10.3969/j.issn.1671-489X.2014.22.050

基于鼠标键盘行为的身份认证*

◆王鑫　王梦莹　梁金

摘　要 安全的身份认证是保证计算机及网络系统安全的基本前提，通过研究计算机输入行为特征，提出并实现基于鼠标键盘行为的身份认证系统。该系统运用优化的SVM算法分析用户信息，结合鼠标和键盘双指标复合认证。系统设计了友好的训练场景，明显缩短了训练和认证的行为数据采集时间。实验证明，双指标复合认证与单指标相比较而言，准确性有了较大提高，系统误识率和拒识率都有所降低。

关键词 支持向量机；行为特征；身份认证；双指标

中图分类号：TP393　　**文献标识码**：B

文章编号：1671-489X(2014)22-0050-04

Identity Authentication based on Mouse and Keyboard Actions//
WANG Xin, WANG Mengying, LIANG Jin

Abstract The security of identity authentication is the basic premise of the safety of computer and network system. This paper proposed and realized the identity authentication system based on mouse and keyboard actions by studying the trait of behavior, using the optimized SVM algorithm to analysis user information, adapting the mouse and keyboard double indicator compound authentication. The system shortened the time of training and acquisition certification of behavior data greatly by designing a friendly training scene. Experiments proved that the double indicator compound authentication accuracy had a larger increase compared with single index, and decreased the system course rate and false reject rate.

Key words SVM; trait of behavior; identity authentication; double indicator

1 引言

身份认证是计算机安全、网络安全的第一道防线。长久以来，人们一直在研究这方面的技术，试图寻找一种安全的、可靠的、可行的和简易的身份认证方式来满足安全需求。目前市面上应用比较广泛的身份认证方式有三种[1]：1）传统字符密码；2）生物特征认证（指纹识别、虹膜识别等）；3）辅助设备（加密狗、口令卡等）。从安全性、方便性、廉价性分析，这些身份认证方式或多或少都存在不足。如传统的字符型密码容易被间谍软件窃取、密钥存储问题、容易被猜出；生物特征识别（如指纹、虹膜识别）需要添加额外的设备，这会增加额外的成本开销；使用辅助设备来进行身份识别（如口令卡、加密狗等），如果这些辅助设备丢失，就无法完成身份认证。因此，研究人员仍然不断在寻找新的认证方法。

2 相关工作

基于计算机输入行为特征的认证方法，具有鲜明的特点，如：行为方式难以模仿；行为方式无需记忆；行为数据数据量多，行为密码不具有明显的特征；无需额外设备，并且在当前的大多数计算机系统中可以直接部署。正是因为这些特点，使得这种认证方法逐渐成为研究的热点[2-3]。

由于基于计算机输入行为特征的认证的优点，国内外学者在此基础上做了很多研究。从2003年开始，国外有四个小组对鼠标的使用行为特征进行了初步的研究[4-8]，内容多为基于统计的鼠标行为特征。文献[9]提出一种基于鼠标运动的用户再认证方案，他们提出对于每个用户的每个请求都用决策树分类器建立不同的模型。文献[10]提出并实现设计一个为击键特征识别系统设计的压力感应键盘，将击键时间和击键压力融合。文献[11]提出基于实时击键序列的主机入侵检测，该方法能独立于用户名和口令之外对用户进行身份认证。但是这些都仅基于击键或是单考虑鼠标的行为特征，鲜有文章综合鼠标、键盘两者的行为进行身份认证。

本文提出并实现基于鼠标键盘行为的身份认证系统。系统通过常用的输入设备——鼠标和键盘采集用户训练和认证的行为数据，采用优化过的SVM（Support Vector Machine，支持向量机）算法结合鼠标、键盘双指标对登录操作系统的用户进行身份认证。系统采用灵活的结构布局，允许用户在系统原有部件的基础上进行二次开发，有很好的移植性。实验证明，本系统采用的双指标复合认证有较高的准确率。

3 鼠标键盘行为特征

基于鼠标、键盘行为的身份认证系统中有一个基本假

*基金项目：国家级大学生创新创业训练计划项目资助（201210595004）。

作者：王鑫，副教授，桂林电子科技大学海洋信息工程学院副院长，桂林电子科技大学计算机科学与工程学院，硕士研究生导师，主要研究方向为网络与信息安全；王梦莹，桂林电子科技大学计算机科学与工程学院，主要研究方向为无线传感器网络；梁金，桂林电子科技大学计算机科学与工程学院，主要研究方向为数据库系统（541004）。

设：对每个用户而言，其鼠标、键盘操作都存在与其他用户显著不同的模式。每个用户由于个人原因可能造成在使用鼠标、键盘时有不同的习惯，比如鼠标轨迹、鼠标滚轮操作速度、鼠标左右键的使用习惯、击键时间间隔和击键迟延时间。不同的用户的操作行为有较大的区别，如不同的用户在移动鼠标时的力度以及准确定位的能力不同，或是在点击键盘时一次按键到下次按键之间的时间间隔习惯不同。鼠标、键盘的行为特征即是指用户操作鼠标、键盘的习惯。

4 优化的 SVM 算法

SVM 算法在解决小样本、非线性极高危模式识别中表现出许多特有的优势，主要解决的是两类问题。先来看下面的目标函数：

$W*X+b=0$ （1）

显然这是一条直线，在确定的 W 值和 b 值下，只有唯一的 X 值可以满足（1）式，其他的 X 值带入后，要么大于零，要么小于零，这样其他的 X 值便被分成两类，不可分的是满足（1）式的 X。

以上 W 和 X 值是一维的情况，SVM 在此基础上给这参数赋予全新定义，将原先的 W 值和 X 值扩展到多维，得到如下目标函数：

$X=(x_1, x_2, x_3, \ldots\ldots, x_n)$ $n=1,2,3\ldots n$ （2）

$Y=(y_1, y_2, y_3, \ldots\ldots, y_n)$ $n=1,2,3\ldots n$ （3）

$A=(a_1, a_2, a_3, \ldots\ldots, a_n)$ $n=1,2,3\ldots n$ （4）

$W=y_1*a_1*X_1+y_2*a_2*X_2+\ldots+y_n*a_n*X_n$ $n=1,2,3\ldots n$ （5）

$\langle W, X\rangle+b=1-C*E$ C 和 E 为实数 （6）

这里定义了两个新值 Y 和 A，同时也不难看出 X 和 W 都变成了多维，（6）式相当于原来的（1）式，将不满足（6）式的多维 X 分成两类。

用户定义为一个 n 维的向量 X，取值为用户鼠标或键盘操作特征，如用户 =（左键单击时间 $t1$）。假设有两个用户甲和乙。甲有三组鼠标键盘特征样本 X_1，X_2，X_3，一组样本就对应一个式（2），如 $X_1=(x_1, x_2, x_3)$，$X_2=(x_1, x_2, x_3)$，$X_3=(x_1, x_2, x_3)$。应该注意不同样本中 x_n 值不一定相等，但都是同一类型的值（如鼠标或键盘特征），而且括号中的数据个数必须相等。至于数据个数为自定义，比如增加一个双击数据记为 x_4，则 $X_1=(x_1, x_2, x_3, x_4)$，$X_1=(x_1, x_2, x_3, x_4)$，$X_1=(x_1, x_2, x_3, x_4)$。同理，乙有三组鼠标键盘特征样本 X_4，X_5，X_6。甲乙样本数可以不等，但样本中的鼠标或键盘特征必须一一对应，现在总样本数为 6。

再来看（3）（4）式，由于总样本为 6，因此，（3）（4）式子括号中的 n 为 6，即由样本总数确定，（3）（4）括号中的值要与 6 个样本一一对应，如 X_1 对应 Y 中的 y_1 以及 A 中的 a_1，其他样本以此类推，现将以上对应关系整理如下：

甲：$X_1=(x_1, x_2, x_3)$ y_1 a_1

$X_2=(x_1, x_2, x_3)$ y_2 a_2

$X_3=(x_1, x_2, x_3)$ y_3 a_3

乙：$X_4=(x_1, x_2, x_3)$ y_4 a_4

$X_5=(x_1, x_2, x_3)$ y_5 a_5

$X_6=(x_1, x_2, x_3)$ y_6 a_6

至于 Y 中的值，做这样的处理：对应甲的所有 y 都赋 1，对应乙的为 -1，也可以反过来，即 $y_1=y_2=y_3=1$，$y_4=y_5=y_6=-1$；或者，$y_1=y_2=y_3=-1$，$y_4=y_5=y_6=1$。A 中的值是在用户训练时程序对用户各样本组计算后自动产生的。本例以用户甲的 y 为 1 作为标准。

再来看(5)式，发现将以上整理的甲乙关系数据代入(5)式右边，便可得出左边的 W 值了。(5)式便是认证的式子，(6)式中的 W 便是（5）式的 W，b 为用户训练时程序自动产生，C 和 E 为本算法自定义值，X 代表待认证用户的一组新样本。注意：这里的待认证用户只能是决定 W 的两个用户，这是因为（6）式和（1）式一样，解决的是两类问题，如本例中 W 由甲乙的样本而得来，则只能认证甲乙，之后数据都带入（6）式，大于式子左边，认证通过。

那么，要对所收集来的行为数据进行整理，并把它们转化为符合算法要求的格式。由于 SVM 算法是解决两类样本的分类问题的，因而只用一条分类线可以区分两个样本。对于多用户的区分，采取这种方式判定当前用户的身份。如对用户 1 进行判定，且当前训练过的用户样本有 6 个时，那么用户 1 的训练样本与其余 5 个用户的训练样本分别进行分类线计算，然后用这 5 条分类线进行样本判定。这里需要注意的是：训练样本必须保持奇数个，若当前用户训练量不足奇数个，系统会自动调用存根样本去补足。若分类线判定时投给用户 1 的票数超过一半（一条分类线相当于一次投票），那么就判定为 1 用户。

算法的优化情况：现在再将以上（2）～（6）式的参数值确定情况整理如表 1 所示。由表 1 看出，可以在这些参数中作改变的为 X（上面已讲过 X 可自定义，即鼠标指标数自定义且鼠标指标自定义）、C（程序自定义具体值），而剩下的参数要么固定，要么由其他参数算出。

5 基于鼠标键盘行为的身份认证系统

系统概述 系统对收集到的用户鼠标键盘行为信息进

表 1 各式子参数值确定情况表

	X	y	a	W	b	C	E
值确定情况	由用户鼠标数据直接确定，为已知条件	根据 X 对应用户确定 1 或 -1，是定值，为已知条件	将对应 X，y 带入程序自动生成，值不确定	将所有 X，y，a 带入程序自动生成，值不确定	将所有 X，y，a 带入程序自动生成，值不确定	由程序自定义	固定为 0.01
对应式子	（2）（5）（6）	（3）（5）	（4）（5）	（5）（6）	（6）	（6）	(6)

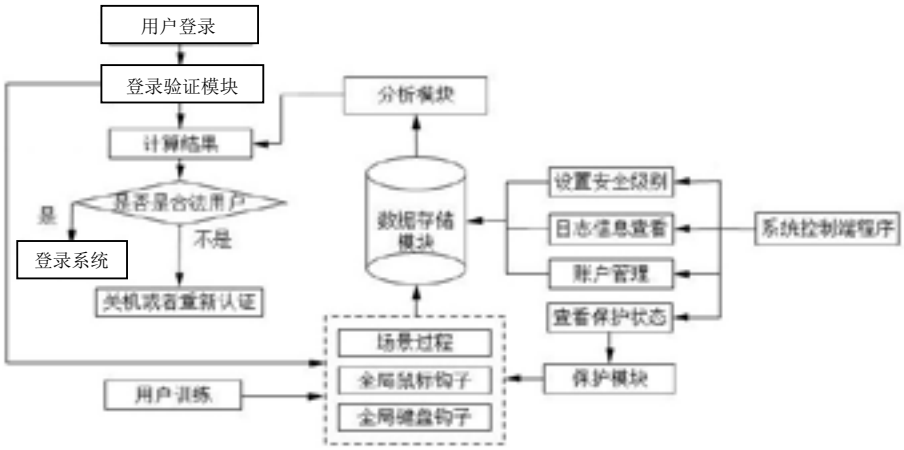

图 1　系统工作原理图

行处理、保存，当有用户登录系统时，从数据存储中分析用户是否合法，从而达到用户身份认证。并且根据数据库提供的各种安全措施来保障数据的安全。目前该系统已经封装成 DLL，系统工作原理图如图 1 所示。

本系统有两个主要过程，分别是训练模块和认证模块。训练和认证都必须在特定的场景内进行，每一个场景都对应于一个特定的鼠标或键盘的行为指标，如左键单击、左键双击、右键单击、滚轮等，系统共提供 9 个场景用于训练和认证。用户必须根据场景给出的提示完成指定的操作，对于非指定操作、认证或训练，系统将会自动屏蔽对应的行为数据，并且用户无法完成当前认证或训练。由于 SVM 算法的特性，系统可以根据用户当前的训练数据给出一个评估，即根据当前训练的数据，对在进行认证时通过的可能性进行估计。用户可以根据此评估结果决定是否再次进行训练。训练状态下，用户除了完成全部场景的训练任务，还需要赋予当前账户的权限和设置认证时所需的安全级别。认证状态下，用户需要提供自己账户的账号和密码，并根据提示完成对应安全级别的认证场景过程。训练产生数据的计算、认证数据的分类判断，全部由优化过的 SVM 算法完成。系统软件流程图如图 2 所示。

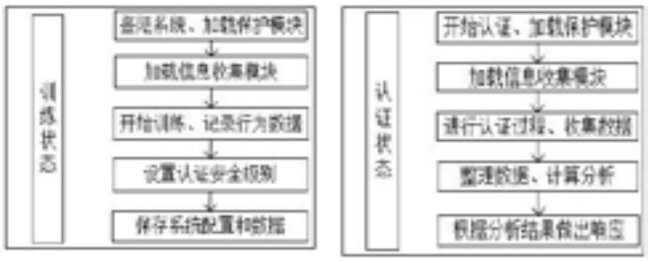

图 2　系统软件流程

用户信息收集和分析　本系统共提供了 9 个场景用于用户的训练和认证。每一个场景的设计都是建立在一定的背景之上的，避免了使用过程中的枯燥和复杂。它们分别是：爷爷的大餐（双击图片）、按点划线（鼠标轨迹相关）、弟弟的数数测试（单击图片）、力量训练（鼠标滚轮操作）、妹妹的购物（右击图片）、口算训练（键盘按键相关指标）、宠物市场大抢购（鼠标轨迹相关）、我的密码手势（轨迹相关）、传统密码字符（击键特征）。每一个场景都对应一个或多个特定的行为指标，在场景过程中，用户必须按照场景给出的提示完成相应的操作，对于非指定操作，场景将给出操作失误的提示，该操作所对应的行为数据也将被丢弃。场景指标都是经过测试后选出来的。用这些指标进行身份认证时，需要的样本数据量小，自然数据收集时间也相应的减少，这样在训练和认证的过程中就省了不少的时间。所选出的这些指标，在进行身份判别测试中，它们的效果也是不一样的。因此，用安全级别对认证时这些指标的使用进行划分归类：对于效果比较好的指标，将归入高的安全级别；对于效果一般的指标，将归入一般或较低的安全级别。这样在进行认证时，用户只需通过预先设定好的安全级别所对应的场景，而不需要通过全部的场景，相比训练时间又减少了不少。

通过使用特定的场景环境，再结合全局的键盘钩子（WH_KEYBOARD_LL）和全局鼠标钩子（WH_MOUSE_LL），可以精确地收集到想要的行为数据。在场景环境中，用户需要根据场景所给出的提示完成一系列的操作，如单击鼠标左键、单击鼠标右键、双击鼠标左键、滚动鼠标滚轮、移动鼠标光标、输入预定的密码字符串等。在场景中，单场景中的每一步操作、场景间的切换都具有可控性，所以可以很精确地定位用户当前时间段所产生的行为数据是什么，进而可以对这些行为数据进行分类整理，并按计算分析模块的数据格式要求进行初步的整理。

用户行为数据的分析是整个系统的核心。采用前文所描述的优化的 SVM 算法对收集到的用户行为数据进行处理，然后对登录操作系统的用户进行身份认证。

系统特点　本文讨论的基于鼠标键盘行为的身份认证系统具有以下优点。

1）训练时间短。系统在收集用户行为数据时根据特定

场景引导用户操作，相应地减少了训练的时间。

2）精确性。与单指标认证相比，本系统采用的是双指标复合认证，认证的准确性有了较大的提高，系统误识率和拒识率都有所降低。

3）适应性。系统的各功能模块封装成 DLL，在遵循一定的调用规则的前提下，可以在原有模块的基础上进行二次开发。

6 性能测试

随机选取 6 名用户，让他们分别进行全场景训练操作，收集他们的行为数据。然后使用这些行为数据，进行规定次数的分指标认证，记录相应的认证结果并对该结果做出分析。

由于 $C > 32$ 后分类结果不好，因此，本工程只考虑 C 大于等于 1 且小于等于 32 的情况。现工程共定义 5 个指标：左键单击时间、左键双击时间、右键单击时间、滚轮移动速率、击键时间。这些指标共有 31 种组合方式，下面给出部分组合的测试数据，如表 2、表 3 所示。

表 2　单指标准确率（X 指标数 =1，C=1）

指标	本用户正确识别率	本用户拒识率	其他用户正确拦截率	其他用户误判率
左键单击	80%	20%	65%	35%
左键双击	61%	39%	77.3%	33.7%
右键单击	81.6%	18.4	72.3%	27.7%
滚轮移动	80%	20%	58.9%	42.1%
击键时间	71%	29%	70.1%	29.9%

表 3　单指标准确率（X 指标数 =1，C=32）

指标	本用户正确识别率	本用户拒识率	其他用户正确拦截率	其他用户误判率
左键单击时间	68.3%	31.7%	73.6%	26.4%
左键双击时间	61.6%	38.4%	80.6%	19.4%
右键单击时间	81.6%	18.4	68.3%	31.7%
滚轮移动速率	96.6%	3.4%	55.3%	44.7%
击键时间	86%	14%	60%	40%

通过表 2 和表 3 结果的比较，不难看出本用户正确识别率和其他用户正确拦截率在 C 值变化下成此消彼长的关系，经过 C 值变化结果的分析，C 值设为 1 为最优。

X 维数和指标对实验结果的影响比较复杂，有太多可能性。从实际对 31 种组合的测试结果中选择最优的组合，最后选择的是：X 指标数 =2，指标是右键单击时间和击键时间，C 值为 1，测试结果如表 4 所示。

表 4　双指标准确率（X 指标数 =1，C=32）

指标	本用户正确识率	本用户拒识率	其他用户正确拦截率	其他用户误判率
右键单击时间和击键时间	91.6%	8.9%	80.6%	19.4%

从表 2、表 3、表 4 中可以看出，在双指标时系统对用户身份认证有较好的效果。实验证明，与单指标认证相比，系统采用的双指标复合认证的准确性有了较大的提高，系统误识率和拒识率都有所降低。

7 结束语

本文提出并实现一种基于鼠标、键盘行为的身份认证系统，通过实验验证结合鼠标、键盘两种指标的行为特征进行身份认证，效果更好。系统通过特定场景缩短用户训练时间，并采用优化的 SVM 算法实现用户信息分析，同时该系统具有较好的移植性。接下来的工作放在提高认证的准确度方面，将继续对鼠标、键盘的行为数据进行深入分析和测试，寻找效果更加稳定的指标。■

参考文献

[1] 郭晓静 . 基于统计学的击键序列身份认证算法研究与改进 [D]. 北京 : 北京邮电大学 , 2013.

[2] 沈超 , 蔡忠闽 , 管晓宏 , 等 . 基于鼠标行为特征的用户身份认证与监控 [J]. 通信学报 , 2010(7):68-75.

[3] 张世雷 . 基于击键特征的身份识别研究 [D]. 哈尔滨 : 哈尔滨工业大学 , 2011.

[4]PUSARA M, BRODLEY C E. User re-authentication via mouse movements[C]//Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, DMSEC Session. Washington DC, USA:2004:1-8.

[5]GAMBOA H, FRED A. A behavioral biometric system based on human computer interaction[J]. Proceedings of SPIE,2004, 54:4-36.

[6]AHMED A A E, TRAORE I. Anomaly intrusion detection based on biometrics[C]. Proceedings of 6th IEEE Information Assurance Workshop. New York, USA:2005:452-453.

[7]AHMED A A E, TRAORE I. Detecting computer intrusions using behavioral biometrics[C]. 3rd Annual Conference on Privacy, Security and Trust, St. Andrews, Canada:2005:91-98.

[8]GARG A, VIDYARAMAN S, UPADHYAYA S, et al. USim: a user behavior simulation framework for training and testing IDSes in GUI based systems[C]. Proceedings of 39th Annual Simulation Symposium. Huntsville, AL:2006:196-203.

[9]PUSARA M, BRODLEY C E. User re-authentication via mouse movements[C]. Proceedings of the 2004 ACM Workshop on Visuali-zation and Data Mining for Computer Security, DMSEC Session. Washington DC, USA:2004:1-8.

[10] 房超 , 蔡忠闽 , 沈超 , 等 . 基于鼠标动力学模型的用户身份认证与监控 [J]. 西安交通大学学报 , 2008(10):1235-1239.

[11] 高艳 , 管晓宏 , 孙国基 , 等 . 基于实时击键序列的主机入侵检测 [J]. 计算机学报 , 2004(3):396-401.