

应用数据挖掘关联技术研究温湿度 对冲击谱试验的影响

冯雪梅, 卢来洁, 马爱军, 刘洪英
(中国航天员科研训练中心, 北京 100094)

摘要: 冲击谱试验随机性强, 难度较大, 必须了解环境温度、湿度对试验的影响。通过对数据预处理并建立冲击谱试验评价模型, 应用数据挖掘中关联分析方法提取数据背后隐藏的有用知识, 获得了温度、湿度对冲击谱试验各个调试因素的影响, 并分析了其间存在的某种规律性, 找出了适用于冲击谱试验的环境温度、湿度范围。研究结果可为今后的产品冲击环境试验提供有力的技术支持, 进一步提高冲击谱试验的试验质量和试验中的产品安全性。

关键词: 冲击谱试验; 温度; 湿度; 数据挖掘; 关联分析

中图分类号: V416.2; V416.5 **文献标识码:** A **文章编号:** 1673-1379(2007)02-0109-04

1 引言

冲击谱试验主要是模拟爆炸冲击环境的力学试验, 其试验环境复杂, 影响因素较多, 随机性强, 试验难度较大。随着环境试验数据量的增加, 有必要用数据库管理系统来存储数据, 用机器学习的方法来分析数据, 充分挖掘利用已获得试验数据中隐含的有用信息, 以求完善试验方法, 提高试验质量。数据挖掘就是从大量不完全的、有噪声的、模糊且随机的实际应用数据中提取隐含在其中的预先未知但又是潜在有用的信息和知识的过程^[1]。本文旨在将试验人员的经验和知识结合到数据挖掘过程中, 对冲击谱试验机的试验数据进行整理, 并对冲击数据进行预处理, 构建冲击谱试验机数据仓库, 以便在数据挖掘过程中迅速获得完整、正确的数据。利用数据挖掘工具 (SPSS Clementine) 对数据仓库内容进行关联分析, 研究冲击谱试验中环境温度、湿度对试验的影响, 找出适用于冲击谱试验的环境温度、湿度范围。

2 数据预处理

高质量的决策依赖于高质量的数据, 因此数据预处理是数据挖掘过程中的重要步骤。在构建冲击谱试验机数据仓库过程中使用的数据是经过

预处理 (即去掉坏的数据, 消除数据之间的矛盾, 集成来自不同数据源的数据, 进行数据转换等) 的历史数据, 所以提高了所获规则的可靠性^[2]。

对冲击谱试验数据源做必要的整理, 包括以下内容:

(1) 数据获取: 整理记录冲击谱试验机的历史试验及调试数据;

(2) 数据取样: 按冲击谱试验的不同试验规范、产品质量分类取样;

(3) 数据筛选: 滤除不希望要的试验数据, 去除数据中的噪音并纠正其不一致。当属性出现缺少值、奇异值或数据存在不一致时, 采用数据分箱、人机交互检查以及回归等方法进行检测;

(4) 数据转换: 将数据转换成适合于挖掘的形式。将冲击谱试验机试验数据中的时域信息转换为冲击响应谱信息以及与谱规范相比得出的误差信息, 并根据各频率点的误差值建立冲击谱试验评价模型, 然后将转换后的信息作为新的变量存放在样本数据中; 对数据作规范化、平滑和泛化处理, 进而提高其后挖掘过程的精度和速度。

3 建立冲击谱试验评价模型

冲击谱是指单自由度系统在冲击作用下的最

收稿日期: 2006-11-07

作者简介: 冯雪梅 (1974-), 女, 工程师, 主要从事计算机测控技术研究。联系电话: (010)66365643-801; E-mail: fengxuemei507@yahoo.com.cn

大响应同系统固有频率之间的关系。冲击谱是频率的函数，反映了冲击激励在频域中的信息^[3]。在冲击谱试验机数据挖掘方法研究中需建立冲击谱试验评价模型，用以评价冲击试验的成功度。

针对冲击谱分析的特点，在建立冲击谱试验评价模型过程中对冲击谱试验曲线（图 1）的每一点根据其所处频带跨度设定不同的加权值，进行逐步逻辑回归，获得评价参数。第 i 点的误差值 ΔE_i 定义为

$$\Delta E_i = 20 \log \left(\frac{g_{i,sc}}{g_{i,ck}} \right), \quad (1)$$

式中： $g_{i,sc}$ 为实测的加速度值， g ； $g_{i,ck}$ 为参考的加速度值， g 。

冲击谱评价参数 ΔE_{SRS} 定义为：

$$\Delta E_{SRS} = \frac{\sum_{i=1}^n (|\Delta E_i - \Delta E_{i,rc}| \cdot \Delta f_{(i-1,i)})}{f_n - f_1}, \quad (2)$$

式中： $\Delta E_{i,rc}$ 为冲击谱第 i 点的容差值，dB； $\Delta f_{(i-1,i)}$ 为相邻两个频率段之间的差值，Hz； f_n 为分析频率的最高值，Hz； f_1 为分析频率的最低值，Hz。

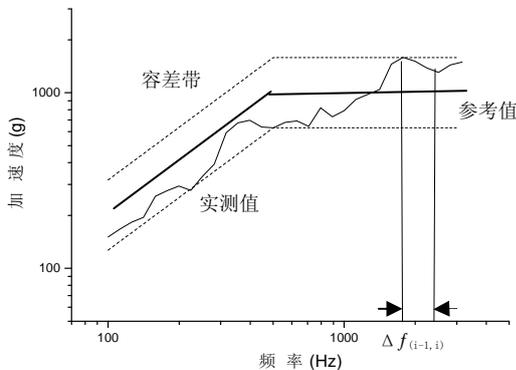


图 1 冲击谱试验曲线

Fig.1 Shock spectrum test curve

4 关联分析

4.1 分析依据

两个或两个以上变量的取值之间存在某种规律性称作关联。数据一般都存在关联关系，即一个含有 n 个元组的关系 R ，其属性 A_1, A_2, \dots, A_m ，设属性集 $I = \{A_{i1}, A_{i2}, \dots, A_{ik}\}$ 和 $J = \{A_{j1}, A_{j2}, \dots, A_{jk}\}$ 为两个属性值的集合。如果满足 $\text{support}(I \Rightarrow J) = P(I \cup J) > \text{min_support}$ ，且

$\text{confidence}(I \Rightarrow J) = P(J | I) > \text{min_confidence}$ ，则称 $I \Rightarrow J$ 是一个强关联规则。

本文利用数据挖掘中关联分析 Apriori 算法查找并描述环境温湿度与试验中各个调试参数（谐振板厚度、谐振板上下阻尼材料参数、标尺高度等）之间存在的某种规律性。Apriori^[4]是一种广度优先算法，它利用了频繁项集的向下封闭性（即频繁项集的非空子集仍是频繁项集），通过对数据库的多次扫描来发现所有的频繁项目集。在第 1 次扫描中，Apriori 算法计算每条记录中所有单个项目的支持度，生成所有长度为 1 的频繁项目集。在后续的第 k 次扫描中，首先以前一次中所发现的所有频繁项目集为基础，生成所有新的候选项目集（即潜在的频繁项目集）；然后扫描数据库，计算这些候选项目集的支持度，最后确定候选项目集中哪一些成为频繁项目集；重复上述过程直至不再发现新的频繁项目集。

根据冲击谱试验数据的具体情况，在进行关联分析中设定最小支持度为 1%，最小置信度为 60%。

4.2 分析结果

(1) 温湿度与谐振板的关系

不满足最小支持度和最小置信度的关系。

(2) 温湿度与板下材料的关系

不满足最小支持度和最小置信度的关系。

(3) 温湿度与板上材料的关系

得到如图 2 所示 19 条满足最小支持度和最小置信度的关系。

(4) 温湿度与标尺高度的关系

得到如图 3 所示 24 条满足最小支持度和最小置信度的关系。

从以上关联分析中可以看出：环境温湿度对冲击谱试验中谐振板厚度和谐振板下阻尼材料的选择无关，与谐振板上阻尼材料和标尺高度有关。因此在对冲击谱数据进行分类预测建模时，对于谐振板厚度和谐振板下阻尼材料的分类建模可不采用环境温湿度作为输入变量；而对于谐振板上阻尼材料和标尺高度分类建模时，则应该将环境温湿度作为输入变量进行分析。

Instances	Support	Confidence	Lift	Consequent	Antecedent 1	Antecedent 2
1	0.000	100.000	4.222	板上材料 = 4172(5)	湿度加权 = 10	温度加权 = 17
9	0.400	100.000	8.170	板上材料 = 4172(3)	温度加权 = 29	湿度加权 = 50
12	0.600	100.000	2.208	板上材料 = 4172(5宽)	湿度加权 = 29	湿度加权 = 40
9	0.400	100.000	2.208	板上材料 = 4172(5宽)	湿度加权 = 55	温度加权 = 26
16	0.800	100.000	4.222	板上材料 = 4172(5)	湿度加权 = 25	温度加权 = 23
86	4.300	96.500	2.131	板上材料 = 4172(5宽)	湿度加权 = 60	
86	4.300	96.500	2.131	板上材料 = 4172(5宽)	湿度加权 = 60	温度加权 = 26
50	2.500	86.000	3.631	板上材料 = 4172(5)	温度加权 = 17	湿度加权 = 30
12	0.600	83.300	3.518	板上材料 = 4172(5)	温度加权 = 29	湿度加权 = 30
45	2.200	73.300	1.619	板上材料 = 4172(5宽)	湿度加权 = 45	温度加权 = 26
21	1.000	71.400	3.016	板上材料 = 4172(5)	湿度加权 = 10	
84	4.200	70.200	2.965	板上材料 = 4172(5)	温度加权 = 17	
20	1.000	70.000	2.955	板上材料 = 4172(5)	湿度加权 = 10	温度加权 = 20
43	2.100	62.800	5.130	板上材料 = 4172(3)	湿度加权 = 50	温度加权 = 23
16	0.800	62.500	2.639	板上材料 = 4172(5)	湿度加权 = 45	温度加权 = 20
42	2.100	61.900	1.367	板上材料 = 4172(5宽)	湿度加权 = 35	温度加权 = 23
15	0.700	60.000	1.325	板上材料 = 4172(5宽)	湿度加权 = 55	温度加权 = 23
30	1.500	60.000	3.246	板上材料 = 4172(2)	湿度加权 = 55	温度加权 = 20
10	0.500	60.000	1.325	板上材料 = 4172(5宽)	湿度加权 = 50	温度加权 = 20

图 2 温湿度与板上材料的关系

Fig.2 The relationship between temperature, humidity and damping materials of the resonant panel

Instances	Support	Confidence	Lift	Consequent	Antecedent 1	Antecedent 2
1	0.000	100.000	1.888	标尺 = 17.5	湿度加权 = 10	温度加权 = 17
9	0.400	100.000	1.888	标尺 = 17.5	温度加权 = 29	湿度加权 = 50
12	0.600	100.000	4.455	标尺 = 19	湿度加权 = 29	湿度加权 = 40
16	0.800	100.000	1.888	标尺 = 17.5	湿度加权 = 25	温度加权 = 23
14	0.700	100.000	15.288	标尺 = 14	湿度加权 = 40	温度加权 = 20
86	4.300	90.700	5.028	标尺 = 15.5	湿度加权 = 60	
86	4.300	90.700	5.028	标尺 = 15.5	湿度加权 = 60	温度加权 = 26
16	0.800	87.500	1.652	标尺 = 17.5	湿度加权 = 45	温度加权 = 20
21	1.000	85.700	1.618	标尺 = 17.5	湿度加权 = 10	
20	1.000	85.000	1.605	标尺 = 17.5	湿度加权 = 10	温度加权 = 20
15	0.700	73.300	1.384	标尺 = 17.5	湿度加权 = 55	温度加权 = 23
75	3.700	72.000	1.359	标尺 = 17.5	湿度加权 = 25	
42	2.100	71.400	1.348	标尺 = 17.5	湿度加权 = 35	温度加权 = 23
30	1.500	70.000	3.118	标尺 = 19	湿度加权 = 55	温度加权 = 20
9	0.400	66.700	10.192	标尺 = 14	湿度加权 = 55	温度加权 = 26
113	5.600	65.500	1.236	标尺 = 17.5	温度加权 = 23	湿度加权 = 30
43	2.100	65.100	1.229	标尺 = 17.5	湿度加权 = 50	温度加权 = 23
59	2.900	64.400	1.216	标尺 = 17.5	湿度加权 = 25	温度加权 = 20
33	1.600	63.600	1.201	标尺 = 17.5	温度加权 = 17	湿度加权 = 50
569	28.200	63.600	1.201	标尺 = 17.5	温度加权 = 20	湿度加权 = 30
84	4.200	63.100	1.191	标尺 = 17.5	温度加权 = 17	
50	2.500	62.000	1.170	标尺 = 17.5	温度加权 = 17	湿度加权 = 30
834	41.300	61.500	1.161	标尺 = 17.5	湿度加权 = 30	
733	36.300	60.700	1.146	标尺 = 17.5	湿度加权 = 20	

图 3 温湿度与标尺高度的关系

Fig.3 The relationship between temperature, humidity and the height of staff gauge

5 寻找温湿度设定的合适范围

通过关联分析发现环境温湿度对冲击谱试验中某些调试因素产生影响。GJB 150《军用设备环境试验方法》规定环境温度范围为 15~35℃, 环境湿度范围为 20%~80%RH, 这一要求为试验通用条件, 具体试验时需找出满足国军标且最有利于冲击谱试验的环境温湿度范围, 以提高试验质量。

通过建立冲击谱试验评价模型, 可以对每次试验结果进行量化评价, 并对各试验评价结果按极好、好、一般和不好 4 个层次进行分级。为考

察环境温湿度对试验质量的影响, 进行关联分析中设定最小支持度为 1%, 最小置信度为 60%。综合考虑支持度和置信度的因素, 即按支持度与置信度的乘积排序, 得到图 4 所示 29 条满足最小支持度和最小置信度的关系。

使用网状图可以更清晰地表明环境温湿度与试验评价结果之间的强关联关系(见图 5), 图中较强的关系用较粗的线表示。图 4 和图 5 表明适于进行冲击谱试验的温度范围为 20~26℃, 湿度范围为 30%~40%RH。

Instances	Support	Confidence	Lift	Consequent	Antecedent 1	Antecedent 2
411	23.700	60.100	1.017	SRS = 好	湿度加权 = 40	
261	15.000	62.500	1.057	SRS = 好	温度加权 = 23	
198	11.400	74.200	1.257	SRS = 好	湿度加权 = 50	
110	6.300	78.200	1.324	SRS = 好	湿度加权 = 50	温度加权 = 26
84	4.800	88.100	2.152	SRS = 极好	湿度加权 = 60	
84	4.800	88.100	2.152	SRS = 极好	湿度加权 = 60	温度加权 = 26
78	4.500	66.700	1.129	SRS = 好	温度加权 = 17	
76	4.400	67.100	1.136	SRS = 好	湿度加权 = 23	湿度加权 = 30
73	4.200	63.000	1.067	SRS = 好	温度加权 = 23	湿度加权 = 40
55	3.200	70.900	1.200	SRS = 好	湿度加权 = 26	湿度加权 = 30
52	3.000	75.000	1.270	SRS = 好	湿度加权 = 25	
42	2.400	71.400	1.209	SRS = 好	湿度加权 = 25	温度加权 = 20
41	2.400	68.300	1.668	SRS = 极好	湿度加权 = 45	温度加权 = 26
39	2.200	66.700	1.129	SRS = 好	湿度加权 = 35	温度加权 = 23
40	2.300	60.000	1.016	SRS = 好	湿度加权 = 50	温度加权 = 23
29	1.700	79.300	1.343	SRS = 好	温度加权 = 17	湿度加权 = 50
30	1.700	70.000	1.185	SRS = 好	湿度加权 = 29	
16	0.900	87.500	1.481	SRS = 好	湿度加权 = 55	温度加权 = 20
15	0.900	80.000	1.954	SRS = 极好	湿度加权 = 55	温度加权 = 23
17	1.000	70.600	1.195	SRS = 好	湿度加权 = 10	
16	0.900	68.800	1.164	SRS = 好	湿度加权 = 10	温度加权 = 20
10	0.600	100.000	1.693	SRS = 好	湿度加权 = 50	温度加权 = 20
10	0.600	90.000	1.524	SRS = 好	湿度加权 = 25	温度加权 = 23
10	0.600	90.000	1.524	SRS = 好	湿度加权 = 45	温度加权 = 20
12	0.700	75.000	1.270	SRS = 好	温度加权 = 29	湿度加权 = 40
12	0.700	66.700	1.629	SRS = 极好	湿度加权 = 35	温度加权 = 20
9	0.500	88.900	1.505	SRS = 好	温度加权 = 29	湿度加权 = 30
8	0.500	75.000	1.270	SRS = 好	湿度加权 = 55	温度加权 = 26
1	0.100	100.000	1.693	SRS = 好	湿度加权 = 10	温度加权 = 17

图4 温湿度与试验评价结果的关系
Fig.4 The relationship between temperature, humidity and the test evaluation results

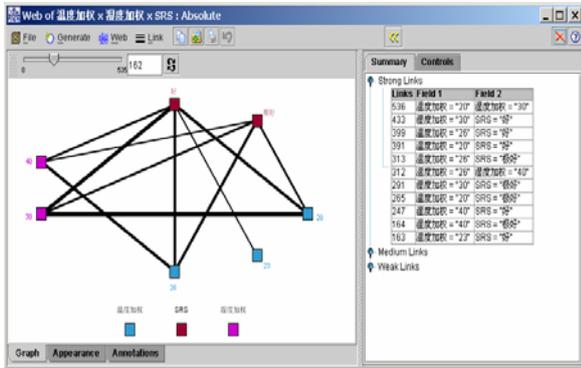


图5 温湿度与试验评价结果关系的网状图
Fig.5 The reticular chart of the test evaluation results with respect to temperature and humidity

6 结束语

通过对冲击谱试验数据的关联分析, 获得环境温湿度对试验调试各因素的不同影响, 为冲击谱数据进行数据挖掘分类预测分析提供输入依据, 并找出适用于冲击谱试验的环境温湿度范围, 为以后产品的冲击环境试验提供有力的技术支持, 可进一步提高冲击谱试验的试验质量和冲击谱试验中产品的安全性。

参考文献 (References)

- [1] Han Jiawei, Kamber M. Data mining: concepts and techniques[M]. USA: Morgan Kaufmann Publishers, 2001. 70-181
- [2] 刘莉, 徐玉生, 马志新. 数据挖掘中数据预处理技术综述[J]. 甘肃科学学报, 2003, 15(1): 117-119[Liu Li, Xu Yusheng, Ma Zhixin. A survey to data preprocessing in data mining[J]. Journal of Gansu Sciences, 2003, 15(1):117-119]
- [3] 马爱军, 冯雪梅, 卢来洁, 等. 一种冲击谱试验机的研制[J]. 航天医学与医学工程, 2003, 16(4): 296-299[Ma Aijun, Feng Xuemei, Lu Laijie, et al. Development of shock spectrum test facility[J]. Space Medicine & Medical Engineering, 2003, 16(4): 296-299]
- [4] 邹志文, 朱金伟. 数据挖掘算法研究与综述[J]. 计算机工程与设计, 2005, 26(9): 2304-2307[Zou Zhiwen, Zhu Jinwei. Research and summary of data mining algorithms[J]. Computer Engineering and Design, 2005, 26(9): 2304-2307]