

# 地铁施工事故的多维关联规则分析

陈伟珂, 李金玲

(天津理工大学 管理学院, 天津 300384)

**摘要:** 针对地铁施工事故频发, 施工事故数据间的关系日益复杂的特点, 提出了多维关联规则在地铁施工事故数据的具体应用方法, 展现了“人—机—环境—管理”和事故发生类型之间如何挖掘强关联规则的过程。通过 30 个地铁施工事故数据的分析, 利用 Excel 软件对数据进行处理, 找出可能导致地铁施工事故发生的频繁因素的组合来, 发现并评论施工事故类型发生的潜在规律, 并将这些规律作为现实中管理者做出预防措施的依据。

**关键词:** 地铁施工; 施工事故; 多维关联规则

**中图分类号:** TU 9 **文献标识码:** A **文章编号:** 2095-0373(2010)04-0086-05

## 0 引言

城市人口的增加以及经济的快速发展, 积极推动了公共交通系统的建设, 地铁作为城市公共交通系统的重要组成部分, 正处在蓬勃发展的阶段。由于建设周期长、不可预见风险因素多等特点, 导致地铁工程建设过程中事故频发, 形势严峻, 令人堪忧<sup>[1]</sup>。据初步统计: 从 2001 年至 2009 年, 我国共发生 95 起地铁事故, 其中地铁施工事故 52 起占 50% 以上。随着城市公共交通建设地建设, 数据资料的数量和种类相应增加, 数据间关系日也益复杂。如何从大量而关系复杂的信息中将有价值的部分提取出来, 并根据提取的有效信息来组织相应对策是一个急需解决的问题<sup>[2]</sup>。针对某一地铁施工事故进行传统的定性分析已不适应大量而复杂数据的需要。利用关联规则的分析工具挖掘地铁施工事故数据的特点与规律, 找到地铁施工事故发生类型与“人—机—环境—管理”各因子之间的强关联规则。研究的结果为施工管理人员提供辅助决策支持, 对在全国的范围内有效减少由于不确定因素造成的地铁施工事故数量有一定的现实意义。

## 1 地铁施工事故的关联规则模型

地铁施工事故的因子涉及“人—机—环境—管理”等等, 一般情况下, 单独一种因子很难引发施工事故, 往往在多种因子的相互作用下导致重大的地铁施工事故。数据挖掘是一项可以从大量数据中获取有价值信息的技术分析工具, 关联规则是数据挖掘中的一项重要技术, 它可以做到分解地铁施工事故引发的因子, 并通过检验各因子组合引发事故的频繁谓词集得到有价值的强关联规则。

### 1.1 关联规则

Agrawal 在 1993 提出了挖掘顾客交易数据库中项集间的关联规则问题<sup>[3]</sup>。关联规则技术是能推导出“ $A \Rightarrow B$ ”的模式, 例如: 可以将  $A$  事件定义为“支护结构损坏”, 而  $B$  定义为“塌方”, 那么  $A \Rightarrow B$  可理解为“当地铁施工过程中支护结构发生损坏的情况下可能会出现塌方事故”。

设有事务数据库  $D = \{T_1, T_2, \dots, T_n\}$ ,  $T_j (j = 1, 2, \dots, n)$  称为事务  $T$ ; 构成  $T$  的元素  $i_k (k = 1, 2, \dots, p)$  被称为项; 设  $D$  中所有项的集合为  $I = \{i_1, i_2, \dots, i_m\}$ , 显然  $T \subseteq I$ 。

$A$  和  $B$  都是  $D$  的项集, 且  $A \cap B = \emptyset$  关联规则  $A \Rightarrow B$  由两个参数来约束, 支持度 (Support) 与可置信度 (Confidence)。支持度是同时包含项集  $A$  和项集  $B$  的事务在  $D$  中的所有事务中所占的百分比, 它表示规则

收稿日期: 2010-04-13

作者简介: 陈伟珂 女 1961 年出生 教授

基金项目: 天津市科技支撑计划项目 (102CKF501200)

出现的频度; 可信度就是同时包含项集  $A$  和项集  $B$  的事务在所有包含项集  $A$  的事务中所占的百分比, 它表示规则的强度。

如果  $A$  的支持度满足最小支持度阈值  $\text{min-Support}$ , 即  $\text{Support}(A) \geq \text{min-Support}$ , 则  $A$  称为  $D$  中的频繁项集。

如果存在关联规则  $A \Rightarrow B$ , 其支持度和可信度分别满足预设的最小支持度阈值 ( $\text{min-Support}$ ) 和最小可信度阈值 ( $\text{min-Confidence}$ ), 称其为强关联规则。

### 1.2 分析流程

地铁施工事故关联规则分析的具体流程如图 1 所示。

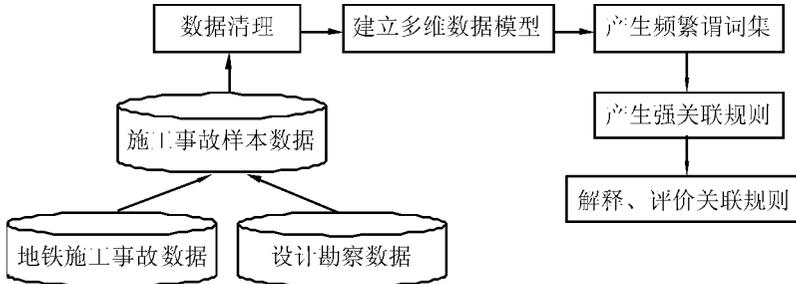


图 1 关联规则分析流程

(1) 收集与挖掘有关的历史灾害数据和设计勘察资料, 得到地铁施工事故相关的数据。

(2) 对地铁施工事故数据进行数据清理, 保证数据的准确规范。在地铁施工事故数据记录不完整, 不一致, 还有错误的信息等等, 因此, 为保证以后分析中数据的有效, 这一阶段需要对此类数据进行清理, 主要解决数据文件建立中的人为误差, 以及数据文件中一些对统计分析结果影响较大的特殊数值。

(3) 建立多维数据模型。根据地铁施工事故的特点 事故发生的原因, 设置数据属性。其中数据的属性如表 1 所示。

表 1 地铁施工事故的数据属性

| 地铁施工事故 ( $A$ ) | 时间 ( $T$ ) | 地点 ( $L$ ) | 事故情况 ( $D$ ) | 施工人员 ( $P$ ) | 施工器具 ( $I$ ) | 环境风险 ( $E$ ) | 管理 ( $M$ ) |
|----------------|------------|------------|--------------|--------------|--------------|--------------|------------|
| A001           |            |            |              |              |              |              |            |
| A002           |            |            | ...          |              |              |              |            |
| ...            |            |            |              |              |              |              |            |

(4) 产生频繁谓词集。通过扫描模型中发的数据, 利用改进的经典 Apriori<sup>[4]</sup> 算法产生频繁谓词集。

(5) 产生强关联规则。通过关联规则中的属性——最小可信度, 在已产生的频繁谓词集中生成强关联规则。

(6) 解释、评价关联规则。关联规则是否有用, 要根据地铁施工的实际情况, 解释得出的有意义的强关联规则并对其进行解释与评价。

### 1.3 样本数据

地铁施工事故数据<sup>[5-6]</sup> (根据文献及相关网络记录得出以下数据, 不能按照地铁施工相关标准而造成的日常安全事故不作为地铁施工事故数据), 如表 2 所示。

### 1.4 多维关联规则挖掘

关联规则可分为单维和多维的。单维的关联规则是处理单个属性中的关系; 多维关联规则是处理各个属性之间的某些关系。本文涉及到的多维关联规则挖掘就是从“人—机—环境—管理”等不同角度来考虑地铁施工事故发生的原因。

关联规则按处理变量的类型, 分为布尔型和数值型。布尔型关联规则处理的值多是离散的, 种类化的, 显示了变量之间的关系; 对数值型字段进行处理, 将其进行动态的分割, 或者直接对原始的数据处理。

多维关联规则挖掘过程可以概括为以下两个阶段: (1) 从事务数据库中找出所有的频繁谓词集; (2) 由频繁谓词集产生强关联规则, 即满足最小支持度与最小可信度。从无穷多的关联规则中发现有意义的关联规则, 为今后的决策提供依据。

表 2 地铁施工事故样本数据表

| 事故 (A) | 时间 (T)   | 地点 (L) | 事故情况 (D)  | 行为人 (P) | 施工器具 (I) | 环境风险 (E) | 管理 (M) |
|--------|----------|--------|-----------|---------|----------|----------|--------|
| A 001  | 2003. 7  | 上海     | 建筑物倾斜, 下沉 | 监理不当    | 设备故障     | 地层       | 开挖顺序错误 |
| A 002  | 2004. 4  | 广州     | 车站沉降      | 监测失误    | 因外力损坏    | 雨水过多     | 沟通不好   |
| A 003  | 2004. 8  | 广州     | 钻破煤气管道    | 操作失误    | 操作失误     | 沿线管道     | 信息不畅   |
| A 004  | 2005. 3  | 北京     | 路面塌陷      | 经验不足    | 操作失误     | 地层沉降     | 沟通不畅   |
| A 005  | 2005. 11 | 北京     | 基坑坍塌      | 操作失误    | 管线年久     | 形成水囊     | 管理失误   |
| A 006  | 2006. 1  | 广州     | 工地发生地陷    | 操作过失    | 设备操作不当   | 地面下沉     | 管理不当   |
| A 007  | 2006. 2  | 北京     | 吊斗坠落      | 监测失误    | 钢缆断裂     | 其它原因     | 管理不当   |
| A 008  | 2007. 5  | 南京     | 基坑土体滑坡    | 监测 人为过失 | 其它原因     | 雨水       | 管理漏洞   |
| A 009  | 2008. 4  | 深圳     | 模板坍塌      | 监测失误    | 安装不合格    | 沿线交通     | 管理失误   |
| A 010  | 2009. 1  | 上海     | 吊车侧翻      | 操作不当    | 设备故障     | 交通不畅     | 工人大意   |
| A 011  | 2003. 1  | 北京     | 地梁钢筋倾覆    | 操作失误    | 其它原因     | 其它原因     | 施工纪律不严 |
| A 012  | 2006. 1  | 北京     | 涌水致塌坍     | 未监测到渗漏  | 水管断裂     | 水囊       | 管理失误   |
| A 013  | 2007. 3  | 北京     | 塌方        | 现场监测失误  | 其它原因     | 塌陷       | 管理意识弱  |
| A 014  | 2004. 9  | 上海     | 楼房倾斜下沉    | 未监测渗水   | 其它原因     | 地下水文     | 方案不妥   |
| A 015  | 2009. 1  | 上海     | 火灾        | 线路短路    | 电器设备老化   | 其它原因     | 安全意识弱  |
| A 016  | 2004. 3  | 广州     | 围护结构塌方    | 操作失误    | 围护设备故障   | 环境复杂     | 信息沟通不畅 |
| A 017  | 2004. 9  | 广州     | 工地塌方      | 司机失误    | 管道断裂     | 管线涌水     | 管理不足   |
| A 018  | 2005. 11 | 广州     | 电缆坠落      | 监理过失    | 电缆断裂     | 环境复杂     | 施工管理不足 |
| A 019  | 2005. 11 | 广州     | 塌方        | 打桩失误    | 其它原因     | 溶洞地质     | 勘察不足   |
| A 020  | 2006. 4  | 广州     | 爆炸        | 操作失误    | 风管裂开     | 环境复杂     | 施工管理不足 |
| A 021  | 2008. 1  | 广州     | 桥路面下陷     | 现场监测失误  | 管道断裂     | 涌水       | 施工方案不足 |
| A 022  | 2005. 9  | 北京     | 龙门吊倾倒     | 现场监测失误  | 其它原因     | 地面塌陷     | 安全管理失误 |
| A 023  | 2005. 1  | 北京     | 渗水塌陷      | 未监测到渗漏  | 设备故障     | 地下水文     | 方案不妥   |
| A 024  | 2006. 1  | 北京     | 塌陷        | 施工不当    | 管道断裂     | 沿线设施     | 管理不足   |
| A 025  | 2006. 6  | 北京     | 塌陷        | 现场监测失误  | 其它原因     | 土体疏松     | 管理不足   |
| A 026  | 2006. 6  | 北京     | 竖井塌坍      | 监测失误    | 其它原因     | 地质       | 管理不足   |
| A 027  | 2007. 3  | 北京     | 出入口塌方     | 监测失误    | 其它原因     | 地质       | 管理不足   |
| A 028  | 2007. 5  | 北京     | 涌水淹没路口    | 未监测到渗漏  | 水管裂开     | 沿线设施     | 安全检查不足 |
| A 029  | 2007. 12 | 南京     | 道路塌方      | 监测失误    | 施工面断裂    | 地质       | 勘察不足   |
| A 030  | 2008. 11 | 杭州     | 塌坍        | 监测失误    | 其它原因     | 环境复杂     | 勘察不足   |

## 2 应用及评价

### 2.1 数据清理

在表 2 的样本数据中, 根据要找到“原因推出事故类型”的规则, 假定数据记录完整, 但为方便提出相关规则, 不需要具体的事故编号、时间值和地点, 因此去除事故、时间与地点数据列。同时, 将不完整和信息不足的劣数据剔除。

### 2.2 多维属性分类

通过数据分析, 将属性的原始值用区间值或较高层的概念替换, 即概念分层<sup>[7]</sup>。事故类型的分类及标号<sup>[8]</sup>为坍塌 =  $T_1$ , 对第三方的破坏 =  $T_2$  和器具类破坏三类 =  $T_3$ ; 行为人风险因素的分类及标号为监理过失 =  $P_1$ , 现场监测失误 =  $P_2$ , 操作失误 =  $P_3$ , 其它意外行为 =  $P_4$ ; 施工器具风险的分类及标号为设备故障 =  $I_1$ , 操作失误 =  $I_2$ , 设备质量 =  $I_3$ ; 环境风险的分类及标号为地层沉降 =  $E_1$ , 地下水或相邻江河水 =  $E_2$ , 雨水 =  $E_3$ , 沿线设施 =  $E_4$ , 地质 =  $E_5$ , 沿线交通 =  $E_6$ ; 管理因素的分类及标号为信息沟通失真 =  $M_1$ , 员工违纪 =  $M_2$ , 管理漏洞 =  $M_3$ 。将地铁施工数据各维度的分类见图 2 所示。

原始数据可清理为 28 个数据, 如表 3 所示。

表 3 地铁施工事故经清理后的数据表

| T | T <sub>2</sub> | T <sub>3</sub> | T <sub>2</sub> | T <sub>1</sub> | T <sub>2</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>3</sub> | T <sub>1</sub> | T <sub>2</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>3</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>3</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>2</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>1</sub> | T <sub>1</sub> |                |
|---|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| P | P <sub>1</sub> | P <sub>3</sub> | P <sub>2</sub> | P <sub>2</sub> | P <sub>3</sub> | P <sub>3</sub> | P <sub>1</sub> | P <sub>3</sub> | P <sub>1</sub> | P <sub>3</sub> | P <sub>2</sub> | P <sub>1</sub> | P <sub>2</sub> | P <sub>3</sub> | P <sub>2</sub> | P <sub>2</sub> | P <sub>1</sub> | P <sub>1</sub> | P <sub>2</sub> | P <sub>2</sub> | P <sub>2</sub> | P <sub>2</sub> | P <sub>3</sub> | P <sub>1</sub> | P <sub>1</sub> | P <sub>2</sub> | P <sub>2</sub> |                |
| I | I <sub>1</sub> | I <sub>1</sub> | I <sub>3</sub> | I <sub>1</sub> | I <sub>3</sub> | I <sub>3</sub> | I <sub>1</sub> | I <sub>2</sub> | I <sub>2</sub> | I <sub>3</sub> | I <sub>1</sub> | I <sub>3</sub> | I <sub>2</sub> | I <sub>1</sub> | I <sub>3</sub> | I <sub>3</sub> | I <sub>1</sub> | I <sub>1</sub> | I <sub>3</sub> | I <sub>3</sub> | I <sub>3</sub> | I <sub>3</sub> | I <sub>1</sub> | I <sub>3</sub> | I <sub>3</sub> | I <sub>2</sub> | I <sub>3</sub> |                |
| E | E <sub>5</sub> | E <sub>6</sub> | E <sub>2</sub> | E <sub>3</sub> | E <sub>4</sub> | E <sub>5</sub> | E <sub>6</sub> | E <sub>4</sub> | E <sub>6</sub> | E <sub>5</sub> | E <sub>4</sub> | E <sub>1</sub> | E <sub>4</sub> | E <sub>6</sub> | E <sub>6</sub> | E <sub>4</sub> | E <sub>4</sub> | E <sub>1</sub> | E <sub>6</sub> | E <sub>2</sub> | E <sub>5</sub> | E <sub>5</sub> | E <sub>5</sub> | E <sub>4</sub> | E <sub>3</sub> | E <sub>5</sub> | E <sub>6</sub> | E <sub>6</sub> |
| M | M <sub>3</sub> | M <sub>3</sub> | M <sub>3</sub> | M <sub>1</sub> | M <sub>1</sub> | M <sub>3</sub> | M <sub>1</sub> | M <sub>3</sub> | M <sub>3</sub> | M <sub>3</sub> | M <sub>3</sub> | M <sub>2</sub> | M <sub>3</sub> | M <sub>1</sub> | M <sub>2</sub> | M <sub>1</sub> | M <sub>3</sub> | M <sub>3</sub> | M <sub>3</sub> | M <sub>1</sub> | M <sub>1</sub> | M <sub>1</sub> | M <sub>1</sub> | M <sub>3</sub> |

### 2.3 频繁谓词集的产生

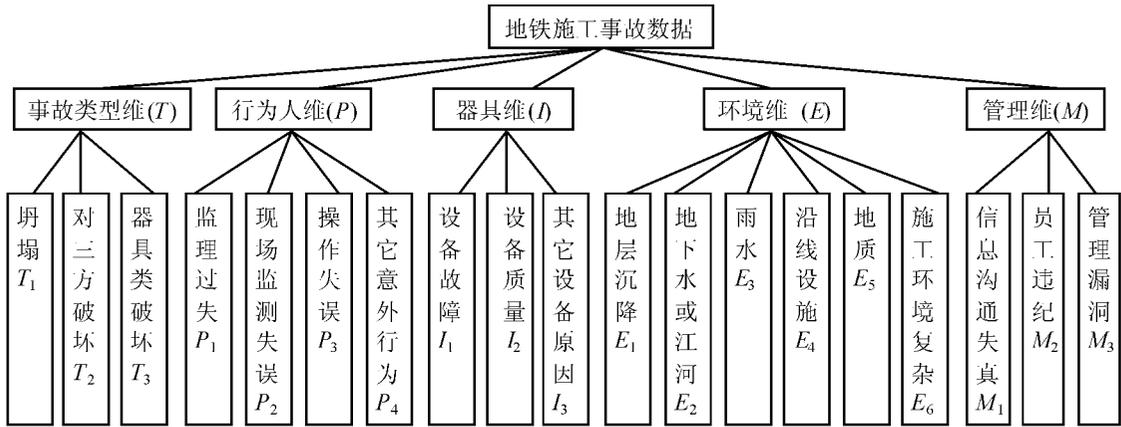


图 2 相关维度的设定

利用 Excel 软件, 根据文献 [4] 的算法来寻找频繁谓词集, 其中假设最小的支持计数为 2。那么样本的数据可表示为: 寻找频繁谓词集的过程如图 3 所示。由于篇幅所限, 只给出由四维到五维的寻找过程。

| 谓词集  | 支持度计数 |
|--|-------|
| {T <sub>1</sub> , P <sub>2</sub> , I <sub>3</sub> , E <sub>5</sub> } | 2     |
| {T <sub>1</sub> , P <sub>3</sub> , I <sub>3</sub> , E <sub>5</sub> } | 2     |
| {T <sub>1</sub> , P <sub>2</sub> , I <sub>1</sub> , E <sub>5</sub> } | 2     |
| {T <sub>1</sub> , P <sub>2</sub> , I <sub>3</sub> , E <sub>5</sub> } | 3     |
| {T <sub>1</sub> , P <sub>2</sub> , I <sub>3</sub> , E <sub>6</sub> } | 4     |
| {T <sub>1</sub> , P <sub>3</sub> , I <sub>3</sub> , E <sub>5</sub> } | 2     |
| {T <sub>1</sub> , P <sub>3</sub> , I <sub>3</sub> , E <sub>6</sub> } | 3     |
| {T <sub>1</sub> , P <sub>2</sub> , E <sub>4</sub> , E <sub>6</sub> } | 2     |
| {T <sub>1</sub> , P <sub>2</sub> , E <sub>5</sub> , E <sub>6</sub> } | 3     |
| {T <sub>1</sub> , P <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 2     |
| {T <sub>1</sub> , I <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 3     |
| {T <sub>2</sub> , P <sub>2</sub> , I <sub>1</sub> , E <sub>4</sub> } | 2     |
| {T <sub>2</sub> , P <sub>2</sub> , E <sub>4</sub> , E <sub>6</sub> } | 2     |
| {T <sub>2</sub> , P <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 2     |
| {T <sub>2</sub> , I <sub>1</sub> , E <sub>4</sub> , E <sub>6</sub> } | 2     |
| {P <sub>2</sub> , I <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 2     |
| {P <sub>2</sub> , I <sub>1</sub> , E <sub>4</sub> , E <sub>6</sub> } | 3     |
| {P <sub>2</sub> , I <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 3     |
| {P <sub>2</sub> , I <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 2     |

谓词的维数增加并扫描样本数据, 以最小支持数为限, 选择有效谓词集

以最小支持数为限, 选择有效谓词集

| 谓词集   | 支持度计数 |
|---|-------|
| {T <sub>1</sub> , P <sub>2</sub> , I <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 3     |
| {T <sub>1</sub> , P <sub>3</sub> , I <sub>3</sub> , E <sub>5</sub> , E <sub>6</sub> } | 2     |
| {T <sub>2</sub> , P <sub>2</sub> , I <sub>1</sub> , E <sub>4</sub> , E <sub>6</sub> } | 2     |

图 3 频繁谓词集的寻找过程

### 2.4 由频繁谓词集产生地铁事故的关联规则

根据样本数据找到的频繁谓词集为:

(1)  $S_1 = \{T_1, P_2, I_3, E_5, M_1\}$ , 其子集非空。从中可提取的关于施工事故类型的关联规则为  $P_2, I_3, E_5, M_1 \Rightarrow T_1$ ,  $Support(P_2, I_3, E_5, M_1 \Rightarrow T_1) = 11\%$ ,  $Confidence(P_2, I_3, E_5, M_1 \Rightarrow T_1) = 3/3 = 100\%$ 。

(2)  $S_2 = \{T_1, P_3, I_3, E_5, M_3\}$ , 其子集非空。从中可提取的关于施工事故类型的关联规则为  $P_3, I_3, E_5, M_3 \Rightarrow T_1$ ,  $Support(P_3, I_3, E_5, M_3 \Rightarrow T_1) = 7\%$ ,  $Confidence(P_3, I_3, E_5, M_3 \Rightarrow T_1) = 2/2 = 100\%$ 。

(3)  $S_3 = \{T_2, P_2, I_1, E_4, M_3\}$ , 其子集非空。从中可提取的关于施工事故类型的关联规则为  $P_2, I_1, E_4, M_3 \Rightarrow T_2$ ,  $Support(P_2, I_1, E_4, M_3 \Rightarrow T_2) = 7\%$ ,  $Confidence(P_2, I_1, E_4, M_3 \Rightarrow T_2) = 2/3 = 67\%$ 。

假设最小置信度阈值为 50%, 则所提取三个关联规则均为强关联规则。

### 2.5 对强关联规则的评价

对上述频繁谓词集进行评价:

(1) 强关联规则  $(P_2, I_3, E_5, M_1 \Rightarrow T_1)$  可表示为: {现场监测失误, 其它设备原因, 地质原因, 信息沟通不畅}  $\Rightarrow$  坍塌 ( $Support = 11\%$ ,  $Confidence = 100\%$ )。即当现场对地质灾害等监测失误同时各管理层面上的信息沟通不畅时, 那么地铁施工现场一般会发生坍塌事故。因此, 在地铁施工现场要加强实时信息的监测力度, 从而对地铁施工安全事故起到防范作用。

(2) 强关联规则 ( $P_3, I_3, E_5, M_3 \Rightarrow T_1$ ) 可表示为: {现场施工人员操作不当, 其它设备原因, 地质原因, 管理上漏洞  $\Rightarrow$  坍塌 (Support= 7%, Confidence= 100%) }。即当地铁施工现场出现管理上的漏洞, 例如, 无安全检查等等, 出现地质灾害时会发生坍塌的事故。针对这种类型的地铁施工事故, 管理层要制定相当的规章制度约束各方行为, 同时重视施工环境的调查评估, 积极应对异常情况, 从而避免或削弱施工事故的发生。

(3) 强关联规则 ( $P_3, I_3, E_4, M_3 \Rightarrow T_2$ ) 可表示为: {现场监测失误, 设备故障, 沿线设施复杂, 管理漏洞  $\Rightarrow$  对第三方的破坏 (Support= 7%, Confidence= 67%) }。即当地铁施工现场沿线设施, 管线铺设复杂, 勘察阶段方案不足时, 施工时对地下管线破坏后又出现监测失误, 那么地铁施工现场会对第三方的破坏行为, 例如楼房倾倒, 管线破裂等等事故。针对此类施工事故, 在今后的地铁施工过程中重视设计勘察阶段的调查评估, 提高施工技术水平, 在各方管理、操作及施工措施都严格执行, 从而遏制潜在的安全事故发生。

### 3 总结

由于地铁施工事故频发, 记录地铁施工事故的数据也在日益增多, 本文试图从数据挖掘的角度为地铁施工单位提供一种辅助决策的工具。在建立多维关联规则模型后, 利用相关的样本数据, 挖掘地铁施工事故中的规律, 选择地铁施工事故类型和“人一机一环境一管理”等维度来挖掘有意义的强关联规则。在地铁施工管理的实际应用中, 最小支持度与最小置信度等参数的设定影响数据中强关联规则的挖掘质量, 因此需要与地铁施工方面的相关专家或者有经验的人员共同确定。挖掘地铁施工过程中有价值的规律, 可以在一定程度上改变地铁施工中事故多发的现状, 起到预警的作用, 同时帮助决策都做出正确、有效的决策。

## 参 考 文 献

- [1] 钱七虎, 戎晓力. 中国地下工程安全风险管理的现状、问题及相关建议 [J]. 岩石力学与工程学报, 2008(4): 649-655.
- [2] 韩家炜, 堪博. 数据挖掘概念与技术 [M]. 北京: 机械工业出版社, 2007.
- [3] Agrawal R, Imielinski T, Swami A N. Mining association rules between sets of items in large databases [C] // Proc ACM SIGMOD Conference on Management of data. Hington [s n], 1993: 207-216.
- [4] 高学东, 王文贤, 武森. 基于数据立方体的多维关联规则的挖掘方法 [J]. 计算机工程, 2003, 8(14): 74-76.
- [5] 侯艳娟, 张顶立, 李鹏飞. 北京地铁施工安全事故分析及防治对策 [J]. 北京交通大学学报, 2009, 33(3): 52-59.
- [6] 周洁静. 基于 WBS-RBS 结构的地铁施工风险研究 [J]. 价值工程, 2009, 11: 76-80.
- [7] 熊馨, 王卫平, 叶跃祥. 基于概念分层的个性化推荐算法 [J]. 计算机应用, 2005, 25(5): 1006-1015.
- [8] 谭学清, 罗琳, 周洞汝. 数据立方体上的约束关联规则挖掘 [J]. 情报科学, 2007, 25(1): 129-160.

## Analysis of Multidimensional Association Rule in Subway Construction Accidents

Chen Weike Li Jinling

(School of Management, Tianjin University of Technology, Tianjin 300384, China)

**Abstract** Subway construction accidents have taken place from time to time recently, and the relationship among construction accident data is becoming increasingly complicate. This paper introduces a famous method called multidimensional association rule to help analyze a large amount of data in subway construction accidents. With this method, this procedure shows how to find out strong association rules between “person-instrument-environment management” and the types of accidents. Further, this research is able to figure out the potential rules in construction accidents through searching the combination of frequent factors that probably lead to subway construction accidents. Besides, with these rules discovered, the managers would be able prohibit the anticipated accidents in subway construction.

**Key words** subway construction accidents of construction multidimensional association rules