

Improving the Performance of Mouse Dynamics Based Authentication Using Machine Learning Algorithm

S. Suganya, G. Muthumari, and C. Balasubramanian

Department of Computer Science and Engineering,
P.S.R.Renasamy College of Engineering for women,
Sivakasi, Tamilnadu, India

Copyright © 2016 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: To diminish the behavioral variability of mouse dynamics, the machine learning algorithm was proposed. Mouse dynamics is the process of identifying the user based on their mouse operating behavior. The dataset includes co-ordinates values, time stamp value and mouse operation. From this dataset, the schematic features, holistic features and motor-skill features like average speed, average distance, mean, standard deviation and mouse silence ratio, velocity, slope angle, curvature were extracted to obtain feature vector. The obtained feature vector can be applied to the dimensionality reduction based approach, diffusion map to reduce the dimension of the feature vector that compared with ISOMAP (Isometric Feature Mapping). Without dimensionality reduction based method the classification process was difficult. The machine learning algorithm i.e.) hop field network to be used to identify whether the given input sample was authenticated user (or) unauthenticated.

KEYWORDS: behavioral variability, mouse dynamics, feature vector, dimensionality reduction, diffusion map, hop field network.

1 INTRODUCTION

Authentication may take place a vital role in computer security and network security. In today's web-centered world, the activities of each user authentication and verification is more essential than that it before [1][2]. Authentication is the process of recognizing the individuality of each user. In many industry and organization authentication can be provided by typing username and password. But this type of authentication is not much secure. Because there are two issues 1)password cracking 2)password theft [1]. So the system authenticate the user based on the biometric technologies [2].

Biometric technology may present the individuality of the authentication with the use of physical characteristics i.e.) identify the user based on their physical characters like finger print, iris, face, hand geometry and behavioral characteristics [7] i.e.) identify the user based on their behavioral character like voice, signature, keystroke, mouse dynamics [15]. Need to address the behavioral variability problem in mouse dynamics which means each user can differently operate the mouse at different time [15]. Extract the features to identify the uniqueness of the user. Dimensionality reduction based approach to be applied to reduce the dimension and also reduce the complexity.

2 BACKGROUND

In this section explained about the existing work related to mouse operating behavior and some methodologies to analyze the performance of the existing system and to improve the performance of the proposed system.

In [2]-[5] the work only focused the comparison of keystroke dynamics and mouse dynamics. They conclude that mouse dynamic provided the much better performance.

Bala ganesh, Soniya[18] proposed the authentication system based on mouse operating behavior. The mouse behavior used to verify the computer user authenticity and also to improve the efficiency of the authentication process. The mouse operation task can be performed before authenticate the user. Identify the features from the data that collected from mouse operation and characterized the feature in fine grained manner. One class learning algorithm was to be used to the Eigen space computation for the authentication task. The result proves that it provides high accuracy.

Issa Traore, woungang, mohammad [12] addressed the risk based authentication system in web communication on the basis of source IP address, velocity of transaction performed by a account. The new-online authentication system provided more user identity information by combining the keystroke and mouse dynamics biometric techniques. It improved the reliability and robustness of the system. The experimental result shown that the new framework provides better authentication and the error rate measured as 8.21%.

Chao Shen, Z.Cai, X.Guan [13] addressed behavioral variability issue with 10 computer user. Silence ratio, elapsed time of single click, movement speed movement direction, cursor position distribution features were extracted. Dimensionality reduction based approach, Isometric Feature Mapping (ISOMAP) was proposed and compared with principal component analysis (PCA) technique. The performance can be measured as FAR 5.12% and FRR 4.23%.

Qi Liu, Xiuzi ye, Yin zhang [5] addressed the Hopfield network algorithm to predict the RNA molecule in a secondary structure. Several RNA molecule can represent a structure. Hopfield network predict the exact match of the RNA molecule accurately to control the variation rate of the inhibitory. Hopfield network can be compared with the two classical prediction method Zuker and Nussinov's method. The test result shows that the Hopfield network was more sensitive and specific than the Zuker and Nussinov's method. The algorithm is very efficient to predict the information.

The performance results of the previous work demonstrated with error rates and accuracy rate. A technique related to the work is proposed which defines schematic features, motor-skill features and Isometric Feature Mapping for Eigen space computation with One-Class Support Vector Machine for classification.

3 SYSTEM DESIGN

The proposed system contains the following methodologies to perform user authentication. The main objective is to improve the performance of the existing system with the methodologies. From the dataset [19] the system may lead the following steps.

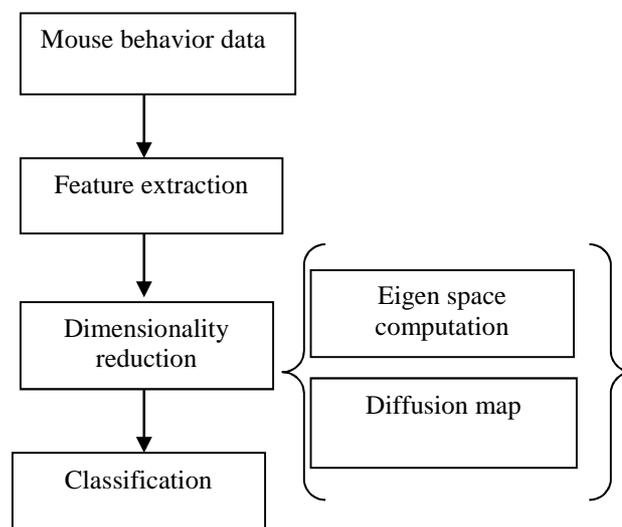


Fig 1: System description

The dataset consists of mouse operation, co-ordinates values, timestamp value. The feature extraction includes schematic features and motor-skill features and curve based features were extracted from the collected mouse behavior dataset. Dimensionality reduction based approach i.e.) diffusion map was proposed to reduce the behavioral variability [12] and also reduce the dimension of the feature space. Classification methodology, Hopfield network was applied to classify whether the input sample is legitimate or imposter.

4 METHODS

This section describes the implementation of the proposed work. The proposed work consists of the following methods of implementation. The dataset contains the static mouse behavior data of 20 users. Dataset consists of the mouse operation, x co-ordinate value, y co-ordinate value and timestamp value [19]. The dataset may have 20 user samples with 50 samples for each user.

Table 1. SAMPLE DATASET

Mouse operation	X Co-ordinate	Y Coordinate	Timestamp(ms)
512	166	100	1512335875
512	166	99	1512335890
512	166	98	1512335890
513	165	97	1512336000
514	165	97	1512336109
512	172	95	1512336218

The table 1 shows the sample dataset for a single user with only of 10 samples. Similarly the dataset consists of 20 user data with 50 samples for each single user. Each sample consists of thousands of mouse movements. The mouse operation represented as numeric representation. Each number represents one mouse operation. For example 512 represent the mouse movement operation [14].

From the collected dataset the following feature can be derived. The feature can be derived to identify the identity or uniqueness of the user.

4.1 FEATURE EXTRACTION

From the collected dataset [17], we needed to extract the feature to identify the identity of each user because from the dataset we could not identify the individuality of the user. The following schematic feature and motor-skill feature can be extracted [15].

4.1.1 MOUSE OPERATION FREQUENCY

Number of frequent occurrence of each mouse operation (left, right, double click, drag and drop).

4.1.2 MOUSE SILENCE RATIO

The particular period of time the mouse should be ideal to the session.

$$silence\ ratio = \frac{\text{Amount of time that the mouse is ideal}}{\text{Total amount of time taken by mouse}} \quad (1)$$

4.1.3 MOVEMENT ELAPSED TIME

Elapsed time is time difference between starting point and ending point of a mouse movement.

4.1.4 MOVEMENT OFFSET

Offset is the distance between the practical mouse trajectory and the ideal mouse trajectory for each movement.

4.1.5 AVERAGE MOVEMENT SPEED

Speed can be calculated as the ratio of the distance between the start and end point of the mouse and the time difference between the start and end time.

$$speed = \frac{\text{distance}}{\text{time}} \quad (4)$$

4.1.6 DISTRIBUTION OF CURSOR POSITION

Mean and standard deviation for both X co-ordinate and Y co-ordinate were calculated as,

$$\text{Mean} = \frac{\sum_i^m \text{co-ordinate value}}{\text{number of co-ordinate value}} \quad (5)$$

4.1.7 AVERAGE MOVEMENT DISTANCE

Mean and standard deviation for distance (distance between two co-ordinate values) were computed as,

$$Dm = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (6)$$

4.1.8 HORIZONTAL VELOCITY

The ratio between the difference between the adjacent x co-ordinate values and the difference between the timestamp value of corresponding co-ordinate values.

$$VH = \frac{X_{m+1} - X_m}{T_{m+1} - T_m} \quad (7)$$

4.1.9 VERTICAL VELOCITY

The ratio between the difference between the adjacent Y co-ordinate values and the difference between the timestamp values of corresponding co-ordinate values.

$$VV = \frac{Y_{m+1} - Y_m}{T_{m+1} - T_m} \quad (9)$$

4.1.10 TANGENTIAL VELOCITY

Tangential velocity can be calculated from the horizontal and vertical velocity of the data.

$$TV = \sqrt{VV^2 + VH^2} \quad (10)$$

4.1.11 TANGENTIAL ACCELERATION

The ratio between the tangential velocity and the difference between the timestamp values.

$$TA = \frac{TV_{m+1} - TV_m}{T_{m+1} - T_m} \quad (11)$$

4.1.12 TANGENTIAL JERK

The ratio of the tangential acceleration to the difference between the timestamp values.

$$Tjerk = \frac{TV_{m+1} - TV_m}{T_{m+1} - T_m} \quad (12)$$

4.1.13 SLOPE ANGLE

Slope angle of the tangent is the inverse tangent of the X and Y co-ordinate, calculated as

$$\text{Slope}(\phi_m) = \arctan\left(\frac{Y_m}{X_m}\right) \quad (13)$$

4.1.14 CURVATURE

The ratio of the difference between the slope angle of the tangent to the difference between the feature length.

$$Curvature(C_m) = \left(\frac{\Delta\phi_m}{\Delta l_m} \right) \quad (14)$$

Where $\Delta\phi_m$ is difference between the slope.

Δl_m is difference between the feature length which is defined as,

$$\Delta l_m = \sqrt{X^2 + Y^2} \quad (15)$$

The extracted feature may generate the feature vector space.

4.2 DIMENSIONALITY REDUCTION

This section explains about the dimensionality reduction approach, used to reduce the dimension of the extracted feature vector. The diffusion map algorithm was proposed to reduce the dimension of feature vector.

Diffusion map embeds the samples into the Euclidean space whose co-ordinate can be derived from the Eigen value of the diffusion operator. Euclidean distance is equal to the diffusion distance. The algorithm focused on manifolding the data that had been sampled from the diffusion operator. It could integrating the local similarities at different scale [20]. Diffusion map have the same steps of Isomap, additionally diffusion map can be used.

Step 1: Similarity matrix can be calculated.

Step 2: Normalize the similarity matrix according to α

$$H^{(\alpha)} = B^{-\alpha} H B^{\alpha} \quad (16)$$

Step 3: Normalized matrix can be formed from the matrix as follows,

$$K = (B^{(\alpha)})^{-1} H^{(\alpha)} \quad (17)$$

Step 4: K-largest eigen value can be calculated for the matrix K^n .

Step 5: Diffusion map can be used to get embedded sample with τ_n .

4.2.1 CONNECTIVITY

It exploits the relationship between heat diffusion and random walk Markov chain. Probability can be specified by using Gaussian kernel function as,

$$G(X_m, Y_n) = e^{-\left(\frac{(X-Y)^2}{\epsilon}\right)} \quad (18)$$

4.2.2 DIFFUSION PROCESS

Transition matrix of a Markov chain (B) on X can be constructed from the diffusion matrix H.

$$H_{m,n} = G(X_m, Y_n) \quad (19)$$

Eigen decomposition of the diffusion matrix can be calculated as,

$$K_{m,n}^t = \sum_a \delta_a^t \varphi_a(X_m) \phi_a(X_n) \quad (20)$$

4.2.3 DIFFUSION DISTANCE

Diffusion distance at the particular time interval t between the co-ordinates can be measured as,

$$D_t(X_m, X_n)^2 = \sum_Y \frac{(p(Y, t|X_m) - p(Y, t|X_n))^2}{\rho(Y)} \quad (21)$$

Where $\rho(Y)$ is the stationary distribution of the markov chain. It can be derived from the first left eigen vector of K.

$$\rho(Y) = \frac{\partial(Y)}{\sum_{Z \in X} \partial(Z)} \quad (22)$$

Diffusion distance may have the following features based on scale parameter t.

- 1) The co-ordinate points were highly connected in graph it could form the cluster.
- 2) Diffusion distance is robust to noise and it depends on all possible paths.
- 3) Diffusion distance may focus on inference algorithm based on the majority of preponderance.

4.2.4 LOW DIMENSIONAL EMBEDDING

Diffusion distance can be calculated from the eigenvector as,

$$D_t(X_m, X_n) = \sum_s \mu_s^{2t} \left((\varphi_s(X_m) - \varphi_s(X_n)) \right)^2 \quad (23)$$

Diffusion map is defined as,

$$\varphi_t(X) = (\mu_1^t \varphi_{1(X)}, \mu_2^t \varphi_{2(X)}, \dots, \mu_l^t \varphi_{l(X)}) \quad (24)$$

Embed the sample with diffusion map operator then get the feature vector with low dimension.

4.3 CLASSIFICATION

In this section explained the implementation part of classification. Classification is the continuous process of recognizing to which a set of groups a new observed sample belongs on the basis of a training set of data containing observation whose category membership is known to the system [15]. This section considered the output of the diffusion map is the input of Hopfield network [9]. The network structure looks as follows

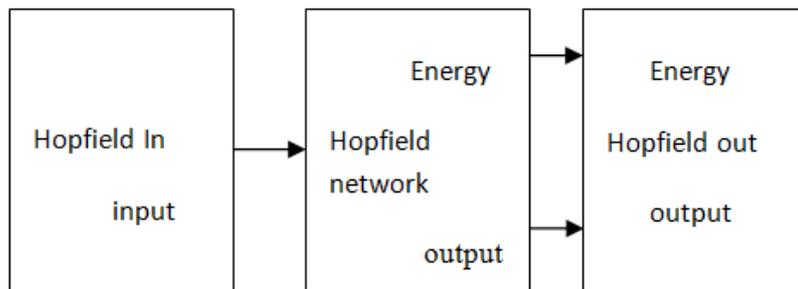


Fig 2: Hopfield network model

Classification can be done by the neural network classifier. Neural network is defined as a computing system made up of a number of simple and highly interconnected processing elements, which process the information by their dynamic state response to external inputs [8]. Hopfield neural network can be used for classification. Hopfield network is a model of associative memory [1]. It is based on Hebbian learning but uses binary neurons.

Hopfield network may use the following relationship with the use of threshold value to identify the legitimate user.

$$a_i = \begin{cases} 0 & \text{if } \sum_j w_{ij} s_j > \theta_j \\ 1 & \text{otherwise} \end{cases} \quad (25)$$

The network result is 0 then the new sample is imposter i.e.) unauthorized user. Then the system could not authenticate that user. The result is 1 the sample is authorized user.

5 PERFORMANCE

The performance of the ISOMA with one-class classifier can be evaluated using the following measurement

FAR: The instance of a security system incorrectly verifying an unauthorized user. It measured as False acceptance rate[15].

$$FAR = \frac{\text{Number of false acceptance}}{\text{Number of invalid sample}} \quad (26)$$

FRR: The instance of a security system failing to verify an authorized user. It is measured as False rejection rate [1].

$$FRR = \frac{\text{Number of false rejection}}{\text{Number of valid sample}} \quad (27)$$

Hopfield network classifier with diffusion map algorithm can achieved the performance with FAR as 5.25 and FRR as 4.26.

6 RESULT AND DISCUSSION

This section illustrates the result and performance analysis of the proposed system. Table 2 describes the performance measure of existing and proposed system. Performance can be measured as FAR and FRR.

Table 2. PERFORMANCE (FAR AND FRR) ANALYSIS

Algorithm	FRR (%)	FAR (%)
Hopfield network with diffusion map	4.15	5.05
One-class SVM with Isomap	4.25	5.25
Original feature space	10.25	11.25

The performance result proves that the Hopfield network with diffusion map provide better performance than the existing system.

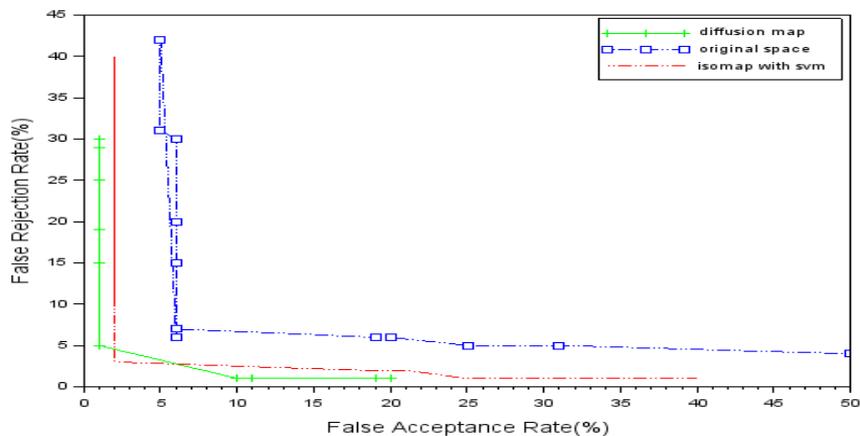


Fig3: Performance analysis of diffusion map with Isomap

The performance analysis may shows the result and diffusion map with Hopfield network can be compared to the ISOMAP with one-class SVM beside the ROC curve [9]. ROC curve is a curve which may plot the true positive rate against false positive rate for various cut points [15]. The test result proves that diffusion map with diffusion operator with one-class classification may provide the better result than the existing method ISOMAP with SVM classifier.

7 CONCLUSION

Authentication based on mouse dynamics biometric is a most powerful technique that present the authentication within a small period of time and provides the better accuracy [11]. Feature can be extracted from the mouse operating data and dimension can be reduced using diffusion map technique with the help of the diffusion operator and the selected feature can

be applied to the Hopfield network classifier to identify the identity of the user. The performance can be measured as FAR 4.15% and FRR 5.05%. This provides the reduced authentication time and good accuracy than the existing system. Diffusion map technique is computationally inexpensive and robust to noise. The test results, shows that the diffusion map with Hopfield network classifier may provide the good and better performance than the existing system (ISOMAP with SVM).

REFERENCES

- [1] Yi sun, "Hopfield neural network algorithm for image restoration reconstruction part II performance analysis" in IEEE transaction on signal processing vol 48. No.7 July 2000.
- [2] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Roux, and M. Ouimet, "Out-of-sample extensions for LLE, ISOMAP, MDS, eigenmaps, and spectral clustering," in *Advances in Neural Information Processing Systems*. vol. 16, Cambridge, MA, USA: MIT Press, 2004.
- [3] Ahmed and I. Traore, "Anomaly intrusion detection based on biometrics," in *Proc. IEEE Inform. Assurance Workshop*, West Point, NY, USA, 2005, pp. 452–453.
- [4] S. Hashia, C. Pollett, and M. Stamp, "On using mouse movements as a biometric," in *Proc. Int. Conf. Comput. Sci. Appl.*, Singapore, 2005, pp. 143–147.
- [5] Qi Lui, Xiuzi Ye, Yi Zhang, "Hopfield neural network based algorithm for RNA secondary structure prediction" in Intenation multi symposium on computer and computational science (IMSCC), 2006, pp.7695-2581.
- [6] Mikio, Joachim, Robert, "Denosing and dimension reduction in feature space" in *Proc. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 2006 pp.185-192.
- [7] Y. Nakkabi, L. Traore, and A. A. E. Ahmed, "Improving mouse dynamics biometric performance using variance reduction via extractors with separate features," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 6, pp. 1345–1353, Nov. 2010.
- [8] A.A.E. Ahmed and I.Traore, "A new biometric technology based on mouse dynamics," *IEEE Trans Depend .Secure Comput.*, vol. 4, no.3, pp.165–179, Jul.–Sep. 2007.
- [9] Shen, Z. M. Cai, X. H. Guan, H. L. Sha, and J. Z. Du, "Feature analysis of mouse dynamics in identity authentication and monitoring," in *Proc. IEEE Int. Conf. Communication (ICC)*, Dresden, Germany, 2009, pp. 1–5.
- [10] Y. Aksari and H. Artuner, "Active authentication by mouse movements," in *Proc. 24th Int. Symp. Comput. Inform. Sci.*, Guzelyurt, Turkey, 2009, pp. 571–574.
- [11] C. Shen, Z. M. Cai, and X. H. Guan, "Continuous authentication for mouse dynamics: A pattern-growth approach," in *Proc. IEEE /IFIP Conf. Dependable Syst. Netw.*, Boston, MA, USA, 2012, pp. 1–12.
- [12] Issa Traore, Iris Lai, "Combining mouse and keystroke dynamics biometrics based authentication in web environment" in 4th international conference on digital home, pp 7695-4899 2012.
- [13] Chao Shen, Zhongmin Cai, Xiaohong Guan and Roy A. Maxion, " User authentication through mouse dynamics," *IEEE Transaction on Information Forensics and security*, Vol. 8,no.1,JAN 2013.
- [14] Zhongmin Cai, Chao Shen, and Xiaohong Guan "Mitigating Behavioral Variability for Mouse Dynamics: A Dimensionality-Reduction-Based Approach" in IEEE April 2014
- [15] G. Muthumari, R. Shenbagaraj, M. Blessa Binolin Pepsi " Mouse Gesture Based Authentication Using Machine Learning Algorithm", 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)
- [16] Dr. P.M. Balaganesh, A. Sonia, "A survey of authentication based on mouse behaviour" in international journal of advanced information science and technology vol.22, No.22 Feb 2014
- [17] <http://nskeylab.xjtu.edu.cn/projects/mousedynamics/behavior-data-set>.
- [18] http://en.wikipedia.org/wiki/Network_security.