

文章编号:1673-0062(2016)01-0061-05

## 基于统计理论的人因事件关联分析

蒋建军<sup>1</sup>, 彭玉元<sup>2</sup>, 李敏<sup>3</sup>, 马聪敏<sup>4</sup>, 李一帆<sup>4</sup>, 章琪<sup>4</sup>

(1.南华大学经济管理学院人因研究所,湖南 衡阳 421001; 2.广州商学院 计算机系,广东 广州 510830;  
3.南华大学 网络中心,湖南 衡阳 421001;4.南华大学 经济管理学院,湖南 衡阳 421001)

**摘要:**近年来,随着人误事件不断上升,人们开始大力关注这类事故的发生.为有效减少人误事件,该文提出人误事件的关联规则分析方法,该方法以统计学权重关联规则为基础,计算人误事件中的支持度,置信度及关联程度.所提出的方法通过实验进行测试及应用举例,从结果中,得到核电厂蒸汽发生器传热管断裂人误事件之间的高失误率事件,高置信度及高关联程度事件,以此为根据,对这些事件进行改善,以减少人因事故的发生.

**关键词:**人误事件;权重关联规则;传热管断裂

**中图分类号:**TL364<sup>+</sup>.1 **文献标识码:**B

## Association Analysis of Human Factor Events Based on Statistics Theory

JIANG Jian-jun<sup>1</sup>, PENG Yu-yuan<sup>2</sup>, LI Ming<sup>3</sup>, MA Cong-min<sup>4</sup>,  
LI Yi-fan<sup>4</sup>, ZHANG Qi<sup>4</sup>

(1.Human Factors Institute, the College of Economic & Management, University of South China, Hengyang, Hunan 421001, China; 2.Computer Science and Engineering Department, Guangzhou College of Commerce, Guangzhou, Guangdong 510830, China; 3.Networks Center, University of South China, Hengyang, Hunan 421001, China; 4.School of Economics and Management, University of South China, Hengyang, Hunan 421001, China)

**Abstract:** Recently, with human failure events rising, people begin to pay much attention to it. To effectively decrease human failure events, this paper proposes the method of association rules analysis of human error events. The proposed method is based on a weight association rule based on statistics to calculate support, confidence and correlation degree between human error events. The proposed methods are tested by experiences and are ap-

收稿日期:2015-07-28

基金项目:湖南省自然科学基金资助项目(2016JJ6129;14JJ7046);湖南省教育厅基金资助项目(14C0974);南华大学  
科研支撑基金资助项目(2012XQD51)

作者简介:蒋建军(1977-),男,湖南衡阳人,南华大学经济管理学院讲师,博士.主要研究方向:人因工程,数据挖掘.

plied to an example. From results of experience, we can easily obtain higher error rate events, higher confidence and correlation degree events between human failure events in SGTR event of nuclear power plant. According to those experiment results, we will improve the human events in order to decrease human events.

**key words:** human error events; a weight association rule; SGTR

## 0 引言

目前,在各类事故中,60%~90%的事故都是由人本身的失误引起<sup>[1]</sup>,由人引起的事故已占主要地位,如1979年的三里岛及1986年的切尔诺贝利核电站事故等等<sup>[2]</sup>.面对这种状况,本文提出带权重的关联规则对人误事件(Human Error Event, HEE)进行预测,以能够有效预测在时刻 $t$ 人误事件之间的相关性,从而有效地防止人误事件.

关联规则已得到研究者的极大关注,这方面已存在许多研究成果,例如:关联因果结构挖掘<sup>[3]</sup>,无候选集挖掘<sup>[4]</sup>,多支持系统关联规则挖掘<sup>[5]</sup>,查询语言关联规则挖掘<sup>[6]</sup>,定量关联规则挖掘<sup>[7-8]</sup>,兴趣关联规则挖掘<sup>[9]</sup>,模糊集关联规则挖掘<sup>[10]</sup>,基于权重和文本挖掘技术<sup>[11]</sup>,行为挖掘规则<sup>[12]</sup>等等.然而,在关联规则的研究和应用中,把关联规则应用到人误事件的研究很少.

## 1 基于统计理论的人因事件关联规则

系统在运行过程中,当发生人因事件时,如果后续的操作能被正确执行,事件就能得到恢复.但如果发生人因事故后,下一操作或多个操作没正确执行,系统定会出现人因事故,从而造成不可预料的后果.本研究就是希望得到人因事件之间的关联程度,也就是说,当发生某一人因事故时,预测发生的下一个人因事的可能性,如果得到的关联性较高,就需在平时的模拟训练中多训练、多注意、多总结.人因事件的权重关联规程就是用来得到人因事件之间的联系程度,从而避免连续第二次,第三次等发生人因事件.

关联规则用于从大量数据中发现隐含的信息及规律,可以利用这些信息对事件进行分析,决策.设: $I = \{i_1, i_2, i_3, \dots, i_N\}$ ,为一系列数据集;设: $D$ 为一系列事务,每个事物 $T(T \in I)$ 由许多项集构成,每个事务用TID表示,因此可以得到等式(1)及等式(2):

$$D = \{t_1, t_2, \dots, t_n\} \quad (1)$$

$$I = \{i_1, i_2, i_3, \dots, i_N\} \quad (2)$$

这里, $t_i$ 表示每个人误事件(HEE)的一个事务, $I$ 为项集的一个数据集, $t_i \in I$ .

假设, $P(X)$ 为一系列人误事件中 $X$ 出现的概率,那么 $p(Y|X)$ 为在给定 $x$ 条件下,人误事件 $Y$ 出现的概率.可以得到带权重的人因事件规则 $X \rightarrow Y$ 的支持度,如下:

$$\text{Support}(X \rightarrow Y) = P(X \cup Y)W(H_i(t)) \quad (3)$$

如果 $x$ 和 $y$ 相互独立,根据概率相关知识,等式(3)可以变为:

$$\text{Support}(X \rightarrow Y) = [p(x) + p(y) - p(x)p(y)]W(H_i(t)) \quad (4)$$

这里, $W(H_i(t))$ 表示时刻 $t$ 人因的权重值; $P(x)$ 表示事件 $x$ 的失误率; $P(y)$ 表示事件 $y$ 的失误率.

条件关联规则 $X \rightarrow Y$ 可以看成是条件置信度,表示成 $\text{Confidence}(X \rightarrow Y)$ ,定义如下:

$$\text{Confidence}(X \rightarrow Y) = W(H_i(t)) \times \frac{\text{support}(x \rightarrow y) - \text{support}(x)\text{support}(y)}{\text{support}(x)[1 - \text{support}(y)]} \quad (5)$$

其中, $\text{support}(x) = p(x)$ , $\text{support}(y) = p(y)$ .

对于两个人误事件 $x$ 和 $y$ ,关联度表示成 $\text{corr}(x, y)$ :

$$\text{corr}(x, y) = w(H_i(t)) \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}} = w(H_i(t)) \times$$

$$\frac{s_{xy}}{n\sigma_x\sigma_y} \quad x \neq y \quad (6)$$

其中:

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n(\bar{x})^2 \quad (7)$$

$$s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n(\bar{y})^2 \quad (8)$$

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \quad (9)$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (10)$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (11)$$

$\bar{x}$ 表示所有属于第 $x$ 类事件的失误均值; $x_i$ 表

示第  $i$  个事件  $x$  的失误率;  $\bar{y}$  表示所有属于第  $y$  类事件的失误均值;  $y_i$  表示第  $i$  个事件  $y$  的失误率。

如果  $x = y$ , 那么  $corr(x, y) = 0$ . 关联性越强,  $corr(x, y)$  的值越大. 现对  $corr(x, y)$  进行如下定义: 如果  $corr(x, y) \geq 1$ , 那么事件  $x, y$  看成是强相关; 如果  $0.5 \leq corr(x, y) < 1$ , 认为事件  $x, y$  是一般性相关; 如果  $corr(x, y) < 0.5$ , 那么就认为事件  $x, y$  弱相关; 如果  $corr(x, y) = 0$ , 就认为事件  $x, y$  不相关。

从前面的式中不难看出, 权重失误率值  $W(H_i(t)), P(x), P(y), \bar{x}, \bar{y}, x_i, y_i$  实际上都表示某个事件的失误率或某类事件的平均失误率, 其计算方法如下<sup>[13]</sup>:

$$P(event\_error) = [e^{\lambda\Delta} - 1]e^{-\lambda\Delta} \quad (12)$$

已有研究表明<sup>[14]</sup>, 监视过程中的人因事故受七个因子影响. 式(12)表明失误分布为连续分布,

事实上, 操纵员在处理事故时为离散性的, 这时需要对处理事故的时间进行分段, 认为在某个很小的时间段是连续的, 时间段之间是离散的, 因此, 式(12)进一步变为: 式

$$P(event\_error) = \sum_{j=1}^7 [\prod_{k=1}^n (e^{\lambda_j \Delta_i} - 1) * e^{-\lambda_j \Delta_i}] \quad (13)$$

式(13)中,  $j$  表示人的影响因子;  $n$  为处理事故时间的分段数;  $\Delta = T/n$ ;  $\lambda_j$  为影响因子对应的参数值。

## 2 应用

以核电厂蒸汽传热管断裂(SGTR)事故为例进行分析. 该例的所有事故都是通过试验获得, 均来自秦山核电站 300MW 机组操纵员事故响应的测试数据<sup>[15]</sup>, 如表 1 所示。

表 1 蒸汽传热管断裂人因事故

Table 1 The human events about steam generator tube rupture

序号	蒸汽传热管断裂人因事故	时间均值( $\mu$ )	时间方差( $\sigma$ )
1	SGTR+失辅助给水一台发现	8.00	1.59
2	SGTR+失辅助给水一台估计泄露率	22.60	4.25
3	SGTR+失辅助给水一台降负荷	47	14.84
4	SGTR+失辅助给水一台辅助给水泵未启动	173.36	37.20
5	SGTR+失辅助给水一台蒸汽管隔离	214.62	61.67
6	SGTR+放射性报警失灵发现泄露	28.06	9.12
7	SGTR+放射性报警失灵估计泄露率	58.07	33.72
8	SGTR+放射性报警失灵降负荷/停堆	120.71	44.55
9	SGTR+放射性报警失灵判断 SGTR	334.00	239.19
10	SGTR+放射性报警失灵隔离 SG 时间	355.19	201.28

根据已有研究<sup>[15]</sup>,  $t = 20$ . 使用表 1 中的原始数据、利用式(3)、式(4)、式(5)、式(6)、式(7)、式(8)、式(9)、式(10)、式(11)及式(13)的计算方法, 试验平台为: Matlab6.0, Celeron(R) Dual-Core 1.80 GHz, RAM 1.00 G, Windows 7, 得到的试验结果见表 2, 表 3, 表 4, 表 5。

表 2 蒸汽传热管断裂人因事故中每个因子概率

Table 2 Each factor probability of steam generator tube rupture human events

序号	概率	序号	概率
1	0.011	6	0.056
2	0.004	7	0.052
3	0.07	8	0.047
4	0.044	9	0.088
5	0.05	10	0.004

从计算数据知, (1)表 2 中序号为 9 的人误事件概率最大, 即: SGTR+放射性报警失灵判断 SGTR 事故出现的可能性最大, 需要加强训练和注意; (2)表 3 中显示, 支持度最大的是  $supp(1 \rightarrow 3)$ , 最小的是  $supp(2 \rightarrow 10)$ . 说明同时出现事故序号为 1 和 3 占总事务的比例较高, 应加强事故序号 1 和 3 同时出现的防范措施; (3)置信度最大的为 0.484, 最少为 0.0046, 说明在出现事故序号为 1 的情况下出现事故 3 的次数与出现事故 1 的次数的比率较高, 应加强若出现事故序号 1 时, 及时防范事故 3 的出现; (4)关联度最大为  $corr(1 \rightarrow 2)$ , 最少为  $corr(7 \rightarrow 9)$ , 说明事故 1 与事故 2 紧密相关, 也就是说事故 2 的出现很大程度上是由事故 1 的出现引起的, 因此需加强对事故 1 的防范措施和模拟训练. 获得实验数据能对高失误率, 高置信度事件进行重要预防, 在人因事故中

更注意高相关度事件,以至减少核电站 SGTR 事件中人因事故的发生.

表 3  $Support(x \rightarrow y)$   
Table 3  $Support(x \rightarrow y)$

$Support(x \rightarrow y)$	1	2	3	4	5	6	7	8	9	10
1	0.0016	0.0011	0.0802	0.0040	0.0045	0.0049	0.0046	0.0043	0.0074	0.0011
2	0.0011	0.0005	0.0055	0.0032	0.0040	0.0044	0.0041	0.0038	0.0069	0.0005
3	0.0802	0.0055	0.0103	0.0083	0.0087	0.0091	0.0089	0.0085	0.0049	0.0055
4	0.004	0.0032	0.0083	0.0064	0.0068	0.0073	0.0070	0.0067	0.0096	0.0032
5	0.0045	0.0040	0.0087	0.0068	0.0073	0.0077	0.0075	0.0070	0.0687	0.0040
6	0.0049	0.0044	0.0091	0.0073	0.0077	0.0081	0.0079	0.0075	0.0100	0.0044
7	0.0046	0.0041	0.0089	0.0070	0.0075	0.0079	0.0076	0.0712	0.0101	0.0041
8	0.0043	0.0038	0.0085	0.0067	0.0070	0.0075	0.0712	0.0069	0.0094	0.0038
9	0.0074	0.0069	0.0113	0.0096	0.0687	0.0100	0.0101	0.0094	0.0126	0.0069
10	0.0011	0.0005	0.0055	0.0032	0.0040	0.0044	0.0041	0.0038	0.0069	0.0005

表 4  $Confidence(x \rightarrow y)$   
Table 4  $Confidence(x \rightarrow y)$

$Confidence(x \rightarrow y)$	1	2	3	4	5	6	7	8	9	10
1	0.010	0.050	0.484	0.025	0.028	0.030	0.029	0.010	0.048	0.007
2	0.020	0.009	0.105	0.059	0.075	0.083	0.077	0.071	0.135	0.009
3	0.860	0.006	0.006	0.006	0.0059	0.0059	0.0059	0.0059	0.0060	0.0056
4	0.0061	0.005	0.010	0.080	0.0082	0.0087	0.0085	0.0083	0.0107	0.0050
5	0.0060	0.006	0.0083	0.0072	0.0076	0.0078	0.0078	0.0073	0.1058	0.0057
6	0.0058	0.006	0.0074	0.0068	0.0069	0.0070	0.0070	0.0068	0.0074	0.0056
7	0.0059	0.005	0.0081	0.0084	0.0074	0.0076	0.0074	0.0067	0.0087	0.0056
8	0.0061	0.006	0.0089	0.0077	0.0078	0.0082	0.0075	0.0079	0.0092	0.0060
9	0.0055	0.006	0.0047	0.0051	0.0577	0.0046	0.0050	0.0047	0.0046	0.006
10	0.020	0.009	0.105	0.059	0.075	0.0083	0.077	0.0710	0.135	0.009

表 5  $Corr(x \rightarrow y)$   
Table 5  $Corr(x \rightarrow y)$

$Corr(x \rightarrow y)$	1	2	3	4	5	6	7	8	9	10
1	0	2.036	0.324	1.795	1.339	1.206	0.662	1.042	0.537	0.679
2	2.036	0	1.294	1.868	1.388	1.227	0.687	1.081	0.557	0.704
3	0.324	1.294	0	1.134	0.847	0.749	0.419	0.660	0.340	0.430
4	1.795	1.868	1.134	0	1.216	1.096	0.602	0.947	0.488	0.617
5	1.339	1.388	0.847	1.216	0	0.803	0.449	0.707	0.364	0.461
6	1.206	1.227	0.749	1.096	0.803	0	0.397	0.625	0.322	0.407
7	0.662	0.687	0.419	0.602	0.449	0.397	0	0.350	0.180	0.228
8	1.042	1.081	0.660	0.947	0.707	0.625	0.350	0	0.284	0.359
9	0.537	0.557	0.340	0.488	0.364	0.322	0.180	0.284	0	0.184
10	0.679	0.704	0.430	0.617	0.461	0.407	0.228	0.359	0.184	0

#### 参考文献:

- [1] 何旭洪,黄祥瑞.工业系统中人的可靠性分析:原理、方法与应用[M].北京:清华大学出版社,2007.
- [2] Carvalho P V R, Santos I L D, Vidal M C R. Safety implications of culture and cognitive issues in nuclear power plant operation[J]. Applied Ergonomics, 2006, 37(2): 211-223.
- [3] Zhang X Z, Hu Y, Xie K, et al. A causal feature algorithm for stock prediction modeling[J]. Neurocomputing, 2014, 142(22): 48-59.

- [4] Vo B, Hong T P, Le B. DBV-Miner: A Dynamic Bit-Vector approach for fast mining frequent closed itemsets [J]. *Expert System with Applications*, 2012, 39 ( 8 ): 7196-7206.
- [5] Tran A, Truong T, Le B. Simultaneous mining of frequent closed itemsets and their generators: Foundation and algorithm [J]. *Engineering Applications of Artificial Intelligence*, 2014, 36: 64-80.
- [6] Chandanan A K, Shukla M K. Removal of duplicate rules for association rule mining from multilevel dataset [J]. *Procedia Computer Science*, 2015, 45: 143-149.
- [7] Sahoo J, Das A K, Goswami A. An efficient approach for mining association rules from high utility itemsets [J]. *Expert System with Application*, 2015, 42( 13 ): 5754-5778.
- [8] Narveka M, Syed S F. An Optimized Algorithm for Association Rule Mining Using FP Tree [J]. *Procedia Computer Science*, 2015, 45: 101-110.
- [9] Chakrabarti S, Sarawagi S, Dom B. Mining surprising patterns using temporal description length [C] // Proc. of the 24th Intl. Conf. on Very Large Data Bases, New York, USA. 1998: 606-617.
- [10] Chen C L, Tseng F S C, Liang T. Mining fuzzy frequent itemsets for hierarchical document clustering [J]. *Information Processing & Management*, 2010, 46 ( 2 ): 193-211.
- [11] Huang C J, Liao J J, Yang D X, et al. Realization of a news dissemination agent based on weighted association rules and text mining techniques [J]. *Expert Systems with Applications*, 2010, 37( 9 ): 6409-6413.
- [12] He Z Y, Xu X F, Deng S C, et al. Mining action rules from scratch [J]. *Expert Systems with Applications*, 2005, 29( 3 ): 691-699.
- [13] Jiang J J, Zhang Li, Wang Y Q, et al. Markov reliability model research of monitoring process in digital main control room of nuclear power plant [J]. *Safety Science*, 2011, 49( 6 ): 843-851.
- [14] 蒋建军, 张力, 王以群, 等. 考虑人因的核电厂主控室认知可靠性模型研究 [J]. *核动力工程*, 2012, 33( 1 ): 66-73.
- [15] 张力. 概率安全评价中人因可靠性分析技术 [M]. 北京: 原子能出版社, 2006.