Int. J. Advanced Networking and Applications
Volume: 04, Issue: 04, Pages: 27-31, Jan – Feb (2013)
Special Issue on "Computational Intelligence – A Research Perspective" held on "21st -22nd Feburary, 2013"

Applying Computational Intelligence Techniques for Personal Authentication using Keystroke Dynamics

D. Shanmugapriya
Asst. professor, Dept. of Information Technology
Email: shanmugapriya.adu@gmail.com
Dr. G. Padmavathi
Prof and Head, Dept. of Computer science
*Avinashilingam Institute for Home Science and Higher Education for Women,
Coimbatore, India
Email: ganapathi.padmavathi@gmail.com

ABSTRACT

Username with password is the commonly used authentication mechanism for securing huge data transactions being carried out every day via Internet. 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 inexpensive and strong behavioral biometric technologies, which identifies the authenticity of a user when the user is working via a keyboard. The paper uses computational intelligence techniques such as Genetic algorithm, Ant Colony Optimization and Particle Swarm Optimization for feature subset selection and Back propagation Neural Network for classification. From the results, it is observed that ACO wrapped with BPN outperforms the other methods.

Keywords - Ant Colony Optimization, Back Propagation Neural Network, Genetic Algorithm, Keystroke Dynamics, Particle Swarm Optimization

I. INTRODUCTION

Unauthorized access of confidential information can be secured by providing user authentication. User Authentication is normally classified into three categories [5] 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. 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 [1] 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 most commonly measured from the keystroke. The huge features often lead to degradation of performance of the system. So wrapper [12] based computational intelligence techniques such as Genetic algorithm, Ant Colony Optimization and Particle Swarm Optimization with Back propagation Neural Network (BPN) are used for dimensionality reduction and the selected features are classified using BPN.

II. METHODOLOGY

The proposed method is attempted to provide better classification accuracy and to reduce training and testing times. The entire methodology is divided into four phases namely, feature extraction phase, normalization phase, feature subset selection phase and classification phase after obtaining the raw keystroke data (press time and release time).

2.1 Feature Extraction

The feature extraction characterizes the attributes common to all patterns belonging to a class. 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 templates. For the benchmark dataset [13], the features like Duration or Dwell time, Flight time or Latencies, Digraph and Trigraph are extracted. In addition to all the timing features, a new feature called Virtual Key Force has been introduced. The force of the key types is usually measured with the help of special force sensitive keyboard which increases the cost of the equipment. The Virtual Key Force is a feature which is measured from the available keyboard without any extra cost. It is measured from the distance of the key on the keyboard and the time taken to press the key.

2.2 Normalization

The typing pattern of the user varies from time to time even for the same user. There will also be a significant difference in the keystroke patterns exhibited from person to person, even when they are asked to type the same words. Therefore data normalization is done between 0 and 1 to preprocess the data. Z-Score, Min-Max, Zero Mean and Standard Deviation, Median - MAD, and Tanh normalizations are applied and Z-score normalization is found to be effective compared to other normalization methods.

2.3 Feature Subset Selection

Feature subset selection identifies the most selective features. It reduces the dimensionality of features which improves the accuracy and decreases the computation time. Feature subset selection is fundamentally an optimization problem, which concerns searching the space of possible features to be familiar with one that is optimum or near-optimal in accordance with some performance measure [9]. The purpose is to obtain any subset that reduces or to improve a particular measure. After preprocessing with Z-Score normalization, feature subset selection is done for reducing the features for further processing.

Some of the computational intelligent techniques that are suggested to select the subset of features of extracted features are Ant Colony Optimization (ACO), Particle Swam Optimization (PSO), and Genetic Algorithm (GA). These techniques are wrapped with Back Propagation Neural Network (BPN) which can routinely select an appropriate subset of features and the rest will not be considered, thus concluding in a more comprehensive model.

2.3.1 Particle Swarm Optimization – Back propagation Neural Network (BPN) Wrapper approach

Particle Swarm Optimization (PSO) [3, 6, 8 and 10] 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 move 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 particle attempts 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) [71]. The velocity $v_i(t)$ and the positions $x_i(t)$ of the particles are updated with the following equations.

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

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

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 updated using the following equation:

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

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

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

where s = number of particles in the swarm.

In the proposed PSO-BPN Wrapper approach, 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 in order to evaluate the fitness of features. The process is repeated until the best solution is obtained. The Algorithm for proposed PSO-BPN wrapper approach for feature subset selection is given below. The fitness function of the PSO is evaluated using BPN.

Input: *Duration, Latency, Trigraph, Digraph, Virtual Key Force*
Output: *Subset feature values.*
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 BPN.*
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.

2.3.2 Ant Colony Optimization – Back Propagation Neural Network (ACO-ELM) Wrapper approach

Ant Colony Optimization (ACO) [7] algorithms have been introduced based on the observation of the real ant colonies. A unique characteristic of ant colonies is their foraging behavior. The ability of the ants to find the shortest route between their nest and a food source is a very significant character of the ant colonies. Ant colony optimization algorithms have been used to provide better solutions to many search problems.

A wrapper based Ant Colony optimization with BPN has been proposed in this paper for feature subset selection. The Algorithm for ACO-BPN is given below

Input: *Duration, Latency, Trigraph, Digraph, Virtual Key Force*
Output: *Subset feature values.*
Step 1: *Extract the feature values Duration, Latency, Trigraph, Digraph and Virtual Key Force.*
Step 2: *Initialize Number of iterations, Number of Ants, Initial pheromone value associated with each feature and pheromone evaporation rate.*
Step 3: *Select a random feature value for each ant with the criteria that the particular feature value should not have been selected previously by the ant and generate a subset for each Ant.*
Step 4: *Evaluate the selected subset of each ant by* ***BPN.***
Step 5: *Exit, if the number of iterations is more than the maximum number of iterations, otherwise continue.*
Step 6: *Update the pheromone value for features which are selected in step 4.*
Step 7: *Generate new ants by removing the old ants.*
Step 8: *Repeat steps 3 to 7 until the last iteration.*
Step 9: *Set best feature subset as global best.*

2.3.3 Genetic Algorithm – Back Propagation Neural Network (GA-BPN) Wrapper Approach

Genetic Algorithm (GA) [2, 6, 10, 11 and 12] is a stochastic search technique based on natural selection of the population. In Genetic Algorithm, the fitness of all individuals in the population is calculated in every generation by choosing several individuals from the existing population according to their fitness value, and updated to generate a new population. Genetic algorithms combine selection, crossover, and mutation operators in order to find the best solution to a problem. Chromosomes are selected from the population to be parents to crossover.

In the proposed GA-ELM wrapper approach, duration, latency, trigraph, digraph and the proposed virtual key force are given as input features. Feature subset selection is done after crossover and mutation and selected features are evaluated by BPN to find fitness value. The process is repeated until the best solution is obtained. The algorithm of GA-BPN wrapper approach is given below.

Input: *Duration, Latency, Trigraph, Digraph, Virtual Key Force*
Output: *Subset feature values.*
Step 1: *Extract the feature values Duration, Latency, Trigraph, Digraph and Virtual Key Force.*
Step 2: *Initialize Number of generations, Initial population, Crossover rate and mutation rate.*
Step 3: *Generate pool of candidate feature subset.*
Step 4: *Perform crossover and mutation operations and generate a new pool of candidate feature subset.*
Step 5: *Evaluate the selected subset by* ***BPN.***
Step 6: *Selected feature subset by BPN are put back into the population.*
Step 7: *Repeat Steps 3 to 6 until the best solution or maximum iteration is reached.*

The best solutions are achieved in the end.

2.4 Classification

The purpose of classification is to find the best class that is closest to the classified pattern. Neural Networks [4] are one of the most widely used computational intelligence techniques for classification. Neural networks are a computational model motivated by the connectivity of neurons in living nervous systems. Artificial Neural Network (ANN) is observed to be very useful in classification approaches with better accuracy and performance. Among ANN, Back Propagation Neural Network (BPN) [7] is observed to provide the best results in terms of accuracy. The selected features are given as input to classifiers such as Back Propagation Neural Network (BPN). The Algorithm for BPN is given below.

Step 1: *Assign the feature values to input neurons.*
Step 2: *Initialize the weights randomly.*
Step 3: *Calculate the input to the hidden layers by multiplying the weight with input value given to input neuron using the Sigmoid function.*
Step 4: *Calculate the input to the output layers by multiplying the weight with the output value of hidden neuron.*
Step 5: *Calculate the output from output neurons using the Sigmoid function.*
Step 6: *Calculate the delta value by subtracting the output from the target value.*
Step 7: *Update the hidden to output layer weights.*
Step 8: *Update the input to hidden layer weights.*
Step 9: *Repeat steps 3 to 8 until the target output is equal to the desired output.*

III RESULTS AND DISCUSSIONS

The proposed methodology is tested with benchmark dataset [13] which consists of keystroke-timing information from 51 subjects each typing a password '.tie5Roanl' 400 times. Timing features such as Duration, Latency and Virtual Key Force (VKF) are calculated. The password contains 11 characters including shift key to press R. Therefore, a total of 71 features is measured from the password with VKF and 60 features are measured without VKF.

Table 1 shows the number of features selected using the wrapper based feature subset selection methods.

Table 1 Number of Features Selected for DSL Dataset

	PSO				GA				ACO			
	W.O. VKF		W. VKF		W.O. VKF		W. VKF		W.O. VKF		W. VKF	
	F	FS	F	FS	F	FS	F	FS	F	FS	F	FS
BPN	60	55	71	60	60	45	71	50	60	40	71	38

W.O.VKF – Without VKF, W.VKF-With VKF, F- No. of Features, FS – Features Selected

From the table 1, it is observed that ACO-BPN wrapper has reduced the existing 71 features to 38 features and has shown about 46.48% reduction of the total features which are shown in the table 2.

Table 2. Feature Selection Percentage with VKF for DSL Dataset

	PSO	Selection %	GA	Selection %	ACO	Selection %
BPN	60	84.51	50	70.42	38	53.52

After feature subset selection, classification is done. 70% of the selected features are used as training set and the remaining is used as test set. Table 3 shows the average classification accuracy obtained using the wrapper based classification by wrapping PSO, GA and ACO with the classifiers BPN.

Table 3 Percentage of Accuracy obtained with feature subset selection

	BPN	
	Without VKF	With VKF
PSO	89.44	90.66
GA	91.49	91.51
ACO	90.60	91.17

From the table 3, it is observed that the ACO-BPN approach yields the accuracy of 91.17% which outperforms the accuracy obtained using other methods.

IV. CONCLUSION

Back Propagation Neural Network is wrapped with PSO, ACO and GA in order to reduce the number of features and for further classification. ACO-BPsN selects less number of features and gives an accuracy of 91.17% and outperforms other computational intelligence methods. Hence Computational Intelligent techniques provide desirable accuracy for the widely used text based authentication mechanism.

REFERENCES

[1] Ahmed Awad E. Ahmed, and Issa Traore , "Anomaly Intrusion Detection based on Biometrics", *Proceedings of 6th IEEE Information Assurance Workshop*, pp. 452-453, 2005.

[2] Azevedo, G.L.F., Cavalcanti, G.D.C., Carvalho Filho, E.C.B., "Hybrid Solution for the Feature Selection in Personal Identification Problems through Keystroke Dynamics", *International Joint Conference on Neural Networks*, pp. 1947–1952, 2007.

[3] Azevedo, G.L.F., Cavalcanti, G.D.C., Filho, E.C.B., "An Approach to Feature Selection for Keystroke Dynamics Systems based on PSO and Feature Weighting", *IEEE Congress on Evolutionary Computation*, pp. 3577–3584, 2007.

[4] Chen Change Loy, Weng Kin Lai, Chee Peng Lim, "Keystroke Patterns Classification using the ARTMAP-FD Neural Network", *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Vol. 1, pp. 61–64, 2007.

[5] Chen, W., Chang, W., "Applying Hidden Markov Models to Keystroke Pattern Analysis for Password Verification", *Proceedings of IEEE International Conference on Information Reuse and Integration*, pp. 467–474, 2004.

[6] Karnan, M., Akila, M., "Identity Authentication based on Keystroke Dynamics using Genetic Algorithm and Particle Swarm Optimization", *Proceedings of 2nd IEEE International Conference on Computer Science and Information Technology*, pp. 203–207, 2009.

[7] Karnan. M and M. Akila, "Personal Authentication based on Keystroke Dynamics using Ant Colony Optimization and Back Propagation Neural Network", *International Journal of Computer and Network Security (IJCNS)*, Vol. 1, No. 2, pp. 8-15, 2009.

[8] Kennedy. J and Eberhart. R, "Particle Swarm Optimization", *Proceedings of International Conference on Neural Networks*, pp. 1942-1948, 1995.

[9] Robinson, J.A., Liang, V.W., Chambers, J.A.M., MacKenzie, C.L., "Computer User Verification using Login String Keystroke Dynamics", *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol. 28, No. 2, pp. 236–241, 1998.

[10] Sidhartha Panda, N. P. Padhy, "Comparison of Particle Swarm Optimization and Genetic Algorithm for TCSC-based Controller Design", *World Academy of Science, Engineering and Technology*, Vol. 27, pp. 497-505, 2007.

[11] Yang, J., Honavar, V., "Feature Subset Selection using a Genetic Algorithm", *IEEE Intelligent Systems and their Applications*, Vol. 13, No. 2, pp. 44–49, 1998.

[12] Yu. E, Sungzoon Cho, "GA-SVM Wrapper Approach for Feature Subset Selection in Keystroke Dynamics Identity Verification", *Proceedings of the International Joint Conference on Neural Networks*, Vol. 3, pp. 2253–2257, 2003.

[13] http://www.cs.cmu.edu/~keystroke/