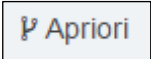


## 3.4.6 关联规则

### 3.4.6.1 Apriori

图标: 

图标:

**描述:** Apriori 是关联规则里一项基本算法, 目的就是在一个数据集中找出项与项之间的关系, 也被称为购物篮分析 (Market Basket analysis)。

#### 字段属性

**项:** 必选。包含若干个项目的集合 (一次事务中的), 一般会大于 0 个。可为所有项目所在的列, 必须要求每一行为一个事务, 事务内项与项之间以空格隔开, 格式如图 337 所示。

item
A2 B1 C3 D3 E1 F1 H1
A2 B1 C3 D3 E1 F1 H1
A2 B1 C3 D3 E1 F1 H1
A2 B1 C3 D3 E1 F1 H1
A2 B2 C3 D3 E1 F1 H1
A1 B2 C1 D1 E1 F1 H1
A1 B1 C1 D1 E1 F1 H1
A1 B2 C1 D1 E1 F1 H1

图 337

#### 参数设置

**最小支持度:** 必选。默认 0.5, 以[A,B]这个关联规则为例来说明,表示 A、B 同时使用的人数占所有用户数的比例。

**最小置信度:** 必选。默认 0.7, 以[A,B]这个关联规则为例来说明,表示使用 A 的用户中同时使用 B 的比例, 即同时使用 A 和 B 的人占使用 A 的人的比例。

**提升度:** 默认 2, 以[A,B]这个关联规则为例来说明,表示“使用 A 的用户中同时使用 B

的比例”与“使用 B 的用户比例”的比值。

最大项目数：默认为 1。

### 输出

表结果：规则（前项、后项、支持度、置信度、提升度）。

报告：无。

### 示例

下面对某数据进行关联分析。原数据如图 338 所示。

item
A2 B1 C3 D3 E1 F1 H1
A2 B1 C3 D3 E1 F1 H1
A2 B1 C3 D3 E1 F1 H1
A2 B1 C3 D3 E1 F1 H1
A2 B2 C3 D3 E1 F1 H1
A1 B2 C1 D1 E1 F1 H1
A1 B1 C1 D1 E1 F1 H1
A1 B2 C1 D1 E1 F1 H1

图 338

- 首先选择关联分析所需的数据，如图 339 所示。
- 设置相应参数，如图 340 所示。
- 运行成功，可选择查看数据，如图 341 所示。

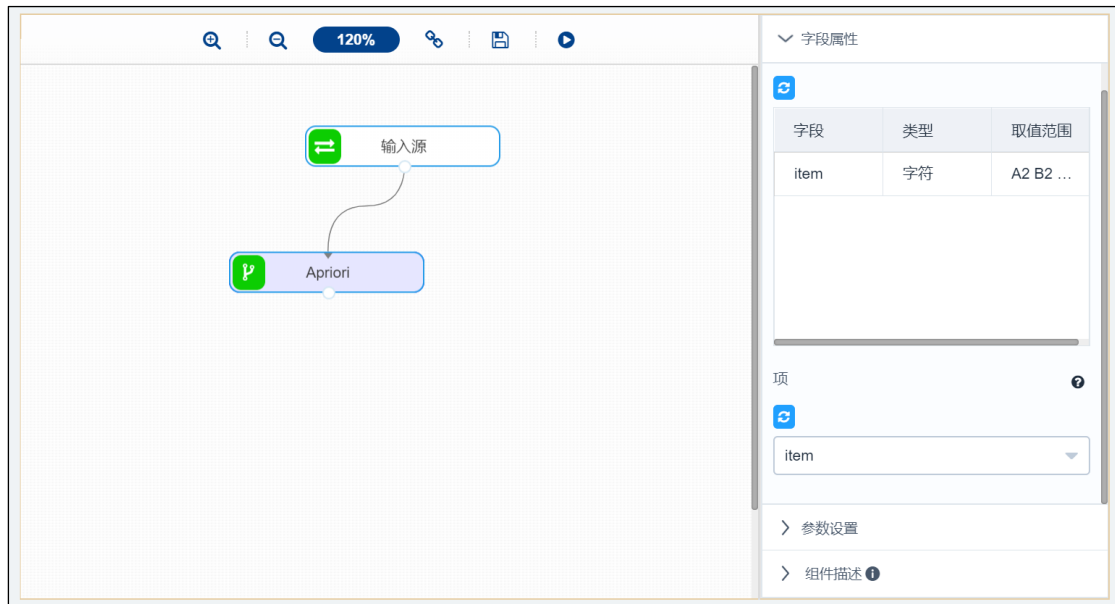


图 339

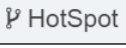


图 340

预览数据				
lhs	rhs	support	confidence	lift
{F4,A3}	H4	0.07849462365591398	0.8795180722891566	1.9709682101901582
{B2,F4}	H4	0.06236559139784946	0.7945205479452054	1.7804918303350388
{E3,C2}	D2	0.09247311827956989	0.7543859649122807	1.8708771929824564
{F4,C3}	H4	0.07526881720430108	0.875	1.9608433734939759
{D2,H4,F3}	A2	0.06236559139784946	0.7532467532467533	1.9732943113224803

图 341

### 3.4.6.2 HotSpot

图标: 

**描述:** HotSpot 是一种主要用于挖掘数值型特征与类别型标签之间的关联规则的算法。HotSpot 能够发现在某个取值区间内, 某个类别发生的概率。

#### 字段属性

前项: 请选择数值型数据。

后项: 请选择类别特征型数据。

如图 342 所示。



The screenshot shows the configuration interface for the HotSpot algorithm. It includes a table of available fields and two dropdown menus for selecting '前项' (Antecedent) and '后项' (Consequent).

字段	类型	取值范围
sepal_width	数值	2.2-3.9
petal_length	数值	1.2-6.7
petal_width	数值	0.1-2.5
species	字符	setosa,...

The '前项' (Antecedent) dropdown menu is set to 'petal\_length' and 'petal\_width'. The '后项' (Consequent) dropdown menu is set to 'species'.

图 342

参数设置

最小支持度：前项与后项同时发生的概率，默认为 0.05。

最小置信度：前项发生的情况下后项发生的概率，默认为 0.2。

步长：步长越大，运行越快，模型效果越低。取值范围为 0-100，建议为 5-20 之间。

如图 343 所示。



参数设置	
最小支持度	?
0.05	
最小置信度	?
0.2	
步长	?
5	

图 343

### 输出

表结果：规则。

报告：无。

### 示例

下面对某数据进行 Hotspot 关联规则分析。

- 选择两列序列作为前项，数据必须为数值型；选择类别特征型序列作为后项。如图 344 所示。
- 点击参数设置，设置如图 345 所示。
- 运行该组件，对组件右击，选择查看数据，结果如图 346 所示。

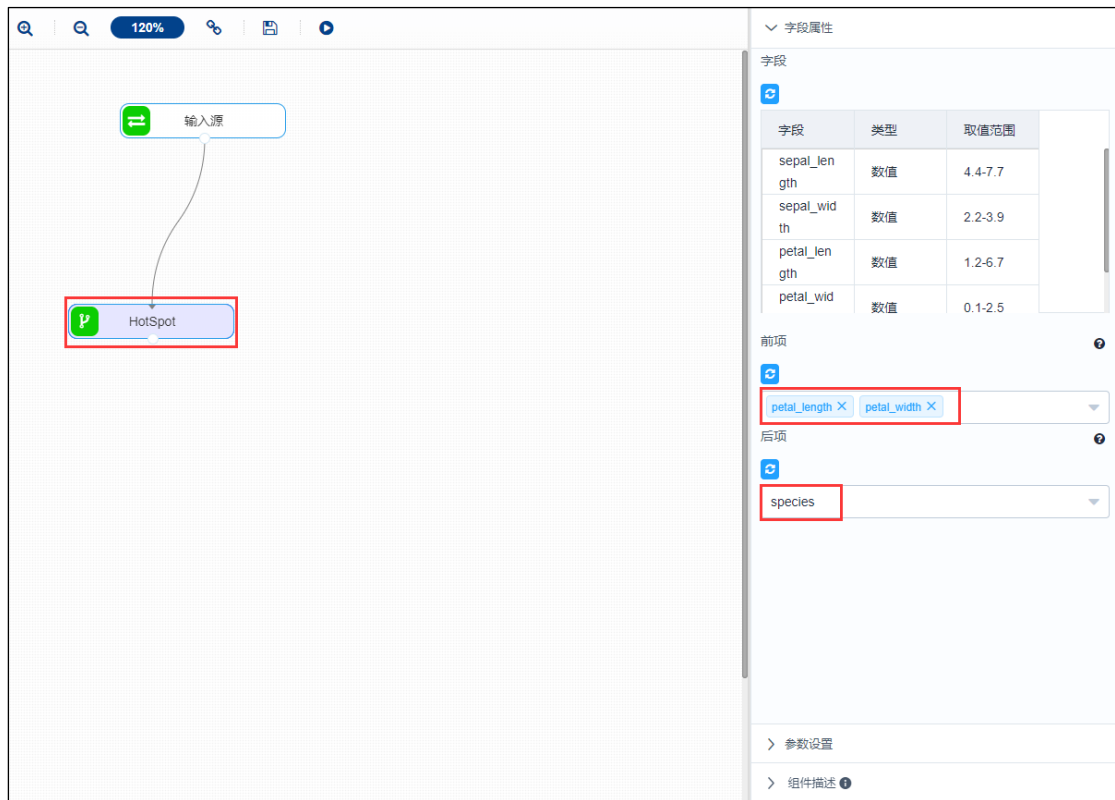


图 344



图 345

预览数据			
conf	left_obj	right_obj	sup
1	((1.0, 1.42), (0.1, 0.205000000000002))	setosa	0.12
1	((3.66, 4.29), (0.415000...000000004, 1.5550000000000002))	versicolor	0.11333333333333333
1	((3.66, 4.29), (0.31, 2.08))	versicolor	0.11333333333333333
1	((3.66, 4.29), (0.31, 2.185))	versicolor	0.11333333333333333
1	((3.66, 4.29), (0.31, 2.29))	versicolor	0.11333333333333333
1	((3.66, 4.29), (0.31, 2.394999999999996))	versicolor	0.11333333333333333

共 32188 条 25 条/页 < 1 2 3 4 5 6 ... 1288 > 前往 1 页

图 346

### 3.4.6.3 FP-growth

图标:

FP-growth

**描述:** FP-growth 将事务数据表中的各个事务数据项按照支持度排序后, 把每个事务中的数据项按降序依次插入到一棵以 NULL 为根结点的树中, 同时在每个结点处记录该结点出现的支持度。

**字段属性**

**项:** 必选。包含若干个项目的集合 (一次事务中的), 一般会大于 0 个。可为所有项目所在的列, 必须要求每一行为一个事务, 事务内项与项之间以空格隔开。

**参数设置**

**最小支持度:** 必选。A、B 同时包含的概率。

**最小置信度:** 必选。发生概率大于多少时认为可信。

**输出**

**表结果:** 规则、频繁项集。

**报告:** 无。

**示例**

- 首先选择关联分析所需的数据，如图 347 所示。
- 设置相应参数，如图 348 所示。
- 运行成功，可选择查看数据，如图 349、图 350 所示。

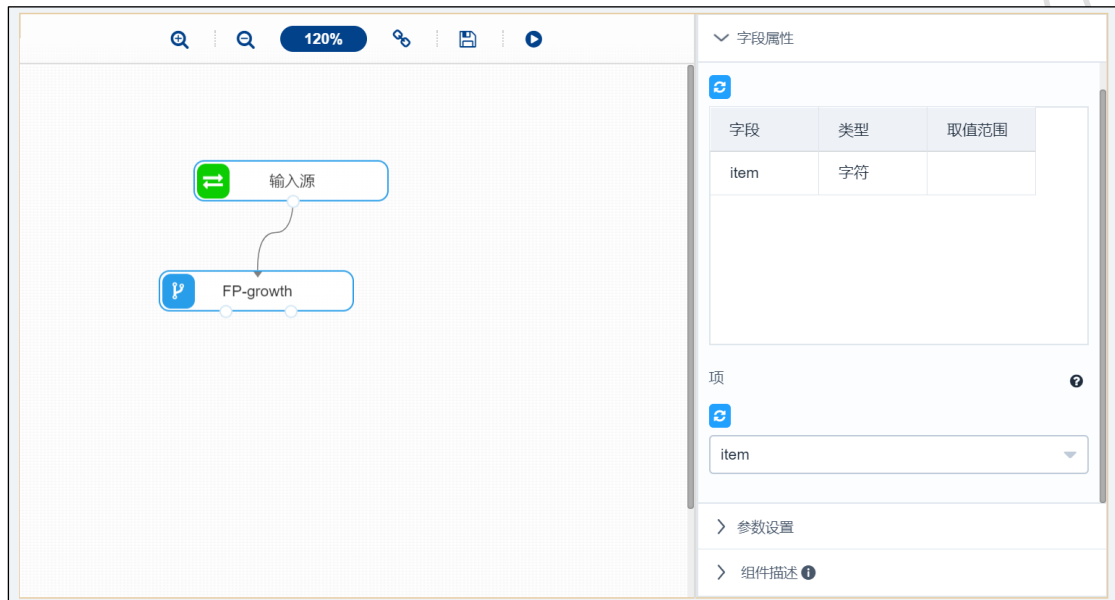


图 347

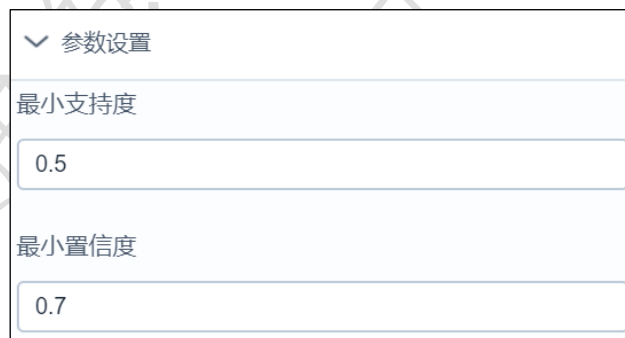


图 348



预览数据		
lhs	rhs	confi
A1	E5	1
F6	E5	1
B2	(C3,D4,E5)	1
E5	(B2,C3,D4)	0.75
C3	(B2,D4,E5)	1
(B2,C3)	(D4,E5)	1
(B2,E5)	(C3,D4)	1
(C3,E5)	(B2,D4)	1

共 16 条  < 1 > 前往  页

图 349

预览数据	
item	support
A1	0.5
(A1,E5)	0.5
F6	0.5
(E5,F6)	0.5
B2	0.75
(B2,E5)	0.75
C3	0.75
(B2,C3)	0.75

共 19 条  < 1 > 前往  页

图 350