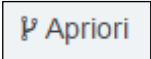


3.4.6 关联规则

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。

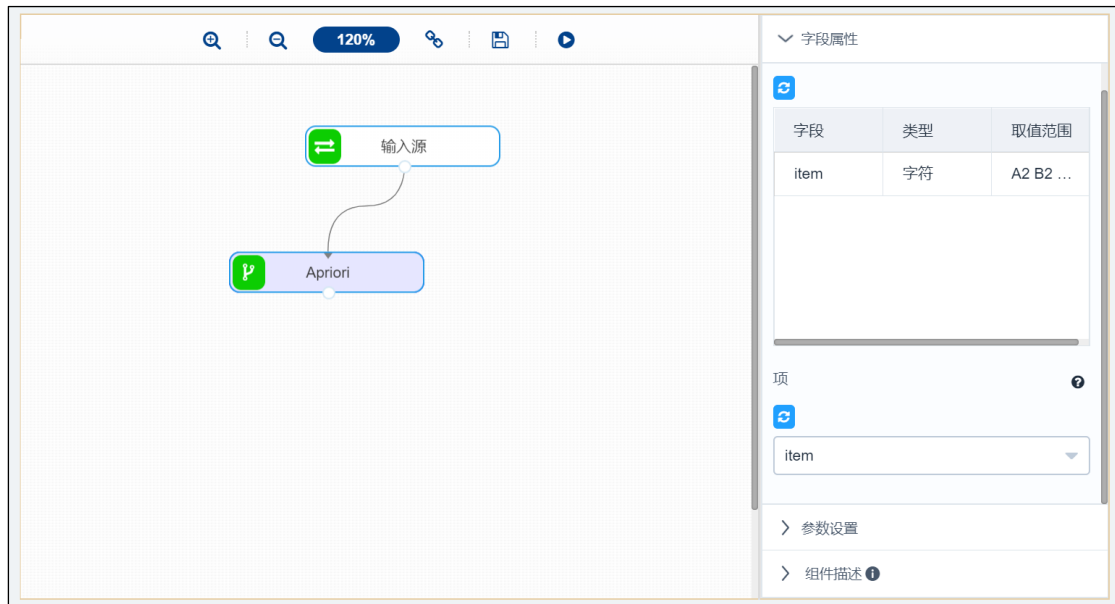


图 339



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