or`

关键字使用不当通常会生成 `SyntaxError`。

代码块

代码块是在期望单个语句的位置使用的语句组。代码块可以跟随下列任何语句：`if`、`elif`、`else`、`for`、`while`、`try`、`except`、`def` 和 `class`。这些语句将引入带有冒号字符 (`:`) 的代码块，例如：

```
if x == 1:
    y = 2
    z = 3
elif:
    y = 4
    z = 5
```

使用缩进对代码块进行定界，而不是像 Java 一样使用花括号。代码块中的所有行都必须缩进到同一位置。这是因为对缩进的更改指示代码块结束。通常，每一级缩进四个空格。建议使用空格而不是制表符来缩进行。不得混用空格和制表符。模块的最外层块中的行必须从第一列开始，否则将发生 `SyntaxError`。

组成代码块的语句（以及冒号后面的语句）也可以包括在一行中，并以分号分隔，例如：

```
if x == 1: y = 2; z = 3;
```

将参数传递给脚本

将参数传递给脚本非常有用，因为这表示可以重复使用脚本而无需进行修改。在命令行中传递的参数将作为列表 `sys.argv` 中的值进行传递。使用命令 `len(sys.argv)` 可以获取所传递的值的数目。例如：

```
import sys
print "test1"
print sys.argv[0]
print sys.argv[1]
print len(sys.argv)
```

在此示例中，`import` 命令用于导入整个 `sys` 类，以便可以使用这个类中存在的方法，例如 `argv`。

可以使用以下行调用此示例中的脚本：

```
/u/mjloos/test1 mike don
```

结果为以下输出：

```
/u/mjloos/test1 mike don
test1
mike
don
3
```

示例

`print` 关键字将打印紧跟其后的参数。如果语句后跟逗号，那么不会在输出中新增一行。例如：

```
print "This demonstrates the use of a",  
print " comma at the end of a print statement."
```

这将生成以下输出：

```
This demonstrates the use of a comma at the end of a print statement.
```

`for` 语句用于迭代代码块。例如：

```
mylist1 = ["one", "two", "three"]  
for lv in mylist1:  
    print lv  
    continue
```

在此示例中，将为列表 `mylist1` 分配 3 个字符串。然后，将打印该列表的元素，每个元素占用一行。这将生成以下输出：

```
one  
two  
three
```

在此示例中，迭代器 `lv` 将依次采用列表 `mylist1` 中每个元素的值，因为 `for` 循环用于实现每个元素的代码块。迭代器可以是任意长度的任何有效标识。

`if` 语句是条件语句。该语句将对条件进行求值，并根据求值结果返回 `true` 或 `false`。例如：

```
mylist1 = ["one", "two", "three"]  
for lv in mylist1:  
    if lv == "two"  
        print "The value of lv is ", lv  
    else  
        print "The value of lv is not two, but ", lv  
    continue
```

在此示例中，对迭代器 `lv` 进行了求值。如果 `lv` 的值为 `two`，那么将返回一个字符串，该字符串不同于 `lv` 的值不是 `two` 时返回的字符串。这将生成以下输出：

```
The value of lv is not two, but one  
The value of lv is two  
The value of lv is not two, but three
```

数学方法

您可以从 `math` 模块访问有用的数学方法。下表列出了其中一些方法。除非另有说明，否则所有值将作为浮点数返回。

表 7. 数学方法

方法	用法
<code>math.ceil(x)</code>	将 <code>x</code> 的上限作为浮点数返回，即大于或等于 <code>x</code> 的最小整数
<code>math.copysign(x, y)</code>	返回带有 <code>y</code> 的符号的 <code>x</code> 。 <code>copysign(1, -0.0)</code> 将返回 <code>-1</code>
<code>math.fabs(x)</code>	返回 <code>x</code> 的绝对值
<code>math.factorial(x)</code>	返回 <code>x</code> 阶乘。如果 <code>x</code> 是负数或非整数，那么将发生 <code>ValueError</code> 。
<code>math.floor(x)</code>	将 <code>x</code> 的下限作为浮点数返回，即小于或等于 <code>x</code> 的最大整数

表 7. 数学方法 (续)

方法	用法
<code>math.frexp(x)</code>	将 x 的尾数 (m) 和指数 (e) 作为 (m, e) 对返回。 m 是浮点数, e 是整数, 这样刚好满足 $x == m * 2^{**}e$ 。如果 x 为 0, 那么此方法将返回 $(0.0, 0)$, 否则将返回 $0.5 \leq \text{abs}(m) < 1$ 。
<code>math.fsum(iterable)</code>	返回 <code>iterable</code> 中值的精确浮点总和
<code>math.isinf(x)</code>	检查浮点数 x 是正不定式还是负不定式
<code>math.isnan(x)</code>	检查浮点数 x 是否为 NaN (非数字)
<code>math.ldexp(x, i)</code>	返回 $x * (2^{**}i)$ 。此方法本质上是函数 <code>frexp</code> 的反函数。
<code>math.modf(x)</code>	返回 x 的小数和整数部分。这两个结果都带有 x 的符号, 并且都是浮点数。
<code>math.trunc(x)</code>	返回已截断为 Integral 的 Real 值 x 。
<code>math.exp(x)</code>	返回 $e^{**}x$
<code>math.log(x[, base])</code>	返回以给定值 <code>base</code> 为底的 x 的对数。如果未指定 <code>base</code> , 那么将返回 x 的自然对数。
<code>math.log1p(x)</code>	返回 $1+x$ (<code>base e</code>) 的自然对数
<code>math.log10(x)</code>	返回以 10 为底的 x 的对数
<code>math.pow(x, y)</code>	返回 x 的 y 次幂。 <code>pow(1.0, x)</code> 和 <code>pow(x, 0.0)</code> 将始终返回 1, 即使 x 为 0 或非数字时也是如此。
<code>math.sqrt(x)</code>	返回 x 的平方根

除数学函数外, 还提供了一些有用的三角函数法。下表列出了这些方法。

表 8. 三角函数法

方法	用法
<code>math.acos(x)</code>	返回以弧度表示的 x 的反余弦
<code>math.asin(x)</code>	返回以弧度表示的 x 的正弦
<code>math.atan(x)</code>	返回以弧度表示的 x 的正切
<code>math.atan2(y, x)</code>	返回以弧度表示的 <code>atan(y / x)</code> 。
<code>math.cos(x)</code>	返回以弧度表示的 x 的余弦。
<code>math.hypot(x, y)</code>	返回欧几里得范数 <code>sqrt(x*x + y*y)</code> 。这是从原点到点 (x, y) 的向量的长度。
<code>math.sin(x)</code>	返回以弧度表示的 x 的正弦
<code>math.tan(x)</code>	返回以弧度表示的 x 的正切
<code>math.degrees(x)</code>	将角度 x 从弧度转换为度
<code>math.radians(x)</code>	将角度 x 从度转换为弧度
<code>math.acosh(x)</code>	返回 x 的反双曲余弦值
<code>math.asinh(x)</code>	返回 x 的反双曲正弦值
<code>math.atanh(x)</code>	返回 x 的反双曲正切值
<code>math.cosh(x)</code>	返回 x 的双曲余弦值
<code>math.sinh(x)</code>	返回 x 的双曲正弦值
<code>math.tanh(x)</code>	返回 x 的双曲正切值

还有两个数学常量。math.pi 的值为数学常量 pi。math.e 的值为数学常量 e。

使用非 ASCII 字符

要使用非 ASCII 字符，Python 需要明确地将字符串编码和解码为 Unicode。在 IBM SPSS Modeler 中，假定 Python 脚本采用 UTF-8 进行编码，这是支持非 ASCII 字符的标准 Unicode 编码。以下脚本将执行编译，这是因为 SPSS Modeler 已将 Python 编译器设置为 UTF-8。

```
stream = modeler.script.stream()
filenode = stream.createAt("variablefile", "テストノード", 96, 64)
```

但是，生成的节点将具有不正确的标签。



图 3. 错误显示的包含非 ASCII 字符的节点标签

标签不正确，因为 Python 已将字符串字面值自身转换为 ASCII 字符串。

Python 通过在字符串字面值前添加 u 字符前缀来支持指定 Unicode 字符串字面值：

```
stream = modeler.script.stream()
filenode = stream.createAt("variablefile", u"テストノード", 96, 64)
```

这将创建 Unicode 字符串，并且将正确显示标签。



图 4. 正确显示的包含非 ASCII 字符的节点标签

使用 Python 和 Unicode 是一个非常大的主题，它超出了本文档的范围。提供了许多对此主题进行了更详细介绍的书籍和在线资源。

面向对象的程序设计

面向对象的程序设计基于在程序中创建目标问题模型的概念。面向对象的程序设计减少了编程错误并促进了代码的复用。Python 是一种面向对象的语言。以 Python 定义的对象具有下列特征:

- **身份**。每个对象都必须截然不同，并且必须可以对此特征进行测试。is 和 is not 测试可用于此目的。
- **状态**。每个对象都必须能够存储状态。属性（例如字段和实例变量）可用于此目的。
- **行为**。每个对象都必须能够处理其状态。方法可用于此目的。

Python 提供了支持面向对象的程序设计的下列特征:

- **基于类的对象创建**。类是用于创建对象的模板。对象是具有关联行为的数据结构。
- **多态性继承**。Python 支持单继承和多重继承。所有 Python 实例方法都具有多态性，并且可以由子类覆盖。
- **具有隐藏数据的封装**。Python 允许隐藏属性。隐藏后，将只能通过类的方法从类外部访问这些属性。类实现了用于修改数据的方法。

定义类

在 Python 类中，可以定义变量和方法。与 Java 不同，您可以在 Python 中对每个源文件（或模块）定义任意数目的公共类。因此，可以认为 Python 中的模块类似于 Java 中的软件包。

在 Python 中，类是使用 class 语句定义的。class 语句的格式如下:

```
class name (superclasses): statement
```

或

```
class name (superclasses):  
    assignment  
    :  
    :  
    function  
    :  
    :
```

定义类时，您可以选择提供零个或零个以上的赋值语句。这些赋值语句将创建该类的所有实例共享的类属性。您还可以提供零个或零个以上的函数定义。这些函数定义将创建方法。超类列表是可选的。

在同一作用域中（即模块、函数或类中），类名应该唯一。您可以将多个变量定义为引用同一类。

创建类实例

类用于保存类（或共享）属性，或者用于创建类实例。要创建某个类的实例，请将该类作为函数进行调用。例如，请考虑以下类:

```
class MyClass:  
    pass
```

这里使用了 pass 语句，因为需要一个语句来完成类，但不需要以编程方式执行任何操作。

以下语句将创建类 MyClass 的实例:

```
x = MyClass()
```

向类实例添加属性

与 Java 不同，客户机可以在 Python 中向类实例添加属性。只有一个实例会发生更改。例如，要向实例 x 添加属性，请在该实例上设置新值:

```
x.attr1 = 1
x.attr2 = 2
.
.
x.attrN = n
```

定义类属性和方法

任何绑定在类中的变量都是类属性。任何在类中定义的函数都是方法。方法接收类的实例（通常称为 `self`）作为第一个自变量。例如，要定义一些类属性和方法，您可以输入以下代码：

```
class MyClass
    attr1 = 10      #class attributes
    attr2 = "hello"

    def method1(self):
        print MyClass.attr1  #reference the class attribute

    def method2(self):
        print MyClass.attr2  #reference the class attribute

    def method3(self, text):
        self.text = text      #instance attribute
        print text, self.text #print my argument and my attribute

    method4 = method3  #make an alias for method3
```

在类中，您应该使用类名限定所有对类属性的引用，例如 `MyClass.attr1`。应该使用 `self` 变量限定所有对实例属性的引用；例如 `self.text`。在类外部，您应该使用类名（例如 `MyClass.attr1`）或使用类的实例（例如 `x.attr1`，其中 `x` 是类的实例）限定所有对类属性的引用。在类外部，应该使用类的实例限定所有对实例变量的引用；例如 `x.text`。

隐藏变量

可以通过创建专用变量来隐藏数据。专用变量只能由类自身进行访问。如果您声明格式为 `__xxx` 或 `__xxx_yyy`（即，带有两个前置下划线）的名称，那么 Python 解析器将自动向声明的名称添加类名以创建隐藏变量，例如：

```
class MyClass:
    __attr = 10  #private class attribute

    def method1(self):
        pass

    def method2(self, p1, p2):
        pass

    def __privateMethod(self, text):
        self.__text = text  #private attribute
```

与 Java 不同，在 Python 中必须使用 `self` 限定所有对实例变量的引用；未隐舍地使用 `this`。

继承

从类进行继承的功能是面向对象的程序设计的基础。Python 同时支持单继承和多重继承。单继承表示只能存在一个超类。多重继承表示可以存在多个超类。

继承通过对其他类划分子类来实现。许多 Python 类都可以是超类。在 Python 的 Jython 实现中，只能从一个 Java 类进行直接或间接继承。无需提供超类。

超类中的任何属性或方法也包含在任何子类中，并且只要属性或方法未隐藏，类自身或任何客户机就可以使用这些属性或方法。可以在任何位置使用子类的任何实例，并且可以使用超类的实例；这是多态性示例。这些特征支持复用和轻松扩展。

示例

```
class Class1: pass    #no inheritance

class Class2: pass

class Class3(Class1): pass    #single inheritance

class Class4(Class3, Class2): pass    #multiple inheritance
```

第 3 章 在 IBM SPSS Modeler 中进行脚本编制

脚本类型

在 IBM SPSS Modeler 中，有 3 种类型的脚本：

- *流脚本*，用于控制单个流的执行并且它存储在流中。
- *超节点脚本*，用于控制超节点的行为。
- *独立或会话脚本*，可用于跨多个不同的流协调执行。

提供了在 IBM SPSS Modeler 中的脚本内使用的多种方法，您可以使用这些方法来访问各种 SPSS Modeler 功能。另外，这些方法还在第 31 页的第 4 章，『脚本编制 API』中用于创建更高级的函数。

流、超节点流和图

在大多数情况下，术语流表示同一内容，而与流是从文件装入还是在超节点中使用无关。通常，它表示连接在一起的节点的集合，并且可以执行。但是，在脚本编制中，并非所有操作在任何位置都受支持，这意味着脚本作者应该了解其所使用的流变体。

流

流是主要的 IBM SPSS Modeler 文档类型。可以保存、查找、编辑和执行流。流还可以具有参数、全局值、脚本以及其他与之关联的信息。

超节点流

*超节点流*是一种在超节点中使用的流类型。与普通流一样，它包含链接在一起的节点。超节点流与普通流之间存在一些差异：

- 参数及所有脚本与超节点流所在的超节点进行关联，而不是与超节点流自身关联。
- 超节点流具有附加的输入和输出连接器节点（具体取决于超节点的类型）。这些连接器节点用于将信息流入和流出超节点流，并且在创建超节点时自动创建。

图

术语图涵盖普通流和超节点流支持的功能，例如添加和除去节点，以及修改节点之间的连接。

执行流

以下示例将运行流中的所有可执行节点，并且它是最简单的流脚本类型：

```
modeler.script.stream().runAll(None)
```

以下示例也将运行流中的所有可执行节点：

```
stream = modeler.script.stream()  
stream.runAll(None)
```

在此示例中，流存储在名为 `stream` 的变量中。将流存储在变量中非常有用，因为脚本通常用于修改流或流中的节点。创建用于存储流的变量将生成一个更加简洁的脚本。

脚本编制上下文

`modeler.script` 模块提供了在其中执行脚本的上下文。此模块将在运行时自动导入 SPSS Modeler 脚本中。此模块定义了以下 4 个函数，这些函数提供可以访问其执行环境的脚本：

- `session()` 函数，用于返回脚本的会话。此会话定义语言环境等信息以及将用于运行任何流的 SPSS Modeler 后端（本地进程或联网 SPSS 建模器服务器）。
- `stream()` 函数，此函数可以与流脚本及超节点脚本配合使用。此函数将返回拥有正在运行的流脚本或超节点脚本的流。
- `diagram()` 函数，此函数可以与超节点脚本配合使用。此函数将返回超节点内的图。对于其他脚本类型，此函数返回的内容与 `stream()` 函数相同。
- `supernode()` 函数，此函数可以与超节点脚本配合使用。此函数将返回拥有正在运行的脚本的超节点。

下表概述了这四个函数及其输出。

表 9. `modeler.script` 函数摘要

脚本类型	<code>session()</code>	<code>stream()</code>	<code>diagram()</code>	<code>supernode()</code>
独立	返回会话	在调用该脚本时返回当前受管流（例如，通过批处理方式 <code>-stream</code> 选项传递的流）或 <code>None</code> 。	与 <code>stream()</code> 相同	不适用
流	返回会话	返回流	与 <code>stream()</code> 相同	不适用
超节点	返回会话	返回流	返回超节点流	返回超节点

`modeler.script` 模块还定义了一种使用退出码终止脚本的方式。`exit(exit-code)` 函数将停止执行脚本并返回所提供的整数退出码。

为流定义的其中一种方法是 `runAll(List)`。此方法将运行所有可执行节点。执行节点时生成的任何模型或输出都将添加到所提供的列表中。

流执行通常生成模型和图形等输出以及其他输出。要捕获此输出，脚本可以提供一個初始化为列表的变量，例如：

```
stream = modeler.script.stream()
results = []
stream.runAll(results)
```

执行完成后，可以从 `results` 列表访问执行所生成的任何对象。

引用现有节点

流通常是使用一些参数预先构建的，在执行流之前必须先修改这些参数。修改这些参数涉及下列任务：

1. 在相关流中找到节点。
2. 更改节点和/或流设置。

查找节点

流提供了多种查找现有节点的方式。下表概述了这些方法。

表 10. 查找现有节点的方法

方法	返回类型	描述
<code>s.findAll(type, label)</code>	集合	返回具有指定类型和标签的所有节点的列表。类型或标签可以为 <code>None</code> ，在这种情况下将使用其他参数。
<code>s.findAll(filter, recursive)</code>	集合	返回指定过滤器所接受的所有节点的集合。如果递归标志为 <code>True</code> ，那么还将搜索指定流内的任何超节点。
<code>s.findById(id)</code>	节点	返回具有所提供标识的节点或 <code>None</code> （如果不存在此类节点）。搜索范围限制为当前流。
<code>s.findByName(type, label)</code>	节点	返回具有所提供类型和/或标签的节点。类型或名称可以为 <code>None</code> ，在这种情况下将使用其他参数。如果有多个节点匹配，那么将选择并返回任意一个节点。如果没有匹配的节点，那么返回值为 <code>None</code> 。
<code>s.findDownstream(fromNodes)</code>	集合	从所提供的节点列表中进行搜索，并返回所提供节点下游的节点集合。返回的列表包括最初提供的节点。
<code>s.findUpstream(fromNodes)</code>	集合	从所提供的节点列表中进行搜索，并返回所提供节点上游的节点集合。返回的列表包括最初提供的节点。

例如，如果流包含脚本需要访问的单个“过滤”节点，那么可以使用以下脚本来找到该“过滤”节点：

```
stream = modeler.script.stream()
node = stream.findByName("filter", None)
...
```

另外，如果已经知道节点的标识（如节点对话框的“注释”选项卡中所示），那么可以使用此标识来查找节点，例如：

```
stream = modeler.script.stream()
node = stream.findById("id32FJT71G2") # the filter node ID
...
```

设置属性

节点、流、模型和输出都具有可以访问并在大多数情况下可以设置的属性。这些属性通常用于修改对象的行为或外观。下表概述了可用于访问和设置对象属性的方法。

表 11. 用于访问和设置对象属性的方法

方法	返回类型	描述
<code>p.getPropertyValue(propertyName)</code>	对象	返回指定属性的值或 <code>None</code> （如果不存在此属性）。
<code>p.setPropertyValue(propertyName, value)</code>	不适用	设置指定属性的值。

表 11. 用于访问和设置对象属性的方法 (续)

方法	返回类型	描述
<code>p.setPropertyValues(properties)</code>	不适用	设置指定属性的值。属性图中的每个条目都包含一个表示属性名的键，以及应指定给该属性的值。
<code>p.getKeyedPropertyValue(propertyName, keyName)</code>	对象	返回指定属性的值及关联的键或 <code>None</code> （如果不存在此属性或键）。
<code>p.setKeyedPropertyValue(propertyName, keyName, value)</code>	不适用	设置指定属性的值和键。

例如，如果要设置位于流的开始位置的“变量文件”节点的值，那么可以使用以下脚本：

```
stream = modeler.script.stream()
node = stream.findByType("variablefile", None)
node.setPropertyValue("full_filename", "$CLEO/DEMOS/DRUG1n")
...
```

或者，您可能希望根据“过滤”节点来过滤字段。在这种情况下，还将根据字段名称键入值，例如：

```
stream = modeler.script.stream()
# Locate the filter node ...
node = stream.findByType("filter", None)
# ... and filter out the "Na" field
node.setKeyedPropertyValue("include", "Na", False)
```

创建节点以及修改流

在某些情况下，您可能希望向现有流添加新节点。向现有流添加节点通常涉及下列任务：

1. 创建节点。
2. 将节点链接到现有流。

创建节点

流提供了多种创建节点的方式。下表概述了这些方法。

表 12. 创建节点的方法

方法	返回类型	描述
<code>s.create(nodeType, name)</code>	节点	创建具有指定类型的节点并将其添加到指定的流中。
<code>s.createAt(nodeType, name, x, y)</code>	节点	创建具有指定类型的节点并将其添加到指定流中的指定位置。如果 $x < 0$ 或 $y < 0$ ，那么未设置位置。
<code>s.createModelApplier(modelOutput, name)</code>	节点	创建派生自所提供的模型输出对象的模型应用器节点。

例如，要在流中创建新的“类型”节点，您可以使用以下脚本：

```
stream = modeler.script.stream()
# Create a new type node
node = stream.create("type", "My Type")
```


链接和取消链接节点

在流中创建新节点时，这个新节点必须先连接到节点序列，然后才可使用。流提供了多种链接节点和取消链接节点的方法。下表概述了这些方法。

表 13. 用于链接和取消链接节点的方法

方法	返回类型	描述
<code>s.link(source, target)</code>	不适用	在源节点与目标节点之间创建新链接。
<code>s.link(source, targets)</code>	不适用	在源节点与所提供列表中的每个目标节点之间创建新链接。
<code>s.linkBetween(inserted, source, target)</code>	不适用	连接两个其他节点实例（源节点和目标节点）之间的节点，并将已插入节点的位置设置为位于这两个节点实例之间。将首先除去源节点与目标节点之间的任何直接链接。
<code>s.linkPath(path)</code>	不适用	在节点实例之间创建新路径。第一个节点将链接到第二个节点，而第二个节点将链接到第三个节点，依此类推。
<code>s.unlink(source, target)</code>	不适用	除去源节点与目标节点之间的任何直接链接。
<code>s.unlink(source, targets)</code>	不适用	除去源节点与目标列表中每个对象之间的任何直接链接。
<code>s.unlinkPath(path)</code>	不适用	除去节点实例之间存在的任何路径。
<code>s.disconnect(node)</code>	不适用	除去所提供节点与指定流中任何其他节点之间的任何链接。
<code>s.isValidLink(source, target)</code>	布尔值	如果在指定的源节点与目标节点之间创建链接是有效的，那么此方法将返回 True。此方法将检查这两个对象是否属于指定流，源节点是否可以提供链接以及目标节点是否可以接收链接，并确认创建此类链接不会在流中引起循环。

下面的示例脚本将执行 5 项任务：

1. 创建“变量文件”输入节点、“过滤”节点和“表”输出节点。
2. 将这些节点连接到一起。
3. 在“变量文件”输入节点上设置文件名。
4. 根据生成的输出过滤字段“药品”。
5. 执行“表”节点。

```
stream = modeler.script.stream()
filenode = stream.createAt("variablefile", "My File Input ", 96, 64)
filternode = stream.createAt("filter", "Filter", 192, 64)
tablenode = stream.createAt("table", "Table", 288, 64)
stream.link(filenode, filternode)
stream.link(filternode, tablenode)
filenode.setPropertyValue("full_filename", "$CLEO_DEMOS/DRUG1n")
filternode.setKeyedPropertyValue("include", "Drug", False)
results = []
tablenode.run(results)
```

导入、替换和删除节点

与创建和连接节点一样，通常需要替换和删除流中的节点。下表概述了可用于导入、替换和删除节点的方法。

表 14. 用于导入、替换和删除节点的方法

方法	返回类型	描述
<code>s.replace(originalNode, replacementNode, discardOriginal)</code>	不适用	替换指定流中的指定节点。原始节点和替换节点都必须属于指定流。
<code>s.insert(source, nodes, newIDs)</code>	列表	在所提供的列表中插入节点的副本。假定所提供列表中的所有节点都包含在指定流中。 <code>newIDs</code> 标志指示应该为每个节点生成新标识，还是应该复制并使用现有标识。假定流中所有节点都具有唯一标识，这样在源流与指定流相同的情况下，必须将此标志设置为 <code>True</code> 。此方法将返回新插入节点的列表，此列表中未定义节点的顺序（即，顺序不一定与输入列表中节点的顺序相同）。
<code>s.delete(node)</code>	不适用	从指定流中删除指定节点。该节点必须属于指定流。
<code>s.deleteAll(nodes)</code>	不适用	从指定流中删除所有指定节点。集合中的所有节点都必须属于指定流。
<code>s.clear()</code>	不适用	从指定流中删除所有节点。

遍历流中的节点

标识特定节点上游或下游的节点是一项常见需求。流提供了一些可用于标识这些节点的方法。下表概述了这些方法。

表 15. 用于标识上游节点和下游节点的方法

方法	返回类型	描述
<code>s.iterator()</code>	迭代器	返回针对指定流中包含的节点对象的迭代器。如果在 <code>next()</code> 函数的两次调用之间对流进行了修改，那么将取消定义迭代器的行为。
<code>s.predecessorAt(node, index)</code>	节点(N)	返回所提供节点的指定直接前趋或 <code>None</code> （如果索引超出范围）。
<code>s.predecessorCount(node)</code>	<i>int</i>	返回所提供节点的直接前趋数。
<code>s.predecessors(node)</code>	列表	返回所提供节点的直接前趋。
<code>s.successorAt(node, index)</code>	节点(N)	返回所提供节点的指定直接后继或 <code>None</code> （如果索引超出范围）。
<code>s.successorCount(node)</code>	<i>int</i>	返回所提供节点的直接后继数。
<code>s.successors(node)</code>	列表	返回所提供节点的直接后继。

获取节点的相关信息

节点分为多种不同的类别，例如数据导入和导出节点、模型构建节点和其他类型的节点。每个节点都提供了一些可用于查找该节点的相关信息的方法。

下表概述了可用于获取节点的标识、名称和标签的方法。

表 16. 用于获取节点的标识、名称和标签的方法

方法	返回类型	描述
n.getLabel()	字符串	返回指定节点的显示标签。只有在属性 custom_name 是非空字符串并且属性 use_custom_name 未进行设置的情况下，标签才是前一属性的值；否则，标签是 getName() 的值。
n.setLabel(label)	不适用	设置指定节点的显示标签。如果新标签为非空字符串，那么会将其指定给属性 custom_name，并且会将 False 指定给属性 use_custom_name，以使指定的标签优先；否则，会将空字符串指定给属性 custom_name，并将 True 指定给属性 use_custom_name。
n.getName()	字符串	返回指定节点的名称。
n.getID()	字符串	返回指定节点的标识。每次创建新节点时都会创建一个新标识。将节点作为流的组成部分进行保存时，标识将随节点一起持久存储，以便在打开流时保留节点标识。但是，如果将已保存的节点插入流中，那么会将已插入的节点视为新对象并为其分配一个新标识。

下表概述了可用于获取节点的其他相关信息的方法。

表 17. 用于获取节点的相关信息的方法

方法	返回类型	描述
n.getTypeName()	字符串	返回此节点的脚本编制名称。此名称即为可用于创建该节点的新实例的名称。
n.isInitial()	布尔值	如果此节点是初始节点（即，位于流的开始位置的节点），那么此方法将返回 True。
n.isInline()	布尔值	如果此节点是内嵌节点（即，位于流中部的节点），那么此方法将返回 True。
n.isTerminal()	布尔值	如果此节点是终端节点（即，位于流的结束位置的节点），那么此方法将返回 True。
n.getXPosition()	整数	返回节点在流中的 x 位置偏移量。
n.getYPosition()	整数	返回节点在流中的 y 位置偏移量。
n.setXYPosition(x, y)	不适用	设置节点在流中的位置。

表 17. 用于获取节点的相关信息的方法 (续)

方法	返回类型	描述
n.setPositionBetween(source, target)	不适用	设置节点在流中的位置，以使其位于所提供的节点之间。
n.isCacheEnabled()	布尔值	如果已启用高速缓存，那么此方法将返回 True，否则将返回 False。
n.setCacheEnabled(val)	不适用	对此对象启用或禁用高速缓存。如果高速缓存已满并且高速缓存功能进入禁用状态，那么高速缓存将进行清仓。
n.isCacheFull()	布尔值	如果已禁用高速缓存，那么此方法将返回 True，否则将返回 False。
n.flushCache()	不适用	清空此节点的高速缓存。如果高速缓存未启用或者未滿，那么此方法不起任何作用。

第 4 章 脚本编制 API

脚本编制 API 简介

脚本编制 API 提供对各种不同的 SPSS Modeler 功能的访问。目前描述的所有方法是 API 的组成部分，可以在脚本内隐式地访问这些方法而不执行进一步的导入。但是，如果要引用 API 类，那么必须使用以下语句显式导入 API:

```
import modeler.api
```

此导入语句是许多脚本编制 API 示例所必需的。

示例: 使用定制过滤器搜索节点

第 25 页的『查找节点』一节提供了关于使用节点的类型名称作为搜索条件在流中搜索节点的示例。在某些情况下，需要执行更加通用的搜索，并且此搜索可使用 `NodeFilter` 类和流 `findAll()` 方法来实现。这种搜索包括以下两个步骤:

1. 创建用于扩展 `NodeFilter` 并实现 `accept()` 方法的定制版本的新类。
2. 使用此新类的实例来调用流 `findAll()` 方法。这将返回所有满足 `accept()` 方法中定义的条件节点。

以下示例显示如何在流中搜索已启用节点高速缓存的节点。所返回的节点列表可用于清空或禁用这些节点的高速缓存。

```
import modeler.api

class CacheFilter(modeler.api.NodeFilter):
    """A node filter for nodes with caching enabled"""
    def accept(this, node):
        return node.isCacheEnabled()

cachingnodes = modeler.script.stream().findAll(CacheFilter(), False)
```

元数据: 关于数据的信息

由于节点在流中连接在一起，因此提供了关于在各个节点提供的列或字段的信息。例如，在 `Modeler UI` 中，这将使您能够选择作为排序或汇总依据的字段。此信息称为数据模型。

另外，脚本还可以通过查看传入或传出节点的字段来访问数据模型。对于某些节点，输入数据模型与输出数据模型相同，例如，排序节点仅用于对记录进行重新排序，而不更改数据模型。某些节点（例如“派生”节点）可以添加新字段。其他节点（例如“过滤”节点）可以重命名或删除字段。

在下面的示例中，脚本采用标准 `IBM SPSS Modeler druglearn.str` 流，并为每个字段都构建一个删除了其中一个输入字段的模型。它通过以下过程来完成此操作:

1. 从“类型”节点访问输出数据模型。
2. 对输出数据模型中的各个字段进行遍历。
3. 针对各个输入字段修改“过滤”节点。
4. 更改所构建的模型名称。
5. 运行模型构建节点。

注：请记住，在 druglean.str 流中运行脚本之前，请先将脚本编制语言设置为 Python（该流在先前版本的 IBM SPSS Modeler 中创建，因此流脚本编制语言设置为旧脚本编制语言）。

```
import modeler.api

stream = modeler.script.stream()
filternode = stream.findByType("filter", None)
typenode = stream.findByType("type", None)
c50node = stream.findByType("c50", None)
# Always use a custom model name
c50node.setPropertyValue("use_model_name", True)

lastRemoved = None
fields = typenode.getOutputDataModel()
for field in fields:
    # If this is the target field then ignore it
    if field.getModelingRole() == modeler.api.ModelingRole.OUT:
        continue

    # Re-enable the field that was most recently removed
    if lastRemoved != None:
        filternode.setKeyedPropertyValue("include", lastRemoved, True)

    # Remove the field
    lastRemoved = field.getColumnName()
    filternode.setKeyedPropertyValue("include", lastRemoved, False)

    # Set the name of the new model then run the build
    c50node.setPropertyValue("model_name", "Exclude " + lastRemoved)
    c50node.run([])
```

DataModel 对象提供了多种用于访问关于数据模型内的字段或列的信息的方法。下表概述了这些方法。

表 18. 用于访问关于字段或列的信息的 DataModel 对象方法

方法	返回类型	描述
d.getColumnCount()	int	返回数据模型中的列数。
d.columnIterator()	迭代器	返回一个迭代器，用于以“自然”插入顺序返回各列。此迭代器将返回列实例。
d.nameIterator()	迭代器	返回一个迭代器，用于以“自然”插入顺序返回各列的名称。
d.contains(name)	布尔值	如果此 DataModel 中存在具有指定名称的列，那么此方法返回 True，否则，返回 False。
d.getColumn(name)	Column	返回具有指定名称的列。
d.getColumnGroup(name)	ColumnGroup	返回指定的列组或 None（如果不存在这样的列组）。
d.getColumnGroupCount()	int	返回此数据模型中的列组数。
d.columnGroupIterator()	迭代器	返回一个迭代器，用于依次返回各个列组。
d.toArray()	Column[]	将数据模型作为列数组返回。列按其“自然”插入顺序进行排序。

每个字段（列对象）都提供了多种用于访问关于列的信息的方法。下表列出了其中一些方法。

表 19. 用于访问关于列的信息的列对象方法

方法	返回类型	描述
<code>c.getColumnName()</code>	<i>string</i>	返回列的名称。
<code>c.getColumnLabel()</code>	<i>string</i>	返回列的标签或空字符串（如果不存在与该列关联的标签）。
<code>c.getMeasureType()</code>	MeasureType	返回列的度量类型。
<code>c.getStorageType()</code>	StorageType	返回列的存储类型。
<code>c.isMeasureDiscrete()</code>	布尔值	返回 True（如果列是独立的）。作为集合或标志的列将被视作独立的列。
<code>c.isModelOutputColumn()</code>	布尔值	返回 True（如果列是模型输出列）。
<code>c.isStorageDatetime()</code>	布尔值	返回 True（如果列的存储为时间、日期或时间戳记值）。
<code>c.isStorageNumeric()</code>	布尔值	返回 True（如果列的存储为整数或实数）。
<code>c.isValidValue(value)</code>	布尔值	返回 True（如果指定的值对于此存储有效）和 <code>valid</code> （如果有效的列值已知）。
<code>c.getModelingRole()</code>	ModelingRole	返回列的建模角色。
<code>c.getSetValues()</code>	Object[]	返回列的有效值阵列或 None（如果值未知或者列不是集合）。
<code>c.getValueLabel(value)</code>	<i>string</i>	返回列中的值的标签或空字符串（如果不存在与该值关联的标签）。
<code>c.getFalseFlag()</code>	对象	返回列的“false”指标值或 None（如果值未知或者列不是标志）。
<code>c.getTrueFlag()</code>	对象	返回列的“true”指标值或 None（如果值未知或者列不是标志）。
<code>c.getLowerBound()</code>	对象	返回列中的值的下限值或 None（如果值未知或者列不连续）。
<code>c.getUpperBound()</code>	对象	返回列中的值的上限值或 None（如果值未知或者列不连续）。

请注意，大多数用于访问关于列的信息的方法都在 `DataModel` 对象本身中定义了等效方法。例如，下列两个语句是等效语句：

```
dataModel.getColumn("someName").getModelingRole()
dataModel.getModelingRole("someName")
```

访问已生成的对象

执行某个流通常涉及生成附加输出对象。这些附加对象可能是新模型，也可能是提供要在后续执行中使用的信息的输出。

在以下示例中，`druglearn.str` 流再次用作流的起始点。在此示例中，将执行流中的所有节点，并且结果将存储在列表中。然后，脚本将遍历这些结果，并且执行所产生的任何模型输出都将保存为 `IBM SPSS Modeler` 模型（.gm）文件，而模型将以 `PMML` 格式导出。

```
import modeler.api
stream = modeler.script.stream()
```

```

# Set this to an existing folder on your system.
# Include a trailing directory separator
modelFolder = "C:/temp/models/"

# Execute the stream
models = []
stream.runAll(models)

# Save any models that were created
taskrunner = modeler.script.session().getTaskRunner()
for model in models:
    # If the stream execution built other outputs then ignore them
    if not(isinstance(model, modeler.api.ModelOutput)):
        continue

    label = model.getLabel()
    algorithm = model.getModelDetail().getAlgorithmName()

    # save each model...
    modelFile = modelFolder + label + algorithm + ".gm"
    taskrunner.saveModelToFile(model, modelFile)

    # ...and export each model PMML...
    modelFile = modelFolder + label + algorithm + ".xml"
    taskrunner.exportModelToFile(model, modelFile, modeler.api.FileFormat.XML)

```

任务运行器类提供了一种运行各项常见任务的便捷方式。下表概述了此类中提供的方法。

表 20. 用于执行常见任务的任务运行器类方法

方法	返回类型	描述
t.createStream(name, autoConnect, autoManage)	流	创建并返回新的流。请注意，必须以不公开方式创建流而不向用户显示这些流的代码应该将 autoManage 标志设置为 False。
t.exportDocumentToFile(documentOutput, filename, fileFormat)	不适用	使用指定的文件格式将流描述导出至文件。
t.exportModelToFile(modelOutput, filename, fileFormat)	不适用	使用指定的文件格式将模型导出至文件。
t.exportStreamToFile(stream, filename, fileFormat)	不适用	使用指定的文件格式将流导出至文件。
t.insertNodeFromFile(filename, diagram)	节点	从指定文件中读取并返回节点，然后将其插入所提供的图中。请注意，此方法可用于同时读取节点对象和超节点对象。
t.openDocumentFromFile(filename, autoManage)	文档输出	从指定文件读取并返回文档。
t.openModelFromFile(filename, autoManage)	模型输出	从指定文件读取并返回模型。
t.openStreamFromFile(filename, autoManage)	流	从指定文件读取并返回流。
t.saveDocumentToFile(documentOutput, filename)	不适用	将文档保存到指定的文件位置。
t.saveModelToFile(modelOutput, filename)	不适用	将模型保存到指定的文件位置。

表 20. 用于执行常见任务的任务运行器类方法 (续)

方法	返回类型	描述
<code>t.saveStreamToFile(stream, filename)</code>	不适用	将流保存到指定的文件位置。

处理错误

Python 语言提供了通过 `try...except` 代码块执行的错误处理方法。可以在脚本内使用此方法来捕获异常，并处理将导致脚本终止的问题。

在以下示例脚本中，已尝试从 IBM SPSS Collaboration and Deployment Services Repository 中检索模型。此操作可能会导致抛出异常，例如可能未正确设置存储库登录凭证或者存储库路径错误。在此脚本中，这可能会导致抛出 `ModelerException` (IBM SPSS Modeler 生成的所有异常都派生自 `modeler.api.ModelerException`)。

```
import modeler.api

session = modeler.script.session()
try:
    repo = session.getRepository()
    m = repo.retrieveModel("/some-non-existent-path", None, None, True)
    # print goes to the Modeler UI script panel Debug tab
    print "Everything OK"
except modeler.api.ModelerException, e:
    print "An error occurred:", e.getMessage()
```

注：某些脚本编制操作可能会导致抛出标准 Java 异常；这些异常并非派生自 `ModelerException`。要捕获这些异常，可以使用附加的 `except` 块来捕获所有 Java 异常，例如：

```
import modeler.api

session = modeler.script.session()
try:
    repo = session.getRepository()
    m = repo.retrieveModel("/some-non-existent-path", None, None, True)
    # print goes to the Modeler UI script panel Debug tab
    print "Everything OK"
except modeler.api.ModelerException, e:
    print "An error occurred:", e.getMessage()
except java.lang.Exception, e:
    print "A Java exception occurred:", e.getMessage()
```

流、会话和超节点参数

参数提供了一种在运行时传递值而不是在脚本中直接对这些值进行硬编码的有用方式。参数及其值的定义方式与流相同，即定义为流或超节点的参数表中的条目或命令行中的参数。流类和超节点类实现了一组由 `ParameterProvider` 对象定义的函数，如下表所示。会话提供了 `getParameters()` 调用，此调用将返回定义这些函数的对象。

表 21. 由 `ParameterProvider` 对象定义的函数

方法	返回类型	描述
<code>p.parameterIterator()</code>	迭代器	返回此对象的参数名的迭代器。

表 21. 由 *ParameterProvider* 对象定义的函数 (续)

方法	返回类型	描述
<code>p.getParameterDefinition(parameterName)</code>	参数定义	返回具有指定名称的参数的参数定义或 <code>None</code> (如果此提供程序中不存在此类参数)。结果可以是调用此方法时的定义快照, 并且不需要反映通过此提供程序对该参数进行的任何后续修改。
<code>p.getParameterLabel(parameterName)</code>	字符串	返回指定参数的标签或 <code>None</code> (如果不存在此类参数)。
<code>p.setParameterLabel(parameterName, label)</code>	不适用	设置指定参数的标签。
<code>p.getParameterStorage(parameterName)</code>	参数存储	返回指定参数的存储或 <code>None</code> (如果不存在此类参数)。
<code>p.setParameterStorage(parameterName, storage)</code>	不适用	设置指定参数的存储。
<code>p.getParameterType(parameterName)</code>	参数类型	返回指定参数的类型或 <code>None</code> (如果不存在此类参数)。
<code>p.setParameterType(parameterName, type)</code>	不适用	设置指定参数的类型。
<code>p.getParameterValue(parameterName)</code>	对象	返回指定参数的值或 <code>None</code> (如果不存在此类参数)。
<code>p.setParameterValue(parameterName, value)</code>	不适用	设置指定参数的值。

在以下示例中, 脚本汇总了一些 `Telco` 数据以查找具有最低平均收入数据的区域。然后, 将使用此区域设置一个流参数。接下来, 将在“选择”节点中使用该流参数从数据中排除此区域, 然后根据剩余的数据构建流失模型。

此示例并不真实, 这是因为脚本本身生成了“选择”节点, 并因此已经在“选择”节点表达式中直接生成正确的值。但是, 流通常是预先构建的, 因此通过这种方式设置参数提供了有用的示例。

示例脚本的第一部分用于创建流参数, 该流参数将包含平均收入最低的区域。另外, 此脚本还将在汇总分支和模型构建分支中创建节点, 并将这些节点连接在一起。

```
import modeler.api

stream = modeler.script.stream()

# Initialize a stream parameter
stream.setParameterStorage("LowestRegion", modeler.api.ParameterStorage.INTEGER)

# First create the aggregation branch to compute the average income per region
statisticsimportnode = stream.createAt("statisticsimport", "SPSS File", 114, 142)
statisticsimportnode.setPropertyValue("full_filename", "$CLEO_DEMOS/telco.sav")
statisticsimportnode.setPropertyValue("use_field_format_for_storage", True)

aggregatenode = modeler.script.stream().createAt("aggregate", "Aggregate", 294, 142)
aggregatenode.setPropertyValue("keys", ["region"])
aggregatenode.setKeyedPropertyValue("aggregates", "income", ["Mean"])

tablenode = modeler.script.stream().createAt("table", "Table", 462, 142)

stream.link(statisticsimportnode, aggregatenode)
stream.link(aggregatenode, tablenode)
```

```

selectnode = stream.createAt("select", "Select", 210, 232)
selectnode.setPropertyValue("mode", "Discard")
# Reference the stream parameter in the selection
selectnode.setPropertyValue("condition", "'region' = '$P-LowestRegion'")

typenode = stream.createAt("type", "Type", 366, 232)
typenode.setKeyedPropertyValue("direction", "churn", "Target")

c50node = stream.createAt("c50", "C5.0", 534, 232)

stream.link(statisticsimportnode, selectnode)
stream.link(selectnode, typenode)
stream.link(typenode, c50node)

```

此示例脚本将创建以下流。

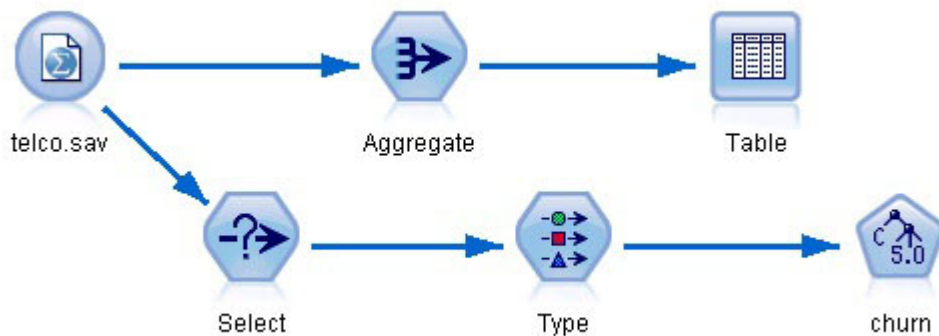


图 5. 示例脚本生成的流

示例脚本的以下部分用于执行位于汇总分支末尾的“表”节点。

```

# First execute the table node
results = []
tablenode.run(results)

```

示例脚本的以下部分用于访问执行“表”节点所生成的表输出。随后，此脚本将对表中的各行执行迭代，以查找平均收入最低的区域。

```

# Running the table node should produce a single table as output
table = results[0]

# table output contains a RowSet so we can access values as rows and columns
rowset = table.getRowSet()
min_income = 1000000.0
min_region = None

# From the way the aggregate node is defined, the first column
# contains the region and the second contains the average income
row = 0
rowcount = rowset.getRowCount()
while row < rowcount:
    if rowset.getValueAt(row, 1) < min_income:
        min_income = rowset.getValueAt(row, 1)
        min_region = rowset.getValueAt(row, 0)
    row += 1

```

脚本的以下部分使用平均收入最低的区域来设置先前创建的“LowestRegion”流参数。然后，在从训练数据中排除了指定区域的情况下，此脚本将运行模型构建器。

```
# Check that a value was assigned
if min_region != None:
    stream.setParameterValue("LowestRegion", min_region)
else:
    stream.setParameterValue("LowestRegion", -1)

# Finally run the model builder with the selection criteria
c50node.run([])
```

完整的示例脚本如下所示。

```
import modeler.api

stream = modeler.script.stream()

# Create a stream parameter
stream.setParameterStorage("LowestRegion", modeler.api.ParameterStorage.INTEGER)

# First create the aggregation branch to compute the average income per region
statisticsimportnode = stream.createAt("statisticsimport", "SPSS File", 114, 142)
statisticsimportnode.setPropertyValue("full_filename", "$CLEO_DEMOS/telco.sav")
statisticsimportnode.setPropertyValue("use_field_format_for_storage", True)

aggregatenode = modeler.script.stream().createAt("aggregate", "Aggregate", 294, 142)
aggregatenode.setPropertyValue("keys", ["region"])
aggregatenode.setKeyedPropertyValue("aggregates", "income", ["Mean"])

tablenode = modeler.script.stream().createAt("table", "Table", 462, 142)

stream.link(statisticsimportnode, aggregatenode)
stream.link(aggregatenode, tablenode)

selectnode = stream.createAt("select", "Select", 210, 232)
selectnode.setPropertyValue("mode", "Discard")
# Reference the stream parameter in the selection
selectnode.setPropertyValue("condition", "'region' = '$P-LowestRegion'")

typenode = stream.createAt("type", "Type", 366, 232)
typenode.setKeyedPropertyValue("direction", "churn", "Target")

c50node = stream.createAt("c50", "C5.0", 534, 232)

stream.link(statisticsimportnode, selectnode)
stream.link(selectnode, typenode)
stream.link(typenode, c50node)

# First execute the table node
results = []
tablenode.run(results)

# Running the table node should produce a single table as output
table = results[0]

# table output contains a RowSet so we can access values as rows and columns
rowset = table.getRowSet()
min_income = 1000000.0
min_region = None

# From the way the aggregate node is defined, the first column
# contains the region and the second contains the average income
row = 0
rowcount = rowset.getRowCount()
while row < rowcount:
```

```

    if rowset.getValueAt(row, 1) < min_income:
        min_income = rowset.getValueAt(row, 1)
        min_region = rowset.getValueAt(row, 0)
    row += 1

# Check that a value was assigned
if min_region != None:
    stream.setParameterValue("LowestRegion", min_region)
else:
    stream.setParameterValue("LowestRegion", -1)

# Finally run the model builder with the selection criteria
c50node.run([])

```

全局值

全局值用于计算指定字段的各项汇总统计。可以在流内部的任何位置访问这些汇总值。在流中，可以按名称访问全局值，这一点与流参数相似。全局值与流参数的差异在于，关联值将在“设置全局值”节点运行时自动进行更新，而不是通过脚本编制或命令行来指定。可以通过调用流的 `getGlobalValues()` 方法来访问该流的全局值。

`GlobalValues` 对象定义下表中显示的函数。

表 22. `GlobalValues` 对象所定义的函数

方法	返回类型	描述
<code>g.fieldNameIterator()</code>	迭代器	返回至少具有一个全局值的每个字段名称的迭代器。
<code>g.getValue(type, fieldName)</code>	对象	返回指定类型和字段名称的全局值或 <code>None</code> （如果找不到值）。虽然未来的功能可能会返回各种值类型，但通常期望返回值为数字。
<code>g.getValues(fieldName)</code>	图	返回包含指定字段名称的已知条目的图或 <code>None</code> （如果该字段没有现有条目）。

`GlobalValues.Type` 定义可用的汇总统计的类型。可用的汇总统计如下：

- `MAX`: 字段的最大值。
- `MEAN`: 字段的均值。
- `MIN`: 字段的最小值。
- `STDDEV`: 字段的标准差。
- `SUM`: 字段中值的总和。

例如，以下脚本将访问“收入”字段的均值，此均值由“设置全局值”节点计算：

```

import modeler.api

globals = modeler.script.stream().getGlobalValues()
mean_income = globals.getValue(modeler.api.GlobalValues.Type.MEAN, "income")

```

使用多个流 - 独立脚本

要使用多个流，必须使用独立脚本。可以在 `IBM SPSS Modeler UI` 内编辑和运行独立脚本，也可以在批处理方式下将独立脚本作为命令行参数进行传递。

以下独立脚本将打开两个流。其中一个流用于构建模型，而第二个流用于绘制预测值的分布。

```

# Change to the appropriate location for your system
demosDir = "C:/Program Files/IBM/SPSS/Modeler/16/DEMOS/streams/"

session = modeler.script.session()
tasks = session.getTaskRunner()

# Open the model build stream, locate the C5.0 node and run it
buildstream = tasks.openStreamFromFile(demosDir + "druglearn.str", True)
c50node = buildstream.findByType("c50", None)
results = []
c50node.run(results)

# Now open the plot stream, find the Na_to_K derive and the histogram
plotstream = tasks.openStreamFromFile(demosDir + "drugplot.str", True)
derivenode = plotstream.findByType("derive", None)
histogramnode = plotstream.findByType("histogram", None)

# Create a model applier node, insert it between the derive and histogram nodes
# then run the histogram
applyc50 = plotstream.createModelApplier(results[0], results[0].getName())
applyc50.setPositionBetween(derivenode, histogramnode)
plotstream.linkBetween(applyc50, derivenode, histogramnode)
histogramnode.setPropertyValue("color_field", "$C-Drug")
histogramnode.run([])

# Finally, tidy up the streams
buildstream.close()
plotstream.close()

```

第 5 章 脚本编写技巧

本节提供了使用脚本的技巧和方法概述，其中包括修改流执行以及在脚本中使用经过编码的密码。

修改流执行

运行时，将按缺省情形下的优化顺序来执行其终端节点。某些情况下，您可能更喜欢以其他顺序来执行。要修改流的执行顺序，请在流属性对话框的“执行”选项卡上完成以下步骤：

1. 打开一个空脚本。
2. 单击工具栏上的**追加缺省脚本**按钮来添加缺省流脚本。
3. 将缺省流脚本中语句的顺序更改为您希望的执行顺序。

使用模型

如果在 IBM SPSS Modeler 中启用了自动模型替换，那么将通过 IBM SPSS Modeler 用户界面执行模型构建器节点，链接到模型构建器节点的现有模型块将由新模型块替换。如果使用脚本执行了模型构建器节点，那么不会替换现有已链接的模型块。要替换现有模型块，必须在脚本中明确指定块的替换。

生成加密密码

某些情况下，可能需要在脚本中包含密码，例如，您可能需要访问受密码保护的数据源。加密密码可用在：

- 数据库源和输出节点的节点属性
- 登录到服务器的命令行自变量
- 存储在 *.par* 文件（由导出节点的“发布”选项卡生成的参数文件）中的数据库连接属性

通过此用户界面，可以使用一个工具根据 Blowfish 算法来生成加密密码（有关详细信息，请参阅 <http://www.schneier.com/blowfish.html>）。进行编码后，可以复制密码并将其存储到脚本文件和命令行自变量中。用于 database 和 databaseexport 的节点属性 epassword 的作用是存储经过编码的密码。

1. 要生成加密密码，请从“工具”菜单中选择：

对密码进行编码...

2. 在“密码”文本框中指定一个密码。
3. 单击**编码**，以便为您的密码生成随机编码。
4. 单击“复制”按钮，以便将经过编码的密码复制到剪贴板。
5. 将此密码粘贴到所需的脚本或参数中。

脚本检查

通过单击“独立脚本”对话框工具栏上的红色检查按钮，可以快速检查所有类型脚本的语法。



图 6. 流脚本工具栏图标

脚本检查将就您编码中的错误发出警报并给出改进建议。要查看错误行，请单击该对话框下半部分的反馈。此时将以红色突出显示错误。

从命令行编写脚本

通过编写脚本可以运行通常在用户界面中执行的操作。启动 IBM SPSS Modeler 时，只需在命令行中指定和运行一个独立流。

例如:

```
client -script scores.py -execute
```

-script 标记表示装入指定脚本，而 -execute 标记表示执行该脚本文件中的所有命令。

指定文件路径

指定目录和文件的文件路径时，您可以使用单个正斜杠 (/) 或两个反斜杠 (\) 作为目录分隔符，例如:

```
c:/demos/druglearn.str
```

或

```
c:\\demos\\druglearn.str
```

与早期版本的兼容性

通常，在不进行更改的情况下，IBM SPSS Modeler 的前发行版中创建的旧脚本应当可以在当前发行版中正常运行。要使这些脚本正常运行，必须在“流属性”对话框或“独立脚本”对话框中的流脚本选项卡上选择旧。现在，可以自动将模型块插入流中（此为缺省设置），并这些模型块可替代或补充流中该类型的现有模型块。实际发生的行为取决于将模型添加到流中和替换原有模型选项（工具 > 选项 > 用户选项 > 通知）的设置。例如，您可能需要修改以前版本中的脚本，在该版本中模型块替换是通过删除现有模型块并插入新的模型块来完成。

在当前版本中创建的脚本在以前的版本中可能无法正常运行。

在当前发行版中创建的 Python 脚本将无法在先前的发行版中正常运行。

如果在旧版本中创建的脚本使用了已被替换（或不支持）的命令，那么使用旧形式命令的脚本仍然会得到支持，但将显示一条警告消息。例如，旧的 generated 关键字已被 model 替换，且 clear generated 已被 clear generated palette 替换。沿用旧形式的脚本仍然可以运行，但将显示一条警告消息。

第 6 章 命令行自变量

调用软件

您可以使用操作系统的命令行来如下启动 IBM SPSS Modeler:

1. 在安装了 IBM SPSS Modeler 的计算机上, 打开 DOS 或命令提示符窗口。
2. 要以交互方式启动 IBM SPSS Modeler 界面, 请输入 `modelerclient` 命令, 然后输入所需的参数; 例如:
`modelerclient -stream report.str -execute`

可用参数 (标记) 允许您连接到服务器、装入流、运行脚本或根据需要指定其他参数。

命令行自变量的使用

您可以将命令行自变量 (也称为标记) 附加到最初的 `modelerclient` 命令以更改对 IBM SPSS Modeler 的调用。

存在多种可用的命令行自变量类型, 本节的随后内容将对其进行描述。

表 23. 命令行自变量类型.

自变量类型	描述位置
系统自变量	有关更多信息, 请参阅第 44 页的『系统自变量』主题。
参数自变量	有关更多信息, 请参阅第 45 页的『参数自变量』主题。
服务器连接自变量	有关更多信息, 请参阅第 45 页的『服务器连接自变量』主题。
IBM SPSS Collaboration and Deployment Services Repository 连接自变量	有关更多信息, 请参阅第 46 页的『IBM SPSS Collaboration and Deployment Services Repository 连接自变量』主题。

例如, 可以使用 `-server`、`-stream` 和 `-execute` 标记来连接到服务器, 然后装入并运行流, 如下所示:

```
modelerclient -server -hostname myserver -port 80 -username dminer  
-password 1234 -stream mystream.str -execute
```

请注意, 针对本地客户机安装运行时, 不需要指定服务器连接自变量。

可以用双引号括起包含空格的参数值, 例如:

```
modelerclient -stream mystream.str -Pusername="Joe User" -execute
```

另外, 还可以使用 `-script` 标志来通过此方式执行 IBM SPSS Modeler 脚本。

调试命令行自变量

要调试命令行, 请使用 `modelerclient` 命令在指定所需自变量的情况下启动 IBM SPSS Modeler。这样可以验证命令是否将按期望方式执行。另外, 您还可以在“会话参数”对话框 (“工具”菜单 ->“设置会话参数”) 中对通过命令行传递的任何参数的值进行确认。

系统自变量

下表描述可用于用户界面命令行调用的系统自变量。

表 24. 系统自变量

自变量	行为/描述
@ <commandFile>	@ 符号后跟文件名，此文件用于指定命令列表。当 <code>modelerclient</code> 遇到以 @ 开头的参数时，它将在该文件中对命令进行操作，就如同在命令行中一样。有关更多信息，请参阅第 47 页的『组合多个参数』主题。
-directory <dir>	设置缺省工作目录。在本地模式下，该目录将同时用于数据操作和输出。示例： <code>-directory c:/</code> 或 <code>-directory c:\</code>
-server_directory <dir>	为数据设置缺省服务器目录。通过 <code>-directory</code> 标记指定的工作目录将用于输出。
-execute	在启动后执行启动时所装入的流、状态或脚本。如果在流或状态之外还装入了脚本，那么脚本将单独执行。
-stream <stream>	启动时装入指定的流。可以指定多个流，但是最后一个指定的流将被设置为当前流。
-script <script>	启动时装入指定的独立脚本。如下所述，除流或状态之外，此标记还可用于指定脚本，但在启动时仅可装入一个脚本。如果脚本文件后缀为 <code>.py</code> ，那么将假定该文件为 Python 脚本，否则将假定它为旧脚本。
-model <model>	在启动时装入指定的已生成模型（ <code>.gm</code> 格式的文件）。
-state <state>	在启动时，装入指定的已保存状态。
-project <project>	装入指定工程。在启动时仅可装入一个工程。
-output <output>	在启动时装入已保存的输出项目（ <code>.cou</code> 格式的文件）。
-help	显示命令行自变量列表。指定此选项后，将忽略所有其他参数并显示帮助屏幕。
-P <name>=<value>	用于设置启动参数。另外，还可用于设置节点属性（槽参数）。
-scriptlang <python legacy>	此参数可用于指定与 <code>-script</code> 选项关联的脚本语言，而与脚本文件后缀无关。 示例 <pre>client -scriptlang python -script scores.txt -execute</pre> 这将使用 Python 运行所提供的脚本文件，即使文件后缀不是 <code>.py</code> 也是如此。

注意：也可以在用户界面中设置缺省目录。要访问上述选项，请在“文件”菜单中选择**设置工作目录**或**设置服务器目录**。

装入多个文件

命令行模式下，您可以通过在启动时重复输入每个装入对象的相关参数来装入多个流、状态和输出。例如，要装入和运行两个称为 `report.str` 和 `train.str` 的流，您可以使用如下命令：

```
modelerclient -stream report.str -stream train.str -execute
```

从 IBM SPSS Collaboration and Deployment Services Repository 装入对象

因为可以从某个文件或 IBM SPSS Collaboration and Deployment Services Repository（如果已获许可）装入特定对象，可以使用文件名前缀 `spssc:` 以及选择性地使用 `file:`（对于磁盘上的对象）来指示 IBM SPSS Modeler 在什么位置查找对象。前缀可与以下标记配合使用：

- `-stream`
- `-script`
- `-output`

- -model
- -project

您可以使用前缀创建 URI 以指定对象的位置，例如 `-stream "spsscr:///folder_1/scoring_stream.str"`。如果指定了 `spsscr:` 前缀，那么要求已在同一命令中指定了有效的 IBM SPSS Collaboration and Deployment Services Repository 连接。因此，完整的命令应形如以下的示例：

```
modelerclient -spsscr_hostname myhost -spsscr_port 8080
-spsscr_username myusername -spsscr_password mypassword
-stream "spsscr:///folder_1/scoring_stream.str" -execute
```

注意，在命令行中 **必须**使用 URI。不支持像 `REPOSITORY_PATH` 这样的简单路径。（此种路径仅适用于脚本。）

参数自变量

参数可用作在 IBM SPSS Modeler 的命令执行期间的标记。在命令行自变量中，`-P` 标记用于表示形如 `-P <name>=<value>` 的参数。

形式参数可以是：

- **简单参数**
- **槽参数**，也称为**节点属性**。此类参数用于修改流中各个节点的设置。请参阅主题第 50 页的『节点属性概述』以获取更多信息。
- **命令行自变量**，用于更改对 IBM SPSS Modeler 的调用。

例如，您可以提供数据源用户名和密码作为命令行标志，如下所示：

```
modelerclient -stream response.str -P:database.datasource={"ORA 10gR2", user1, mypsw, true}
```

格式与 `database` 节点属性的 `datasource` 参数相同。请参阅主题第 58 页的『`database` 节点属性』以获取更多信息。

服务器连接自变量

`-server` 标记指示 IBM SPSS Modeler 应连接到公共服务器，标记 `-hostname`、`-use_ssl`、`-port`、`-username`、`-password` 和 `-domain` 用于指示 IBM SPSS Modeler 如何连接到公共服务器。如果未指定 `-server` 参数，那么使用缺省 或本地 服务器。

示例

连接到公共服务器：

```
modelerclient -server -hostname myserver -port 80 -username dminer
-password 1234 -stream mystream.str -execute
```

连接到服务器集群：

```
modelerclient -server -cluster "QA Machines" \
-spsscr_hostname pes_host -spsscr_port 8080 \
-spsscr_username asmith -spsscr_epassword xyz
```

请注意，连接到服务器集群需要通过在整个 IBM SPSS Collaboration and Deployment Services 中使用过程协调器，因此 `-cluster` 参数必须与存储库连接选项 (`spsscr_*`) 结合使用。有关更多信息，请参阅第 46 页的『IBM SPSS Collaboration and Deployment Services Repository 连接自变量』主题。

表 25. 服务器连接自变量.

自变量	行为/描述
-server	以服务器模式运行 IBM SPSS Modeler, 同时使用标志 -hostname、-port、-username、-password 和 -domain 连接到公共服务器。
-hostname<name>	服务器的主机名。仅在服务器方式下可用。
-use_ssl	指定连接应使用 SSL (安全套接字层)。此标记为可选项, 缺省设置为不使用 SSL。
-port<number>	所指定服务器的端口号。仅在服务器方式下可用。
-cluster<name>	指定指向服务器集群 (而不是已命名的服务器) 的连接; 此参数可用来替代 hostname、port 和 use_ssl 参数。名称为聚类名, 或标识 IBM SPSS Collaboration and Deployment Services Repository 中聚类的唯一 URI。服务器集群由 IBM SPSS Collaboration and Deployment Services 中的过程协调器管理。有关更多信息, 请参阅「IBM SPSS Collaboration and Deployment Services Repository 连接自变量」主题。
-username<name>	这是用于登录服务器的用户名。仅在服务器方式下可用。
-password<password>	这是用于登录服务器的密码。仅在服务器方式下可用。注意: 如果未使用 -password 自变量, 那么系统将提示您输入密码。
-epassword<encodedpasswordstring>	这是用于登录服务器的经过编码的密码。仅在服务器方式下可用。注意: 可以通过 IBM SPSS Modeler 应用程序的“工具”菜单生成经过编码的密码。
-domain<name>	这是用于登录服务器的域。仅在服务器方式下可用。
-P <name>=<value>	用于设置启动参数。另外, 还可用于设置节点属性 (槽参数)。

IBM SPSS Collaboration and Deployment Services Repository 连接自变量

注意: 访问 IBM SPSS Collaboration and Deployment Services 存储库需要单独的许可证。有关更多信息, 请参阅<http://www.ibm.com/software/analytics/spss/products/deployment/cds/>。

如果想通过命令行来存储或检索 IBM SPSS Collaboration and Deployment Services 中的对象, 那么必须指定一个指向该 IBM SPSS Collaboration and Deployment Services Repository 的有效连接。例如:

```
modelerclient -spsscr_hostname myhost -spsscr_port 8080
-spsscr_username myusername -spsscr_password mypassword
-stream "spsscr:///folder_1/scoring_stream.str" -execute
```

下表列出了可用于建立连接的自变量。

表 26. IBM SPSS Collaboration and Deployment Services Repository 连接自变量

自变量	行为/描述
-spsscr_hostname <hostname or IP address>	安装 IBM SPSS Collaboration and Deployment Services Repository 的服务器的主机名或 IP 地址。
-spsscr_port <number>	IBM SPSS Collaboration and Deployment Services Repository 接受连接的端口号 (通常, 缺省值为 8080)。
-spsscr_use_ssl	指定连接应使用 SSL (安全套接字层)。此标记为可选项, 缺省设置为不使用 SSL。
-spsscr_username<name>	登录到 IBM SPSS Collaboration and Deployment Services Repository 的用户名。
-spsscr_password<password>	登录到 IBM SPSS Collaboration and Deployment Services Repository 的密码。
-spsscr_epassword<encoded password>	登录到 IBM SPSS Collaboration and Deployment Services Repository 的加密密码。

表 26. IBM SPSS Collaboration and Deployment Services Repository 连接自变量 (续)

自变量	行为/描述
-spsscr_domain<name>	登录到 IBM SPSS Collaboration and Deployment Services Repository 所使用的域。此标志是可选项，除非您使用 LDAP 或 Active Directory 进行登录，否则请勿使用此标志。

组合多个参数

通过使用后跟文件名的 @ 符号，可以在调用时指定的单个命令文件中组合使用多个自变量。这使您能够缩短命令行调用，并且可以克服操作系统在命令长度方面的限制。例如，以下启动命令使用了 <commandFileName > 的引用文件中的指定参数。

```
modelerclient @<commandFileName>
```

如果需用空格，那么请用引号将命令文件的文件名和路径括起来，如下所示：

```
modelerclient @ "C:\Program Files\IBM\SPSS\Modeler\mn\scripts\my_command_file.txt"
```

命令文件可以包含先前在启动时逐个指定的所有自变量，并且每个自变量各占一行。例如：

```
-stream report.str
-Porder.full_filename=APR_orders.dat
-Preport.filename=APR_report.txt
-execute
```

当写入和引用命令文件时，必须遵循以下限制：

- 每条命令各占一行。
- 不要在命令文件中嵌入 @CommandFile 参数。

第 7 章 属性参考信息

属性参考信息概述

可以为节点、流、超节点和工程指定多个不同的属性。某些属性在所有节点中通用，例如“名称”、“注释”和“工具提示”，有些属性则只针对某些特定的节点类型。其他属性涉及高级流操作，例如高速缓存或“超节点”行为。可通过标准用户界面（例如当打开对话框编辑节点选项时）访问属性，还可以多种其他方式使用属性。

- 可通过脚本修改属性（如本章所述）。有关更多信息，请参阅第 25 页的『设置属性』。
- 可在“超节点”参数中应用节点属性。
- 启动 IBM SPSS Modeler 时，节点属性还可用作命令行选项（使用 -P 标记）的一部分。

在 IBM SPSS Modeler 的脚本编制上下文中，节点属性和流属性通常称为**槽参数**。在本指南中，它们指的是节点或流的属性。

有关脚本编写语言的详细信息，请参阅第 11 页的第 2 章，『脚本语言』。

缩写

在节点属性语法中使用标准缩写。了解缩写有助于构建脚本。

表 27. 语法中使用的标准缩写。

缩写	含义
abs	绝对值
len	长度
最小	最小值
最大	最大值
correl	相关
covar	协方差
num	数字或数值
pct	百分比
transp	透明度
xval	交叉验证
变量	方差或变量（源节点中）

节点和流属性示例

在 IBM SPSS Modeler 中可以各种方式使用节点和流属性。此类属性经常用作脚本的一部分，作为**独立脚本**的一部分用以实现多个流或操作的自动化；或用作**流脚本**的一部分用以实现单个流内部的过程自动化。还可通过在“超节点”内使用节点参数来指定节点参数。就最基础的水平而言，属性还可用作命令行选项来启动 IBM SPSS Modeler。将 -p 参数用作命令行调用的一部分时，可以使用流属性来更改流设置。

请参阅主题第 35 页的『流、会话和超节点参数』和第 44 页的『系统自变量』以获取更多脚本编制示例。

节点属性概述

每种类型的节点都具有自己的一组合法属性，并且每个属性都具有类型。此类型可以是一般类型（数字、标志或字符串），在这种情况下，属性设置将强制转换为正确类型。如果无法进行强制转换，那么将发生错误。另外，通过属性引用可以指定合法值的范围，例如 Discard、PairAndDiscard 和 IncludeAsText，此时如果采用其他值，那么将出现错误。应使用值 True 和 False 来读取或设置标志属性。在本指南的参考表中，属性描述列对结构化属性进行了说明，并给出了属性的使用格式。

公共节点属性

IBM SPSS Modeler 中的很多属性通用于所有节点（包括超节点）。

表 28. 公共节点属性.

属性名称	数据类型	属性描述
use_custom_name	布尔值	
name	字符串	读取画布中某个节点名称的只读属性（自动或定制）。
custom_name	字符串	指定节点的定制名称。
tooltip	字符串	
annotation	字符串	
keywords	字符串	用于指定与对象关联的关键词列表的结构化槽
cache_enabled	布尔值	
node_type	source_supernode process_supernode terminal_supernode 所有为了进行脚本编制而指定的节点名	按类型引用节点的只读属性。例如，您还可以指定类型（例如 userinput 或 filter），而不是按名称引用节点（例如 real_income）。

超节点属性以及所有其他节点的属性均将单独讨论。有关更多信息，请参阅第 219 页的第 19 章，『超节点属性』主题。

第 8 章 流属性

通过脚本编写可以控制多种流属性。

脚本可以使用 `modeler.script` 模块中的 `stream()` 函数访问当前流，例如：

```
mystream = modeler.script.stream()
```

要引用流属性，您必须使用特殊的流变量，该变量通过在流前面添加一个 `^` 来表示。

`nodes` 属性用于引用当前流中的节点。

流属性的具体说明见于下表。

表 29. 流属性.

属性名称	数据类型	属性描述
<code>execute_method</code>	Normal Script	
<code>date_format</code>	"DDMMYY" "MMDDYY" "YYMMDD" "YYYYMMDD" "YYYYDDD" DAY MONTH "DD-MM-YY" "DD-MM-YYYY" "MM-DD-YY" "MM-DD-YYYY" "DD-MON-YY" "DD-MON-YYYY" "YYYY-MM-DD" "DD.MM.YY" "DD.MM.YYYY" "MM.DD.YY" "MM.DD.YYYY" "DD.MON.YY" "DD.MON.YYYY" "DD/MM/YY" "DD/MM/YYYY" "MM/DD/YY" "MM/DD/YYYY" "DD/MON/YY" "DD/MON/YYYY" MON YYYY q Q YYYY ww WK YYYY	
<code>date_baseline</code>	数字	
<code>date_2digit_baseline</code>	数字	

表 29. 流属性 (续).

属性名称	数据类型	属性描述
time_format	"HHMMSS" "HHMM" "MMSS" "HH:MM:SS" "HH:MM" "MM:SS" "(H)H:(M)M:(S)S" "(H)H:(M)M" "(M)M:(S)S" "HH.MM.SS" "HH.MM" "MM.SS" "(H)H.(M)M.(S)S" "(H)H.(M)M" "(M)M.(S)S"	
time_rollover	布尔值	
import_datetime_as_string	布尔值	
decimal_places	数字	
decimal_symbol	Default Period Comma	
angles_in_radians	布尔值	
use_max_set_size	布尔值	
max_set_size	数字	
ruleset_evaluation	Voting FirstHit	
refresh_source_nodes	布尔值	用于在流执行中自动刷新源节点。
script	字符串	
script_language	Python Legacy	为流脚本设置脚本语言。
annotation	字符串	
encoding	SystemDefault "UTF-8"	
stream_rewriting	布尔值	
stream_rewriting_maximise_sql	布尔值	
stream_rewriting_optimise_clem_execution	布尔值	
stream_rewriting_optimise_syntax_execution	布尔值	
enable_parallelism	布尔值	
sql_generation	布尔值	
database_caching	布尔值	
sql_logging	布尔值	
sql_generation_logging	布尔值	
sql_log_native	布尔值	
sql_log_prettyprint	布尔值	
record_count_suppress_input	布尔值	

表 29. 流属性 (续).

属性名称	数据类型	属性描述
record_count_feedback_interval	整数	
use_stream_auto_create_node_settings	布尔值	如果值为 true, 那么将使用特定于流的设置, 否则将使用用户首选项。
create_model_applier_for_new_models	布尔值	如果值为 true, 那么在模型构建器创建新模型并且没有处于活动状态的更新链接时, 将添加一个新的模型应用器。
create_model_applier_update_links	createEnabled createDisabled doNotCreate	定义自动添加模型应用器节点时创建的链接类型。
create_source_node_from_builders	布尔值	如果值为 true, 那么在源构建器创建新的源输出并且没有处于活动状态的更新链接时, 将添加一个新的源节点。
create_source_node_update_links	createEnabled createDisabled doNotCreate	定义自动添加源节点时创建的链接类型。

第 9 章 源节点属性

源节点公共属性

所有源节点的公共属性如下所示，后面的主题是具体节点的相关信息。

表 30. 源节点公共属性.

属性名称	数据类型	属性描述
direction	Input Target Both None Partition Split Frequency RecordID	字段角色的键控属性。 注意：值 In 和 Out 现已不推荐使用。 在将来的版本中可能取消对这些值的支持。
type	范围 (Range) Flag Set Typeless Discrete Ordered Set Default	字段类型。如果将该属性设置为 <i>Default</i> ，那么将清除所有 <i>values</i> 属性设置，如果将 <i>value_mode</i> 设置为 <i>Specify</i> ，那么它将重新设置为 <i>Read</i> 。如果 <i>value_mode</i> 已设置为 <i>Pass</i> 或 <i>Read</i> ，那么它将不受 <i>type</i> 设置的影响。
storage	Unknown 字符串 整数 实数 Time 日期 Timestamp	字段存储类型的只读键控属性。
check	None Nullify Coerce Discard Warn Abort	字段类型和范围检查的键控属性。
values	[value value]	对于连续型（范围）字段而言，第一个是最小值，后一个是最大值。对于名义（集合）字段，请指定所有值。对标志字段而言，第一个值代表 <i>false</i> ，后一个值代表 <i>true</i> 。设置该属性将自动把 <i>value_mode</i> 属性设置为 <i>Specify</i> 。
value_mode	Read Pass Read+ Current Specify	确定下一次数据传递中设置某个字段值的方式。 注意，不能将此属性直接设置为 <i>Specify</i> ；要使用特定值，需设置 <i>values</i> 属性。

表 30. 源节点公共属性 (续).

属性名称	数据类型	属性描述
default_value_mode	Read Pass	指定用缺省方式设置所有字段值。 该设置可以通过使用 value_mode 属性, 用特定字段进行覆盖。
extend_values	布尔值	当 value_mode 设置为 Read 时将应用。设为 T 则将新读取的值添加到任意现有字段值。设置为 F 则丢弃现有值并添加新读取值。
value_labels	字符串	用于指定值标签。请注意, 必须先指定值。
enable_missing	布尔值	当设置为 T 时, 那么激活对字段缺失值的跟踪。
missing_values	[value value ...]	指定表示缺失数据的数据值。
range_missing	布尔值	此属性设置为 T 时, 指定是否为字段定义缺失值 (空白) 范围。
missing_lower	字符串	range_missing 为 true 时, 此属性指定缺失值范围的下限。
missing_upper	字符串	range_missing 为 true 时, 此属性指定缺失值范围的上限。
null_missing	布尔值	当此属性设置为 T 时, 将用空 (在本软件中显示为 \$null\$ 的未定义值) 表示缺失值。
whitespace_missing	布尔值	当该属性设置为 T 时, 仅包含空白 (空格、制表符和换行符) 的值将被当成缺失值。
description	字符串	用于指定字段标签或描述。
default_include	布尔值	用于指定缺省行为是传递还是过滤字段的键控属性:
include	布尔值	用于指出是包含还是过滤单个字段的键控属性:
new_name	字符串	

asimport 节点属性

Analytic Server 源使您可以在 Hadoop 分布式文件系统 (HDFS) 上运行流。

表 31. asimport 节点属性.

asimport 节点属性	数据类型	属性描述
data_source	string	数据源的名称。
host	string	Analytic Server 主机的名称。
端口	integer	Analytic Server 进行侦听的端口。
tenant	string	在多租户环境中, 这是您所属的租户的名称。在单租户环境中, 此属性的缺省值为 ibm 。
set_credentials	布尔值	如果 Analytic Server 上的用户认证与 SPSS Modeler 服务器相同, 请将此属性设置为 false 。否则, 将其设置为 true 。
user_name	string	用于登录到 Analytic Server 的用户名。仅当 set_credentials 为 true 时才需要。
password	string	用于登录到 Analytic Server 的密码。仅当 set_credentials 为 true 时才需要。

cognosimport 节点属性



IBM Cognos BI 源节点从 Cognos BI 数据库导入数据。

表 32. cognosimport 节点属性.

cognosimport 节点属性	数据类型	属性描述
mode	Data Report	指定是导入 Cognos BI 数据（缺省）还是报告。
cognos_connection	<code>{"string",boolean,"string","string","string"}</code>	这是包含 Cognos 服务器连接详细信息的列表属性。格式为： <code>{"Cognos_server_URL", login_mode, "namespace", "username", "password"}</code> 其中： <i>Cognos_server_URL</i> 是包含源的 Cognos 服务器的 URL <i>login_mode</i> 指示是否使用匿名登录，其值为 true 或 false；如果设置为 true，那么应将下列字段设置为 "" <i>namespace</i> 指定用于登录服务器的安全认证提供程序 <i>username</i> 和 <i>password</i> 为用于登录 Cognos 服务器的用户名和密码
cognos_package_name	string	要从中导入数据对象的 Cognos 数据包的路径和名称，例如：/Public Folders/GOSALES 注意：只有正斜杠有效。
cognos_items	<code>{"field","field", ... ,"field"}</code>	要导入的一个或多个数据对象的名称。字段格式为 <code>[namespace].[query_subject].[query_item]</code>
cognos_filters	字段	导入数据前要应用的一个或多个过滤器的名称。
cognos_data_parameters	列表	数据的提示参数的值。“名称/值”对括在花括号内，并且每个对之间以逗号分隔，而整个字符串括在方括号内。格式： <code>[{"param1", "value"},...,{"paramN", "value"}]</code>
cognos_report_directory	字段	要从中导入报告的文件夹或包的 Cognos 路径，例如：/Public Folders/GOSALES 注意：只有正斜杠有效。
cognos_report_name	字段	要导入的报告的报告位置中的路径和名称。
cognos_report_parameters	列表	报告参数的值。“名称/值”对括在花括号内，并且每个对之间以逗号分隔，而整个字符串括在方括号内。格式： <code>[{"param1", "value"},...,{"paramN", "value"}]</code>

database 节点属性



“数据库”节点可用于通过 ODBC（开放式数据库连接）从各种其他数据包（包括 Microsoft SQL Server、DB2 和 Oracle 等等）中导入数据。

表 33. database 节点属性.

database 节点属性	数据类型	属性描述
mode	Table Query	借助对话框控件，指定将 <i>Table</i> 连接到数据库表，或借助 <i>SQL</i> 来指定用 <i>Query</i> 查询选定数据库。
datasource	字符串	数据库名称（另请参阅下面的注释）。
username	字符串	数据库连接详细信息（另请参阅下面的注释）。
password	字符串	
epassword	字符串	指定经过编码的密码，以代替在脚本中硬编码密码。 有关更多信息，请参阅第 41 页的『生成加密密码』主题。在执行期间，此属性为只读。
tablename	字符串	要访问的表名称。
strip_spaces	None Left Right Both	丢弃字符串中前端和尾部空格的选项。
use_quotes	AsNeeded Always Never	指定向数据库发送查询时，是否将表名和列名括在引号内（例如，如果这些名称包含空格或标点）。
query	字符串	指定要提交查询所对应的 SQL 编码。

注意：如果 `datasource` 属性中的数据库名称包含空格，请勿使用 `datasource`、`username` 和 `password` 等各个属性，而改为使用以下格式的单个数据源属性：

表 34. database 节点属性 - 特定于数据源.

database 节点属性	数据类型	属性描述
datasource	字符串	格式： [database_name,username,password[,true false]] 最后一个参数与经过加密的密码配合使用。如果将其设为 <code>true</code> ，将会在使用之前对密码进行解密。

如果您要更改数据源，也可使用此格式；不过，如果您只想更改用户名或密码，那么可使用 `username` 或 `password` 属性。

datacollectionimport 节点属性



IBM SPSS Data Collection 数据导入节点根据 IBM Corp. 市场调查产品使用的 IBM SPSS Data Collection 数据模型导入调查数据。必须安装 IBM SPSS Data Collection 数据库才可使用此节点。

图 7. 维度数据导入节点

表 35. datacollectionimport 节点属性.

datacollectionimport 节点属性	数据类型	属性描述
metadata_name	字符串	MDSC 的名称。特殊值 DimensionsMDD 表示应使用标准 IBM SPSS Data Collection 元数据文档。其他可能的值包括： mrADODsc mrI2dDsc mrLogDsc mrQdiDrsDsc mrQvDsc mrSampleReportingMDSC mrSavDsc mrSCDsc mrScriptMDSC 特殊值 none 指示不存在 MDSC。
metadata_file	字符串	存储元数据的文件的名称。
casedata_name	字符串	CDSC 的名称。可能的值包括： mrADODsc mrI2dDsc mrLogDsc mrPunchDSC mrQdiDrsDsc mrQvDsc mrRdbDsc2 mrSavDsc mrScDSC mrXm1Dsc 特殊值 none 指示不存在 CDSC。
casedata_source_type	未知 File 文件夹 UDL DSN	指出 CDSC 的源类型。
casedata_file	字符串	当 casedata_source_type 为 File 时，那么指定包含案例数据的文件。
casedata_folder	字符串	当 casedata_source_type 为 Foder 时，那么指定包含案例数据的文件夹。
casedata_udl_string	字符串	当 casedata_source_type 为 UDL 时，那么为包含案例数据的数据源指定 OLD-DB 连接字符串。

表 35. datacollectionimport 节点属性 (续).

datacollectionimport 节点属性	数据类型	属性描述
casedata_dsn_string	字符串	当 casedata_source_type 为 DSN, 那么为数据源指定 ODBC 连接字符串。
casedata_project	字符串	从 IBM SPSS Data Collection 数据库中读取观测值数据时, 可以输入工程的名称。对于所有其他的观测值数据类型, 应将此设置留空。
version_import_mode	All Latest Specify	定义版本处理方式。
specific_version	字符串	当 version_import_mode 为 Specify 时, 那么定义要导入案例数据的版本。
use_language	字符串	定义是否应使用指定语言的标签。
language	字符串	如果 use_language 的值为 True, 那么定义导入时要使用的语言代码。语言代码应为案例数据中的某一可用代码。
use_context	字符串	定义是否应导入特定的上下文。上下文用于区分与响应相关的描述。
context	字符串	如果 use_context 的值为 true, 那么定义导入环境。环境应是案例数据中的某一可用环境。
use_label_type	字符串	定义是否应导入指定标签类型。
label_type	字符串	如果 use_label_type 的值为 true, 那么定义要导入的标签类型。标签类型应是案例数据中的某一可用标签类型。
user_id	字符串	对于要求显式登录的数据库, 您可以提供用于访问数据源的用户 ID 和密码。
password	字符串	
import_system_variables	常用 None All	指定要导入的系统变量。
import_codes_variables	布尔值	
import_sourcefile_variables	布尔值	
import_multi_response	MultipleFlags Single	

excelimport 节点属性



“Excel 导入”节点可以从任何版本的 Microsoft Excel 中导入数据。不需要指定 ODBC 数据源。

表 36. excelimport 节点属性.

excelimport 节点属性	数据类型	属性描述
excel_file_type	Excel2003 Excel2007	
full_filename	字符串	完整文件名（包括路径）。
use_named_range	布尔值	是否使用指定的范围。如果为 true，那么将用 named_range 属性来指定读取范围，但忽略其他工作表和数据范围设置。
named_range	字符串	
worksheet_mode	Index Name	指定是否通过索引或名称来定义工作表。
worksheet_index	整数	要读取工作表的索引，开始时第一个工作表的索引为 0，第二个工作表的索引为 1，依此类推。
worksheet_name	字符串	要读取工作表的名称。
data_range_mode	FirstNonBlank ExplicitRange	指定确定范围的方式。
blank_rows	StopReading ReturnBlankRows	当 data_range_mode 为 FirstNonBlank 时，指定空行处理方式。
explicit_range_start	字符串	当 data_range_mode 为 ExplicitRange 时，指定要读取范围的起点。
explicit_range_end	字符串	
read_field_names	布尔值	指定是否应将指定范围的第一行用作字段（列）名称。

evimport 节点属性



Enterprise View 节点用于创建指向 IBM SPSS Collaboration and Deployment Services Repository 的连接，使您可以将 Enterprise View 数据读入流中，并将模型打包装入其他用户可通过存储库访问的方案。

表 37. evimport 节点属性.

evimport 节点属性	数据类型	属性描述
connection	列表	结构化属性 - 组成 Enterprise View 连接的参数的列表。
tablename	字符串	Application View 中表格的名称。

fixedfile 节点属性



“固定文件”节点从固定字段文本文件（即，文件中的字段不进行定界，而是从同一位置开始且长度固定）中导入数据。机器生成或遗存数据通常以固定字段格式进行存储。

表 38. *fixedfile* 节点属性.

fixedfile 节点属性	数据类型	属性描述
record_len	数字	指定每条记录中的字符数。
line_oriented	布尔值	跳过每条记录尾部的换行符。
decimal_symbol	Default Comma Period	这是数据源中使用的十进制分隔符的类型。
skip_header	数字	指定要在每条记录开头忽略的行数。用于忽略列标题。
auto_recognize_datetime	布尔值	指定在源数据中是否自动标识日期或时间。
lines_to_scan	数字	
fields	列表	结构化属性。
full_filename	字符串	要读取文件的全称（包括目录）。
strip_spaces	None Left Right Both	在导入时丢弃字符串中前端和尾部的空格。
invalid_char_mode	Discard Replace	从数据输入中除去无效字符（空值、0 或当前编码中所没有的字符），或用指定的单字符符号替换无效字符。
invalid_char_replacement	字符串	
use_custom_values	布尔值	
custom_storage	Unknown 字符串 整数 实数 Time 日期 Timestamp	

表 38. fixedfile 节点属性 (续).

fixedfile 节点属性	数据类型	属性描述
custom_date_format	"DDMMYY" "MMDDYY" "YYMMDD" "YYYYMMDD" "YYYYDDD" DAY MONTH "DD-MM-YY" "DD-MM-YYYY" "MM-DD-YY" "MM-DD-YYYY" "DD-MON-YY" "DD-MON-YYYY" "YYYY-MM-DD" "DD.MM.YY" "DD.MM.YYYY" "MM.DD.YY" "MM.DD.YYYY" "DD.MON.YY" "DD.MON.YYYY" "DD/MM/YY" "DD/MM/YYYY" "MM/DD/YY" "MM/DD/YYYY" "DD/MON/YY" "DD/MON/YYYY" MON YYYY q Q YYYY ww WK YYYY	只有在指定了定制存储的情况下，此属性才适用。
custom_time_format	"HHMMSS" "HHMM" "MMSS" "HH:MM:SS" "HH:MM" "MM:SS" "(H)H:(M)M:(S)S" "(H)H:(M)M" "(M)M:(S)S" "HH.MM.SS" "HH.MM" "MM.SS" "(H)H.(M)M.(S)S" "(H)H.(M)M" "(M)M.(S)S"	只有在指定了定制存储的情况下，此属性才适用。
custom_decimal_symbol	字段	只有在指定了定制存储的情况下适用。
encoding	StreamDefault SystemDefault "UTF-8"	指定文本编码方法。

sasimport 节点属性



SAS 导入节点可将 SAS 数据导入到 IBM SPSS Modeler 中。

表 39. sasimport 节点属性.

sasimport 节点属性	数据类型	属性描述
format	Windows UNIX Transport SAS7 SAS8 SAS9	要导入文件的格式。
full_filename	字符串	输入的完整文件名（包括路径）。
member_name	字符串	指定要从特定 SAS 传输文件中导入的成员。
read_formats	布尔值	从指定格式文件中读取数据格式（例如变量标签）。
full_format_filename	字符串	
import_names	NamesAndLabels LabelsasNames	指定在导入时映射变量名称和标签的方法。

simgen 节点属性



“模拟生成”节点提供了一种生成模拟数据的简单方法 - 使用用户指定的统计分布从头开始生成数据，或者使用对现有历史数据运行“模拟拟合”节点而获取的分布自动生成数据。对于模型输入中存在不确定性的情况，此节点在对预测模型的结果进行评估时非常有用。

表 40. simgen 节点属性.

simgen 节点属性	数据类型	属性描述
fields	结构化属性	
correlations	结构化属性	
max_cases	整数	最小值为 1000，最大值为 2,147,483,647
create_iteration_field	布尔值	
iteration_field_name	字符串	
replicate_results	布尔值	
random_seed	整数	
overwrite_when_refitting	布尔值	
parameter_xml	字符串	以字符串形式返回参数 XML

表 40. simgen 节点属性 (续).

simgen 节点属性	数据类型	属性描述
distribution	Bernoulli Beta Binomial Categorical Exponential Fixed Gamma Lognormal NegativeBinomialFailures NegativeBinomialTrials Normal Poisson Range Triangular Uniform Weibull	
bernoulli_prob	数字	$0 \leq \text{bernoulli_prob} \leq 1$
beta_shape1	数字	必须大于或等于 0
beta_shape2	数字	必须大于或等于 0
beta_min	数字	可选。必须小于 beta_max。
beta_max	数字	可选。必须大于 beta_min。
binomial_n	整数	必须 > 0
binomial_prob	数字	$0 \leq \text{binomial_prob} \leq 1$
binomial_min	数字	可选。必须小于 binomial_max。
binomial_max	数字	可选。必须大于 binomial_min。
exponential_scale	数字	必须 > 0
exponential_min	数字	可选。必须小于 exponential_max。
exponential_max	数字	可选。必须大于 exponential_min。
fixed_value	字符串	
gamma_shape	数字	必须大于或等于 0
gamma_scale	数字	必须大于或等于 0
gamma_min	数字	可选。必须小于 gamma_max。
gamma_max	数字	可选。必须大于 gamma_min。
lognormal_shape1	数字	必须大于或等于 0
lognormal_shape2	数字	必须大于或等于 0
lognormal_min	数字	可选。必须小于 lognormal_max。
lognormal_max	数字	可选。必须大于 lognormal_min。
negative_bin_failures_threshold	数字	必须大于或等于 0
negative_bin_failures_prob	数字	$0 \leq \text{negative_bin_failures_prob} \leq 1$
negative_bin_failures_min	数字	可选。必须小于 negative_bin_failures_max。
negative_bin_failures_max	数字	可选。必须大于 negative_bin_failures_min。
negative_bin_trials_threshold	数字	必须大于或等于 0
negative_bin_trials_prob	数字	$0 \leq \text{negative_bin_trials_prob} \leq 1$

表 40. *simgen* 节点属性 (续).

simgen 节点属性	数据类型	属性描述
negative_bin_trials_min	数字	可选。必须小于 negative_bin_trials_max。
negative_bin_trials_max	数字	可选。必须小于 negative_bin_trials_min。
normal_mean	数字	
normal_sd	数字	必须 > 0
normal_min	数字	可选。必须小于 normal_max。
normal_max	数字	可选。必须大于 normal_min。
poisson_mean	数字	必须大于或等于 0
poisson_min	数字	可选。必须小于 poisson_max。
poisson_max	数字	可选。必须大于 poisson_min。
triangular_mode	数字	$\text{triangular_min} \leq \text{triangular_mode} \leq \text{triangular_max}$
triangular_min	数字	必须小于 triangular_mode
triangular_max	数字	必须大于 triangular_mode
uniform_min	数字	必须小于 uniform_max
uniform_max	数字	必须大于 uniform_min
weibull_rate	数字	必须大于或等于 0
weibull_scale	数字	必须大于或等于 0
weibull_location	数字	必须大于或等于 0
weibull_min	数字	可选。必须小于 weibull_max。
weibull_max	数字	可选。必须大于 weibull_min。

相关性可以是介于 +1 与 -1 之间的任何数字。您可以根据需要指定相关性。任何未指定的相关性都将设置为 0。如果存在任何未知字段，那么应该在相关性矩阵（或表）上设置相关性值，并以红色文本显示该值。如果存在未知字段，那么无法执行节点。

statisticsimport 节点属性



IBM SPSS Statistics 文件节点从 IBM SPSS Statistics 使用的 .sav 文件格式以及保存在 IBM SPSS Modeler 中的高速缓存文件（其也使用相同格式）读取数据。

有关此节点属性的信息，请参阅第 215 页的『statisticsimport 节点属性』。

userinput 节点属性



用户输入节点提供了一种用于创建综合数据的简单方式 - 可以从头开始创建也可以通过更改现有数据进行创建。此节点非常有用，例如，在希望为建模创建测试数据集时，即可使用此节点。

表 41. *userinput* 节点属性.

userinput 节点属性	数据类型	属性描述
data		每个字段的数据长度可以不同，但必须与字段存储一致。如果为某个并不存在的字段设置值，那么将创建此字段。另外，如果将某个字段的值设置为空字符串 (" ")，那么将删除指定字段。 注： 针对此属性输入的值必须是字符串，而不是数字。例如，必须以 "1 2 3 4" 的形式输入数字 1、2、3 和 4。
names		设置或返回节点所生成的字段名称列表的结构化通道。
custom_storage	Unknown 字符串 整数 实数 Time 日期 Timestamp	可用于设置或返回某个字段存储的通道。
data_mode	Combined Ordered	如果指定了 Combined，那么设定值以及最小/最大值的每个组合都将生成一个记录。生成的记录数等于每个字段中值的数量的乘积。如果指定了有序，那么从每条记录的每一列中提取一个值来生成数据行。生成的记录数等于一个与字段相关的最大数值。将为所有数据值较少的字段添加空值。
values		此属性已由 data 取代，不应继续使用。

variablefile 节点属性



“变量文件”节点读取自由格式字段文本文件中的数据，即，其记录包含固定数量的字段，但包含不定数量字符的文件。此节点对于具有固定长度标题文本和某些特定类型注解的文件也非常有用。

表 42. *variablefile* 节点属性.

variablefile 节点属性	数据类型	属性描述
skip_header	数字	指定每条记录开头要忽略的字符数。
num_fields_auto	布尔值	自动确定每条记录中的字段数。记录必须以换行符终止。
num_fields	数字	手动指定每条记录中的字段数。
delimit_space	布尔值	指定文件中用于划定字段边界的字符。
delimit_tab	布尔值	
delimit_new_line	布尔值	
delimit_non_printing	布尔值	

表 42. variablefile 节点属性 (续).

variablefile 节点属性	数据类型	属性描述
delimit_comma	布尔值	当逗点在流中同时用作十进制分隔符和字段定界符时, 将 delimit_other 设置为 true, 然后使用其他属性将逗号指定为定界符。
delimit_other	布尔值	允许使用其他属性来指定定制定界符。
other	字符串	在 delimit_other 为 true 时, 指定要使用的定界符。
decimal_symbol	Default Comma Period	指定用于数据源中的十进制分隔符。
multi_blank	布尔值	将多个相邻空格定界符视为一个单一定界符处理。
read_field_names	布尔值	将数据文件的第一行作为列的标签。
strip_spaces	None Left Right Both	在导入时丢弃字符串中前端和尾部的空格。
invalid_char_mode	Discard Replace	从数据输入中除去无效字符 (空值、0 或当前编码中所没有的字符), 或用指定的单字符符号替换无效字符。
invalid_char_replacement	字符串	
break_case_by_newline	布尔值	指定行定界符为换行符。
lines_to_scan	数字	指定具体数据类型的扫描行数。
auto_recognize_datetime	布尔值	指定在源数据中是否自动标识日期或时间。
quotes_1	Discard PairAndDiscard IncludeAsText	指定导入后单引号的处理方式。
quotes_2	Discard PairAndDiscard IncludeAsText	指定导入后双引号的处理方式。
full_filename	字符串	要读取的文件全称 (包括目录)。
use_custom_values	布尔值	
custom_storage	Unknown 字符串 整数 实数 Time 日期 Timestamp	

表 42. variablefile 节点属性 (续).

variablefile 节点属性	数据类型	属性描述
custom_date_format	"DDMMYY" "MMDDYY" "YYMMDD" "YYYYMMDD" "YYYYDDD" DAY MONTH "DD-MM-YY" "DD-MM-YYYY" "MM-DD-YY" "MM-DD-YYYY" "DD-MON-YY" "DD-MON-YYYY" "YYYY-MM-DD" "DD.MM.YY" "DD.MM.YYYY" "MM.DD.YY" "MM.DD.YYYY" "DD.MON.YY" "DD.MON.YYYY" "DD/MM/YY" "DD/MM/YYYY" "MM/DD/YY" "MM/DD/YYYY" "DD/MON/YY" "DD/MON/YYYY" MON YYYY q Q YYYY ww WK YYYY	只有在指定了定制存储的情况下适用。
custom_time_format	"HHMMSS" "HHMM" "MMSS" "HH:MM:SS" "HH:MM" "MM:SS" "(H)H:(M)M:(S)S" "(H)H:(M)M" "(M)M:(S)S" "HH.MM.SS" "HH.MM" "MM.SS" "(H)H.(M)M.(S)S" "(H)H.(M)M" "(M)M.(S)S"	只有在指定了定制存储的情况下适用。
custom_decimal_symbol	字段	只有在指定了定制存储的情况下适用。
encoding	StreamDefault SystemDefault "UTF-8"	指定文本编码方法。

xmlimport 节点属性



“XML 源”节点将 XML 格式的数据导入到流中。可以导入单个文件，也可以导入某个目录中的所有文件。您可以选择性地指定模式文件，以便从中读取 XML 结构。

表 43. xmlimport 节点属性.

xmlimport 节点属性	数据类型	属性描述
read	single directory	读取单个数据文件（缺省），或目录中的所有 XML 文件。
recurse	布尔值	指定是否另外读取指定目录的所有子目录中的 XML 文件。
full_filename	字符串	（必需）要导入的 XML 文件的完整路径和文件名（如果 read = single）。
directory_name	字符串	（必需）要从中导入 XML 文件的目录的完整路径和名称（如果 read = directory）。
full_schema_filename	字符串	要从中读取 XML 结构的 XSD 或 DTD 文件的完整路径和文件名。如果您省略了此参数，将从 XML 源文件中读取结构。
records	字符串	XPath 表达式（例如，/author/name），用以定义记录边界。每次在源文件中遇到此元素时，都将创建新的记录。
mode	read specify	读取所有数据（缺省），或指定要读取的项目。
fields		要导入的项目（元素和属性）列表。列表中的每项为 XPath 表达式。

第 10 章 记录操作节点属性

append 节点属性



“追加”节点用于连接多组记录。另外，也可以用于将结构类似但内容不同的数据集组合到一起。

表 44. append 节点属性.

append 节点属性	数据类型	属性描述
match_by	Position Name	可以根据字段在主数据源中的位置或输入数据集中字段的名称来附加数据集。
match_case	布尔值	匹配字段名称时启用区分大小写。
include_fields_from	Main All	
create_tag_field	布尔值	
tag_field_name	字符串	

aggregate 节点属性



“汇总”节点将一系列输入记录替换为经过摘要和汇总的输出记录。

表 45. aggregate 节点属性.

aggregate 节点属性	数据类型	属性描述
keys	<i>[field field ... field]</i>	列出可用作汇总键的字段。例如，如果 Sex 和 Region 是键字段，M 和 F 与区域 N 和 S 的每个唯一性组合（四个唯一性组合）都将具有一个经过汇总的记录。
contiguous	布尔值	如果您知道输入中所有具有相同键值的记录都分组到一起（例如，如果按键字段对输入进行了排序），请选择此选项。这样做有助于提高性能。
aggregates		一种结构化属性，它列出其值将被汇总的数字字段以及选定的汇总模式。
extension	字符串	对重复的汇总字段指定前缀或后缀（样本如下）。
add_as	Suffix Prefix	
inc_record_count	布尔值	创建一个额外字段，该字段指定为形成每条汇总记录汇总了多少条输入记录。
count_field	字符串	指定记录计数字段的名称。

balance 节点属性



“均衡”节点用于纠正数据集中的不平衡，以使其遵循指定的条件。“均衡”伪指令根据指定系数调整条件成立的记录所占的比例。

表 46. *balance* 节点属性.

balance 节点属性	数据类型	属性描述
directives		根据指定数字平衡字段值比例的结构化属性（参阅下面的示例）。
training_data_only	布尔值	指定只应该对训练数据进行平衡。如果流中不存在分区字段，那么将忽略此选项。

directives 节点属性使用以下格式:

```
[[ number string ] \ { number string } \ ... { number string } ]].
```

注: 如果在表达式中嵌入字符串（括在双引号内），那么必须在字符串前添加转义字符" \ "。" \ "字符同时也是行继续符，这样参数就可以显示在同一行，看起来清楚了。

derive_stb 节点属性



“空间时间限制”节点根据纬度、经度和时间戳记字段派生了空间时间限制。您还可以将频繁的空间时间限制标识为逗留。

表 47. *derive_stb* 节点属性.

derive_stb 节点属性	数据类型	属性描述
mode	IndividualRecords Hangouts	
latitude_field	字段	
longitude_field	字段	
timestamp_field	字段	
hangout_density	密度	单一密度。请参阅 <i>densities</i> 以了解有效的密度值。

表 47. *derive_stb* 节点属性 (续).

derive_stb 节点属性	数据类型	属性描述
densities	[<i>density,density,..., density</i>]	每个密度都是一个字符串，例如 STB_GH8_1DAY。 注：对于哪些密度有效，存在限制。对于 geohash，可以使用 GH1 到 GH15 中的值。对于 temporal 部分，可以使用下列值： EVER 1YEAR 1MONTH 1DAY 12HOURS 8HOURS 6HOURS 4HOURS 3HOURS 2HOURS 1HOUR 30MINS 15MINS 10MINS 5MINS 2MINS 1MIN 30SECS 15SECS 10SECS 5SECS 2SECS 1SEC
id_field	字段	
qualifying_duration	1DAY 12HOURS 8HOURS 6HOURS 4HOURS 3HOURS 2Hours 1HOUR 30MIN 15MIN 10MIN 5MIN 2MIN 1MIN 30SECS 15SECS 10SECS 5SECS 2SECS 1SECS	必须是字符串。
min_events	<i>integer</i>	最小有效整数值为 2。
qualifying_pct	<i>integer</i>	必须介于 1 与 100 之间。
add_extension_as	Prefix Suffix	
name_extension	<i>string</i>	

distinct 节点属性



“区分”节点通过将第一个区分记录传递到数据流，或者通过丢弃第一个记录并将任何重复记录传递到数据流，除去重复的记录。

表 48. distinct 节点属性.

distinct 节点属性	数据类型	属性描述
mode	Include Discard	既可以将第一条区分记录包括在数据流中，也可以丢弃第一条区分记录并将任何重复记录传送到数据流。
grouping_fields	[<i>field field field</i>]	列出用于确定记录是否相同的字段。 注: 从 IBM SPSS Modeler 16 起，不推荐使用此属性。
composite_value	结构化槽	
composite_values	结构化槽	
inc_record_count	布尔值	创建一个额外字段，该字段指定为形成每条汇总记录汇总了多少条输入记录。
count_field	字符串	指定记录计数字段的名称。
sort_keys	结构化槽。	注: 从 IBM SPSS Modeler 16 起，不推荐使用此属性。
default_ascending	布尔值	
low_distinct_key_count	布尔值	指定只存在少量记录以及/或者键字段的少量唯一值。
keys_pre_sorted	布尔值	指定具有相同键值的所有记录在输入中分组在一起。
disable_sql_generation	布尔值	

merge 节点属性



“合并”节点使用多个输入记录，并创建包含某些或全部输入字段的单个输出记录。这对于合并来源不同的数据 非常有用，例如内部客户数据和已购买人群统计数据。

表 49. merge 节点属性.

merge 节点属性	数据类型	属性描述
method	Order Keys Condition	指定记录是否按它们在数据文件中的列示顺序进行合并，是否使用一个或多个键字段来合并键字段中包含相同值的记录，或者是否在满足指定条件时合并记录。
condition	字符串	如果 method 设置为 Condition，指定包括或丢弃记录的条件。
key_fields	[<i>field field field</i>]	
common_keys	布尔值	

表 49. merge 节点属性 (续).

merge 节点属性	数据类型	属性描述
join	Inner FullOuter PartialOuter Anti	
outer_join_tag.n	布尔值	在此属性中, <i>n</i> 是“选择数据集”对话框中显示的标记名。注意, 可以指定多个标记名, 因为任何数量的数据集都无法提供完整记录。
single_large_input	布尔值	指定是否进行优化, 以使一个输入与其他输入相比具有一个相对较大的输入值。
single_large_input_tag	字符串	按“选择较大数据集”对话框中的显示指定标记名。请注意, 此属性的用法与 outer_join_tag 属性的用法略有不同 (布尔值与字符串), 因为前者只能指定一个输入数据集。
use_existing_sort_keys	布尔值	指定输入值是否已根据一个或多个键字段进行排序。
existing_sort_keys	[[<i>string</i> Ascending] \ { <i>string</i> Descending}]	指定已排序的字段及其排序方向。

rfmaggregate 节点属性



使用“近因、频率和货币 (RFM) 汇总”节点, 您可以采用客户的 历史记录事务处理数据, 删除所有无用数据以及将所有他们保留的事务处理数据组合成一行, 且该行中列出了他们与您上次谈业务的时间、所完成的交易量以及这些交易的总货币价值。

表 50. rfmaggregate 节点属性.

rfmaggregate 节点属性	数据类型	属性描述
relative_to	Fixed Today	指定计算交易近因的日期。
reference_date	日期	仅在 relative_to 中选择 Fixed 时才可用。
contiguous	布尔值	如果您的数据进行了预先排序, 以便所有 ID 相同的记录一起出现在数据流中, 那么选择此选项可以加快处理速度。
id_field	字段	指定该字段以用来识别客户及其交易。
date_field	字段	指定将要用来计算近因的日期字段。
value_field	字段	指定该字段以用来计算货币值。
extension	字符串	为重复汇总字段指定前缀或后缀。
add_as	Suffix Prefix	指定是否应作为前缀或后缀来添加 extension。
discard_low_value_records	布尔值	启用使用 discard_records_below 设置。
discard_records_below	数字	可在计算 RFM 总计时, 指定一个最小值, 凡低于该值的交易详细信息都不再被使用。值单位与所选 value 字段相关。

表 50. *rfmaggregate* 节点属性 (续).

rfmaggregate 节点属性	数据类型	属性描述
only_recent_transactions	布尔值	启用使用 <code>specify_transaction_date</code> 或 <code>transaction_within_last</code> 设置。
specify_transaction_date	布尔值	
transaction_date_after	日期	只有选中 <code>specify_transaction_date</code> 时才可用。指定交易日期以在分析时包含其之后的记录。
transaction_within_last	数字	只有选中 <code>transaction_within_last</code> 时才可用。指定从计算相对于以下内容的近因日期字段所返回的周期数和周期类型（天、周、月或年），在此日期之后的记录将被包含在您的分析中。
transaction_scale	Days Weeks Months Years	只有选中 <code>transaction_within_last</code> 时才可用。指定从计算相对于以下内容的近因日期字段所返回的周期数和周期类型（天、周、月或年），在此日期之后的记录将被包含在您的分析中。
save_r2	布尔值	显示每个客户第二个最近交易的日期。
save_r3	布尔值	只有选中 <code>save_r2</code> 时才可用。显示每个客户第三个最近交易的日期。

Rprocess 节点属性



通过使用您自己的定制 R 脚本，可以使用“R 进程”节点从 IBM(r) SPSS(r) Modeler 流中获取数据并进行修改。修改数据后，会将数据返回到流中。

表 51. *Rprocess* 节点属性.

Rprocess 节点属性	数据类型	属性描述
syntax	字符串	
convert_flags	StringsAndDoubles LogicalValues	
convert_datetime	布尔值	
convert_datetime_class	POSIXct POSIXlt	
convert_missing	布尔值	

sample 节点属性



“样本”节点用于选择一部分记录。支持各种样本类型，包括分层、聚类和非随机（结构化）样本。采样对于提高性能以及选择相关记录组或事务组进行分析十分有用。

表 52. *sample* 节点属性.

sample 节点属性	数据类型	属性描述
method	Simple Complex	
mode	Include Discard	包括或丢弃满足指定条件的记录。
sample_type	First OneInN RandomPct	指定抽样方法。
first_n	整数	将包括或丢弃直到指定截止点的记录。
one_in_n	数字	每隔 <i>n-1</i> 条记录包括或丢弃一条记录。
rand_pct	数字	指定要包括或丢弃记录的百分比。
use_max_size	布尔值	启用使用 <code>maximumiscard_records_below</code> 设置。
maximum_size	整数	指定要包括在数据流中或丢弃的最大样本量。此选项是冗余选项，因此指定 <code>First</code> 和 <code>Include</code> 时会被禁用。
set_random_seed	布尔值	启用随机种子设置。
random_seed	整数	指定用作随机种子的值。
complex_sample_type	Random Systematic	
sample_units	Proportions Counts	
sample_size_proportions	Fixed Custom Variable	
sample_size_counts	Fixed Custom Variable	
fixed_proportions	数字	
fixed_counts	整数	
variable_proportions	字段	
variable_counts	字段	
use_min_stratum_size	布尔值	
minimum_stratum_size	整数	仅当抽取复杂样本 <code>Sample units=Proportions</code> 时才应用此选项。
use_max_stratum_size	布尔值	
maximum_stratum_size	整数	仅当抽取复杂样本 <code>Sample units=Proportions</code> 时才应用此选项。
clusters	字段	
stratify_by	[<i>field1</i> ... <i>fieldN</i>]	
specify_input_weight	布尔值	
input_weight	字段	
new_output_weight	字符串	

表 52. *sample* 节点属性 (续).

sample 节点属性	数据类型	属性描述
sizes_proportions	[[{string string value}{string string value}...]	如果 sample_units=proportions 且 sample_size_proportions=Custom, 指定层字段值每个可能组合的值。
default_proportion	数字	
sizes_counts	[[{string string value}{string string value}...]	指定层字段值每个可能的组合值。用法与 sizes_proportions 的用法相似, 但指定的是整数, 而非比例。
default_count	数字	

select 节点属性



“选择”节点根据特定条件从数据流中选择或废弃一部分记录。例如, 可以选择与特定销售区域相关的记录。

表 53. *select* 节点属性.

select 节点属性	数据类型	属性描述
mode	Include Discard	指定是包括还是丢弃选定记录。
condition	字符串	包括或丢弃记录的条件。

sort 节点属性



“排序”节点根据一个或多个字段的值按升序或降序对记录进行排序。

表 54. *sort* 节点属性.

sort 节点属性	数据类型	属性描述
keys	[[{string Ascending} \ {string Descending}]]	指定要作为排序依据的字段。如果未指定方向, 那么会使用缺省值。
default_ascending	布尔值	指定缺省排序顺序。
use_existing_keys	布尔值	指定是否使用以前已排序字段的排序顺序来优化现在的排序。
existing_keys		指定已排序的字段及其排序方向。使用的格式与 keys 属性相同。

streamingts 节点属性



“流式 TS”节点在某个步骤中构建时间序列模型并对其进行评估，而不需要“时间区间”节点。

表 55. streamingts 节点属性.

streamingts 节点属性	数据类型	属性描述
custom_fields	布尔值	如果 custom_fields=false, 那么将使用上游“类型”节点的当前设置。如果 custom_fields=true, 那么必须指定 targets 和 inputs。
targets	[field1...fieldN]	
inputs	[field1...fieldN]	
method	ExpertModeler Exsmooth Arima	
calculate_conf	布尔值	
conf_limit_pct	real	
use_time_intervals_node	布尔值	如果 use_time_intervals_node=true, 那么将使用上游“时间区间”节点的设置。否则, 必须指定 interval_offset_position、interval_offset 和 interval_type。
interval_offset_position	LastObservation LastRecord	LastObservation 是指最后一个有效观测值。LastRecord 是指从最后一个记录计数。
interval_offset	number	
interval_type	Periods Years Quarters Months WeeksNonPeriodic DaysNonPeriodic HoursNonPeriodic MinutesNonPeriodic SecondsNonPeriodic	
events	字段	
expert_modeler_method	AllModels Exsmooth Arima	
consider_seasonal	布尔值	
detect_outliers	布尔值	
expert_outlier_additive	布尔值	
expert_outlier_level_shift	布尔值	
expert_outlier_innovational	布尔值	
expert_outlier_transient	布尔值	
expert_outlier_seasonal_additive	布尔值	
expert_outlier_local_trend	布尔值	
expert_outlier_additive_patch	布尔值	

表 55. *streamingts* 节点属性 (续).

streamingts 节点属性	数据类型	属性描述
exsmooth_model_type	SimpleHoltsLinearTrend BrownsLinearTrend DampedTrend SimpleSeasonal WintersAdditive WintersMultiplicative	
exsmooth_transformation_type	None SquareRoot NaturalLog	
arma_p	integer	对于“时间序列”建模节点是同一属性
arma_d	integer	对于“时间序列”建模节点是同一属性
arma_q	integer	对于“时间序列”建模节点是同一属性
arma_sp	integer	对于“时间序列”建模节点是同一属性
arma_sd	integer	对于“时间序列”建模节点是同一属性
arma_sq	integer	对于“时间序列”建模节点是同一属性
arma_transformation_type	None SquareRoot NaturalLog	对于“时间序列”建模节点是同一属性
arma_include_constant	布尔值	对于“时间序列”建模节点是同一属性
tf_arma_p.fieldname	integer	对于“时间序列”建模节点是同一属性。用于转换函数。
tf_arma_d.fieldname	integer	对于“时间序列”建模节点是同一属性。用于转换函数。
tf_arma_q.fieldname	integer	对于“时间序列”建模节点是同一属性。用于转换函数。
tf_arma_sp.fieldname	integer	对于“时间序列”建模节点是同一属性。用于转换函数。
tf_arma_sd.fieldname	integer	对于“时间序列”建模节点是同一属性。用于转换函数。
tf_arma_sq.fieldname	integer	对于“时间序列”建模节点是同一属性。用于转换函数。
tf_arma_delay.fieldname	integer	对于“时间序列”建模节点是同一属性。用于转换函数。
tf_arma_transformation_type.fieldname	None SquareRoot NaturalLog	
arma_detect_outlier_mode	None Automatic	
arma_outlier_additive	布尔值	
arma_outlier_level_shift	布尔值	
arma_outlier_innovational	布尔值	
arma_outlier_transient	布尔值	
arma_outlier_seasonal_additive	布尔值	
arma_outlier_local_trend	布尔值	
arma_outlier_additive_patch	布尔值	
deployment_force_rebuild	布尔值	

表 55. *streamingts* 节点属性 (续).

streamingts 节点属性	数据类型	属性描述
deployment_rebuild_mode	Count Percent	
deployment_rebuild_count	<i>number</i>	
deployment_rebuild_pct	<i>number</i>	
deployment_rebuild_field	<字段>	

第 11 章 字段操作节点属性

anonymize 节点属性



“匿名化”节点用于转换字段名和字段值在下游的表示方式，从而掩饰原始数据。如果要允许其他用户构建含有敏感数据（例如客户名称或其他详细信息）的模型，那么这种节点十分有用。

表 56. anonymize 节点属性.

anonymize 节点属性	数据类型	属性描述
enable_anonymize	布尔值	设置为 T 时，可激活匿名化字段值（相当于在“匿名值”列中为该字段选择是）。
use_prefix	布尔值	设置为 T 时，如果已指定定制前缀，那么将使用该前缀。适用于将通过杂凑法被匿名化的字段，而且相当于在“替换值”对话框中为该字段选择定制单选按钮。
prefix	字符串	相当于在“替换值”对话框的文本框中输入前缀。如果未指定其他任何值，那么缺省前缀即是该缺省值。
transformation	Random Fixed	确定通过转换法匿名化的字段的转换参数是随机的还是固定的。
set_random_seed	布尔值	设置为 T 时，将使用指定的种子值（如果转换也设置为 Random）。
random_seed	整数	当 set_random_seed 设置为 T 时，该值是随机数的种子。
scale	数字	当转换设置为“固定”时，该值用于“转换尺度”。通常，最大尺度值为 10，但可能会被减小以避免溢出。
translate	数字	当转换设置为“固定”时，该值用于“转换”。通常，最大转换值为 1000，但可能会被减小以避免溢出。

autodataprep 节点属性



“自动数据准备 (ADP)”节点可分析您的数据并标识修正，筛选出存在问题或可能无用的字段，并在适当的情况下派生新的属性，通过智能筛选和抽样技术改进性能。您可以采用完全自动化方式使用此节点，从而允许此节点选择并应用修订，另外也可以在应用修订前预览更改并根据需要接受、拒绝或进行修改。

表 57. autodataprep 节点属性.

autodataprep 节点属性	数据类型	属性描述
objective	Balanced Speed Accuracy Custom	

表 57. autodataprep 节点属性 (续).

autodataprep 节点属性	数据类型	属性描述
custom_fields	布尔值	如果为 true, 那么允许您为当前节点指定目标字段、输入字段和其他字段。如果为 false, 那么将使用上游“类型”节点的当前设置。
target	字段	指定单个目标字段。
inputs	[field1 ... fieldN]	模型所使用的输入字段或预测变量字段。
use_frequency	布尔值	
frequency_field	字段	
use_weight	布尔值	
weight_field	字段	
excluded_fields	Filter None	
if_fields_do_not_match	StopExecution ClearAnalysis	
prepare_dates_and_times	布尔值	控制对所有日期与时间字段的访问
compute_time_until_date	布尔值	
reference_date	Today Fixed	
fixed_date	日期	
units_for_date_durations	Automatic Fixed	
fixed_date_units	Years Months Days	
compute_time_until_time	布尔值	
reference_time	CurrentTime Fixed	
fixed_time	time	
units_for_time_durations	Automatic Fixed	
fixed_date_units	Hours Minutes Seconds	
extract_year_from_date	布尔值	
extract_month_from_date	布尔值	
extract_day_from_date	布尔值	
extract_hour_from_time	布尔值	
extract_minute_from_time	布尔值	
extract_second_from_time	布尔值	
exclude_low_quality_inputs	布尔值	
exclude_too_many_missing	布尔值	
maximum_percentage_missing	数字	
exclude_too_many_categories	布尔值	

表 57. autodataprep 节点属性 (续).

autodataprep 节点属性	数据类型	属性描述
maximum_number_categories	数字	
exclude_if_large_category	布尔值	
maximum_percentage_category	数字	
prepare_inputs_and_target	布尔值	
adjust_type_inputs	布尔值	
adjust_type_target	布尔值	
reorder_nominal_inputs	布尔值	
reorder_nominal_target	布尔值	
replace_outliers_inputs	布尔值	
replace_outliers_target	布尔值	
replace_missing_continuous_inputs	布尔值	
replace_missing_continuous_target	布尔值	
replace_missing_nominal_inputs	布尔值	
replace_missing_nominal_target	布尔值	
replace_missing_ordinal_inputs	布尔值	
replace_missing_ordinal_target	布尔值	
maximum_values_for_ordinal	数字	
minimum_values_for_continuous	数字	
outlier_cutoff_value	数字	
outlier_method	Replace Delete	
rescale_continuous_inputs	布尔值	
rescaling_method	MinMax ZScore	
min_max_minimum	数字	
min_max_maximum	数字	
z_score_final_mean	数字	
z_score_final_sd	数字	
rescale_continuous_target	布尔值	
target_final_mean	数字	
target_final_sd	数字	
transform_select_input_fields	布尔值	
maximize_association_with_target	布尔值	
p_value_for_merging	数字	
merge_ordinal_features	布尔值	
merge_nominal_features	布尔值	
minimum_cases_in_category	数字	
bin_continuous_fields	布尔值	
p_value_for_binning	数字	
perform_feature_selection	布尔值	

表 57. *autodataprep* 节点属性 (续).

autodataprep 节点属性	数据类型	属性描述
p_value_for_selection	数字	
perform_feature_construction	布尔值	
transformed_target_name_extension	字符串	
transformed_inputs_name_extension	字符串	
constructed_features_root_name	字符串	
years_duration_name_extension	字符串	
months_duration_name_extension	字符串	
days_duration_name_extension	字符串	
hours_duration_name_extension	字符串	
minutes_duration_name_extension	字符串	
seconds_duration_name_extension	字符串	
year_cyclical_name_extension	字符串	
month_cyclical_name_extension	字符串	
day_cyclical_name_extension	字符串	
hour_cyclical_name_extension	字符串	
minute_cyclical_name_extension	字符串	
second_cyclical_name_extension	字符串	

binning 节点属性



“分箱”节点根据一个或多个现有连续（数字范围）字段的值自动创建新的名义（集合）字段。例如，您可以将连续收入字段转换为一个包含各组收入（作为与均值之间的偏差）的新分类字段。为新字段创建分箱后，即可根据分割点生成“派生”节点。

表 58. *binning* 节点属性.

binning 节点属性	数据类型	属性描述
fields	[field1 field2 ... fieldn]	待转换的连续（数字范围）字段。可以同时多个字段进行分箱。
method	FixedWidth EqualCount Rank SDev Optimal	用于为新字段分箱（类别）确定分割点的方法。
rcalculate_bins	Always IfNecessary	指定是每次执行节点时都重新计算分箱并将数据放入相关分箱，还是仅将数据添加到现有分箱和任何已添加的新分箱。
fixed_width_name_extension	字符串	缺省扩展名为 <i>_BIN</i> 。
fixed_width_add_as	Suffix Prefix	指定是将扩展名添加到字段名末尾（后缀）还是开头（前缀）。缺省扩展名为 <i>income_BIN</i> 。
fixed_bin_method	Width Count	

表 58. binning 节点属性 (续).

binning 节点属性	数据类型	属性描述
fixed_bin_count	整数	指定用于确定新字段的固定宽度分箱（类别）数的整数。
fixed_bin_width	实数	这是用于计算分箱宽度的值（整数或实数）。
equal_count_name_extension	字符串	缺省扩展名为 <i>_TILE</i> 。
equal_count_add_as	Suffix Prefix	指定针对使用标准 p-tile 法生成的字段名使用的扩展名（后缀或前缀）。缺省扩展名为 <i>_TILE</i> 加上 <i>N</i> ，其中 <i>N</i> 是分位数。
tile4	布尔值	生成四分位数分箱，每个分箱中包含 25% 的观测值。
tile5	布尔值	生成五分位数分箱。
tile10	布尔值	生成十分位数分箱。
tile20	布尔值	生成二十分位数分箱。
tile100	布尔值	生成百分位数分箱。
use_custom_tile	布尔值	
custom_tile_name_extension	字符串	缺省扩展名为 <i>_TILEN</i> 。
custom_tile_add_as	Suffix Prefix	
custom_tile	整数	
equal_count_method	RecordCount ValueSum	RecordCount 方法是每个分箱分配相同数目的记录，而 ValueSum 方法是使分配记录后每个分箱中值的总和相等。
tied_values_method	Next Current Random	指定要输入的分箱结值数据。
rank_order	Ascending Descending	此属性包括 Ascending（最低值标记为 1）或 Descending（最高值标记为 1）。
rank_add_as	Suffix Prefix	此选项适用于排序、分数排序和百分比排序。
rank	布尔值	
rank_name_extension	字符串	缺省扩展名为 <i>_RANK</i> 。
rank_fractional	布尔值	对个案进行排秩，其中新字段的值等于排秩值除以非缺失个案的权重之和。分数排序值介于 0-1 之间。
rank_fractional_name_extension	字符串	缺省扩展名为 <i>_F_RANK</i> 。
rank_pct	布尔值	每个排秩值除以具有有效值的记录数，再乘以 100。百分比分数秩处于 1-100 范围内。
rank_pct_name_extension	字符串	缺省扩展名为 <i>_P_RANK</i> 。
sdev_name_extension	字符串	
sdev_add_as	Suffix Prefix	

表 58. binning 节点属性 (续).

binning 节点属性	数据类型	属性描述
sdev_count	One Two Three	
optimal_name_extension	字符串	缺省扩展名为 <i>_OPTIMAL</i> 。
optimal_add_as	Suffix Prefix	
optimal_supervisor_field	字段	作为监督字段选择的字段，为分箱选择的字段与之相关。
optimal_merge_bins	布尔值	指定将所有具有较小观测值计数的分箱添加到更大的相邻分箱。
optimal_small_bin_threshold	整数	
optimal_pre_bin	布尔值	表示要进行数据集的预分箱。
optimal_max_bins	整数	指定上限以避免创建过大分箱数。
optimal_lower_end_point	Inclusive Exclusive	
optimal_first_bin	Unbounded Bounded	
optimal_last_bin	Unbounded Bounded	

derive 节点属性



“派生”节点修改数据值或者根据一个或多个现有字段创建新字段。它可以创建类型为公式、标志、名义、状态、计数和条件的字段。

表 59. derive 节点属性.

derive 节点属性	数据类型	属性描述
new_name	字符串	新字段的名称。
mode	Single Multiple	指定单个或多个字段。
fields	[field field field]	仅用于在“多重”方式下选择多个字段。
name_extension	字符串	指定新字段名的扩展名。
add_as	Suffix Prefix	将扩展名添加为字段名的前缀（开头）或后缀（末尾）。
result_type	Formula Flag Set State Count Conditional	可创建的六种类型的新字段。

表 59. derive 节点属性 (续).

derive 节点属性	数据类型	属性描述
formula_expr	字符串	这是用于在“派生”节点中计算新字段值的表达式。
flag_expr	字符串	
flag_true	字符串	
flag_false	字符串	
set_default	字符串	
set_value_cond	字符串	这是一个结构化属性，用于提供与给定值相关联的条件。
state_on_val	字符串	指定满足 On 条件时新字段的值。
state_off_val	字符串	指定满足 Off 条件时新字段的值。
state_on_expression	字符串	
state_off_expression	字符串	
state_initial	On Off	为新字段的每个记录分配初始值 On 或 Off。可在满足每个条件时更改此值。
count_initial_val	字符串	
count_inc_condition	字符串	
count_inc_expression	字符串	
count_reset_condition	字符串	
cond_if_cond	字符串	
cond_then_expr	字符串	
cond_else_expr	字符串	

ensemble 节点属性



“整体”节点对两个或两个以上模型块进行组合，这样所获得的预测比通过任意一个模型获得的预测更为准确。

表 60. ensemble 节点属性.

ensemble 节点属性	数据类型	属性描述
ensemble_target_field	字段	为在整体中使用的所有模型指定目标字段。
filter_individual_model_output	布尔值	指定是否应抑制各个模型的评分结果。
flag_ensemble_method	Voting ConfidenceWeightedVoting RawPropensityWeightedVoting AdjustedPropensityWeightedVoting HighestConfidence AverageRawPropensity AverageAdjustedPropensity	指定用于确定整体得分的方法。仅当选定的目标是标志字段时，才会应用此设置。

表 60. ensemble 节点属性 (续).

ensemble 节点属性	数据类型	属性描述
set_ensemble_method	Voting ConfidenceWeightedVoting HighestConfidence	指定用于确定整体得分的方法。仅当选定的目标是名义字段时，才会应用此设置。
flag_voting_tie_selection	Random HighestConfidence RawPropensity AdjustedPropensity	如果已选定投票方法，那么指定解决结的方法。仅当选定的目标是标志字段时，才会应用此设置。
set_voting_tie_selection	Random HighestConfidence	如果已选定投票方法，那么指定解决结的方法。仅当选定的目标是名义字段时，才会应用此设置。
calculate_standard_error	布尔值	如果目标字段是连续的，那么缺省情况下将运行标准误差计算，以计算测量值或估算值与真实值之间的差，并显示这些估算值的匹配程度。

filler 节点属性



“填充器”节点用于替换字段值并更改存储。您可以选择基于 CLEM 条件（例如 @BLANK (@FIELD)）的替换值。另外，也可以选择将所有空白值或 Null 值替换为特定值。“填充器”节点经常与“类型”节点配合使用，以替换缺失值。

表 61. filler 节点属性.

filler 节点属性	数据类型	属性描述
fields	[field field field]	数据集中其值将被检查并替换的字段。
replace_mode	Always Conditional Blank Null BlankAndNull	可以替换所有值、空白值或空值，也可以根据指定条件进行替换。
condition	字符串	
replace_with	字符串	

filter 节点属性



“过滤”节点用于过滤（废弃）字段、对字段进行重命名以及将字段从一个源节点映射到另一个节点。

使用 **default_include** 属性。请注意，设置 default_include 属性的值不会自动包括或排除所有字段，而只是确定针对当前所选字段的缺省行为。在功能上，此属性相当于单击“过滤节点”对话框中的缺省情况下包括字段按钮。

表 62. filter 节点属性.

filter 节点属性	数据类型	属性描述
default_include	布尔值	这是用于指定缺省行为是传递还是过滤字段的键控属性。请注意，设置此属性不会自动包括或排除所有字段，而只是确定缺省情况下是包括还是排除选定字段。
include	布尔值	这是用于包括和除去字段的键控属性。
new_name	字符串	

history 节点属性



“历史记录”节点创建新字段，这些字段包含先前记录中的字段的数据。“历史记录”节点最常用于顺序数据，例如时间序列数据。在使用“历史记录”节点前，可以使用“排序”节点对数据进行排序。

表 63. history 节点属性.

history 节点属性	数据类型	属性描述
fields	[field field field]	需要其历史记录的字段的。
offset	数字	指定要从中提取历史记录字段值的最新记录（当前记录之前的记录）。
span	数字	指定要从中提取值的以前记录的数目。
unavailable	Discard Leave Fill	处理不含历史记录值的记录时，通常参考没有以前的记录作为历史记录的前几个记录（位于数据集顶部）。
fill_with	字符串 数字	指定要用于无历史记录值可用的记录的或字符串。

partition 节点属性



分区节点可生成分区字段，该字段可将数据分割为单独的子集以便在模型构建的训练、测试和验证阶段使用。

表 64. partition 节点属性.

partition 节点属性	数据类型	属性描述
new_name	字符串	由节点生成的分区字段的名称。
create_validation	布尔值	指定是否应创建验证分区。
training_size	整数	要分配给训练分区的记录所占的百分比 (0-100)。
testing_size	整数	要分配给测试分区的记录所占的百分比 (0-100)。
validation_size	整数	要分配给验证分区的记录所占的百分比 (0-100)。如果未创建验证分区，那么忽略此属性。
training_label	字符串	训练分区的标签。

表 64. *partition* 节点属性 (续).

partition 节点属性	数据类型	属性描述
testing_label	字符串	测试分区的标签。
validation_label	字符串	验证分区的标签。如果未创建验证分区，那么忽略此属性。
value_mode	System SystemAndLabel Label	指定用于表示数据中每个分区的值。例如，训练样本可以表示为系统整数 1、标签 Training 或二者的组合 1_Training。
set_random_seed	布尔值	指定是否应使用用户指定的随机种子。
random_seed	整数	用户指定的随机种子值。如果要使用此值，set_random_seed 必须设置为 True。
enable_sql_generation	布尔值	指定是否使用 SQL 回送以分配记录到分区。
unique_field		指定输入字段，用以确保以随机但可重复的方式将记录分配到分区。如果要使用此值，enable_sql_generation 必须设置为 True。

reclassify 节点属性



“重新分类”节点将一组分类值转换为另一组值。对于折叠类别或者进行数据重新分组以执行分析而言，重新分类非常有用。

表 65. *reclassify* 节点属性.

reclassify 节点属性	数据类型	属性描述
mode	Single Multiple	Single 对一个字段的类别进行重新分类。Multiple 将激活用于同时对多个字段进行转换的选项。
replace_field	布尔值	
field	字符串	仅在 Single 模式下使用。
new_name	字符串	仅在 Single 模式下使用。
fields	[field1 field2 ... fieldn]	仅在“多重”方式下使用。
name_extension	字符串	仅在“多重”方式下使用。
add_as	Suffix Prefix	仅在“多重”方式下使用。
reclassify	字符串	字段值的结构化属性。
use_default	布尔值	使用缺省值。
default	字符串	指定缺省值。
pick_list	[string string ... string]	允许用户导入已知新值的列表以填充表中的下拉列表。

reorder 节点属性



“字段重新排序器”节点定义用于显示下游字段的自然顺序。此顺序将影响字段在各种位置（例如表、列表和字段选择器）的显示方式。处理宽数据集时，此操作有助于使所需字段更为直观。

表 66. reorder 节点属性.

reorder 节点属性	数据类型	属性描述
mode	Custom Auto	可以自动对值进行排序，也可以指定定制顺序。
sort_by	Name Type Storage	
ascending	布尔值	
start_fields	[field1 field2 ... fieldn]	新字段插入到这些字段之后。
end_fields	[field1 field2 ... fieldn]	新字段插入到这些字段之前。

restructure 节点属性



“重构”节点将名义字段或标志字段转换为一组字段，这组字段可以使用另一字段的值进行填充。例如，给定一个名为 支付类型的字段，其值为 贷方、现金和 借方，那么将创建三个新字段（贷方、现金、借方），每个字段可能包含实际支付的值。

表 67. restructure 节点属性.

restructure 节点属性	数据类型	属性描述
fields_from	[category category category] all	
include_field_name	布尔值	指示是否在重新结构化的字段名中使用字段名。
value_mode	OtherFields Flags	表示用于为重新结构化字段指定值的模式。如果选择 OtherFields，必须指定要使用哪些字段（参阅下文）。如果选择 Flags，那么值为数值标志。
value_fields	[field field field]	如果 value_mode 是 OtherFields，那么此属性是必需的。指定使用哪些字段作为值字段。

rfmanalysis 节点属性



通过近因、频率和货币 (RFM) 分析节点，您可以检查客户最近一次购买您产品或服务的时间（近因）、客户购买的频率（频率）以及客户支付的所有交易金额（货币），确定可能成为最佳客户的数量。

表 68. rfmanalysis 节点属性.

rfmanalysis 节点属性	数据类型	属性描述
recency	字段	指定近因字段。它有可能是日期、时间戳记或简单的数值。
frequency	字段	指定频率字段。
monetary	字段	指定货币字段。
recency_bins	整数	指定要生成的近因分箱数量。
recency_weight	数字	指定应用于近因数据的权重。缺省值为 100。
frequency_bins	整数	指定要生成的频率分箱数量。
frequency_weight	数字	指定应用于频率数据的权重。缺省值为 10。
monetary_bins	整数	指定要生成的货币分箱数量。
monetary_weight	数字	指定应用于货币数据的权重。缺省值为 1。
tied_values_method	Next Current	指定要输入的分箱结值数据。
recalculate_bins	Always IfNecessary	
add_outliers	布尔值	只有当 recalculate_bins 设为 IfNecessary 时才可用。如果已设置，那么将位于下限分箱以下的记录添加到下限分箱中，并且将最高分箱以上的记录添加到最高分箱中。
binned_field	Recency Frequency Monetary	
recency_thresholds	value value	只有当 recalculate_bins 设为 Always 时才可用。指定近因分箱的下限阈值和上限阈值。一个分箱的上限阈值用作下一个分箱的下限阈值 例如，[10 30 60] 可定义两个分箱，第一个分箱的上限阈值和下限阈值分别为 10 和 30，第二个分箱的两个阈值分别为 30 和 60。
frequency_thresholds	value value	只有当 recalculate_bins 设为 Always 时才可用。
monetary_thresholds	value value	只有当 recalculate_bins 设为 Always 时才可用。

settoflag 节点属性



“设为标志”节点根据针对一个或多个名义字段定义的分类值派生多个标志字段。

表 69. settoflag 节点属性.

settoflag 节点属性	数据类型	属性描述
fields_from	[category category category] all	

表 69. settoflag 节点属性 (续).

settoflag 节点属性	数据类型	属性描述
true_value	字符串	指定设置标志时节点所使用的真值。缺省值为 T。
false_value	字符串	指定设置标志时节点所使用的假值。缺省值为 F。
use_extension	布尔值	使用扩展名作为新标志字段的后缀或前缀。
extension	字符串	
add_as	Suffix Prefix	指定所添加的扩展名是后缀还是前缀。
aggregate	布尔值	根据键字段将记录分组。如果有任何记录被设置为 true，那么会启用组中的所有标志字段。
keys	[field field field]	键字段。

statisticstransform 节点属性



“Statistics 转换”节点针对 IBM SPSS Modeler 中的数据源运行选择的 IBM SPSS Statistics 语法命令。此节点需要 IBM SPSS Statistics 的许可副本。

有关此节点属性的信息，请参阅第 215 页的『statisticstransform 节点属性』。

timeintervals 节点属性



“时间区间”节点指定区间，创建用于对时间序列数据 进行建模的标签（如果需要）。如果各个值的间隔不均匀，那么此节点可以根据需要填充值或者将值汇总，以使记录之间的区间均匀。

表 70. timeintervals 节点属性.

timeintervals 节点属性	数据类型	属性描述
interval_type	None Periods CyclicPeriods Years Quarters Months DaysPerWeek DaysNonPeriodic HoursPerDay HoursNonPeriodic MinutesPerDay MinutesNonPeriodic SecondsPerDay SecondsNonPeriodic	

表 70. *timeintervals* 节点属性 (续).

timeintervals 节点属性	数据类型	属性描述
mode	Label Create	指定是要连续标记记录还是要根据指定日期、时间戳记或时间字段构建序列。
field	字段	当根据数据构建序列时，指定表示每个记录的日期或时间的字段。
period_start	整数	指定周期或循环周期的起始区间
cycle_start	整数	循环周期的起始周期。
year_start	整数	对于适用的区间类型，指第一个区间所属的年份。
quarter_start	整数	对于适用的区间类型，指第一个区间所属的季度。
month_start	January February March April May June July August September October November December	
day_start	整数	
hour_start	整数	
minute_start	整数	
second_start	整数	
periods_per_cycle	整数	对于循环周期，指每个周期中的期间数。
fiscal_year_begins	January February March April May June July August September October November December	对于季度区间，指定财政年度开始的月份。
week_begins_on	Sunday Monday Tuesday Wednesday Thursday Friday Saturday Sunday	对于周期性区间（一周中的天、一天中的小时、一天中的分钟和一天中的秒），指定一周开始的那一天。

表 70. *timeintervals* 节点属性 (续).

timeintervals 节点属性	数据类型	属性描述
day_begins_hour	整数	对于周期性区间（一天中的小时、一天中的分钟和一天中的秒），指定一天开始的小时。可以与 day_begins_minute 和 day_begins_second 结合起来使用，以指定一个准确时间，例如 8:05:01。请参见下面的使用示例。
day_begins_minute	整数	对于周期性区间（一天中的小时、一天中的分钟和一天中的秒），指定一天开始的分钟（例如 8:05中的 5）。
day_begins_second	整数	对于周期性区间（一天中的小时、一天中的分钟和一天中的秒），指定一天开始的秒（例如 8:05:17 中的 17）。
days_per_week	整数	对于周期性区间（一周中的天、一天中的小时、一天中的分钟和一天中的秒），指定一周中的天数。
hours_per_day	整数	对于周期性区间（一天中的小时、一天中的分钟和一天中的秒），指定一天中的小时数。
interval_increment	1 2 3 4 5 6 10 15 20 30	对于一天中的分钟和一天中的秒，指定为每个记录增加的分钟数或秒数。
field_name_extension	字符串	
field_name_extension_as_prefix	布尔值	

表 70. *timeintervals* 节点属性 (续).

timeintervals 节点属性	数据类型	属性描述
date_format	"DDMMYY" "MMDDYY" "YYMMDD" "YYYYMMDD" "YYYYDDD" DAY MONTH "DD-MM-YY" "DD-MM-YYYY" "MM-DD-YY" "MM-DD-YYYY" "DD-MON-YY" "DD-MON-YYYY" "YYYY-MM-DD" "DD.MM.YY" "DD.MM.YYYY" "MM.DD.YY" "MM.DD.YYYY" "DD.MON.YY" "DD.MON.YYYY" "DD/MM/YY" "DD/MM/YYYY" "MM/DD/YY" "MM/DD/YYYY" "DD/MON/YY" "DD/MON/YYYY" MON YYYY q Q YYYY ww WK YYYY	
time_format	"HHMMSS" "HHMM" "MMSS" "HH:MM:SS" "HH:MM" "MM:SS" "(H)H:(M)M:(S)S" "(H)H:(M)M" "(M)M:(S)S" "HH.MM.SS" "HH.MM" "MM.SS" "(H)H.(M)M.(S)S" "(H)H.(M)M" "(M)M.(S)S"	
aggregate	Mean Sum Mode Min Max First Last TrueIfAnyTrue	指定字段的汇总方法。
pad	Blank MeanOfRecentPoints true False	指定字段的填充方法。

表 70. *timeintervals* 节点属性 (续).

timeintervals 节点属性	数据类型	属性描述
agg_mode	All Specify	指定是根据需要使⽤缺省函数汇总或填充所有字段，还是指定要使⽤的字段和函数。
agg_range_default	Mean Sum Mode Min Max	指定汇总连续字段时要使⽤的缺省函数。
agg_set_default	Mode First Last	指定汇总名义字段时要使⽤的缺省函数。
agg_flag_default	TrueIfAnyTrue Mode First Last	
pad_range_default	Blank MeanOfRecentPoints	指定填充连续字段时要使⽤的缺省函数。
pad_set_default	Blank MostRecentValue	
pad_flag_default	Blank true False	
max_records_to_create	整数	指定填充序列时要创建的最大记录数。
estimation_from_beginning	布尔值	
estimation_to_end	布尔值	
estimation_start_offset	整数	
estimation_num_holdouts	整数	
create_future_records	布尔值	
num_future_records	整数	
create_future_field	布尔值	
future_field_name	字符串	

transpose 节点属性



“转置”节点交换行和列中的数据，以便记录变成字段，字段变成记录。

表 71. *transpose* 节点属性.

transpose 节点属性	数据类型	属性描述
transposed_names	Prefix Read	可以根据指定前缀⾃动生成新字段名，也可以从数据的现有字段中读取新字段名。
prefix	字符串	

表 71. transpose 节点属性 (续).

transpose 节点属性	数据类型	属性描述
num_new_fields	整数	使用前缀时, 指定要创建的新字段的最大数目。
read_from_field	字段	从中读取名称的字段。此字段必须是一个实例化字段, 否则执行节点时将出错。
max_num_fields	整数	当从某个字段中读取名称时, 指定上限以避免创建过大的字段数。
transpose_type	数字 字符串 Custom	缺省情况下, 只能转置连续 (数字范围) 字段, 但也可以选择数字字段的定制子集或转置所有字符串字段。
transpose_fields	[field field field]	指定使用定制选项时转置的字段。
id_field_name	字段	

type 节点属性



“类型”节点指定字段元数据和属性。例如, 您可以指定每个字段的测量级别 (连续、名义、有序或标志)、设置用于处理缺失值和系统 Null 值的选项、设置用于建模的字段的角色、指定字段标签和值标签以及为字段指定值。

注意, 某些情况下可能需要完全实例化“类型”节点才能使其他节点正常运行, 例如, “设为标志”节点的 fields from 属性。您可以只连接“表”节点并执行该节点以对这些字段进行实例化。

表 72. type 节点属性.

type 节点属性	数据类型	属性描述
direction	Input 目标 (Target) Both 无 (None) 分区 Split 频率 (Frequency) RecordID	字段角色的键控属性。 注意: 值 In 和 Out 现已不推荐使用。 在将来的版本中可能取消对这些值的支持。
type	范围 (Range) 标记 集合 (Set) 无类型 Discrete OrderedSet Default	字段的测量级别 (以前称为字段的“类型”)。 将 type 设置为 Default 会清除所有的 values 参数设置, 如果 value_mode 的值为 Specify, 那么其值将被重置为 Read。如果将 value_mode 设置为 Pass 或 Read, 那么设置 type 将不会影响 value_mode。 注意: 内部使用的数据类型与类型节点中的那些变量不同。对应关系如下: Range -> Continuous Set -> Nominal OrderedSet -> Ordinal Discrete- -> Categorical

表 72. type 节点属性 (续).

type 节点属性	数据类型	属性描述
storage	未知 字符串 整数 实数 时间 日期 时间戳记	字段存储类型的只读键控属性。
check	无 (None) Nullify Coerce Discard Warn Abort	字段类型和范围检查的键控属性。
values	[value value]	对于连续型字段而言, 第一个是最小值, 后一个是最大值。对于名义字段, 指定所有值。对标志字段而言, 第一个值代表 <i>false</i> , 后一个值代表 <i>true</i> 。设置该属性将自动把 <i>value_mode</i> 属性设置为 <i>Specify</i> 。
value_mode	Read Pass 读取+ 当前 Specify	确定值的设置方式。注意, 不能将此属性直接设置为 <i>Specify</i> ; 要使用特定值, 需设置 <i>values</i> 属性。
extend_values	<i>boolean</i>	当 <i>value_mode</i> 设置为 <i>Read</i> 时将应用。设为 <i>T</i> 则将新读取的值添加到任意现有字段值。设置为 <i>F</i> 则丢弃现有值并添加新读取值。
enable_missing	<i>boolean</i>	当设置为 <i>T</i> 时, 那么激活对字段缺失值的跟踪。
missing_values	[value value ...]	指定表示缺失数据的数据值。
range_missing	<i>boolean</i>	指定是否为字段定义缺失值 (空白) 范围。
missing_lower	<i>string</i>	<i>range_missing</i> 为 <i>true</i> 时, 此属性指定缺失值范围的下限。
missing_upper	<i>string</i>	<i>range_missing</i> 为 <i>true</i> 时, 此属性指定缺失值范围的上限。
null_missing	<i>boolean</i>	设置为 <i>T</i> 时, 空 (在软件中显示为 <i>\$null\$</i> 的未定义值) 被视为缺失值。
whitespace_missing	<i>boolean</i>	设置为 <i>T</i> 时, 仅包含空白 (空格、制表符和换行符) 的值被视为缺失值。
description	<i>string</i>	为字段指定说明。
value_labels	[[Value LabelString] { Value LabelString} ...]	用于为值对指定标签。
display_places	<i>integer</i>	为字段设置显示的小数位数 (仅用于以 <i>REAL</i> 存储的字段)。值为 <i>-1</i> 时, 将使用流缺省值。
export_places	<i>integer</i>	为字段设置导出时的小数位数 (仅用于以 <i>REAL</i> 存储的字段)。值为 <i>-1</i> 时, 将使用流缺省值。

表 72. type 节点属性 (续).

type 节点属性	数据类型	属性描述
decimal_separator	DEFAULT PERIOD COMMA	为字段设置十进制分隔符（仅用于以 REAL 存储的字段）。
date_format	"DDMMYY" "MMDDYY" "YYMMDD" "YYYYMMDD" "YYYYDDD" DAY MONTH "DD-MM-YY" "DD-MM-YYYY" "MM-DD-YY" "MM-DD-YYYY" "DD-MON-YY" "DD-MON-YYYY" "YYYY-MM-DD" "DD.MM.YY" "DD.MM.YYYY" "MM.DD.YY" "MM.DD.YYYY" "DD.MON.YY" "DD.MON.YYYY" "DD/MM/YY" "DD/MM/YYYY" "MM/DD/YY" "MM/DD/YYYY" "DD/MON/YY" "DD/MON/YYYY" MON YYYY q Q YYYY ww WK YYYY	为字段设置日期格式（仅用于以 DATE 或 TIME-STAMP 存储的字段）。
time_format	"HHMMSS" "HHMM" "MMSS" "HH:MM:SS" "HH:MM" "MM:SS" "(H)H:(M)M:(S)S" "(H)H:(M)M" "(M)M:(S)S" "HH.MM.SS" "HH.MM" "MM.SS" "(H)H.(M)M.(S)S" "(H)H.(M)M" "(M)M.(S)S"	为字段设置时间格式（仅用于以 TIME 或 TIME-STAMP 存储的字段）。
number_format	DEFAULT STANDARD SCIENTIFIC CURRENCY	为字段设置数字显示格式。
standard_places	<i>integer</i>	为字段设置以标准格式显示时的小数位数。值为 -1 时，将使用流缺省值。请注意，现有的 display_places 通道也会更改此属性，但目前已不再使用。
scientific_places	<i>integer</i>	为字段设置以科学计数格式显示时的小数位数。值为 -1 时，将使用流缺省值。

表 72. type 节点属性 (续).

type 节点属性	数据类型	属性描述
currency_places	<i>integer</i>	为字段设置以货币格式显示时的小数位数。值为 -1 时，将使用流缺省值。
grouping_symbol	DEFAULT NONE LOCALE PERIOD COMMA SPACE	为字段设置分组符号。
column_width	<i>integer</i>	为字段设置列宽度。值为 -1 标识将列宽度设置为 Auto。
justify	AUTO CENTER LEFT RIGHT	为字段设置列对齐格式。

第 12 章 图形节点属性

图形节点公共属性

本节介绍图形节点的可用属性，包括公共属性和每种节点类型特有的属性。

表 73. 公共图形节点属性.

公共图形节点属性	数据类型	属性描述
title	字符串	指定标题。示例: "This is a title."
caption	字符串	指定文字说明。示例: "This is a caption."
output_mode	Screen File	指定是显示图形节点的输出还是将其写到文件中。
output_format	BMP JPEG PNG HTML output (.cou)	指定输出类型。允许每个节点使用的确切输出类型是不同的。
full_filename	字符串	为从图形节点生成的输出指定目标路径和文件名。
use_graph_size	布尔值	控制是否使用下面的宽度和高度属性明确调整图形的大小。只影响输出到屏幕的图形。不适用于“分布”节点。
graph_width	数字	当 use_graph_size 为 True 时，以像素为单位设置图形宽度。
graph_height	数字	当 use_graph_size 为 True 时，以像素为单位设置图形高度。

附注

关闭可选字段。 通过将属性值设置为 " " (空字符串)，可以关闭可选字段 (例如图的交叠字段)。

指定颜色。 使用十六进制字符串 (从井号 (#) 开始)，可指定标题、标注、背景和标签的颜色。

前两位数指定红色内容; 中间两位数指定绿色内容; 最后两位数指定蓝色内容。每位数可获取一个位于范围 0-9 或 A-F 内的值。这些值在一起可以指定红-绿-蓝 (即 RGB) 颜色。

注意: 以 RGB 形式指定颜色时，可以使用用户界面中的字段选择器确定正确的颜色代码。只需将鼠标停留在颜色上面就可激活含所需信息的工具提示。

collection 节点属性



“收集”节点显示一个数字字段的值相对于另一个数字字段的值的分布。(它创建类似于直方图的图形。) 图示说明值不断变化的变量或字段时，它是有用的。使用 3-D 图形表示时，还可以包括一个按类别显示分布的符号轴。

表 74. collection 节点属性.

collection 节点属性	数据类型	属性描述
over_field	字段	
over_label_auto	布尔值	
over_label	字符串	
collect_field	字段	
collect_label_auto	布尔值	
collect_label	字符串	
three_D	布尔值	
by_field	字段	
by_label_auto	布尔值	
by_label	字符串	
operation	Sum Mean Min Max SDev	
color_field	字符串	
panel_field	字符串	
animation_field	字符串	
range_mode	Automatic UserDefined	
range_min	数字	
range_max	数字	
bins	ByNumber ByWidth	
num_bins	数字	
bin_width	数字	
use_grid	布尔值	
graph_background	颜色	本节在开头处介绍了标准图形颜色。
page_background	颜色	本节在开头处介绍了标准图形颜色。

distribution 节点属性



“分布”节点显示符号（分类）值（例如抵押类型或性别）的出现次数。通常，您可以使用“分布”节点来显示数据中的不平衡，然后可以在创建模型前使用“均衡”节点来纠正此类不平衡。

表 75. distribution 节点属性.

distribution 节点属性	数据类型	属性描述
plot	SelectedFields Flags	
x_field	字段	

表 75. *distribution* 节点属性 (续).

distribution 节点属性	数据类型	属性描述
color_field	字段	交叠字段。
normalize	布尔值	
sort_mode	ByOccurrence Alphabetic	
use_proportional_scale	布尔值	

evaluation 节点属性



“评估”节点有助于评估和比较预测模型。评估图表显示模型预测特定结果的优劣程度。它根据预测值和预测置信度对记录进行排序。它将记录分成若干个相同大小的组（分位数），然后从高到底为每个分位数划分业务标准值。在散点图中，以不同的线条显示多个模型。

表 76. *evaluation* 节点属性.

evaluation 节点属性	数据类型	属性描述
chart_type	Gains Response Lift Profit ROI ROC	
inc_baseline	布尔值	
field_detection_method	Metadata Name	
use_fixed_cost	布尔值	
cost_value	数字	
cost_field	字符串	
use_fixed_revenue	布尔值	
revenue_value	数字	
revenue_field	字符串	
use_fixed_weight	布尔值	
weight_value	数字	
weight_field	字段	
n_tile	Quartiles Quintiles Deciles Vingtiles Percentiles 1000-tiles	
cumulative	标志	
style	Line Point	

表 76. *evaluation* 节点属性 (续).

evaluation 节点属性	数据类型	属性描述
point_type	Rectangle Dot Triangle Hexagon Plus Pentagon Star BowTie HorizontalDash VerticalDash IronCross Factory House Cathedral OnionDome ConcaveTriangle OblateGlobe CatEye FourSidedPillow RoundRectangle Fan	
export_data	布尔值	
data_filename	字符串	
delimiter	字符串	
new_line	布尔值	
inc_field_names	布尔值	
inc_best_line	布尔值	
inc_business_rule	布尔值	
business_rule_condition	字符串	
plot_score_fields	布尔值	
score_fields	[field1 ... fieldN]	
target_field	字段	
use_hit_condition	布尔值	
hit_condition	字符串	
use_score_expression	布尔值	
score_expression	字符串	
caption_auto	布尔值	

graphboard 节点属性



“图形板”节点在单个节点中提供许多不同类型的图形。使用此节点，可以选择要探索的数据字段，然后从适用于选定数据的字段中选择一个图形。此节点自动过滤掉所有不适用于字段选项的图形类型。

注：如果您设置对图形类型无效的属性（例如，为直方图指定 `y_field`），那么将忽略该属性。

表 77. graphboard 节点属性.

graphboard 节点属性	数据类型	属性描述
graph_type	2DDotplot 3DArea 3DBar 3DDensity 3DHistogram 3DPie 3DScatterplot Area ArrowMap Bar BarCounts BarCountsMap BarMap BinnedScatter Boxplot Bubble ChoroplethMeans ChoroplethMedians ChoroplethSums ChoroplethValues ChoroplethCounts CoordinateMap CoordinateChoroplethMeans CoordinateChoroplethMedians CoordinateChoroplethSums CoordinateChoroplethValues CoordinateChoroplethCounts Dotplot Heatmap HexBinScatter Histogram Line LineChartMap LineOverlayMap Parallel Path Pie PieCountMap PieCounts PieMap PointOverlayMap PolygonOverlayMap Ribbon Scatterplot SPLOM Surface	标识图形类型。
x_field	字段	为 x 轴指定定制标签。只适用于标签。
y_field	字段	为 y 轴指定定制标签。只适用于标签。

表 77. graphboard 节点属性 (续).

graphboard 节点属性	数据类型	属性描述
z_field	字段	用于某些 3-D 图。
color_field	字段	在热图中使用。
size_field	字段	在气泡散点图中使用。
categories_field	字段	
values_field	字段	
rows_field	字段	
columns_field	字段	
fields	字段	
start_longitude_field	字段	与参考图中的箭头配合使用。
end_longitude_field	字段	
start_latitude_field	字段	
end_latitude_field	字段	
data_key_field	字段	用于各种图。
panelrow_field	字符串	
panelcol_field	字符串	
animation_field	字符串	
longitude_field	字段	与图中的坐标配合使用。
latitude_field	字段	
map_color_field	字段	

histogram 节点属性



“直方图”节点显示数字字段的值的出现次数。此节点经常用来在进行数据操作和模型构建前探索数据。与“分布”节点相似，“直方图”节点经常用来揭示数据中的不平衡。

表 78. histogram 节点属性.

histogram 节点属性	数据类型	属性描述
field	字段	
color_field	字段	
panel_field	字段	
animation_field	字段	
range_mode	Automatic UserDefined	
range_min	数字	
range_max	数字	
bins	ByNumber ByWidth	
num_bins	数字	
bin_width	数字	

表 78. histogram 节点属性 (续).

histogram 节点属性	数据类型	属性描述
normalize	布尔值	
separate_bands	布尔值	
x_label_auto	布尔值	
x_label	字符串	
y_label_auto	布尔值	
y_label	字符串	
use_grid	布尔值	
graph_background	颜色	本节在开头处介绍了标准图形颜色。
page_background	颜色	本节在开头处介绍了标准图形颜色。
normal_curve	布尔值	指出是否应在输出中显示正态分布曲线。

multiplot 节点属性



“多重散点图”节点创建在单个 X 字段上显示多个 Y 字段的散点图。Y 字段被绘制为彩色的线；每条线相当于“样式”设置为线且“X 模式”设置为排序的散点图节点。在探索多个变量随时间推移的变化情况时，多重散点图非常有用。

表 79. multiplot 节点属性.

multiplot 节点属性	数据类型	属性描述
x_field	字段	
y_fields	[field field field]	
panel_field	字段	
animation_field	字段	
normalize	布尔值	
use_overlay_expr	布尔值	
overlay_expression	字符串	
records_limit	数字	
if_over_limit	PlotBins PlotSample PlotAll	
x_label_auto	布尔值	
x_label	字符串	
y_label_auto	布尔值	
y_label	字符串	
use_grid	布尔值	
graph_background	颜色	本节在开头处介绍了标准图形颜色。
page_background	颜色	本节在开头处介绍了标准图形颜色。

plot 节点属性



散点图节点可显示数字字段间的关系。可通过使用点（散点）或线创建散点图。

表 80. plot 节点属性.

plot 节点属性	数据类型	属性描述
x_field	字段	为 x 轴指定定制标签。只适用于标签。
y_field	字段	为 y 轴指定定制标签。只适用于标签。
three_D	布尔值	为 y 轴指定定制标签。只适用于 3 维图形中的标签。
z_field	字段	
color_field	字段	交叠字段。
size_field	字段	
shape_field	字段	
panel_field	字段	指定用于为每个类别绘制单独图表的名义字段或标志字段。图表一起平铺在一个输出窗口中。
animation_field	字段	指定以图说明数据值类别（通过使用动画创建一系列按顺序显示的图表来说明）时所使用的名义字段或标志字段。
transp_field	字段	指定以图说明数据值类别（通过为每个类别使用不同级别的透明度来说明）时所使用的字段。不适用于线散点图。
overlay_type	None Smoother Function	指定是显示重叠函数还是 LOESS 平滑器。
overlay_expression	字符串	指定当 overlay_type 设置为 Function 时使用的表达式。
style	Point Line	
point_type	Rectangle Dot Triangle Hexagon Plus Pentagon Star BowTie HorizontalDash VerticalDash IronCross Factory House Cathedral OnionDome ConcaveTriangle OblateGlobe CatEye FourSidedPillow RoundRectangle Fan	

表 80. plot 节点属性 (续).

plot 节点属性	数据类型	属性描述
x_mode	Sort Overlay AsRead	
x_range_mode	Automatic UserDefined	
x_range_min	数字	
x_range_max	数字	
y_range_mode	Automatic UserDefined	
y_range_min	数字	
y_range_max	数字	
z_range_mode	Automatic UserDefined	
z_range_min	数字	
z_range_max	数字	
jitter	布尔值	
records_limit	数字	
if_over_limit	PlotBins PlotSample PlotAll	
x_label_auto	布尔值	
x_label	字符串	
y_label_auto	布尔值	
y_label	字符串	
z_label_auto	布尔值	
z_label	字符串	
use_grid	布尔值	
graph_background	颜色	本节在开头处介绍了标准图形颜色。
page_background	颜色	本节在开头处介绍了标准图形颜色。
use_overlay_expr	布尔值	该属性已由 overlay_type 替代。

timeplot 节点属性



“时间散点图”节点显示一组或多组时间序列数据。通常情况下，您首先要使用“时间区间”节点创建一个 *TimeLabel* 字段，该字段用于为 *x* 轴设置标签。

表 81. timeplot 节点属性.

timeplot 节点属性	数据类型	属性描述
plot_series	Series Models	

表 81. timeplot 节点属性 (续).

timeplot 节点属性	数据类型	属性描述
use_custom_x_field	布尔值	
x_field	字段	
y_fields	[field field field]	
panel	布尔值	
normalize	布尔值	
line	布尔值	
points	布尔值	
point_type	Rectangle Dot Triangle Hexagon Plus Pentagon Star BowTie HorizontalDash VerticalDash IronCross Factory House Cathedral OnionDome ConcaveTriangle OblateGlobe CatEye FourSidedPillow RoundRectangle Fan	
smoother	布尔值	只有将 panel 设置为 True , 才可将其平滑添加到散点图中。
use_records_limit	布尔值	
records_limit	整数	
symbol_size	数字	指定符号大小。
panel_layout	Horizontal Vertical	

web 节点属性



Web 节点说明两个或两个以上符号（分类）字段的值之间的关系强度。此图使用不同粗细的线条来表示连接强度。例如，您可以使用 Web 节点来探索电子商务网站上一组商品的购买之间的关系。

表 82. web 节点属性.

web 节点属性	数据类型	属性描述
use_directed_web	布尔值	
fields	[field field field]	
to_field	字段	

表 82. web 节点属性 (续).

web 节点属性	数据类型	属性描述
from_fields	[field field field]	
true_flags_only	布尔值	
line_values	Absolute OverallPct PctLarger PctSmaller	
strong_links_heavier	布尔值	
num_links	ShowMaximum ShowLinksAbove ShowAll	
max_num_links	数字	
links_above	数字	
discard_links_min	布尔值	
links_min_records	数字	
discard_links_max	布尔值	
links_max_records	数字	
weak_below	数字	
strong_above	数字	
link_size_continuous	布尔值	
web_display	Circular 网络 Directed Grid	
graph_background	颜色	本节在开头处介绍了标准图形颜色。
symbol_size	数字	指定符号大小。

第 13 章 建模节点属性

公共建模节点属性

以下属性通用于某些或所有建模节点。所有例外情况均根据需要记录在各个建模节点的文档中。

表 83. 公共建模节点属性.

属性	值	属性描述
custom_fields	布尔值	如果为 true, 那么允许您为当前节点指定目标字段、输入字段和其他字段。如果为 false, 那么将使用上游“类型”节点的当前设置。
target 或 targets	字段 或 [field1 ... fieldN]	根据模型类型指定一个目标字段或多个目标字段。
inputs	[field1 ... fieldN]	模型所使用的输入字段或预测变量字段。
partition	字段	
use_partitioned_data	布尔值	如果定义了分区字段, 那么此选项可确保仅训练分区的数据用于构建模型。
use_split_data	布尔值	
splits	[field1 ... fieldN]	指定一个或多个用于分割建模的字段。仅在 use_split_data 设置为真时有效。
use_frequency	布尔值	特定模型所使用的权重字段和频率字段 (参见每种模型类型的说明)。
frequency_field	字段	
use_weight	布尔值	
weight_field	字段	
use_model_name	布尔值	
model_name	字符串	新模型的定制名称。
mode	Simple Expert	

anomalydetection 节点属性



Anomaly Detection 节点确定不符合“正常”数据格式的异常观测值 (离群值)。即使离群值不匹配任何已知格式或用户不清楚自己的查找对象, 也可以使用此节点来确定离群值。

表 84. anomalydetection 节点属性.

anomalydetection 节点属性	值	属性描述
inputs	[field1 ... fieldN]	“异常检测”模型根据指定的输入字段对记录进行筛选。它们不使用目标字段。另外，也不使用权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
mode	Expert Simple	
anomaly_method	IndexLevel PerRecords NumRecords	指定用于将记录标记为异常的分界值的确定方法。
index_level	数字	指定用于标记异常的最小分界值。
percent_records	数字	根据训练数据中的记录百分比来设置用于标记记录的阈值。
num_records	数字	根据训练数据中的记录数目来设置用于标记记录的阈值。
num_fields	整数	针对每条异常记录报告的字段数。
impute_missing_values	布尔值	
adjustment_coeff	数字	此值用于对计算距离时赋予连续型字段和分类字段的相对权重进行平衡。
peer_group_num_auto	布尔值	自动计算对等组数。
min_num_peer_groups	整数	指定 peer_group_num_auto 设置为 True 时使用的对等组的最小数。
max_num_per_groups	整数	指定对等组的最大数目。
num_peer_groups	整数	指定 peer_group_num_auto 设置为 False 时使用的对等组数。
noise_level	数字	确定创建聚类期间处理离群值的方式。指定值必须为 0 到 0.5 之间。
noise_ratio	数字	指定分配给应该用于噪声缓存的组件的内存比例。指定值必须为 0 到 0.5 之间。

apriori 节点属性



“先验”节点从数据抽取一组规则，即抽取信息内容最多的规则。Apriori 节点提供五种选择规则的方法并使用复杂的索引模式来高效地处理大数据集。对于较大的问题，Apriori 训练的速度通常较快；它对可保留的规则数量没有任何限制，而且可处理最多带有 32 个前提条件的规则。“先验”要求输入和输出字段均为分类型字段，但因为它专为处理此类型数据而进行优化，因而处理速度快得多。

表 85. apriori 节点属性.

apriori 节点属性	值	属性描述
consequents	字段	Apriori 模型使用“结果”和“前提条件”代替标准的目标字段和输入字段。不使用权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
antecedents	[field1 ... fieldN]	
min_supp	数字	
min_conf	数字	
max_antecedents	数字	
true_flags	布尔值	
optimize	Speed Memory	
use_transactional_data	布尔值	
contiguous	布尔值	
id_field	字符串	
content_field	字符串	
mode	Simple Expert	
evaluation	RuleConfidence DifferenceToPrior ConfidenceRatio InformationDifference NormalizedChiSquare	
lower_bound	数字	
optimize	Speed Memory	用于指定模型构建是针对速度还是内存进行优化。

autoclassifier 节点属性



“自动分类器”节点用于创建和对比二元结果（是或否，流失或不流失等）的若干不同模型，使用户可以选择给定分析的最佳处理方法。由于支持多种建模算法，因此可以对用户希望使用的方法、每种方法的特定选项以及对比结果的标准进行选择。节点根据指定的选项生成一组模型并根据用户指定的标准排列最佳候选项的顺序。

表 86. autoclassifier 节点属性.

autoclassifier 节点属性	值	属性描述
target	字段	对于标志目标，“自动分类器”节点需要单个目标字段以及一个或多个输入字段。另外，也可以使用权重和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。

表 86. autoclassifier 节点属性 (续).

autoclassifier 节点属性	值	属性描述
ranking_measure	Accuracy Area_under_curve Profit Lift Num_variables	
ranking_dataset	Training Test	
number_of_models	整数	这是要包括在模型块中的模型数。指定整数必须为 1 到 100 之间。
calculate_variable_importance	布尔值	
enable_accuracy_limit	布尔值	
accuracy_limit	整数	介于 0 与 100 之间的整数。
enable_area_under_curve_limit	布尔值	
area_under_curve_limit	数字	介于 0.0 与 1.0 之间的实数。
enable_profit_limit	布尔值	
profit_limit	数字	大于 0 的整数。
enable_lift_limit	布尔值	
lift_limit	数字	这是大于 1.0 的实数。
enable_number_of_variables_limit	布尔值	
number_of_variables_limit	数字	大于 0 的整数。
use_fixed_cost	布尔值	
fixed_cost	数字	这是大于 0.0 的实数。
variable_cost	字段	
use_fixed_revenue	布尔值	
fixed_revenue	数字	这是大于 0.0 的实数。
variable_revenue	字段	
use_fixed_weight	布尔值	
fixed_weight	数字	这是大于 0.0 的实数。
variable_weight	字段	
lift_percentile	数字	介于 0 与 100 之间的整数。
enable_model_build_time_limit	布尔值	
model_build_time_limit	数字	设置为分钟数的整数，用于限制构建各个模型所花费的时间。
enable_stop_after_time_limit	布尔值	
stop_after_time_limit	数字	这是设置为小时数的实数，用于限制运行自动分类器的总耗时。
enable_stop_after_valid_model_produced	布尔值	
use_costs	布尔值	
<algorithm>	布尔值	允许或禁止使用特定算法。
<algorithm>.<property>	字符串	设置特定算法的属性值。请参阅主题第 121 页的『设置算法属性』以获取更多信息。

设置算法属性

用于自动分类器节点的算法名称有 cart、chaid、quest、c50、logreg、decisionlist、bayesnet、discriminant、svm 和 knn。

用于自动数字节点的算法名称有 cart、chaid、neuralnetwork、genlin、svm、regression、linear 和 knn。

“自动聚类”节点的算法名称有 twostep、k-means 和 kohonen。

属性名是各算法节点的文档中记录的标准名称。

包含句点或其他标点符号的算法属性必须使用单引号括起。

也可以为属性分配多个值。

注:

- 设置 true 和 false 值时，必须使用下档字符（而不是 False）。
- 如果某些算法选项在“自动分类器”节点中不可用，或者只能指定单个值而不能指定值范围，那么编写脚本时的限制与采用标准方式访问节点时的限制相同。

autocluster 节点属性



“自动聚类”节点估算和比较识别具有类似特征记录组的聚类模型。节点工作方式与其他自动建模节点相同，使您在一次建模运行中即可试验多个选项组合。模型可使用基本测量进行比较，以尝试过滤聚类模型的有效性以及对其进行排序，并提供一个基于特定字段的重要性的测量。

表 87. autocluster 节点属性.

autocluster 节点属性	值	属性描述
evaluation	字段	注：仅适用于“自动聚类”节点。标识要计算重要性值的字段。另外，可用于标识聚类对此字段的值进行区分的良好程度，从而标识模型预测此字段的良好程度。
ranking_measure	Silhouette Num_clusters Size_smallest_cluster Size_largest_cluster Smallest_to_largest Importance	
ranking_dataset	Training Test	
summary_limit	整数	要在报告中列出的模型的数目。指定整数必须为 1 到 100 之间。
enable_silhouette_limit	布尔值	
silhouette_limit	整数	介于 0 与 100 之间的整数。
enable_number_less_limit	布尔值	
number_less_limit	数字	介于 0.0 与 1.0 之间的实数。

表 87. *autocluster* 节点属性 (续).

autocluster 节点属性	值	属性描述
enable_number_greater_limit	布尔值	
number_greater_limit	数字	大于 0 的整数。
enable_smallest_cluster_limit	布尔值	
smallest_cluster_units	Percentage Counts	
smallest_cluster_limit_percentage	数字	
smallest_cluster_limit_count	整数	大于 0 的整数。
enable_largest_cluster_limit	布尔值	
largest_cluster_units	Percentage Counts	
largest_cluster_limit_percentage	数字	
largest_cluster_limit_count	整数	
enable_smallest_largest_limit	布尔值	
smallest_largest_limit	数字	
enable_importance_limit	布尔值	
importance_limit_condition	Greater_than Less_than	
importance_limit_greater_than	数字	介于 0 与 100 之间的整数。
importance_limit_less_than	数字	介于 0 与 100 之间的整数。
<algorithm>	布尔值	允许或禁止使用特定算法。
<algorithm>.<property>	字符串	设置特定算法的属性值。请参阅主题第 121 页的『设置算法属性』以获取更多信息。

autonumeric 节点属性



自动数字节点使用多种不同方法估计和对比模型的连续数字范围结果。此节点和自动分类器节点的工作方式相同，因此可以选择要使用和要在单个建模传递中使用多个选项组合进行测试的算法。受支持的算法包括神经网络、C&R 树、CHAID、线性回归、广义线性回归以及支持向量机 (SVM)。可基于相关度、相对错误或已用变量数对模型进行对比。

表 88. *autonumeric* 节点属性.

autonumeric 节点属性	值	属性描述
custom_fields	布尔值	如果为 True，将使用定制字段设置代替“类型”节点设置。
target	字段	“自动数字”节点要求单个目标字段以及一个或多个输入字段。另外，也可以使用权重和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
inputs	[field1 ... field2]	
partition	字段	
use_frequency	布尔值	

表 88. *autonumeric* 节点属性 (续).

autonumeric 节点属性	值	属性描述
frequency_field	字段	
use_weight	布尔值	
weight_field	字段	
use_partitioned_data	布尔值	如果定义了分区字段，那么仅将训练数据用于模型构建。
ranking_measure	Correlation NumberOfFields	
ranking_dataset	Test Training	
number_of_models	整数	这是要包括在模型块中的模型数。指定整数必须为 1 到 100 之间。
calculate_variable_importance	布尔值	
enable_correlation_limit	布尔值	
correlation_limit	整数	
enable_number_of_fields_limit	布尔值	
number_of_fields_limit	整数	
enable_relative_error_limit	布尔值	
relative_error_limit	整数	
enable_model_build_time_limit	布尔值	
model_build_time_limit	整数	
enable_stop_after_time_limit	布尔值	
stop_after_time_limit	整数	
stop_if_valid_model	布尔值	
<algorithm>	布尔值	允许或禁止使用特定算法。
<algorithm>.<property>	字符串	设置特定算法的属性值。请参阅主题第 121 页的『设置算法属性』以获取更多信息。

bayesnet 节点属性



通过贝叶斯网络节点，你可以利用对真实世界认知的判断力并结合所观察和记录的证据来构建概率模型。该节点重点应用了树扩展简单贝叶斯 (TAN) 和马尔可夫覆盖网络，这些算法主要用于分类问题。

表 89. *bayesnet* 节点属性.

bayesnet 节点属性	值	属性描述
inputs	<i>[field1 ... fieldN]</i>	贝叶斯网络模型使用单个目标字段以及一个或多个输入字段。连续字段将自动进行分箱。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
continue_training_existing_model	布尔值	

表 89. bayesnet 节点属性 (续).

bayesnet 节点属性	值	属性描述
structure_type	TAN MarkovBlanket	请选择要在构建贝叶斯网络时使用的结构。
use_feature_selection	布尔值	
parameter_learning_method	Likelihood Bayes	指定父节点值已知的节点之间的条件概率表的预测方法。
mode	Expert Simple	
missing_values	布尔值	
all_probabilities	布尔值	
independence	Likelihood Pearson	指定用于确定两个变量的成对观测值是否相互独立的方法。
significance_level	数字	指定用于确定独立性的分界值。
maximal_conditioning_set	数字	设置用于独立性测试的条件变量的最大数目。
inputs_always_selected	[field1 ... fieldN]	指定构建贝叶斯网络时始终使用的数据集字段。 注意: 目标字段始终处于选中状态。
maximum_number_inputs	数字	指定构建贝叶斯网络时使用的输入字段的最大数目。
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

buildr 节点属性



“R 构建”节点使您能够输入定制 R 脚本，以执行 IBM SPSS Modeler 中部署的模型构建和模型评分。

表 90. buildr 属性.

buildr 节点属性	值	属性描述
build_syntax	字符串	这是用于进行模型构建的 R 脚本语法。
score_syntax	字符串	这是用于进行模型评分的 R 脚本语法。
convert_flags	StringsAndDoubles LogicalValues	此选项用于转换标志字段。
convert_datetime	布尔值	此选项用于将具有日期或日期时间格式的变量转换为 R 日期/时间格式。
convert_datetime_class	POSIXct POSIXlt	这些选项用于指定要将日期或日期时间格式的变量转换为什么格式。
convert_missing	布尔值	此选项用于将缺失值转换为 R NA 值。

表 90. buildr 属性 (续).

buildr 节点属性	值	属性描述
output_html	布尔值	此选项用于在 R 模型块中的选项卡上显示图形。
output_text	布尔值	此选项用于将 R 控制台文本输出写至 R 模型块中的选项卡。

c50 节点属性



C5.0 节点构建决策树或规则集。该模型的工作原理是根据在每个级别提供最大信息收获的字段分割样本。目标字段必须为分类字段。允许进行多次多于两个子组的分割。

表 91. c50 节点属性.

c50 节点属性	值	属性描述
target	字段	C50 模型使用单个目标字段以及一个或多个输入字段。另外，还可以指定权重字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
output_type	DecisionTree RuleSet	
group_symbolics	布尔值	
use_boost	布尔值	
boost_num_trials	数字	
use_xval	布尔值	
xval_num_folds	数字	
mode	Simple Expert	
favor	Accuracy Generality	优先选择准确性还是通用性。
expected_noise	数字	
min_child_records	数字	
pruning_severity	数字	
use_costs	布尔值	
costs	结构化	这是结构化属性。
use_winning	布尔值	
use_global_pruning	布尔值	缺省为“启用 (True)。”
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

carma 节点属性



CARMA 模型在不要求用户指定输入或目标字段的情况下从数据抽取一组规则。与 Apriori 不同，CARMA 节点提供构建规则设置支持（前项和后项支持），而不仅仅是前项支持。这就意味着生成的规则可以用于更多应用程序，例如用于查找产品或服务（前项）的列表，这些产品或服务的后项为想在节日期间促销的商品。

表 92. carma 节点属性.

carma 节点属性	值	属性描述
inputs	[field1 ... fieldn]	CARMA 模型使用输入字段列表，但不使用目标字段。不使用权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
id_field	字段	此字段是用于进行模型构建的 ID 字段。
contiguous	布尔值	用于指定 ID 字段中的 ID 是否连续。
use_transactional_data	布尔值	
content_field	字段	
min_supp	数字（百分比）	与规则支持相关，而不是与前提条件支持相关。缺省值为 20%。
min_conf	数字（百分比）	缺省值为 20%。
max_size	数字	缺省值为 10。
mode	Simple Expert	缺省值为 Simple。
exclude_multiple	布尔值	排除具有多个结果的规则。缺省值为 False。
use_pruning	布尔值	缺省值为 False。
pruning_value	数字	缺省值为 500。
vary_support	布尔值	
estimated_transactions	整数	
rules_without_antecedents	布尔值	

cart 节点属性



分类和回归 (C&R) 树节点生成可用于预测或分类未来观测值的决策树。该方法通过在每个步骤最大限度降低不纯度，使用递归分区来将训练记录分割为组。如果树中某个节点中 100% 的观测值都属于目标字段的一个特定类别，那么该节点将被认定为“纯洁”。目标和输入字段可以是数字范围或分类（名义、有序或标志）；所有分割均为二元分割（即仅分割为两个子组）。

表 93. cart 节点属性.

cart 节点属性	值	属性描述
target	字段	C&R 树模型需要单个目标字段以及一个或多个输入字段。另外，还可以指定频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
continue_training_existing_model	布尔值	

表 93. cart 节点属性 (续).

cart 节点属性	值	属性描述
objective	Standard Boosting Bagging psm	psm 用于非常大的数据集, 同时需要 Server 连接。
model_output_type	Single InteractiveBuilder	
use_tree_directives	布尔值	
tree_directives	字符串	指定用于使树生长的伪指令。这些伪指令可以括在三重引号中, 以避免转义新行或引号。请注意, 伪指令可能对数据或建模选项的细微变化高度敏感, 并且可能无法通用于其他数据集。
use_max_depth	Default Custom	
max_depth	整数	最大树深度从 0 到 1000。只在 use_max_depth = Custom 时使用。
prune_tree	布尔值	修剪树, 以避免过度拟合。
use_std_err	布尔值	使用最大风险差值 (标准误差)。
std_err_multiplier	数字	最大差值。
max_surrogates	数字	最大代用项。
use_percentage	布尔值	
min_parent_records_pc	数字	
min_child_records_pc	数字	
min_parent_records_abs	数字	
min_child_records_abs	数字	
use_costs	布尔值	
costs	结构化	结构化属性。
priors	Data Equal Custom	
custom_priors	结构化	结构化属性。
adjust_priors	布尔值	
trails	数字	用于推进或组装的组件模型数。
set_ensemble_method	Voting HighestProbability HighestMeanProbability	分类目标的缺省组合规则。
range_ensemble_method	Mean Median	连续目标的缺省组合规则。
large_boost	布尔值	对非常大型的数据集应用推进。
min_impurity	数字	
impurity_measure	Gini Twoing Ordered	

表 93. cart 节点属性 (续).

cart 节点属性	值	属性描述
train_pct	数字	防止过度拟合集合。
set_random_seed	布尔值	复制结果选项。
seed	数字	
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

chaid 节点属性



CHAID 使用卡方统计来生成决策树，以确定最佳的分割。CHAID 与 C&R 树和 QUEST 节点不同，它可以生成非二元树，这意味着有些分割将有多于两个的分支。目标和输入字段可以是数字范围（连续）或分类。Exhaustive CHAID 是 CHAID 的修正版，它对所有分割进行更彻底的检查，但计算时间比较长。

表 94. chaid 节点属性.

chaid 节点属性	值	属性描述
target	字段	CHAID 模型需要单个目标字段以及一个或多个输入字段。另外，还可以指定频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
continue_training_existing_model	布尔值	
objective	Standard Boosting Bagging psm	psm 用于非常大的数据集，同时需要 Server 连接。
model_output_type	Single InteractiveBuilder	
use_tree_directives	布尔值	
tree_directives	字符串	
method	Chaid ExhaustiveChaid	
use_max_depth	Default Custom	
max_depth	整数	最大树深度从 0 到 1000。只在 use_max_depth = Custom 时使用。
use_percentage	布尔值	
min_parent_records_pc	数字	
min_child_records_pc	数字	
min_parent_records_abs	数字	
min_child_records_abs	数字	

表 94. chaid 节点属性 (续).

chaid 节点属性	值	属性描述
use_costs	布尔值	
costs	结构化	结构化属性。
trails	数字	用于推进或组装的组件模型数。
set_ensemble_method	Voting HighestProbability HighestMeanProbability	分类目标的缺省组合规则。
range_ensemble_method	Mean Median	连续目标的缺省组合规则。
large_boost	布尔值	对非常大的数据集应用推进。
split_alpha	数字	用于执行分割的显著性水平。
merge_alpha	数字	用于执行合并的显著性水平。
bonferroni_adjustment	布尔值	使用 Bonferroni 法调整显著性值。
split_merged_categories	布尔值	允许对合并的类别进行再分割。
chi_square	Pearson LR	这是用于计算卡方统计的方法: Pearson 或似然比
epsilon	数字	期望单元格频率的最小变化值。
max_iterations	数字	收敛的最大迭代次数。
set_random_seed	整数	
seed	数字	
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	
maximum_number_of_models	整数	

coxreg 节点属性



使用 Cox 回归节点, 您可以在已有的检查记录中建立时间事件的生存模型。该模型会生成一个生存函数, 该函数可预测在给定时间 (t) 内对于所给定的输入变量值相关事件的发生概率。

表 95. coxreg 节点属性.

coxreg 节点属性	值	属性描述
survival_time	字段	Cox 回归模型需要单个包含生存时间的字段。
target	字段	Cox 回归模型需要单个目标字段以及一个或多个输入字段。有关更多信息, 请参阅第 117 页的『公共建模节点属性』主题。

表 95. coxreg 节点属性 (续).

coxreg 节点属性	值	属性描述
method	Enter Stepwise BackwardsStepwise	
groups	字段	
model_type	MainEffects Custom	
custom_terms	["BP*Sex" "BP*Age"]	
mode	Expert Simple	
max_iterations	数字	
p_converge	1.0E-4 1.0E-5 1.0E-6 1.0E-7 1.0E-8 0	
p_converge	1.0E-4 1.0E-5 1.0E-6 1.0E-7 1.0E-8 0	
l_converge	1.0E-1 1.0E-2 1.0E-3 1.0E-4 1.0E-5 0	
removal_criterion	LR Wald Conditional	
probability_entry	数字	
probability_removal	数字	
output_display	EachStep LastStep	
ci_enable	布尔值	
ci_value	90 95 99	
correlation	布尔值	
display_baseline	布尔值	
survival	布尔值	
hazard	布尔值	
log_minus_log	布尔值	

表 95. *coxreg* 节点属性 (续).

coxreg 节点属性	值	属性描述
one_minus_survival	布尔值	
separate_line	字段	
value	数字或字符串	如果未对某个字段指定值, 那么将对该字段使用缺省选项“均值”。

decisionlist 节点属性



决策列表节点可标识子组或段, 显示与总体相关的给定二元结果的似然度的高低。例如, 您或许在寻找那些最不可能流失的客户或最有可能对某个商业活动作出积极响应的客户。通过定制段和并排预览备选模型来比较结果, 您可以将自己的业务知识体现在模型中。决策列表模型由一组规则构成, 其中每个规则具备一个条件和一个结果。规则依顺序应用, 相匹配的第一个规则将决定结果。

表 96. *decisionlist* 节点属性.

decisionlist 节点属性	值	属性描述
target	字段	“决策列表”模型使用单个目标以及一个或多个输入字段。另外, 还可以指定频率字段。有关更多信息, 请参阅第 117 页的『公共建模节点属性』主题。
model_output_type	Model InteractiveBuilder	
search_direction	Up Down	与查找段相关; 其中 Up 相当于“高概率”, 而 Down 相当于“低概率”。
target_value	字符串	如果未指定, 那么标志将采用 true 值。
max_rules	整数	除余数以外的最大段数。
min_group_size	整数	最小段大小。
min_group_size_pct	数字	最小段大小 (以百分比表示)。
confidence_level	数字	为了使输入字段符合添加到段定义的条件, 输入字段必须将响应似然度提高 (提升) 的最小阈值。
max_segments_per_rule	整数	
mode	Simple Expert	
bin_method	EqualWidth EqualCount	
bin_count	数字	
max_models_per_cycle	整数	列表的搜索宽度。
max_rules_per_cycle	整数	段规则的搜索宽度。
segment_growth	数字	
include_missing	布尔值	
final_results_only	布尔值	

表 96. *decisionlist* 节点属性 (续).

decisionlist 节点属性	值	属性描述
reuse_fields	布尔值	允许重复使用属性（出现在规则中的输入字段）。
max_alternatives	整数	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

discriminant 节点属性



判别分析所做的假设比 logistic 回归的假设更严格，但在符合这些假设时，判别分析可以作为 logistic 回归分析的有用替代项或补充。

表 97. *discriminant* 节点属性.

discriminant 节点属性	值	属性描述
target	字段	“判别”模型需要单个目标字段以及一个或多个输入字段。不使用权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
method	Enter Stepwise	
mode	Simple Expert	
prior_probabilities	AllEqual ComputeFromSizes	
covariance_matrix	WithinGroups SeparateGroups	
means	布尔值	“高级输出”对话框中的统计选项。
univariate_anovas	布尔值	
box_m	布尔值	
within_group_covariance	布尔值	
within_groups_correlation	布尔值	
separate_groups_covariance	布尔值	
total_covariance	布尔值	
fishers	布尔值	
unstandardized	布尔值	
casewise_results	布尔值	“高级输出”对话框中的分类选项。
limit_to_first	数字	缺省值为 10。
summary_table	布尔值	
leave_one_classification	布尔值	

表 97. discriminant 节点属性 (续).

discriminant 节点属性	值	属性描述
combined_groups	布尔值	
separate_groups_covariance	布尔值	矩阵选项类协方差。
territorial_map	布尔值	
combined_groups	布尔值	散点图选项 联合组。
separate_groups	布尔值	散点图选项 独立组。
summary_of_steps	布尔值	
F_pairwise	布尔值	
stepwise_method	WilksLambda UnexplainedVariance MahalanobisDistance SmallestF RaosV	
V_to_enter	数字	
criteria	UseValue UseProbability	
F_value_entry	数字	缺省值为 3.84。
F_value_removal	数字	缺省值为 2.71。
probability_entry	数字	缺省值为 0.05。
probability_removal	数字	缺省值为 0.10。
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

factor 节点属性



“PCA/因子”节点提供用于降低数据复杂程度的强大数据降维技术。主成份分析（PCA）可找出输入字段的线性组合，该组合最好地捕获了整个字段集中的方差，且组合中的各个成分相互正交（相互垂直）。因子分析则尝试识别底层因素，这些因素说明了观测的字段集合内的相关性模式。对于这两种方法，其共同的目标是找到可对原始字段集中的信息进行有效总结的少量派生字段。

表 98. factor 节点属性.

factor 节点属性	值	属性描述
inputs	[field1 ... fieldN]	“PCA/因子”模型使用输入字段列表，但不使用目标字段。不使用权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。

表 98. factor 节点属性 (续).

factor 节点属性	值	属性描述
method	PC ULS GLS ML PAF Alpha Image	
mode	Simple Expert	
max_iterations	数字	
complete_records	布尔值	
matrix	Correlation Covariance	
extract_factors	ByEigenvalues ByFactors	
min_eigenvalue	数字	
max_factor	数字	
rotation	None Varimax DirectOblimin Equamax Quartimax Promax	
delta	数字	如果选择 DirectOblimin 作为旋转数据的类型，那么可以指定 delta 的值。 如果未指定一个值，那么将使用 delta 的缺省值。
kappa	数字	如果选择 Promax 作为旋转数据的类型，那么可以指定 kappa 的值。 如果未指定一个值，那么将使用 kappa 的缺省值。
sort_values	布尔值	
hide_values	布尔值	
hide_below	数字	

featureselection 节点属性



“特征选择”节点根据一组条件（例如缺失值百分比）筛选要除去的输入字段，然后，相对于指定目标对余下的输入的重要性进行排序。例如，假如某个给定数据集有上千个潜在输入，那么哪些输入最有可能用于对患者结果进行建模呢？

表 99. *featureselection* 节点属性.

featureselection 节点属性	值	属性描述
target	字段	“特征选择”模型相对于指定的目标对预测变量进行排序。不使用权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
screen_single_category	布尔值	如果为 True，那么将筛选相对于记录总数而言同个类别中具有过多记录的字段。
max_single_category	数字	指定 screen_single_category 为 True 时使用的阈值。
screen_missing_values	布尔值	如果为 True，那么将筛选具有过多缺失值的字段，字段数表示为记录总数的百分比。
max_missing_values	数字	
screen_num_categories	布尔值	如果为 True，那么将筛选相对于记录总数而言具有过多类别的字段。
max_num_categories	数字	
screen_std_dev	布尔值	如果为 True，那么将筛选标准差小于或等于指定最小值的字段。
min_std_dev	数字	
screen_coeff_of_var	布尔值	如果为 True，那么将筛选方差系数小于或等于指定最小值的字段。
min_coeff_of_var	数字	
criteria	Pearson Likelihood CramersV Lambda	根据分类目标对分类预测变量进行排序时，指定重要性值所依据的度量。
unimportant_below	数字	指定用于将变量排序为“重要”、“边际”或“不重要”的 p 阈值。接受从 0.0 到 1.0 的值。
important_above	数字	接受从 0.0 到 1.0 的值。
unimportant_label	字符串	指定“不重要”排序的标签。
marginal_label	字符串	
important_label	字符串	
selection_mode	ImportanceLevel ImportanceValue TopN	
select_important	布尔值	在 selection_mode 设置为 ImportanceLevel 时，指定是否选择“重要”字段。
select_marginal	布尔值	在 selection_mode 设置为 ImportanceLevel 时，指定是否选择“边际”字段。
select_unimportant	布尔值	在 selection_mode 设置为 ImportanceLevel 时，指定是否选择“不重要”字段。

表 99. *featureselection* 节点属性 (续).

featureselection 节点属性	值	属性描述
importance_value	数字	在 selection_mode 设置为 ImportanceValue 时, 指定要使用的分界值。接受从 0 到 100 的值。
top_n	整数	在 selection_mode 设置为 TopN 时, 指定要使用的分界值。接受从 0 到 1000 的值。

genlin 节点属性



“广义线性”模型对一般线性模型进行了扩展, 这样因变量通过指定的关联函数与因子和协变量线性相关。而且, 该模型还允许因变量为非正态分布。它包括统计模型大部分的功能, 其中包括线性回归、logistic 回归、用于计数数据的对数线性模型以及区间删失生存模型。

表 100. *genlin* 节点属性.

genlin 节点属性	值	属性描述
target	字段	广义线性模型要求单个目标字段 (必须是一个集合或标志), 以及一个或多个输入字段。另外, 还可以指定权重字段。有关更多信息, 请参阅第 117 页的『公共建模节点属性』主题。
use_weight	布尔值	
weight_field	字段	字段类型只有“连续”。
target_represents_trials	布尔值	
trials_type	Variable FixedValue	
trials_field	字段	字段类型包括“连续”、“标志”或“有序”。
trials_number	数字	缺省值为 10。
model_type	MainEffects MainAndAllTwoWayEffects	
offset_type	Variable FixedValue	
offset_field	字段	字段类型只有“连续”。
offset_value	数字	必须是实数。
base_category	Last First	
include_intercept	布尔值	
mode	Simple Expert	

表 100. genlin 节点属性 (续).

genlin 节点属性	值	属性描述
distribution	BINOMIAL GAMMA IGAUSS NEGBIN NORMAL POISSON TWEEDIE MULTINOMIAL	IGAUSS: 逆高斯。 NEGBIN: 负二项式。
negbin_para_type	Specify Estimate	
negbin_parameter	数字	缺省值为 1。必须包含非负实数。
tweedie_parameter	数字	
link_function	IDENTITY CLOGLOG LOG LOGC LOGIT NEGBIN NLOGLOG ODDSPower PROBIT POWER CUMCAUCHIT CUMCLOGLOG CUMLOGIT CUMNLOGLOG CUMPROBIT	CLOGLOG: 互补双对数。 LOGC: 对数补数。 NEGBIN: 负二项式。 NLOGLOG: 负双对数。 CUMCAUCHIT: 累积 Cauchit。 CUMCLOGLOG: 累积互补双对数。 CUMLOGIT: 累积 Logit。 CUMNLOGLOG: 累积负双对数。 CUMPROBIT: 累积 Probit。
power	数字	值必须是非零实数。
method	Hybrid Fisher NewtonRaphson	
max_fisher_iterations	数字	缺省值为 1; 只允许使用正整数。
scale_method	MaxLikelihoodEstimate Deviance PearsonChiSquare FixedValue	
scale_value	数字	缺省值为 1; 必须大于 0。
covariance_matrix	ModelEstimator RobustEstimator	
max_iterations	数字	缺省值为 100; 只允许使用非负整数。
max_step_halving	数字	缺省值为 5; 只允许使用正整数。
check_separation	布尔值	
start_iteration	数字	缺省值为 20; 只允许使用正整数。
estimates_change	布尔值	
estimates_change_min	数字	缺省值为 1E-006; 只允许使用正数。

表 100. genlin 节点属性 (续).

genlin 节点属性	值	属性描述
estimates_change_type	Absolute Relative	
loglikelihood_change	布尔值	
loglikelihood_change_min	数字	只允许使用正数。
loglikelihood_change_type	Absolute Relative	
hessian_convergence	布尔值	
hessian_convergence_min	数字	只允许使用正数。
hessian_convergence_type	Absolute Relative	
case_summary	布尔值	
contrast_matrices	布尔值	
descriptive_statistics	布尔值	
estimable_functions	布尔值	
model_info	布尔值	
iteration_history	布尔值	
goodness_of_fit	布尔值	
print_interval	数字	缺省值为 1; 必须是正整数。
model_summary	布尔值	
lagrange_multiplier	布尔值	
parameter_estimates	布尔值	
include_exponential	布尔值	
covariance_estimates	布尔值	
correlation_estimates	布尔值	
analysis_type	TypeI TypeIII TypeIAndTypeIII	
statistics	Wald LR	
citype	Wald Profile	
tolerancelevel	数字	缺省值为 0.0001。
confidence_interval	数字	缺省值为 95。
loglikelihood_function	Full Kernel	
singularity_tolerance	1E-007 1E-008 1E-009 1E-010 1E-011 1E-012	

表 100. *genlin* 节点属性 (续).

genlin 节点属性	值	属性描述
value_order	Ascending Descending DataOrder	
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

glmm 节点属性



广义线性混合模型 (GLMM) 扩展了线性模型, 使得目标可以有非正态分布, 通过指定的关联函数与因子和协变量线性相关, 并且观测值可能相关。广义线性混合模型涵盖从简单线性回归模型到非正态纵向数据的复杂多级模型的各种模型。

表 101. *glmm* 节点属性.

glmm 节点属性	值	属性描述
residual_subject_spec	结构化	这是指定的分类字段的值组合, 此组合唯一地定义数据集中的主体。
repeated_measures	结构化	这些字段用于标识重复观测值。
residual_group_spec	[field1 ... fieldN]	这些字段用于定义重复效应协方差参数的独立集合。
residual_covariance_type	对角线 (Diagonal) AR1 ARMA11 COMPOUND_SYMMETRY IDENTITY TOEPLITZ UNSTRUCTURED VARIANCE_COMPONENTS	指定残值的协方差结构。
custom_target	boolean	指明是使用在上游节点定义的目标 (false) 还是由 target_field 指定的定制目标 (true)。
target_field	字段	要用作目标的字段 (如果 custom_target 为 true)。
use_trials	boolean	指示目标响应是一组试验中发生的众多事件时, 是否使用用于指定试验数的附加字段或值。缺省值为 false。
use_field_or_value	字段 值 (Value)	指示是使用字段 (缺省) 还是值来指定试验数。
trials_field	字段	此字段用于指定试验数。
trials_value	integer	此值用于指定试验数。如果指定此属性, 那么最小值为 1。

表 101. glmm 节点属性 (续).

glmm 节点属性	值	属性描述
use_custom_target_reference	<i>boolean</i>	指示将定制参考类别用于分类目标。缺省值为 <i>false</i> 。
target_reference_value	<i>string</i>	要使用的参考类别 (如果 <i>use_custom_target_reference</i> 为 <i>true</i>)。
dist_link_combination	名义 (Nominal) Logit GammaLog BinomialLogit PoissonLog BinomialProbit NegbinLog BinomialLogC 自定义 (Custom)	目标值的分布的公共模型。选择 Custom 可以指定 <i>target_distribution</i> 所提供的列表中的分布。
target_distribution	正态 (Normal) 二项 Multinomial 伽玛 (Gamma) 逆模型 NegativeBinomial 泊松 (Poisson)	当 <i>dist_link_combination</i> 为 Custom 时目标值的分布。
link_function_type	IDENTITY LOGC LOG CLOGLOG LOGIT NLOGLOG PROBIT POWER CAUCHIT	这是用于使目标值与预测变量相关的关联函数。如果 <i>target_distribution</i> 为 Binomial, 那么您可以使用任何列出的关联函数。如果 <i>target_distribution</i> 为 Multinomial, 那么您可以使用 CLOGLOG、CAUCHIT、LOGIT、NLOGLOG 或 PROBIT。如果 <i>target_distribution</i> 不是 Binomial 或 Multinomial, 那么您可以使用 IDENTITY、LOG 或 POWER。
link_function_param	<i>number</i>	要使用的关联函数参数值。仅当 <i>normal_link_function</i> 或 <i>link_function_type</i> 为 POWER 时才适用。
use_predefined_inputs	<i>boolean</i>	指示固定效应字段是定义为输入字段的上流 (<i>true</i>) 还是来自 <i>fixed_effects_list</i> (<i>false</i>)。缺省值为 <i>false</i> 。
fixed_effects_list	结构化	如果 <i>use_predefined_inputs</i> 为 <i>false</i> , 那么指定将输入字段用作固定效应字段。
use_intercept	<i>boolean</i>	如果为 <i>true</i> (缺省), 那么在模型中包括截距。
random_effects_list	结构化	作为随机效应指定的字段列表。
regression_weight_field	字段	此字段用作分析权重字段。
use_offset	无 (None) offset_value offset_field	指示如何指定平移。值 None 表示不使用平移。
offset_value	<i>number</i>	<i>use_offset</i> 设置为 <i>offset_value</i> 时使用的平移值。

表 101. glmm 节点属性 (续).

glmm 节点属性	值	属性描述
offset_field	字段	use_offset 设置为 offset_field 时用于平移值的字段。
target_category_order	升序 (Ascending) 降序 (Descending) Data	分类目标的排序顺序。值 Data 指定使用数据中的排序顺序。缺省值为 Ascending。
inputs_category_order	升序 (Ascending) 降序 (Descending) Data	分类预测变量的排序顺序。值 Data 指定使用数据中的排序顺序。缺省值为 Ascending。
max_iterations	integer	此算法要执行的最大迭代次数。非负整数；缺省值为 100。
confidence_level	integer	这是用于计算模型系数的区间估计值的置信度级别。非负整数；最大值为 100，缺省值为 95。
degrees_of_freedom_method	Fixed Varied	指定如何计算自由度以进行显著性检验。
test_fixed_effects_coefficients	模型 (Model) Robust	这是用于计算参数估计协方差矩阵的方法。
use_p_converge	布尔值	用于参数收敛的选项。
p_converge	number	空白或任何正值。
p_converge_type	Absolute Relative	
use_l_converge	布尔值	用于对数似然收敛的选项。
l_converge	number	空白或任何正值。
l_converge_type	Absolute Relative	
use_h_converge	布尔值	用于 Hessian 收敛的选项。
h_converge	number	空白或任何正值。
h_converge_type	Absolute Relative	
max_fisher_steps	integer	
singularity_tolerance	number	
use_model_name	boolean	指示是为模型指定定制名称 (true) 还是使用系统生成的名称 (false)。缺省值为 false。
model_name	string	如果 use_model_name 为 true, 那么指定使用的模型名称。
confidence	onProbability onIncrease	计算评分置信度值的基础: 最高预测概率或者最高与次高预测概率之差。
score_category_probabilities	boolean	如果为 true, 那么为分类目标生成预测概率。缺省值为 false。
max_categories	integer	如果 score_category_probabilities 为 true, 那么指定保存最大类别数。
score_propensity	boolean	如果为 true, 那么为标记目标字段生成倾向得分, 指示字段结果为“true”的可能性。

表 101. glmm 节点属性 (续).

glmm 节点属性	值	属性描述
emeans	structure	对于固定效应列表中的每个分类字段, 指定是否生成估计边际均值。
covariance_list	structure	对于固定效应列表中的每个连续字段, 指定计算估计边际均值时是使用均值还是定制值。
mean_scale	Original 已转换 (Transformed)	指定是根据目标的原始尺度 (缺省) 还是根据关联函数转换来计算估计边际均值。
comparison_adjustment_method	LSD SEQBONFERRONI SEQSIDAK	对多个对比执行假设检验时使用的调整方法。

kmeans 节点属性



K-Means 节点将数据集聚类到不同分组 (或聚类)。此方法将定义固定的聚类数量, 将记录迭代分配给聚类, 以及调整聚类中心, 直到进一步优化无法再改进模型。*k-means* 节点作为一种非监督学习机制, 它并不试图预测结果, 而是揭示隐含在输入字段集中的模式。

表 102. kmeans 节点属性.

kmeans 节点属性	值	属性描述
inputs	[field1 ... fieldN]	K-means 模型在一系列输入字段上执行聚类分析, 但并不使用目标字段。不使用权重字段和频率字段。有关更多信息, 请参阅第 117 页的『公共建模节点属性』主题。
num_clusters	数字	
gen_distance	布尔值	
cluster_label	字符串 数字	
label_prefix	字符串	
mode	Simple Expert	
stop_on	Default Custom	
max_iterations	数字	
tolerance	数字	
encoding_value	数字	
optimize	Speed Memory	用于指定模型构建是针对速度还是内存进行优化。

knn 节点属性



The k -最近相邻元素 (KNN) 节点将新的个案关联到预测变量空间中与其最邻近的 k 个对象的类别或值 (其中 k 为整数)。类似个案相互靠近, 而不同个案相互远离。

表 103. knn 节点属性.

knn 节点属性	值	属性描述
analysis	PredictTarget IdentifyNeighbors	
objective	Balance Speed Accuracy Custom	
normalize_ranges	布尔值	
use_case_labels	布尔值	此复选框用于启用下一个选项。
case_labels_field	字段	
identify_focal_cases	布尔值	此复选框用于启用下一个选项。
focal_cases_field	字段	
automatic_k_selection	布尔值	
fixed_k	整数	只有当 automatic_k_selectio 为 False 时才启用。
minimum_k	整数	只有当 automatic_k_selectio 为 True 时才启用。
maximum_k	整数	
distance_computation	Euclidean CityBlock	
weight_by_importance	布尔值	
range_predictions	Mean Median	
perform_feature_selection	布尔值	
forced_entry_inputs	[field1 ... fieldN]	
stop_on_error_ratio	布尔值	
number_to_select	整数	
minimum_change	数字	
validation_fold_assign_by_field	布尔值	
number_of_folds	整数	只有当 validation_fold_assign_by_field 为 False 时才启用。
set_random_seed	布尔值	
random_seed	数字	
folds_field	字段	只有当 validation_fold_assign_by_field 为 True 时才启用。
all_probabilities	布尔值	
save_distances	布尔值	

表 103. knn 节点属性 (续).

knn 节点属性	值	属性描述
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

kohonen 节点属性



Kohonen 节点会生成一种神经网络，此神经网络可用于将数据集聚类到各个差异组。此网络训练完成后，相似的记录应在输出映射中紧密地聚集，差异大的记录则应彼此远离。您可以通过查看模型块 中每个单元所捕获观测值的数量来找出规模较大的单元。这将让您对聚类的相应数量有所估计。

表 104. kohonen 节点属性.

kohonen 节点属性	值	属性描述
inputs	[field1 ... fieldN]	Kohonen 模型使用输入字段的列表，但不使用目标。不使用频率字段和权重字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
continue	布尔值	
show_feedback	布尔值	
stop_on	Default 时间	
time	数字	
optimize	Speed Memory	用于指定模型构建是针对速度还是内存进行优化。
cluster_label	布尔值	
mode	Simple Expert	
width	数字	
length	数字	
decay_style	线性 (Linear) Exponential	
phase1_neighborhood	数字	
phase1_eta	数字	
phase1_cycles	数字	
phase2_neighborhood	数字	
phase2_eta	数字	
phase2_cycles	数字	

linear 节点属性



线性回归模型根据目标与一个或多个预测变量之间的线性关系预测连续目标。

表 105. linear 节点属性.

linear 节点属性	值	属性描述
target	字段	指定单个目标字段。
inputs	[field1 ... fieldN]	模型使用的预测变量字段。
continue_training_existing_model	布尔值	
objective	Standard Bagging Boosting psm	psm 用于非常大的数据集，同时需要 Server 连接。
use_auto_data_preparation	布尔值	
confidence_level	数字	
model_selection	ForwardStepwise BestSubsets None	
criteria_forward_stepwise	AICC Fstatistics AdjustedRSquare ASE	
probability_entry	数字	
probability_removal	数字	
use_max_effects	布尔值	
max_effects	数字	
use_max_steps	布尔值	
max_steps	数字	
criteria_best_subsets	AICC AdjustedRSquare ASE	
combining_rule_continuous	Mean Median	
component_models_n	数字	
use_random_seed	布尔值	
random_seed	数字	
use_custom_model_name	布尔值	
custom_model_name	字符串	
use_custom_name	布尔值	
custom_name	字符串	

表 105. linear 节点属性 (续).

linear 节点属性	值	属性描述
tooltip	字符串	
keywords	字符串	
annotation	字符串	

logreg 节点属性



Logistic 回归是一种统计方法，它可根据输入字段的值对记录进行分类。它类似于线性回归，但采用的是类别目标字段而非数字范围。

表 106. logreg 节点属性.

logreg 节点属性	值	属性描述
target	字段	Logistic 回归模型需要一个目标字段以及一个或多个输入字段。不使用频率字段和权重字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
logistic_procedure	Binomial Multinomial	
include_constant	布尔值	
mode	Simple Expert	
method	Enter Stepwise Forwards Backwards BackwardsStepwise	
binomial_method	Enter Forwards Backwards	
model_type	MainEffects FullFactorial Custom	将 FullFactorial 指定为模型类型时，即使指定了步进方法，步进方法也不会运行。而是使用 Enter 方法。 如果将模型类型设置为 Custom，但未指定定制字段，那么将构建主效应模型。
custom_terms	[{BP Sex}{BP}{Age}]	
multinomial_base_category	字符串	指定如何确定参考类别。
binomial_categorical_input	字符串	

表 106. logreg 节点属性 (续).

logreg 节点属性	值	属性描述
binomial_input_contrast	Indicator Simple Difference Helmert Repeated Polynomial Deviation	这是分类输入的键控属性，用于指定如何确定对比。
binomial_input_category	First Last	这是分类输入的键控属性，用于指定如何确定参考类别。
scale	None UserDefined Pearson Deviance	
scale_value	数字	
all_probabilities	布尔值	
tolerance	1.0E-5 1.0E-6 1.0E-7 1.0E-8 1.0E-9 1.0E-10	
min_terms	数字	
use_max_terms	布尔值	
max_terms	数字	
entry_criterion	Score LR	
removal_criterion	LR Wald	
probability_entry	数字	
probability_removal	数字	
binomial_probability_entry	数字	
binomial_probability_removal	数字	
requirements	HierarchyDiscrete HierarchyAll Containment None	
max_iterations	数字	
max_steps	数字	
p_converge	1.0E-4 1.0E-5 1.0E-6 1.0E-7 1.0E-8 0	

表 106. logreg 节点属性 (续).

logreg 节点属性	值	属性描述
l_converge	1.0E-1 1.0E-2 1.0E-3 1.0E-4 1.0E-5 0	
delta	数字	
iteration_history	布尔值	
history_steps	数字	
summary	布尔值	
likelihood_ratio	布尔值	
asymptotic_correlation	布尔值	
goodness_fit	布尔值	
parameters	布尔值	
confidence_interval	数字	
asymptotic_covariance	布尔值	
classification_table	布尔值	
stepwise_summary	布尔值	
info_criteria	布尔值	
monotonicity_measures	布尔值	
binomial_output_display	at_each_step at_last_step	
binomial_goodness_of_fit	布尔值	
binomial_parameters	布尔值	
binomial_iteration_history	布尔值	
binomial_classification_plots	布尔值	
binomial_ci_enable	布尔值	
binomial_ci	数字	
binomial_residual	outliers all	
binomial_residual_enable	布尔值	
binomial_outlier_threshold	数字	
binomial_classification_cutoff	数字	
binomial_removal_criterion	LR Wald Conditional	
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	

neuralnet 节点属性

注意: 在此发行版中提供了具有增强功能的新版本的神经网络建模节点，并将在下一节 (*neuralnetwork*) 中进行介绍。尽管您仍然可以使用先前版本来构建模型并对其评分，但我们建议您将脚本更新为使用新版本。这里保留了先前版本的详细信息供您参考。

表 107. *neuralnet* 节点属性.

neuralnet 节点属性	值	属性描述
targets	[field1 ... fieldN]	“神经网络”节点需要一个或多个目标字段以及一个或多个输入字段。将忽略频率和权重字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
method	Quick Dynamic Multiple Prune ExhaustivePrune RBFN	
prevent_overtrain	布尔值	
train_pct	数字	
set_random_seed	布尔值	
random_seed	数字	
mode	Simple Expert	
stop_on	Default Accuracy Cycles 时间	停止方式。
accuracy	数字	停止准确性。
cycles	数字	训练周期数。
time	数字	训练时间（分钟数）。
continue	布尔值	
show_feedback	布尔值	
binary_encode	布尔值	
use_last_model	布尔值	
gen_logfile	布尔值	
logfile_name	字符串	
alpha	数字	
initial_eta	数字	
high_eta	数字	
low_eta	数字	
eta_decay_cycles	数字	
hid_layers	One Two Three	
h1_units_one	数字	

表 107. neuralnet 节点属性 (续).

neuralnet 节点属性	值	属性描述
hl_units_two	数字	
hl_units_three	数字	
persistence	数字	
m_topologies	字符串	
m_non_pyramids	布尔值	
m_persistence	数字	
p_hid_layers	One Two Three	
p_hl_units_one	数字	
p_hl_units_two	数字	
p_hl_units_three	数字	
p_persistence	数字	
p_hid_rate	数字	
p_hid_pers	数字	
p_inp_rate	数字	
p_inp_pers	数字	
p_overall_pers	数字	
r_persistence	数字	
r_num_clusters	数字	
r_eta_auto	布尔值	
r_alpha	数字	
r_eta	数字	
optimize	Speed Memory	用于指定模型构建是针对速度还是内存进行优化。
calculate_variable_importance	布尔值	注：此属性取代了先前版本中使用的 sensitivity_analysis 属性。仍然支持旧属性，但建议使用 calculate_variable_importance。
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

neuralnetwork 节点属性

神经网络节点使用的模型是对人类大脑处理信息的方式简化了的模型。此模型通过模拟大量类似于神经元的抽象形式的互连简单处理单元而运行。神经网络是功能强大的一般函数估计器，只需要最少的统计或数学知识就可以对其进行训练或应用。

表 108. neuralnetwork 节点属性.

neuralnetwork 节点属性	值	属性描述
targets	[field1 ... fieldN]	指定目标字段。
inputs	[field1 ... fieldN]	模型使用的预测变量字段。
splits	[field1 ... fieldN]	指定一个或多个用于分割建模的字段。
use_partition	布尔值	如果定义了分区字段，那么此选项可确保仅训练分区的数据用于构建模型。
continue	布尔值	继续训练现有模型。
objective	Standard Bagging Boosting psm	psm 用于非常大的数据集，同时需要 Server 连接。
method	MultilayerPerceptron RadialBasisFunction	
use_custom_layers	布尔值	
first_layer_units	数字	
second_layer_units	数字	
use_max_time	布尔值	
max_time	数字	
use_max_cycles	布尔值	
max_cycles	数字	
use_min_accuracy	布尔值	
min_accuracy	数字	
combining_rule_categorical	Voting HighestProbability HighestMeanProbability	
combining_rule_continuous	Mean Median	
component_models_n	数字	
overfit_prevention_pct	数字	
use_random_seed	布尔值	
random_seed	数字	
missing_values	listwiseDeletion missingValueImputation	
use_custom_model_name	布尔值	
custom_model_name	字符串	
confidence	onProbability onIncrease	
score_category_probabilities	布尔值	
max_categories	数字	
score_propensity	布尔值	
use_custom_name	布尔值	

表 108. neuralnetwork 节点属性 (续).

neuralnetwork 节点属性	值	属性描述
custom_name	字符串	
tooltip	字符串	
keywords	字符串	
annotation	字符串	

quest 节点属性



QUEST 节点可提供用于构建决策树的二元分类法，此方法的设计目的是减少大型 C&R 树分析所需的处理时间，同时也减少在分类树方法中发现的趋势以便支持允许有多个分割的输入。输入字段可以是数字范围（连续），但目标字段必须是分类。所有分割都是二元的。

表 109. quest 节点属性.

quest 节点属性	值	属性描述
target	字段	QUEST 模型需要一个目标字段以及一个或多个输入字段。另外，还可以指定频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
continue_training_existing_model	布尔值	
objective	Standard Boosting Bagging psm	psm 用于非常大的数据集，同时需要 Server 连接。
model_output_type	Single InteractiveBuilder	
use_tree_directives	布尔值	
tree_directives	字符串	
use_max_depth	Default Custom	
max_depth	整数	最大树深度从 0 到 1000。只在 use_max_depth = Custom 时使用。
prune_tree	布尔值	修剪树，以避免过度拟合。
use_std_err	布尔值	使用最大风险差值（标准误差）。
std_err_multiplier	数字	最大差值。
max_surrogates	数字	最大代用项。
use_percentage	布尔值	
min_parent_records_pc	数字	
min_child_records_pc	数字	
min_parent_records_abs	数字	
min_child_records_abs	数字	
use_costs	布尔值	
costs	结构化	结构化属性。

表 109. quest 节点属性 (续).

quest 节点属性	值	属性描述
priors	Data Equal Custom	
custom_priors	结构化	结构化属性。
adjust_priors	布尔值	
trails	数字	用于推进或组装的组件模型数。
set_ensemble_method	Voting HighestProbability HighestMeanProbability	分类目标的缺省组合规则。
range_ensemble_method	Mean Median	连续目标的缺省组合规则。
large_boost	布尔值	对非常大的数据集应用推进。
split_alpha	数字	用于执行分割的显著性水平。
train_pct	数字	防止过度拟合集合。
set_random_seed	布尔值	复制结果选项。
seed	数字	
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

regression 节点属性



线性回归是一种通过拟合直线或平面以实现汇总数据和预测的普通统计方法，它可使预测值和实际输出值之间的差异最小化。

注：在未来的发行版中，“回归”节点将替换为“线性”节点。我们建议您从现在开始使用线性模型进行线性回归。

表 110. regression 节点属性.

regression 节点属性	值	属性描述
target	字段	“回归”模型需要单个目标字段以及一个或多个输入字段。另外，还可以指定权重字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
method	Enter Stepwise Backwards Forwards	
include_constant	布尔值	

表 110. regression 节点属性 (续).

regression 节点属性	值	属性描述
use_weight	布尔值	
weight_field	字段	
mode	Simple Expert	
complete_records	布尔值	
tolerance	1.0E-1 1.0E-2 1.0E-3 1.0E-4 1.0E-5 1.0E-6 1.0E-7 1.0E-8 1.0E-9 1.0E-10 1.0E-11 1.0E-12	请使用双引号将自变量括起来。
stepping_method	useP useF	useP: 使用 F 的概率 useF: 使用 F 值
probability_entry	数字	
probability_removal	数字	
F_value_entry	数字	
F_value_removal	数字	
selection_criteria	布尔值	
confidence_interval	布尔值	
covariance_matrix	布尔值	
collinearity_diagnostics	布尔值	
regression_coefficients	布尔值	
exclude_fields	布尔值	
durbin_watson	布尔值	
model_fit	布尔值	
r_squared_change	布尔值	
p_correlations	布尔值	
descriptives	布尔值	
calculate_variable_importance	布尔值	

sequence 节点属性



序列节点可发现连续数据或与时间有关的数据中的关联规则。序列是一系列可能会以可预测顺序发生的项目集合。例如，一个购买了剃刀和须后水的顾客可能在下次购物时购买剃须膏。序列节点基于 CARMA 关联规则算法，该算法使用一个有效的两次传递方法查找序列。

表 111. *sequence* 节点属性.

sequence 节点属性	值	属性描述
id_field	字段	要创建序列规则集，您需要指定一个 ID 字段以及一个可选的时间字段，以及一个或多个内容字段。不使用权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
time_field	字段	
use_time_field	布尔值	
content_fields	[field1 ... fieldn]	
contiguous	布尔值	
min_supp	数字	
min_conf	数字	
max_size	数字	
max_predictions	数字	
mode	Simple Expert	
use_max_duration	布尔值	
max_duration	数字	
use_gaps	布尔值	
min_item_gap	数字	
max_item_gap	数字	
use_pruning	布尔值	
pruning_value	数字	
set_mem_sequences	布尔值	
mem_sequences	整数	

slrm 节点属性



自学响应模型（SLRM）节点可用于构建一个包含单个新观测值或少量新观测值的模型，通过此模型，无需使用全部数据对模型进行重新训练即可对模型进行重新评估。

表 112. *slrm* 节点属性.

slrm 节点属性	值	属性描述
target	字段	目标字段必须是名义字段或标志字段。另外，还可以指定频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
target_response	字段	类型必须是标志。
continue_training_existing_model	布尔值	
target_field_values	布尔值	Use all: 使用来自源的全部值。 Specify: 选择所需的值。

表 112. slrm 节点属性 (续).

slrm 节点属性	值	属性描述
target_field_values_specify	[field1 ... fieldN]	
include_model_assessment	布尔值	
model_assessment_random_seed	数字	必须是实数。
model_assessment_sample_size	数字	必须是实数。
model_assessment_iterations	数字	迭代数。
display_model_evaluation	布尔值	
max_predictions	数字	
randomization	数字	
scoring_random_seed	数字	
sort	Ascending Descending	指定是先显示得分最高还是最低的报价。
model_reliability	布尔值	
calculate_variable_importance	布尔值	

statisticsmodel 节点属性



“Statistics 模型”节点使您能够通过运行将会生成 PMML 的 IBM SPSS Statistics 过程来分析和处理数据。此节点需要 IBM SPSS Statistics 的许可副本。

有关此节点属性的信息，请参阅第 216 页的『statisticsmodel 节点属性』。

svm 节点属性



使用支持向量机 (SVM) 节点，可以将数据分为两组，而无需过度拟合。SVM 可以与宽数据集配合使用，例如那些含有大量输入字段的数据集。

表 113. svm 节点属性.

svm 节点属性	值	属性描述
all_probabilities	布尔值	
stopping_criteria	1.0E-1 1.0E-2 1.0E-3 (default) 1.0E-4 1.0E-5 1.0E-6	确定何时停止优化算法。
regularization	数字	也称为 C 参数。
precision	数字	仅当目标字段的测量级别为 Continuous 时才使用。

表 113. svm 节点属性 (续).

svm 节点属性	值	属性描述
kernel	RBF (缺省) Polynomial Sigmoid Linear	用于转换的内核函数的类型。
rbf_gamma	数字	仅在 kernel 为 RBF 时使用。
gamma	数字	仅在 kernel 为 Polynomial 或 Sigmoid 时使用。
bias	数字	
degree	数字	仅在 kernel 为 Polynomial 时使用。
calculate_variable_importance	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	
adjusted_propensity_partition	Test Validation	

timeseries 节点属性



“时间序列”节点估计时间序列数据的指数平滑模型、单变量自回归整合移动平均值 (ARIMA) 模型和多变量 ARIMA (即转换函数) 模型, 并生成未来性能的预测数据。在“时间序列”节点之前必须有“时间区间”节点。

表 114. timeseries 节点属性.

timeseries 节点属性	值	属性描述
targets	字段	“时间序列”节点可以预测一个或多个目标, 可以选择使用一个或多个输入字段作为预测变量。不使用频率字段和权重字段。有关更多信息, 请参阅第 117 页的『公共建模节点属性』主题。
continue	布尔值	
method	ExpertModeler Exsmooth Arima Reuse	
expert_modeler_method	布尔值	
consider_seasonal	布尔值	
detect_outliers	布尔值	
expert_outlier_additive	布尔值	
expert_outlier_level_shift	布尔值	
expert_outlier_innovational	布尔值	
expert_outlier_level_shift	布尔值	

表 114. timeseries 节点属性 (续).

timeseries 节点属性	值	属性描述
expert_outlier_transient	布尔值	
expert_outlier_seasonal_additive	布尔值	
expert_outlier_local_trend	布尔值	
expert_outlier_additive_patch	布尔值	
exsmooth_model_type	Simple HoltLinearTrend BrownsLinearTrend DampedTrend SimpleSeasonal WintersAdditive WintersMultiplicative	
exsmooth_transformation_type	None SquareRoot NaturalLog	
arma_p	整数	
arma_d	整数	
arma_q	整数	
arma_sp	整数	
arma_sd	整数	
arma_sq	整数	
arma_transformation_type	None SquareRoot NaturalLog	
arma_include_constant	布尔值	
tf_arma_p. <i>fieldname</i>	整数	用于转换函数。
tf_arma_d. <i>fieldname</i>	整数	用于转换函数。
tf_arma_q. <i>fieldname</i>	整数	用于转换函数。
tf_arma_sp. <i>fieldname</i>	整数	用于转换函数。
tf_arma_sd. <i>fieldname</i>	整数	用于转换函数。
tf_arma_sq. <i>fieldname</i>	整数	用于转换函数。
tf_arma_delay. <i>fieldname</i>	整数	用于转换函数。
tf_arma_transformation_type. <i>fieldname</i>	None SquareRoot NaturalLog	用于转换函数。
arma_detect_outlier_mode	None Automatic	
arma_outlier_additive	布尔值	
arma_outlier_level_shift	布尔值	
arma_outlier_innovational	布尔值	
arma_outlier_transient	布尔值	
arma_outlier_seasonal_additive	布尔值	
arma_outlier_local_trend	布尔值	

表 114. *timeseries* 节点属性 (续).

timeseries 节点属性	值	属性描述
arma_outlier_additive_patch	布尔值	
conf_limit_pct	实数	
max_lags	整数	
events	<i>fields</i>	
scoring_model_only	布尔值	用于包含大量（数万）时间序列的模型。

twostep 节点属性



“二阶”节点使用两步聚类方法。第一步完成简单数据处理，以便将原始输入数据压缩为可管理的子聚类集合。第二步使用层级聚类方法将子聚类一步一步合并为更大的聚类。“二阶”具有一个优点，就是能够为训练数据自动估计最佳聚类数。它可以高效处理混合的字段类型和大型的数据集。

表 115. *twostep* 节点属性.

twostep 节点属性	值	属性描述
inputs	<i>[field1 ... fieldN]</i>	“二阶”模型使用输入字段列表，但不使用目标字段。不识别权重字段和频率字段。有关更多信息，请参阅第 117 页的『公共建模节点属性』主题。
standardize	布尔值	
exclude_outliers	布尔值	
percentage	数字	
cluster_num_auto	布尔值	
min_num_clusters	数字	
max_num_clusters	数字	
num_clusters	数字	
cluster_label	字符串 数字	
label_prefix	字符串	
distance_measure	Euclidean Loglikelihood	
clustering_criterion	AIC BIC	

第 14 章 模型块节点属性

模型块节点具有与其他节点相同的公共属性。有关更多信息，请参阅第 50 页的『公共节点属性』主题。

applyanomalydetection 节点属性

您可以使用“异常检测”建模节点来生成“异常检测”模型块。此模型块的脚本编制名称为 *applyanomalydetection*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 117 页的『anomalydetection 节点属性』。

表 116. *applyanomalydetection* 节点属性.

applyanomalydetection 节点属性	值	属性描述
anomaly_score_method	FlagAndScore FlagOnly ScoreOnly	确定创建哪些输出用于评分。
num_fields	整数	要报告的字段。
discard_records	布尔值	指示是否从输出中丢弃记录。
discard_anomalous_records	布尔值	指示是丢弃异常记录还是丢弃 非异常记录。缺省状态为 off，表示丢弃非异常记录。否则，如果状态为 on，那么丢弃异常记录。仅当启用 discard_records 属性时，才会启用此属性。

applyapriori 节点属性

您可以使用 Apriori 建模节点来生成 Apriori 模型块。此模型块的脚本编制名称为 *applyapriori*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 118 页的『apriori 节点属性』。

表 117. *applyapriori* 节点属性.

applyapriori 节点属性	值	属性描述
max_predictions	数字（整数）	
ignore_unmated	布尔值	
allow_repeats	布尔值	
check_basket	NoPredictions Predictions NoCheck	
criterion	Confidence Support RuleSupport Lift Deployability	

applyautoclassifier 节点属性

您可以使用“自动分类器”建模节点来生成“自动分类器”模型块。此模型块的脚本编制名称为 *applyautoclassifier*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 119 页的『*autoclassifier* 节点属性』。

表 118. *applyautoclassifier* 节点属性.

applyautoclassifier 节点属性	值	属性描述
flag_ensemble_method	Voting ConfidenceWeightedVoting RawPropensityWeightedVoting HighestConfidence AverageRawPropensity	指定用于确定整体得分的方法。仅当选定的目标是标志字段时，才会应用此设置。
flag_voting_tie_selection	Random HighestConfidence RawPropensity	如果已选定投票方法，那么指定解决结的方法。仅当选定的目标是标志字段时，才会应用此设置。
set_ensemble_method	Voting ConfidenceWeightedVoting HighestConfidence	指定用于确定整体得分的方法。仅当选定的目标是集合字段时，才会应用此设置。
set_voting_tie_selection	Random HighestConfidence	如果已选定投票方法，那么指定解决结的方法。仅当选定的目标是名义字段时，才会应用此设置。

applyautocluster 节点属性

您可以使用“自动聚类”建模节点来生成“自动聚类”模型块。此模型块的脚本编制名称为 *applyautocluster*。此模型块不具有任何其他属性。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 121 页的『*autocluster* 节点属性』。

applyautonumeric 节点属性

您可以使用“自动数字”建模节点来生成“自动数字”模型块。此模型块的脚本编制名称为 *applyautonumeric*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 122 页的『*autonumeric* 节点属性』。

表 119. *applyautonumeric* 节点属性.

applyautonumeric 节点属性	值	属性描述
calculate_standard_error	布尔值	

applybayesnet 节点属性

您可以使用“贝叶斯网络”建模节点来生成“贝叶斯网络”模型块。此模型块的脚本编制名称为 *applybayesnet*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 123 页的『*bayesnet* 节点属性』。

表 120. *applybayesnet* 节点属性.

applybayesnet 节点属性	值	属性描述
all_probabilities	布尔值	
raw_propensity	布尔值	
adjusted_propensity	布尔值	

表 120. *applybayesnet* 节点属性 (续).

applybayesnet 节点属性	值	属性描述
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applyc50 节点属性

您可以使用 C5.0 建模节点来生成 C5.0 模型块。此模型块的脚本编制名称为 *applyc50*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 125 页的『c50 节点属性』。

表 121. *applyc50* 节点属性.

applyc50 节点属性	值	属性描述
sql_generate	Never NoMissingValues	用于设置规则集执行期间使用的 SQL 生成选项。
calculate_conf	布尔值	启用 SQL 生成时可用；此属性将置信度的计算包括在生成的树中。
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applycarma 节点属性

您可以使用 CARMA 建模节点来生成 CARMA 模型块。此模型块的脚本编制名称为 *applycarma*。此模型块不具有任何其他属性。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 126 页的『carma 节点属性』。

applycart 节点属性

您可以使用“C&R 树”建模节点来生成“C&R 树”模型块。此模型块的脚本编制名称为 *applycart*。有关对此建模节点本身进行脚本编制的更多信息，请参阅主题第 126 页的『cart 节点属性』。

表 122. *applycart* 节点属性.

applycart 节点属性	值	属性描述
sql_generate	Never MissingValues NoMissingValues	用于设置规则集执行期间使用的 SQL 生成选项。
calculate_conf	布尔值	启用 SQL 生成时可用；此属性将置信度的计算包括在生成的树中。
display_rule_id	布尔值	在得分输出中添加一个字段，表示每个记录分配到的终端节点的 ID。
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applychaid 节点属性

您可以使用 CHAID 建模节点来生成 CHAID 模型块。此模型块的脚本编制名称为 *applychaid*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 128 页的『chaid 节点属性』。

表 123. *applychaid* 节点属性.

applychaid 节点属性	值	属性描述
sql_generate	Never MissingValues	
calculate_conf	布尔值	
display_rule_id	布尔值	在得分输出中添加一个字段，表示每个记录分配到的终端节点的 ID。
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applycoxreg 节点属性

您可以使用 Cox 建模节点来生成 Cox 模型块。此模型块的脚本编制名称为 *applycoxreg*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 129 页的『coxreg 节点属性』。

表 124. *applycoxreg* 节点属性.

applycoxreg 节点属性	值	属性描述
future_time_as	Intervals Fields	
time_interval	数字	
num_future_times	整数	
time_field	字段	
past_survival_time	字段	
all_probabilities	布尔值	
cumulative_hazard	布尔值	

applydecisionlist 节点属性

您可以使用“决策列表”建模节点来生成“决策列表”模型块。此模型块的脚本编制名称为 *applydecisionlist*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 131 页的『decisionlist 节点属性』。

表 125. *applydecisionlist* 节点属性.

applydecisionlist 节点属性	值	属性描述
enable_sql_generation	布尔值	值为 true 时，IBM SPSS Modeler 会尝试将决策列表模型回推到 SQL。
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applydiscriminant 节点属性

您可以使用“判别”建模节点来生成“判别”模型块。此模型块的脚本编制名称为 *applydiscriminant*。有关对此建模节点本身进行脚本编制的更多信息，请参阅主题第 132 页的『discriminant 节点属性』。

表 126. *applydiscriminant* 节点属性.

applydiscriminant 节点属性	值	属性描述
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applyfactor 节点属性

您可以使用“PCA/因子”建模节点来生成“PCA/因子”模型块。此模型块的脚本编制名称为 *applyfactor*。此模型块不具有任何其他属性。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题第 133 页的『factor 节点属性』。

applyfeatureselection 节点属性

您可以使用“特征选择”建模节点来生成“特征选择”模型块。此模型块的脚本编制名称为 *applyfeatureselection*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题第 134 页的『featureselection 节点属性』。

表 127. *applyfeatureselection* 节点属性.

applyfeatureselection 节点属性	值	属性描述
selected_ranked_fields		指定要在模型浏览器中检查哪些已排序的字段。
selected_screened_fields		指定要在模型浏览器中检查哪些已筛选的字段。

applygeneralizedlinear 节点属性

您可以使用“广义线性 (genlin)”建模节点来生成“广义线性”模型块。此模型块的脚本编制名称为 *applygeneralizedlinear*。有关对此建模节点本身进行脚本编制的更多信息，请参阅主题第 136 页的『genlin 节点属性』。

表 128. *applygeneralizedlinear* 节点属性.

applygeneralizedlinear 节点属性	值	属性描述
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applyglm 节点属性

您可以使用 GLMM 建模节点来生成 GLMM 模型块。此模型块的脚本编制名称为 *applyglm*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 139 页的『glm 节点属性』。

表 129. *applyglm* 节点属性.

applyglm 节点属性	值	属性描述
confidence	onProbability onIncrease	计算评分置信度值的基础：最高预测概率或者最高与次高预测概率之差。
score_category_probabilities	布尔值	如果设置为 true，那么为分类目标生成预测概率。为每个类别创建一个字段。缺省值为 False。
max_categories	整数	要预测概率的最大类别数目。仅当 score_category_probabilities 为 True 时才使用。
score_propensity	布尔值	如果设置为 True，那么为具有标志目标的模型生成原始倾向得分（“True”结果的可能性）。如果分区处于有效，那么模型还会根据测试分区产生调整后的倾向得分。缺省值为 False。

applykmeans 节点属性

您可以使用 K-Means 建模节点来生成 K-Means 模型块。此模型块的脚本编制名称为 *applykmeans*。此模型块不具有任何其他属性。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 142 页的『kmeans 节点属性』。

applyknn 节点属性

您可以使用 KNN 建模节点来生成 KNN 模型块。此模型块的脚本编制名称为 *applyknn*。有关对此建模节点本身进行脚本编制的更多信息，请参阅主题第 143 页的『knn 节点属性』。

表 130. *applyknn* 节点属性.

applyknn 节点属性	值	属性描述
all_probabilities	布尔值	
save_distances	布尔值	

applykohonen 节点属性

您可以使用 Kohonen 建模节点来生成 Kohonen 模型块。此模型块的脚本编制名称为 *applykohonen*。此模型块不具有任何其他属性。有关对此建模节点本身进行脚本编制的更多信息，请参阅主题第 144 页的『kohonen 节点属性』。

applylinear 节点属性

您可以使用“线性”建模节点来生成“线性”模型块。此模型块的脚本编制名称为 *applylinear*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 145 页的『linear 节点属性』。

表 131. *applylinear* 节点属性.

applylinear 节点属性	值	属性描述
use_custom_name	布尔值	
custom_name	字符串	
enable_sql_generation	布尔值	

applylogreg 节点属性

您可以使用“Logistic 回归模型”建模节点来生成“Logistic 回归模型”模型块。此模型块的脚本编制名称为 *applylogreg*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 146 页的『logreg 节点属性』。

表 132. *applylogreg* 节点属性.

applylogreg 节点属性	值	属性描述
calculate_raw_propensities	布尔值	
calculate_conf	布尔值	
enable_sql_generation	布尔值	

applyneuralnet 节点属性

您可以使用“神经网络”建模节点来生成“神经网络”模型块。此模型块的脚本编制名称为 *applyneuralnet*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 149 页的『neuralnet 节点属性』。

注意：在此发行版中提供了具有增强功能的新版本的神经网络模型块，并将在下一节 (*applyneuralnetwork*) 中进行介绍。尽管先前版本仍然可用，但我们建议您更新脚本以使用新的版本。此处保留了先前版本的详细信息以供参考，但会在将来的发行版中不再支持。

表 133. *applyneuralnet* 节点属性.

applyneuralnet 节点属性	值	属性描述
calculate_conf	布尔值	启用 SQL 生成时可用；此属性将置信度的计算包括在生成的树中。
enable_sql_generation	布尔值	
nn_score_method	Difference SoftMax	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applyneuralnetwork 节点属性

您可以使用“神经网络”建模节点来生成“神经网络”模型块。此模型块的脚本编制名称为 *applyneuralnetwork*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 150 页的『neuralnetwork 节点属性』。

表 134. *applyneuralnetwork* 节点属性.

applyneuralnetwork 节点属性	值	属性描述
use_custom_name	布尔值	
custom_name	字符串	
confidence	onProbability onIncrease	
score_category_probabilities	布尔值	
max_categories	数字	
score_propensity	布尔值	

applyquest 节点属性

您可以使用 QUEST 建模节点来生成 QUEST 模型块。此模型块的脚本编制名称为 *applyquest*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 152 页的『quest 节点属性』。

表 135. *applyquest* 节点属性.

applyquest 节点属性	值	属性描述
sql_generate	Never MissingValues NoMissingValues	
calculate_conf	布尔值	
display_rule_id	布尔值	在得分输出中添加一个字段，表示每个记录分配到的终端节点的 ID。
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applyregression 节点属性

您可以使用“线性回归”建模节点来生成“线性回归”模型块。此模型块的脚本编制名称为 *applyregression*。此模型块不具有任何其他属性。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 153 页的『regression 节点属性』。

applyr 节点属性

您可以使用“R 构建”节点来生成 R 模型块。此模型块的脚本编制名称为 *applyr*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 124 页的『buildr 节点属性』。

表 136. *applyr* 节点属性

applyr 节点属性	值	属性描述
score_syntax	字符串	这是用于进行模型评分的 R 脚本语法。

表 136. *applyr* 节点属性 (续)

applyr 节点属性	值	属性描述
convert_flags	StringsAndDoubles LogicalValues	此选项用于转换标志字段。
convert_datetime	布尔值	此选项用于将具有日期或日期时间格式的变量转换为 R 日期/时间格式。
convert_datetime_class	POSIXct POSIXlt	这些选项用于指定要将日期或日期时间格式的变量转换为什么格式。
convert_missing	布尔值	此选项用于将缺失值转换为 R NA 值。

applyselflearning 节点属性

您可以使用“自学响应模型 (SLRM)”建模节点来生成 SLRM 模型块。此模型块的脚本编制名称为 *applyselflearning*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 155 页的『slrm 节点属性』。

表 137. *applyselflearning* 节点属性.

applyselflearning 节点属性	值	属性描述
max_predictions	数字	
randomization	数字	
scoring_random_seed	数字	
sort	ascending descending	指定是先显示得分最高还是最低的报价。
model_reliability	布尔值	将“设置”选项卡中的模型可靠性选项考虑在内。

applysequence 节点属性

您可以使用“序列”建模节点来生成“序列”模型块。此模型块的脚本编制名称为 *applysequence*。此模型块不具有任何其他属性。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 154 页的『sequence 节点属性』。

applysvm 节点属性

您可以使用 SVM 建模节点来生成 SVM 模型块。此模型块的脚本编制名称为 *applysvm*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 156 页的『svm 节点属性』。

表 138. *applysvm* 节点属性.

applysvm 节点属性	值	属性描述
all_probabilities	布尔值	
calculate_raw_propensities	布尔值	
calculate_adjusted_propensities	布尔值	

applytimeseries 节点属性

您可以使用“时间序列”建模节点来生成“时间序列”模型块。此模型块的脚本编制名称为 *applytimeseries*。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 157 页的『*timeseries* 节点属性』。

表 139. *applytimeseries* 节点属性.

applytimeseries 节点属性	值	属性描述
calculate_conf	布尔值	
calculate_residuals	布尔值	

applytwostep 节点属性

您可以使用“二阶”建模节点来生成“二阶”模型块。此模型块的脚本编制名称为 *applytwostep*。此模型块不具有任何其他属性。有关对此建模节点自身进行脚本编制的更多信息，请参阅主题 第 159 页的『*twostep* 节点属性』。

第 15 章 数据库建模节点属性

IBM SPSS Modeler 支持与多家数据库供应商的数据挖掘和建模工具集成，这包括 Microsoft SQL Server Analysis Services、Oracle Data Mining、IBM DB2 InfoSphere Warehouse 和 IBM Netezza Analytics。。您可以使用 IBM SPSS Modeler 应用程序自有的数据库算法来构建模型并对模型进行评分。还可以使用本节介绍的属性通过编写脚本来构建和操纵数据库模型。

Microsoft 建模的节点属性

Microsoft 建模节点属性

公共属性

Microsoft 数据库建模节点的公共属性如下所示。

表 140. 公共 Microsoft 节点属性.

公共 Microsoft 节点属性	值	属性描述
analysis_database_name	字符串	Analysis Services 数据库的名称。
analysis_server_name	字符串	Analysis Services 主机的名称。
use_transactional_data	布尔值	指定输入数据是采用表格式还是事务格式。
inputs	[field field field]	表数据的输入字段。
target	字段	预测字段（不适用于“MS 聚类”或“序列聚类”节点）。
unique_field	字段	键字段。
msas_parameters	结构化	算法参数。有关更多信息，请参阅第 172 页的『算法参数』主题。
with_drillthrough	布尔值	“进行深入钻取”选项。

MS 决策树

未对类型为 `mstree` 的节点定义具体属性。请参阅本节开头的公共 Microsoft 属性。

MS 聚类

未对类型为 `mscluster` 的节点定义具体属性。请参阅本节开头的公共 Microsoft 属性。

MS 关联规则

以下具体属性可用于类型为 `msassoc` 的节点:

表 141. `msassoc` 节点属性.

<code>msassoc</code> 节点属性	值	属性描述
id_field	字段	标识数据中的每项事务。
trans_inputs	[field field field]	事务数据的输入字段。
transactional_target	字段	预测字段（事务数据）。

MS 朴素贝叶斯

未对类型为 `msbayes` 的节点定义具体属性。请参阅本节开头的公共 Microsoft 属性。

MS 线性回归

未对类型为 `msregression` 的节点定义具体属性。请参阅本节开头的公共 Microsoft 属性。

MS 神经网络

未对类型为 `msneuralnetwork` 的节点定义具体属性。请参阅本节开头的公共 Microsoft 属性。

MS Logistic 回归

未对类型为 `mslogistic` 的节点定义具体属性。请参阅本节开头的公共 Microsoft 属性。

MS 时间序列

未对类型为 `mstimeseries` 的节点定义具体属性。请参阅本节开头的公共 Microsoft 属性。

MS 序列聚类

以下具体属性可用于类型为 `mssequencecluster` 的节点：

表 142. `mssequencecluster` 节点属性.

<code>mssequencecluster</code> 节点属性	值	属性描述
<code>id_field</code>	字段	标识数据中的每项事务。
<code>input_fields</code>	<code>[field field field]</code>	事务数据的输入字段。
<code>sequence_field</code>	字段	序列标识。
<code>target_field</code>	字段	预测字段（表数据）。

算法参数

每种 Microsoft 数据库模型类型都有可以使用 `msas_parameters` 属性设置的特定参数。

这些参数源自 SQL Server。要查看每个节点的相关参数，请执行如下操作：

1. 将数据库源节点放入画布中。
2. 打开该数据库源节点。
3. 从**数据源**下拉列表选择一个有效源。
4. 从**表名称**列表选择一个有效表。
5. 单击**确定**以关闭该数据库源节点。
6. 附加要列出其属性的 Microsoft 数据库建模节点。
7. 打开该数据库建模节点。
8. 选择**专家**选项卡。

此时会显示该节点的可用 `msas_parameters` 属性。

Microsoft 模型块属性

使用 Microsoft 数据库建模节点创建的模型块具有下列属性。

MS 决策树

表 143. MS 决策树属性.

aplymstree 节点属性	值	描述
analysis_database_name	字符串	可以直接在流中对此节点进行评分。 此属性用于标识 Analysis Services 数据库的名称。
analysis_server_name	字符串	Analysis 服务器主机的名称。
datasource	字符串	SQL Server ODBC 数据源名称 (DSN) 的名称。
sql_generate	布尔值	启用 SQL 生成。

MS 线性回归

表 144. MS 线性回归属性.

aplymsregression 节点属性	值	描述
analysis_database_name	字符串	可以直接在流中对此节点进行评分。 此属性用于标识 Analysis Services 数据库的名称。
analysis_server_name	字符串	Analysis 服务器主机的名称。

MS 神经网络

表 145. MS 神经网络属性.

aplymsneuralnetwork 节点属性	值	描述
analysis_database_name	字符串	可以直接在流中对此节点进行评分。 此属性用于标识 Analysis Services 数据库的名称。
analysis_server_name	字符串	Analysis 服务器主机的名称。

MS Logistic 回归

表 146. MS Logistic 回归属性.

aplymslogistic 节点属性	值	描述
analysis_database_name	字符串	可以直接在流中对此节点进行评分。 此属性用于标识 Analysis Services 数据库的名称。
analysis_server_name	字符串	Analysis 服务器主机的名称。

MS 时间序列

表 147. MS 时间序列属性.

aplymstimeseries 节点属性	值	描述
analysis_database_name	字符串	可以直接在流中对此节点进行评分。 此属性用于标识 Analysis Services 数据库的名称。
analysis_server_name	字符串	Analysis 服务器主机的名称。

表 147. MS 时间序列属性 (续).

aplymstimeseries 节点属性	值	描述
start_from	new_prediction historical_prediction	指定是进行未来预测还是历史预测。
new_step	数字	定义未来预测的开始时间段。
historical_step	数字	定义历史预测的开始时间段。
end_step	数字	定义预测结束时间段。

MS 序列聚类

表 148. MS 序列聚类属性.

aplymssequencecluster 节点属性	值	描述
analysis_database_name	字符串	可以直接在流中对此节点进行评分。 此属性用于标识 Analysis Services 数据库的名称。
analysis_server_name	字符串	Analysis 服务器主机的名称。

Oracle 建模的节点属性

Oracle 建模节点属性

Oracle 数据库建模节点的公共属性如下所示。

表 149. 公共 Oracle 节点属性.

公共 Oracle 节点属性	值	属性描述
target	字段	
inputs	字段的列表	
partition	字段	此字段用于将数据分区为不同的样本，以用于模型构建的训练、检验和验证阶段。
datasource		
username		
password		
epassword		
use_model_name	布尔值	
model_name	字符串	新模型的定制名称。
use_partitioned_data	布尔值	如果定义了分区字段，那么此选项可确保仅训练分区的数据用于构建模型。
unique_field	字段	
auto_data_prep	布尔值	启用或禁用 Oracle 自动数据准备功能（仅适用于 11g 数据库）。
costs	结构化	结构化属性。
mode	Simple Expert	如在各个节点属性中注释的那样，如果设置为 Simple，会导致忽略某些属性。
use_prediction_probability	布尔值	
prediction_probability	字符串	

表 149. 公共 Oracle 节点属性 (续).

公共 Oracle 节点属性	值	属性描述
use_prediction_set	布尔值	

Oracle 朴素贝叶斯

以下属性可用于类型为 oranb 的节点。

表 150. oranb 节点属性.

oranb 节点属性	值	属性描述
singleton_threshold	数字	0.0-1.0.*
pairwise_threshold	数字	0.0-1.0.*
priors	Data Equal Custom	
custom_priors	结构化	结构化属性。

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle Adaptive Bayes

以下属性可用于类型为 oraabn 的节点。

表 151. oraabn 节点属性.

oraabn 节点属性	值	属性描述
model_type	SingleFeature MultiFeature NaiveBayes	
use_execution_time_limit	布尔值	*
execution_time_limit	整数	值必须大于 0。*
max_naive_bayes_predictors	整数	值必须大于 0。*
max_predictors	整数	值必须大于 0。*
priors	Data Equal Custom	
custom_priors	结构化	结构化属性。

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle 支持向量机

以下属性可用于类型为 orasvm 的节点。

表 152. orasvm 节点属性.

orasvm 节点属性	值	属性描述
active_learning	Enable Disable	

表 152. orasvm 节点属性 (续).

orasvm 节点属性	值	属性描述
kernel_function	Linear Gaussian System	
normalization_method	zscore minmax none	
kernel_cache_size	整数	仅适用于高斯内核。值必须大于 0。*
convergence_tolerance	数字	值必须大于 0。*
use_standard_deviation	布尔值	仅适用与高斯内核。*
standard_deviation	数字	值必须大于 0。*
use_epsilon	布尔值	仅适用于回归模型。*
epsilon	数字	值必须大于 0。*
use_complexity_factor	布尔值	*
complexity_factor	数字	*
use_outlier_rate	布尔值	仅适用于单类变体。*
outlier_rate	数字	仅适用于单类变体。0.0-1.0.*
weights	Data Equal Custom	
custom_weights	结构化	结构化属性。

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle 广义线性模型

以下属性可用于类型为 oraglm 的节点。

表 153. oraglm 节点属性.

oraglm 节点属性	值	属性描述
normalization_method	zscore minmax none	
missing_value_handling	ReplaceWithMean UseCompleteRecords	
use_row_weights	布尔值	*
row_weights_field	字段	*
save_row_diagnostics	布尔值	*
row_diagnostics_table	字符串	*
coefficient_confidence	数字	*
use_reference_category	布尔值	*
reference_category	字符串	*

表 153. oraglm 节点属性 (续).

oraglm 节点属性	值	属性描述
ridge_regression	Auto Off On	*
parameter_value	数字	*
vif_for_ridge	布尔值	*

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle 决策树

以下属性可用于类型为 oradecisiontree 的节点。

表 154. oradecisiontree 节点属性.

oradecisiontree 节点属性	值	属性描述
use_costs	布尔值	
impurity_metric	Entropy Gini	
term_max_depth	整数	2-20.*
term_minpct_node	数字	0.0-10.0.*
term_minpct_split	数字	0.0-20.0.*
term_minrec_node	整数	值必须大于 0。*
term_minrec_split	整数	值必须大于 0。*
display_rule_ids	布尔值	*

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle O-Cluster

以下属性可用于类型为 oraocluster 的节点。

表 155. oraocluster 节点属性.

oraocluster 节点属性	值	属性描述
max_num_clusters	整数	值必须大于 0。
max_buffer	整数	值必须大于 0。*
sensitivity	数字	0.0-1.0.*

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle KMeans

以下属性可用于类型为 orakmeans 的节点。

表 156. orakmeans 节点属性.

orakmeans 节点属性	值	属性描述
num_clusters	整数	值必须大于 0。

表 156. orakmeans 节点属性 (续).

orakmeans 节点属性	值	属性描述
normalization_method	zscore minmax none	
distance_function	Euclidean Cosine	
iterations	整数	0-20.*
conv_tolerance	数字	0.0-0.5.*
split_criterion	Variance Size	缺省值为 Variance。*
num_bins	整数	值必须大于 0。*
block_growth	整数	1-5.*
min_pct_attr_support	数字	0.0-1.0.*

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle NMF

以下属性可用于类型为 oranmf 的节点。

表 157. oranmf 节点属性.

oranmf 节点属性	值	属性描述
normalization_method	minmax none	
use_num_features	布尔值	*
num_features	整数	0-1。缺省值由算法根据数据估计得出。*
random_seed	数字	*
num_iterations	整数	0-500.*
conv_tolerance	数字	0.0-0.5.*
display_all_features	布尔值	*

* 如果 mode 设置为 Simple, 那么忽略属性。

Oracle Apriori

以下属性可用于类型为 oraapriori 的节点。

表 158. oraapriori 节点属性.

oraapriori 节点属性	值	属性描述
content_field	字段	
id_field	字段	
max_rule_length	整数	2-20。
min_confidence	数字	0.0-1.0。

表 158. oraapriori 节点属性 (续).

oraapriori 节点属性	值	属性描述
min_support	数字	0.0-1.0。
use_transactional_data	布尔值	

Oracle 最小描述长度 (MDL)

未对类型为 oramd1 的节点定义具体属性。请参阅本章节开头部分的通用 Oracle 属性。

Oracle 属性重要性 (AI)

以下属性可用于类型为 oraai 的节点。

表 159. oraai 节点属性.

oraai 节点属性	值	属性描述
custom_fields	布尔值	如果为 true, 那么允许您为当前节点指定目标字段、输入字段和其他字段。如果为 false, 那么将使用上游“类型”节点的当前设置。
selection_mode	ImportanceLevel ImportanceValue TopN	
select_important	布尔值	在 selection_mode 设置为 ImportanceLevel 时, 指定是否选择“重要”字段。
important_label	字符串	指定“重要”排序的标签。
select_marginal	布尔值	在 selection_mode 设置为 ImportanceLevel 时, 指定是否选择“边际”字段。
marginal_label	字符串	指定“边际”排序的标签。
important_above	数字	0.0-1.0。
select_unimportant	布尔值	在 selection_mode 设置为 ImportanceLevel 时, 指定是否选择“不重要”字段。
unimportant_label	字符串	指定“不重要”排序的标签。
unimportant_below	数字	0.0-1.0。
importance_value	数字	在 selection_mode 设置为 ImportanceValue 时, 指定要使用的分界值。接受从 0 到 100 的值。
top_n	数字	在 selection_mode 设置为 TopN 时, 指定要使用的分界值。接受从 0 到 1000 的值。

Oracle 模型块属性

使用 Oracle 模型创建的模型块具有下列属性。

Oracle 朴素贝叶斯

未对类型为 applyoranb 的节点定义具体属性。

Oracle Adaptive Bayes

未对类型为 applyoraabn 的节点定义具体属性。

Oracle 支持向量机

未对类型为 `applyorasvm` 的节点定义具体属性。

Oracle 决策树

以下属性可用于类型为 `applyoradecisiontree` 的节点。

表 160. `applyoradecisiontree` 节点属性.

applyoradecisiontree 节点属性	值	属性描述
<code>use_costs</code>	布尔值	
<code>display_rule_ids</code>	布尔值	

Oracle O-Cluster

未对类型为 `applyoraocluster` 的节点定义具体属性。

Oracle KMeans

未对类型为 `applyorakmeans` 的节点定义具体属性。

Oracle NMF

以下属性可用于类型为 `applyoranmf` 的节点:

表 161. `applyoranmf` 节点属性.

applyoranmf 节点属性	值	属性描述
<code>display_all_features</code>	布尔值	

Oracle Apriori

该模型块不能应用于脚本。

Oracle MDL

该模型块不能应用于脚本。

IBM DB2 建模节点属性

IBM DB2 建模节点属性

IBM InfoSphere Warehouse (ISW) 数据库建模节点的公共属性如下所示。

表 162. 公共 ISW 节点属性.

公共 ISW 节点属性	值	属性描述
<code>inputs</code>	字段的列表	
<code>datasource</code>		
<code>username</code>		
<code>password</code>		
<code>epassword</code>		

表 162. 公共 ISW 节点属性 (续).

公共 ISW 节点属性	值	属性描述
enable_power_options	布尔值	
power_options_max_memory	整数	值必须大于 32。
power_options_cmdline	字符串	
mining_data_custom_sql	字符串	
logical_data_custom_sql	字符串	
mining_settings_custom_sql		

ISW 决策树

以下属性可用于类型为 db2imtree 的节点。

表 163. db2imtree 节点属性.

db2imtree 节点属性	值	属性描述
target	字段	
perform_test_run	布尔值	
use_max_tree_depth	布尔值	
max_tree_depth	整数	大于 0 的值。
use_maximum_purity	布尔值	
maximum_purity	数字	介于 0 与 100 之间的数字。
use_minimum_internal_cases	布尔值	
minimum_internal_cases	整数	大于 1 的值。
use_costs	布尔值	
costs	结构化	结构化属性。

ISW 关联

以下属性可用于类型为 db2imassoc 的节点。

表 164. db2imassoc 节点属性.

db2imassoc 节点属性	值	属性描述
use_transactional_data	布尔值	
id_field	字段	
content_field	字段	
data_table_layout	basic limited_length	
max_rule_size	整数	值必须大于 2。
min_rule_support	数字	0-100%
min_rule_confidence	数字	0-100%
use_item_constraints	布尔值	
item_constraints_type	Include Exclude	
use_taxonomy	布尔值	

表 164. db2imassoc 节点属性 (续).

db2imassoc 节点属性	值	属性描述
taxonomy_table_name	字符串	这是用于存储分类法详细信息的 DB2 表的名称。
taxonomy_child_column_name	字符串	分类法表中子列的名称。该子列包含项目名或类别名。
taxonomy_parent_column_name	字符串	分类法表中父列的名称。该父列包含类别名。
load_taxonomy_to_table	布尔值	控制是否应在构建模型时将 IBM SPSS Modeler 中存储的分类法信息上载至分类法表。请注意，如果分类法表已经存在，那么会将其丢弃。分类法信息存储在模型构建节点中，可以使用 编辑类别 和 编辑分类法 按钮进行编辑。

ISW 序列

以下属性可用于类型为 db2imsequence 的节点。

表 165. db2imsequence 节点属性.

db2imsequence 节点属性	值	属性描述
id_field	字段	
group_field	字段	
content_field	字段	
max_rule_size	整数	值必须大于 2。
min_rule_support	数字	0-100%
min_rule_confidence	数字	0-100%
use_item_constraints	布尔值	
item_constraints_type	Include Exclude	
use_taxonomy	布尔值	
taxonomy_table_name	字符串	这是用于存储分类法详细信息的 DB2 表的名称。
taxonomy_child_column_name	字符串	分类法表中子列的名称。该子列包含项目名或类别名。
taxonomy_parent_column_name	字符串	分类法表中父列的名称。该父列包含类别名。
load_taxonomy_to_table	布尔值	控制是否应在构建模型时将 IBM SPSS Modeler 中存储的分类法信息上载至分类法表。请注意，如果分类法表已经存在，那么会将其丢弃。分类法信息存储在模型构建节点中，可以使用 编辑类别 和 编辑分类法 按钮进行编辑。

ISW 回归

以下属性可用于类型为 db2imreg 的节点。

表 166. db2imreg 节点属性.

db2imreg 节点属性	值	属性描述
target	字段	
regression_method	transform linear polynomial rbf	请参阅下表，以了解仅当 regression_method 设置为 rbf 时才适用的属性。
perform_test_run	字段	

表 166. db2imreg 节点属性 (续).

db2imreg 节点属性	值	属性描述
limit_rsquared_value	布尔值	
max_rsquared_value	数字	介于 0.0 与 1.0 之间的值。
use_execution_time_limit	布尔值	
execution_time_limit_mins	整数	大于 0 的值。
use_max_degree_polynomial	布尔值	
max_degree_polynomial	整数	
use_intercept	布尔值	
use_auto_feature_selection_method	布尔值	
auto_feature_selection_method	normal adjusted	
use_min_significance_level	布尔值	
min_significance_level	数字	
use_min_significance_level	布尔值	

下列属性只有在 regression_method 设置为 rbf 时才适用。

表 167. db2imreg 节点属性 (如果 regression_method 设置为 rbf) .

db2imreg 节点属性	值	属性描述
use_output_sample_size	布尔值	如果为 true, 那么将值自动设置为缺省值。
output_sample_size	整数	缺省值为 2。 最小值为 1。
use_input_sample_size	布尔值	如果为 true, 那么将值自动设置为缺省值。
input_sample_size	整数	缺省值为 2。 最小值为 1。
use_max_num_centers	布尔值	如果为 true, 那么将值自动设置为缺省值。
max_num_centers	整数	缺省值为 20。 最小值为 1。
use_min_region_size	布尔值	如果为 true, 那么将值自动设置为缺省值。
min_region_size	整数	缺省值为 15。 最小值为 1。
use_max_data_passes	布尔值	如果为 true, 那么将值自动设置为缺省值。
max_data_passes	整数	缺省值为 5。 最小值为 2。
use_min_data_passes	布尔值	如果为 true, 那么将值自动设置为缺省值。
min_data_passes	整数	缺省值为 5。 最小值为 2。

ISW 聚类

以下属性可用于类型为 db2imcluster 的节点。

表 168. db2imcluster 节点属性.

db2imcluster 节点属性	值	属性描述
cluster_method	demographic kohonen birch	
kohonen_num_rows	整数	
kohonen_num_columns	整数	
kohonen_passes	整数	
use_num_passes_limit	布尔值	
use_num_clusters_limit	布尔值	
max_num_clusters	整数	大于 1 的值。
birch_dist_measure	log_likelihood euclidean	缺省值为 log_likelihood。
birch_num_cfleaves	整数	缺省值为 1000。
birch_num_refine_passes	整数	缺省值为 3; 最小值为 1。
use_execution_time_limit	布尔值	
execution_time_limit_mins	整数	大于 0 的值。
min_data_percentage	数字	0-100%
use_similarity_threshold	布尔值	
similarity_threshold	数字	介于 0.0 与 1.0 之间的值。

ISW 朴素贝叶斯

以下属性可用于类型为 db2imnbs 的节点。

表 169. db2imnb 节点属性.

db2imnb 节点属性	值	属性描述
perform_test_run	布尔值	
probability_threshold	数字	缺省值为 0.001。 最小值为 0; 最大值为 1.000
use_costs	布尔值	
costs	结构化	结构化属性。

ISW Logistic 回归

以下属性可用于类型为 db2imlog 的节点。

表 170. db2imlog 节点属性.

db2imlog 节点属性	值	属性描述
perform_test_run	布尔值	
use_costs	布尔值	
costs	结构化	结构化属性。

ISW 时间序列

注意: 输入字段参数不用于此节点。如果在脚本中找到输入字段参数, 那么显示一个警告, 说明该节点有时间和目标作为输入字段, 但没有输入字段。

以下属性可用于类型为 `db2imtimeseries` 的节点。

表 171. `db2imtimeseries` 节点属性.

db2imtimeseries 节点属性	值	属性描述
<code>time</code>	字段	允许使用整数、时间或日期。
<code>targets</code>	字段的列表	
<code>forecasting_algorithm</code>	arima exponential_smoothing seasonal_trend_decomposition	
<code>forecasting_end_time</code>	auto 整数 日期 time	
<code>use_records_all</code>	布尔值	如果假, 必须设置 <code>use_records_start</code> 和 <code>use_records_end</code> 。
<code>use_records_start</code>	整数/时间/日期	取决于时间字段的类型
<code>use_records_end</code>	整数/时间/日期	取决于时间字段的类型
<code>interpolation_method</code>	none linear exponential_splines cubic_splines	

IBM DB2 模型块属性

使用 IBM DB2 ISW 模型创建的模型块具有下列属性。

ISW 决策树

未对类型为 `applydb2imtree` 的节点定义具体属性。

ISW 关联

该模型块不能应用于脚本。

ISW 序列

该模型块不能应用于脚本。

ISW 回归

未对类型为 `applydb2imreg` 的节点定义具体属性。

ISW 聚类

未对类型为 `applydb2imcluster` 的节点定义具体属性。

ISW 朴素贝叶斯

未对类型为 applydb2imnb 的节点定义具体属性。

ISW Logistic 回归

未对类型为 applydb2imlog 的节点定义具体属性。

ISW 时间序列

该模型块不能应用于脚本。

IBM Netezza Analytics 建模节点属性

Netezza 建模节点属性

IBM Netezza 数据库建模节点的公共属性如下所示。

表 172. 公共 Netezza 节点属性。

公共 Netezza 节点属性	值	属性描述
custom_fields	布尔值	如果为 true, 那么允许您为当前节点指定目标字段、输入字段和其他字段。如果为 false, 那么将使用上游“类型”节点的当前设置。
inputs	[field1 ... fieldN]	模型所使用的输入字段或预测变量字段。
target	字段	目标字段 (连续或分类)。
record_id	字段	要用作唯一记录标识的字段。
use_upstream_connection	布尔值	如果为 true (缺省), 那么连接详细信息在上游节点中指定。在指定了 move_data_to_connection 时不使用。
move_data_connection	布尔值	如果为 true, 那么将数据移动到由 connection 指定的数据库。在指定了 use_upstream_connection 时不使用。
connection	结构化	这是用于存储模型的 Netezza 数据库的连接字符串。格式如下的结构化属性: ['odbc' '<dsn>' '<username>' '<psw>' '<catname>' '<conn_attribs>' {true false}] 其中: <dsn> 是数据源名称 <username> 和 <psw> 是数据库的用户名和密码 <catname> 是目录名称 <conn_attribs> 是连接属性 true false 指示是否需要密码。
table_name	字符串	这是用于存储模型的数据库表的名称。
use_model_name	布尔值	如果为 true, 使用由 model_name 指定的名称作为模型名称, 否则采用系统创建的模型名称。
model_name	字符串	新模型的定制名称。
include_input_fields	布尔值	如果为 true, 向下游传递所有输入字段, 否则仅传递模型产生的 record_id 和字段。

Netezza 决策树

以下属性可用于类型为 netezzadectree 的节点。

表 173. netezadectree 节点属性.

netezadectree 节点属性	值	属性描述
impurity_measure	Entropy Gini	杂质测量，用于评估分割树的最佳位置。
max_tree_depth	整数	树可以增长到的最大级别数。缺省值为 62（最大可能值）。
min_improvement_splits	数字	进行分割前必须满足的最低杂质改进。缺省值为 0.01。
min_instances_split	整数	可以进行分割前余下的最小未分割记录数。缺省值为 2（最小可能值）。
weights	结构化	各个类的相对权重。结构化属性。缺省情况是所有类的权重均为 1。
pruning_measure	Acc wAcc	缺省值为 Acc（准确性）。如果要在应用修剪时将类权重考虑在内，可使用 wAcc（加权精确度）替代。
prune_tree_options	allTrainingData partitionTrainingData useOtherTable	缺省情况下，使用 allTrainingData 来估计模型精确度。使用 partitionTrainingData 来指定要使用训练数据的百分比，或 useOtherTable 来使用源自指定数据库表的训练数据集。
perc_training_data	数字	如果 prune_tree_options 设置为 partitionTrainingData，那么指定用于训练的数据所占的百分比。
prune_seed	整数	在 prune_tree_options 设置为 partitionTrainingData 时，用于重复分析结果的随机种子，缺省值为 1。
pruning_table	字符串	这是用于估算模型精确性的独立修剪数据集的表名称。
compute_probabilities	布尔值	如果为 true，那么将生成置信度级别（概率）字段以及预测字段。

Netezza K-Means

以下属性可用于类型为 netezakmeans 的节点。

表 174. netezakmeans 节点属性.

netezakmeans 节点属性	值	属性描述
distance_measure	Euclidean Manhattan Canberra maximum	这是用于对数据点之间的距离进行测量的方法。
num_clusters	整数	要创建的聚类数；缺省值为 3。
max_iterations	整数	算法迭代次数，模型训练在此之后停止；缺省值为 5。
rand_seed	整数	这是用于复制分析结果的随机种子；缺省值为 12345。

Netezza Bayes 网络

以下属性可用于类型为 `netezababes` 的节点。

表 175. `netezababes` 节点属性.

netezababes 节点属性	值	属性描述
<code>base_index</code>	整数	对第一个输入字段指定的数字标识，用于进行内部管理；缺省值为 777。
<code>sample_size</code>	整数	属性数目非常大时采用的样本大小；缺省值为 10,000。
<code>display_additional_information</code>	布尔值	如果为 <code>true</code> ，那么在消息对话框中显示额外的进度信息。
<code>type_of_prediction</code>	best neighbors nn-neighbors	要使用的预测算法类型：best（最相关的相邻值）、neighbors（相邻值的加权预测）或 nn-neighbors（非空相邻值）。

Netezza 朴素贝叶斯

以下属性可用于类型为 `netezanaivebayes` 的节点。

表 176. `netezanaivebayes` 节点属性.

netezanaivebayes 节点属性	值	属性描述
<code>compute_probabilities</code>	布尔值	如果为 <code>true</code> ，那么将生成置信度级别（概率）字段以及预测字段。
<code>use_m_estimation</code>	布尔值	如果为 <code>true</code> ，那么使用 m-estimation 技术以避免估算期间的零概率。

Netezza KNN

以下属性可用于类型为 `netezaknn` 的节点。

表 177. `netezaknn` 节点属性.

netezaknn 节点属性	值	属性描述
<code>weights</code>	结构化	这是用于对各个类指定权重的结构化属性。
<code>distance_measure</code>	Euclidean Manhattan Canberra Maximum	这是用于对数据点之间的距离进行测量的方法。
<code>num_nearest_neighbors</code>	整数	特定个案的最近相邻元素数；缺省值为 3。
<code>standardize_measurements</code>	布尔值	如果为 <code>true</code> ，那么在计算距离值之前，对连续输入字段的测量值进行标准化。
<code>use_coresets</code>	布尔值	如果为 <code>true</code> ，那么对大型数据集使用核心集采样以提高计算速度。

Netezza 分裂式聚类

以下属性可用于类型为 `netezadivcluster` 的节点。

表 178. netezadivcluster 节点属性.

netezadivcluster 节点属性	值	属性描述
distance_measure	Euclidean Manhattan Canberra Maximum	这是用于对数据点之间的距离进行测量的方法。
max_iterations	整数	在模型训练停止前执行的最大算法迭代次数; 缺省值为 5。
max_tree_depth	整数	可以将数据集拆分为的最大级别数; 缺省值为 3。
rand_seed	整数	随机种子, 用于复制分析; 缺省值为 12345。
min_instances_split	整数	可以拆分的最小记录数, 缺省值为 5。
level	整数	要将记录评分到的层次结构级别; 缺省值为 -1。

Netezza PCA

以下属性可用于类型为 netezzapca 的节点。

表 179. netezzapca 节点属性.

netezzapca 节点属性	值	属性描述
center_data	布尔值	如果为 true (缺省值), 那么先执行数据集中 (也称为“平均值消去法”), 然后再执行分析。
perform_data_scaling	布尔值	如果为 true, 那么在分析前执行数据换算。这样做可以减小不同变量以不同单位测量时的分析任意性。
force_eigensolve	布尔值	如果为 true, 那么使用不太准确但较快的方法来查找主成份。
pc_number	整数	要将数据集精简为主成份数; 缺省值为 1。

Netezza 回归树

以下属性可用于类型为 netezzaregtree 的节点。

表 180. netezzaregtree 节点属性.

netezzaregtree 节点属性	值	属性描述
max_tree_depth	整数	树在根节点下可以增长到的最大层数; 缺省值为 10。
split_evaluation_measure	Variance	类杂质测量, 用于评估分割树的最佳位置, 缺省值 (当前唯一选项) 是 Variance。
min_improvement_splits	数字	在树中进行新拆分前要将杂质减少到的最小数量。
min_instances_split	整数	可以拆分的最小记录数。
pruning_measure	mse r2 pearson spearman	要使用的修剪方法

表 180. netezzaregtree 节点属性 (续).

netezzaregtree 节点属性	值	属性描述
prune_tree_options	allTrainingData partitionTrainingData useOtherTable	缺省情况下, 使用 allTrainingData 来估计模型精确度。使用 partitionTrainingData 来指定要使用训练数据的百分比, 或 useOtherTable 来使用源自指定数据库表的训练数据集。
perc_training_data	数字	如果 prune_tree_options 设置为 PercTrainingData, 那么指定用于训练的数据所占的百分比。
prune_seed	整数	在 prune_tree_options 设置为 PercTrainingData 时, 用于重复分析结果的随机种子, 缺省值为 1。
pruning_table	字符串	这是用于估算模型精确性的独立修剪数据集的表名称。
compute_probabilities	布尔值	如果为 true, 那么指定应该包括在输出中的指定类的方差。

Netezza 线性回归

以下属性可用于类型为 netezzalinereregression 的节点。

表 181. netezzalinereregression 节点属性.

netezzalinereregression 节点属性	值	属性描述
use_svd	布尔值	如果为 true, 那么使用“奇异值分解”矩阵代替原始矩阵, 以便提高速度和数字准确性。
include_intercept	布尔值	如果为 true (缺省值), 那么提高解的整体准确性。
calculate_model_diagnostics	布尔值	如果为 true, 那么对模型计算诊断信息。

Netezza 时间序列

以下属性可用于类型为 netezzatimeseries 的节点。

表 182. netezzatimeseries 节点属性.

netezzatimeseries 节点属性	值	属性描述
time_points	字段	这是包含时间序列的日期或时间值的输入字段。
time_series_ids	字段	此输入字段包含时间序列 ID; 在输入包含多个时间序列时使用。
model_table	字段	这是用于存储 Netezza 时间序列模型的数据库表的名称。
description_table	字段	这是包含时间序列名称和描述的输入表的名称。
seasonal_adjustment_table	字段	这是一个输出表的名称, 该表用于存储指数平滑或季节性趋势分解算法所计算的按季度调整值。

表 182. netezzatimeseries 节点属性 (续).

netezzatimeseries 节点属性	值	属性描述
algorithm_name	SpectralAnalysis 或 spectral ExponentialSmoothing 或 esmoothing ARIMA SeasonalTrendDecomposition 或 std	这是用于时间序列建模的算法。
trend_name	N A DA M DM	指数平滑的趋势类型: N - 无 A - 加性 DA - 衰减加性 M - 乘性 DM - 衰减乘性
seasonality_type	N A M	指数平滑的季节性类型: N - 无 A - 加性 M - 乘性
interpolation_method	linear cubicspline exponentialspline	要使用的插值方法。
timerange_setting	SD SP	用于设置要使用的时间范围: SD - 由系统确定 (使用时间序列数据的完整范围) SP - 用户通过 earliest_time 和 latest_time 指定
earliest_time	日期	开始时间和结束时间 (如果 timerange_setting 为 SP)。 格式: <yyyy>-<mm>-<dd>
latest_time		
arima_setting	SD SP	用于设置 ARIMA 算法 (仅当 algorithm_name 设置为 ARIMA 时才使用): SD - 由系统确定 SP - 由用户指定 如果 arima_setting = SP, 请使用下列参数来设置季节性值和非季节性值。
p_symbol	less eq lesseq	ARIMA - 参数 p、d、q、sp、sd 和 sq 的运算符: less - 小于 eq - 等于 lesseq - 小于或等于
d_symbol		
q_symbol		
sp_symbol		
sd_symbol		
sq_symbol		
p	整数	ARIMA - 非季节性自动关联度。
q	整数	ARIMA - 非季节性派生值。
d	整数	ARIMA - 模型中的非季节性移动平均阶数。
sp	整数	ARIMA - 季节性自动关联度。

表 182. netezatimeseries 节点属性 (续).

netezatimeseries 节点属性	值	属性描述
sq	整数	ARIMA - 季节性派生值。
sd	整数	ARIMA - 模型中的季节性移动平均值阶数。
advanced_setting	SD SP	确定如何处理高级设置: SD - 由系统确定 SP - 由用户通过 period、units_period 和 forecast_setting 指定。
period	整数	季节周期的长度, 与 units_period 一起指定。不适用于谱分析。
units_period	ms s min h d wk q y	period 的表示单位: ms - 毫秒 s - 秒 min - 分钟 h - 小时 d - 天 wk - 星期 q - 季度 y - 年 例如, 对于每周时间序列, 请对 period 使用 1, 并对 units_period 使用 wk。
forecast_setting	forecasthorizon forecasttimes	指定如何进行预测。
forecast_horizon	字符串	如果 forecast_setting = forecasthorizon, 那么指定预测结束点。 格式: <yyyy>-<mm>-<dd>
forecast_times	[{'date'}, {'date'},..., {'date'}]	如果 forecast_setting = forecasttimes, 那么指定用于进行预测的时间。 格式: <yyyy>-<mm>-<dd>
include_history	布尔值	指示是否将历史值包括在输出中。
include_interpolated_values	布尔值	指示是否将内插值包括在输出中。如果 include_history 为 false, 那么不适用。

Netezza 广义线性

以下属性可用于类型为 netezaglm 的节点。

表 183. netezzaglm 节点属性.

netezzaglm 节点属性	值	属性描述
dist_family	bernoulli gaussian poisson negativebinomial wald gamma	分布类型; 缺省值为 bernoulli。
dist_params	数字	要使用的分布参数值。仅当 distribution 为 Negativebinomial 时才适用。
trials	整数	仅当 distribution 为 Binomial 时才适用。当目标响应为发生在一组试验中的事件数时, target 字段包含事件数, trials 字段包含试验数。
model_table	字段	这是用于存储 Netezza 广义线性模型的数据库表的名称。
maxit	整数	算法应执行的最大迭代次数; 缺省值为 20。
eps	数字	指定最大误差值 (以科学记数法表示), 达到此值后, 算法应停止查找最佳匹配模型。缺省值为 -3, 这表示 1E-3, 即 0.001。
tol	数字	设置数值 (用科学表示法), 低于此值的所有误差均被视为 0 值。缺省值为 -7, 表示误差值若低于 1E-7 (或 0.0000001), 那么被视为不显著。
link_func	identity inverse invnegative invsquare sqrt power oddspower log clog loglog cloglog logit probit gaussit cauchit canbinom cangeom cannegbinom	要使用的联接函数; 缺省值为 logit。
link_params	数字	要使用的关联函数参数值。仅当 link_function 为 power 或 oddspower 时才适用。

表 183. netezzaglm 节点属性 (续).

netezzaglm 节点属性	值	属性描述
interaction	[[[colnames1],[levels1]],[colnames2],[levels2]],...,[colnamesN],[levelsN]],]	指定字段之间的交互。colnames 是输入字段的列表，而对于每个字段，level 始终为 0。
intercept	布尔值	如果为 true，那么在模型中包括截距。

Netezza 模型块属性

Netezza 数据库模型块的公共属性如下所示。

表 184. 公共 Netezza 模型块属性.

公共 Netezza 模型块属性	值	属性描述
connection	字符串	这是用于存储模型的 Netezza 数据库的连接字符串。
table_name	字符串	这是用于存储模型的数据库表的名称。

其他模型块的属性与相应建模节点的属性相同。

模型块的脚本名称如下所示。

表 185. Netezza 模型块的脚本名称.

模型块	脚本名称
决策树	applynetezzagtree
K-Means	applynetezzagkmeans
贝叶斯网络	applynetezzagbayes
朴素贝叶斯	applynetezzagnaivebayes
KNN	applynetezzagknn
分裂式聚类	applynetezzagdivcluster
PCA	applynetezzagpca
回归树	applynetezzagregtree
线性回归	applynetezzaglineregression
时间序列	applynetezzagtimeseries
广义线性	applynetezzagglm

第 16 章 输出节点属性

输出节点的属性与其他“类型”节点的属性略有不同。输出节点属性不是指特定的节点选项，而是存储对输出对象的引用。这在从表中提取值并将其设置为流参数时非常有用。

本节说明输出节点的可用的脚本属性。

analysis 节点属性



“分析”节点评估预测模型生成准确预测的能力。“分析”节点执行一个或多个模型块的预测值和实际值之间的各种比较。“分析”节点也可以对比各个预测模型。

表 186. analysis 节点属性.

analysis 节点属性	数据类型	属性描述
output_mode	Screen File	用于指定输出节点中生成的输出的目标位置。
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	如果 use_output_name 为 true ，那么指定使用的名称。
output_format	Text (.txt) HTML (.html) Output (.cou)	用于指定输出类型。
by_fields	[field field field]	
full_filename	字符串	如果是磁盘、数据或 HTML 输出，那么此属性指输出文件的名称。
coincidence	布尔值	
performance	布尔值	
evaluation_binary	布尔值	
confidence	布尔值	
threshold	数字	
improve_accuracy	数字	
inc_user_measure	布尔值	
user_if	表达式	
user_then	表达式	
user_else	表达式	
user_compute	[Mean Sum Min Max SDev]	

dataaudit 节点属性



“数据审核”节点提供有关数据的全面概览，包括每个字段的汇总统计、直方图和分布以及有关离群值、缺失值和极值的信息。结果显示在易于读取的矩阵中，该矩阵可以排序并且可以用于生成完整大小的图表和数据准备节点。

表 187. dataaudit 节点属性.

dataaudit 节点属性	数据类型	属性描述
custom_fields	布尔值	
fields	[field1 ... fieldN]	
overlay	字段	
display_graphs	布尔值	用于打开或关闭输出矩阵中图形的显示。
basic_stats	布尔值	
advanced_stats	布尔值	
median_stats	布尔值	
calculate	Count Breakdown	用于计算缺失值。选择两种计算方法中的一种、两种，或均不选择。
outlier_detection_method	std iqr	用于指定离群值和极值的检测方法。
outlier_detection_std_outlier	数字	如果 outlier_detection_method 是 std，那么指定用于定义离群值的数值。
outlier_detection_std_extreme	数字	如果 outlier_detection_method 是 std，那么指定用于定义极值的数值。
outlier_detection_iqr_outlier	数字	如果 outlier_detection_method 是 iqr，那么指定用于定义离群值的数值。
outlier_detection_iqr_extreme	数字	如果 outlier_detection_method 是 iqr，那么指定用于定义极值的数值。
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	如果 use_output_name 为 true，那么指定使用的名称。
output_mode	Screen File	用于指定输出节点中生成的输出的目标位置。
output_format	Formatted (.tab) Delimited (.csv) HTML (.html) Output (.cou)	用于指定输出类型。
paginate_output	布尔值	当 output_format 是 HTML 时，使输出分页。
lines_per_page	数字	与 paginate_output 一起使用时，指定每个输出页中的行数。

表 187. dataaudit 节点属性 (续).

dataaudit 节点属性	数据类型	属性描述
full_filename	字符串	

matrix 节点属性



“矩阵”节点创建一个表，用于显示字段之间的关系。此节点最常用于显示两个符号字段之间的关系，但也可用于显示标志字段或数字字段之间的关系。

表 188. matrix 节点属性.

matrix 节点属性	数据类型	属性描述
fields	Selected Flags Numerics	
row	字段	
column	字段	
include_missing_values	布尔值	指定在行和列输出中是否包含用户缺失值（空白）和系统缺失值（空值）。
cell_contents	CrossTabs Function	
function_field	字符串	
function	Sum Mean Min Max SDev	
sort_mode	Unsorted Ascending Descending	
highlight_top	数字	如果非零，那么为 true。
highlight_bottom	数字	如果非零，那么为 true。
display	[Counts 期望值(E) 残差 (Residuals) RowPct ColumnPct TotalPct]	
include_totals	布尔值	
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	如果 use_output_name 为 true，那么指定使用的名称。
output_mode	Screen File	用于指定输出节点中生成的输出的目标位置。

表 188. matrix 节点属性 (续).

matrix 节点属性	数据类型	属性描述
output_format	Formatted (.tab) Delimited (.csv) HTML (.html) Output (.cou)	用于指定输出类型。Formatted 和 Delimited 格式都可使用修饰符 transposed, 此符号可转置表中的行和列。
paginate_output	布尔值	当 output_format 是 HTML 时, 使输出分页。
lines_per_page	数字	与 paginate_output 一起使用时, 指定每个输出页中的行数。
full_filename	字符串	

means 节点属性



“均值”节点在独立组之间或者相关字段的配对之间进行均值比较, 以检验是否存在显著差别。例如, 您可以比较开展促销 前后的平均收入, 或者将来自未接受促销客户的收入与接受促销客户的收入进行比较。

表 189. means 节点属性.

means 节点属性	数据类型	属性描述
means_mode	BetweenGroups BetweenFields	指定要在数据上执行的均值统计的类型。
test_fields	[field1 ... fieldn]	指定当 means_mode 设置为 BetweenGroups 时的检验字段。
grouping_field	字段	指定分组字段。
paired_fields	[{field1 field2} {field3 field4} ...]	指定当 means_mode 设置为 BetweenFields 时要使用的字段对。
label_correlations	布尔值	指定在输出中是否显示相关标签。仅在当 means_mode 设置为 BetweenFields 时才应用此设置。
correlation_mode	Probability Absolute	指定用概率还是绝对值标注相关。
weak_label	字符串	
medium_label	字符串	
strong_label	字符串	
weak_below_probability	数字	当 correlation_mode 设置为 Probability 时, 指定弱相关的分界值。这必须是 0 到 1 之间的一个值, 例如 0.90。
strong_above_probability	数字	强相关的分界值。
weak_below_absolute	数字	当 correlation_mode 设置为 Absolute 时, 指定弱相关的分界值。这必须是 0 到 1 之间的一个值, 例如 0.90。

表 189. means 节点属性 (续).

means 节点属性	数据类型	属性描述
strong_above_absolute	数字	强相关的分界值。
unimportant_label	字符串	
marginal_label	字符串	
important_label	字符串	
unimportant_below	数字	低字段重要性的分界值。这必须是 0 到 1 之间的一个值，例如 0.90。
important_above	数字	
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	使用的名称。
output_mode	Screen File	指定从输出节点中生成的输出的目标位置。
output_format	Formatted (.tab) Delimited (.csv) HTML (.html) Output (.cou)	指定输出类型。
full_filename	字符串	
output_view	Simple Advanced	指定在输出中显示简单视图还是高级视图。

report 节点属性



“报告”节点可创建格式化报告，其中包含固定文本、数据及得自数据的其他表达式。您可以使用文本模板指定报告的格式，以定义固定文本和数据输出构造。通过在模板中使用 HTML 标记以及在“输出”选项卡上设置选项，可以提供定制文本格式。通过使用模板中的 CLEM 表达式，可以包括数据值和其他条件输出。

表 190. report 节点属性.

report 节点属性	数据类型	属性描述
output_mode	Screen File	用于指定输出节点中生成的输出的目标位置。
output_format	HTML (.html) Text (.txt) Output (.cou)	用于指定输出类型。
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	如果 use_output_name 为 true，那么指定使用的名称。
text	字符串	
full_filename	字符串	
highlights	布尔值	
title	字符串	
lines_per_page	数字	

Routput 节点属性



通过使用您自己的定制 R 脚本，可以使用“R 输出”节点来分析模型评分的数据和结果。分析的输出可以是文本，也可以是图形。输出将添加到管理器窗格的输出选项卡中；另外，可以将输出重定向到文件。

表 191. Routput 节点属性.

Routput 节点属性	数据类型	属性描述
syntax	字符串	
convert_flags	StringsAndDoubles LogicalValues	
convert_datetime	布尔值	
convert_datetime_class	POSIXct POSIXlt	
convert_missing	布尔值	
output_name	Auto Custom	
custom_name	字符串	
output_to	Screen File	
output_type	Graph Text	
full_filename	字符串	
graph_file_type	HTML COU	
text_file_type	HTML TXT COU	

setglobals 节点属性



“设置全局值”节点扫描数据并计算可在 CLEM 表达式中使用的汇总值。例如，可以用该节点为一个名为年龄的字段计算统计并通过插入函数 @GLOBAL_MEAN(age) 在 CLEM 表达式中使用年龄的总均值。

表 192. setglobals 节点属性.

setglobals 节点属性	数据类型	属性描述
globals	[Sum Mean Min Max SDev]	结构化属性
clear_first	布尔值	
show_preview	布尔值	

simeval 节点属性



“模拟评估”节点对指定的预测目标字段进行评估，并显示有关该目标字段的分布和相关性信息。

表 193. simeval 节点属性.

simeval 节点属性	数据类型	属性描述
target	字段	
iteration	字段	
presorted_by_iteration	布尔值	
max_iterations	数字	
tornado_fields	[field1...fieldN]	
plot_pdf	布尔值	
plot_cdf	布尔值	
show_ref_mean	布尔值	
show_ref_median	布尔值	
show_ref_sigma	布尔值	
num_ref_sigma	数字	
show_ref_pct	布尔值	
ref_pct_bottom	数字	
ref_pct_top	数字	
show_ref_custom	布尔值	
ref_custom_values	[number1...numberN]	
category_values	Category Probabilities Both	
category_groups	Categories Iterations	
create_pct_table	布尔值	
pct_table	Quartiles Intervals Custom	
pct_intervals_num	数字	
pct_custom_values	[number1...numberN]	

simfit 节点属性



“模拟拟合”节点检查每个字段中的数据的统计分布，并生成（或更新）“模拟生成”节点，同时将最佳拟合分布分配给每个字段。然后，可以使用“模拟生成”节点来生成模拟数据。

表 194. *simfit* 节点属性.

simfit 节点属性	数据类型	属性描述
build	Node XMLExport Both	
use_source_node_name	布尔值	
source_node_name	字符串	正在生成或更新的源节点的定制名称。
use_cases	All LimitFirstN	
use_case_limit	整数	
fit_criterion	AndersonDarling KolmogorovSmirnov	
num_bins	整数	
parameter_xml_filename	字符串	
generate_parameter_import	布尔值	

statistics 节点属性



“统计”节点提供有关数字字段的基本汇总信息。它计算各个字段的汇总统计以及字段间的相关性。

表 195. *statistics* 节点属性.

statistics 节点属性	数据类型	属性描述
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	如果 use_output_name 为 true , 那么指定使用的名称。
output_mode	Screen File	用于指定输出节点中生成的输出的目标位置。
output_format	Text (.txt) HTML (.html) Output (.cou)	用于指定输出类型。
full_filename	字符串	
examine	[field field field]	
correlate	[field field field]	
statistics	[Count Mean Sum Min Max Range Variance SDev SErr Median Mode]	
correlation_mode	Probability Absolute	指定用概率还是绝对值标注相关。
label_correlations	布尔值	
weak_label	字符串	
medium_label	字符串	

表 195. *statistics* 节点属性 (续).

statistics 节点属性	数据类型	属性描述
strong_label	字符串	
weak_below_probability	数字	当 correlation_mode 设置为 Probability 时, 指定弱相关的分界值。这必须是 0 到 1 之间的一个值, 例如 0.90。
strong_above_probability	数字	强相关的分界值。
weak_below_absolute	数字	当 correlation_mode 设置为 Absolute 时, 指定弱相关的分界值。这必须是 0 到 1 之间的一个值, 例如 0.90。
strong_above_absolute	数字	强相关的分界值。

statisticsoutput 节点属性



“Statistics 输出”节点用于调用 IBM SPSS Statistics 过程, 以分析 IBM SPSS Modeler 数据。可以访问许多不同的 IBM SPSS Statistics 分析过程。此节点需要 IBM SPSS Statistics 的许可副本。

有关此节点属性的信息, 请参阅第 216 页的『statisticsoutput 节点属性』。

table 节点属性



“表”节点以表格格式显示数据, 这些数据还可以写入到文件中。每当您需要检查数据值或者采用可轻松阅读的格式导出这些数据值时, 此节点非常有用。

表 196. *table* 节点属性.

table 节点属性	数据类型	属性描述
full_filename	字符串	如果是磁盘、数据或 HTML 输出, 那么此属性指输出文件的名称。
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	如果 use_output_name 为 true, 那么指定使用的名称。
output_mode	Screen File	用于指定输出节点中生成的输出的目标位置。
output_format	Formatted (.tab) Delimited (.csv) HTML (.html) Output (.cou)	用于指定输出类型。
transpose_data	布尔值	导出前转置数据, 使行表示字段, 列表示记录。

表 196. table 节点属性 (续).

table 节点属性	数据类型	属性描述
paginate_output	布尔值	当 output_format 是 HTML 时, 使输出分页。
lines_per_page	数字	与 paginate_output 一起使用时, 指定每个输出页中的行数。
highlight_expr	字符串	
output	字符串	只读属性, 可保留对由节点构建的最后一个表的引用。
value_labels	<i>[{Value LabelString}</i> <i>{Value LabelString} ...]</i>	用于为值对指定标签。
display_places	整数	为字段设置显示的小数位数 (仅用于以 REAL 存储的字段)。值为 -1 时将使用流缺省值。
export_places	整数	为字段设置导出的小数位数 (仅用于以 REAL 存储的字段)。值为 -1 时将使用流缺省值。
decimal_separator	DEFAULT PERIOD COMMA	为字段设置十进制分隔符 (仅用于以 REAL 存储的字段)。
date_format	"DDMMYY" "MMDDYY" "YMMDD" "YYYYMMDD" "YYYYDDD" DAY MONTH "DD-MM-YY" "DD-MM-YYYY" "MM-DD-YY" "MM-DD-YYYY" "DD-MON-YY" "DD-MON-YYYY" "YYYY-MM-DD" "DD.MM.YY" "DD.MM.YYYY" "MM.DD.YY" "MM.DD.YYYY" "DD.MON.YY" "DD.MON.YYYY" "DD/MM/YY" "DD/MM/YYYY" "MM/DD/YY" "MM/DD/YYYY" "DD/MON/YY" "DD/MON/YYYY" MON YYYY q Q YYYY ww WK YYYY	为字段设置日期格式 (仅用于以 DATE 或 TIMESTAMP 存储的字段)。

表 196. table 节点属性 (续).

table 节点属性	数据类型	属性描述
time_format	"HHMMSS" "HHMM" "MMSS" "HH:MM:SS" "HH:MM" "MM:SS" "(H)H:(M)M:(S)S" "(H)H:(M)M" "(M)M:(S)S" "HH.MM.SS" "HH.MM" "MM.SS" "(H)H.(M)M.(S)S" "(H)H.(M)M" "(M)M.(S)S"	为字段设置时间格式（仅用于以 TIME 或 TIMESTAMP 存储的字段）。
column_width	整数	为字段设置列宽度。值为 -1 表示将列宽度设置为 Auto。
justify	AUTO CENTER LEFT RIGHT	为字段设置列对齐格式。

transform 节点属性



“转换”节点允许您先选择并以可视方式预览转换结果，然后再将这些转换应用于选择的字段。

表 197. transform 节点属性.

transform 节点属性	数据类型	属性描述
fields	[field1... fieldn]	要在转换中使用的字段。
formula	All Select	表示应计算所有转换还是选定的转换。
formula_inverse	布尔值	表示是否应使用逆转换。
formula_inverse_offset	数字	表示公式中要使用的数据偏移量。除非用户指定，否则缺省情况下设置为 0。
formula_log_n	布尔值	表示是否应使用 \log_n 转换。
formula_log_n_offset	数字	
formula_log_10	布尔值	表示是否应使用 \log_{10} 转换。
formula_log_10_offset	数字	
formula_exponential	布尔值	表示是否应使用指数 (e^x) 转换。
formula_square_root	布尔值	表示是否应使用平方根转换。

表 197. transform 节点属性 (续).

transform 节点属性	数据类型	属性描述
use_output_name	布尔值	指定是否使用定制的输出名。
output_name	字符串	如果 use_output_name 为 true , 那么指定使用的名称。
output_mode	Screen File	用于指定输出节点中生成的输出的目标位置。
output_format	HTML (.html) Output (.cou)	用于指定输出类型。
paginate_output	布尔值	当 output_format 是 HTML 时, 使输出分页。
lines_per_page	数字	与 paginate_output 一起使用时, 指定每个输出页中的行数。
full_filename	字符串	表示要在文件输出中使用的文件名。

第 17 章 导出节点属性

公共导出节点属性

以下属性通用于所有导出节点。

表 198. 公共导出节点属性.

属性	值	属性描述
publish_path	字符串	输入要用于所发布的图像和参数文件的 rootname 名称。
publish_metadata	布尔值	指定是否生成元数据文件（用于描述图像的输入和输出以及它们的数据模型）。
publish_use_parameters	布尔值	指定是否在 *.par 文件中包含流参数。
publish_parameters	字符串列表	指定要包括的参数。
execute_mode	export_data publish	指定是执行节点而不发布流，还是在执行节点时自动发布流。

asexport 节点属性

通过 Analytic Server 导出，可以在分布式文件系统 (HDFS) 上运行流。

表 199. asexport 节点属性.

asexport 节点属性	数据类型	属性描述
data_source	string	数据源的名称。
export_mode	string	指定是将导出的数据追加到现有数据源，还是覆盖现有数据源。
host	string	Analytic Server 主机的名称。
端口	integer	Analytic Server 进行侦听的端口。
tenant	string	在多租户环境中，这是您所属的租户的名称。在单租户环境中，此属性的缺省值为 ibm 。
set_credentials	布尔值	如果 Analytic Server 上的用户认证与 SPSS Modeler 服务器相同，请将此属性设置为 false 。否则，将其设置为 true 。
user_name	string	用于登录到 Analytic Server 的用户名。仅当 set_credentials 为 true 时才需要。
password	string	用于登录到 Analytic Server 的密码。仅当 set_credentials 为 true 时才需要。

cognosexport 节点属性



IBM Cognos BI 导出节点以 Cognos BI 数据库所能够读取的格式导出数据。

注意: 对于此节点, 必须定义 Cognos 连接和 ODBC 连接。

Cognos 连接

Cognos 连接的属性如下。

表 200. cognosexport 节点属性.

cognosexport 节点属性	数据类型	属性描述
cognos_connection	<code>{"field","field", ... ,"field"}</code>	这是包含 Cognos 服务器连接详细信息的列表属性。格式为: <code>{"Cognos_server_URL", login_mode, "namespace", "username", "password"}</code> 其中: <i>Cognos_server_URL</i> 您要导出数据的目标 Cognos 服务器的 URL <i>login_mode</i> 指示是否使用匿名登录, 其值为 true 或 false; 如果设置为 true, 那么应将下列字段设置为 "" <i>namespace</i> 指定用于登录服务器的安全认证提供程序 <i>username</i> 和 <i>password</i> 为用于登录 Cognos 服务器的用户名和密码
cognos_package_name	字符串	您要将数据导出到的 Cognos 数据包的路径和名称, 例如: /Public Folders/MyPackage
cognos_datasource	字符串	
cognos_export_mode	发布 ExportFile	
cognos_filename	字符串	

ODBC 连接

除了 datasource 属性无效之外, ODBC 连接的属性与下一节中针对 databaseexport 列出的属性完全相同。

databaseexport 节点属性



“数据库导出”节点将数据写入符合 ODBC 标准的关系数据源。要写入 ODBC 数据源, 该数据源必须存在, 且您对该数据源必须具有写许可权。

表 201. databaseexport 节点属性.

databaseexport 节点属性	数据类型	属性描述
datasource	字符串	
username	字符串	
password	字符串	
epassword	字符串	在执行期间, 此槽为只读。要生成经过编码的密码, 请使用“工具”菜单中的“密码工具”。有关更多信息, 请参阅第 41 页的『生成加密密码』主题。
table_name	字符串	
write_mode	Create Append Merge	
map	字符串	用于将流字段名称映射到数据库列名称 (仅在 write_mode 为 Merge 的情况下才有效)。对于合并, 所有字段必须经过映射才能导出。数据库中不存在的字段名称将添加为新列。
key_fields	[field field ... field]	指定用作键的流字段; map 属性指示了此字段在数据库中的对应项。
join	Database Add	
drop_existing_table	布尔值	
delete_existing_rows	布尔值	
default_string_size	整数	
type		用于设置模式类型的结构属性。
generate_import	布尔值	
use_custom_create_table_command	布尔值	使用 custom_create_table 通道修改标准 CREATE TABLE SQL 命令。
custom_create_table_command	字符串	指定字符串命令代替标准 CREATE TABLE SQL 命令使用。
use_batch	布尔值	下列属性是用于数据库批量装入的高级选项。Use_batch 为 true 时将关闭向数据库逐行提交的功能。
batch_size	数字	指定在提交到内存前发送到数据库的记录数。
bulk_loading	Off ODBC External	指定批量装入类型。下面列出了 ODBC 和 External 的其他选项。
not_logged	布尔值	
odbc_binding	Row Column	指定通过 ODBC 批量装入时使用逐行绑定或逐列绑定。

表 201. databaseexport 节点属性 (续).

databaseexport 节点属性	数据类型	属性描述
loader_delimit_mode	Tab Space Other	对于通过外部程序的批量装入, 指定定界符的类型。选择 Other 连同 loader_other_delimiter 属性以指定定界符, 例如逗号 (,)。
loader_other_delimiter	字符串	
specify_data_file	布尔值	标志为 true 时可激活下面的 data_file 属性, 在该属性中可以指定批量装入到数据库时写入的文件名和路径。
data_file	字符串	
specify_loader_program	布尔值	标志为 true 时可激活下面的 loader_program 属性, 在该属性中可以指定外部装入程序脚本或程序的名称和位置。
loader_program	字符串	
gen_logfile	布尔值	标志为 true 时可激活下面的 logfile_name, 在该属性中可以指定服务器上的文件的名称以生成错误日志。
logfile_name	字符串	
check_table_size	布尔值	标志为 true 时允许进行表检查以确保数据库表大小的增加与从 IBM SPSS Modeler 导出的行数相符。
loader_options	字符串	指定装入程序的其他参数, 例如 -comment 和 -specialdir。
export_db_primarykey	布尔值	指定给定字段是否为主键。
use_custom_create_index_command	布尔值	如果标志为 true, 那么为所有索引启用定制 SQL。
custom_create_index_command	字符串	指定启用定制 SQL 后用于创建索引的 SQL 命令。(创建特定索引时, 该值可被覆盖, 如下所示。)
indexes.INDEXNAME.fields		必要时创建指定的索引, 并列上要包括在该索引中的字段名。
indexes.INDEXNAME.use_custom_create_index_command	布尔值	用于启用或禁用特定索引的定制 SQL。
indexes.INDEXNAME.custom_create_command		指定用于指定索引的定制 SQL。
indexes.INDEXNAME.remove	布尔值	如果标志为 true, 那么从索引集中删除指定的索引。
table_space	字符串	指定将创建表空间。
use_partition	布尔值	指定将使用分布散列字段。
partition_field	字符串	指定分布散列字段的内容。

注意: 对于某些数据库, 您可以指定以导出时进行压缩的方式来创建数据库表 (例如, SQL 中的等效语句 CREATE TABLE MYTABLE (...) COMPRESS YES;)。为了支持此功能, 提供了属性 use_compression 和 compression_mode, 如下所示。

表 202. 使用压缩功能的 *databaseexport* 节点属性.

databaseexport 节点属性	数据类型	属性描述
use_compression	布尔值	如果设置为 true, 那么将以导出时进行压缩的方式创建表。
compression_mode	Row Page	设置 SQL Server 数据库的压缩级别。
	Default Direct_Load_Operations All_Operations Basic OLTP Query_High Query_Low Archive_High Archive_Low	设置 Oracle 数据库的压缩级别。请注意, 值 OLTP、Query_High、Query_Low、Archive_High 和 Archive_Low 至少需要 Oracle 11gR2。

datacollectionexport 节点属性



IBM SPSS Data Collection 导出节点以 IBM SPSS Data Collection 市场调查软件使用的格式输出数据。必须安装 IBM SPSS Data Collection 数据库才可使用此节点。

表 203. *datacollectionexport* 节点属性.

datacollectionexport 节点属性	数据类型	属性描述
metadata_file	字符串	要导出的元数据文件的名称。
merge_metadata	Overwrite MergeCurrent	
enable_system_variables	布尔值	指定导出的 .mdd 文件是否应包括 IBM SPSS Data Collection 系统变量。
casedata_file	字符串	要导出观测数据的 .sav 文件的名称。
generate_import	布尔值	

excelexport 节点属性



Excel 导出节点以 Microsoft Excel 格式 (.xls) 输出数据。您还可选择在执行完此节点后自动启动 Excel 并打开导出的文件。

表 204. *excelexport* 节点属性.

excelexport 节点属性	数据类型	属性描述
full_filename	字符串	

表 204. *excelexport* 节点属性 (续).

excelexport 节点属性	数据类型	属性描述
excel_file_type	Excel2003 Excel2007	
export_mode	Create Append	
inc_field_names	布尔值	指定字段名是否可以包含在工作表的第一行中。
start_cell	字符串	指定导出的开始单元格。
worksheet_name	字符串	要写入的工作表的名称。
launch_application	布尔值	指定是否应该对生成的文件调用 Excel。请注意，必须在“帮助应用程序”对话框（“工具”菜单 >“帮助应用程序”）中指定用于启动 Excel 的路径。
generate_import	布尔值	指定是否应生成用于读取所导出数据文件的“Excel 导入”节点。

outfile 节点属性



“平面文件导出”节点将数据输出到定界文本文件中。这对于导出可以由其他分析或电子表格软件读取的数据非常有用。

表 205. *outfile* 节点属性.

outfile 节点属性	数据类型	属性描述
full_filename	字符串	输出文件的名称。
write_mode	Overwrite Append	
inc_field_names	布尔值	
use_newline_after_records	布尔值	
delimit_mode	Comma Tab Space Other	
other_delimiter	字符	
quote_mode	None Single Double Other	
other_quote	布尔值	
generate_import	布尔值	

表 205. *outputfile* 节点属性 (续).

outputfile 节点属性	数据类型	属性描述
encoding	StreamDefault SystemDefault "UTF-8"	

sasexport 节点属性



“SAS 导出”节点以 SAS 格式输出数据，以便将该数据读入 SAS 或者与 SAS 兼容的软件包。共有三种可用的 SAS 文件格式：SAS for Windows/OS2、SAS for UNIX 和 SAS V7/V8。

表 206. *sasexport* 节点属性.

sasexport 节点属性	数据类型	属性描述
format	Windows UNIX SAS7 SAS8	可变属性标签字段。
full_filename	字符串	
export_names	NamesAndLabels NamesAsLabels	用于将字段名从 IBM SPSS Modeler 的导出中映射到 IBM SPSS Statistics 或 SAS 的变量名中。
generate_import	布尔值	

statisticsexport 节点属性



“Statistics 导出”节点以 IBM SPSS Statistics *.sav* 格式输出数据。*.sav* 文件可由 IBM SPSS Statistics Base 和其他产品读取。这种格式也用于 IBM SPSS Modeler 中的某些缓存文件。

有关此节点属性的信息，请参阅第 216 页的『*statisticsexport* 节点属性』。

xmlexport 节点属性



“XML 导出”节点将数据以 XML 格式输出到文件。您可以选择性地创建“XML 源”节点，以便将导出的数据重新读取到流中。

表 207. *xmlexport* 节点属性.

xmlexport 节点属性	数据类型	属性描述
full_filename	字符串	(必需) XML 导出文件的完整路径和文件名。

表 207. *xmlexport* 节点属性 (续).

xmlexport 节点属性	数据类型	属性描述
use_xml_schema	布尔值	指定是否使用 XML 模式 (XSD 或 DTD 文件) 控制导出数据的结构。
full_schema_filename	字符串	要使用的 XSD 或 DTD 文件的完整路径和文件名。如果 use_xml_schema 设为 true, 那么为必需。
generate_import	布尔值	生成用于将导出的数据文件读回到流中的“XML 源”节点。
records	字符串	表示记录边界的 XPath 表达式。
map	字符串	将字段名映射到 XML 结构。

第 18 章 IBM SPSS Statistics 节点属性

statisticsimport 节点属性



“Statistics 文件”节点从 IBM SPSS Statistics 使用的 .sav 文件格式以及保存在 IBM SPSS Modeler 中的高速缓存文件（这些文件使用同一格式）中读取数据。

表 208. statisticsimport 节点属性.

statisticsimport 节点属性	数据类型	属性描述
full_filename	字符串	完整文件名（包括路径）。
import_names	NamesAndLabels LabelsAsNames	处理变量名和标签的方法。
import_data	DataAndLabels LabelsAsData	处理值和标签的方法。
use_field_format_for_storage	布尔值	指定导入时是否使用 IBM SPSS Statistics 字段格式信息。

statisticstransform 节点属性



“Statistics 转换”节点针对 IBM SPSS Modeler 中的数据源运行选择的 IBM SPSS Statistics 语法命令。此节点需要 IBM SPSS Statistics 的许可副本。

表 209. statisticstransform 节点属性.

statisticstransform 节点属性	数据类型	属性描述
syntax	字符串	
check_before_saving	布尔值	保存输入项之前验证已输入的语法。如果语法无效，那么会显示一条错误消息。
default_include	布尔值	有关更多信息，请参阅第 90 页的『filter 节点属性』主题。
include	布尔值	有关更多信息，请参阅第 90 页的『filter 节点属性』主题。
new_name	字符串	有关更多信息，请参阅第 90 页的『filter 节点属性』主题。

statisticsmodel 节点属性



“Statistics 模型”节点使您能够通过运行将会生成 PMML 的 IBM SPSS Statistics 过程来分析和处理数据。此节点需要 IBM SPSS Statistics 的许可副本。

statisticsmodel 节点属性	数据类型	属性描述
syntax	字符串	
default_include	布尔值	有关更多信息，请参阅第 90 页的『filter 节点属性』主题。
include	布尔值	有关更多信息，请参阅第 90 页的『filter 节点属性』主题。
new_name	字符串	有关更多信息，请参阅第 90 页的『filter 节点属性』主题。

statisticsoutput 节点属性



“Statistics 输出”节点用于调用 IBM SPSS Statistics 过程，以分析 IBM SPSS Modeler 数据。可以访问许多不同的 IBM SPSS Statistics 分析过程。此节点需要 IBM SPSS Statistics 的许可副本。

表 210. statisticsoutput 节点属性.

statisticsoutput 节点属性	数据类型	属性描述
mode	Dialog Syntax	选择“IBM SPSS Statistics 对话框”选项或语法编辑器
syntax	字符串	
use_output_name	布尔值	
output_name	字符串	
output_mode	Screen File	
full_filename	字符串	
file_type	HTML SPV SPW	

statisticsexport 节点属性



“Statistics 导出”节点以 IBM SPSS Statistics .sav 格式输出数据。.sav 文件可由 IBM SPSS Statistics Base 和其他产品读取。这种格式也用于 IBM SPSS Modeler 中的某些缓存文件。

表 211. *statisticsexport* 节点属性.

statisticsexport 节点属性	数据类型	属性描述
full_filename	字符串	
launch_application	布尔值	
export_names	NamesAndLabels NamesAsLabels	用于将字段名从 IBM SPSS Modeler 的导出中映射到 IBM SPSS Statistics 或 SAS 的变量名中。
generate_import	布尔值	

第 19 章 超节点属性

下列各表描述特定于超节点的属性。注意公共节点属性也可应用于超节点。

表 212. 终端超节点属性.

属性名称	属性类型/值列表	属性描述
execute_method	Script 正态 (Normal)	
script	string	
script_language	PythonLegacy	为超节点脚本设置脚本语言。

超节点参数

您可以使用修改流参数时所用的函数来使用脚本创建或设置超节点参数。请参阅主题第 35 页的『流、会话和超节点参数』，了解更多信息。

设置已封装节点的属性

要设置有关超节点内的节点的属性，您必须访问该超节点所拥有的图，然后使用各种 find 方法（例如 findByName() 和 findById()）来找到这些节点。例如，在包含单个“类型”节点的超节点脚本中：

```
supernode = modeler.script.supernode()
diagram = supernode.getCompositeProcessorDiagram()
# Find the type node within the supernode internal diagram
typenode = diagram.findByName("type", None)
typenode.setKeyedProperty("direction", "Drug", "Input")
typenode.setKeyedProperty("direction", "Age", "Target")
```

超节点脚本的局限性。超节点无法处理其他流，也无法更改当前流。

附录 A. 节点名引用

此部分提供 IBM SPSS Modeler 中节点的脚本编制名称引用。

模型块名称

模型块（也称为生成的模型）可以按类型进行引用，就好像节点和输出对象一样。下表列出模型对象的引用名称。

请注意，这些名称专用于引用模型选用板（位于 IBM SPSS Modeler 窗口的右上角）中的模型块。要引用已经添加到流中进行评分的模型节点，那么使用另外一套以 `apply...` 为前缀的名称。有关更多信息，请参阅第 161 页的第 14 章，『模型块节点属性』主题。

注：在通常情况下，建议按名称和类型来引用模型，以避免引起混淆。

表 213. 模型块名称（建模选用板）。

模型名称	模型
anomalydetection	异常
apriori	Apriori
autoclassifier	自动分类器
autocluster	自动聚类
autonumeric	自动数字
bayesnet	贝叶斯网络
c50	C5.0
carma	Carma
cart	C&R 树
chaid	CHAID
coxreg	Cox 回归
decisionlist	决策列表
discriminant	判别
factor	PCA/因子
featureselection	特征选择
genlin	广义线性回归
glmm	GLMM
kmeans	K-Means
knn	k-最近相邻元素
kohonen	Kohonen
linear	线性
logreg	Logistic 回归
neuralnetwork	类神经网络
quest	QUEST
regression	线性回归

表 213. 模型块名称（建模选用板）（续）.

模型名称	模型
sequence	序列
slrm	自学响应模型
statisticsmodel	IBM SPSS Statistics 模型
SVM	支持向量机
timeseries	时间序列
twostep	二阶

表 214. 模型块名称（数据库建模选用板）.

模型名称	模型
db2imcluster	IBM ISW 聚类
db2imlog	IBM ISW Logistic 回归
db2imnb	IBM ISW 朴素贝叶斯
db2imreg	IBM ISW 回归
db2imtree	IBM ISW 决策树
msassoc	MS 关联规则
msbayes	MS 朴素贝叶斯
mscluster	MS 聚类
mslogistic	MS Logistic 回归
msneuralnetwork	MS 神经网络
msregression	MS 线性回归
mssequencecluster	MS 序列聚类
mstimeseries	MS 时间序列
mstree	MS 决策树
netezzabayes	Netezza Bayes 网络
netezzadectree	Netezza 决策树
netezzadivcluster	Netezza 分裂式聚类
netezzaglm	Netezza 广义线性
netezzakmeans	Netezza K-Means
netezzaknn	Netezza KNN
netezzalinegression	Netezza 线性回归
netezzanaiivebayes	Netezza 朴素贝叶斯
netezzapca	Netezza PCA
netezzaregtree	Netezza 回归树
netezzatimeseries	Netezza 时间序列
oraabn	Oracle Adaptive Bayes
oraai	Oracle AI
oradecisiontree	Oracle 决策树
oraglm	Oracle GLM
orakmeans	Oracle <i>k</i> -Means
oranb	Oracle 朴素贝叶斯

表 214. 模型块名称 (数据库建模选用板) (续).

模型名称	模型
oranmf	Oracle NMF
oraocluster	Oracle O-Cluster
orasvm	Oracle SVM

避免重复的模型名称

使用脚本对生成的模型进行操作时, 务必注意: 允许重复的模型名称可能会导致歧义引用。为了避免这种情况的发生, 最好在编制脚本时要求对生成的模型使用唯一的名称。

要为重复模型名称设置选项:

1. 从菜单中选择:

工具 > 用户选项

2. 单击通知选项卡。
3. 选择替换原有模型以限制生成的模型的重复命名。

存在不明确的模型引用时, 脚本执行行为在 SPSS Modeler 与 IBM SPSS Collaboration and Deployment Services 之间可能有所不同。SPSS Modeler 客户机提供了“替换先前模型”选项, 此选项将自动替换同名的模型 (例如, 脚本通过循环执行迭代, 以便每次都生成不同的模型)。但是, 在 IBM SPSS Collaboration and Deployment Services 中运行同一脚本时, 此选项不可用。通过将每次迭代中生成的模型重命名以避免对模型进行不明确的引用, 或者通过在循环结束前清除当前模型 (例如, 添加 `clear generated palette` 语句), 可以避免这种情况。

输出类型名称

下表列出了所有的输出对象类型和创建它们的节点。有关每种输出类型可用的导出格式完整列表, 请参阅创建该输出类型的节点的属性描述 (第 105 页的『图形节点公共属性』和第 195 页的第 16 章, 『输出节点属性』)。

表 215. 输出对象类型以及创建这些类型的节点.

输出对象类型	节点
analysisoutput	分析
collectionoutput	集合
dataauditoutput	数据审核
distributionoutput	分布
evaluationoutput	评估
histogramoutput	直方图
matrixoutput	矩阵
meansoutput	平均值
multiplotoutput	多重散点图
plotoutput	散点图
qualityoutput	质量
reportdocumentoutput	此对象类型不属于节点; 它是工程报告创建的输出
reportoutput	报告

表 215. 输出对象类型以及创建这些类型的节点 (续).

输出对象类型	节点
statisticsprocedureoutput	Statistics 输出
statisticsoutput	统计
tableoutput	表
timeplotoutput	时间散点图
weboutput	Web

附录 B. 从旧脚本编制迁移到 Python 脚本编制

旧脚本迁移概述

本节提供 IBM SPSS Modeler 中 Python 脚本编制与旧脚本编制之间的差异摘要，并提供有关如何将旧脚本迁移为 Python 脚本的信息。在本节中，您将找到标准 SPSS Modeler 旧命令和等效的 Python 命令的列表。

一般差异

旧脚本编制的设计在很大程度上借鉴了操作系统命令脚本。尽管包含一些块结构（例如 `if...then...else...endif` 和 `for...endfor`），但旧脚本编制面向行，并且缩进通常没有意义。

在 Python 脚本编制中，缩进有意义，并且属于同一逻辑块的行必须在同一级别进行缩进。

注：复制和粘贴 Python 代码时，请务必小心操作。在编辑器中，使用 `tab` 缩进的行可能与使用空格缩进的行看起来一样。但是，Python 脚本将生成错误，这是因为未将这些行视作缩进相同的行。

脚本编制上下文

脚本编制上下文定义了将在其中执行脚本的环境，例如，用于执行脚本的流或超节点。例如，在旧脚本编制中，上下文是隐式的，这意味着假定流脚本中的所有节点引用都包含在执行该脚本的流中。

在 Python 脚本编制中，脚本编制上下文通过 `modeler.script` 模块以显式方式提供。例如，Python 流脚本可以使用以下代码访问执行该脚本的流：

```
s = modeler.script.stream()
```

然后，可以通过返回的对象来调用与流相关的函数。

命令与函数

旧脚本编制面向命令。这意味着脚本的每一行通常以后跟参数进行运行的命令开始，例如：

```
connect 'Type':typenode to :filternode  
rename :derivencode as "Compute Total"
```

Python 使用通常通过定义函数的对象（模块、类或对象）所调用的函数，例如：

```
stream = modeler.script.stream()  
typenode = stream.findByName("type", "Type")  
filternode = stream.findByName("filter", None)  
stream.link(typenode, filternode)  
derive.setLabel("Compute Total")
```

文字和注释

IBM SPSS Modeler 中一些常用的文字和注释命令在 Python 脚本编制中具有等效命令。这可以帮助您将现有 SPSS Modeler 旧脚本转换为 Python 脚本，以便在 IBM SPSS Modeler 16 中使用。

表 216. 文字和注释的旧脚本编制到 Python 脚本编制的映射。

旧脚本编制	Python 脚本编制
整数, 例如 4	相同
浮点数, 例如 0.003	相同
加单引号的字符串, 例如 'Hello'	相同 注: 包含非 ASCII 字符的字符串面值必须以 u 作为前缀, 以确保它们表示为 Unicode。
加双引号的字符串, 例如 "Hello again"	相同 注: 包含非 ASCII 字符的字符串面值必须以 u 作为前缀, 以确保它们表示为 Unicode。
长字符串, 例如 """This is a string that spans multiple lines"""	相同
列表, 例如 [1 2 3]	[1, 2, 3]
变量引用, 例如 set x = 3	x = 3
行继续符 (\), 例如 set x = [1 2 \ 3 4]	x = [1, 2,\n3, 4]
块注释, 例如 /* This is a long comment over a line.*/	/* This is a long comment over a line.*/
行注释, 例如 set x = 3 # make x 3	x = 3 # make x 3
undef	无
true	True
false	False

运算符

IBM SPSS Modeler 中一些常用的运算符命令在 Python 脚本编制中具有等效命令。这可以帮助您将现有 SPSS Modeler 旧脚本转换为 Python 脚本，以便在 IBM SPSS Modeler 16 中使用。

表 217. 运算符的旧脚本编制到 Python 脚本编制的映射。

旧脚本编制	Python 脚本编制
NUM1 + NUM2 LIST + ITEM LIST1 + LIST2	NUM1 + NUM2 LIST.append(ITEM) LIST1.extend(LIST2)
NUM1 - NUM2 LIST - ITEM	NUM1 - NUM2 LIST.remove(ITEM)
NUM1 * NUM2	NUM1 * NUM2
NUM1 / NUM2	NUM1 / NUM2
= ==	==

表 217. 运算符的旧脚本编制到 Python 脚本编制的映射 (续).

旧脚本编制	Python 脚本编制
/= /==	!=
X ** Y	X ** Y
X < Y X <= Y X > Y X >= Y	X < Y X <= Y X > Y X >= Y
X div Y X rem Y X mod Y	X // Y X % Y X % Y
and or not(EXPR)	and or not EXPR

条件语句和循环

IBM SPSS Modeler 中一些常用的条件和循环命令在 Python 脚本编制中具有等效命令。这可以帮助您将现有 SPSS Modeler 旧脚本转换为 Python 脚本，以便在 IBM SPSS Modeler 16 中使用。

表 218. 条件语句和循环的旧脚本编制到 Python 脚本编制的映射.

旧脚本编制	Python 脚本编制
for VAR from INT1 to INT2 ... endfor	for VAR in range(INT1, INT2): ... 或 VAR = INT1 while VAR <= INT2: ... VAR += 1
for VAR in LIST ... endfor	for VAR in LIST: ...
for VAR in_fields_to NODE ... endfor	for VAR in NODE.getInputDataModel(): ...
for VAR in_fields_at NODE ... endfor	for VAR in NODE.getOutputDataModel(): ...
if...then ... elseif...then ... else ... endif	if ...: ... elif ...: ... else: ...
with TYPE OBJECT ... endwith	无等效项
var VAR1	不需要变量声明

变量

在旧脚本编制中，引用变量之前对变量进行了声明，例如：

```
var mynode
set mynode = create typenode at 96 96
```

在 Python 脚本编制中，变量在首次引用时进行创建，例如：

```
mynode = stream.createAt("type", "Type", 96, 96)
```

在旧脚本编制中，必须使用 ^ 运算符显式除去对变量的引用，例如：

```
var mynode
set mynode = create typenode at 96 96
set ^mynode.direction."Age" = Input
```

与大大对数脚本编制语言一样，在 Python 脚本编制中，这不是必需操作，例如：

```
mynode = stream.createAt("type", "Type", 96, 96)
mynode.setKeyedPropertyValue("direction", "Age", "Input")
```

节点、输出和模型类型

在旧脚本编制中，各种对象类型（节点、输出和模型）通常向对象类型追加了类型。例如，“派生”节点具有 `derivemode` 类型：

```
set feature_name_node = create derivemode at 96 96
```

Python 中的 IBM SPSS Modeler API 未包含 `node` 后缀，因此“派生”节点具有 `derive` 类型，例如：

```
feature_name_node = stream.createAt("derive", "Feature", 96, 96)
```

旧脚本编制与 Python 脚本编制的类型名称中的唯一差异在于缺少类型后缀。

属性名

在旧脚本编制和 Python 脚本编制中，属性名相同。例如，在这两种脚本编制环境中，变量文件节点中用于定义文件位置的属性为 `full_filename`。

节点引用

许多旧脚本使用隐式搜索来查找和访问要修改的节点。例如，下列命令用于在当前流中搜索带有“类型”标签的“类型”节点，然后将“年龄”字段的方向（或建模角色）设置为输入，并将“药品”字段设置为目标（也就是要预测的值）：

```
set 'Type':typenode.direction."Age" = Input
set 'Type':typenode.direction."Drug" = Target
```

在 Python 脚本编制中，必须先显式查找节点对象，然后再调用用于设置属性值的函数，例如：

```
typenode = stream.findByType("type", "Type")
typenode.setKeyedPropertyValue("direction", "Age", "Input")
typenode.setKeyedPropertyValue("direction", "Drug", "Target")
```

注：在本例中，“Target”必须包含在字符串引号中。

另外，Python 脚本可使用 `modeler.api` 软件包中的 `ModelingRole` 枚举。

虽然 Python 脚本编制版本可能更为繁琐，但它能够实现更佳运行时性能，这是因为通常仅执行一次搜索节点。在旧脚本编制示例中，将针对各个命令搜索节点。

另外，还支持按标识查找节点（可以在节点对话框的“注释”选项卡中查看节点标识）。例如，在旧脚本编制中：

```
# id65EMPB9VL87 is the ID of a Type node
set @id65EMPB9VL87.direction."Age" = Input
```

以下脚本显示 Python 脚本编制中的同一示例：

```
typenode = stream.findByID("id65EMPB9VL87")
typenode.setKeyedPropertyValue("direction", "Age", "Input")
```

获取并设置属性

旧脚本编制使用 `set` 命令来指定值。`set` 命令后跟的词汇可以是属性定义。以下脚本显示了两种可能的用于设置属性的脚本格式：

```
set <node reference>.<property> = <value>
set <node reference>.<keyed-property>.<key> = <value>
```

在 Python 脚本编制中，通过使用函数 `setProperty()` 和 `setKeyedPropertyValue()`，可以实现同一结果，例如：

```
object.setProperty(property, value)
object.setKeyedPropertyValue(keyed-property, key, value)
```

在旧脚本编制中，可以使用 `get` 命令来实现访问属性值，例如：

```
var n v
set n = get node :filternode
set v = ^n.name
```

在 Python 脚本编制中，通过使用函数 `getPropertyValue()`，可以实现同一结果，例如：

```
n = stream.findByName("filter", None)
v = n.getPropertyValue("name")
```

编辑流

在旧脚本编制中，`create` 命令用于创建新节点，例如：

```
var agg select
set agg = create aggregatenode at 96 96
set select = create selectnode at 164 96
```

在 Python 脚本编制中，流具有多种创建节点的方法，例如：

```
stream = modeler.script.stream()
agg = stream.createAt("aggregate", "Aggregate", 96, 96)
select = stream.createAt("select", "Select", 164, 96)
```

在旧脚本编制中，`connect` 命令用于创建节点之间的链接，例如：

```
connect ^agg to ^select
```

在 Python 脚本编制中，`link` 方法用于创建节点之间的链接，例如：

```
stream.link(agg, select)
```

在旧脚本编制中，`disconnect` 命令用于除去节点之间的链接，例如：

```
disconnect ^agg from ^select
```

在 Python 脚本编制中，`unlink` 方法用于除去节点之间的链接，例如：

```
stream.unlink(agg, select)
```

在旧脚本编制中，`position` 命令用于将节点放在流画布上或其他节点之间，例如：

```
position ^agg at 256 256  
position ^agg between ^myselect and ^mydistinct
```

在 Python 脚本编制中，通过使用两种不同的方法（`setXYPosition` 和 `setPositionBetween`），可以实现同一结果。例如：

```
agg.setXYPosition(256, 256)  
agg.setPositionBetween(myselect, mydistinct)
```

节点操作

IBM SPSS Modeler 中一些常用的节点操作命令在 Python 脚本编制中具有等效命令。这可以帮助您将现有 SPSS Modeler 旧脚本转换为 Python 脚本，以便在 IBM SPSS Modeler 16 中使用。

表 219. 节点操作的旧脚本编制到 Python 脚本编制的映射。

旧脚本编制	Python 脚本编制
<code>create nodespec at x y</code>	<code>stream.create(type, name)</code> <code>stream.createAt(type, name, x, y)</code> <code>stream.createBetween(type, name, preNode, postNode)</code> <code>stream.createModelApplier(model, name)</code>
<code>connect fromNode to toNode</code>	<code>stream.link(fromNode, toNode)</code>
<code>delete node</code>	<code>stream.delete(node)</code>
<code>disable node</code>	<code>stream.setEnabled(node, False)</code>
<code>enable node</code>	<code>stream.setEnabled(node, True)</code>
<code>disconnect fromNode from toNode</code>	<code>stream.unlink(fromNode, toNode)</code> <code>stream.disconnect(node)</code>
<code>duplicate node</code>	<code>node.duplicate()</code>
<code>execute node</code>	<code>stream.runSelected(nodes, results)</code> <code>stream.runAll(results)</code>
<code>flush node</code>	<code>node.flushCache()</code>
<code>position node at x y</code>	<code>node.setXYPosition(x, y)</code>
<code>position node between node1 and node2</code>	<code>node.setPositionBetween(node1, node2)</code>
<code>rename node as name</code>	<code>node.setLabel(name)</code>

循环

在旧脚本编制中，主要支持下列两种循环选项：

- 计数循环，在此循环中，下标变量在两个整数范围之间进行移动。
- 序列循环，此循环对值序列进行遍历，以便将当前值绑定到循环变量。

以下脚本是旧脚本编制中的计数循环示例：

```
for i from 1 to 10  
  println ^i  
endfor
```

以下脚本是旧脚本编制中的序列循环示例:

```
var items
set items = [a b c d]

for i in items
  println ^i
endfor
```

另外, 还可以使用其他类型的循环:

- 对模型选用板中的模型或输出选用板中的输出执行迭代。
- 对传入或传出节点的字段执行迭代。

Python 脚本编制还支持其他类型的循环。以下脚本是 Python 脚本编制中的计数循环示例:

```
i = 1
while i <= 10:
  print i
  i += 1
```

以下脚本是 Python 脚本编制中的序列循环示例:

```
items = ["a", "b", "c", "d"]
for i in items:
  print i
```

序列循环非常灵活, 并且在与 IBM SPSS Modeler API 方法结合后, 此循环可支持多个脚本编制用例。以下示例显示如何使用 Python 脚本编制中的序列循环对传出节点的字段执行迭代:

```
node = modeler.script.stream().findByType("filter", None)
for column in node.getOutputDataModel().columnIterator():
  print column.getColumnname()
```

执行流

执行流的过程中, 生成的模型或输出对象将添加到其中一个对象管理器中。在旧脚本编制中, 脚本必须在对象管理器中找到构建对象, 或者从生成输出的节点访问最新生成的输出。

在 Python 中, 执行流的过程有所不同: 执行生成的任何模型或输出将以传递到执行函数的列表形式返回。这使得可以更加轻松地访问流执行结果。

旧脚本编制支持下列三种流执行命令:

- `execute_all`, 用于执行流中的所有可执行终端节点。
- `execute_script`, 用于执行流脚本 (与脚本执行的设置无关)。
- `execute node`, 用于执行指定的节点。

Python 脚本编制支持一组类似的函数:

- `stream.runAll(results-list)`, 用于执行流中的所有可执行终端节点。
- `stream.runScript(results-list)`, 用于执行流脚本 (与脚本执行的设置无关)。
- `stream.runSelected(node-array, results-list)`, 用于按节点的提供顺序执行指定的一组节点。
- `node.run(results-list)`, 用于执行指定的节点。

在旧脚本中, 可以使用带有可选整数代码的 `exit` 命令来终止流执行, 例如:

```
exit 1
```

在 Python 脚本编制中，使用以下脚本可实现同一结果：

```
modeler.script.exit(1)
```

通过文件系统和存储库访问对象

在旧脚本编制中，您可以使用 `open` 命令打开现有流、模型或输出对象，例如：

```
var sset s = open stream "c:/my streams/modeling.str"
```

在 Python 脚本编制中，存在一个可从会话进行访问的 `TaskRunner` 类，并且这个类可用于执行类似的任务，例如：

```
taskrunner = modeler.script.session().getTaskRunner()
s = taskrunner.openStreamFromFile("c:/my streams/modeling.str", True)
```

在旧脚本编制中，要保存对象，您可以使用 `save` 命令，例如：

```
save stream s as "c:/my streams/new_modeling.str"
```

等效的 Python 脚本方法将使用 `TaskRunner` 类，例如：

```
taskrunner.saveStreamToFile(s, "c:/my streams/new_modeling.str")
```

基于 IBM SPSS Collaboration and Deployment Services Repository 的操作在旧脚本编制中通过 `retrieve` 和 `store` 命令受支持，例如：

```
var sset s = retrieve stream "/my repository folder/my_stream.str"
store stream ^s as "/my repository folder/my_stream_copy.str"
```

在 Python 脚本编制中，可以通过与会话关联的存储库对象来访问等效功能，例如：

```
session = modeler.script.session()
repo = session.getRepository()
s = repo.retrieveStream("/my repository folder/my_stream.str", None, None, True)
repo.storeStream(s, "/my repository folder/my_stream_copy.str", None)
```

注：存储库访问要求使用有效存储库连接对会话进行了配置。

流操作

IBM SPSS Modeler 中一些常用的流操作命令在 Python 脚本编制中具有等效命令。这可以帮助您将现有 SPSS Modeler 旧脚本转换为 Python 脚本，以便在 IBM SPSS Modeler 16 中使用。

表 220. 流操作的旧脚本编制到 Python 脚本编制的映射。

旧脚本编制	Python 脚本编制
create stream <i>DEFAULT_FILENAME</i>	<code>taskrunner.createStream(name, autoConnect, autoManage)</code>
close stream	<code>stream.close()</code>
clear stream	<code>stream.clear()</code>
get stream <i>stream</i>	无等效项
load stream <i>path</i>	无等效项
open stream <i>path</i>	<code>taskrunner.openStreamFromFile(path, autoManage)</code>
save <i>stream</i> as <i>path</i>	<code>taskrunner.saveStreamToFile(stream, path)</code>
retrive stream <i>path</i>	<code>repository.retriveStream(path, version, label, autoManage)</code>
store <i>stream</i> as <i>path</i>	<code>repository.storeStream(stream, path, label)</code>

模型操作

IBM SPSS Modeler 中一些常用的模型操作命令在 Python 脚本编制中具有等效命令。这可以帮助您将现有 SPSS Modeler 旧脚本转换为 Python 脚本，以便在 IBM SPSS Modeler 16 中使用。

表 221. 模型操作的旧脚本编制到 Python 脚本编制的映射。

旧脚本编制	Python 脚本编制
<code>open model path</code>	<code>taskrunner.openModelFromFile(path, autoManage)</code>
<code>save model as path</code>	<code>taskrunner.saveModelToFile(model, path)</code>
<code>retrieve model path</code>	<code>repository.retrieveModel(path, version, label, autoManage)</code>
<code>store model as path</code>	<code>repository.storeModel(model, path, label)</code>

文档输出操作

IBM SPSS Modeler 中一些常用的文档输出操作命令在 Python 脚本编制中具有等效命令。这可以帮助您将现有 SPSS Modeler 旧脚本转换为 Python 脚本，以便在 IBM SPSS Modeler 16 中使用。

表 222. 文档输出操作的旧脚本编制到 Python 脚本编制的映射。

旧脚本编制	Python 脚本编制
<code>open output path</code>	<code>taskrunner.openDocumentFromFile(path, autoManage)</code>
<code>save output as path</code>	<code>taskrunner.saveDocumentToFile(output, path)</code>
<code>retrieve output path</code>	<code>repository.retrieveDocument(path, version, label, autoManage)</code>
<code>store output as path</code>	<code>repository.storeDocument(output, path, label)</code>

旧脚本编制与 Python 脚本编制之间的其他差异

旧脚本提供对处理 IBM SPSS Modeler 工程的支持。Python 脚本编制当前不提供此支持。

旧脚本编制提供了某种装入状态对象（流和模型的组合）的支持。IBM SPSS Modeler 8.0 后的版本不推荐使用状态对象。Python 脚本编制不支持状态对象。

Python 脚本编制提供了下列附加功能，旧脚本编制中未提供这些功能：

- 类和函数定义
- 错误处理
- 更复杂的输入/输出支持
- 外部模块和第三方模块

声明

这些信息开发用于在全球提供的产品和服务。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您所在区域当前可获得的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务的操作，由用户自行负责。

IBM 可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并不意味着授予用户使用这些专利的任何许可。您可以用书面形式将许可查询寄往：

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

以下段落对于英国和与当地法律有不同规定的其他国家或地区均不适用：INTERNATIONAL BUSINESS MACHINES CORPORATION“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。

本信息可能包含技术方面不够准确的地方或印刷错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可以随时对本出版物中描述的产品和/或程序进行改进和/或更改，而不另行通知。

在本信息材料中对任何非 IBM 网站的引用仅为了方便用户，并不以任何方式表明对这些网站的认可。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：(i) 使其能够在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及 (ii) 使其能够对已经交换的信息进行相互使用，请与下列地址联系：

IBM 软件部ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

本文档中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际程序许可协议或任何同等协议中的条款提供。

此处所含的性能数据均在受控环境下决定。因此，在其他操作环境中获得的结果可能差异较大。有些测量可能在开发级的系统中进行，不保证这些测量结果与常用系统上的测量结果相同。另外，有些测量结果可能通过推断来估计得出。实际结果可能有所差异。此文档的用户应针对其具体环境验证适用的数据。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

有关 IBM 未来方向或意向的所有声明均可能未经通知即变更或撤销，并且仅代表目标和目的。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

如果您正在查阅此信息的软拷贝，照片和彩色插图可能不会显示。

商标

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corp.，在全球许多管辖区域的商标或注册商标。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，可从 Web 站点 www.ibm.com/legal/copytrade.shtml 上『版权和商标信息』部分获取。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国和@3B72其他国家或地区的注册商标。

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

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

Java 和所有基于 Java 的商标及徽标皆为 Oracle 和/或其附属公司的商标或注册商标。

其他产品和服务名称可能是 IBM 或其他公司的商标。

索引

[A]

安全性
 加密密码 41, 45

[B]

贝叶斯网络模型
 节点脚本编制属性 123, 162
备注 14
遍历节点 28
变量
 脚本编制 11
变量文件节点
 属性 67
标识 15
标志
 命令行自变量 43
 组合多个标志 47

[C]

参数 3, 49, 51
 超节点 219
 脚本编制 11
槽参数 3, 49, 50
查找和替换 8
查找节点 25
超节点 49
 参数 219
 脚本 1, 3, 23
 脚本编写 219
 流 23
 设置属性 219
 属性 219
重排节点
 属性 93
传递参数 15
创建节点 26, 27, 28
创建类 19
错误检查
 脚本编制 41

[D]

代码块 15
导出节点
 节点脚本编制属性 207
迭代变量
 脚本中的循环 6

迭代关键字
 脚本中的循环 5
定向 Web 节点
 属性 114
定义方法 20
定义类 19
定义属性 20
独立脚本 1, 3, 23

[E]

二阶模型
 节点脚本编制属性 159, 170

[F]

非 ASCII 字符 18
分区节点
 属性 91
服务器
 命令行自变量 45

[G]

广义线性模型
 节点脚本编制属性 136, 165

[H]

函数
 对象引用 226
 节点操作 230
 流操作 232
 模型操作 233
 条件语句 227
 文档输出操作 233
 文字 226
 循环 227
 运算符 226
 注释 226

[J]

继承 20
加密密码
 添加至脚本 41
建模节点
 节点脚本编制属性 117
脚本
 保存 1

脚本 (续)

 从文本文件导入 1
 迭代变量 6
 迭代关键字 5
 条件执行 3, 7
 选择字段 6
 循环 3, 4

脚本编写

 超节点流 23
 迭代变量 6
 迭代关键字 5
 旧脚本编制 226, 227, 230, 232, 233
 可见循环 3, 4
 流 23
 条件执行 3, 7
 图 23
 文件路径 42
 选择字段 6
 语法 18
 Python 脚本编制 226, 227, 230, 232, 233

脚本编制

 超节点脚本 1, 23
 从命令行 42
 错误检查 41
 独立脚本 1, 23
 概述 1, 11
 公共属性 50
 建模节点执行 41
 旧脚本编制 226
 流 1, 23
 流执行顺序 41
 模型替换 41
 上下文 24
 输出节点 195
 所用缩写 49
 图形节点 105
 用户界面 1, 3, 8
 与早期版本的兼容性 42
 语法 11, 12, 13, 14, 15, 16, 19, 20
 在超节点中 3
 执行 7
 中断 7
 Python 脚本编制 226
脚本编制 API
 超节点参数 35
 处理错误 35
 独立脚本 39
 多个流 39
 访问已生成的对象 33
 会话参数 35

- 脚本编制 API (续)
 - 简介 31
 - 流参数 35
 - 全局值 39
 - 示例 31
 - 搜索 31
 - 元数据 31
- 节点
 - 导入 28
 - 链接节点 27
 - 名称引用 221
 - 取消链接节点 27
 - 删除 28
 - 替换 28
 - 信息 29
- 节点脚本编制属性 171
 - 导出节点 207
 - 建模节点 117
 - 模型块 161
- 决策列表模型
 - 节点脚本编制属性 131, 164

[K]

- 块
 - 节点脚本编制属性 161

[L]

- 类型节点
 - 属性 100
- 列表 12
- 流
 - 多重集合命令 49
 - 脚本编写 1, 23
 - 脚本编制 1, 23
 - 属性 51
 - 条件执行 3, 7
 - 修改 26
 - 循环 3, 4
 - 执行 23
- 流的条件执行 3, 7
- 流执行顺序
 - 用脚本更改 41
- 流中的循环 3, 4

[M]

- 密码
 - 添加至脚本 41
 - 已编码 45
- 面向对象 19
- 命令行
 - 参数 45
 - 参数列表 44, 45, 46

- 命令行 (续)
 - 多个参数 47
 - 脚本编制 42
 - 运行 IBM SPSS Modeler 43
- 模型对象
 - 脚本编写名称 221, 223
- 模型块
 - 脚本编写名称 221, 223
 - 节点脚本编制属性 161

[N]

- 匿名化节点
 - 属性 83

[P]

- 派生节点
 - 属性 88
- 判别模型
 - 节点脚本编制属性 132, 165
- 评估节点
 - 属性 107
- 平面文件节点
 - 属性 212

[Q]

- 迁移
 - 编辑流 229
 - 变量 228
 - 存储库 232
 - 访问对象 232
 - 概述 225
 - 函数 225
 - 获取属性 229
 - 脚本编制上下文 225
 - 节点类型 228
 - 节点引用 228
 - 命令 225
 - 模型类型 228
 - 其他 233
 - 设置属性 229
 - 输出类型 228
 - 属性名 228
 - 文件系统 232
 - 循环 230
 - 一般差异 225
 - 执行流 231

[S]

- 散点图节点
 - 属性 112
- 设置属性 25

- 神经网络
 - 节点脚本编制属性 150, 168
- 神经网络模型
 - 节点脚本编制属性 149, 167
- 生成的关键字 42
- 时间序列模型
 - 节点脚本编制属性 157, 170
- 示例 16
- 输出对象
 - 脚本编写名称 223
- 输出节点
 - 脚本属性 195
- 数据库导出节点
 - 属性 208
- 数据库建模 171
- 数据库节点
 - 属性 58
- 属性
 - 超节点 219
 - 脚本编制 49, 50, 117, 161, 207
 - 流 51
 - 数据库建模节点 171
 - 通用脚本编写 50
- 数学方法 16

[T]

- 特征选择模型
 - 节点脚本编制属性 134, 165
- 添加属性 19
- 图 23
- 图形节点
 - 脚本属性 105

[X]

- 系统
 - 命令行自变量 44
- 线性回归模型
 - 节点脚本编制属性 153, 168
- 线性模型
 - 节点脚本编制属性 145, 167
- 修改流 26, 28
- 序列模型
 - 节点脚本编制属性 154, 169

[Y]

- 已生成的模型
 - 脚本编写名称 221, 223
- 异常检测模型
 - 节点脚本编制属性 117, 161
- 隐藏变量 20
- 引用节点 24
- 查找节点 25

引用节点 (续)
 设置属性 25
用户输入节点
 属性 66
语句 14
源节点
 属性 55
运算 12

[Z]

正则表达式 8
支持向量机模型
 节点脚本编制属性 156, 169
执行脚本 7
执行流 23
执行顺序
 用脚本更改 41
中断脚本 7
自变量
 服务器连接 45
 命令文件 47
 系统 44
 IBM SPSS Collaboration and
 Deployment Services Repository 连接
 46
自动分类器节点
 节点脚本编制属性 119
自动分类器模型
 节点脚本编制属性 162
自动聚类模型
 节点脚本编制属性 162
自动数据准备
 属性 83
自动数字模型
 节点脚本编制属性 122, 162
字段
 在脚本中关闭 105
字符串 13
自学响应模型
 节点脚本编制属性 155, 169
最近相邻元素模型
 节点脚本编制属性 143

A

aggregate 节点
 属性 71
aggregate 节点属性 71
analysis 节点
 属性 195
analysis 节点属性 195
Analytic Server 源节点
 属性 56
anomalydetection 节点属性 117

anonymize 节点属性 83
append 节点属性 71
applyanomalydetection 节点属性 161
applyapriori 节点属性 161
applyautoclassifier 节点属性 162
applyautocluster 节点属性 162
applyautonumeric 节点属性 162
applybayesnet 节点属性 162
applyc50 节点属性 163
applycarma 节点属性 163
applycart 节点属性 163
applychaid 节点属性 164
applycoxreg 节点属性 164
applydb2imcluster 节点属性 185
applydb2imlog 节点属性 185
applydb2imnb 节点属性 185
applydb2imreg 节点属性 185
applydb2imtree 节点属性 185
applydecisionlist 节点属性 164
applydiscriminant 节点属性 165
applyfactor 节点属性 165
applyfeatureselection 节点属性 165
applygeneralizedlinear 节点属性 165
applyglmm 节点属性 166
applykmeans 节点属性 166
applyknn 节点属性 166
applykohonen 节点属性 166
applylinear 节点属性 167
applylogreg 节点属性 167
applymlogistic 节点属性 173
applymsneuralnetwork 节点属性 173
applymsregression 节点属性 173
applymssequencecluster 节点属性 173
applymstimeseries 节点属性 173
applymstree 节点属性 173
applynetezababes 节点属性 194
applynetezadectree 节点属性 194
applynetezadivcluster 节点属性 194
applynetezakmeans 节点属性 194
applynetezaknn 节点属性 194
applynetezalinereregression 节点属性 194
applynetezanaivebayes 节点属性 194
applynetezapca 节点属性 194
applynetezaregtree 节点属性 194
applyneuralnet 节点属性 167
applyneuralnetwork 节点属性 168
applyoraabn 节点属性 179
applyoradecisiontree 节点属性 179
applyorakmeans 节点属性 179
applyoranb 节点属性 179
applyoranmf 节点属性 179
applyoracluster 节点属性 179
applyorasvm 节点属性 179
applyquest 节点属性 168
applyr 属性 168
applyregression 节点属性 168

applyselflearning 节点属性 169
applysequence 节点属性 169
applysvm 节点属性 169
applytimeseries 节点属性 170
applytwestep 节点属性 170
apriori 节点属性 118
Apriori 模型
 节点脚本编制属性 118, 161
asexport 节点属性 207
asimport 节点属性 56
autoclassifier 节点属性 119
autocluster 节点属性 121
autodataprep 节点属性 83
autonumeric 节点属性 122

B

balance 节点
 属性 72
balance 节点属性 72
bayesnet 节点属性 123
binning 节点
 属性 86
binning 节点属性 86
buildr 属性 124

C

c50 节点属性 125
C5.0 模型
 节点脚本编制属性 125, 163
carma 节点属性 126
CARMA 模型
 节点脚本编制属性 126, 163
cart 节点属性 126
chaid 节点属性 128
CHAID 模型
 节点脚本编制属性 128, 164
clear generated palette 命令 42
cognosimport 节点属性 57
collection 节点
 属性 105
collection 节点属性 105
Cox 回归模型
 节点脚本编制属性 129, 164
coxreg 节点属性 129
C&R 树模型
 节点脚本编制属性 126, 163

D

dataaudit 节点属性 196
database 节点属性 58
databaseexport 节点属性 208
datacollectionexport 节点属性 211

datacollectionimport 节点属性 59
db2imassoc 节点属性 180
db2imcluster 节点属性 180
db2imlog 节点属性 180
db2imnb 节点属性 180
db2imreg 节点属性 180
db2imsequence 节点属性 180
db2imtimeseries 节点属性 180
db2imtree 节点属性 180
decisionlist 节点属性 131
derive 节点属性 88
derive_stb 节点属性 72
directedweb 节点属性 114
discriminant 节点属性 132
distinct 节点属性 74
distribution 节点属性 106

E

ensemble 节点属性 89
Enterprise View 节点属性 61
evaluation 节点属性 107
evimport 节点属性 61
Excel 导出节点属性 211
excelexport 节点属性 211
excelimport 节点属性 60

F

factor 节点属性 133
featureselection 节点属性 134
filler 节点属性 90
filter 节点属性 90
fixedfile 节点属性 61
flatfilenode 属性 212

G

genlin 节点属性 136
glm 节点属性 139
GLMM 模型
节点脚本编制属性 139, 166
graphboard 节点属性 108

H

histogram 节点属性 110
history 节点属性 91

I

IBM Cognos BI 源节点属性 57
IBM DB2 模型
节点脚本编制属性 180
IBM ISW 关联模型
节点脚本编制属性 180, 185
IBM ISW 回归模型
节点脚本编制属性 180, 185
IBM ISW 聚类模型
节点脚本编制属性 180, 185
IBM ISW 决策树模型
节点脚本编制属性 180, 185
IBM ISW 朴素贝叶斯模型
节点脚本编制属性 180, 185
IBM ISW 时间序列模型
节点脚本编制属性 180
IBM ISW 序列模型
节点脚本编制属性 180, 185
IBM ISW Logistic 回归模型
节点脚本编制属性 180, 185
IBM SPSS Collaboration and Deployment Services Repository
命令行自变量 46
IBM SPSS Data Collection 导出节点属性 211
IBM SPSS Data Collection 源节点属性 59
IBM SPSS Modeler
从命令行运行 43
IBM SPSS Statistics 模型
节点脚本编制属性 216
IBM SPSS Statistics 源节点属性 215

J

Jython 11

K

kmeans 节点属性 142
knn 节点属性 143
KNN 模型
节点脚本编制属性 166
kohonen 节点属性 144
Kohonen 模型
节点脚本编制属性 166
kohonen 模型
节点脚本编制属性 144
K-Means 模型
节点脚本编制属性 142, 166

L

linear 节点属性 145
Logistic 回归模型
节点脚本编制属性 146, 167
logreg 节点属性 146

M

matrix 节点属性 197
means 节点属性 198
merge 节点属性 74
Microsoft 模型
节点脚本编制属性 171, 173
models
脚本编写名称 221, 223
MS 决策树
节点脚本编制属性 171, 173
MS 神经网络
节点脚本编制属性 171, 173
MS 时间序列
节点脚本编制属性 173
MS 线性回归
节点脚本编制属性 171, 173
MS 序列聚类
节点脚本编制属性 173
MS Logistic 回归
节点脚本编制属性 171, 173
msassoc 节点属性 171
msbayes 节点属性 171
mscluster 节点属性 171
mslogistic 节点属性 171
msneuralnetwork 节点属性 171
msregression 节点属性 171
mssequencecluster 节点属性 171
mstimeseries 节点属性 171
mstree 节点属性 171
multiplot 节点属性 111

N

Netezza 贝叶斯网络模型
节点脚本编制属性 186, 194
Netezza 分裂式聚类模型
节点脚本编制属性 186, 194
Netezza 广义线性模型
节点脚本编制属性 186
Netezza 回归树模型
节点脚本编制属性 186, 194
Netezza 决策树模型
节点脚本编制属性 186, 194
Netezza 模型
节点脚本编制属性 186
Netezza 朴素贝叶斯模型
节点脚本编制属性 186, 194

Netezza 时间序列模型
 节点脚本编制属性 186
Netezza 线性回归模型
 节点脚本编制属性 186, 194
Netezza KNN 模型
 节点脚本编制属性 186, 194
Netezza K-Means 模型
 节点脚本编制属性 186, 194
Netezza PCA 模型
 节点脚本编制属性 186, 194
netezababes 节点属性 186
netezadectree 节点属性 186
netezadivcluster 节点属性 186
netezzaglm 节点属性 186
netezzakmeans 节点属性 186
netezzaknn 节点属性 186
netezzalinerregression 节点属性 186
netezzanaivebayes 节点属性 186
netezzapca 节点属性 186
netezzaregtree 节点属性 186
netezzatimeseries 节点属性 186
neuralnet 节点属性 149
neuralnetwork 节点属性 150
numericpredictor 节点属性 122

O

oraabn 节点属性 174
oraai 节点属性 174
oraapriori 节点属性 174
Oracle 广义线性模型
 节点脚本编制属性 174
Oracle 决策树模型
 节点脚本编制属性 174, 179
Oracle 模型
 节点脚本编制属性 174
Oracle 朴素贝叶斯模型
 节点脚本编制属性 174, 179
Oracle 支持向量机模型
 节点脚本编制属性 174, 179
Oracle 自适应贝叶斯模型
 节点脚本编制属性 174, 179
Oracle AI 模型
 节点脚本编制属性 174
Oracle Apriori 模型
 节点脚本编制属性 174, 179
Oracle KMeans 模型
 节点脚本编制属性 174, 179
Oracle MDL 模型
 节点脚本编制属性 174, 179
Oracle NMF 模型
 节点脚本编制属性 174, 179
Oracle O-Cluster
 节点脚本编制属性 174, 179
oradecisiontree 节点属性 174
oraglm 节点属性 174

orakmeans 节点属性 174
oramdl 节点属性 174
oranb 节点属性 174
oranmf 节点属性 174
oraoccluster 节点属性 174
orasvm 节点属性 174
outputfile 节点属性 212

P

partition 节点属性 91
PCA 模型
 节点脚本编制属性 133, 165
PCA/因子模型
 节点脚本编制属性 133, 165
plot 节点属性 112
Python 11
 脚本编制 11

Q

quest 节点属性 152
QUEST 模型
 节点脚本编制属性 152, 168

R

R 输出节点
 属性 200
reclassify 节点属性 92
regression 节点属性 153
reorder 节点属性 93
report 节点属性 199
restructure 节点属性 93
RFM 分析节点
 属性 93
RFM 汇总节点
 属性 75
rfmaggregate 节点属性 75
rfmanalysis 节点属性 93
Routput 节点属性 200
Rprocessnode 节点属性 76

S

sample 节点属性 76
sasexport 节点属性 213
sasimport 节点属性 64
select 节点属性 78
sequence 节点属性 154
setglobals 节点属性 200
settoflag 节点属性 94
simeval 节点属性 201
simfit 节点属性 201
simgen 节点属性 64

slrm 节点属性 155
SLRM 模型
 节点脚本编制属性 155, 169
sort 节点属性 78
statistics 节点属性 202
statisticsexport 节点属性 216
statisticsimport 节点属性 215
statisticsmodel 节点属性 216
statisticsoutput 节点属性 216
statisticstransform 节点属性 215
streamingts 节点属性 79
svm 节点属性 156
SVM 模型
 节点脚本编制属性 156

T

table 节点属性 203
timeintervals 节点属性 95
timeplot 节点属性 113
timeseries 节点属性 157
transform 节点属性 205
transpose 节点属性 99
twostep 节点属性 159
type 节点属性 100

U

userinput 节点属性 66

V

variablefile 节点属性 67

W

Web 节点
 属性 114
web 节点属性 114

X

xmlexport 节点属性 213
xmlimport 节点属性 70

[特别字符]

“报告”节点
 属性 199
“表”节点
 属性 203
“重构”节点
 属性 93

“重新分类”节点
 属性 92
 “多重散点图”节点
 属性 111
 “分布”节点
 属性 106
 “固定文件”节点
 属性 61
 “过滤”节点
 属性 90
 “合并”节点
 属性 74
 “矩阵”节点
 属性 197
 “均值”节点
 属性 198
 “空间时间限制”节点
 属性 72
 “历史记录”节点
 属性 91
 “流式时间序列”节点
 属性 79
 “模拟拟合”节点
 属性 201
 “模拟评估”节点
 属性 201
 “模拟生成”节点
 属性 64
 “排序”节点
 属性 78
 “区分”节点
 属性 74
 “设为标志”节点
 属性 94
 “设置全局值”节点
 属性 200
 “时间区间”节点
 属性 95
 “时间散点图”节点
 属性 113
 “数据审核”节点
 属性 196
 “填充器”节点
 属性 90
 “统计”节点
 属性 202
 “图形板”节点
 属性 108
 “选择”节点
 属性 78
 “样本”节点
 属性 76
 “整体”节点
 属性 89
 “直方图”节点
 属性 110
 “转换”节点
 属性 205
 “转置”节点
 属性 99
 “追加”节点
 属性 71
 “自动聚类”节点
 节点脚本编制属性 121
 “字段重新排序器”节点
 属性 93
 “Excel 源”节点
 属性 60
 “IBM SPSS Statistics 导出”节点
 属性 216
 “IBM SPSS Statistics 输出”节点
 属性 216
 “IBM SPSS Statistics 转换”节点
 属性 215
 “R 构建”节点
 节点脚本编制属性 124
 “R 进程”节点
 属性 76
 “SAS 导出”节点
 属性 213
 “SAS 源”节点
 属性 64
 “XML 导出”节点
 属性 213
 “XML 源”节点
 属性 70



Printed in China