GA SVM Wrapper Ensemble for Keystroke Dynamics Authentication

Ki-seok Sung and Sungzoon Cho*

Department of Industrial Engineering, Seoul National University,
San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-744, Korea
{zoro81, zoon}@snu.ac.kr
http://dmlab.snu.ac.kr

Abstract. User authentication based on keystroke dynamics is concerned with accepting or rejecting someone based on the way the person types. A timing vector is composed of the keystroke duration times interleaved with the keystroke interval times. Which times or features to use in a classifier is a classic feature selection problem. Genetic algorithm based wrapper approach does not only solve the problem, but also provides a population of "fit" classifiers which can be used in ensemble. In this paper, we propose to add uniqueness term in the fitness function of genetic algorithm. Preliminary experiments show that the proposed approach performed better than two phase ensemble selection approach and prediction based diversity term approach.

1 Introduction

Keystroke dynamics based authentication (KDA) is concerned with accepting or rejecting someone based on the way that person types. In typing a phrase or a string of characters, the keystroke dynamics or its timing pattern can be measured and used for identity verification. More specifically, a timing vector consists of the keystroke duration times interleaved with the keystroke interval times. The times can be measured in a scale of milliseconds (ms). When a key is stroked before a previous key is released, a negative interval results. When a password of n characters is typed, a (2n +1) dimensional timing vector results, which consists of n keystroke duration times and (n+1) keystroke interval times, with the return key included (see Figure 1).

Feature selection, a major step in pattern classification, determines the minimum number of essential features to be used in building a classifier. There have been some works investigating which elements are useful in KDA, but it seems there is not a clear winner [1]. There are two different feature selection approaches, filter and wrapper approach [2]. In wrapper approach, a subset of features is tentatively selected and fed to a classifier. And this process repeats until a good subset is found (see Figure 2). Combinatorial optimization in the search process is often performed by genetic algorithm, thus it is called GA based wrapper [3]. GA wrapper results in not just one subset of features, but a set of subsets of features (see Figure 3). Repetitive application of genetic operators such as crossover and mutation transforms a randomly generated

* Corresponding author.

D. Zhang and A.K. Jain (Eds.): ICB 2006, LNCS 3832, pp. 654 – 660, 2005.

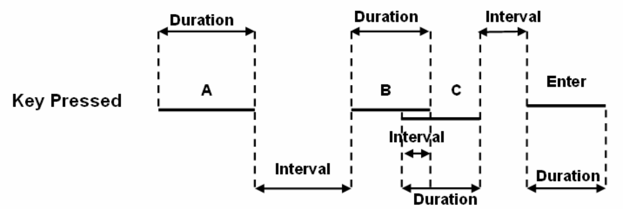

Fig. 1. Timing vector of Password "ABC"

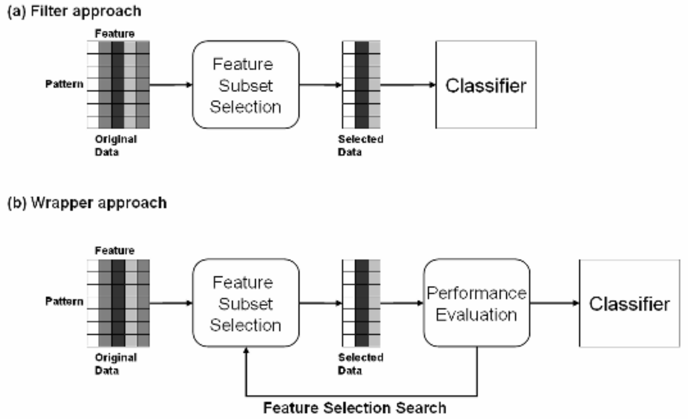

Fig. 2. (a) filter and (b) wrapper approach in feature selection

population of classifiers into a population of highly fit classifiers. In KDA, given a set of timing vectors of D dimension, feature selection tries to find "reduced" yet "optimal" timing vectors of d dimension where d < D. By optimal, we mean achieving the minimum error or highest accuracy of the classifier which employs the reduced set of features. In GA based wrapper approach, a candidate is represented by a D bit binary string. The value of an element is 0 or 1 when the corresponding feature is absent or present, respectively. Started with a randomly generated population of D bit chromosomes, GA process repeats application of evolutionary operations to the population. In the end, fit chromosomes are expected to emerge. The classifiers that correspond to the fit chromosomes are identified and used in the ensemble.

Ensemble is a set of classifiers trained differently: by different data sets, by different features, or by different models [4]. After individual classifiers are trained, they are combined by either majority voting or averaging to output a single value. The performance of an ensemble classifier has been found to be quite high in practice in a variety of applications. Bagging and Boosting are two of the most popular methods [5, 6, 7]. Individual classifiers participating in an ensemble have to be accurate as well as diverse in order to result in a accurate ensemble. It is only natural to combine

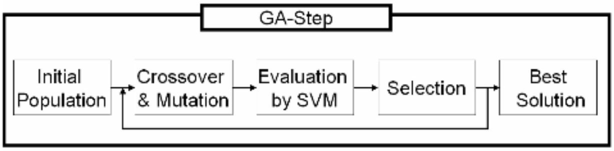

Fig. 3. GA wrapper based feature subset selection

GA wrapper and ensemble since the former generates a population of accurate classifiers. Of course, it has be made sure that they are diverse. So called Genetic Ensemble Feature Selection (GEFS) proposed by Opitz [8], adds a diversity term in the fitness function of GA. The fitness function of genetic algorithm has two terms, the accuracy and diversity:

$$\text{Fitness(x)=A(x)+}\lambda\text{D(x)}. \tag{1}$$

where A denotes accuracy and D denotes diversity with lambda a constant weighing between the two terms.

The accuracy measures how well each neural network predicted of each validation pattern. The diversity measures how different each neural network's prediction is from that of the ensemble. Specifically, the algorithm involves finding a population of neural networks, each of which differs from each other in terms of the predictions. The GEFS performed better than AdaBoost and Bagging for the data sets tested. But major disadvantage of GEFS is that the approach indirectly tries to diversify the population through the difference in prediction. A more direct approach would consider the difference in the features actually employed in each neural network.

Recently, a similar but more elaborate approach has been proposed for KDA by Yu and Cho [9]. Other differences include use of SVM as base classifier for quick training and a different fitness function for GA.

$$\text{Fitness(x)=}\alpha\text{A(x)+}\beta\frac{1}{\text{LrnT(x)}}+\gamma\frac{1}{\text{DimRat(x)}}. \tag{2}$$

where A refers to false rejection rate, LrnT training time, and DimRat dimensionality reduction ratio. If the dimensionality of full feature set was 15, and the dimensionality of currently selected feature subset is 6, for instance, then DimRat(x) = 6/15 = 40%. Since the fitness function clearly does not force diversity, the post processing step was required. Major disadvantage of this approach is that the post processing step involves a time consuming heuristic procedure.

Here in this paper, we propose one step approach similar to that of GEFS, yet with a more direct diversity term in the fitness function and SVM as base classifier and similar to that of Yu and Cho, yet with a diversity term and no more post processing step. In particular, so called "uniqueness" term is used in a fitness function, measuring how unique each classifier is from others in terms of the features used.

This paper is structured as follows. The next section presents the proposed approach. Then, experimental settings and results follow. Finally, a conclusion and future work is discussed.

2 Proposed Method

Contrary to the ordinary GA, GA wrapper has to find not only good strings but also diverse strings. In order to enforce diversity, the fitness function needs a diversity term as in GEFS. What we propose to use here is "uniqueness" term, which measures for each chromosome how different it is from other chromosomes. Since more unique chromosomes are preferred, uniqueness is simply added to accuracy just like the diversity term in GEFS.

Before defining uniqueness, let us define S-distance between the two chromosomes. The S-distance between two chromosomes i and j, $S(d_{ij})$ is defined as follows:

$$S(d_{ij}) = \begin{cases} (\frac{d_{ij}}{C})^2, \text{ if } d_{ij} < C; \\ 1, \qquad\qquad \text{otherwise .} \end{cases} \tag{3}$$

where d_{ij} denotes the Hamming distance between two chromosomes and C a constant. Inspired by sharing function proposed in [11], S-distance is upper bounded at 1.

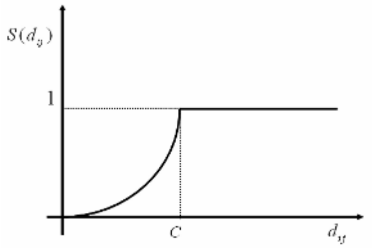

Fig. 4. $S(d_{ij})$ against d_{ij}

Now the uniqueness of xth chromosome is defined as an arithmetic average of S-distances to all other chromosomes.

$$U(\text{x}) = \frac{\sum_{\text{x} \neq j} S(d_{\text{x}j})}{n-1}. \tag{4}$$

Finally, the fitness of chromosome x is defined as a simple sum of accuracy and uniquness:

$$\text{Fitness(x)=A(x)+U(x).} \tag{5}$$

Of course, accuracy A here represents 1 – false-rejection-rate (1-FRR) since only the user's patterns are available in training.

The proposed approach differs from that of Opitz [8] in that diversity is not measured by the indirect approach, difference in the predictions, but by the direct approach, difference in the actual features selected and. The proposed approach differs from that of Yu and Cho [9] in that diversity is introduced in wrapper GA step through the use of uniqueness term so that the subsequent post processing is not necessary and it makes term qualitatively simple.

3 Experimental Setting

The proposed method was applied to 21 sets of password typing patterns used in other research [9, 10]. Even though the original data sets contain hundreds of user's typing patterns, only 50 patterns were used in order to improve the reality of the experiments. Generally, it is hard to expect a user to type a password hundreds of times in enrollment. Out of 50 patterns for each password, 35 of them were used for training while 15 of them were used for validation, in particular to measure FRR in the fitness function of GA wrapper.

It has to be noted that one timing vector set was found to be very poor in its consistency. Figure 5 compares the mean timing vectors of training and test patterns. For "90200jdg" on the left, they are quite different. In particular, note the first, second, sixth and eighth interval values. They are all negative for the test while they are all positive for the training. It is obvious that the user was originally not quite familiar to the password, but later on, after hundreds of typing "practice," he became familiar to it. There is no way to discriminate user and impostor based on the user's past typing patterns if they changed over the time. Thus, we removed this particular password, "90200jdg," in the experiment.

In order to understand the performance of the proposed approach, we also implemented related approaches: the work of Opitz and that of Yu and Cho. Even though Yu and Cho also used the same data set, they used a randomly selected 50 patterns. So we performed experiment again with the different 50 patterns.

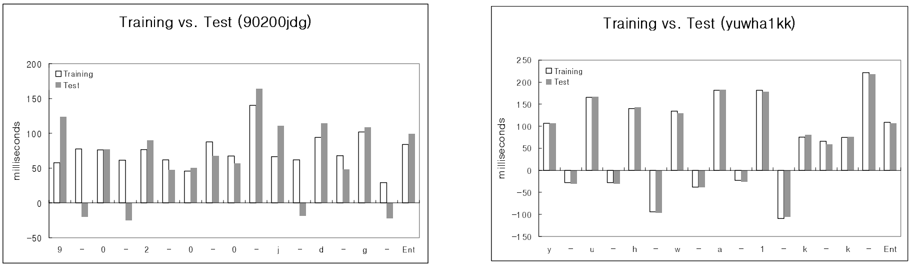

Fig. 5. Comparison of training and test timing vectors of two passwords "90200jdg" and "yuhwa1kk"

A population of 100 chromosomes was run 50 generations with cross over rate of 0.2 and mutation rate of 0.01. The SVM employed Gaussian kernel. The values of its parameters γ, cost, υ were determined in an empirical way. Of course, these values were shared also by all three approaches. The C value for the proposed approach was set to 30% of the original dimension. Early stopping criterion and classifier HD percentage used by Yu and Cho approach was set to 0.2 and 30%, respectively.

4 Results

Table 1 shows the performance of three approaches for 20 password timing vector sets. Since GA is stochastic in nature, five GA runs were made for each. The every entry in the table is an average from the five runs. There are 75 user's test patterns and 75 impostor patterns. They were used to calculate the accuracy, false acceptance rate and false negative rate. The number of ensemble denotes the number of classifiers in ensemble. The proposed approach and Opitz approach has a same fixed number, but Yu Cho approach has various numbers since it is the post processing phase that determines the exact number of classifiers in ensemble. On average, the proposed approach results in the best numbers, closely followed by that of Opitz and Yu and Cho in that order, although the difference may not be statistically significant. The FAR is much smaller than the FRR in the proposed approach, which is quite desirable considering that FAR is much more costly than FRR. By comparing best performing approach of each password, the proposed approach was best by coming first in nine passwords.

Table 1. Performance of three approaches

Password	Models											
	Sung-Cho Fitness = A(x) + U(x)				Yu-Cho Fitness = 10A(x) + 1/(100 * LrnT(x)) + 1/DimRat(x)				Opitz Fitness = A(x)+D(x)			
	Ensemble Accuracy	FAR	FRR	Num of Ensemble	Ensemble Accuracy	FAR	FRR	Num of Ensemble	Ensemble Accuracy	FAR	FRR	Num of Ensemble
ahrfus88	89.60	5.86	14.93	31	80.00	36.26	3.73	10.20	89.60	4.80	16.00	31
anehwksu	90.53	1.60	17.33	31	86.26	12.80	14.66	13.40	90.53	3.46	15.46	31
autumnman	93.60	0.00	12.80	31	92.00	10.93	5.06	11.40	92.93	0.53	13.60	31
beaupowe	86.00	17.33	10.66	31	78.00	38.66	5.33	9.20	85.33	21.86	7.46	31
c.s.93/ksy	93.20	1.33	12.26	31	92.66	6.66	8.00	23.60	93.60	1.33	11.46	31
dhfpql.	94.40	0.00	11.20	31	95.73	1.33	7.20	11.20	96.00	0.00	8.00	31
dirdhfmw	96.93	0.00	6.13	31	98.13	0.80	2.93	11.20	96.26	0.00	7.46	31
dlfjs wp	85.46	0.00	29.06	31	93.06	1.86	12.00	12.40	85.60	0.00	28.80	31
dltjdgml	90.93	0.00	18.13	31	95.73	1.60	6.93	10.60	91.86	0.00	16.26	31
drizzle	92.13	6.66	9.06	31	87.46	21.06	4.00	11.60	91.33	6.13	11.20	31
dusru427	90.13	0.00	19.73	31	93.06	1.33	12.53	15.40	90.53	0.00	18.93	31
i love 3	94.93	1.06	9.06	31	91.06	10.66	7.20	8.60	95.06	1.06	8.80	31
love wjd	88.80	14.13	8.26	31	84.40	27.20	4.00	11.80	86.13	11.20	16.53	31
loveis.	92.13	8.00	7.73	31	89.06	20.00	1.86	12.40	91.06	7.20	10.66	31
manseiii	83.06	18.40	15.46	31	74.00	46.13	5.86	13.00	81.33	24.53	12.80	31
rhkdwo	93.06	0.53	13.33	31	93.60	4.53	8.26	7.80	92.53	0.80	14.13	31
rla sua	97.20	1.86	3.73	31	89.73	16.80	3.73	10.80	95.86	3.46	4.80	31
tjddmswjd	90.93	0.26	17.86	31	91.20	2.40	15.20	14.40	90.13	0.00	19.73	31
tmdwnsl1	90.26	0.00	19.46	31	93.60	1.60	11.20	11.00	91.20	0.00	17.60	31
yuhwa1kk	97.06	0.00	5.86	31	97.33	0.00	5.33	11.80	96.53	0.00	6.93	31
Min	83.06	0.00	3.73	31	74.00	0.00	1.86	7.8	81.33	0.00	4.80	31
Max	97.20	18.40	29.06	31	98.13	46.13	15.20	23.6	96.53	24.53	28.80	31
Average	91.52	3.85	13.10	31	89.80	13.13	7.25	12.09	91.17	4.32	13.33	31

5 Conclusion and Future Work

In this paper, we proposed a GA based wrapper approach to be applied to keystroke dynamics based authentication. Compared to the previous work by Yu and Cho, we proposed to introduce diversity of the population by adding a term in fitness function that measures the uniqueness of a chromosome. This renders a rather complicated post processing unnecessary. Compared to the work by Opitz, we used one class SVM as base classifier and forced diversity through the uniqueness of each chromosome. A preliminary experiment involving 20 passwords shows that the proposed approach performed best. It is our contribution that a simpler approach produced a slightly better or similar performance.

There are limitations to the approach. First, the SVM used as a base classifier does not involve a threshold thus a balance between FAR and FRR cannot be controlled. We can indirectly control FRR in training, by using training parameters, like γ and cost. Second, fitness is computed as a sum of accuracy and diversity. A multi-objective optimization technique can be used instead. Third, removing outliers from user's training patterns might help achieve better performance.

Acknowledgements

This work was supported by grant No.R01-2005-000-103900-0 from the Basic Research Program of the Korea Science and Engineering Foundation.

References

1. Araujo, L., Sucupira, L., Lizarraga, M., Ling, L., and Yabu-Uti, J.: User Authentication through Typing Biometrics Features. IEEE Transactions on Signal Processing. 53(2), (2005) 851-855
2. Liu, H., Motoda, H. : Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic Publishers (1998)
3. Yang, J., Honavar, V. : Feature Subset Selection using a Genetic Algorithm. in Feature Selection for Knowledge Discovery and Data Mining. Liu, H. and Motoda, H. (eds.), Kluwer Academic Publishers, (1998) 117-136
4. Dietterich, T. G. : Ensemble methods in machine learning. First International Workshop on Multiple Classifier Systems, (2000) 1-15.
5. Breiman, L. : Bagging predictors. Machine Learning. 24(2), (1996) 123-140
6. Freund, Y., Schapire, R.E. : Experiments with a new boosting algorithm. Proceedings of the 13th International Conference on Machine Learning. Morgan Kaufmann, (1996) 148–156.
7. Sullivan, J., Langford, J., Caruana, R., Blum, A. : Featureboost: A meta-learning algorithm that improves model robustness. Proceedings of the Seventeenth International Conference on Machine Learning. (2000)
8. Opitz, D. : Feature selection for ensembles. AAAI/IAAI, (1999) 379-384.
9. Yu, E., Cho, S.: Keystroke dynamics identity verification - its problems and practical solutions. Computers and Security. 23(5) (2004) 428-440
10. Cho, S., Han, C., Han, D., Kim, H.: Web-based keystroke dynamics identity verification using neural network. J. Organizational computing and electronic commerce. 10(4) (2000) 295-307
11. Srinivas, N., Deb, K. : Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms, Evolutionary Computation, 2(3) (1994) 221-248