

基于智慧实验室的安全事故分析与预测

楚丹琪¹, 李睿智², 高洪皓³, 张 康¹

(1. 上海大学 实验设备处, 上海 200444;

2. 上海大学 计算机学院, 上海 200444;

3. 上海大学 计算中心, 上海 200444)

摘 要:随着中国高校信息化水平的不断提高和信息化数据的积累,智慧实验室应运而生。实验室的主体是人,而人也是构成实验室安全事故的重要因素,所以实验室安全防范、事故预测显得十分必要。文中提出了安全信用的概念,提出了一种新型的安全教育模式,并通过对上海大学思安网信息化平台提供的信息化数据的研究与分析,从用户特质、行为偏好、学习能力、考试成绩、课程考勤和事故历史六个维度来对实验用户进行考察。实验室进入者的安全知识和安全技能的掌握程度可以反映实验室的安全隐患,因此基于安全信用将用户进行分类,挖掘高危人群和潜在危险人群的安全技能和安

关键词:智慧实验室;安全信用;事故预测;数据分析

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2018)05-0107-05

doi:10.3969/j.issn.1673-629X.2018.05.025

Analysis and Prediction of Safety Accidents Based on Smart Laboratory

CHU Dan-qi¹, LI Rui-zhi², GAO Hong-hao³, ZHANG Kang¹

(1. Equipment Office of Shanghai University, Shanghai 200444, China;

2. School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China;

3. Computing Center, Shanghai University, Shanghai 200444, China)

Abstract: With the continuous improvement of the informatization level and the accumulation of data in Chinese universities, the intelligent laboratory emerges. The human, as the main body of the laboratory, is the important factor of the laboratory safety accidents, so the protection of laboratory and the prediction of accidents are very significant. In this paper we put forward the concept of security credit and a new model for safety education, and evaluate the users from user's preferences, behavior characteristics, learning ability, test scores, attendance and accident history through the research and analysis of informatization data from Shanghai University Sian Informatization Platform. The laboratory safety hazard can be reflected by master degree of safety knowledge and skills of lab entrants, so we categorize users based on safety credit and mine the vulnerabilities of safety knowledge and skills of population with high risk and potential risk, and then forecast the security accident. From the large data a new way for the smart laboratory management is given.

Key words: smart laboratory; safety credit; accident prediction; data analysis

1 概 述

1.1 实验室现状及问题

实验室是科技的摇篮,安全、健康、环保、节能是实验室的基本要素^[1-3]。实验室安全教育是安全管理的重要内容之一,充实的安全教育可以提高人们的认识和安全素质,也是保障实验人员安全的第一道也是最重要的屏障。为了防患于未然,全方位保护实验人员

和实验设备的安全十分必要。智慧实验室是以“互联网+”为基础,采用先进的信息化技术手段和软硬件设备,实现对实验教学、科研和管理等实验室信息的收集、处理、监管、存储、传输和调用^[4-5],使实验室资源得以充分优化利用。

高校实验室的信息化存在以下亟待解决的问题:

(1)设备管理非自动化。教学仪器和设备分布离

收稿日期:2017-05-23

修回日期:2017-09-27

网络出版时间:2018-02-08

基金项目:上海市科研计划项目(14590500500);2016年全国高等院校计算机基础教育研究会计算机基础教学改革课题(2016066);2016年赛尔网络下一代互联网技术创新项目(NGII20160210,NGII20160614,NGII20160325)

作者简介:楚丹琪(1958-),女,副教授,研究方向为信息化建设、软件工程;李睿智(1990-),男,硕士,研究方向为云计算和大数据。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180207.1913.072.html>

散、管理难度大、无专人管理和保养等;信息获取被动,存在空间障碍,对设备数量和种类繁多的实验室,管理人员难以准确快速定位、跟踪设备使用和维护。

(2)安全管理非智能化。实验室安全管理难以形成稳固的体系,对各类潜在危险的警告系统尚不完善^[6],只能依赖于管理人员的责任心。随着大量开放式实验教学的开展,对人员的智能监控要求越来越高。

(3)环境管理非低碳化。机房的温湿度、粉尘、噪声等指标未实现统一监控和管理;实验室照明、设备供电、机器用电难以实现自动化调节,造成资源浪费。

(4)资源管理非共享化。不同院系实验室之间资源利用和分配不协调,资源共享难以实现,存在“信息孤岛”现象。

1.2 实验室安全预防策略现状

目前,国内高校实验室安全工作多数由多个部门共同负责,相关处室工作侧重点不同。在这种不同职能处室分块管理的管理体制下,容易出现部分管理职责交叉重复的现象,导致管理效率不高^[7-8]。同时,每个实验室面对的用户群体不同,使用的仪器设备不同,如果全都进行相同的实验室安全教育,会浪费大量的人力、物力和财力,并达不到相应的教育和预防效果。

在信息化不断发展的时代,良好、真实的数据统计与挖掘可以为政策的制定提供有力的支持^[9-10]。上海大学思安网是上海大学针对实验室安全和保护展开的全方位平台。上海大学思安网不仅整合了上海大学用户信息、在线学习、在线考试、上课考勤,而且记录分析用户的操作习惯和事故记录。借助以上真实的数据,文中提出安全信用的模型,通过这个模型可以将用户的安全知识和安全技能的掌握情况进行量化,从而为每个实验室具有针对性的安全事故预测奠定了基础。

2 安全信用分析模型

安全信用,是上海大学思安网运用大数据分析技术客观呈现个人的安全信用状况。安全信用的评测通过对海量信息数据的综合处理和评估,主要包含了用户特质、行为偏好、学习能力、考试成绩、课程考勤和事故历史六个维度。最后,通过归一化处理综合计算得出用户的安全信用分数。

2.1 评价维度

安全信用主要包含用户特质、行为偏好、学习能力、考试成绩、课程考勤、事故记录六个维度,分别从用户的各个方面反映其安全意识和安全行为。单维度评价方式中每一维度的最高分为 100 分,最低分 0 分。

用户特质主要从用户身份、用户背景综合分析。用户身份主要分为本科生、硕士生和博士生;相较于本科生而言,硕士生和博士生在安全信用方面具有知识

积累的优势;用户背景主要考察用户的学院和专业背景,尤其对于文科专业会有一定的背景优势;同时,不同的专业年级对安全知识积累沉淀也会有所不同^[11]。

定义 1:用户特质 S_1 。

$$S_1 = \alpha \times (P_e + P_g) + \beta \times (P_m + P_r) \quad (1)$$

其中, α 表示学历层次和专业年级在评价中所占权重, β 表示专业性质在评价中所占权重,并且满足条件 $\alpha + \beta = 1$; P_e , P_g , P_m , P_r 表示学历、年级、专业和高危专业的权重。

行为偏好主要从用户的操作行为进行分析,用户的操作习惯往往可以反映用户现实生活中的习惯,比如:用户是否具有良好的登录和注销的习惯,用户学习和考试是否存在作弊现象。

定义 2:行为偏好 S_2 。

$$S_2 = \sum \alpha \times N_e - \sum \beta \times N_e \quad (2)$$

其中, α 和 β 分别表示某类良好习惯和不良习惯发生的权重; N_e 和 N_e 表示某类良好习惯和不良习惯发生的次数。

学习能力主要从用户在线学习进行分析,人的注意力集中的时间跟年龄、学习时间和学习难度有联系。用户的年龄直接影响用户学习的注意力,为了方便计算,针对年龄采用线性变换。

定义 3:年龄 S_3 。

$$S_3 = 100 - U_a \quad (3)$$

其中, U_a 表示用户的年龄。

用户每天平均学习时间的长短、学习时间段均是影响用户学习程度的因素,下面给出学习时间的定义。

定义 4:学习时间 S_4 。

$$S_4 = \eta \times (- (t - T)^2 + 100) \quad (4)$$

其中, T 表示正常学习时间中间值,按照正常学习时间范围 40 ~ 60 min 之间,故 T 应该取值 50; η 表示不同时间段的权重系数,将 24 小时划分为 4 部分:深夜、上午、下午、夜晚,每部分的权重系数分别为 0.1、0.8、0.6、0.5。

上海大学思安网针对每类资源均定义了学习难度,学习的内容难度越大,表示用户的学习能力越高,于是下面给出了学习难度的定义。

定义 5:学习难度 S_5 。

$$S_5 = \sum_{i=1}^N \sigma_i / N \quad (5)$$

其中, N 表示学习资源总数; σ_i 表示第 i 个资源的难度系数。

所有用户的学习能力可以看作是符合正态分布,同时根据定义 3 ~ 5,给出学习能力的定义。

定义 6:用户学习能力 S_6 。

$$S_6 = 100 \times \frac{\alpha}{\sqrt{2\pi}} e^{-\frac{(c_1 \times S_3 + c_2 \times S_4 + c_3 \times S_5 - \mu)^2}{2}}$$

(6)

其中, c_1, c_2, c_3 表示用户年龄、学习时间、学习难度的权重系数; α 表示整体权重系数; μ 表示正态分布的位置参数, $\mu = 75$ 。

考试成绩主要从用户参加考试的考试结果进行分析, 用户的考试成绩越高, 其安全知识掌握程度越高。

定义 7: 考试成绩 S_7 。

$$S_7 = \frac{\sum_{i=1}^N \text{Score}(i)}{N}$$

(7)

其中, N 表示用户参加的考试数目; $\text{Score}(i)$ 表示用户参加的第 i 个安全考试的考试成绩。

课程考勤主要从用户上课考勤情况进行分析, 上课情况可以从侧面反映出用户的学习态度。

定义 8: 课程考勤 S_8 :

$$S_8 = \frac{100 \times \sum_{i=1}^N (n_i / N_i)}{T}$$

(8)

其中, n 表示出勤次数; N 表示应该出勤次数; T 表示课程总数。

信用历史主要从用户过去发生过的安全事故和安全预防的事件进行计算, 这是最客观反映用户安全信用的方式。文中信用历史采用指数函数表示, 以提高初始的变化率。

定义 9: 信用历史 S_9 。

$$S_9 = -e^{-(\alpha \times N_p - \beta \times N_a)} + 100$$

(9)

其中, α, β 分别表示预防事故的权重和事故发生权重, $\alpha + \beta = 1$; N_p, N_a 分别表示预防事故的数量和事故发生数量。

2.2 安全信用归一化

文中的安全信用评价方式主要是通过对每一个维度分别进行评价, 最后综合计算得出用户的安全信用分数。安全信用分数的高低可以反映出用户的安全知识和安全技术的掌握情况。那么, 如何将六个维度归一化成安全信用, 文中采用六个维度来计算安全信用, 并将这六个维度放置到一个六维坐标系中。安全信用值的计算如下所示:

$$S = \sqrt{S_1^2 + S_2^2 + S_6^2 + S_7^2 + S_8^2 + S_9^2}$$

(10)

3 安全事故分析与预测算法

实验室事故预测主要是基于用户安全信用的六个维度来进行。首先根据用户六个维度, 将用户聚集为三类: 高危人群、潜在危险人群、安全人群, 下面会详细介绍用户的分类方法; 接着根据实验室面对的人群重点分析高危人群和潜在问题人群, 从学习内容和考试

成绩中重点分析知识漏洞, 从而预测安全事故。

3.1 用户聚类

根据安全信用公式计算出用户的安全信用后, 根据用户安全信用进行第一次聚类^[12-13], 这次聚类主要分为两类人群: 安全人群和危险人群。聚类原则如下:

(1) 安全人群应该满足安全信用分大于等于 220 分, 安全信用分达到这个层次可以表明用户的六个维度的平均分不低于 90 分, 用户具有良好的安全知识和安全技术的积累。

(2) 危险人群满足安全信用分数低于 220 分, 这部分用户在安全意识方面还有一定的欠缺。

根据第一次用户聚类方式可以区分安全人群和危险人群, 但是在危险人群中, 并非每一个用户需要进行关注。那么, 文中采用 K-means 算法对第一次聚类中的危险人群进行了第二次聚类^[14], 这次聚类主要分为: 高危人群和潜在问题人群。聚类算法如下:

(1) 首先根据用户信用分与阈值 220 分做比较, 如果大于等于 220 分则将该用户划分至安全人群中, 如果低于 220 分则划分至危险人群中;

(2) 确定用户行为向量, 经过实际分析, 公式为:

$$V = \{a_1, a_2, \dots, a_6\}$$

(11)

其中, a_1, a_2, \dots, a_6 分别为某个用户的六个维度。

(3) 随机选择 k 个初始聚类中心:

$$\text{Center}_k = \{C_{1k}, C_{2k}, \dots, C_{6k}\}$$

(12)

其中, k 应满足 $k = 2$ 。

(4) $s \in S$, 其中 S 为样本集合, 根据欧几里得公式计算 S 距离所有聚类中心的距离, 将 s 划分到距离最近的中心 Center 的集合内。

(5) 重新计算聚类 Center 的中心, 使用各个点的坐标的平均值来求点群新的中心点。若产生新的中心点, 执行第 3 步; 若中心点不再发生改变, 执行第 5 步。

(6) 使用欧几里得公式计算 2 个聚类中心距离原点的距离, 聚类中心距离原点更近的聚类簇为高危人群, 距离较远的聚类簇为潜在问题人群。

聚类算法如下:

算法 1: 用户聚类算法。

输入: 用户安全信用的六个维度信息及 $S_1, S_2, S_6, S_7, S_8, S_9$ 评价分 S

输出: 安全人群 Groupsafe、高危人群 Groupdanger 和潜在问题人群 Grouppotential

```
BEGIN
n = 0
FOR EACH i = 1 TO m
IFSi ≥ 220 THEN
Groupsafe = Groupsafe ∪ i
ELSE
vi = {a1, a2, a6, a7, a8, a9}
```

```
n = n + 1
ENDIF
ENDFROEACH
FOR EACHi = 1 TO 2
Random( Centeri )
ENDFROEACH
Group1 = {}
Group2 = {}
WHILE ( Center is not change )
FOR EACHi IN n
DisPoint&Center1 =  $\sqrt{\text{sum}((v_i - \text{Center}_1)^2)}$ 
DisPoint&Center2 =  $\sqrt{\text{sum}((v_i - \text{Center}_2)^2)}$ 
IF DisCenter1 ≥ DisCenter2 THEN
Group1 = Group1 ∪ i
ELSE
Group2 = Group2 ∪ i
ENDIF
ENDFOREACH
FOR EACHi = 1 TO 2
Update( Centeri )
ENDIF
ENDWHILE
DisO&Center1 =  $\sqrt{\text{sum}((\text{Center}_1)^2)}$ 
DisO&Center2 =  $\sqrt{\text{sum}((\text{Center}_2)^2)}$ 
IF DisO&Center1 ≥ DisO&Center2 THEN
Grouppotential = Group1
Groupdanger = Group2
ELSE
Groupdanger = Group1
Grouppotential = Group2
ENDIF
END
```

3.2 安全事故预测

根据 3.1 节,将用户聚类为高危人群、潜在危险人群、安全人群三类,本节的安全事故预测算法将重点关注高危人群和潜在危险人群。将安全事故定义为五类:火或水灾性事故、毒害性事故、爆炸性事故、机械性事故和电气性事故;上海大学思安网中的安全知识学习和考试也分为六类:通识类、化学类、辐射类、医学生物类、电气类和机械类。定义的对对应关系如表 1 所示。

表 1 安全事故、学习和考试分类对应关系

学习与考试类目 (category)	安全事故类目 (type of accident)
通识类	火或水灾性事故
化学类	毒害性事故、爆炸性事故
辐射类	毒害性事故
医学生物类	毒害性事故
电气类	电气事故
机械类	机械事故

在现实生活中每一个实验室或高危场所都面对一些固定的用户群,根据 3.1 节可以将固定的用户群进行用户类型划分,从而抽取高危人群和潜在危险人群的数据进行分析。每一类事故发生的概率取决于两类人群发生事故的概率,于是得到每一类事故发生的概率公式:

$$P_c = \alpha \times P_d + \beta \times P_p$$
 (13)

其中, P_d 和 P_p 表示高危人群和潜在危险人群在该类事故中发生的概率; α 和 β 表示权重系数。

那么,某类人群发生某类安全事故的概率,则可以直接由这类人群学习情况和考试情况来进行反映,于是可以得到如下公式:

$$P_L = \frac{\sum_{i=1}^{group} (1 - \frac{Learn_i}{Total})}{group}$$
 (14)

$$P_E = \frac{\sum_{i=1}^{group} (1 - \frac{\sum_{j=1}^N Exam_{ij}}{Total})}{group}$$
 (15)

其中, $Learn_i$ 表示第 i 个用户学习某类资源的数量; $Total$ 表示某类学习资源的总量; $Exam_{ij}$ 表示第 i 个用户参加某类第 j 次考试的成绩; N 表示该用户参加考试的总数; $group$ 表示人群总数。

可得到某类人群发生某类安全事故的概率公式:

$$P_{group} = \eta \times P_L + \sigma \times P_E$$
 (16)

其中, η , σ 表示权重系数,并且满足 $\eta + \sigma = 1$ 。

4 实验分析

文中的安全信用评价体系和实验室安全事故预测算法均在上海大学思安网中实现,实际运行过程中分析相关数据量可以参考表 2(截止 2016 年 7 月 25 日星期一)。算法中相关权重系数对的参数初始化可以参考表 3。

表 2 算法分析相关数据量

参数名称	取值
用户数/人	16 281
学习记录/条	87 507
考试记录/条	5 172 270
操作日志/条	152 338
事故记录/条	112
考勤记录/条	85 423

通过对上海大学所有用户进行安全信用评价,并且针对 5 个实验室进行安全事故预测来验证算法的合理性。上海大学思安网自动根据用户的相关数据进行六个维度的评价(相关维度定义参考 2.1 节),最终给出用户的安全信用维度评价,可以参考图 1。

表3 相关权重系数对的参数初始化

参数名称	取值
用户特质(α, β)	65,35
行为偏好(α, β)	5,5
学习时间(η)	不同时间段权重系数不同,文中划分为深夜、上午、下午、晚上,权限分别为0.1/0.8/0.6/0.5
学习能力($\alpha, \mu, c_1, c_2, c_3$)	2.5,75,0.2,0.4,0.4
信用历史(α, β)	0.5,0.5
安全信用归一化(η)	0.1

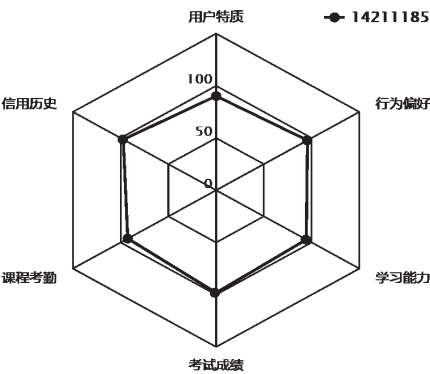


图1 用户的安全信用维度示意图

上海大学思安网在完成维度评价后,根据2.2节进行归一化处理。系统默认展示用户近半年的安全信用变化情况,可以参考表4。

表4 安全信用变化情况(学生14721185)

时间	取值
2017年01月	221.2
2017年02月	216.7
2017年03月	218.9
2017年04月	221.4
2017年05月	225.3

安全事故预测的重点是根据用户群的知识漏洞来预测实验室事故。表5展示的是4个实验室安全事故预测统计情况,标注了事故类型和事故发生的概率。

表5 实验室安全事故预测统计情况 %

事故类型	索朗光伏材料实验室	高性能计算实验室	生产工程试验中心	分析测试中心
火或水灾性事故	12.57	6.98	11.27	9.58
毒害性事故	51.28	12.36	12.19	48.31
爆炸性事故	16.12	22.15	19.36	22.38
电气事故	10.69	53.48	15.75	10.78
机械事故	9.34	5.03	41.43	8.95

5 结束语

上海大学实验室安全教育综合管理平台是针对高

校实验室安全的问题提出的一种新型教育模式;同时在智慧实验的背景下,提出了安全信用的概念,将用户的安全知识和安全技能的掌握情况进行量化,形成一套完整的评价体系。根据用户的安全信用,将用户进行分类,找出危险用户和知识漏洞,从而预测实验室安全事故。在后续的工作中将把安全信用评价应用到更加广泛的实验室预约等功能中,为更安全、更便捷的实验室管理提供保障。

参考文献:

[1] MASSIE D L, CAMPBELL K L, WILLIAMS A F. Traffic accident involvement rates by driver age and gender[J]. Accident Analysis and Prevention,1995,27(1):73-87.

[2] KOO J, KIM S, KIM H, et al. A systematic approach towards accident analysis and prevention[J]. Drug Discovery Today, 2015,20(7):856-862.

[3] WU Lirong, JIANG Zhongan, CHENG Weimin, et al. Major accident analysis and prevention of coal mines in China from the year of 1949 to 2009[J]. Mining Science & Technology, 2011,21(5):693-699.

[4] AKYUZ E, CELIK M, AKYUZ E, et al. Utilisation of cognitive map in modelling human error in marine accident analysis and prevention[J]. Safety Science,2014,70:19-28.

[5] 郑禄,常佳.基于Arduino与Python的实验室联动门禁报警系统设计[J].软件导刊,2017,16(1):76-77.

[6] 张博.基于GA-BP神经网络的高校实验室安全评价研究[J].微处理机,2016,37(2):51-55.

[7] MARSAGLIA G. Evaluating the normal distribution[J]. Journal of Statistical Software,2004,11(5):1-11.

[8] RANI A J M, PARTHIPAN L. Clustering analysis by improved particle swarm optimization and K-means algorithm[C]//IET Chennai 3rd international conference on sustainable energy and intelligent systems. [s. l.]:[s. n.],2012:83-88.

[9] 王国强,吴敏.实验室安全防范智能化系统的组成与应用[J].实验技术与管理,2008,25(4):151-155.

[10] 吴文华,马升灯,沈水富.提高安全意识,预防高校实验室安全事故发生[J].科技风,2015(23):189.

[11] 丁智斌,袁方,董贺伟.数据挖掘在高校学生学习成绩分析中的应用[J].计算机工程与设计,2006,27(4):590-592.

[12] 马闯,吴涛,段梦雅.基于K近邻隶属度的聚类算法研究[J].计算机工程与应用,2016,52(10):55-58.

[13] 徐向平,鲁海燕,徐迅.基于环形邻域的混沌粒子群聚类算法[J].计算机工程与应用,2016,52(2):54-60.

[14] HAMIDA E B, JAVED M A. Channel-aware ECDSA signature verification of basic safety messages with K-Means clustering in VANETs[C]//International conference on advanced information networking and applications. Crans - Montana, Switzerland:IEEE,2016:603-610.