Feature Subset Selection Using PSO-ELM-ANP Wrapper Approach for Keystroke Dynamics

D. Shanmugapriya and G. Padmavathi

Abstract The security of computer access is important today because of huge transactions being carried out every day via the Internet. Username with password is the commonly used authentication mechanism. Most of the text based authentication methods are vulnerable to many attacks as they depend on text and can be strengthened more by combining password with key typing manner of the user. Keystroke Dynamics is one of the famous and inexpensive behavioral biometric technologies, which identifies the authenticity of a user when the user is working via a keyboard. The paper uses a new feature Virtual Key Force along with the commonly extracted timing features. Features are normalized using Z-Score method. For feature subset selection, a wrapper based approach using Particle Swarm Optimization—Extreme Learning Machine combined with Analytic Network Process (PSO-ELM-ANP) is proposed. From the results, it is observed that PSO-ELM-ANP selects less number of features for further processing.

Keywords Keystroke dynamics · Particle swarm optimization · Extreme learning machine · Analytic network process

---

D. Shanmugapriya (✉)
Department of Information Technology, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India
e-mail: shanmugapriya.adu@gmail.com

G. Padmavathi
Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India
e-mail: ganapathi.padmavathi@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 221, DOI: 10.1007/978-81-322-0997-3_55,

1 Introduction

The confidential information can be secured from unauthorized users by providing authentication. User authentication is defined as the process of verifying the identity claimed by an individual. User Authentication is normally classified into three categories [1] namely, Knowledge based Authentication, Object based Authentication, and Biometric based Authentication. Biometric based authentication provides a very reliable method of authenticating a user and is divided into physiological and behavioral types [2]. Physiological biometrics refer to what the person is or they measure physical parameters of certain parts of the body and Behavioral biometrics shows how the person is using the body for authentication. Keystroke dynamics is a strong behavioral biometric [3] and it is the process of analyzing the way a user types at a terminal by monitoring the keyboard in order to identify the user based on habitual typing rhythm patterns. Unlike other biometric systems, which may be expensive to implement, keystroke dynamics is almost free and it does not require any sophisticated hardware as the only hardware required is the keyboard. Timing features are the one most commonly measured from the keystroke. The huge features often lead to degradation of performance of the system. Hence an attempt is made in this paper to reduce the dimensionality of features using wrapper approaches.

The rest of the paper is organized as follows: the next section discusses the related works done in feature selection, Sect. 3 discusses the methodology and the results are discussed in the Sect. 4. Final section concludes the work with future directions.

2 Related Works

This section discusses the major feature subset selection methods applied in keystroke dynamics. Optimization techniques such as Genetic Algorithm (GA) and Azevedo et al. [4, 5] evaluated the feature subset selection based on Support Vector Machine (SVM) with Genetic Algorithm (GA) and a variation in Particle Swam Optimization (PSO). In this study, wrapper based approach with SVM-PSO variation outperformed SVM-GA with regard to classification error and processing time. Bleha and Obaidat [6] used a reduction technique based on Fisher analysis. However, the technique considered by keeping m − 1 dimension for each vector, with 'm' number of users in the system. Boechat et al. [7] used weighted probability measure to select a subset of 'N' features with the minors of standard deviation. The study eliminates less significant features by keeping 70 % of the features. Experiments are done at Zero False Acceptance Rate. False Rejection Rate reduces when the number of selected features increases. Particle Swarm Optimization (PSO), Genetic Algorithm (GA) and Ant Colony Optimization (ACO) are used by Karnan and Akila [8] for feature subset selection.

Back propagation Neural Network (BPN) has been used for classification. ACO gives better performance than PSO and GA with regard to feature reduction rate and classification accuracy. Sung and Cho [9] adds a uniqueness term in the fitness function of genetic algorithm and used SVM as classifier for classification and Genetic Algorithm (GA) to select the data. The experiments performed better than two phase ensemble selection approach and prediction based diversity term approach. Enzhe Yu and Sungzoon Cho [10] used wrapper based Support Vector Machines (SVM) and Genetic Algorithms (GA) to reduce the feature size. SVM is applied for training and GA is used for searching. GA-SVM wrapper approach gives good accuracy. Feature Selection (FS)—Ensemble is proposed to deal with the over fitting problems.

3 Methodology

The entire methodology is divided into three phases namely, feature extraction phase, normalization phase and feature subset selection phase after obtaining the raw keystroke data (press time and release time). The key function of feature extraction in keystroke dynamics is to extract the fundamental features from the timestamp collected from raw keystroke data for creation of template. The password '.tieRo-anl' has been collected from 100 users each typed the password 10 times. The features Duration or Dwell time, Flight time or Latencies, Digraph, Trigraph and Virtual Key Force [11] are measured. After feature extraction, data normalization using Z-Score [12] is carried out to bring the data in a particular range (0–1) as the time and speed of typing varies for same user even for the same text.

Feature subset selection identifies the most discriminating features. It also reduces the dimensionality of features which improves the assumption accuracy and decreases the computation time. Feature subset selection is applied to high dimensional data before preceding the classification since the increased dimensionality of features makes testing and training of classification method difficult. Feature subset selection is divided into Filter and Wrapper approach [13]. In Filter approach, the feature subset selection is done independently of the learning algorithm. The wrapper approach uses the method of classification to measure the importance of features set.

3.1 Wrapper Based Feature Subset Selection using PSO-ELM-ANP

Particle Swarm Optimization (PSO) is wrapped with Extreme Learning Machine (ELM)—Analytic Network Process to select the subset of features from the extracted features. These proposed wrapper based approaches select an appropriate

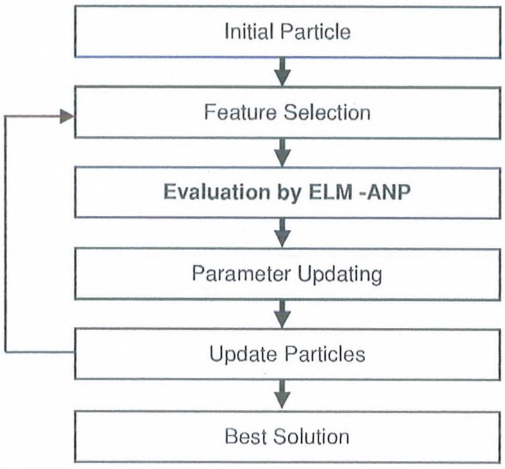

Fig. 1 Flowchart of proposed PSO-ELM-ANP wrapper feature subset selection

subset of features and the rest will not be considered, thus resulting in a more comprehensive model. Flow chart for PSO-ELM-ANP wrapper approach for feature subset selection is given in Fig. 1.

3.1.1 Particle Swarm Optimization

Particle Swarm Optimization (PSO) [14] is a stochastic search technique that aims to optimize an objective function, motivated by social activities of birds gathering or fish schooling. PSO is a population dependent search algorithm in which each individual is indicated as a particle and represents a candidate solution. All the particles in PSO moves through the search space with an adjustable velocity that is dynamically transformed based on its individual moving experience and also the moving experience of the other particles. In PSO each particles attempt to enhance themselves by imitating traits from their successful peers. The position in proportion to the best fitness is known as pbest (Particle best) and the overall best out of all the particles in the population is called gbest (Global best) [14]. The velocity $v_i(t)$and the position $x_i(t)$ of the particles are updated with the following equations.

$$v_i(t+1) = w * v_i(t) + c_1 r_1 (y_i(t) - x_i(t)) + c_2 r_2 (\hat{y}(t) - x_i(t)) \quad (1)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2)$$

where i = 1, 2, …, n

$v_i(t)$ Velocity of agent i at iteration t and must lie in the range

$$V_{min} \leq V_i(t) \leq V_{max} \quad (3)$$

W: Inertia weight factor
c_1, c_2 : Cognitive and social acceleration factors respectively,
r_1, r_2 : Uniformly distributed random number between 0 and 1,
$x_i(t)$: Current position of agent i at iteration t,
$y_i(t)$: Personal best (pbest) and is calculated as

$$y_i(t+1) = \begin{cases} y_i(t), f(x_i(t+1)) \geq f(y_i(t)) \\ x_i(t+1), f(x_i(t+1)) < f(y_i(t)) \end{cases} \quad (4)$$

$\hat{y}(t)$ Global best (gbest) and is calculated using the following equation:

$$\hat{y}_i(t+1) = \text{argmin}\{f(y_i(t+1))\}, i \leq i \leq s \quad (5)$$

where s = number of particles in the swarm.

3.1.2 Extreme Learning Machine

Extreme learning machine (ELM), an emerging technology proposed by [15] overcomes the challenges such as slow learning speed, trivial human intervene and poor computational scalability faced by the computational intelligent techniques such as Back Propagation Neural Network (BPN) and Support Vector Machines (SVMs). ELM works for single-hidden layer feed forward networks (SLFNs). The hidden layer of SLFNs need not be tuned which is an advantage of Extreme Learning Machine. The learning speed of ELM is faster and with least human intervene which gives better performance compared with other computational intelligent techniques. The Algorithm for ELM [15] is narrated below:

Given a training set $\aleph = (x_i, y_i)$, $x_i \in R^{d_1}$, $y_i \in R^{d_2}$ an activation function $f : R \mapsto R$ and the number of hidden nodes H.

- Randomly assign input weights w_i and biases $b_i, i \in [1, H]$..
- Calculate the hidden layer output matrix H.
- Calculate output weight matrix$\beta = H^{\dagger} Y$.

3.1.3 Analytic Network Process

The input weights and the hidden bias are randomly chosen in Extreme Learning machine. The output weights are analytically determined by Moore Penrose inverse. ELM overcomes the problems such as slow learning speed, number of epochs etc., which are being faced by the computational intelligence techniques such as BPN and SVMs and has faster learning rate. However, ELM requires more number of hidden neurons than that of human intervene based algorithms that are mentioned above. The input weights are generated using Analytic Network Process (ANP).

ANP proposed by [16] facilitates for more complex interrelationships among the decision levels and attributes and does not require any strict hierarchical structure. The ANP approach handles interdependent relationships among the elements by obtaining the composite weights developed by the supermatrix. The supermatrix concept contains parallels to the Markov chain process where relative importance weights are adjusted by forming a supermatrix from the eigenvectors of these relative importance weights. The weights are adjusted by determining products of the supermatrix. In this work, the input weights of ELM are determined using ANP technique.

3.1.4 Proposed PSO: ELM: ANP Wrapper Approach Algorithm

In the proposed PSO-ELM-ANP feature subset selection wrapper approach PSO algorithm is wrapped with ELM-ANP approach and used for feature subset selection. Duration, Latency, Trigraph, Digraph and the proposed Virtual Key Force are given as input features. Feature subset selection is done and the selected features are evaluated by Extreme Learning Machine- Analytic Network Process. The process is repeated until the best solution is obtained. The training algorithm used for ELM is modified Levenberg–Marquardt Algorithm [17]. In standard Levenberg–Marquardt (LM) algorithm, the learning parameter μ, is a constant number where modified LM algorithm modifies the learning parameter as

$$\mu = 0.01e^{T}e \tag{6}$$

where **e** is a k × 1 matrix.

The algorithm of PSO-ELM-ANP Wrapper approach is given as follows:

Step 1: Initialize the number of Iterations, Number of particles, Weight, c_1, c_2, r_1,r_2, $v_i(t)$.

Step 2: Compute the feature values of $x_i(t)$ (Duration, Latency, Digraph, Trigraph, and Virtual Key Force).

Step 3: Evaluate fitness for each feature value using **ELM-ANP**.

Step 4: The following is repeated for number of iterations:

1. Check if p >=pbest then, pbest = p else pbest = pbest.
2. If pbest >=gbest then, gbest = pbest else gbest = gbest.
3. Update velocity by (1) and position is updated by (2), pbest and gbest position are updated using (4) and (5).

Step 4 is repeated until gbest is optimum value.

Table 1 Features selected

Induction Algorithms	PSO			
	W.O. VKF		W. VKF	
	F	FS	F	FS
ELM	60	50	71	53
ELM—ANP	60	46	71	51

Table 2 Features selection percentage

Induction Algorithms	PSO	Feature Selection (%)
ELM	53	74.7
ELM—ANP	51	71.8

4 Experimental Results

In the proposed method, the features are selected by PSO and are given as input vectors to the input layers. The input weights and hidden bias are generated by Analytic Network Process. Extreme Learning Machine analytically determines the output weights and the weights and bias are updated using Modified Levenberg–Marquardt (LM) algorithm which is used to train the network.

The proposed methodology is tested with 100 users typing the password '.tieRoanl' 10 times. The key press and release time of the characters of the passwords are extracted. Timing features such as Duration, Latency and Virtual Key Force are measured. The password contains 11 characters including shift key to press R. Therefore, total of 71 features are measured from the password with VKF and 60 features are measured without VKF. The feature subset selection methods, PSO-ELM and PSO-ELM-ANP approaches with and without the feature Virtual Key Force are experimented. The following Table 1 shows the features selected

F-Features measured, FS-Features Selected, W.O. VKF- Without VKF, W.WKF—With VKF

From Table 1, it is observed that PSO—ELM—ANP wrapper has reduced the existing 71 features to 51 features. The proposed PSO—ELM—ANP wrapper based approach has selected 49.3 % of total 71 features and has reduced 28.2 % (100–71.8 %) of the total features which is shown in the Table 2.

5 Conclusion and Future Directions

The Extreme Learning machine integrated with Analytic Network Process is wrapped with Particle Swarm Optimization to reduce the number of features. The reduced features can be used for further classification. In future more optimization techniques such as Ant Colony Optimization, Genetic Algorithm etc., can be explored.

References

1. O'Gorman L (2003) Comparing passwords. Tokens Biometrics User Authentication Proc IEEE 91(12):2019–2040
2. Francesco B, Gunetti D, Picardi C (2002) User authentication through keystroke dynamics. ACM Trans Inf Syst Secur 5(4):367–397
3. Ahmed AAE, Traore I (2005) Anomaly Intrusion detection based on biometrics. In: Proceedings of 6th IEEE information assurance, workshop, pp. 452–453
4. Azevedo GLF, Cavalcanti GDC, Carvalho Filho ECB (2007) Hybrid solution for the feature selection in personal identification problems through keystroke dynamics. In: International joint conference on neural networks, pp 1947–1952
5. Azevedo GLF, Cavalcanti GDC, Filho ECB (2007) An approach to feature selection for keystroke dynamics systems based on PSO and feature weighting. In: IEEE Congress on, Evolutionary Computation, pp 3577–3584
6. Bleha S, Obaidat M (1991) Dimensionality reduction and feature extraction applications in identifying computer users. IEEE Trans Syst Man Cybern 21(2):452–456
7. Boechat G, Ferreira J, Carvalho E (2006) Using the keystrokes dynamic for systems of personal security. Proc World Acad Sci Eng Technol 18:200–205
8. Karnan M, Akila M (2010) Personal authentication based on keystroke dynamics using soft computing techniques. In: Second international conference on communication software and networks. pp 334–338
9. Sung KS, Cho S (2006) GA SVM wrapper ensemble for keystroke dynamics authentication. In: Proceedings of the international conference on biometrics, China, vol 3832, pp 654–660
10. Yu E, Sungzoon Cho (2003) GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification. In: Proceedings of the international joint conference on neural networks, vol 3, pp 2253–2257
11. Shanmugapriya D, Padmavathi G (2011) Virtual key force—a new feature for keystroke. Int J Eng Sci Technol 3(10):7738–7743
12. Jain A, Nandakumar K, Ross A (2005) Score normalization in multimodal biometric systems. Pattern Recogn 38:2270–2285
13. Yu E, Cho S (2004) Keystroke dynamics identity verification: its problems and practical solutions. Comput Secur 23:428–440
14. Panda S, Padhy NP (2007) Comparison of particle swarm optimization and genetic algorithm for TCSC-based controller design. World Acad Sci Eng Technol 27:497–505
15. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: a new learning scheme of feedforward neural networks. In: Proceedings of 2004 international joint conference on neural networks
16. Saaty TL (1996) Decision making with dependence and feedback, the analytic network process. RWS Publications, Pittsburgh
17. Suratgar AA, Tavakoli MB, Hoseinabadi A (2005) Modified Levenberg-Marquardt method for neural networks training. World Acad Sci Eng Technol 6:46–48