대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002년 5월 3일-4일

4층 자기연상 다층퍼셉트론을 이용한 키스트로크 기반 사용자 인증
A 4-layer AaMLP for Keystroke Dynamics Identity Verification

우은철 조성준
{ enzhe@snu.ac.kr, zoon@snu.ac.kr }

서울대학교 산업공학과, 서울 관악구 신림동 산 56-1 151-744

ABSTRACT Password typing is the most widely used identity verification method in computer security domain. However, due to its simplicity, it is vulnerable to imposter attacks. Keystroke dynamics adds a shield to password. Discriminating imposters from owners is a novelty detection problem. Auto-Associative Multilayer Perceptron (AaMLP) has been proved to be a good novelty detector. However, the widely used 2-layer AaMLP cannot identify nonlinear boundaries, which can result in serious problems in computer security. In this paper, a nonlinear model, i.e. 4-layer AaMLP, is proposed to serve as the novelty detector, which can remedy the limitations of 2-layer AaMLP.

Key Words: User Authentication, Keystroke Dynamics, Novelty Detection, Autoassociative MLP.

1. INTRODUCTION

Keystroke dynamics is a biometric-based approach that utilizes the manner and rhythm in which each individual types passwords to create a biometric template. It measures the keystroke rhythm of a user in order to develop a template that identifies the authorized user. When a user types a word, for instance a password, the keystroke dynamics can be characterized by a "timing vector", consisting of the durations of keystrokes and the time intervals between them. The owner's timing vectors are collected and used to build a model that discriminates between the owner and imposters. This idea steps originally from the observations that a user's keystroke pattern is highly repeatable and distinct from others'. The only disadvantage has been a relatively low rate of accuracy.

All biometrics-based approaches have two types of error, the false acceptance rate (FAR) and the false rejection rate (FRR). FAR denotes the rate that an imposter is allowed access, and FRR denotes the rate that the legitimate user is denied access. Because one type of error can be reduced at the expense of the other, an appropriate tradeoff point is usually used as a threshold based on the relative cost of the errors.

In 1980, Gaines et al. [4] first proposed the approach using keystroke dynamics for user authentication. Experiments with a population of 7 candidates were conducted. Later on, Leggett et al.[8] conducted similar experiments by applying a long string of 537 characters, and reported a result of 5.0% FAR and 5.5% FRR. Recently, through the use of neural networks, a comparable performance of 12% to 21% was achieved using short strings such as real-life names [4]. Obaidat et al. reported a 0% error rate in user verification using 7-character-long login name [9]. However, both the imposter's typing patterns and the owner's patterns were used training, and the training data set was much larger (6300 owners and 112 negatives). Also, the training and test patterns were not chronologically separated. In [3], a novelty detection model was built by training owner's patterns only, and was used to detect imposters using some sort of a similarity measure, and a 1.0% FRR and 0%FAR was reported. Furthermore, some products have been marketed, such as Net Nanny's BioPassword (http://www.biopassword.com).

In this paper, we propose a 4-layer autoassociative multilayer perceptron (AaMLP) to improve the performance of the novelty detector. Timing vectors from an owner were collected and used to build a neural network model that outperformed a generally applied neural network model, 2-layer AaMLP.

This paper is structured as follows. In session 2, descriptions on the neural network based novelty detector , the descriptions of the limitations of 2-layer AaMLP and the proposed 4-layer AaMLP are presented. After the explanation of the data, experimental results are shown in session 3. A summary and acknowledge conclude this paper.

2. AUTOASSOCIATIVE MULTILAYER PERCEPTRON NOVELTY DETECTOR

2.1 AaMLP for User Authentication

From the pattern classification viewpoint, user authentication can be regarded as a two-class (owner vs. imposter) problem. Yet the patterns from only one class, the owner's, are available in advance. Because there are millions of potential imposters, it is not practical to obtain enough patterns from all kinds of imposters. Also, it is not desirable to publicize one's password to collect potential imposters' timing vectors. The only solution is to build a model of the owner's keystroke dynamics and use this to detect imposters using some sort of a similarity measure. This type of problem has been known as partially exposed environment or novelty detection. Usually, a model of normal conditions is built and then used to detect abnormality or novelty. In novelty detection, many neural network approaches have been adopted. Among them, major approaches are auto-associative multi-layer perceptron (AaMLP) and SOM, which build models for owners only.

In an AaMLP, the input vectors are also used as targets during training, and the network is forced to encode the input vector in the hidden layer and then decode it back in the output layer. This model can be used for identity verification as follows. The owner's patterns are use to train the network to become an autoassociator by employing a timing vector as both an input and output. The AaMLP is trained to learn to encode certain properties only present in the owner's timing vectors at the hidden layer. When a previously 'unseen' timing vector for the owner arrives, the network will output a vector that is reasonably close to the input. When an imposter's pattern arrives, the network will output a vector that is far from the input. Then, it is possible to distinguish a pattern as either genuine or forged. This can be measured by the closeness of the vectors to the owner's pattern, that is, a timing vector X is classified as the owner's if and only if,

$$\|X-M(X)\| < \varepsilon,$$

where *M(X)* and ε denote the MLP's output for *X* and a threshold.

Several applications have reported satisfactory performance of AaMLP in novelty application [8, 9]. Especially, Cho et al. applied a 2-layer AaMLP for user authentication through keystroke dynamics [3].

2.2 Limitations of 2-layer AaMLP

In [5], Hwang and Cho studied the properties of AaMLP that are essential for a novelty detector: (1) Uncountably infinite input vectors exist for which AaMLP produces the same output vector; (2) The "output-constrained hyperplane" exists on which all the output vectors are projected. As long as AaMLP uses a bounded activation function such as a step function, the output-constrained hyperplane is bounded; (3) Minimizing the error function leads the hyperplane to be located in the vicinity of the training pattern, etc.

However, a 2-layer MLP is computationally limited since all the output vectors are projected onto a hyperplane. In a situation like Figure 1, where the distribution of the training patterns (shaded area) is concave or nonlinear, the misclassification rate will increase greatly. The output constraining hyperplane is denoted by O. All the patterns located inside the surrounding ellipse are classified as "normal." Thus, these patterns from the areas denoted as A, B, and C are incorrectly classified. In a security problem like computer access or electronic commerce, such false acceptance of imposters, i.e. A, B and C, is very dangerous and must be avoided. In order to overcome such shortcomings of 2-layer AaMLP, a 4-layer AaMLP model, which is capable of nonlinear reconstruction, is proposed.

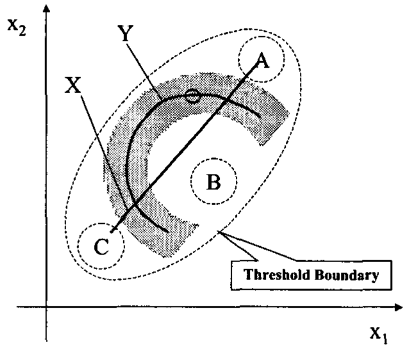

Figure 1 Misclassification resulted from a concave distribution patterns

2.3 The Proposed Approach – 4-layer AaMLP

In a 2-layer AaMLP, although a sigmoid activation function is used in the hidden layer, but it only plays the role of bounding the output value, and the model cannot reflect the nonlinearity of the input patterns. The serious problem shown in Figure 1 is due to the incapability of 2-layer AaMLP's nonlinear pattern mapping. However, if the network can map the input patterns onto the curve Y, rather than the line X, then such misclassification problem is solved.

A 4-layer AaMLP, is supposed to model the nonlinear input patterns, therefore improve the novelty detection capability of the network. The structure of a 4-layer AaMLP is shown in Figure 2. In the *mapping* and *de-mapping* process, a sigmoid activation functions are used, and linear activation functions are applied to other layers. Specially, the output b_i of the unit *i* in the bottleneck layer, and the output y_l' of unit *i* in the output layer are computed as follows:

$$b_k = \sum_j w_{kj} f_\sigma\left(\sum y_i w_i\right) \qquad (1)$$

$$y_l' = \sum_j w_{lj} f_\sigma\left(\sum w_{ji} b_i\right) \qquad (2)$$

where $f_\sigma(x) = \dfrac{1}{1+e^{-x}}$, and w_{kj} is the connection strength from unit *j* to unit *k*.

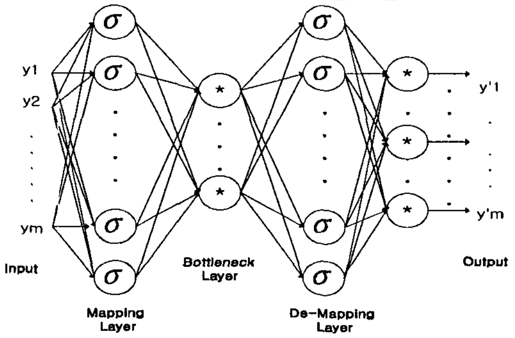

Figure 2 The structure of 4-layer AaMLP

The bottleneck layer is the one that has the least number of nodes. Through learning, a "redundancy compression and non-redundancy differentiation" effect appears. Ikbal et al. [6] studied the characteristics of AaMLP, and argued that as network size increases, the nonlinearity level of the subspace and hyper-surface increases accordingly.

In the past researches, 4-layer AaMLP is mainly applied for the purpose of dimension reduction [7]. The compressed dimensions are from the bottleneck layer, which is extracted by way of nonlinear PCA (NLPCA). Some researches have reported the limitations of 4-layer AaMLP in dimension reduction. One of the limitations is that the 4-layer AaMLP shows strong capability in interpolation, but is weak in extrapolation. However, given enough *normal* patterns, a 4-layer AaMLP is supposed to give excellent novelty detection performance. Accordingly, the limitation raised from dimension compression will play an important role in novelty detection. In [1], a 4-layer AaMLP was applied in financial prediction as a long-rising pattern detector. A simple investment strategy based on the detector achieved a two-year return of 39.4% in comparison with 19.1% return from a sell and hold strategy.

3. EXPERIMENT RESULTS

Experiments were carried out to compare the novelty detection capability of the proposed 4-layer AaMLP with that of the linear model -- 2-layer AaMLP.

3.1 Data Collection

The data was captured by a program in X window environment on a Sun Sparcstation, in which the keystroke duration times and interval times is measured. The keystroke duration and interval times are captured at the accuracy of milliseconds (ms). A timing vector consists of keystroke duration times and interval times. A password with n-character long, length of the timing vector would be *(2n+1)*, where the *Enter* key is also included. For instance, a password *abcd*, which is 4-character long (n=4), together with the *Enter* key, results in a timing vector of 9 dimensions. An example of a timing vector is [30, 60, 70, 135, 60, -35, 75, 40, 55]. When the next key is stroked before the previous key is released, the keystroke interval time is represented as negative (<0).

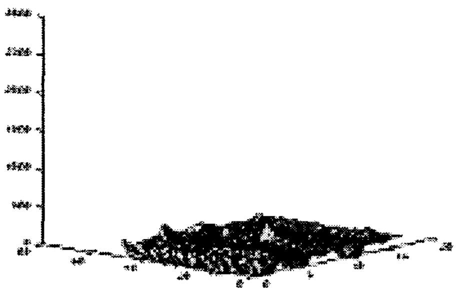

Figure 3 Example of owner's patterns

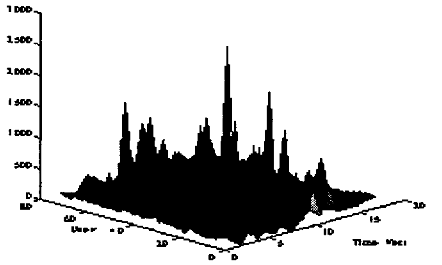

Figure 4 Example of imposters' patterns

The data for both the owners and the imposters were collected. The owners' data was collected from 25 participants with different passwords, whose length ranges from 6 to 10. Each participant was asked to type his password 150 to 400 times, and the last 75 timing vectors were collected for testing, whereas the remaining ones were used as training patterns. As for the novelty data, 15 imposters were asked to type each of the given 21 passwords 5 times without any practice, resulting in 75 impostor timing vectors for each password. We call those imposters as *"imposters without practice."* Together with the owners' test patterns mentioned above, two groups of 75 patterns, i.e. normal and novelty, are ready for each password.

Furthermore, these imposters were given passwords beforehand, and were asked to practice typing these passwords. After that, 21 sets of timing vectors from practiced imposters were collected, each of which consists of 75 timing vectors as was mentioned above. These are called 'imposter with practice.' Figure 3 and Figure 4 illustrated timing vectors of a certain password for the owner and imposters, respectively. The practiced imposters generated very similar patterns to those of the owners'. With the limitations described in *section 2.2*, a 2-layer AaMLP cannot report a good novelty detection performance here. However, the proposed 4-layer AaMLP, which is a nonlinear autoassociator, is supposed to give satisfactory results. In fact, the experiment results reported in the later section proved this assumption.

3.2 Data Preprocessing

A novelty detection model is built under the assumption that the owner's typing follows a consistent pattern. However, there appeared some problems with the original data due to owner's inconsistency, which are illustrated in Figure 5.

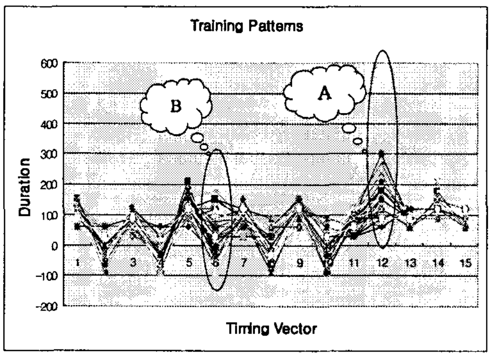

Figure 5 Problems with the training patterns

In situation A, some elements showed large deviation from others. Patterns with these elements were regarded as outliers and discarded. In fact, an upper 10% was adopted as a cut-criterion in this situation. Furthermore, if such deviation happens on both sides, i.e. upper and lower, a 5% was adopted as a cut-criterion for each. In situation B, elements in certain dimension are mainly bifurcated into 2 directions. In such case, we have to decide the main 'direction' for that dimension. Patterns with the opposite direction were all discarded. In fact, building a more complex network for such problematic data, i.e. situation B, would also be very interesting. An ensemble method, in which individual networks are constructed for each direction, is supposed to model such patterns.

According to the owner's consistency, a discard rate of 20%~50% was applied. Since in model building, only the experienced owners are considered, those with low quality patterns were not included in out experiment. As a result, 21 owners and corresponding imposters' patterns were used for the experiment.

3.3 Experiment Results

Two models were built for every owner, i.e. 2-layer AaMLP and 4-layer AaMLP. All the networks, i.e. 2-layer and 4-layey, were trained with *Resilient backpropagation* algorithm, with a learning rate of 0.1, a momentum term of 0.25.

For the 2-layer networks with the structure of $N - h$ - N, the number of hidden nodes ranged from 6 to 8 according to the performance of the network over different patterns. A 4-layer AaMLP has the structure of $N - l - h - l$ -N, where N is the input dimension, l is the mapping or de-mapping layer, and h is the bottleneck. Depending on the input pattern, number of the nodes in l ranged from 12 to 25, and that in h ranged from 6 to 10. As we can see from *Table 8.1*, some owners only have a small number of useful patterns after data cleaning, say less than 100. In such a case, a *10-fold crossvalidation* method was applied. The final structure of the network is determined by the validation set. Performances of the models were measured by (a) the *reconstruction error*, i.e. MSE, and (b) FRR, when FAR is reduced to zero.

Let *min(imposter)* denotes the minimum value of imposters' reconstruction errors, and m*ax(owner)* the maximum value of the owner's reconstruction errors. The s*eparation degree,* i.e. value of (m*in(imposter)* - *max(owner))*, is applied as a measure for evaluating the performance of the model. The larger the *separation degree* is, the better the model is. Figure 6 illustrates the distribution of reconstruction errors and the concept of *separation degree.*

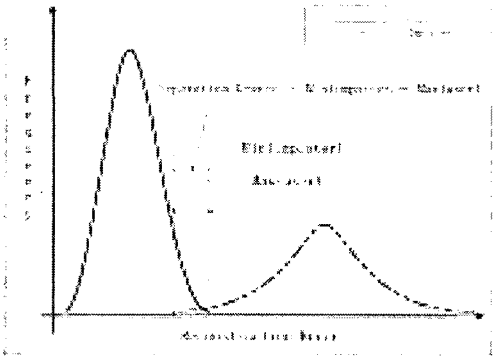

Figure 6 Histogram of Reconstruction Error

Such *separation degree* is measured in two situations: (1) *owners vs. imposters without practice*, and (2) *owners vs. imposters with practice.* 10 experiments were carried out for each ID, and the performances are evaluated by the average values.

For *situation (1)*, *separation degree* is shown in Figure 7 and Figure 8: Figure 7 for 2-layer AaMLP and Figure 8 for 4-layer AaMLP. As we can see from these figures, 4-layer AaMLPs outperformed 2-layer networks greatly in terms of *separation degree.* Almost all the separation degrees of 2-layer AaMLP are negative (<0), which means that there are overlaps between owner and imposter test vector histogram, whereas the 4-layer models gave fewer overlaps.

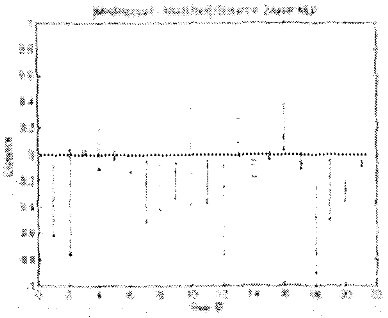

Figure 7 Separation degree of 21 Ids with unpracticed imposters : 2-layer MLP

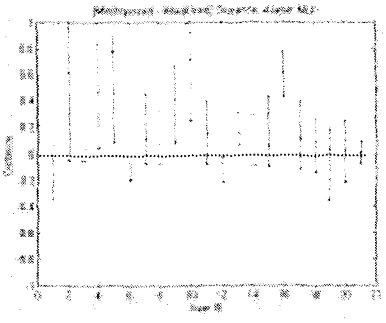

Figure 8 Separation degree of 21 Ids with unpracticed imposters: 4-layer MLP

Similar comparisons for *situation (2)*, that is, *owners vs imposters with practice*, are also shown in Figure 9 and Figure 10, respectively. 4-layer AaMLP beat 2-layer AaMLP in all the owners' passwords except two – *owner 3* and *owner 11*. This is probably due to the inconsistency of the owners' test patterns. As was pointed out, one of the characteristics of 4-layer AaMLP is that it is strong in interpolation, whereas weak in extrapolation.

When the owner typed inconsistently, those patterns will be regarded as novelties, thus result in very large errors. With such a characteristic and sufficient training data, 4-layer AaMLP is supposed to have a much better performance. For *owner 4,10,13, 15,16,17* and *18*, the 4-layer AaMLPs gave perfect authentication (no overlap between owner and imposter test vector histograms) while 2-layer AaMLP only reported on perfect performance, i.e. for *owner 10*.

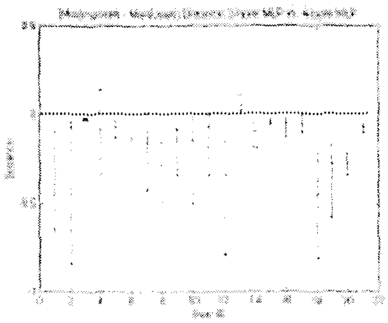

Figure 9 Separation degree of 21 Ids with practiced imposters: 2-layer MLP

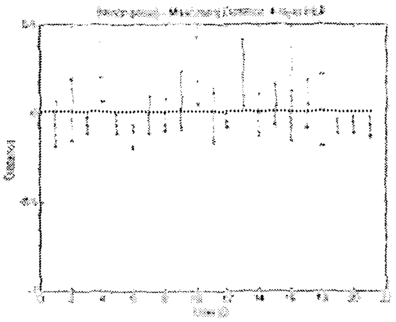

Figure 10 Separation degree of 21 Ids with practiced imposters: 4-layer MLP

Shown in Table 1 are the error rate comparisons for 2-layer AaMLP and 4-layer AaMLP, in terms of FRR when FAR was reduced to *zero*.

For the *unpracticed imposters,* 4-layer AaMLP showed 18 perfect authentications out of 21 owners. The worst performance was with the error rate of 2.67%. The average error rate was 0.25%. However, 2-layer AaMLP only achieved 6 perfect authentications. The worst performance was 4.00%, and the average error rate was 1.71%. In this situation, i.e. owner *vs* unpracticed imposters, though 4-layer AaMLP performed better in general, no significant difference was shown between the two models.

In the situation of *owners vs practiced imposters,* 4-layer AaMLP showed its advantages over 2-layer AaMLP. 4-layer AaMLP showed an average performance of FRR=1.21, with the worst error rate of 4.00%, average error rate of 1.21% and 9 perfect authentications. While 2-layer AaMLP reported worse: the maximum error rate of 17.33%, average error rate of 5.71%.

The paired comparison hypothesis tests were performed for both situations, i.e. *owners vs unpracticed imposters*, *owners vs practiced imposters*. The hypothesis was set as H0: $\mu_d = 0$, and H1: $\mu_d > 0$, where random variable d denotes the difference of error rates from the two approaches, that is, $e_{2\text{-layer}} - e_{4\text{-layer}}$. The t-statistic values of 5.6405 much larger than 2.528 = $t_{0.01}$ for *owners vs unpracticed imposters,* and 4.4191 much larger than 2.528 = $t_{0.01}$ for *owners vs practiced imposters,* indicate that H0 is rejected with much more than 99.5% confidence. In conclusion, the superiority of MLP approach's performance is statistically significant.

As a whole, the proposed nonlinear 4-layer AaMLP showed its advantages in novelty detection over 2-layer AaMLP. When the imposters are not practiced beforehand, the two models showed similar performance. However, with the practiced imposters, whose patterns are very similar to the owners', 2-layer AaMLP cannot report satisfactory performances, while 4-layer AaMLP did.

Table 1 Performance Comparison for the Respective Modes

Owner ID	Dimension of timing vector	Number of training Patterns	FRR* (FAR=0) imposter without practice		FRR* (FAR=0) imposter with practice	
			2L AaMLP	4L AaMLP	2L AaMLP	4L AaMLP
Atom	15	178	4.00	1.33	10.67	2.67
Bubugi	17	312	1.33	0.00	4.00	2.67
Celavie	17	330	0.00	0.00	1.33	4.00
Crapas	19	165	0.00	0.00	2.67	0.00
Dry	19	328	0.00	0.00	1.33	1.33
Flower	13	202	2.67	1.33	2.67	1.33
Gmother	17	101	2.67	0.00	4.00	1.33
Gusegi	15	231	2.67	0.00	9.33	1.33
Jmin	17	95	2.67	0.00	5.33	0.00
June	17	144	0.00	0.00	2.67	0.00
Jywoo	15	297	1.33	0.00	1.33	1.33
Megadeth	17	329	4.00	2.67	14.6	1.33
Oscar	17	365	0.00	0.00	0.00	0.00
Perfect	17	86	2.67	0.00	4.00	0.00
Shlee	17	309	1.33	0.00	2.67	0.00
Sjlee	13	205	0.00	0.00	6.67	0.00
Woo	13	143	1.33	0.00	4.00	0.00
Wooks	17	81	4.00	0.00	17.33	0.00
Yanwenry	17	108	2.67	0.00	9.33	2.67
Ysoya	17	260	1.33	0.00	10.67	1.33
Zeronine	21	135	1.33	0.00	5.33	2.67
Average			1.71	0.25	5.71	1.21
Minimum			0.00	0.00	0.00	0.00
Maximum			4.00	2.67	17.33	4.00

* FRR = False Rejection Rate, FAR = False Acceptance Rate, measured by *percentage* (%).

4. CONCLUSION

In this article, a nonlinear novelty detector, i.e. 4-layer AaMLP is proposed to improve the performance of user authentication using keystroke dynamics. The performance of the proposed model was compared with the commonly used 2-layer AaMLP, which is a linear model.

Experiments were carried out in two situations, i.e. (a) owner *vs* unpracticed imposters, and (b) owner *vs* practiced imposters. 4-layer AaMLP beat 2-layer AaMLP in both situations. Especially, in situation (b), where the imposters' patterns are very similar with the owners', 4-layer AaMLP reported an average error rate of 1.21%, with 9 perfect authentications. However, 2-layer AaMLP performed badly with an average error rate of 5.71%, with maximum one 17.33%. Generally, the proposed 4-layer AaMLP beat 4-layer AaMLP in all aspects.

Further investigation is necessary regarding the following issues: First, the quality of the owner's patterns must be satisfied. If possible, more participants are preferred for experiments. Second, in order to reduce the complexity of the model, a feature extraction or dimension reduction method shall be applied. Third, model selection process needs to be automated. Since a different password requires a different network structure, an automated optimization method, i.e. genetic algorithm, for the selection of network structure is also a requirement. Fourth, further study on nonlinear pattern modeling is needed. In the current study, no multi-modal patterns are taken into consideration. An Ensemble method may be effective.

5. ACKNOWLEDGEMENT

This research was supported by Brain Science and Engineering Research Program sponsored by Korean Ministry of Science and Technology.

REFERENCES

[1] Jinwoo Baek and Sungzoon Cho, Time to Jump in?: Long Rising Pattern Detection in KOSPI 200 Future Using an Auto-Associative Neural Network, 160~165, ICONIP 2001, Shanghai, China, Nov. 14-17, 2001.

[2] M. Brown and S. J. Rogers, User identification via keystroke characteristics of typed names using neural networks, *International Journal of Man-Machine Studies, vol. 39*, pp. 999-1014, 1993.

[3] Sungzoon Cho, C. Han, D. Han, and H. Kim, Web-based keystroke dynamics identity verification using neural network, *Journal of organizational computing and electronic commerce 10(4)*, 295-307, 2000.

[4] R. Gaines, W. Lisowski, S. Press, and N. Shapiro. Authentication by keystroke timing: some preliminary results. Rand Report R-256-NSF. Rand Corporation, 1980.

[5] B. Hwang and S. Cho, "Output characteristics of autoassociative MLP and its application in novelty detection," Proc. Of Korea Inforamtion Science Society, vol. 25, no. 11, pp. 581-583, 1998.

[6] M. Ikbal, H. Misra, and B. Yegnanarayana, Analysis of sutoassociative mapping neural networks, Proceedings of International Joint Conference on Neural Networks, #854, 1999.

[7] M. Kramer, Nonlinear principal component analysis using autoassociative neural networks, *AIChE Journal, Vol. 37, No. 2,* Feb. 1991.

[8] J. Leggett, G. Williams, M. Usnick, and M. Longnecker, Dynamic identity verification via keystroke characteristics, *International Journal of Man-Machine Studies, vol. 35*, pp. 859-870, 1991.

[9] M. Obaidat and S. Sadoun, Verification of computer users using keystroke dynamics, *IEEE Transactions on Systems, Man and Cybernetics, Part B:P Cybernetics, vol. 27*, no. 2, pp. 261-269, 1997.