**数据挖掘模型训练，预测过程使用说明**

作者：张青松

目录

[1. 模型训练 2](#_Toc423340980)

[1.1. 输入数据 2](#_Toc423340981)

[1.1.1. 文件读取数据 2](#_Toc423340982)

[1.1.2. 数据库读取数据 2](#_Toc423340983)

[1.2. 训练模型 3](#_Toc423340984)

[2. 模型预测 3](#_Toc423340985)

[2.1. 输入预测数据 3](#_Toc423340986)

[2.2. 使用预测模型进行预测 3](#_Toc423340987)

[2.3. 查看训练结果 4](#_Toc423340988)

**摘要**

一个数据挖掘完整的工作流包含训练和预测两个过程。

1. 模型训练过程包含：输入训练数据、使用训练模型训练，得到预测模型。
2. 预测过程包含：输入预测数据、使用训练得到的预测模型进行预测。

# 模型训练

## 输入数据

在DataStudio中，输入数据的方式分为两种：

### 文件读取数据

使用文件读取节点输入数据，在节点选择中的数据源页签中，选择文件节点。



根据文件读取节点使用说明设置，并读取数据：

<http://172.16.51.106/pages/viewpage.action?pageId=919692>

### 数据库读取数据

使用数据库节点输入数据，在节点选择中的数据源页签中，选择数据库节点。



根据数据库读取节点使用说明设置，并读取数据：

<http://172.16.51.106/pages/viewpage.action?pageId=920024>

## 训练模型

根据实际需要在节点选择面板的建模页签下选择模型节点

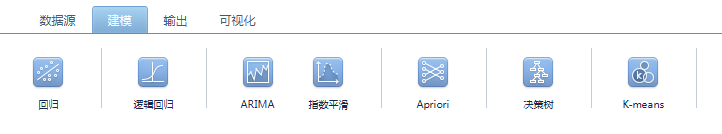


图1 训练模型节点

在工作区中设置选中的模型参数，并点击执行，对开始训练，训练完成后在左下的模型窗口生成对应的训练结果。

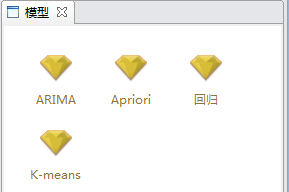


图2 预测模型管理

各模型节点设置方法参照：

<http://172.16.51.106/pages/viewpage.action?pageId=917583>

# 模型预测

## 输入预测数据

输入预测数据的方法仍然采用上文中1.1介绍的方法读取数据。

## 使用预测模型进行预测

对于训练生成的模型，主要分为两种：一种是生成的模型可以用于预测，这样的节点需要输入预测数据，并进行设置，然后对未知的结果进行预测，比如指数平滑、ARIMA等。另一种节点是不需要预测，只需要查看训练结果，比如K-means、Apriori等。

为了统一工作流的工作方式，规定两种训练节点生成的模型采用相同的工作方式，即从模型管理面板拖入预测模型节点，连接数据源，设置，并运行。

预测模型的设置及运行方法见对应的节点使用说明：

<http://172.16.51.106/pages/viewpage.action?pageId=917583>

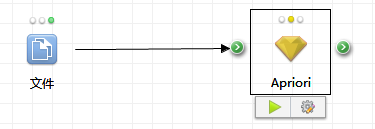


图3 预测工作流

## 查看训练结果

对于2.2中介绍的两种预测模型节点，使用相同的方式查看结果。预测模型设置面板右肩上的执行并预览按钮查看预测模型的结果的二维表，或者执行后在预测模型的右键菜单中的“预览：预测结果集”菜单查看结果。



图4 右肩执行预览按钮

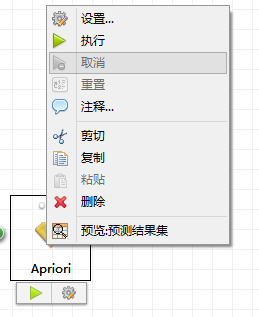


图5 右键预览结果集

为了更加直观的查看预测结果，可以选择使用可视化节点，以图形的方式展示预测结果。选择合适的可视化节点，并调节参数，展示预测结果。

可视化节点设置方法，详见各节点使用说明：

<http://172.16.51.106/pages/viewpage.action?pageId=917583>

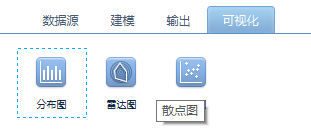


图6 可视化节点