

基于机器学习的 Web 用户行为认证

毋泽南¹, 田立勤^{1,2}, 王志刚²

(1. 华北科技学院计算机学院, 北京 101601;
2. 青海师范大学计算机学院, 青海 西宁 810008)

摘要: 针对 Web 用户信息的安全问题, 结合机器学习的方法, 对用户行为进行分析和认证。首先通过主成分分析法对原始数据集做降维处理, 然后利用 SVM 算法, 让计算机对历史用户行为证据进行学习, 得到一个判别用户身份的模型。实际应用和理论分析表明, 该模型在用户行为认证判别上, 可以更准确和高效地分类出危险用户和可信用用户, 为诸如电子商务、网络金融等关键网络应用用户行为的高性能分析奠定坚实的理论和实践基础。

关键词: 信息安全; 用户行为认证; 支持向量机; 主成分分析

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2018011

Behavior authentication of Web users based on machine learning

WU Zenan¹, TIAN Liqin^{1,2}, WANG Zhigang²

1. School of Computer Science and Technology, North China Institute of Science and Technology, Beijing 101601, China
2. School of Computer Science and Technology, Qinghai Normal University, Xining 810008, China

Abstract: According to the security problem of Web user information, the user behavior was analyzed and authenticated by the method of machine learning. First of all, through the principal component analysis to reduce the dimension of the original data set, then use the SVM algorithm to allow the computer to learn the history of user behavior evidence, to get a model to identify the user's identity. The practical application and theoretical analysis show that the model in user behavior identification authentication, can be more accurate and efficient classification of dangerous users and trusted users, analysis lay a solid theoretical and practical basis for the high performance user behavior such as electronic commerce, network finance and other key of Internet applications.

Key words: information security, user behavior authentication, support vector machine, principal component analysis

1 引言

近年来, 互联网技术得到了迅猛发展, 社会

各行业以及人们的生活越来越离不开互联网。人们在享用互联网技术带来便利的同时, 也感受到互联网所面临的诸多挑战和问题。互联网的开放

收稿日期: 2017-11-07; 修回日期: 2018-01-05

通信作者: 田立勤, 18511465255@163.com

基金项目: 国家自然科学基金资助项目 (No.61472137); 中央高校基本科研业务费基金资助项目 (No.3142015022, No.3142017053); 青海省重点研发、应用基础研究基金资助项目 (No.2016-SF-130, No.2017-ZJ-752); 河北省物联网监控工程技术研究中心基金资助项目 (No.3142016020); 青海省物联网重点实验室基金资助项目 (No.2017-ZJ-Y21)

Foundation Items: The National Natural Science Foundation of China (No.61472137), The Central University Basic Research Fees Fund Project (No.3142015022, No.3142017053), The Key Research and Development Projects of Qinghai Province (No.2016-SF-130, No.2017-ZJ-752), Internet of Things Monitoring Engineering Research Center of Hebei Province (No.3142016020), Key Laboratory of Internet of Things of Qinghai Province (No.2017-ZJ-Y21)

性、动态性和共享性,使不法分子利用互联网危害社会,损害人们的利益。如何有效地保护用户信息、抵御黑客攻击以及预防电信诈骗已经成为当今社会关注的重点。

文献[1]中首次提出了可信网络的概念,主要阐述了对行为状况的监控、行为的分析与评估、异常行为的监管要求;但只是宏观上研究了用户行为可信的总体架构和信任评估机制等,并没有对此提出具体的研究方法。在用户行为的认证方面,文献[2]提出了云计算环境下的用户行为认证模型,该模型结合随机 Petri 网进行数学建模,缺点是对用户行为认证集分析得不够全面。文献[3]提出了权重均衡优化的 Web 用户行为的信任评估,通过将主观权重与客观权重相结合,得出集成权重,该方法有效地权衡了主客观对用户行为分析带来的影响。缺点是该评估方法没有对用户行为的历史证据集做降维操作,不利于对用户行为进行分析和认证,结果准确率较低。文献[4~8]建立了动态评估的用户行为模型,且通过分析大量用户行为证据来甄别系统中行为异常的用户,但是整体模型架构复杂度较高,且不易通过编程完成;同时,当用户行为证据的数量偏大时,整个模型的运算速度很慢,这会影响到用户行为评估的实时性,因此该类模型的运算性能不高,导致推广价值不高。

本文借鉴了上述相关研究成果,提出了基于主成分分析法和支持向量机(SVM, support vector machine)相结合的最优化机器学习模型。以笔者开发的用户行为信任评估平台系统所获取的真实数据为例,对 PCA-SVM 方法进行用户动态行为认证建模过程进行实验分析。通过多次实验计算和分析说明,相对于其他的用户行为认证方法,PCA-SVM 有更高的评估检测准确率和分类效率。

2 用户行为证据

2.1 用户行为证据的获得

用户行为证据通常是指用户进行一系列操作的具体表现形式,如何获取有效的证据是进行用户行为认证的前提。针对如何获得大量、可靠的用户行为数据,文献[9]给出了常用的几种方法。

1) PC 机上的入侵检测系统,如 Sguil 和

Tcpdump,可以用于网络安全分析,搜集实时的事件活动,获得用户访问次数等。

2) 利用网络流量分析工具,如 Bandwidthd,可以获取网关的各种协议,查看数据分组的传送速率。

3) 通过 Web 日志分析用户行为。

4) 基于 Ajax 的单击流捕获工具,可以获得用户动作信息,如用户对鼠标的一系列操作以及对网页的某些操作。

本文为了更好地验证用户行为认证模型的准确率,文中实验所用的数据来源于笔者开发的一个电子商务网站,该系统基于 JavaScript 和 form 表单提交,从用户登录直到退出,所有的一系列动作会被获取并提交到后台进行处理,然后存入数据库。

2.2 用户行为证据的规范化

由于所获得的用户行为证据的单位大多各不相同,如用户输入密码错误总次数、用户使用操作系统的历史出现率。因此,在进行建模前需要对搜集到的用户历史行为证据做进一步的处理,以便将所获得的用户行为数据转化到[0,1]范围内。本文采用了无量纲化的方法。文中用到的 17 种用户行为的证据类型可分为:数值型、百分比类型和布尔类型。

1) 由于百分比类型和布尔类型的数据本身就在[0,1]范围内,因此可利用式(1)进行处理。

$$\delta_{ij} = \begin{cases} a_{ij}, & \text{当 } a_{ij} \text{ 为安全型证据} \\ 1 - a_{ij}, & \text{当 } a_{ij} \text{ 为非安全型证据} \end{cases} \quad (1)$$

其中, a_{ij} 表示用户第 j 次行为的第 i 项行为证据在规范化处理之前的数值, δ_{ij} 表示规范化处理后的数值。安全型证据表示某个用户对系统进行访问时的操作是否规范,是否跟以往的行为一致。例如,用户使用计算机操作系统的出现率,该数值越大越好,表示用户符合日常的行为习惯;非安全型证据表示用户不按照以往的行为习惯进行操作或者非法操作,如用户购物总值超过系统设定的阈值,则为 1,未超过为 0,该值越小越好。这样,式(1)可以很好地处理 2 种不同证据类型的差别。

2) 整数类型的用户行为证据包括用户输入退格数、输入用户名或者密码错误的总次数、购物下载图书量等,针对这类证据集,使用式(2)进行规范化处理。

$$\delta_{ij} = \begin{cases} \frac{a_{ij} - a_{ij\min}}{a_{ij\max} - a_{ij\min}}, & \text{当 } a_{ij} \text{ 为安全型证据} \\ \frac{a_{ij\max} - a_{ij}}{a_{ij\max} - a_{ij\min}}, & \text{当 } a_{ij} \text{ 为非安全型证据} \end{cases} \quad (2)$$

同样, δ_{ij} 和 a_{ij} 与之前的表述一致, $a_{ij\min}$ 和 $a_{ij\max}$ 分别表示该类型用户行为证据集中的最小值和最大值。

3) 原始数据无量纲化的过程。以“用户输入密码的错误次数”为例,最近一次的行为证据中,得到“用户输入密码的错误次数”为 2,用户行为证据中,该证据的最大值为 3,最小值为 0,该证据值为非安全型证据,根据式(2),可以很容易计算出该证据的无量纲化值为 0.333;依次类推,可以得出其他证据的无量纲化值。

3 PCA-SVM 用户行为信任评估

3.1 PCA 对用户行为证据进行降维

用户行为证据具有较高的维数,本文选取了 17 种,为了提高用户行为信任评估的准确率和效率,本文采用 PCA 对用户行为最初的样本数据做降维处理,提取与分析结果相关的重要特性。PCA 基本原理是首先对原始数据维属性矩阵进行处理,并在空间中进行坐标旋转。通过 PAC 可以使原始样本在不改变数据结构情况下,得到用户行为数据的主成分线性组合,且两两独立相关,这样就可以更好地将原始数据中的有效成分保留下来。下面介绍 PCA 在用户行为证据集中应用的主要步骤。

设 $N=17$ 为用户行为证据集的维数, M 为机器学习训练集的样本总量。

1) 将原始数据按列组成 N 行 M 列的矩阵 X 。

2) 将 X 的每一行(代表用户行为证据中某一属性字段)进行零均值化。

3) 求协方差矩阵 $C = \frac{1}{M} (X_{ij} - u_i)(X_{ij} - u_i)^T$;

其中, X_{ij} 为用户行为证据集中的某一个具体元素; $i=1,2,\dots,N$, $j=1,2,\dots,M$, u_i 为样本某一行为证据特性的均值。

4) 求出协方差矩阵的特征向量和特征值。

5) 协方差矩阵对角化;即使对角化后的矩阵除对角线外,其他元素均为 0,而且对角线上的元素按大小从上到下排列;取前 k 行组成矩阵 P 。

6) $Y = PX$ 即为降维得到的 k 维数据。

本文选取了 17 种用户行为证据集,通过上述计算,最终确定了 6 种主成分。

3.2 支持向量机

支持向量机是一种监督式学习的分类算法,是 Vapnik^[10]于 1995 年首先提出的,多用于统计分类以及回归分析,其优势在于解决小样本、高维模式识别以及非线性的分类。同时,它能够很好地应用推广到函数拟合等其他的机器学习中。综合以上这些优点,本文选用支持向量机对用户在互联网环境中的行为进行研究。

对于线性分类问题,设训练集为 $\{x_i, y_i\}$, x_i 为用户行为证据样本点; y_i 是第 i 个样本的标签, y_i 取 +1 或 -1,当 $y_i = 1$ 时,表示正常用户;当 $y_i = -1$ 时,表示非法用户;其中 $i=1,2,\dots,n$,那么区别两类用户的最优超平面为

$$f(x) = \langle wx \rangle + b$$

其中, w 表示法向量, b 为截距;另外, w 的数学语言表示为样本的某种组合,即为

$$w = \sum_{i=1}^n (a_i y_i x_i), \text{ 其中 } a_i \text{ 是拉格朗日乘子。}$$

如图 1 所示,从直观上希望正常(可信)用户和非法(不可信)用户分得越开越好,即图 1 中 $G(x)$ 和 $H(x)$ 与 $F(x)$ 之间的几何间隔越大越好。这是因为距离分类直线(超平面)越近的样本(可疑用户),分类的置信度越低(实际上几何间隔代表了分类器的误差上界)。SVM 的优化目标,就是最大化最接近分类超平面的点到分类超平面的距离,即将正常用户和非法用户最大限度地分离开。

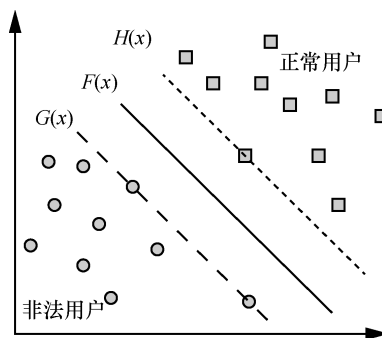


图1 SVM 线性分类

图 1 描述的是线性可分的情况,可大部分时候,待处理的数据并不完全是线性可分的,如

本文用到的用户行为证据集。这个时候很难找到一个合适的分类超平面，将危险用户和可信用户完全分离开。对于这种情况，支持向量机提供了一个核函数 $K(x, y)$ 来处理这个问题，该方法可以将数据映射到高维空间，这样就解决了线性不可分的问题。原始函数 $f(x)$ 就转变为

$$f(x') = \sum_{i=1}^n a_i y_i K(x'_i, x') + b.$$

该方法的关键在于如何选取合适的核函数，一般应用的核函数主要有以下4种。

- 1) 线性核函数: $K(x, y) = xy$ 。
- 2) 多项式核函数: $K(x, y) = (xy + 1)d$ 。
- 3) 径向基函数: $K(x, y) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right)$ 。
- 4) Sigmoid 核: $K(x, y) = \tanh[c_1(x, y) + c_2]$ 。

综上所述，SVM 分类的最终目标就是找到一个最优分类超平面，且这个超平面使2个类别之间的间隔最大。要使这个间隔最大，即求解式(3)所示问题。

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (3)$$

$$\text{s.t. } y_i(w x_i + b) \geq 1 - \xi_i; i = 1, 2, \dots, n, \xi_i \geq 0$$

其中， C 表示惩罚因子，它表示对离群点的重视程度， C 越大表示越重视这些离群点； ξ 表示松弛变量，表示对应点到离群点的距离。在特征数目非常大的情况下，可以将此二次规划问题转化为其对偶问题，通常根据拉格朗日对偶理论，求解其对偶问题式。在线性可分的情况下，其对偶问题表示为

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (4)$$

$$\text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n$$

对于非线性可分的情况，只要令 $x_i x_j = K(x_i, x_j)$ 即可。另外， α_i 为式(3)和式(4)中每个样本对应的 Lagrange 乘子。因此，可看出这是一个不等式约束条件下的二次函数求解最优问题，存在唯一解，且解中将只有很小一部分不为零，符合该条件下所对应的样本即为支持向量。通过对式(3)和式(4)进行求解，就可以得到最优分类函

数，如式(5)所示。

$$\varphi(x) = \text{sgn}(wx + b) \quad (5)$$

本文采用 RBF 作为支持向量机的核函数，所以最终的决策函数为

$$\varphi(x') = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) + b\right) \quad (6)$$

总体来说，对于线性不可分的情况，支持向量机通过选取比较合适的核函数，将输入样本空间映射到高维空间，这样就可以在高维特征空间中构造出用于分类的最优超平面。因此，SVM 不仅可以在平面上对数据进行分类，而且可以对本身不好分的非线性数据进行分类。如图2所示，样本数据在二维空间平面上无法进行有效划分，通过 SVM 的核函数，映射到三维空间中进行分类。

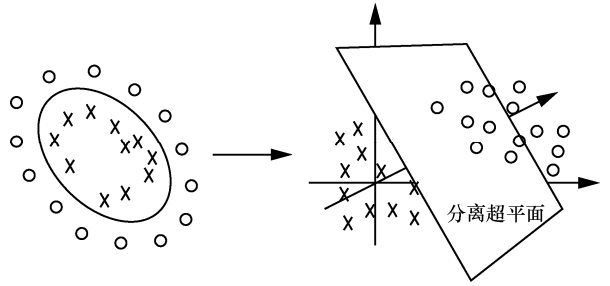


图2 SVM 非线性分类

3.3 SVM 分类的模型选择

关于 SVM 分类模型的选择，目前还没有太多的经验与实力可供参考，因此只能通过大量实验分析及确定。

3.3.1 数据格式转换

本文使用的实验工具是 LIBSVM 包。它是由林智仁博士等^[11]开发的一款开源软件，是目前最流行而且分类效果较好的一款软件，并因其操作简单、效率高、简单实用等特点而被广泛使用。数据来源于开发的一个网上购书系统平台。此外，LIBSVM 对数据格式有统一要求，因此在实验前，需要将原始数据转化成 LIBSVM 要求的数据格式，使最终数据的每一行对应一组用户行为证据特性，每一个向量对应用户行为证据值，用户行为的分类标识（-1 或 1）位于每行最前面，接着是数据对，形如 `key:value`。本文使用宏命令 `FormatDataLibsvm.xls` 将数据转成 LIBSVM 格式。

3.3.2 RBF 核函数参数的确定

由于核参数 σ 和惩罚参数 C 的选取,会直接影响模型的稳定性和泛化能力的强弱,为防止过拟合等情况,通过交叉验证法和网格法寻求核参数和惩罚因子的最优值 C^* 和 σ^* 。使用 C^* 和 σ^* 对样本数据进行训练,同时对训练好的模型进行检验。

4 用户行为认证架构

目前,在认证用户行为时,大部分情况下都是使用用户最近一次的用户行为所产生的数据进行评价,这种方法所得出的结果是不合理、不可靠的,因为用户每次的行为是习惯的一种具体表现,而习惯又是一个人长久养成的生活方式,如果用户的行为出现异常,那么该异常应该体现在需评估用户行为与习惯用户行为的差异上,因此需要综合历史用户行为和需评估用户行为。

使用支持向量机进行用户行为认证的整体过程主要分为:训练学习阶段和实际预测阶段,具体实现过程包括如下几个步骤(如图3所示)。

1) 数据的收集和预处理。本文中实验用到的数据来源于笔者开发的一个用户行为监测系统,该系统是基于 Java 实现的 Web 系统,由 JSP 页面和后台控制程序组成,其主要的功能有图书的借阅、查询、购买、下载和用户个人信息管理等。这些用户行为证据可根据 2.2 节进行规范化处理操作。

2) 主成分提取。本文采用了 PCA 对用户行为的原始样本数据做降维处理,以便获取与检测结果相关的重要行为特征。

3) SVM 的输入与输出。本文对用户行为认证的训练是以用户访问的历史行为数据作为 SVM 的输入,而输出对应的则是用户行为认证结果。通过寻找输入域输出的依赖关系更加准确地预测用户的合法性。

4) 参数调优。选取合适的参数,对训练模型构建起着关键作用,但由于支持向量机中包含多个参数,而且目前关于参数的选择并没有非常有效的方法,更多的是根据平时的经验,首先大致判断出一个范围,然后进行多次训练。

5) 构建评估模型。根据 3.2 节选择不同的核

函数,结合训练样本进行系统训练,通过对比分析,获得最优的用户行为认证模型。

6) 对测试样本进行预测与分类。通过使用 SVM 构造的模型对测试样本进行预测和分析,并输出用户行为认证的结果。

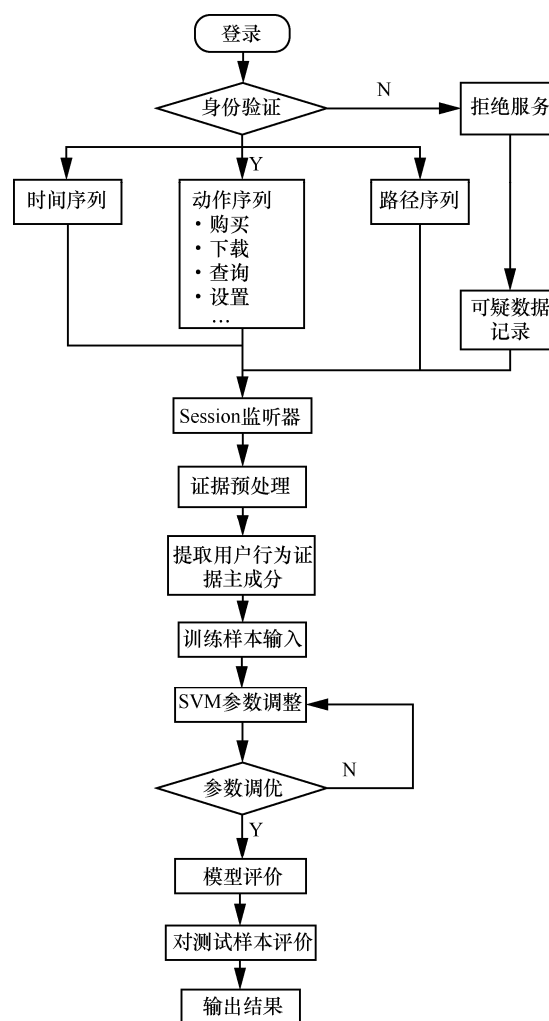


图3 用户行为认证流程

5 实例应用效果分析

5.1 用户行为认证的理论计算实例

本文实验数据来源于自己开发的一个电子图书商务网站,用户可以在上面进行购买、下载、查询等操作,这样系统后台获得的数据较为真实,本文选取 17 种用户行为证据作为研究,分别为: 1) 用户所用操作系统类型的出现率; 2) 用户访问所用浏览器历史出现率; 3) 计算机操作系统版本出现率; 4) 用户 IP 历史出现率; 5) 用户计算机屏幕分辨率历史出现率; 6) 用户登录时所在地

的历史出现率; 7) 输入用户名时的退格数; 8) 输入密码时的退格数; 9) 输入用户名错误总次数; 10) 输入密码错误总次数; 11) 使用银行账号历史出现率; 12) 购买图书总次数; 13) 下载图书总流量; 14) 下载图书总次数; 15) 访问敏感服务总次数; 16) 用户退出时的信任值; 17) 登录时间信任值。

本文根据开发的用户行为评估系统所获得的证据数据, 选出其中 3 000 个已知样本, 为了检验模型预测分类的有效性, 本文随机选取其中的样本作为训练数据和测试数据。采用 SVM 进行机器学习, 并通过网格法和交叉验证的方法对 SVM 的参数进行优化选择, 最后对测试样本进行用户的预测分析。通过选取不同的参数进行实验, 模型的误报率 ρ 和检测率 ν 如表 1 所示。

$$\rho = \frac{\eta}{\mu} \times 100\%$$

$$\nu = \frac{l}{\tau} \times 100\%$$

其中, η 表示被误报为可信的危险用户样本数, μ 表示可信用用户样本数, l 表示检测到的危险用户样本数, τ 表示危险用户样本总数。

表 1 模型的误报率和检测率

训练样本数	测试样本数	C	$g = \sigma^2$	误报率	检测率
1 000	200	100	1	2.35%	81.45%
2 000	300	50	0.5	2.06%	88.53%
3 000	400	200	0.1	1.68%	89.25%

5.2 用户行为认证的算法对比分析

图 4 和图 5 分别表示本文所提出的算法与其他几种常用算法的检测误报率对比和检测成功率对比。其中, 误报率是指认证成功但实际为危险用户占可信用户的比率。因此, 在用户行为认证中, 误报率是一项重要的性能指标, 如果误报率过高, 则会影响系统对用户的判定。本文算法的中心思想是综合 PCA 和 SVM 对用户行为数据进行认证与预测, 采用 PCA 算法可以很好地消除用户行为证据集中的重复信息, 有效降低原始数据的维数, 增强证据间的耦合性, 有助于提高实验检测效率。这也证明, SVM 技术可以很好地适用于用户行为认证中。

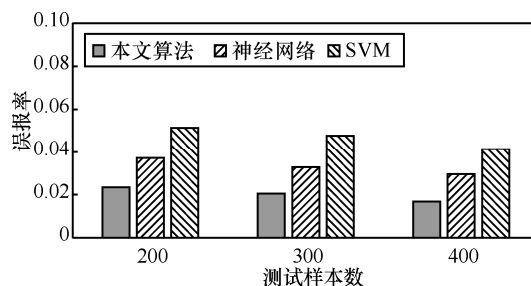


图 4 误报率比较

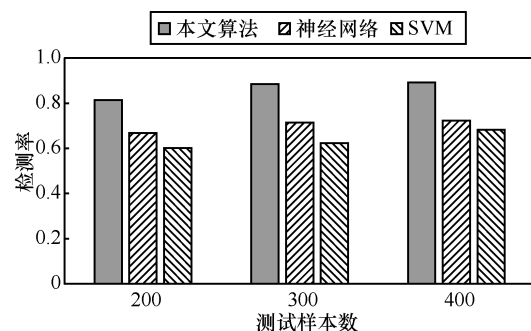


图 5 检测率比较

从图 4 和图 5 的算法比较来看, 本文提出的 PCA-SVM 算法的误报率要好于其他 2 种算法, 说明本文算法中 SVM 的算子和参数选择比较成功。同时, 与神经网络相比, 本文算法的优点在于, 它不仅训练速度较其他 2 种算法更快, 而且当用户行为证据集发生小的变动时, 它并不需要变更算法中的算子和参数, 这样分类效率更高。

6 结束语

本文通过笔者开发的用户行为认证平台, 结合 LIBSVM 工具进行实验, 通过多次实验分析, 选择了 PCA 与 SVM 相结合作为用户动态行为认证的分类器。为了降低实验的工程量, 在确定 SVM 的核函数和核参数时, 首先进行了理论分析与比较, 先确定出参数的大概取值范围, 再通过交叉验证法进行选择。与传统的用户行为认证不同, 本文利用机器学习的方法, 通过训练模型, 使计算机自动对用户行为进行分类, 与传统方法相比, 很好地提升了分类的效率。

实验结果表明, 本文提供的基于 PCA-SVM 的分类器对用户行为认证具有较好的表现, 同时具有较高的检测率。接下来的研究重点是对用户行为认证进行更深一步的学习, 尤其在核函数的参数设置上, 希望能找到最优的核函数, 并通过

与更多的算法进行比较分析, 完善和改进 SVM 算法。

参考文献:

- [1] 林闯, 田立勤, 王元卓. 可信网络中用户行为可信的研究[J]. 计算机研究与发展, 2008, 45(12): 2033-2043.
LIN C, TIAN L Q, WANG Y Z. Study on trusted users' behavior in trusted network[J]. Journal of Computer Research and Development, 2008, 45 (12): 2033-2043.
- [2] 陈亚睿, 田立勤, 杨扬. 云计算环境下动态用户行为认证的机制、模型与分析[J]. 系统仿真学报, 2011, 23(11): 2302-2307.
CHEN Y R, TIAN L Q, YANG Y. Mechanism, model and analysis of dynamic user behavior authentication in cloud computing environment[J]. Journal of System Simulation, 2011, 23 (11): 2302-2307.
- [3] 田立勤, 李君建, 毋泽南. 权重均衡优化的 Web 用户行为的信任评估[J]. 北京邮电大学学报, 2016, 39(6): 99-103.
TIAN L Q, LI J J, WU Z N. Trust assessment of Web user behavior based on weight balance optimization [J]. Journal of Beijing University of Posts and Telecom, 2016, 39 (6): 99-103.
- [4] 郭树凯, 田立勤, 沈学利. FAHP 在用户行为信任评价中的研究[J]. 计算机工程与应用, 2011, 47(12): 59-61.
GUO S K, TIAN L Q, SHEN X L. Research on FAHP in user behavior trust evaluation[J]. Computer Engineering and Application, 2011, 47 (12): 59-61.
- [5] LI W, PING L D, LU K J, et al. Trust model of users' behavior in trustworthy Internet[C]//2009 WASE International Conference on Information Engineering, Piscataway. 2009: 403-406.
- [6] LIU W, REN P, LIU K, DUAN H X. User cooperation trust model and its application in network security management[C]//2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). 2011: 2335-2339.
- [7] BROSSO I, NEVE A L, BRESSAN G, et al. A continuous authentication system based on user behavior analysis[C]// Ares '10 International Conference on Availability, Reliability, and Security. 2010: 380-385.
- [8] NOOR T H, SHENG Q Z, ALFAZI A. Reputation attacks detection for effective trust assessment among cloud services[C]// IEEE International Conference on Trust, Security and Privacy in Computing and Communications. 2013:469-476.
- [9] 冀铁果, 田立勤, 等. 可信网络中一种基于 AHP 的用户行为评分方法[J]. 计算机工程与应用, 2007, 43 (19): 123-127.
JI T G, TIAN L Q, et al. A AHP-based user behavior assessment method in trusted network[J]. Computer Engineering & Applications, 2007, 43 (19): 123-127.
- [10] 万普尼克. 统计学习理论[M]. 北京: 电子工业出版社, 2009.
VAPNIK. Statistical Learning Theory[M]. Beijing: Publishing House of Electronics Industry, 2009.
- [11] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[C]//ACM Transactions on Intelligent Systems and Technology(TIST). 2011.

[作者简介]



毋泽南 (1991-), 男, 河南焦作人, 华北科技学院硕士生, 主要研究方向为信息安全、用户行为认证。



田立勤 (1970-), 男, 陕西定边人, 博士, 华北科技学院教授, 主要研究方向为物联网远程信息监控、大数据有效性审核、网络安全评价与用户行为认证、网络性能评价与优化。



王志刚 (1993-), 男, 河北张家口人, 青海师范大学硕士生, 主要研究方向为大数据、云计算。