

基于改进关联规则的维修专业组合与优化模型

孙志刚, 朱小冬, 王毅刚

(军械工程学院 装备指挥与管理系, 石家庄 050003)

摘要: 针对维修过程中各专业忙闲不均和协同工作效率低下等问题, 提出了一种基于改进关联规则的维修专业组合和优化方法。首先, 以均衡任务量为目标, 建立了维修专业关联度分析模型; 其次, 以功能聚合为目标, 提出了以较大概率共同承担任务且任务量偏小的维修专业组合方法; 最后, 从维修任务和维修专业两方面考虑, 建立了组合维修专业的优化模型。通过具体实例分析, 证明了该方法的有效性和正确性。

关键词: 维修任务; 维修专业; 关联规则; 组合优化

中图分类号: TP311 **文献标志码:** A **文章编号:** 1001-3695(2013)02-0410-04

doi:10.3969/j.issn.1001-3695.2013.02.025

Model of combining and optimizing maintenance specialties based on improved association rule

SUN Zhi-gang, ZHU Xiao-dong, WANG Yi-gang

(Dept. of Equipment Command & Management, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: This paper proposed a method of combining and optimizing maintenance specialties based on improved association rule to solve the problem that task amount of specialties were not balanced and cooperation efficiency was low. Firstly, it established analytic model of correlation degree of maintenance specialties with the goal of balancing the task amount. Secondly, method of combining maintenance specialties that undertook tasks together with great probability and were assigned small amounts of task. At last, considering the maintenance task and specialty, it established the optimized model of combined specialties. According to analyze the example, presented, it proved the correctness and effectiveness of the proposed method.

Key words: maintenance task; maintenance specialty; association rule; combination and optimization

随着装备技术的构成和功能组成越来越复杂, 一个武器装备往往是多系统、多技术的综合体。例如自行火炮和主战坦克都是集车、炮(或导弹)、防化、装甲、通信、弹药于一体^[1], 单个维修专业难以承担损伤装备的全面修复, 需要不同功能的维修专业协作来完成。因此, 维修专业之间以任务为“纽带”产生了某种潜在的关系, 例如维修任务在需要专业 A 的前提下还需要专业 B 或者同时需要专业 A 和 B, 揭示这种关系对于不同功能的专业由“专”到“综”进行聚合、提高专业协同维修的效率具有重要的意义。

关联规则挖掘理论是从大量、不完全、模糊甚至是随机的数据中提取隐含在其中的、人们事先不太清楚但又是潜在有用的信息和知识的过程^[2]。其广泛应用于商业营销^[3]、生物信息^[4]、故障诊断^[5]、医疗诊断^[6]和科学数据分析^[7]等领域。本文旨在利用关联规则挖掘技术从日常积累的维修任务信息中找出形如“专业 A → 专业 B”的关联规则, 并对冲突的专业组合方案进行优化, 为维修专业组合提供数据支持。

1 维修专业组合原则

原则 1 聚合功能, 适应任务需求。

维修专业组合是以任务需求为牵引, 根据不同功能的专业共同承担任务的可能性, 将关联关系较为紧密的专业进行组合。

原则 2 控制规模, 均衡任务量。

考虑到任务量较少的维修专业产生的边际效益不明显, 按照“小专业, 大集中”的维修要求, 优先组合较大概率共同承担任务且任务量偏小的维修专业。

2 维修专业组合模型

2.1 基于任务量的维修专业关联度分析

如果仅仅采用 0-1 二值法来表征维修专业与维修任务的关联关系, 可能导致通过任务共担而组合的维修专业规模过大, 而规模偏小的维修专业则产生的边际效益不明显。因此, 根据原则 2 应当优先组合任务量偏小的专业。

假设维修任务的集合 $T = \{T_1, T_2, \dots, T_m\}$, 对应的维修专业全集 $S = \{S_1, S_2, \dots, S_n\}$, 承担维修任务 $T_i \in T (i = 1, 2, \dots, m)$ 第 j 个维修专业 $S_j \in S (j = 1, 2, \dots, n)$ 所需的修理时间为 t_{ij} , 如表 1 所示。

收稿日期: 2012-06-05; 修回日期: 2012-08-06

作者简介: 孙志刚(1984-), 男, 河北石家庄人, 讲师, 博士, 主要研究方向为装备保障理论与应用(sun1984126@163.com); 朱小冬(1964-), 男, 教授, 博士, 装备保障指挥自动化、软件保障理论与应用; 王毅刚(1976-), 男, 讲师, 博士, 主要研究方向为软件保障理论与应用。

表 1 维修任务、维修专业和维修时间关系

| 任务 | 专业 | | | |
|----------|----------|----------|----------|----------|
| | S_1 | S_2 | ... | S_n |
| | 时间 | | | |
| T_1 | t_{11} | t_{12} | ... | t_{1n} |
| T_2 | - | t_{22} | ... | t_{2n} |
| \vdots | \vdots | - | \vdots | - |
| T_m | t_{m1} | t_{m2} | ... | t_{mn} |

表 1 中“-”表示该维修任务在相应的专业领域内没有维修需求。因此,如何保证任务量较小的维修专业具有较高的组合几率是首先要解决的问题。基本做法是通过对各专业的维修时间进行标准化处理,使得任务量较小的维修专业具有较高的组合几率;反之亦成立。假设专业 $S_j \in S$ 承担维修任务 T_i 的修理时间为 t_{ij} ,采用极小极大值的数学变换标准化为^[8]

$$k_{ij} = \frac{\max_{1 \leq j \leq n} t_{ij} + \min_{1 \leq j \leq n} t_{ij} - t_{ij}}{\max_{1 \leq j \leq n} t_{ij}} \quad (1)$$

其中: k_{ij} 是维修时间 t_{ij} 标准化以后的规范值,表明维修专业 S_j 与其他关联专业的组合几率且 $k_{ij} \in \left[\frac{\min_{1 \leq j \leq n} t_{ij}}{\max_{1 \leq j \leq n} t_{ij}}, 1 \right]$,维修时间 t_{ij} 越小,则被组合的概率或维修专业的关联程度越紧密。

然而,不同维修任务对于专业组合要求也不相同,例如对于维修时间较高的专业组合要求相对“苛刻”一些,而这个要求就通过专业的最小关联度体现,它是相关联的维修专业进行组合的阈值,记做 $MSG(i)$,定义如下:

定义 1 对于维修任务集 $T = \{T_1, T_2, \dots, T_m\}$,其中任意的维修任务 $T_i \in T$ 对于承担其维修的专业集合 $S = \{S_1, S_2, \dots, S_n\}$ 均被赋予组合的最小关联度,记做 $MSG(i)$ 。

定义 2 对于维修专业 $S = \{S_1, S_2, \dots, S_n\}$,承担任务 $T_i \in T$ 的维修时,对应维修专业关联度为 $k_{ij} (j = 1, 2, \dots, n)$,当且仅当 $k_{ij} \geq MSG(i)$ 时,维修专业 $S_j \in S$ 才能够与共同承担任务的相关维修专业组合,即 $S = \{S_j | k_{ij} \geq MSG(i)\}$ 。

然而, k_{ij} 并不是专业组合的唯一条件和约束。事实上,还取决于专业 S_j 与其他专业共同承担任务的频度,即以较大概率共同承担维修任务才使得维修专业组合具有意义。

2.2 多最小支持度的关联规则算法^[9,10]

关联规则最具代表性的算法是 Apriori 算法,可形式化描述为:设 $I = \{i_1, i_2, \dots, i_m\}$ 是 m 个不同项目 (items) 的集合,而 $T = \{t_1, t_2, \dots, t_n\}$ 是针对 I 的所有事务的集合,每一个事务 t_i 包含都是 I 的子集,一般把包含 0 个或多个项的集合称为项集 (itemset),如果一个项集包含 k 个项则称为 K -项集。关联规则表示为 $X \rightarrow Y$,其中 $X \subset I, Y \subset I$,且 $X \cap Y = \emptyset$ 。关联规则强度可以用支持度 (support) 和置信度 (confidence) 来度量。其中,支持度表示 X 和 Y 在给定事务集中出现的频繁程度,即

$$\text{support}(X \rightarrow Y) = |X \cap Y| / |T|$$

置信度表示 Y 在包含 X 的事务中出现的可能性,是 Y 在给定 X 下的条件概率,即

$$\text{confidence}(X \rightarrow Y) = |X \cap Y| / |X|$$

挖掘关联规则问题就是产生支持度和置信度分别大于给定的最小支持度 (min_sup) 和最小可信度 (min_conf) 的有效数据,并将满足最小支持度阈值的项集称做频繁项集。多最小支持度的关联规则挖掘算法是在保持以前支持度定义的基础上根据事务项本身的特点分别规定其各自的最小支持度的挖掘方法,描述如下:

定义 1 对于项集 $I = \{i_1, i_2, \dots, i_m\}$,其中任一事务项均被赋予它自身所需满足的最小支持度,称为最小项支持度,记做 $MIS(i)$ 。

定义 2 对于一挖掘规则: $X_1 \wedge X_2 \wedge \dots \wedge X_k \rightarrow X_{k+1} \wedge \dots \wedge X_n$,其中 $X_i \in I (i = 1, 2, \dots, k, k + 1, \dots, n)$,那么规则所包含的各个项的最小支持度的最小值 $\min(MIS(X_1), MIS(X_2), \dots, MIS(X_n))$ 即为规则所需满足的最小支持度。

对于维修专业而言,使用比较频繁的专业事务项具有较高的最小支持度限制,相反则有较低的最小支持度限制。在维修专业组合的过程中,一方面可以筛选出具有合并意义的维修专业;另一方面不会忽略和遗漏重要的但是使用几率较少的维修专业。

2.3 基于改进关联规则的维修专业组合模型

多最小支持度关联规则算法并没有直接涉及数据自身的属性,常见挖掘算法的样本数据为二元属性,而组合维修专业之间并不是“非此即彼”的情况,具有一定的模糊性和不分明性。鉴于此,提出了模糊数据的多最小支持度关联规则挖掘算法,目的是挖掘符合最小关联度要求且满足最小支持度的关联维修专业。其中,最小支持度定义了项目所需覆盖事件样本数量的下限;而最小可用度则是事务对模糊数据有效性筛选。

模糊数据的多最小支持度关联规则挖掘算法如下:

输入:挖掘数据库 DB,最小支持度项集 $MIS(1), MIS(2), \dots, MIS(n)$;维修专业最小关联度集 $MSG(1), MSG(2), \dots, MSG(m)$ 。

输出:频繁数据集 L_o 。

```

1 SDT = database DB
2 for (i = 1; i ≤ 任务数量 m; i++) do
3 begin
4 for (j = 1; j ≤ 专业数量 n; j++) do
5 Si* = {s | s ∈ SDT ∧ s.kij ≥ MSG(j)} ; //得到满足维修专业最小关联度的项集
6 end
7 L1 = {s ∈ S1*} ; //产生维修专业的项集 L1 (临时)
8. for (k = 2; Lk-1 ≠ ∅; k++) do //产生 k ≥ 2 的维修专业频繁项集
9 begin
10 Sk = apriori-gen(Lk-1) ; //根据维修专业 Lk-1 频繁项集产生 K 维候选项目集
11 Sk* = ∅;
12 for all entries t ∈ Sk-1* do
13 begin
14 St = {s ∈ Sk | (s - s[k]) ∈ t.set_of_itemsets ∧ (s - s[k-1]) ∈ t.set_of_itemsets} ; //验证维修专业集合 Sk 任意子集都是频繁项
15 for all candidates s ∈ Sk-1* do
16 s.count ++;
17 s.support = (s.count) / n ; //计算候选项 Sk 中 s 项的支持度
18 if (St ≠ ∅) then Sk* = <t.TID, St> ; //确定维修专业潜在的频繁项目集 Sk*
19 end
20 for all entries t ∈ Sk do
21 if (St ≠ 0) then
22 begin
23 minsup = min{MIS(Sti), i = 0, 1, 2, ..., k-1} ; //求解 St 项的最小支持度
24 Lk = {St ∈ Sk | s.support ≥ MIS(s)} ;

```

```

25 end
26 end
27  $L_1 = \{s \in S_1^* \mid s.\text{support} \geq \text{MIS}(s)\}$ ; //产生  $L_1$  的频繁集
28  $\text{result} = \cup L_k$ 
    
```

首先,算法第 2~6 步得到满足最小关联度 MSG 的维修专业项集 S_1^* ,然后利用第 7 步算法生成临时的项集 L_1 并计算相应支持度;其次,第 8~18 步算法完成由临时 L_1 生成新的候选项 S_2 ,其中函数 $\text{apriori-gen}(L_{k-1})$ 是维修专业 $(k-1)$ -项频繁集 L_{k-1} 通过自连接操作产生 k -维候选项集 S_k ;算法的第 19~24 步主要是候选项集 S_2 通过与项集 S_2^* 进行比较,求解得到 S_2 每个候选项的支持度;算法第 26 步生成由临时项集 L_1 中满足最小支持度的维修专业组成的正式频繁项集 L_1 。

模糊数据的多最小支持度关联规则挖掘算法的改动主要包括两个方面:a)充分考虑了样本数据的模糊性,提出了有效筛选模糊数据的挖掘策略;b)在生成 L_1 频繁集的算法上进行了改动(步骤 7 和 27),通过临时 L_1 (步骤 7)产生所有的 C_k 和 C_k^* 项集之后,生成正式的频繁 1-项集 L_1 ,避免了在多最小支持度下遗漏或丢失满足最小支持度的项集。

3 组合维修专业优化模型

在关联维修专业挖掘过程中,某一专业可能同时与多个维修专业相关联且满足组合要求,如维修专业 $A \rightarrow B$ 相关联的同时维修专业 $A \rightarrow C$ 也相关联。显然,这些可行的专业组合方案在任务共担上是相互冲突的。就维修任务和维修专业而言,组合维修专业优化主要从以下两方面进行考虑:a)对维修任务全集的影响,其中支持度是描述组合维修专业对于任务全集影响程度的关键参数;b)维修专业自身的关联程度,其中置信度是评估组合维修专业相互影响程度的主要依据。

3.1 基于支持度的组合维修专业优化模型

皮尔森相关系数(Pearson's correlation)是用于分析一些现象或事物之间在一定的数量上存在依存关系的技术方法,就是一种现象发生数量上的变化,就会影响到与它有关的现象也会发生某种程度上的变化^[11]。因此,通过适当转换以支持度作为描述皮尔森相关系数的主要参数,进而作为度量组合维修专业关联程度的基本依据,二元变量的皮尔森相关系数定义^[12]和转换为

$$\alpha(A, B) = \frac{n_{AB} * n_{\bar{A}\bar{B}} - n_{A\bar{B}} * n_{\bar{A}B}}{\sqrt{n_A * n_{\bar{A}} * n_B * n_{\bar{B}}}} = \frac{\frac{1}{N^2} * (n_{AB} * n_{\bar{A}\bar{B}} - n_{A\bar{B}} * n_{\bar{A}B})}{\frac{1}{N^2} * \sqrt{n_A * n_{\bar{A}} * n_B * n_{\bar{B}}}} = \frac{s(A, B) * s(\bar{A}, \bar{B}) - s(A, \bar{B}) * s(\bar{A}, B)}{\sqrt{s(A) * s(\bar{A}) * s(B) * s(\bar{B})}} \quad (2)$$

其中: $\alpha(A, B)$ 表示维修专业 A 和 B 的相关系数且 $\alpha(A, B) \in [-1, 1]$; n_{AB} 表示任务集 T 中同时需要专业 A 和 B 的数量, $n_{\bar{A}\bar{B}}, n_{A\bar{B}}, n_{\bar{A}B}, n_A, n_B$ 等参数也有相类似含义;支持度 $s(A, \bar{B})$ 表示任务集中需要专业 A 但是不需要专业 B 的概率,其他支持度变量也具有类似的含义。从统计意义上来讲,维修专业相关系数 $\alpha(A, B)$ 不同范围的取值可以解释如下:

$$\alpha(A, B) = \begin{cases} \in (0, 1] & \text{表明维修专业 } A \text{ 和 } B \text{ 是正相关} \\ = 0 & \text{表明维修专业 } A \text{ 和 } B \text{ 是独立的} \\ \in [-1, 0) & \text{表明维修专业 } A \text{ 和 } B \text{ 是负相关} \end{cases}$$

a)当 $\alpha(A, B) \in (0, 1]$,表示维修专业 A 和 B 出现的频度互相提升了双方出现的可能性,即两者数量变化的方向是一致的,并且 $\alpha(A, B)$ 值越大表明其关联程度越高。

b)当 $\alpha(A, B) = 0$,维修专业 A 和 B 相互独立,表示专业 A 与 B 之间虽然有联系但是互不影响、各自独立。

c)当 $\alpha(A, B) \in [-1, 0)$,表示两者出现的频度相互降低了双方出现的可能性,并且 $\alpha(A, B)$ 值越小这种趋势越明显。

3.2 基于置信度的组合维修专业优化模型

余弦函数也是度量二元向量之间相似相关性的主要方法之一。如果将维修专业 A 和 B 看做向量,则 $s(A, B) = A \cdot B$ 表示两个向量的点积;而 $|A| = \sqrt{s(A)}$ 和 $|B| = \sqrt{s(B)}$ 表示向量 A 和 B 的大小。根据余弦函数,维修专业 A 和 B 的相关度可以表示为

$$\beta(A, B) = \frac{s(A, B)}{\sqrt{s(A) * s(B)}} = \frac{\sqrt{\frac{s(A, B)}{s(A)} * \frac{s(A, B)}{s(B)}}}{\sqrt{c(A \rightarrow B) * c(B \rightarrow A)}} \quad (3)$$

其中: $c(A \rightarrow B)$ 表示置信度。组合维修专业之间的关联程度则由其置信度共同决定,从式(3)中可以看出,其中任意一方的置信度都可能影响组合维修专业之间的关联程度。换句话说,维修专业之间只有相互影响、相互依存其关联程度才较为紧密。

4 实例

假设维修任务集合 $T = \{T_1, T_2, \dots, T_{20}\}$,涉及机械专业 S_1 、通信专业 S_2 、光学专业 S_3 、电子专业 S_4 、液压专业 S_5 、计算机专业 S_6 等多个维修专业,则令专业集合 $S = \{S_i \mid (i = 1, 2, \dots, 5)\}$ 。其中,每条维修任务可能由若干个维修专业修理,其相应维修时间如表 2 所示。

表 2 维修任务、维修专业以及维修时间表

| TID | S_1 | S_2 | S_3 | S_4 | S_5 | S_6 |
|----------|-------|-------|-------|-------|-------|-------|
| T_1 | 7 | 12 | 3 | 6 | 8 | - |
| T_2 | 6 | 5 | - | 6 | 9 | 11 |
| T_3 | 7 | 17 | 6 | 5 | 10 | 8 |
| T_4 | - | 9 | - | - | 11 | 4 |
| T_5 | - | - | 7 | 7 | 9 | 26 |
| T_6 | 5 | 5 | 7 | 9 | 14 | 11 |
| T_7 | 9 | 8 | - | - | 17 | 9 |
| T_8 | 7 | 12 | 6 | 4 | 10 | - |
| T_9 | - | 5 | 9 | - | 8 | - |
| T_{10} | 7 | 11 | - | 8 | 6 | 8 |
| T_{11} | 6 | 9 | 3 | 7 | 11 | 12 |
| T_{12} | 3 | - | 8 | 11 | 5 | 7 |
| T_{13} | - | 16 | - | - | 8 | 7 |
| T_{14} | 5 | 5 | 4 | 3 | 7 | - |
| T_{15} | - | 24 | - | - | 11 | 8 |
| T_{16} | 8 | 9 | 7 | - | 11 | 14.5 |
| T_{17} | 13 | - | - | - | 9 | 26 |
| T_{18} | - | 14 | 7 | 5 | 15 | 7 |
| T_{19} | - | 5 | 5 | 8 | 18 | 12 |
| T_{20} | 9 | 16 | 11 | 8 | 7 | 19 |

首先,根据式(1)对维修任务在各专业的维修时间进行标准化处理,得到基于任务量的维修专业关联度,如表 3 所示。

表 3 基于任务量的维修专业关联度

| TID | S_1 | S_2 | S_3 | S_4 | S_5 | S_6 | MSG |
|----------|----------|----------|----------|----------|----------|----------|----------|
| T_1 | 0.67 | 0.25 | 1 | 0.75 | 0.58 | - | 0.6 |
| T_2 | 0.91 | 1 | - | 0.91 | 0.6 | 0.45 | 0.9 |
| \vdots |
| MIS | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.8 |

然后,根据各维修专业在维修任务集中出现的频度,可以得到其相应的支持度以及满足最小支持度的频繁项集 L_1 ,如表 4 所示。

表 4 专业频繁项集 L_1 (正式)

| Itemset | Support | Itemset | Support |
|---------|---------|---------|---------|
| S_1 | 0.65 | S_4 | 0.6 |
| S_3 | 0.5 | S_5 | 0.5 |

最后,可以得到满足最小关联度和最小支持度的维修专业组合方案即频繁集,如表 5 所示。

表 5 关联规则输出表

| 关联规则 | 支持度 | 置信度 | 关联规则 | 支持度 | 置信度 |
|-----------------------|------|------|-----------------------|------|------|
| $S_4 \rightarrow S_1$ | 0.45 | 0.75 | $S_4 \rightarrow S_3$ | 0.45 | 0.75 |

其中,关联规则 $S_4 \rightarrow S_1$ 的含义是 45% 的维修任务同时需要光学专业 S_3 和电子专业 S_4 ,75% 的维修任务需要电子专业 S_4 的同时还会需要光学专业 S_3 ,关联规则 $S_4 \rightarrow S_3$ 也具有同样的意义。此外,根据表 2 中可以累计得到各维修专业的任务量,如表 6 所示。

表 6 维修专业任务量累计表 /h

| | 机械 专业 S_1 | 通信 专业 S_2 | 光学 专业 S_3 | 电子 专业 S_4 | 液压 专业 S_5 | 计算机 专业 S_6 |
|-----|----------------|----------------|----------------|----------------|----------------|-----------------|
| 任务量 | 92 | 182 | 83 | 87 | 204 | 189.5 |

由此可见,维修专业 S_4 和 S_3 组合后其任务量为 170;维修专业 S_4 和 S_1 组合后其任务量为 179,与其他维修专业的任务量基本相当。从分析结果来看,电子专业 S_4 既可以与机械专业 S_1 进行组合,也可以与光学专业 S_3 进行组合。显然,专业组合方案 $\{S_1, S_4\}$ 和 $\{S_3, S_4\}$ 在任务共担方面存在冲突,根据式 (2) 和 (3),其评估结果如表 7 所示。

表 7 基于支持度和置信度的评估结果

| | $S_4 \rightarrow S_1$ | $S_4 \rightarrow S_3$ |
|----------------|-----------------------|-----------------------|
| 皮尔森系数 α | 0.25 | 0.61 |
| 余弦函数 β | 0.72 | 0.82 |

可见,随着维修任务的增加,光学专业与电子专业共同承担任务的趋势或几率更大。由此可见,与机械专业 S_1 相比,光学专业 S_3 与电子专业 S_4 的组合是最优的。

(上接第 376 页)往往大于样本观测数目,通常数据中还存在较大的噪声。本文在中位数回归分析方法的基础上提出了一种新的有效的变量选择(或特征提取)方法,并具体给出了估计算法,该算法具有快速计算的特点。从实验结果可以看出,该方法能更加有效地进行变量选择(特征提取),且预测偏差较小,达到了良好的效果。当然,在实际问题中,数据的类型具有多样性、复杂性,噪声的来源也是多渠道的,采用一种方法并不能有效地解决所有的问题。例如,当高维数据中变量是属性数据或离散的计数数据,直接利用本文提出的方法进行分析,效果并不理想。关于这方面的问题还将作进一步研究。

参考文献:

[1] HASTIE T, TIBSHIRANT R, FRIEDMAN J. 统计学习基础:数据挖掘、推理与预测[M]. 范明,等译. 北京:电子工业出版社,2004.
 [2] BüHLMANN P, Van de GEER S. Statistics for high-dimensional da-

5 结束语

传统的数据统计方法难以发现维修专业隐藏在任务信息中潜在的关联关系,而基于数据挖掘的维修专业组合优化方法是切实可行的。通过引入维修专业关联度,利用改进的多最小支持度关联规则挖掘任务信息中维修专业潜在的规律,既实现了任务共担的维修专业有机组合,也均衡了各维修专业的任务量,提高了组合专业协同维修的效率和综合保障能力。

参考文献:

[1] 赵卫民,吴勋,孟宪君,等. 武器装备论证学[M]. 北京:兵器工业出版社,2008:31-32.
 [2] 毛国君,段立娟,王实,等. 数据挖掘原理与算法[M]. 北京:清华大学出版社,2007:67-68.
 [3] MABU S, SHIMADA K. Fuzzy inter transaction class association rule mining using genetic network programming for stock market prediction [J]. IEEE Trans on Electrical and Electronic Engineering, 2011, 6(4): 2697-2750.
 [4] ZHANG Xiao-hui, LIU Zheng-jiang. Analysis of multi-dimension association rule in marine casualties[C]//Proc of the 1st International Conference on Transportation Information and Safety. 2011: 2697-2705.
 [5] 刘晶,季海鹏,朱清香. 改进多重最小支持度关联规则算法在故障诊断中的应用[J]. 工业工程,2010,13(4):108-111.
 [6] ENGLE K M, RADA R. A top-k analysis using multilevel association rule mining for autism treatments[C]//Proc of the 6th International Conference on Universal Access in Human-Computer Interaction. Berlin: Springer-Verlag,2011:328-334.
 [7] 周永生,熊结青,沙宗尧. 基于关联规则挖掘的生化企业数据分析及其应用研究[J]. 计算机与应用化学,2010,27(9):1252-1256.
 [8] 叶义成,柯丽华,黄德育,等. 系统综合评价技术及其应用[M]. 北京:冶金工业出版社,2006:18-19.
 [9] 迈克伦南,唐朝晖,克里沃茨. 数据挖掘原理与应用[M]. 董艳,程文俊,译. 北京:清华大学出版社,2007.
 [10] 何朝阳,赵剑锋. 最大控制的多最小支持度关联规则及其挖掘方法研究[J]. 计算机工程,2006,32(11):103-105.
 [11] 徐建中,朱建新. 应用统计学[M]. 哈尔滨:哈尔滨工程大学出版社. 2001:283-284.
 [12] 孙文爽,陈兰祥. 多元统计分析[M]. 北京:高等教育出版社. 1999:341-343.

ta: methods, theory and applications[M]. Berlin: Springer-Verlag, 2011.
 [3] TIBSHIRANI R. Regression shrinkage and selection via the LASSO [J]. Journal of the Royal Statistical Society: Series B,1996,58(1): 267-288.
 [4] EFRON B, HASTIE T, JOHNSTONE I, et al. Least angle regression [J]. The Annals of Statistics,2004,32(2): 407-489.
 [5] FRIEDMAN J, HATTIE T, HOFLING T, et al. Pathwise coordinate optimization[J]. Annals of Applied Statistics,2007,1(2):302-332.
 [6] ZOU Hui. The adaptive LASSO and its oracle properties[J]. Journal of the American Statistical Association,2006,101(476): 1418-1429.
 [7] WANG Han-sheng, LI Guo-dong, JIANG Guo-hua. Robust regression shrinkage and consistent variable selection through the LAD-Lasso[J]. Journal of Business and Economic Statistics,2007,25(3): 347-355.