User Identity Authentication Based on the Combination of Mouse and Keyboard Behavior

Haitao Tang[1] and Wang Mantao[2,*]

[1]School of Sichuan Agricultural University, Ya'an City Sichuan Province
Zip Code 625014, China
[2]School of Sichuan Agricultural University, Ya'an City Sichuan Province
Zip Code 625014, China
1076092539@qq.com

Abstract

In order to improve the recognition rate of user identity authentication system, a user identity authentication method based on the combination of mouse and keyboard behavior is proposed. First of all, the characteristics of the two indicators of the mouse and keyboard are extracted, and then realizing the use of support to the rationale for the establishment of an identity authentication device, and finally test through a number of user identification and authentication. The results show that this method can improve the recognition rate of the user identity authentication, which greatly reduces the error rate and rejection rate and the results are obviously superior to the traditional method.

Keywords: Security; behavior characteristics; Identity authentication; Support vector machine; Two instruction

1. Introduction

With the increase of people's dependence on computer technology and network information resources, the problem of computer security and user identity authentication is becoming more and more important. Bio-metric authentication technology, which is popular in the past few years, has gradually received attention from the initial fingerprint recognition to the subsequent recognition of the gesture of the voice of the palm. Typing characteristics, that is, the rhythm of people on a computer keyboard. Because of its uniqueness, it is also included in the field of bio-metric authentication technology [1]. Clicking characteristics is the key to the computer keyboard, because it is unique and more important. Meanwhile, it does not need additional equipment, when the software is implemented, it has a hidden feature, that is, the user can not know whether the system uses this technology, so as to avoid illegal invasion of the damage and forgery. In order to improve the security of identity card, a user training and authentication of user training authentication is proposed, which is based on the behavior of mouse keyboard and mouse and keyboard. Then, the user of the vector machine support (SVM) is used to carry out the identity authentication. The system uses a flexible structure layout, allowing the user to carry out the two development on the basis of the original parts of the system, which has a good portability. The experiment proved that the system's using the double indicator composite certification has a higher accuracy.

---

* Corresponding Author

2. Optimized SVM Algorithm

Obviously this is a straight line, in the determination of the W value and b value, the only X value can be satisfied (1), when using the other X value , it's greater than zero or less than zero, so other X values are divided into two categories, not to meet the conditions of formula(1) .

The above W and X values are one-dimensional, and the SVM is given a new definition of this parameter, which extends the original W value and X value to the multidimensional, and gets the following objective function:

$$X = (x_1, x_2, x_3, ..., x_n) \qquad n = 1、2、3...n \tag{2}$$

$$Y = (y_1, y_2, y_3, ..., y_n) \qquad n = 1、2、3...n \tag{3}$$

$$A = (a_1, a_2, a_3, ..., a_n) \qquad n = 1、2、3...n \tag{4}$$

$$W = y_1 * a_1 * X_1 + y_2 * a_2 * X_2 + ... + y_n * a_n * X_n$$
$$n = 1、2、3...n \tag{5}$$

$$< W, X > + b = 1 - C * E \tag{6}$$

In the formula, C and E are real numbers, here are two new values A and Y, and it is not hard to see that W and X have become multidimensional, (6) is equivalent to the original (1), which will not meet (6) type of multidimensional X into two categories.

Vector X user was defined as an n-dimensional value, the characteristics of the user operation of the mouse or keyboard, such as user (left click time = T1). Suppose there are two users: A and B, a three set of mouse keyboard feature sample X1, X2, X3, a set of samples on the corresponding one type (2), such as X1= (x1, X3, x1, X2, X3, x2), X3= (x1, X2, x3). It should be noted that the xn values in different samples are not necessarily equal, but they are the same type of values (*e.g.,* the mouse or keyboard features), and the number of data in parentheses must be equal. As for the number of data for the custom, for example, adding a double click on the data for x4, then X1= (x1, X3, X2, x4, x1, X3, X2, x4), X3= (x1, X2, X3, X4). In the same way, there are three groups of mouse keyboard feature sample X4, X5, X6. The number of samples can be varied, but the mouse or keyboard feature in the sample must be relevant, and the total sample number is 6. Come and see the formula (3), (4), due to a total sample of 6. Therefore, (3), (4) n formula brackets for 6, namely the total number of samples determined it, (3), (4) in parentheses value to six samples with corresponding to the six samples, such as y corresponding to X1 Y1, as well as a A1, other samples can be judged by this kind of push, now the above correspondence are summarized as follows:

$$X_1 = (x_1, x_2, x_3) \qquad y_1 \quad a_1$$
$$X_2 = (x_1, x_2, x_3) \qquad y_2 \quad a_2$$
$$X_3 = (x_1, x_2, x_3) \qquad y_3 \quad a_3$$

Second:

$$X_4 = (x_1, x_2, x_3) \qquad y_4 \quad a_4$$
$$X_5 = (x_1, x_2, x_3) \qquad y_5 \quad a_5$$
$$X_6 = (x_1, x_2, x_3) \qquad y_6 \quad a_6$$

For the value of Y, count like this: the corresponding Y are assigned to 1, corresponding to the -1, or in turn, that is, y1=y2=y3=1, y4=y5=y6=-1 or y1=y2=y3=-1, y4=y5=y6=1. The value in the A is automatically generated by the user training program for the user. This case takes the user a Y as the standard of 1.

Look at formula (5), putting the AB relation data into (5) on the right , you can draw on the left side of the W value. (5) is the certification formula. (6) the formula W is the (5) W, B for user training program automatically generates a, C and e values for the custom algorithms, X represents the user authentication of a new set of samples. Note that it is only the authenticated users that decided W which of the two users. This is because (6) and (1) are in the same rule, the solution to two kinds of problems, such as w in this case is obtained by the AB sample, and it only certified AB. Then putting data into (6), then only it is greater than the formula on the left side of the certification, it can get through.

Sort out the behavior data collected, and convert them into the format of the algorithm. Because the SVM algorithm is the classification problem of solving the two kinds of samples, so it can be used to distinguish two samples from one line. For multiple users, they take this approach to determine the identity of the current user, such as user 1. And the current training sample is in 6, then the user 1 of the training sample and the remaining 5 user training samples were classified was put in line calculation, and then use these 5 categories to determine the sample, where the need to pay attention to that: training sample must be kept odd number, if the current user training is less than the odd number, the system will automatically call stub sample to make up. If the classification of the line to the user 1 of the votes cast more than half (a line is equivalent to a vote), then it is judged as 1 users.

The optimization of the algorithm, the parameters of the (2) - (6) type are determined as shown in Table 1.

Table 1. All Kinds of Sub Parameter Values are Determined

	X	y	a	W
Value determination	Directly determined by the user mouse data, as the condition of the	Depending on the X corresponding to the user to determine the 1 or -1, it is set value for the known conditions	Will correspond to X, y into the program automatically generated, the value is not determined	Will be all X, y, a into the program automatically generated, the value is not determined
The corresponding formula	(2)、(5)、(6)	(3)、(5)	(4)、(5)	(5)、(6)

	b	C	E
Value determination	Will be all X, y, a into the program automatically generated, the value is not determined	Custom by the program	Fixed to 0.01
The corresponding formula	(6)	(6)	(6)

From the above table, we can see that the X can be changed in these parameters (above X can be defined, that is, the number of mouse indicators and the definition of the mouse index), C (program definition), and the rest of the parameters are either fixed or by other parameters.

3. Identity Authentication System based on Mouse Keyboard

3.1. System Description

System has two main processes which are training module and authentication module. Training and certification must be carried out in a specific scene, each scene is corresponding to a specific mouse or keyboard behavior indicators, such as: left click, right click, right click, wheel, *etc*., the system provides 9 scenarios for training and certification. The user must complete the specified operation according to the instructions given by the scene. The system will automatically shield the corresponding behavioral data for non specified operation, authentication or training, and the user can not complete the current certification or training. Because of the characteristics of SVM algorithm, the system can give an evaluation according to the training data. The user can decide whether to train again according to the result of this assessment. In the training condition, the user needs to give the current account privileges and the security level required to set up the authentication when the user completes the whole scene. Authentication status, the user needs to provide their own account and password and password based on the prompt completion of the certification process of the corresponding level of security. The calculation of the training data, the classification and judgment of the authentication data are all optimized by the SVM algorithm. The system software flow chart is shown in Figure 1.
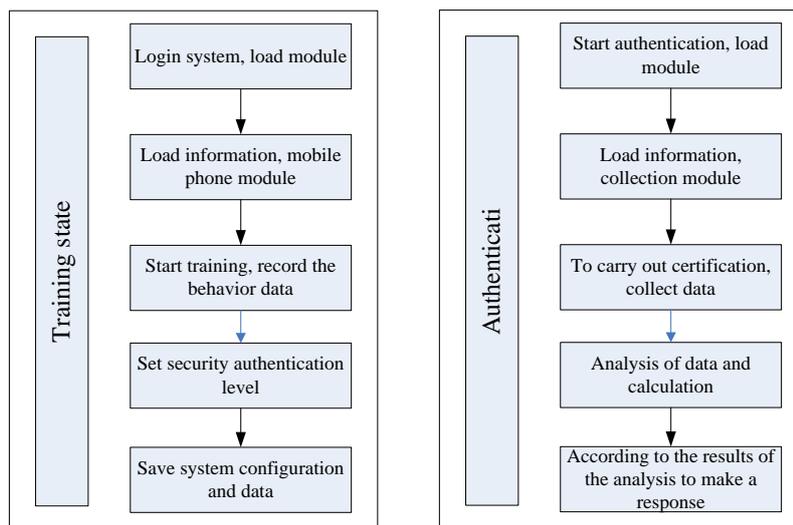

Figure 1. System software process

3.2. User Information Collection and Analysis

System provides a total of 9 scenarios for the user's training and certification. Every scene of the design we are built on a certain background is to avoid the use of the process of boring and complex. They are: Grandpa's meal (double-click the image), according to the dots and dashes (mouse track), brother of the count test (click on the picture), strength training (mouse wheel operations), sister of shopping (right click on the image), mental arithmetic training (keyboard keys related indicators), pet market big rush (mouse track), my gesture password (locus), the traditional password character (keystroke feature). Each scene is corresponding to one or more specific behavior indicators, in the scene, the user must be given in accordance with the tips of the scene to complete the corresponding operation, for non specified operation, the scene will give a hint of operational errors, the corresponding behavior of the operation data will be discarded. The scene indicators are selected after the test. When using these indicators for identity authentication, the sample

data is small, the collection of natural data is also reduced, so that in the process of training and certification, it can save a lot of time. These indicators are selected, in the identification of the test, they are not in the same effect, so we use the safety level of these indicators for the use of these indicators to classify for better results. It will be classified as high security level, the effect of a general or lower level of security. Such users in the authentication, only through a set of good security level of the corresponding scene, do not need to pass through the whole scene, compared to the training time it can also save a lot of time.

By using a specific scene environment, combined with the global keyboard hook (WH_KEYBOARD_LL) and the global mouse hook (WH_MOUSE_LL), we can accurately collect the behavior data we want. In the scene environment. Users need to complete a series of operations according to the tips given by the scene, such as: click the left mouse button, click the right mouse button, double-click the left mouse button, scroll mouse wheel, move the mouse cursor, enter a predetermined password string, *etc.*. In the scene, each step operation in the single scene, the scene of the switch, all of which have ability of control, so it can be very precise positioning of the user's current behavior data generated by the user's current time, and then you can classify these behavioral data according to the calculation of the data format.

The analysis of user behavior data is the core of the whole system. In the third section, the optimized SVM algorithm is used to process the collected user behavior data, and then the user is authenticated.

4. System Simulation

Randomly select 10 users, so that the 10 users were in full scene training operation and their behavior data collected. Then use the behavior data of these six individuals, and carry out the sub index of the number of times to record the corresponding certification results and make an analysis of the results.

Because the classification result is not ideal, so the system only considers C greater than or equal to 1 and less than or equal to 32, this system is defined as 5 indicators: left click the time, left click the time, right click the time, wheel speed, the key time. For these indicators we know that there are a total of 31 combinations, the following test data given in part, as shown in Table 3, Table 2.

Table 2. Standard Rate (X, C=1, =1)

instruction	The correct recognition rate of the user	User rejection rate	Other user correct blocking rate	Other user error rate
Left click	80%	20%	65%	35%
Left double click	61%	39%	77.3%	33.7%
Right click	81.6%	18.4%	72.3%	27.7%
Wheel movement	80%	20%	58.9%	42.1%
clicking time	71%	29%	70.1%	29.9%

Table 3. Standard Rate (X, C=32, =1)

instruction	The correct recognition rate of the user	User rejection rate	Other user correct blocking rate	Other user error rate
left click time	68.3%	31.7%	73.6%	26.4%
left double click time	61.6%	38.4%	80.6%	19.4%
right click time	81.6%	18.4%	68.3%	31.7%
wheel movement rate	96.6%	3.4%	55.3%	44.7%
clicking time	86%	14%	60%	40%

Through the Table 2 and Table 3 result comparison, it is not difficult to see that the correct recognition rate and other users correct interception rate in the C value changes become a reciprocal relationship through the C value analysis resulting in changes in the value of C which is set to 1 optimal.

The X dimension and the index of the experimental results are more complex, there are too many possibilities from the actual test results of 31 kinds of combinations. Select the best combination, the final choice of the optimal combination is: X index number =2, the index is the right click on the time and the key, C value is 1, the test results are shown in Table 4:

Table 4. The Accuracy of the Double =2 (X, C=1)

instruction	The correct recognition rate of the user	User rejection rate	Other user correct blocking rate	Other user error rate
right click time and clicking time	91.6%	6.9%	80.6%	9.4%

From Table 3, 2, 4, we can see that the system has a good effect on the user identity authentication in the double indicator system. The experimental results show that compared with the single index, the system adopts the double index composite authentication, the accuracy of the system is improved and the system error rate and the rejection rate are all reduced.

5. Conclusion

This paper proposed and implemented a kind of identity authentication system based on the behavior of the mouse keyboard. The effect of the identity authentication is better by the experimental verification of the two kinds of indicators. The system has a good portability of the system through the use of a specific scene to shorten the training time and the use of the optimized SVM algorithm to achieve user information analysis. The next work is to improve the accuracy of the certification, which will continue to conduct in-depth analysis and testing of the behavior of the mouse keyboard, looking for more stable indicators.

Acknowledgement

References

[1] K. Wang, "Towards Scalable Distributed Workload Manager with Monitoring-Based Weakly Consistent Resource Stealing", **(2015)**.

[2] Z. Lv, T. Yin, Y. Han, Y. Chen and G. Chen, "WebVR——web virtual reality engine based on P2P network", Journal of Networks, vol. 6, no. 7, **(2011)**, pp. 990-998.

[3] I. Yang, B. Chen, J. Zhou and Z. Lv, "A portable biomedical device for respiratory monitoring with a stable power source", Sensors, **(2015)**.

[4] S. Dang, J. Ju, D. Matthews, X. Feng and C. Zuo, "Efficient solar power heating system based on lenticular condensation", Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on, April 26-28 **(2014)**.

[5] X. Zhang, Y. Han, D. Hao and Z. Lv., "ARPPS: Augmented Reality Pipeline Prospect System", 22th International Conference on Neural Information Processing (ICONIP 2015), Istanbul, Turkey. In press.

[6] K. Wang, "Overcoming Hadoop Scaling Limitations through Distributed Task Execution".

[7] S. Zhang, X. Zhang and X. Ou, "After we knew it: empirical study and modeling of cost-effectiveness of exploiting prevalent known vulnerabilities across iaas cloud", Proceedings of the 9th ACM symposium on Information, computer and communications security. ACM, **(2014)**.

[8] W. Gu, Z. Lv and M. Hao, "Change detection method for remote sensing images based on an improved Markov random field", Multimedia Tools and Applications, **(2016)**.

[9] Z. Lu, C. Esteve, J. Chirivella and P. Gagliardo, "A Game Based Assistive Tool for Rehabilitation of Dysphonic Patients", 3rd International Workshop on Virtual and Augmented Assistive Technology (VAAT) at IEEE Virtual Reality 2015 (VR2015), Arles, France, IEEE, **(2015)**.

[10] Z. Chen, W. Huang and Z. Lv, "Uncorrelated Discriminant Sparse Preserving Projection Based Face Recognition Method", Multimedia Tools and Applications, **(2016)**.

[11] Z. Lv, A. Halawani, S. Feng, H. Li, and S. U. Rehman, **(2014)** "Multimodal Hand and Foot Gesture Interaction for Handheld Devices", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 11, no, 1s, Article 10, **(2014)** October, 19 pages.

[12] K. Leng, W. Shi, J. Chen and Z. Lv, "Designing of a I-shaped less-than-truckload cross-dock: A simulation experiments study", International Journal of Bifurcation and Chaos, **(2015)**.

[13] Y. Lin, J. Yang, Z. Lv and W. Wei, "Houbing Song", A Self-Assessment Stereo Capture Model Applicable to the Internet of Things, Sensors, **(2015)**.

[14] W. Ou, Z. Lv and Z. Xie, "Spatially Regularized Latent topic Model for Simultaneous object discovery and segmentation", The 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC**2015**).

[15] K. Wang, "Using Simulation to Explore Distributed Key-Value Stores for Exascale System Services", 2nd Greater Chicago Area System Research Workshop (GCASR), **(2013)**.

[16] Y. Wang, Y. Su and G. Agrawal, "A Novel Approach for Approximate Aggregations Over Arrays", In Proceedings of the 27th international conference on scientific and statistical database management, ACM, **(2015)**.

[17] Z. Lv, A. Halawani, S. Feng, S. ur Rehman and H. Li, "Touch-less Interactive Augmented Reality Game on Vision Based Wearable Device", Personal and Ubiquitous Computing, **(2015)**.

[18] J. Yang, S. He, Y. Lin and Z. Lv, "Multimedia cloud transmission and storage system based on internet of things", Multimedia Tools and Applications, **(2016)**.

Authors

Tang Haitao, studying at Sichuan Agricultural Uniersity, undergraduate degree. He is mainly engaged in the research of mathematical models, has published "based on DEA and AHP of the regional public library research on performance evaluation of public expenditure" in the Library Journal of CSSCI. He participated not only in the Mathematical Contest In Modeling & Interdisciplinary Contest In Modeling, winning the Honorable Mention, but also received the national awards of Honorable Mention in the China Undergraduate Mathematical Contest in Modeling.