SPSS使用笔记

连续字段（例如年龄字段）包含连续的数字值。

无需分类变量（例如信用评价字段）有两个或多个不同值，例如不良、优良或无信用历史。

有序分类变量（例如收入水平字段）用于描述具有顺序固定的不同值的数据，在本例中为低、中和高。

注意名义字段和有序字段的区别

记录选项主要针对记录进行行上的处理，字段选项主要对记录进行列上的处理

SPSS几个过程功能点如下所示：

• **源。**此类节点可将数据引入 SPSS Modeler。

• **记录选项。**此类节点可对数据记录执行操作，如选择、合并和追加等。

• **字段选项。**此类节点可对数据字段执行操作，如过滤、导出新字段和确定给定字段的测量级别等。

• **图形。**此类节点可在建模前后以图表形式显示数据。图形包括散点图、直方图、网络节点和评估图表。

• **建模。**此类节点可使用 SPSS Modeler 中提供的建模算法，如神经网络、决策树、聚类算法和数据排序等。

• **数据库建模。** 节点使用 Microsoft SQL Server、IBM DB2 和 Oracle 数据库中可用的建模算法。

• **输出。**节点生成可在 SPSS Modeler 中查看的数据、图表和模型等多种输出结果。

• **导出。**节点生成可在外部应用程序（如 IBM® SPSS® Data Collection 或 Excel）中查看的多种输出。

• **SPSS Statistics。** 节点将数据导入 IBM® SPSS® Statistics 或从中导出数据，以及运行 SPSS Statistics 过程。

对于节点较多的数据流可以封装在一起成为一个超节点，右键可以进去超节点进行编辑。

可以在节点上右键，选择缓冲启动，启动缓冲后会在节点的右上角出现一个标记，以后再执行该数据流时，将从节点之后执行，而且缓冲也可以进行保存与下载，需要注意的是当缓存缓存的数据发生变化时，必须右键刷新缓存，数据才会发生变更。

同时spss支持不同的数据流添加注释，方便以后查看

灵活的参数配置，spss中提供了三种参数：

会话参数：工具🡪设置会话参数

数据流参数：文件🡪流属性

超节点参数：双击超节点🡪定义参数

**绕过节点**：可以按住alt键，双击节点，即可绕过该节点。

**禁用节点**：可以鼠标右键选择禁用节点，该节点不绕过流程，只不过不起到任何的处理。

**SPSS中的运行:**

可以在终端节点上邮件选择运行，也可以在非终端节点上右键选择从此处运行，注意两者区别。

SPSS中几个节点过程：

* **源节点：**

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/pevnodeicon.jpg | 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。请参阅 主题 [Enterprise View 节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/pevnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/databasenodeicon.jpg | 数据库节点可用于使用 ODBC（开放数据库连接）从多种其他数据包中导入数据，这些数据包包括 Microsoft SQL Server、DB2、Oracle 等。 请参阅 主题 [数据库源节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/database_overview.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/varfile_node_icon.jpg | 自由格式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。请参阅 主题 [变量文件节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/variablefilenode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/fixedfilenodeicon.jpg | 固定文件节点会从固定字段文本文件（即文件字段不定界而是从相同的位置开始且长度固定）中导入数据。机器生成的数据或遗存数据通常以固定字段格式存储。 请参阅 主题 [固定文件节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/fixedfilenode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/spssfilenodeicon.jpg | Statistics 文件节点从 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。请参阅 主题 [Statistics 文件节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/spssfile_data.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/dimensionsimportnodeicon.jpg | IBM® SPSS® Data Collection 节点从符合 Data Collection 数据模型的市场调查软件所用的各种格式中导入调查数据。必须安装 Data Collection 数据库才可使用此节点。 请参阅 主题 [Data Collection节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/dimensionsdataimportnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/cognos_sourceicon.gif | IBM Cognos BI 源节点从 Cognos BI 数据库导入数据。 请参阅 主题 [IBM Cognos BI 源节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/cognos_source_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/sasfilenodeicon.jpg | SAS 文件节点可将 SAS 数据导入到 SPSS Modeler 中。请参阅 主题 [SAS 源节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/sasimportnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/excelimportnodeicon.jpg | Excel 节点可以从任何版本的 Microsoft Excel 中导入数据。不要求指定 ODBC 数据源。请参阅 主题 [Excel 源节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/excelimportnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/xml_sourcenodeicon.gif | XML 源节点将 XML 格式的数据导入到流中。可以导入某个目录中的单个文件或所有文件。还可选择指定架构文件，以从中读取 XML 结构。 请参阅 主题 [XML 源节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/xmlsourcenode_general.htm) 详细信息。 |

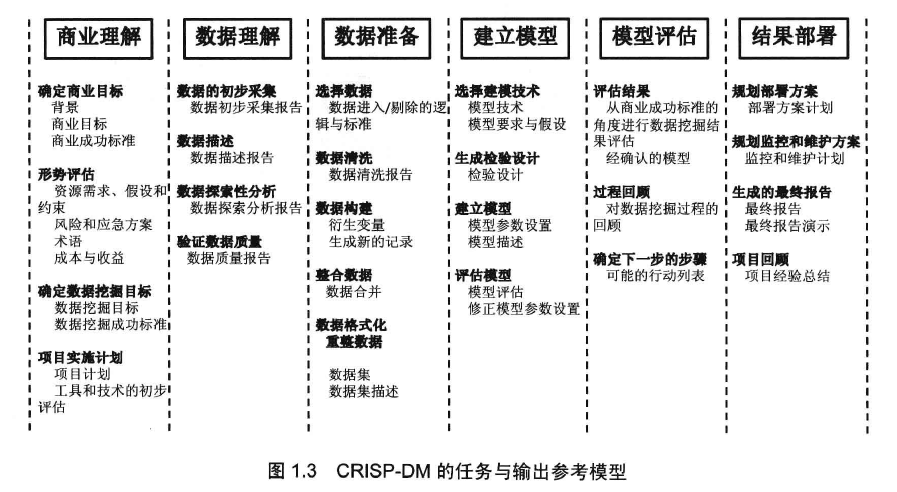
|  |  |
| --- | --- |
| http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/images/userinputnodeicon.jpg | 用户输入节点提供了一种用于创建综合数据的简单方式 — 可以从头开始创建也可以通过更改现有数据进行创建。此节点非常有用，例如，在希望为建模创建测试数据集时，即可使用此节点。请参阅 主题 [用户输入节点](http://127.0.0.1:61726/help/topic/com.ibm.spss.modeler.help/user_general.htm) 详细信息。 |

其中变量文件节点应该可以用于读库csv格式的文件。

Excel读取是需要注意的是：**与 Excel 不同，SPSS Modeler 不允许在字段中有混合的存储类型。要避免此问题，可以在 Excel 电子表格中手动将单元格格式设置为文本，这样将按字符串读取所有的值（包括数字）。**

Modeler中使用$P-pname对参数进行引用，其中pname是参数名，在表达式中，参数必须使用单引号括起来，例如‘$P-pname’

Spss经典的CRISP-DM数据挖掘6个过程



**集中趋势的描述指标：**

**算术均数**：用于描述数据分布集中趋势的统计指标，但它不适应于严重偏态分布的变量进行描述，只有单峰和基本对称的分布资料，使用均数作为集中趋势描述的统计量才是合理的。

**中位数**：中位数适用于任意分布类型的资料，但由于中位数只考虑居中位置，对信息的利用不充分，所以当体谅较小时数值不太稳定，因此对于对称的分布资料，分析者会优先考虑实用均数，仅仅在均数不能使用的情况下才会使用中位数加以描述。

**几何均数**：用G表示，适用于原始数据分布部队称，但经过数据转换后呈对称分布的数据，几何均数实际上就是对数据转换后的数据的logX的算术均数的反对数。

**截尾均数：**由于均数容易受到极端值的影响，因此可以考虑去掉两端的数据，然后再计算均数，常用的截尾均数有5%截尾均数，即两端各去掉5%的数据。

还有其它的指标众数（样本数据中出现频次最大的数字），调和数（观察值X倒数之均数的倒数）日常用的比较少。

**离散趋势的描述指标：**

**全距：**又称为极差，即一组数据中最大值和最小值只差，过于简单，一般只用于预备性检查。

**方差和标准差：**使用最广泛，但是容易受到极值的影响，当数据有较为明显的极端值时不宜使用，实际上方差和标准差试用范围

**百分位数、四分位数与四分位间距：**四分位间距既排除了两侧极值的影响，又能够反应较多数据的离散程度，他是当方差、标准差不适用时较好的离散程度描述指标。

**变异系数：**当需要比较两组数据的离散程度时，如果两组数据的测量尺度相差较大，或者数据的量纲不一致，此时需要使用变异系数，他是标准差与其平均数的比，用CV表示，显然CV没有量纲，这样比较客观。

**分布特征描述：**

正态分布使用偏度和封度来描述分布特征。

**各分类节点的使用说明：**

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/binaryclassifiernodeicon.jpg | “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。 请参阅 主题 [自动分类器节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/binary_classifier_node.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/rangepredictornodeicon.jpg | 自动数值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 Support Vector Machine (SVM)。可基于相关度、相对错误或已用变量数对模型进行对比。 请参阅 主题 [自动数值节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/numeric_predictor_node.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/cartnodeicon.jpg | 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洁度，使用递归分区来将训练记录分割为组。如果节点中 100% 的观测值都属于目标字段的一个特定类别，则树中的该节点将被认定为“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）。 请参阅 主题 [C&R 树节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/cartnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/questnodeicon.jpg | 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。 请参阅 主题 [QUEST 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/questnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/chaidnodeicon.jpg | CHAID 使用卡方统计量来生成决策树，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。 请参阅 主题 [CHAID 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/chaidnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/c50nodeicon.jpg | 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。 请参阅 主题 [C5.0 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/c50node_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/decisionlistnodeicon.jpg | 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。 请参阅 主题 [决策表](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/decisionlistnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/alm_nodeicon.gif | 线性模型根据目标与一个或多个预测变量间的线性关系来预测连续目标。 请参阅 主题 [线性模型](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/idh_alm.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/regressionnodeicon.jpg | 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。 请参阅 主题 [回归节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/regressionnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/pcafactornodeicon.jpg | 因子/主成分分析节点提供了用于降低数据复杂程度的强大数据缩减技术。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合中的方差，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。这两种方式的目标都是找到有效概括原始字段集中的信息的一小部分导出字段。 请参阅 主题 [主成份分析/因子节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/factornode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/featureselectionnodeicon.jpg | “特征选择”节点会根据某组条件（例如缺失值百分比）筛选可删除的输入字段；对于保留的输入，将相对于指定目标对其重要性进行排序。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？ 请参阅 主题 [特征选择节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/featureselectionnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/discriminantnodeicon.jpg | 判别式分析所做的假设比 logistic 回归的假设更严格，但在符合这些假设时，判别式分析可以作为 logistic 回归分析的有用替代项或补充。 请参阅 主题 [判别式节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/discriminant_node_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/logisticnodeicon.jpg | 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。 请参阅 主题 [逻辑节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/nomregnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/genlinnodeicon.jpg | “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。另外，该模型允许因变量呈非正态分布。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。 请参阅 主题 [GenLin 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/genlin_node_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/cox_reg_node_icon.jpg | 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。 请参阅 主题 [Cox 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/coxreg_node_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/svm_icon.jpg | 使用 Support Vector Machine (SVM) 节点，可以将数据分为两组，而无需过度拟合。SVM 可以与大量数据集配合使用，如那些含有大量输入字段的数据集。 请参阅 主题 [SVM 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/svmnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/bayesian_network_icon.jpg | 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) 和马尔可夫毯网络，这些算法主要用于分类问题。 请参阅 主题 [贝叶斯网络节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/bayesian_networks_node_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/selflearn_icon.jpg | 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。 请参阅 主题 [SLRM 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/selflearnnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/timeseriesnodeicon.jpg | 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 (ARIMA) 模型和多变量 ARIMA（即变换函数）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。 请参阅 主题 [时间序列建模节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/timeseriesnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/knn_nodeicon.jpg | 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。 请参阅 主题 [KNN 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/knn_node_general.htm) 详细信息。 |

关联模型查找您数据中的模式，其中一个或多个实体（如事件、购买或属性）与一个或多个其他实体相关联。这些模型构建定义这些关系的规则集。数据中的字段可以作为输入和目标。您可以手动查找这些关联，但关联规则算法可以更快速地完成，并能探索更多复杂的模式。Apriori 和 Carma 模型是使用此类算法的示例。另一种类型的关联模型是序列检测模型，后者可以在按时间建立结构的数据中查找顺序模式。

关联模型在预测多个结果时非常有用，例如，购买了产品 X 的顾客也购买了产品 Y 和 Z。关联模型可以将特定结论（如购买某些产品的决策）与一组条件关联起来。关联规则算法相对于更标准的决策树算法（C5.0 和 C&RT）的优势在于，它可以找到任何属性间存在的关联。决策树算法只使用单一结论来构建规则，而关联算法则试图找到更多规则，且每个规则具有不同的结论。

**关联模型：**﻿

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/apriorinodeicon.jpg | Apriori 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 训练的速度通常快。它对于可以包含的规则数也没有任何限制，可以处理最多带有 32 个预条件的规则。Apriori 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。 请参阅 主题 [Apriori 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/apriorinode_general.htm) 详细信息。 |

﻿

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/carmanodeicon.jpg | 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不同的是：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，例如用于找到后项为想在节日期间促销的商品的产品或服务（前项）的列表。 请参阅 主题 [CARMA 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/carmanode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/sequencenodeicon.jpg | 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，该算法使用有效的两步法来发现序列。 请参阅 主题 [序列节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/sequencenode_general.htm) 详细信息。 |

**细分模型(也就是聚类模型)**

﻿细分模型将数据划分为具有类似输入字段模式的记录段或聚类。细分模型只对输入字段感兴趣，没有输出或目标字段的概念。细分模型的示例为 Kohonen 网络、K-Means 聚类、两步聚类和异常检测等。

在不知道特定结果的情况下（例如，需要识别新犯罪模式或在客户群中识别利益群体时），建议使用细分模型（也称为“聚类模型”）。聚类模型主要用来确定相似记录的组并根据它们所属的组来为记录添加标签。此方法的优点在于，不用提前了解这些组及其特征就可以使用，它使聚类模型（其中没有需要模型预测的预定义输出或目标字段）区别于其他的建模技术。对于这些模型来说，没有正确或错误的结果之分。模型的值由模型捕获数据中感兴趣的分组并提供这些分组的有用说明信息的能力来确定。聚类模型通常用于创建在后续分析中用作输入的聚类或段（例如，将潜在用户分成几个相似的子组）。

**细分节点（也就是聚类模型）：**

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/autoclusternodeicon.jpg | 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。 请参阅 主题 [自动聚类节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/auto_cluster_node_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/kmeansnodeicon.jpg | 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。 请参阅 主题 [K-Means 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/trainkmeansnode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/kohonennodeicon.jpg | 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，有差异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。 请参阅 主题 [Kohonen 节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/kohonennode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/twostepnodeicon.jpg | TwoStep 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。TwoStep 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。 请参阅 主题 [两步聚类节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/clusternode_general.htm) 详细信息。 |

|  |  |
| --- | --- |
| http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/images/anomalydetectionnodeicon.jpg | “异常检测”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。 请参阅 主题 [异常检测节点](http://127.0.0.1:51638/help/topic/com.ibm.spss.modeler.help/anomalydetectionnode_general.htm) 详细信息。 |