

城镇重特大火灾财产损失风险关联性分析

傅智敏¹, 陈云国^{1,2}

1. 中国人民武装警察部队学院消防工程系, 河北廊坊 065000
2. 湖北省公安消防总队随州支队, 湖北随州 432700

[摘要] 运用数据挖掘技术和工具, 对 2000–2005 年全国 1 356 起城镇重特大火灾直接财产损失风险与火灾发生的时间、场所和原因之间的关联性进行了分析。分析结果表明, 厂房电气火灾、11 月份库房火灾、凌晨 02:00–03:00 时段的商场商店火灾、深夜 23:00–24:00 时段的电气火灾、7 月份的电气火灾、7 月份下午 18:00–19:00 时段发生的火灾, 以及 7 月份发生的厂房电气火灾和夜晚 21:00–22:00 时段发生的商场商店电气火灾容易形成高风险直接财产损失火灾。该结果可为有针对性、有重点地采取消防监督管理措施和提高消防执勤战备的有效性提供指导。

[关键词] 城镇重特大火灾; 直接财产损失; 关联规则; 数据挖掘工具

[中图分类号] X932

[文献标识码] A

[文章编号] 1000–7857(2007)15–0035–05

Risk Analysis on Direct Property Damage for Major Fire Incidents in Towns Based on Association Rules

FU Zhimin¹, CHEN Yunguo^{1,2}

1. Department of Fire Protection Engineering, Chinese People's Armed Police Forces Academy, Langfang 065000, Hebei Province, China
2. Suizhou Municipal Fire Protection Department, Suizhou 430070, Hubei Province, China

Abstract: Risk analysis on direct property damage for 1 356 major fire incidents in towns from 2000 to 2005 is carried out based on association rules analysis with respect to time, place and cause by SPSS Clementine 8.1 software. The results show that a fire may easily turn into a major fire incident with direct property damage, in cases of electric fire in workshop, storeroom fire in November, shop fire from 2:00 to 3:00, electric fire from 23:00 to 24:00, electric fire and fire from 18:00 to 19:00 in July, workshop fire due to electric failure in July and shop fire due to electric failure from 21:00 to 22:00. These results may serve as a guidance for taking more effective fire protection measures and improving the efficiency of combat readiness for fire fighting.

Key Words: major fire in town; direct property damage; association rules; data mining tool

CLC Number: X932

Document Code: A

Article ID: 1000–7857(2007)15–0035–05

0 引言

1993 年, IBM Almaden Research Center 的 Rakech Agrawal 等首先提出关联规则挖掘的概念, 关联规则挖掘的目的是发现大量数据中项集之间的相关联系^[1–2]。目前, 关联规则挖掘在保险、商业、电信、电力等行业已得到成功的应用^[3–5], 而在火灾科学领域中的应用还处

于起步阶段。对于城镇火灾这样一个复杂的大系统, 火灾与其起火原因、场所、时间之间有着复杂的相互关系^[6], 这种关系在大量的历史数据记录中表现为火灾报告中各指标数据项之间的定性的关联关系, 将这种关系量化“挖掘”出来就是反映城镇火灾实际情况的定量关联规则。因此, 采用关联规则挖掘方法能够找出

收稿日期: 2007–06–15

作者简介: 傅智敏, 河北省廊坊市中国人民武装警察部队学院消防工程系, 教授, 研究方向为消防工程; E-mail: fuzmin2002@yahoo.com.cn

反映在火灾统计表上不同指标项目之间的规律,为火灾安全管理提供决策情报。

1 表征关联规则的参数

通过关联分析法发现的知识称为关联知识。关联知识就是描述这种在一个事务中,事件之间同时出现的规律的模式。更确切地说,关联规则通过量化的数字描述事件甲的出现对事件乙的出现的程度。它反映了一个事件与其他事件之间依赖或关联的知识。一般用 4 个参数描述 1 个关联规则的属性,其含义见表 1^[7-10]。

表 1 关联规则 4 个参数的含义
Table 1 Definitions of the four parameters in association rules

名称	描述
支持度	项集 A 和项集 B 同时出现的概率
可信度	在项集 A 出现的前提下,项集 B 出现的概率
期望可信度	项集 B 出现的概率
作用度	可信度与期望可信度的比值

支持度是对关联规则重要性的衡量,可信度是对关联规则准确度的衡量。支持度说明了这条规则在所有事务中有多大的代表性,显然支持度越大,关联规则越重要。有些关联规则的可信度虽然很高,但是支持度却很低,说明该关联规则实用的机会很小,因此也不重要。期望可信度描述了在没有项集 A 的作用下,项集 B 本身的支持度;作用度描述了项集 A 对项集 B 的影响力大小。作用度越大,说明项集 B 受项集 A 的影响越大。一般情况下,有用关联规则的作用度应该大于 1,只有关联规则的可信度大于期望可信度,才说明项集 A 的出现对项集 B 的出现有促进作用,也说明了它们之间某种程度的相关性;否则,获得的关联规则也就没有意义了。在关联规则的 4 个属性中,支持度和最小可信度能够比较直接地描述关联规则的性质。从关联规则的定义可以看出,任意给出事务集中的两个项集,它们之间都存在关联规则,只是其属性值有所不同。如果不考虑关联规则的支持度和可信度,在事务库中将存在无穷多的关联规则。事实上,人们一般只对满足一定的支持度和可信度的关联规则感兴趣。本文中,预先设定最小支持度和最小可信度这两个参数分别为 0.01% 和 0.01%。

2 数据来源与处理

本文的数据分析样本为 2000-2005 年我国城镇 1 356 起重特大火灾,来源为《中国火灾统计年鉴(2001-2006)》及公安部消防局火灾统计系统导出的城镇重特大火灾基础数据,它是我国关于火灾的权威统计数据

资源。因关联规则算法是针对离散属性数据提出的,所以首先应对数据进行预处理,包括数据的离散和缺省值的处理等。对数据的预处理按照公安部统一制定的“火灾报告表”,选取记录较为完整的起火时间、场所和原因指标,作为与火灾直接财产损失进行关联性分析的因素进行采集和录入,建立符合 SPSS Clementine 8.1 数据挖掘工具分析的数据源。其中,对起火时间属性中的月份和 24 h 进行了离散;原因属性包括报告中列出的 10 种原因,即放火、电气、违章操作、用火不慎、吸烟、玩火、自燃、雷击、不明和其他;起火场所属性的分类为 6 大类,即建筑物(厂房、村民住宅、居民住宅、办公用房、库房、商场商店、集贸市场、娱乐场所、歌舞厅、宾馆饭店、影剧院、文博馆、车站码头机场、古建筑、学校、医院、其他建筑物),构筑物(加油站、燃气站、可燃气体储罐、可燃液体储罐、露天框架、管道、其他构筑物),堆场(粮食、棉花、草类、其他堆场),交通工具(机动车、列车、船舶、飞机、其他交通工具),山林草原和其他场所火灾。这些指标属性若存在缺省值,可以通过核实相关文献得到填补。对火灾直接损失指标的处理按照《火灾统计管理规定》对火灾等级的界定,依据对火灾直接财产损失额的分类,可将其离散成相应的风险等级,如表 2 所示。火灾直接财产损失风险等级可分为三类,第一类是指发生城镇重特大火灾时,直接财产损失额大于等于 100 万元,为高风险;第二类是直接财产损失额大于等于 30 万元且小于 100 万元,为中风险;第三类是直接财产损失额小于 30 万元,为低风险。

表 2 火灾直接财产损失风险等级分类
Table 2 Risk levels of fire direct property damage in a fire

直接财产损失	<30 万元	≥30 万元且 <100 万元	≥100 万元
直接财产损失 风险级别	低	中	高

3 关联性分析结果

采用 SPSS Clementine 8.1 数据挖掘工具中的 Apriori 算法建立模型,对城镇重特大火灾进行关联性分析。限于篇幅,本文重点讨论了城镇重特大火灾直接财产损失风险与火灾的发生时间、场所和原因之间的双因素和三因素关联性。图 1 为使用 SPSS Clementine 8.1 分析的过程。综合考虑支持度、可信度和作用度的值,以支持度为主,如果支持度和可信度的值比较大,作用度的值 ≥ 1 ,则认为因素与结果之间具有较高的关联度。

3.1 双因素关联性分析

表 3 列出了按火灾直接财产损失风险等级与火灾

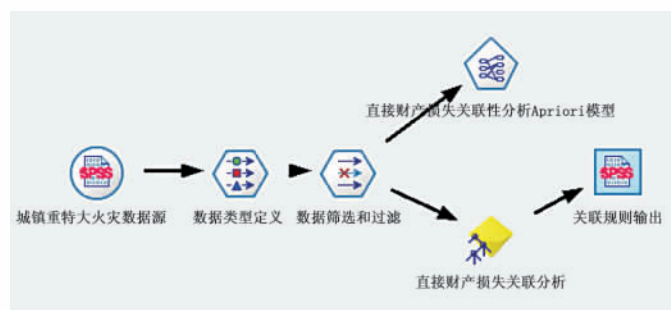


图 1 城镇重特大火灾直接财产损失风险的关联性分析过程

Fig. 1 Schematic representation of risk analysis on direct property damage for major fire incidents in towns based on association rules

发生场所和原因特征双因素关联支持度排列前 5 位的因素组合。因本文关注的重点是高风险直接财产损失,故表 4 和表 5 省略了直接财产损失低、中风险与相关因素的关联结果。

3.1.1 场所和原因

表 3 数据表明,在厂房、商场商店或库房中因电气原因造成的火灾和在厂房、可燃液体储罐场所因违章操作引发的火灾易造成高风险直接财产损失。以厂房中发生电气火灾为例,厂房电气火灾导致高风险直接财产损失的支持度(即在厂房中因电气原因导致的高风险直接财产损失的火灾数在 1 356 起城镇重特大火灾

中所占比率)为 0.88%,可信度(即厂房电气火灾中高风险直接财产损失的火灾数在厂房电气火灾中的比例)为 9.7%,作用度(即厂房电气火灾中高风险直接损失火灾数在厂房电气火灾中所占比率与全部高风险直接损失火灾在 1 356 起城镇重特大火灾中所占比率)为 0.99,因而与高风险直接财产损失火灾具有较高的关联度。虽然可燃液体储罐因违章操作引发的高风险直接财产损失火灾在全部重特大火灾中的比例并不是太高,但在可燃液体储罐场所,一旦有违章操作发生,就容易形成高风险直接财产损失火灾。

表 3 城镇重特大火灾直接财产损失风险与场所和原因的双因素关联性分析

Table 3 Risk analysis on direct property damage for major fire incidents in towns based on association rules of two factors including place and cause

场所	原因	直接财产损失 风险级别	支持度/%	可信度/%	作用度
居民住宅	电气	低	2.53	79.1	2.80
居民住宅	放火		2.10	87.5	3.10
商场商店	电气		1.69	24.5	0.87
居民住宅	用火不慎		1.35	75.0	2.66
娱乐场所	电气		1.20	50.0	1.77
厂房	电气	中	7.26	79.8	1.29
商场商店	电气		4.40	63.8	1.03
库房	电气		3.28	86.3	1.39
厂房	违章操作		3.06	69.5	1.12
厂房	其他		2.88	84.8	1.37
厂房	电气	高	0.88	9.70	0.99
商场商店	电气		0.81	11.7	1.19
厂房	违章操作		0.52	11.9	1.21
库房	电气		0.45	11.8	1.20
可燃液体储罐	违章操作		0.44	54.5	5.56

3.1.2 时间和场所

月周期时间和场所特征双因素关联性分析结果表明,1 月份和 2 月份商场、商店,4 月份和 7 月份厂房发生重特大火灾的概率较大,分别占全部重特大火灾的 2.8%、2.6%、2.2%和 2.2%;11 月份、3 月份和 8 月份库房,以及 7 月份、2 月份厂房易发生高级别直接财产损失火灾,3 月份库房重特大火灾中高级别直接财产损失火灾的比例为 38.5%,表 4 给出了其关联度数据。

从日周期时间和场所特征两个因素看,03:00–04:00 和 01:00–02:00 时段的商场、商店火灾,00:00–01:00 和 02:00–03:00 时段的厂房火灾形成重特大火灾的概率较大,分别占全部重特大火灾的 2.1%、1.6%、1.8%和 1.6%;而 02:00–03:00 和 21:00–22:00 时段的商场、商店火灾,03:00–04:00 时段的库房火灾,以及 00:00–01:00 时段的厂房火灾易形成高级别直接财产损失,03:00–04:00 时段的库房火灾中高级别直接财产损失火灾的比例为 50%,参见表 4。

3.1.3 时间和原因

从月周期时间和原因特征看,1 月份、12 月份、2 月份因电气原因导致的重特大火灾概率较大,分别占全部重特大火灾的 3.4%、2.9%和 2.8%;7 月份的电气火灾和放火火灾、10 月份和 4 月份的电气火灾易造成高级别直接财产损失,7 月份的电气火灾中高级别直接财产损失火灾的比例为 44.4%,见表 4。

从日周期时间和原因特征看,01:00–02:00 时段的电气火灾,00:00–01:00 时段的电气火灾,02:00–03:00 的放火火灾和 02:00–03:00 时段的电气火灾形成重特大火灾的概率较大,分别为 2.4%、2.1%、2.1%和 2.1%;23:00–24:00 和 21:00–22:00 时段的电气火灾,01:00–04:00 时段的放火火灾易造成高级别直接财产损失,其中 21:00–22:00 时段的电气火灾中高级别直接财产损失火灾的比例较高,为 36.4%,见表 4。

表4 城镇重特大火灾高级别直接财产损失风险的双因素关联性分析
Table 4 Risk analysis on high level direct property damage for major fire incidents in towns based on association rules of two factors

因素组合	因素 1	因素 2	支持度/%	可信度/%	作用度
月周期与 场所	11 月	库房	0.490 0	70	7.137
	3 月	库房	0.385 0	38.5	3.921
	7 月	厂房	0.367 4	16.7	1.699
	8 月	库房	0.305 5	23.5	2.399
	2 月	厂房	0.300 6	16.7	1.699
日周期与 场所	02:00-03:00	商场商店	0.357 0	23.8	2.427
	03:00-04:00	库房	0.350 0	50	5.098
	00:00-01:00	厂房	0.288 0	16	1.631
	21:00-22:00	商场商店	0.280 0	40	4.078
	16:00-17:00	其他	0.240 0	60	6.117
月周期与 原因	7 月	电气	0.374 4	20.8	2.124
	7 月	放火	0.310 8	44.4	4.531
	3 月	其他	0.300 0	20.0	2.039
	10 月	电气	0.296 7	12.9	1.316
	4 月	电气	0.292 6	13.3	1.359
日周期与 原因	23:00-24:00	电气	0.375 9	17.9	1.821
	03:00-04:00	放火	0.357 0	23.8	2.427
	21:00-22:00	电气	0.291 2	36.4	3.707
	02:00-03:00	放火	0.224 7	10.7	1.092
	01:00-02:00	放火	0.217 6	13.6	1.39
日周期与 月周期	18:00-19:00	7 月	0.225 0	75.0	7.647
	19:00-20:00	2 月	0.225 0	37.5	3.823
	02:00-03:00	2 月	0.214 5	42.9	4.369
	15:00-16:00	5 月	0.160 0	40.0	4.078
	12:00-13:00	7 月	0.160 0	40.0	4.078

3.1.4 月周期时间与日周期时间

从月周期和日周期特征的关联性分析看,1 月份的凌晨 02:00-03:00,6 月份的凌晨 03:00-04:00 发生重特大火灾的概率较大,分别占全部重特大火灾的 1.4%和 1.1%;7 月份的傍晚 18:00-19:00 和 2 月份的 19:00-20:00 时段易发生高级别直接财产损失火灾,并且 7 月份傍晚 18:00-19:00 时段的高级别直接财产损失火灾比例高达 75.0%,见表 4。

3.2 三因素关联性的分析

3.2.1 场所、原因与月周期时间

场所、原因及月周期特征的三因素关联性分析结果表明,1 月份和 2 月份的商场、商店电气火灾,1 月份

和 2 月份的厂房电气火灾导致重特大火灾的概率较大,占全部重特大火灾的比例均为 0.8%。表 5 中的关联度数据表明,7 月份发生厂房电气火灾造成高级别直接财产损失的概率较大,而 2 月份在商场、商店因吸烟引发的火灾和 9 月份厂房中因焊割引发的火灾中形成高级别直接财产损失火灾的比例均为 66.7%。

3.2.2 场所、原因与日周期时间

场所、原因及日周期特征的三因素关联性分析结果表明,00:00-01:00 时段厂房电气火灾,03:00-04:00 时段商场商店电气火灾,00:00-02:00 时段厂房电气火灾和 03:00-04:00 时段商场商店电气火灾导致重特大火

灾的概率较高,占全部重特大火灾的比例分别为 0.9%、0.8%、0.7%和 0.7%。21:00-22:00 时段发生的商场商店火灾不仅造成高级别直接财产损失的概率大,而且其形成高级别直接财产损失火灾的比例高达 50%,见表 5。

4 讨论与结论

通过上面的分析可以得到以下结论。

1) 商场商店和厂房库房是发生高等级直接财产损失火灾的重点场所,这些场所在不同时间段发生高等级直接财产损失火灾的概率不同,发生火灾的原因主要是电气,其次是放火。

2) 库房发生高等级直接财产损失火灾的主要时间是 11 月、3 月和 8 月;厂房主要是在 7 月和 7 月。

3) 商场商店发生高等级直接财产损失火灾的重点时段是凌晨 02:00-03:00 和夜晚 21:00-22:00;库房的重点时段是凌晨 03:00-04:00;厂房的重点时段是午夜 00:00-01:00。

4) 从月周期时间看,7 月、10 月和 4 月是高等级直接财产损失电气火灾的高发期,7 月还是放火造成高等级直接财产损失火灾的高发期。

5) 高等级直接财产损失电气火灾的高发时段集中在 21:00-22:00 和 23:00-24:00;而凌晨 01:00-04:00 则是纵火引发高等级财产损失火灾的主要时段。

根据关联性分析结果,可以有针对性地采取消防监督措施,并随时做好火灾扑救准备工作,一旦发生火灾,能够迅速采取以下有力措施,将火灾损失控制在最低限度。

1) 对于厂房,在 7 月份应重点对电气设备进行安全检查,并采取有效的预防措施;在 9 月份应对焊割作业进行重点监控。对于商场商店,在 4 月份应注意防止电气火灾;在 2 月份应加强对吸烟的管理。

2) 7 月份傍晚 18:00-19:00 和

表 5 城镇重特大火灾高级别直接财产损失风险与场所和原因及时间三因素的关联分析

Table 5 Risk analysis on high level direct property damage for major fire incidents in towns based on association rules of three factors including place, cause and time

因素组合	因素 1	因素 2	因素 3	支持度/%	可信度/%	作用度
月周期、场所及原因	7 月	厂房	电气	0.16	22.2	2.27
	3 月	其他	其他	0.15	50.0	5.10
	4 月	商场商店	电气	0.15	25.0	2.55
	2 月	商场商店	吸烟	0.13	66.7	6.80
	9 月	厂房	焊割	0.13	66.7	6.80
日周期、场所及原因	21:00-2:00	商场商店	电气	0.20	50.0	5.10
	23:00-24:00	厂房	电气	0.16	40.0	4.08
	03:00-04:00	库房	放火	0.16	40.0	4.08
	23:00-24:00	商场商店	电气	0.15	22.2	2.27
	00:00-01:00	厂房	静电	0.15	50.0	5.10

12:00-13:00 时段, 以及 2 月份 19:00-20:00 时段和凌晨 02:00-03:00 时段易发生高等级直接财产损失火灾, 应对重点场所加强巡查, 做到及时发现火情、及时进行处置。消防部门也应在上述时段加强战备值勤, 对辖区内的重点对象进行排查, 制订灭火预案, 接到报警后以最快的速度到达现场, 控制火势的扩大和进一步发展。

采用数据挖掘方法对火灾统计数据进行分析是一种探索火灾发生规律的新的尝试。由于篇幅所限, 本文针对我国 2000-2005 年 1 356 起城镇重特大火灾进行了高等级直接财产损失火灾与其发生原因、场所和时间等因素之间的关联性分析, 其结果可用于有针对性、有重点地进行消防监督管理, 并可用于指导提高消防执勤战备的有效性。

参考文献 (References)

- [1] 朱绍文, 陆玉昌. 关联规则挖掘技术及发展动向[J]. 计算机工程, 2000(9): 4-7.
ZHU Shaowen, LU Yuchang. Technique and development of association rules mining [J]. Computer Engineering, 2000(9): 4-7.
- [2] 王培吉. 基于 Apriori 算法的关联规则挖掘及改进[D]. 呼和浩特: 内蒙古大学, 2003.
WANG Peiji. Mining association rules based on apriori algorithm and improvement [D]. Hohhot: Inner Mongolia University, 2003.
- [3] 裘国相. 基于关联规则的电厂运行优化目标值确定研究[D]. 大连: 大连海事大学, 2006.
QIU Guoxiang. Research on determining operation optimization value for power plant based on association rules [D]. Dalian: Dalian Maritime University. 2006.
- [4] 谢连雄, 李仲麟. 数据挖掘在电信宽带资源预警中的应用[J]. 广东通信技术, 2005(8): 8-13.
XIE Lianxiong, LI Zhonglin. The application of data mining in telecommunication alarm [J]. Guangdong Communication Technology, 2005(8): 8-13.
- [5] 毕建欣. 数据挖掘技术在我国保险业中的应用研究[D]. 福州: 福州大学, 2005.
BI Jianxin. Research on application of data mining technology in insurance[D]. Fuzhou: Fuzhou University, 2005.
- [6] 吴立志. 城市火灾规律及风险评价技术研究 [D]. 北京: 北京理工大学, 2005.
WU Lizhi. Research on urban fire regularity and risk evaluation [D]. Beijing: Beijing Institute of Technology, 2005.
- [7] 李志洁. 一种基于模式变换的高效关联规则挖掘算法[D]. 大连: 大连理工大学, 2004.
LI Zhijie. An effective algorithm for mining association rules based on stack-transform[D]. Dalian: Dalian Institute of Technology, 2004.
- [8] 安淑芝, 等. 数据仓库与数据挖掘 [M]. 北京: 清华大学出版社, 2005.
AN Shuzhi, et al. Data warehouse and data mining [M]. Beijing: Tsinghua University Press, 2005.
- [9] 性能概述——应用 Clementine 进行数据挖掘 [EB/OL]. 2007-06. <http://www.broadtarget.com/clementine-1.asp>.
Performance summary—data mining by Clementine [EB/OL]. <http://www.broadtarget.com/clementine-1.asp>.
- [10] HAN Jiawei, MICHELINE K. 数据挖掘概念与技术 [M]. 范明, 孟晓峰, 译. 北京: 机械工业出版社, 2001: 226-227.
HAN Jiawei, MICHELINE KAMBER. Data mining: concept and techniques [M]. FAN Ming, MENG Xiaofeng, trans. Beijing: China Machine Press, 2001: 226-227.

(责任编辑 冯峰)