

大型电子商务 web 数据库潜在威胁挖掘模型

高文莲, 高永强

(吕梁学院 计算机科学与技术系, 山西 吕梁 033000)

摘 要: 本文提出基于人工免疫算法的大型电子商务 web 数据库潜在威胁数据挖掘方法, 计算 web 数据库中潜在威胁数据属性的差异性, 获得潜在威胁数据特征, 对潜在威胁数据进行人工免疫, 在免疫的过程中通过计算抗原与抗体的匹配度, 获取潜在威胁数据的关联性, 结合潜在威胁数据的关联性能够建立潜在威胁数据挖掘模型, 实现了对大型电子商务 web 数据库潜在威胁数据的挖掘. 实验结果表明, 利用改进算法进行大型电子商务 web 数据库潜在威胁数据挖掘, 能够提高挖掘的准确性, 效果令人满意.

关键词: 大型电子商务; 潜在威胁; 数据挖掘; 人工免疫算法

中图分类号: TN402

文献标识码: A

文章编号: 1000-7180(2014)05-0151-03

Large Potential Threats E-commerce Web Data Mining Model

GAO Wen-lian, GAO Yong-qiang

(Luliang University, Department of Computer Science, Luliang 033000, China)

Abstract: The article is based on artificial immune algorithm for this potential threat to large e-commerce web database data mining methods. Computing potential threats in web database data attributes, the difference of the potential threat data obtained characteristics of artificial immune to potential threat data, in the process of immune by calculating the matching of antigen and antibody, can obtain potential threat data correlation, combined with the potential threat to potential threat data mining model is set up the correlation of data, has realized the potential threat to large e-commerce web database of data mining. Experimental results show that the improved algorithm to large potential threats e-commerce web database, data mining can improve the accuracy of mining, effect is satisfactory.

Key words: large e-commerce; The potential threat; Data mining; Artificial immune algorithm

1 引言

在大型电子商务 web 数据库潜在威胁数据挖掘的过程中, 由于大型电子商务交易过程中的数据必须进行保密^[1-7], 因此, 在进行潜在威胁数据的挖掘过程中难以计算潜在威胁数据的关联性, 降低了潜在威胁数据挖掘的准确性^[8].

为了避免传统方法的弊端, 提出一种基于人工免疫算法的大型电子商务 web 数据库潜在威胁数据挖掘方法. 实验结果表明, 利用改进算法进行大型电子商务 web 数据库潜在威胁数据察觉, 能够获得

准确的潜在威胁数据, 为大型电子商务安全交易提供了可靠保障.

2 web 数据库潜在威胁数据挖掘优化方法

利用传统的 Apriori 算法进行大型电子商务 web 数据库潜在威胁数据挖掘, 无法避免由于 web 数据库潜在威胁数据必须进行保密导致在挖掘过程中难以计算潜在威胁数据关联性的弊端, 导致 web 数据库潜在威胁数据挖掘准确性降低. 为了避免上述弊端, 提出一种基于人工免疫算法的大型电子商务 web 数据库潜在威胁数据挖掘方法.

收稿日期: 2013-07-23; 修回日期: 2013-09-18

基金项目: 山西省高等学校科技研究开发项目(20121121); 吕梁学院校级自然基金项目(ZRXN201215)

2.1 web 数据库潜在威胁数据属性特征提取方法

设置大型电子商务 web 数据库中全部的潜在威胁数据的属性特征集合能够用 $b = (b_1, b_2, \dots, b_s)^T$ 进行描述, 这些属性特征构成的矩阵能够用 $V_b = [bb^T]$ 进行描述, web 数据库潜在威胁数据属性特征之间额差异性能够用 $G_b = I[(b-b)(b-b)^T]$ 进行描述. 利用下述公式能够描述 web 数据库潜在威胁数据属性特征的差异性:

$$c = Y^T b = (y_1, y_2, \dots, y_s)^T b = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_s^T \end{bmatrix} b \quad (1)$$

式中, $c_m = y_m^T b, m = 1, 2, \dots, s$.

b 的取值如下:

$$b = (Y^T)^{-1} c = Y_c c = Y^T b = (Y^T)^{-1} c = Y_c \quad (2)$$

在对大型电子商务 web 数据库中的潜在威胁数据进行属性特征提取时, 利用下述公式能够计算 web 数据库潜在威胁数据属性特征权重:

$$\hat{c} = \sum_{n=1}^s da_n + \sum_{n=s+1}^t ga_n \quad (3)$$

利用下述公式能够进行大型电子商务 web 数据库潜在威胁数据特征提取:

$$N = \sum_{m=r+1}^s [y_m^T G_b y_m - x_m (y_m^T y_m - 1)] \quad (4)$$

式中, x_m 表示大型电子商务 web 数据库全部潜在威胁数据属性特征矩阵, y_m 表示 web 数据库潜在威胁数据的特征分量.

利用图 1 能够描述 Web 数据库潜在威胁数据与其对应的属性特征之间的关系.

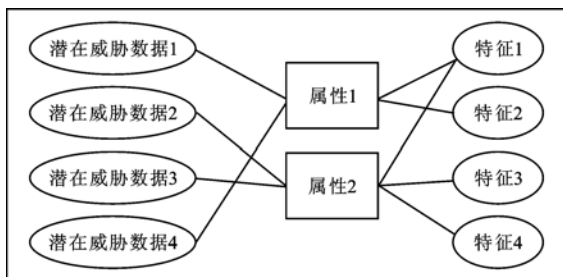


图 1 潜在威胁数据与属性特征的关系

2.2 建立 web 数据库潜在威胁数据挖掘模型

在大型电子商务 web 数据库潜在威胁数据挖掘的过程中, 全部抗原能够用下述公式进行描述:

$$C_i = (C_{i1}, C_{i2}, \dots, C_{ij}), (k = 1, 2, \dots, p) \quad (5)$$

在进行 web 数据库潜在威胁数据挖掘的过程

中, 利用下述公式能够描述抗体描述潜在威胁数据的关联性,

$$C_d = (C_{d1}, C_{d2}, \dots, C_{dk}), (k = 1, 2, \dots, q) \quad (6)$$

在使用人工免疫算法进行大型电子商务 web 数据库潜在威胁数据挖掘的过程中, 利用二进制能够表示潜在威胁数据挖掘过程中的抗原与抗体. 有潜在威胁数据中存在大量的属性特征, 并且这些属性特征之间存在很强的差异性, 因此, 需要对上述潜在威胁数据进行数据挖掘.

在进行 web 数据库潜在威胁数据挖掘的过程中, 利用下述公式能够对全部初始潜在威胁数据集进行重构:

$$Jv(sample) = \{z \in W \mid \exists u \in sample, J(z, u) \leq V\} \quad (7)$$

上述式中, 在 web 数据库潜在威胁数据挖掘的过程中, 集合重构的阈值能够用 V 进行描述.

设置进行 web 数据库潜在威胁数据挖掘的过程中, 全部抗体的数量能够用 Q 进行描述, 第 k 个抗体能够用 N_k 进行描述, 则抗原与抗体的匹配性能够用下述公式进行描述:

$$h_k = \frac{1}{N} \sum_{l=1}^{N_k} R_l \quad (k = 1, 2, \dots, q) \quad (8)$$

式中,

$$R_l = \begin{cases} 1, & \text{匹配} \\ 0, & \text{不匹配} \end{cases}$$

利用人工免疫算法的进行潜在威胁数据挖掘的过程中, 对全部的克隆体进行计算, 能够获得克隆体数量与匹配性之间的关联性, 利用下述公式能够实现这一过程:

$$q_k = \text{round}(q \times \frac{h_k}{u}) \quad (k = 1, 2, \dots, q) \quad (9)$$

对人工免疫过程中的抗体进行变异处理, 能够得到下述潜在威胁数据挖掘模型, 从而实现对 web 数据库潜在威胁数据挖掘:

$$P_{kl} = P_{kl} + \chi q \quad (10)$$

利用上述建立的大型电子商务 web 数据库潜在威胁数据挖掘模型, 能够实现对 web 数据库中潜在威胁数据的挖掘, 为大型电子商务的安全交易提供了可靠的安全保障.

3 实验结果及分析

为了验证改进算法的有效性, 需要进行一次试验. 利用 Visual C++ 6.0 进行编程构建实验环境. 设置大型电子商务 web 数据库中全部的潜在威胁数

量能够用 q 进行描述,全部潜在威胁数据构成的集合能够用 $\{c_1, c_2, \dots, c_q\}$ 进行描述. 潜在威胁数据对应的属性构成的集合能够用 $\{d_1, d_2, \dots, d_q\}$ 进行描述. 在上述潜在威胁数据集中任意选取一个潜在威胁数据 c_k , 该数据的属性对 d_i 的隶属度能够用 μ 进行描述.

大型电子商务 Web 数据库中潜在威胁数据挖掘的精确度,能够作为衡量不同挖掘算法性能优劣的指标. 利用下下述公式能够计算大型电子商务 web 数据库潜在威胁数据挖掘的精确度:

$$\gamma = \frac{\sqrt{c_k - c_{k-1}^2}}{|d_i - 1|} \quad (11)$$

在试验的过程中,设置大型电子商务 web 数据库中全部数据样本数量是 2 000 个,潜在威胁数据的种类是 20 种,从数据库中随机选择 20 种属性的数据作为试验对象.

根据实验结果能够得知,在样本复杂度较高的情况下,利用改进算法进行大型电子商务 web 数据库潜在威胁数据挖掘,精确度要远高于传统算法,充分体现改进算法的优越性.

对实验过程中的实验数据进行整理分析,获得的实验结果能够用表 1 和表 2 进行描述.

表 1 潜在威胁样本复杂度较低时实验结果

实验次数	精确度/%	
	传统算法	改进算法
1	90	94
2	95	95
3	89	93
4	92	89
5	90	92
6	91	92
7	93	93
8	93	93
9	90	92
10	92	93

根据上述实验结果能够得知,利用改进算法进行大型电子商务 web 数据库潜在威胁数据挖掘,能够避免传统算法在数据挖掘过程中难以计算潜在威胁数据关联性导致的潜在威胁数据挖掘准确性较差的弊端,相对传统算法的精确度平均提高 10% 左右,取得了令人满意的效果.

表 2 潜在威胁样本复杂度较高时实验结果

实验次数	精确度/%	
	传统算法	改进算法
1	80	90
2	78	92
3	79	90
4	80	93
5	75	88
6	76	90
7	76	89
8	75	92
9	78	90
10	77	91

4 结束语

针对传统算法难以计算大型电子商务数据库潜在威胁数据关联性,导致潜在威胁数据挖掘准确性较差的弊端,提出一种基于人工免疫算法的大型电子商务 web 数据库潜在威胁数据挖掘方法. 通过计算 web 数据库中潜在威胁数据属性的差异性能够提取潜在威胁数据特征,对潜在威胁数据进行人工免疫,在免疫的过程中,计算抗原与抗体的匹配度能够得到潜在威胁数据的关联性,利用获得的关联性能够建立 web 数据库潜在威胁数据挖掘模型,从而实现了对大型电子商务 web 数据库潜在威胁数据的挖掘. 实验结果表明,利用改进算法进行大型电子商务 web 数据库潜在威胁数据挖掘,挖掘的准确性明显优于传统算法,为大型电子商务交易的安全提供了可靠的安全保障.

参考文献:

- [1] 张新生, 陈永峰. 基于 Web 数据挖掘技术构建自适应站点研究[D]. 西安: 西安建筑科技大学, 2004.
- [2] Marina Fisher, Ray Lai, Sonu Sharman, 等著. 赵睿, 赵俊明, 译. Java EE 和 NET 互操作性[M]. 北京: 机械工业出版社, 2007.
- [3] 熊忠阳, 周亚峰. Web 访问挖掘的预处理技术的研究[J]. 计算机技术与发展, 2007, 17(8): 11-14.
- [4] 张娥, 郑斐峰, 冯耕中. Web 日志数据挖掘的数据预处理方法研究[J]. 计算机应用研究, 2004(2): 59-60.
- [5] He D, Goker A. Detecting session boundaries from Web user logs[C] // Proceedings of the 22nd Annual Colloquium of IR Research. Cambridge: [s. n.], 2005: 57-66.

(下转第 158 页)

升压/反压电荷泵性能分析方法与升压电荷泵类似,这里不再赘述.

5 电荷泵系统整体仿真

按照图 2 将电荷泵系统的整体结构进行仿真. clk1 选取的时钟频率为 50 kHz, clk2、clk3、clk4 选取的频率都为 20 kHz. 依次启动电荷泵,当输出电压都达到较稳定时,70 ms 时在输出端 VLOUT3b 加上负载电流 0.1 mA, 85 ms 时在输出端 VLOUT2a 加上负载电流 0.1 ms. 仿真结果如图 9 所示.

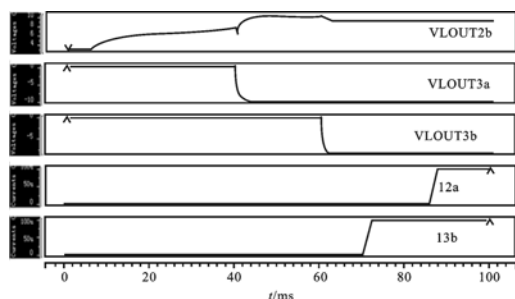


图 9 电荷泵系统电路整体仿真

表 2 中列出电荷泵系统在 VLOUT2a、VLOUT3b 带 0.1 mA 负载时各个输出电压的测量值,求出升压效率. VLOUT1 和 VLOUT3b 的升压效率较高,不少于 94%,而 VLOUT2a、VLOUT2b 和 VLOUT3a 的稍低,但也都都不低于 87%,基本达到电荷泵系统对输出电压的要求.

表 2 电荷泵输出升压效率

	理想值/V	测量值/V	升压效率/%
VLOUT1	5.5	5.17	94.0
VLOUT2a	8.0	7.12	89.0
VLOUT2b	11.0	9.76	88.7
VLOUT3a	-11.0	-9.58	87.1
VLOUT3b	-8.5	-8.18	96.2

6 结束语

本文对对电源电路进行了明确的模块划分和系统设计. 使用 Verilog-A 对系统进行行为级建模,分析了电荷泵如何在充电模式和电荷转移模式实现升压或反压,以及 MOS 管沟道宽度、开关频率、泵电容以及负载对输出电压和消耗功率的影响. 设计了两相不交叠时钟产生电路和相关的电平转移电路. 最后按照电荷泵系统的启动时序,采用 0.16 μm HV 工艺,对整体电路进行了仿真验证. 仿真结果表明,在带负载的情况下,各个输出电压的升压效率都在 87% 以上,有些甚至到 94% 以上,达到了电荷泵系统的设计指标,其功能和性能均满足 AM-OLED 驱动控制芯片的要求.

参考文献:

- [1] Joep Jacobs, Dirk Hente, Eberhard Waffenschmidt. Drives for OLEDs[C]//IEEE Industry Application Conference. IAS Annual meeting. New orleans, IA; IEEE, 2007: 1147-1152.
- [2] Blades CD J, Alison B Walker. Simulation of organic light-emitting diodes[J]. Synthetic Metals, 2005(3): 111-112.
- [3] Hack Mike, Brown J J, Mahon J K. Performance of High Efficiency AM-OLED Displays[J]. Journal of the Society for information Display, 2001,9(3):191-195.

作者简介:

周 静 女,(1989—),硕士研究生. 研究方向是模拟与混合信号 VLSI 设计、AM-OLED 驱动控制芯片的设计与仿真验证.

魏廷存 男,(1960—),博士,教授,博士生导师. 研究方向为 TFT-LCD 驱动控制芯片、高精度稳压电源芯片以及高速接口芯片.

(上接第 153 页)

- [6] Haibin Liu, Vlado Keselj. Combined mining of Web server logs and web contents for classifying user navigation Patterns and Predicting users' future requests[J]. Data & Knowledge Engineering, 2007, 61(2): 304-330.
- [7] 方成效, 袁可风. Web 日志挖掘的数据预处理研究[J]. 计算机与现代化, 2006, 128(4): 81-84.
- [8] 李晓丽. 基于 DCE 的可信 Web 服务度量模型的分析

与研究[J]. 科技通报, 2013, 29(8): 104-106.

作者简介:

高文莲 女,(1968—),硕士研究生,副教授. 研究方向为计算机网络和计算机应用.

高永强 男,(1974—),硕士研究生,讲师. 研究方向为数据库技术.