个人信用数据的关联规则挖掘

教学案例

文件状态：	文件标识：	DMS_03_004
[ ] 草稿	当前版本：	V01.00.000
[√] 正式发布	作　　者：	教培部
[ ] 正在修改	参 与 者：	研发部
[ ] 作废	完成日期：	2010-03-18

太普软件（www.tipdm.com）

目　　录

1. 概述

Apriori 算法是一种最有影响的挖掘布尔关联规则频繁项集的算法。该算法的基本思想是：首先找出所有的频集，这些项集出现的频繁性至少和预定义的最小支持度一样。然后由频集产生强关联规则，这些规则必须满足最小支持度和最小可信度。然后使用第1 步找到的频集产生期望的规则，产生只包含集合的项的所有规则，其中每一条规则的右部只有一项，这里采用的是中规则的定义。一旦这些规则被生成，那么只有那些大于用户给定的最小可信度的规则才被留下来。为了生成所有频集，使用了递推的方法。

2. 案例描述

对个人信用数据进行相关性分析。个人信用数据包含下面属性：年龄、性别、地区、收入、婚姻状态、是否生育、是否有车、按揭时是否有抵押、是否有股权计划。利用Apriori算法挖掘个人信用数据中各个属性之间的关联规则。案例的样本数据如下：

age	sex	region	income	married	children	car	mortgage	pep
48	FEMALE	INNER_CITY	17546	NO	YES	NO	NO	YES
40	MALE	TOWN	30085.1	YES	YES	YES	YES	NO
51	FEMALE	INNER_CITY	16575.4	YES	NO	YES	NO	NO
23	FEMALE	TOWN	20375.4	YES	YES	NO	NO	NO
57	FEMALE	RURAL	50576.3	YES	NO	NO	NO	NO
57	FEMALE	TOWN	37869.6	YES	YES	NO	NO	YES
22	MALE	RURAL	8877.07	NO	NO	NO	NO	YES
58	MALE	TOWN	24946.6	YES	NO	YES	NO	NO
37	FEMALE	SUBURBAN	25304.3	YES	YES	YES	NO	NO
54	MALE	TOWN	24212.1	YES	YES	YES	NO	NO
66	FEMALE	TOWN	59803.9	YES	NO	NO	NO	NO
52	FEMALE	INNER_CITY	26658.8	NO	NO	YES	YES	NO
44	FEMALE	TOWN	15735.8	YES	YES	NO	YES	YES
66	FEMALE	TOWN	55204.7	YES	YES	YES	YES	YES
36	MALE	RURAL	19474.6	YES	NO	NO	YES	NO
38	FEMALE	INNER_CITY	22342.1	YES	NO	YES	YES	NO
37	FEMALE	TOWN	17729.8	YES	YES	NO	YES	NO
46	FEMALE	SUBURBAN	41016	YES	NO	NO	YES	NO
62	FEMALE	INNER_CITY	26909.2	YES	NO	NO	NO	YES
31	MALE	TOWN	22522.8	YES	NO	YES	NO	NO
61	MALE	INNER_CITY	57880.7	YES	YES	NO	NO	YES
50	MALE	TOWN	16497.3	YES	YES	NO	NO	NO
54	MALE	INNER_CITY	38446.6	YES	NO	NO	NO	NO
27	FEMALE	TOWN	15538.8	NO	NO	YES	YES	NO
22	MALE	INNER_CITY	12640.3	NO	YES	YES	NO	NO
56	MALE	INNER_CITY	41034	YES	NO	YES	YES	NO

3. 建模过程

本案例通过太普数据挖掘套件（http://www.tipdm.cn）实现建模过程。

更多关于此软件工具的介绍详见：http://www.tipdm.com

✧ 方案管理

数据管理

预测建模

1、选择算法

从菜单中选择指定算法：

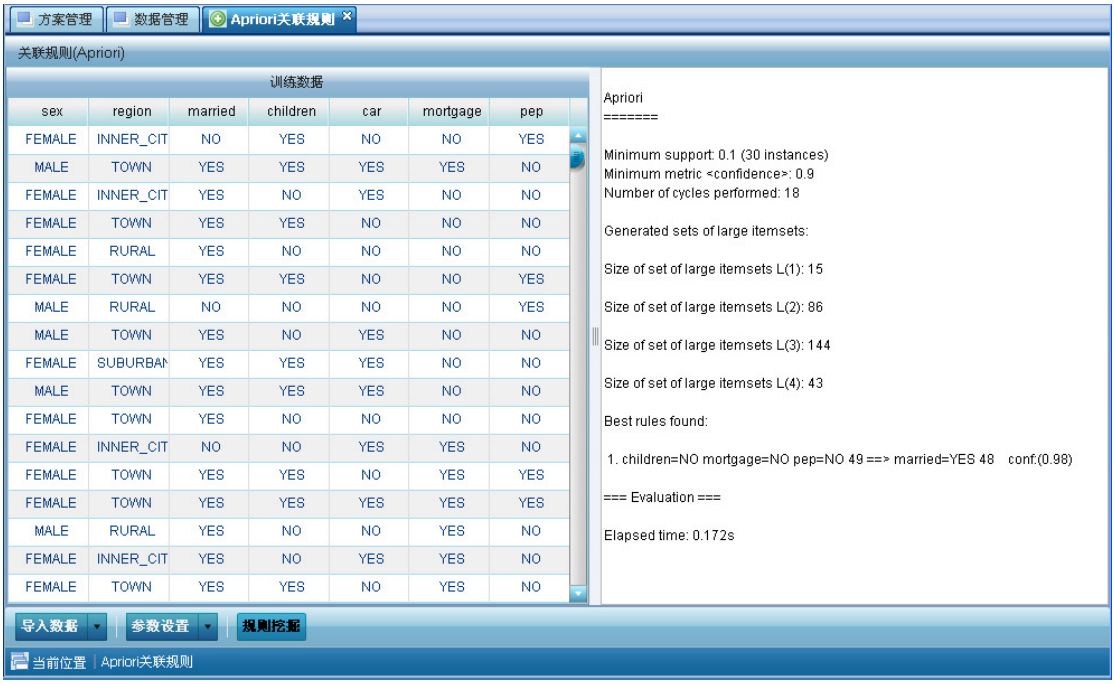
2、数据导入

选择性别（sex）、区域（region）、婚姻状况（married）、是否生育（children）、是否有车（car）、按揭时是否有抵押(mortage)、是否有股权计划(pep)等样本属性，以及样本数据范围，导入样本数据到当前算法操作区。

3、参数设置

1、 规则挖掘

对导入的样本数据使用Apriori关联规则算法对数据进行挖掘。



更多关联规则挖掘应用参见：http://www.tipdm.com/Info.php?barbarism=2