FCM节点使用说明

作者：张青松

目录

[1. FCM算法 2](#_Toc428536449)

[2. FCM节点 2](#_Toc428536450)

# FCM算法

FCM节点采用了Fuzzy -C- Means算法对数据进行聚类的 ，FCM算法是一种基于划分的聚类算法，其思想是是的被划分到同一簇的对象之间的相似度最大，不同簇之间的相似度最小。FCM是普通硬聚类（HCM）的改进，用隶属度确定每个数据点属于某个聚类的程度。

FCM算法是一个简单的迭代过程，在每次迭代中，用欧氏距离量化每个数据点与同一簇数据点的相似度。

从FCM的理论可以知道，FCM算法对于满足正态分布的数据聚类效果较好，并且对于孤立点是敏感的。

# FCM节点

在DataStudio中，FCM在“建模🡪聚类”目录下，与其他建模节点相同，在工作流中为FCM节点添加数据输入节点。如下图：

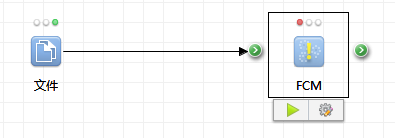


图1 FCM工作流

在输入数据中，必须有数值类型的字段。本文中的输入数据字段类型如图：

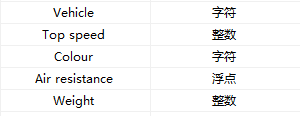


图2 输入数据字段信息

打开FCM设置面板，在设置面板的表格中显示了输入数据中所有数值类型的字段，选择要用作聚类计算的字段，此处做全选。如图：

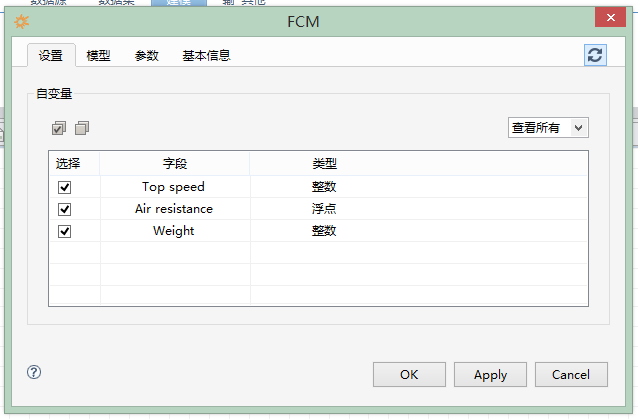


图3 选择参与聚类的数值字段

切换到模型页签，设置FCM算法参数，如图：

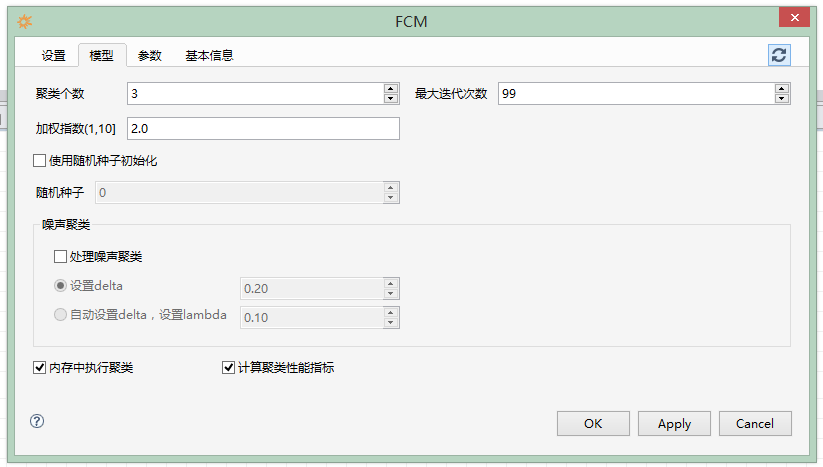


图4 FCM模型参数设置

根据数据以及聚类要求设置各项的值，其中

1. 加权指数的取值范围在
2. 对于孤立点的处理，可以选择使用噪声聚类，在噪声聚类中，可以直接设置delta的值，也可设置lambda，通过lamdba计算适合每一个数据点的delta值。
3. 选择是否在内存中聚类和是否在结果中显示聚类性能指标等选项。

执行后查看聚类结果,如下图：

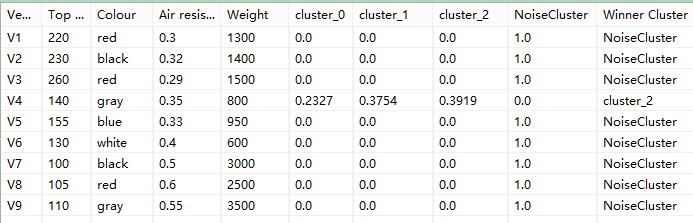


表1 聚类结果

本例中选择了噪声聚类，并将delta设置为0.2；表中前5列数据来自于输入数据，第六列的cluster\_0到cluster\_2是根据设置产生的三个聚类，每一列对应该对象对于该聚类的隶属度，由于已经对每个数据点的所有聚类的隶属度做了归一化处理，所以每个数据点的所有聚类隶属度之和为1。NoiseCluster表示噪声聚类，如果不选择噪声聚类，不会产生这一列。最后一列表示，根据隶属度确定该数据点所属的聚类。

查看模型中对于聚类结果的描述信息，如下





图5 聚类信息

在左侧生成的模型中查看FCM聚类信息，可以从中查看FCM设置的聚类参数，以及聚类中心。如果选择了聚类性能指标，则还会显示聚类内差异、聚类间差异、划分系数、划分熵、Xie-Beni指标和Fuzzy Hypervolumns等聚类信息。