

附录三、RFL 的安装与使用

RFL (Rule Filtering Library) 是由本文作者实现的关联规则筛选类库，支持关联规则的持久化、评价、筛选与查看等功能。下面简单介绍其安装、配置与使用方法。

1. RFL 运行环境的安装与配置

RFL 是采用 Java 语言和 Groovy 语言编写的，编译为 Java 字节码并打包为 Jar 格式，可以在 Java 虚拟机 (Java Virtual Machine, JVM) 上运行。因此，要使用 RFL，首先要安装 Java 运行时环境 (Java Runtime Environment, JRE)；需要以该类库为基础进行开发时通常还需要安装 Java 开发工具包 (Java Development Kit, JDK)。Oracle (已收购推出 Java 语言的 Sun Microsystems 公司) 的 JRE 和 JDK 安装程序可在 <http://www.oracle.com/technetwork/java/javase/downloads/index.html> 下载；也可以使用开源版本的 Java (即 OpenJDK)，其安装说明 (含 JRE 和 JDK) 见 <http://openjdk.java.net/install/index.html>。

RFL 可以以命令行交互方式使用，此时需要安装 Groovy 语言。Groovy 是一种基于 JVM 的动态语言，用该语言编写的代码既可以作为脚本使用 (即以命令行交互方式使用)，也可以编译为可在 JVM 上运行的字节码。Groovy 的安装程序可以在 <http://groovy.codehaus.org/Download> 下载；Eclipse 的 Groovy 插件可以通过 <http://groovy.codehaus.org/Eclipse+Plugin> 或 Eclipse Marketplace 获取。

2. RFL 安装与使用

RFL 源代码、API 文档和库文件可以在 <http://www.c2001.net/rfl.html> 下载。要使用 RFL，只需要确保 Java 类库路径中包含 c2001.jar 和 rflgroovy.jar 即可。下面给出在 Groovy 命令行交互环境中使用 RFL 的示例：

- (1) 启动 Groovy。将类库涉及的 jar 文件拷贝到 Groovy 的 lib 路径下，然后执行 groovysh.bat 或 groovysh.exe 即可启动 Groovy 的命令行交互环境 (Linux/Unix 下则是执行 groovysh)。
- (2) 读取并显示已保存的关联规则。首先要引入相关的类，执行以下语句：

```
import java.util.*;
```

```
import net.c2001.dm.ar.garmf.*;
import net.c2001.dm.ar.rfl.*;
import net.c2001.dm.ar.rfl.base.*;
import net.c2001.dm.ar.ui.ruleviewer.*;
```

然后，读取保存在/tmp/rulesfi.obj 中的关联规则：

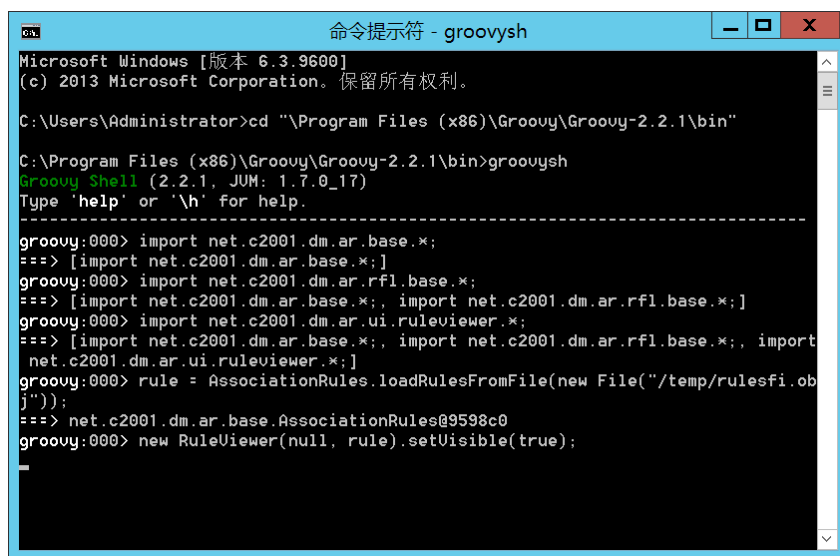
```
rule = AssociationRules.loadRulesFromFile(new File("/temp/rulesfi.obj"));
```

最后，显示这些规则：

```
new RuleViewer(null, rule).setVisible(true);
```

读取的文件中共有 1446 条规则。此时命令行窗口和规则查看窗口如图

1、图 2 所示：

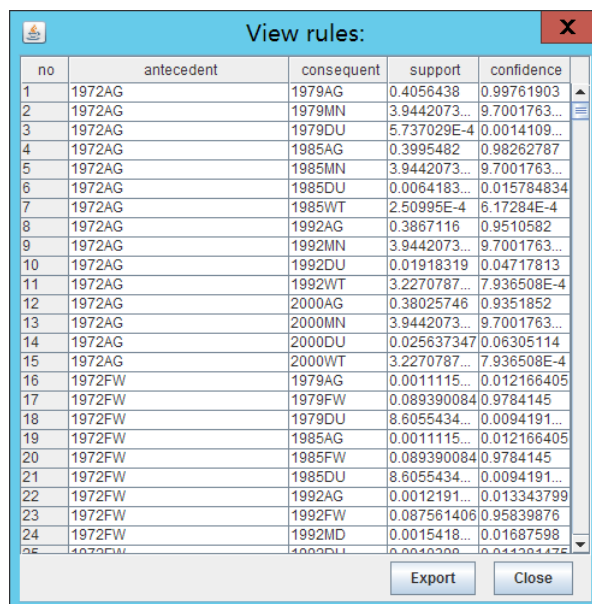


```
命令提示符 - groovysh
Microsoft Windows [版本 6.3.9600]
(c) 2013 Microsoft Corporation. 保留所有权利。

C:\Users\Administrator>cd "\\Program Files (x86)\Groovy\Groovy-2.2.1\bin"

C:\Program Files (x86)\Groovy\Groovy-2.2.1\bin>groovysh
Groovy Shell (2.2.1, JVM: 1.7.0_17)
Type 'help' or '\h' for help.
-----
groovy:000> import net.c2001.dm.ar.base.*;
===> [import net.c2001.dm.ar.base.*;]
groovy:000> import net.c2001.dm.ar.rfl.base.*;
===> [import net.c2001.dm.ar.base.*; , import net.c2001.dm.ar.rfl.base.*;]
groovy:000> import net.c2001.dm.ar.ui.ruleviewer.*;
===> [import net.c2001.dm.ar.base.*; , import net.c2001.dm.ar.rfl.base.*; , import
net.c2001.dm.ar.ui.ruleviewer.*;]
groovy:000> rule = AssociationRules.loadRulesFromFile(new File("/temp/rulesfi.obj"));
===> net.c2001.dm.ar.base.AssociationRules@9598c0
groovy:000> new RuleViewer(null, rule).setVisible(true);
```

图 1. Groovy shell 窗口



| no | antecedent | consequent | support | confidence |
|----|------------|------------|--------------|--------------|
| 1 | 1972AG | 1979AG | 0.4056438 | 0.99761903 |
| 2 | 1972AG | 1979MN | 3.9442073... | 9.7001763... |
| 3 | 1972AG | 1979DU | 5.737029E-4 | 0.0014109... |
| 4 | 1972AG | 1985AG | 0.3995482 | 0.98262787 |
| 5 | 1972AG | 1985MN | 3.9442073... | 9.7001763... |
| 6 | 1972AG | 1985DU | 0.0064183... | 0.015784834 |
| 7 | 1972AG | 1985WT | 2.50995E-4 | 6.17284E-4 |
| 8 | 1972AG | 1992AG | 0.3867116 | 0.9510582 |
| 9 | 1972AG | 1992MN | 3.9442073... | 9.7001763... |
| 10 | 1972AG | 1992DU | 0.01918319 | 0.04717813 |
| 11 | 1972AG | 1992WT | 3.2270787... | 7.936508E-4 |
| 12 | 1972AG | 2000AG | 0.38025746 | 0.9351852 |
| 13 | 1972AG | 2000MN | 3.9442073... | 9.7001763... |
| 14 | 1972AG | 2000DU | 0.025637347 | 0.06305114 |
| 15 | 1972AG | 2000WT | 3.2270787... | 7.936508E-4 |
| 16 | 1972FW | 1979AG | 0.0011115... | 0.012166405 |
| 17 | 1972FW | 1979FW | 0.089390084 | 0.9784145 |
| 18 | 1972FW | 1979DU | 8.6055434... | 0.0094191... |
| 19 | 1972FW | 1985AG | 0.0011115... | 0.012166405 |
| 20 | 1972FW | 1985FW | 0.089390084 | 0.9784145 |
| 21 | 1972FW | 1985DU | 8.6055434... | 0.0094191... |
| 22 | 1972FW | 1992AG | 0.0012191... | 0.013343799 |
| 23 | 1972FW | 1992FW | 0.087561406 | 0.95839876 |
| 24 | 1972FW | 1992MD | 0.0015418... | 0.01687598 |
| 25 | 1972FW | 1992DU | 0.0012191... | 0.013343799 |

图 2. 规则查看窗口

(3) 利用白名单（white list）对规则进行筛选。首先是创建白名单（即一个项

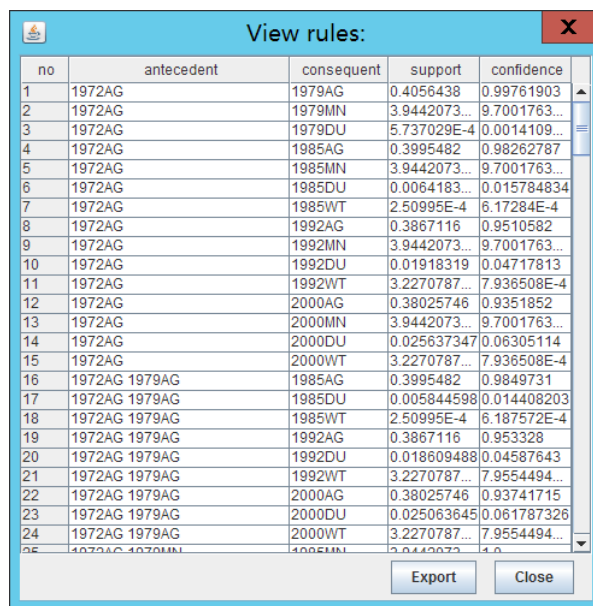
集模板)：

```
whitelist = [];  
iset = ItemSet.parseItemSet("1");  
whitelist<<iset;
```

然后，利用白名单筛选出前件包含第一个字段（即 1972AG）的规则并查看：

```
selector = MemRuleSelector.getInstance();  
newRule = selector.getRulesByContent(rule, whitelist, true, false, null, false,  
false);  
new RuleViewer(null, newRule).setVisible(true);
```

至此，共筛选出符合要求的规则 192 条，筛选如图 3 所示：



| no | antecedent | consequent | support | confidence |
|----|---------------|------------|--------------|--------------|
| 1 | 1972AG | 1979AG | 0.4056438 | 0.99761903 |
| 2 | 1972AG | 1979MN | 3.9442073... | 9.7001763... |
| 3 | 1972AG | 1979DU | 5.737029E-4 | 0.0014109... |
| 4 | 1972AG | 1985AG | 0.3995482 | 0.98262787 |
| 5 | 1972AG | 1985MN | 3.9442073... | 9.7001763... |
| 6 | 1972AG | 1985DU | 0.0064183... | 0.015784834 |
| 7 | 1972AG | 1985WT | 2.50995E-4 | 6.17284E-4 |
| 8 | 1972AG | 1992AG | 0.3867116 | 0.9510582 |
| 9 | 1972AG | 1992MN | 3.9442073... | 9.7001763... |
| 10 | 1972AG | 1992DU | 0.01918319 | 0.04717813 |
| 11 | 1972AG | 1992WT | 3.2270787... | 7.936508E-4 |
| 12 | 1972AG | 2000AG | 0.38025746 | 0.9351852 |
| 13 | 1972AG | 2000MN | 3.9442073... | 9.7001763... |
| 14 | 1972AG | 2000DU | 0.025637347 | 0.06305114 |
| 15 | 1972AG | 2000WT | 3.2270787... | 7.936508E-4 |
| 16 | 1972AG 1979AG | 1985AG | 0.3995482 | 0.9849731 |
| 17 | 1972AG 1979AG | 1985DU | 0.005844598 | 0.014408203 |
| 18 | 1972AG 1979AG | 1985WT | 2.50995E-4 | 6.187572E-4 |
| 19 | 1972AG 1979AG | 1992AG | 0.3867116 | 0.953328 |
| 20 | 1972AG 1979AG | 1992DU | 0.018609488 | 0.04587643 |
| 21 | 1972AG 1979AG | 1992WT | 3.2270787... | 7.9554494... |
| 22 | 1972AG 1979AG | 2000AG | 0.38025746 | 0.93741715 |
| 23 | 1972AG 1979AG | 2000DU | 0.025063645 | 0.061787326 |
| 24 | 1972AG 1979AG | 2000WT | 3.2270787... | 7.9554494... |

图 3. 利用模板筛选出的规则

- (4) 计算规则的提升度 (lift)，筛选出提升度大于 1 的规则，查看后保存筛选结果至文件。首先是创建关联规则评价工具 Measures 类的实例并与刚刚查看的规则绑定：

```
M = new Measures();  
M.setFSets(rule.getFrequentItemset());
```

然后，计算提升度并选出提升度大于 1 的规则并查看：

```
selset = [];  
lift= [];  
newRule.getRules().each({m = M.L(it);  
if(m > 1){selset<<it;lift<<m;}});  
values = new float[lift.size()]  
values = (float[])lift.toArray();  
van = new ValueAndName(values, "Lift");
```

```
newRule.setRules(setset);
new UniViewer(null, newRule, van).setVisible(true);
```

经过这一步筛选最终还剩下 146 条规则，如图 4 所示：

| no | antecedent | consequent | support | confidence | Lift |
|-----|---------------------|------------|--------------|------------|-----------|
| 119 | 1972AG 1979AG 19... | 2000AG | 0.38025746 | 0.9833102 | 2.5672662 |
| 120 | 1972AG 1979AG 19... | 2000DU | 0.01276489 | 1.0 | 7.193449 |
| 121 | 1972AG 1979AG 19... | 2000WT | 7.1712864... | 1.0 | 6.1173506 |
| 122 | 1972AG 1979AG 19... | 1992AG | 0.38025746 | 1.0 | 2.567575 |
| 123 | 1972AG 1979AG 19... | 1992DU | 0.01276489 | 0.66417915 | 5.0458436 |
| 124 | 1972AG 1979AG 19... | 1992WT | 7.1712864... | 1.0 | 6.1361938 |
| 125 | 1972AG 1979AG 19... | 2000DU | 0.0058445... | 1.0 | 7.193449 |
| 126 | 1972AG 1979AG 19... | 1992DU | 0.0058445... | 1.0 | 7.597112 |
| 127 | 1972AG 1979AG 19... | 2000WT | 2.50995E-4 | 1.0 | 6.1173506 |
| 128 | 1972AG 1979AG 19... | 1992WT | 2.50995E-4 | 1.0 | 6.1361938 |
| 129 | 1972AG 1979AG 19... | 1985AG | 0.38025746 | 1.0 | 2.4863153 |
| 130 | 1972AG 1979AG 19... | 1985AG | 0.0064541... | 1.0 | 2.4863153 |
| 131 | 1972AG 1979AG 19... | 1985AG | 0.01276489 | 0.6859345 | 1.7054495 |
| 132 | 1972AG 1979AG 19... | 1985DU | 0.0058445... | 0.3140655 | 2.6462154 |
| 133 | 1972AG 1979AG 19... | 1985WT | 2.50995E-4 | 0.7777778 | 4.774696 |
| 134 | 1972AG 1979MN 1... | 2000MN | 3.9442073... | 1.0 | 49.713013 |
| 135 | 1972AG 1979MN 1... | 1992MN | 3.9442073... | 1.0 | 48.00172 |
| 136 | 1972AG 1979MN 1... | 1985MN | 3.9442073... | 1.0 | 48.00172 |
| 137 | 1972AG 1979DU 1... | 2000DU | 5.737029E-4 | 1.0 | 7.193449 |
| 138 | 1972AG 1979DU 1... | 1992DU | 5.737029E-4 | 1.0 | 7.597112 |
| 139 | 1972AG 1979DU 1... | 1985DU | 5.737029E-4 | 1.0 | 8.42568 |
| 140 | 1972AG 1985AG 19... | 1979AG | 0.38025746 | 1.0 | 2.4552338 |
| 141 | 1972AG 1985AG 19... | 1979AG | 0.0064541... | 1.0 | 2.4552338 |
| 142 | 1972AG 1985AG 19... | 1979AG | 0.01276489 | 1.0 | 2.4552338 |
| 143 | 1972AG 1985AG 19... | 1979AG | 7.1712864... | 1.0 | 2.4552338 |
| 144 | 1972AG 1985MN 1... | 1979MN | 3.9442073... | 1.0 | 48.00172 |
| 145 | 1972AG 1985DU 1... | 1979AG | 0.0058445... | 0.9106145 | 2.2357714 |
| 146 | 1972AG 1985WT 1... | 1979AG | 2.50995E-4 | 1.0 | 2.4552338 |

图 4. 根据评价指标筛选出的规则

最后保存规则至文件：

```
newRule.saveToFile(new File("/temp/newrules.obj"));
```

RFL 中模板使用方法请参见本文第五章；RFL 支持的项集评价指标见表 1：

表 1 项集评价指标（名称与文献）

| 评价指标名称 | | | 公式 | 取值范围 | 建议阈值 | 参考文献 |
|--------|--------------------|-----|-------------------------|----------|------|------|
| 中文 | 英文 | 简称 | | | | |
| 支持度 | support | S | $P(I)$ | [0,1] | 大于 0 | [1] |
| 势 | cardinality | n | $ I $ | N^+ | - | [2] |
| 全支持度 | all-confidence | All | $S(I)/maxSs$ | [0,1] | 大于 0 | [3] |
| 任意支持度 | any-confidence | Any | $S(I)/minSs$ | [0,1] | 大于 0 | [3] |
| 期望 | expectation | E | $P(I)$ | [0,1] | 大于 0 | [4] |
| 方差 | variance | D | $P(I)(1 - P(I))$ | [0,0.25] | - | [4] |
| 标准差 | standard deviation | SD | $\sqrt{P(I)(1 - P(I))}$ | [0,0.5] | - | [4] |
| 期望 2 | expectation2 | EI | $\prod_{x=1}^n P(i_x)$ | [0,1] | - | [5] |

| | | | | | | |
|-----------|---------------------|---|---|---|---|-----|
| 信息获取 度 | information gain | I | $\prod_{x=1}^n -\log_2(P(i_x)) \cdot w_x$ | - | - | [6] |
|-----------|---------------------|---|---|---|---|-----|

这些评价指标的详细介绍可以参见本文第五章或对应参考文献，其中信息获取度公式进行了变更，其余则基本一致。

RFL 实现了的规则评价指标见表 2:

表 2 规则评价指标

| 评价指标名称 | | | 公式 | 取值范围 | 阈值 | 参考文献 |
|--------|----------------------------|------|--|----------------|------------|------|
| 中文 | 英文 | 简称 | | | | |
| 支持度 | support | RS | $P(AB)$ | [0,1] | (0,1] | [1] |
| 置信度 | confidence | C | $P(B A)$ | [0,1] | 大于 0.5 | [1] |
| 协方差 | covariance | corr | $P(AB)-P(A)P(B)$ | [-0.25,0.25] | 大于 0 | [7] |
| 相关系数 | correlation coefficient | r | $\frac{\text{corr}(A \rightarrow B)}{SD(A)SD(B)}$ | [-1,1] | 大于 0 | [4] |
| 提升度 | lift | L | $\frac{P(AB)}{P(A)P(B)}$ | $[0, +\infty)$ | 大于 1 | [8] |
| 确信度 | conviction | Conv | $\frac{P(A)P(\bar{B})}{P(A\bar{B})}$ | $(0, +\infty)$ | 大于 1 | [8] |
| 兴趣度 | interestingness | I1 | $P(B A)-P(B)$ | (-1,1) | 大于 0 | [7] |
| 兴趣度 | interestingness | I2 | $\frac{P(B A)-P(B)}{\max(P(B A), P(B))}$ | [-1,1] | 大于 0 | [9] |
| IS 兴趣度 | IS | IS | $\sqrt{P(A B)P(B A)}$ | [0,1] | 大于 S 阈值 | [4] |
| 获取度 | gain | G | $P(AB) - \text{minconf} \cdot P(A)$ | [-1,1] | 大于 0 | [10] |
| 有效度 | validity | V | $P(AB)-P(\bar{A}B)$ | [-1,1] | 大于 0 | [11] |
| 改进度 | improvement | Imp | $\min(\Delta C_1, \dots, \Delta C_m)$ | (-1,1) | 大于 0 | [12] |
| 常见度 | commonness | Com | $RS(A \rightarrow B)C(A \rightarrow B)$ | [0,1] | 大于 0 | [13] |
| 对比影响度 | contrast influence | CI | $\begin{cases} -1 + \frac{\text{corr}(A \rightarrow B)}{\text{corr}(A \rightarrow B)} & \text{corr}(A \rightarrow B) < 1 \\ 1 - \frac{\text{corr}(A \rightarrow \bar{B})}{\text{corr}(A \rightarrow B)} & \text{corr}(A \rightarrow B) \geq 1 \end{cases}$ | [-1,1] | 大于 0 | [14] |

| | | | | | | |
|-----|-----------------|----|---|----------------------|-----|------|
| 信任度 | trust | T | $\frac{P(AB)}{P(A \cup B)}$ | [0,1] | - | [15] |
| 参考度 | consult | C1 | $\frac{P(B A)-P(B \bar{A})}{\max(P(B A)-P(B \bar{A}))}$ | [-1,1] | 大于0 | [16] |
| 参考度 | consult | C2 | $L(A \rightarrow B)L(\bar{A} \rightarrow B)S(B)$ | $(-\infty, +\infty)$ | 大于0 | [17] |
| 兴趣度 | interest | I3 | $P(AB)-minconf$ | - | - | [18] |
| 提升率 | upgrade rate | Up | $\frac{P(B A)}{P(B \bar{A})}$ | $[0, +\infty)$ | 大于1 | [19] |
| 影响率 | influence ratio | IR | $\frac{P(B A)-P(B \bar{A})}{P(B A)+P(B \bar{A})}$ | [-1,1] | 大于0 | [20] |

其中部分评价指标本文正文部分未进行介绍，其详细介绍可参考对应文献。

参考文献:

- [1] AGRAWAL R, IMIELŃSKI T, SWAMI A. Mining association rules between sets of items in large databases[C]//Proc of the 1993 ACM SIGMOD international conference on Management of data. New York: ACM Press, 1993: 207-216.
- [2] 苏占东, 游福成, 杨炳儒. 关联规则的综合评价方法研究与实例验证[J]. 计算机应用, 2004, 24(10): 17-20.
- [3] OMIECINSKI ER. Alternative interest measures for mining associations in databases[J]. IEEE Trans on Knowledge and Data Engineering, 2003, 15(1): 57-69.
- [4] TAN P N, KUMAR V, SRIVASTAVA J. Selecting the right interestingness measure for association patterns[C]//Proc of the 8th ACM SIGKDD conference on Knowledge discovery and data mining. New York:ACM Press, 2002: 32-41
- [5] AGGARWAL C C, YU P S. A new framework for itemset generation[C]//Proc of the 17th ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems. New York: ACM Press, 1998: 18-24.
- [6] 郝雷, 王咏, 盛焕焯. 通过先验知识挖掘更有意义的关联规则[J]. 计算机仿真, 2005, 22(3): 99-102.
- [7] CHEN M S, HAN J, YU P S. Data mining: an overview from a database perspective[J]. IEEE Trans on Knowledge and data Engineering, 1996, 8(6): 866-883.
- [8] BRIN S, MOTWANI R, ULLMAN J D, et al. Dynamic itemset counting and implication rules for market basket data[C]// Proc of the 1997 ACM SIGMOD international conference on Management of data. New York: ACM Press, 1997: 255-264.

- [9] 周欣, 沙朝锋, 朱扬勇. 兴趣度——关联规则的又一个阈值[J]. 计算机研究与发展, 2000, 37(5): 627-633.
- [10] BRIN S, RASTOGI R, SHIM K. Mining optimized gain rules for numeric attributes[J]. IEEE Trans on Knowledge and Data Engineering, 2003, 15(2): 324-338.
- [11] 罗可, 吴杰. 怎样获得有效的关联规则[J]. 小型微型计算机系统, 2002, 23(6): 711-713.
- [12] BAYARDO Jr. R J, AGRAWAL R, GUNOPULOS D. Constraint-based rule mining in large, dense databases[C]// Proc of the 15th International Conference on Data Engineering. IEEE Press, 1999: 188-197.
- [13] KLEMETTINEN M, MANNILA H, RONKAINEN P, et al. Finding interesting rules from large sets of discovered association rules[C]//Proc of the 3rd international conference on Information and knowledge management. New York: ACM Press, 1994: 401-407.
- [14] 郑尚志, 梁宝华, 赵小龙, 等. 正负关联规则量化方法[J]. 计算机工程, 2009, 35(15): 74-75.
- [15] 刘浩钊, 汤娜, 冯智圣. 关联规则衡量标准的改进模型[J]. 广东自动化与信息工程, 2003, 24(2): 26-28.
- [16] 林嘉宜, 彭宏, 郑启伦, 等. 基于参考度的关联规则挖掘[J]. 计算机应用, 2005, 25(8): 1827-1829.
- [17] 柳景超, 宋胜锋. 基于参考度的有效关联规则挖掘[J]. 火力与指挥控制, 2011, 36(5): 79-81.
- [18] 汪慎文, 刘坤起, 石艳丽. 基于兴趣度的多值关联规则挖掘[J]. 微计算机信息, 2008, 24(24): 98-99.
- [19] 马占欣, 王新社, 黄维通, 等. 对最小置信度门限的置疑[J]. 计算机科学, 2007, 34(6): 216-218.
- [20] 马占欣, 黄维通, 陆玉昌. 相关度计算方法存在的问题及修正[J]. 计算机工程, 2007, 33(11): 67-69.

