

# 신경망을 이용한 휴먼 타이핑 패턴 인식

## Recognition of Human Typing Pattern Using Neural Network

배중기\*, 김병환\*\*, 이상규\*\*\*  
 Junggi Bae, Byungwhan Kim, Sang Kyu Lee

**Abstract** - With the increasing danger of personal information being exposed, a technique to protect personal information by identifying a non-user in case it is exposed. A study to construct a neural network recognizer for developing a economical and effective user protecting system. For this, time variables regarding user typing patterns from a pattern extraction device. With the variations in the standard deviation for the collected time variables, non-user patterns were generated. The recognition performance increased with the increase in the standard deviation and a higher recognition was achieved at 2.5. Also, five types of training data were generated and the recognition performance was examined as a function of the number of non-user patterns. With the increase in non-user patterns, the recognition error quantified in the root mean square error (RMSE) was reduced. The smallest RMSE was obtained at the type 5 and 90 non-user patterns. In overall, the type 3 model yielded the highest recognition accuracy. Particularly, a perfect recognition of 100% was achieved at 45 non-user patterns.

**Key Words** : 신경망, 타이핑 패턴, 모델, 보안

### 1. 서론

컴퓨터 등의 IT 기기에서 사용자 인식을 위해서 ID 와 PASS 등의 개인 정보가 사용되고 있다. IT가 발달하고 개인 정보의 노출 빈도가 높아짐에 따라 개인 정보의 유출 가능성이 증대되고 있다. 개인 정보의 유출로 인한 정신적, 경제적 고통 방지하기 위해서 사용자 인식을 통한 유출된 개인 정보의 이용을 방지할 수 있는 기술 개발이 요구되고 있다. 사용자 인식을 위해 지문 인식, 홍채 인식 등을 위한 별도의 인식 시스템을 설치할 수 있으나 추가적인 하드웨어가 필요하여 구성이 복잡해지며 비용이 증가함에 따라 현실적인 제약이 따른다. 이에 기존의 하드웨어를 그대로 사용하고 추가적인 사용자의 개인정보를 요구하지 않는 경제적이고 효과적인 사용자 인식 시스템의 개발이 요구되고 있다. 이러한 시스템의 개발을 위해 신경망의 응용을 제안한 바 있다 [1-2]. 미국 특허에서는 단일 출력을 가지는 신경망을 이용하였으며, 포항공대가 제안한 구조는 다중 출력을 가지는 신경망 인식기를 제안한 바 있다. 인식기 성능은 신경망 학습인자 이외에 학습 데이터 구성방식에 의해 상당한 차이가 있지만, 이 두 특허에서는 구체적인 방식을 제시하지 못하고 있어 이에 대한 체계적인 연구가 요구된다.

본 연구에서는 감독 학습 기능을 가진 신경망과 사용자 고유의 사용자 타이핑 패턴에 대한 시간변수를 사용하여 새로운 사용자 인식 모델을 개발하였다. 사용자 패턴의 시간변수의 평균과 표준편차를 이용하여 비특이 패턴을 생성하여 표준편차에 따른 모델 성능을 검증하였다. 수집된 사용자 타이핑 패턴의 시간변수를 5 종류 유형으로 분류하여 신경망의 입력으로 이용하였다. 인식 오류를 접근 권한이 있는 사용자를 침입자로 판단하는 1종 인식 오류와 접근 권한이 없는 사용자를 접근 권한이 없는 사용자로 판단하는 2종 인식 오류로 구분하여 인식 성능을 평가하였다.

### 2. 데이터 수집

신경망 인식 모델의 개발을 위해서 수집되는 사용자 타이핑 패턴의 시간변수는 표 1에 나와 있는 3종류의 시간변수가 수집되었다.

표 1 사용자 패턴의 시간 변수

수집된 시간 변수의 종류	단위
1) 각 키의 press / release Time Gap	msec
2) 현재 키 Press 와 다음 키 Press Time Gap	msec
3) 2)번 Time Gap 들의 합계	msec

저자 소개

\* 배중기 : 世宗大學 電子工學科 碩士課程  
 \*\* 김병환 : 世宗大學 電子工學科 副教授  
 \*\*\* 이상규 : (주) 디오텍

사용자 타이핑 패턴의 시간변수는 동일한 사용자 패턴을 10번 반복하여 입력받아 수집하였으며 표 2에 나타내었다.

표 2 수집된 사용자 패턴

특성 인자	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
press / release Time Gap	94	94	62	62	62	94	62	94	63	78
	78	63	63	62	62	78	93	62	78	62
	78	78	78	93	78	63	63	62	78	78
	63	47	63	78	62	62	47	47	78	63
	141	157	156	141	141	141	156	140	140	157
press / press Time Gap	94	94	63	110	94	94	109	110	79	94
	172	141	156	125	141	125	125	141	141	156
	157	172	172	172	203	187	187	172	156	156
	187	203	203	156	156	157	156	172	141	156
	234	281	203	203	187	218	188	203	203	172
합계	110	125	125	109	110	110	125	109	140	125

3. 역전파 신경망

인식 모델 개발을 위해서 역전파 신경망 (Backpropagation Neural Network-BPNN) [3]을 이용하였다. 은닉층은 1개이며 은닉 뉴런 수는 5이다. 학습 규칙으로는 일반화된 델타 규칙을 적용하였다.

$$W_{i,j,k}(m+1) = W_{i,j,k}(m) + \eta \Delta W_{i,j,k}(m) \quad (1)$$

여기서  $W_{i,j,k}$  는 (k-1) 층의 j 번째 뉴런과 k층의 i 번째 뉴런 사이의 연결강도 이고  $\Delta W_{i,j,k}$  는 학습편차를 줄이기 위해 인가되는 연결강도의 변이이며 다음과 같이 정의 된다.

$$\Delta W_{i,j,k} = \frac{\partial E}{\partial W_{i,j,k}} \quad (2)$$

4. 비특이 패턴 생성

신경망 인식 모델을 개발하기 위한 첫 번째 단계는 학습데이터와 테스트 데이터를 구성하는 것이다. 신경망 인식 모델의 인식 성능을 향상시키기 위해 비특이 패턴의 생성에 대한 연구가 요구된다. 수집된 10번의 사용자 패턴 중 #1,#3,#5#7,#9 번을 학습을 위해서, 나머지 5번을 테스트를 위해서 사용한다.

표 3 학습 패턴의 평균과 표준편차

특성 인자	평균	표준편차
press / release Time Gap	74.8	17.52712
	65.6	6.94982
	81	6.708204
	62.6	10.96814
	147.2	8.497058
Press / Press Time Gap	91	17.11724
	147	17.76232
	175.2	16.8434
	181	23.73815
	221.6	37.32024
합계	115.8	8.408329
합계	840.6	

비특이 패턴 생성에는 표3의 학습 패턴의 평균과 표준편차가 이용되며 (3)에 의하여 생성하였다.

$$\text{비특이 패턴} = \text{평균} \pm (\alpha * \text{표준편차}) \quad (3)$$

생성된 비특이 패턴의  $\alpha$  에 따른 영향 분석하기 위해서 각 5, 10, 15, 20 개의 비특이 패턴을 생성하여 학습패턴에 추가하여 학습데이터를 구성한다. 검출 성능을 측정하여 인식 성능을 표4에 나타내었으며 RMSE(Root Mean Squared Error)를 표5에 표시하였다.

표 4 비특이 패턴 수와  $\alpha$  에 따른 테스트 패턴 인식 성능

$\alpha$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
비특이 패턴 수	20	0/5	0/5	0/5	4/5	5/5	5/5	5/5	5/5	5/5
	15	0/5	0/5	2/5	3/5	5/5	5/5	5/5	5/5	5/5
	10	0/5	0/5	1/5	3/5	4/5	4/5	5/5	5/5	5/5
	5	4/5	4/5	5/5	5/5	5/5	5/5	5/5	5/5	5/5

표 5 비특이 패턴 수와  $\alpha$  에 따른 테스트 패턴 RMSE

$\alpha$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
비특이 패턴 수	20	0.993	0.997	0.953	0.361	0.279	0.218	0.191	0.179	0.172
	15	0.998	0.974	0.587	0.440	0.321	0.261	0.229	0.209	0.196
	10	0.998	0.994	0.734	0.551	0.409	0.311	0.257	0.227	0.209
	5	0.509	0.434	0.228	0.089	0.056	0.049	0.049	0.063	0.061

진제적으로  $\alpha$ 가 증가할수록 RMSE 값이 감소하는 경향성을 보였으며  $\alpha = 2.5$  이상에서 뛰어난 테스트 패턴 인식 성능을 보이기 시작했다.

5. 학습데이터 Type에 따른 모델 성능 비교

신경망 입력데이터 구성을 위해서 타이핑 패턴 추출 창에서 수집되는 시간 변수는 는 각 키의 press/release Time Gap 과 현재 키의 Press 와 다음 키의 Press Time Gap 의 조합이다. 수집된 3 종류의 사용자 타이핑 패턴의 시간 변수를 표 6와 같이 5가지 타입으로 구분하여 신경망 학습데이터를 구성한다.

표 6 학습 데이터 Type

Type	Type 1	Type 2	Type 3	Type 4	Type 5
press/release time gap	사용	X	X	사용	사용
press/press time gap	X	사용	사용	사용	사용
press/press time gap 의 총 합	X	X	사용	X	사용

5 가지 타입의 학습데이터에 타입에 맞는 20~100개의 비특이 패턴을 생성하여 학습데이터에 첨가하여 총 13번의 신경망 학습을 수행하였다. 학습데이터 구성 기법을 테스트하기 위해서 다른 환경(학습에 사용된 사용자 패턴이 수집된 기기

와 다른 기기에서 타이핑 패턴을 수집)에서 사용자의 타이핑 패턴을 수집한다. 각각 2가지의 다른 환경에서 10번의 타이핑 패턴을 수집한다. 입력 데이터 구성에 따른 2종 인식 오류를 확인하기 위해서 15개의 비특이 패턴을 생성하여 테스트 패턴에 추가하였다. 비특이 패턴은 3장에서 결과를 바탕으로 표준편차 2.5 이상에서 생성하였다. 5 번의 테스트 패턴, 다른 환경에서 수집된 20번의 사용자 패턴과 15개의 비특이 패턴을 모두 사용하여 총 40번의 테스트 패턴을 구성하여 테스트를 진행한다. 비특이 패턴 수와 신경망 입력 데이터 구성에 따른 1종 인식 오류와 2종 인식 오류는 표7에 있으며 RMSE의 변화를 그림 1에 나타내었다.

표 7 비특이 패턴 수와 Type 에 따른 error 수 측정

비특이 패턴수	Type 1		Type 2		Type 3		Type 4		Type 5	
	1종 error	2종 error	1종 error	2종 error	1종 error	2종 error	1종 error	2종 error	1종 error	2종 error
100	4	0	0	0	0	0	1	1	0	0
90	0	0	0	0	0	0	0	1	0	0
80	0	0	0	0	0	0	1	1	0	0
70	4	0	0	0	0	0	1	2	0	0
60	0	0	0	0	0	0	1	2	0	0
55	0	0	0	1	0	0	1	2	0	0
50	0	0	0	1	0	0	1	2	0	0
45	0	0	0	0	0	0	1	2	1	2
40	0	0	0	0	1	1	1	2	1	2
35	0	0	0	0	1	1	1	2	1	2
30	1	0	0	0	0	0	1	1	1	2
25	0	3	0	2	0	0	0	2	1	2
20	0	2	0	2	1	2	0	5	1	4

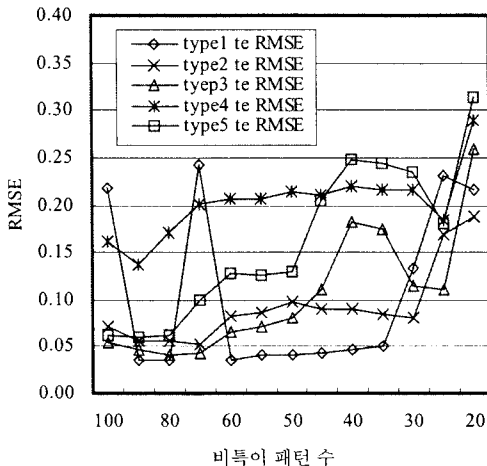


그림 1 비특이 패턴 수와 Type에 따른 RMSE 변화

타이핑 패턴 신경망 인식 모델을 개발한 결과 모든 Type에서 비특이 패턴 수가 많아질수록 RMSE 값이 낮아지는 경향성을 보였으며 Type 1과 Type 4를 제외한 Type에서는 60 이상의 비특이 패턴 수에서 인식률 100% 뛰어난 인식 성

능을 보였으며 전체적인 인식 성능과 안정성 측면에서 봤을 때 Type 3가 가장 좋은 인식 성능을 보였다.

## 6. 결론

본 연구에서는 사용자 인식을 위한 타이핑 패턴 인식 모델을 개발하였다. 2.5 이상의 표준편차에서 비특이 패턴을 생성한 결과 사용자 패턴 인식에 뛰어난 성능을 보였다. 수집된 사용자 타이핑 패턴의 시간 변수의 신경망 입력 구성에 대한 연구 결과 사용자 인식을 위하여 사용자의 현재 키 Press 와 다음 키 Press 간의 Time Gap 과 그 합계로 학습데이터를 구성할 때 가장 좋은 인식 성능을 보였다. 본 인식 기법은 사용자 인증을 위한 타이핑이 필요한 모든 경제 산업적 시스템에서 공통적으로 적용할 수 있으며 추가적인 하드웨어가 요구 되지 않으므로 효과적이며 경제적인 사용자 인식 방식이 될 수 있을 것이다.

## 감사의 글

본 연구는 중소기업청의 기술혁신개발사업으로 지원되었습니다.

## 참 고 문 헌

- [1] M. E. Brown and S. J. Rogers, Method and apparatus for verification of computer user's identification based on keystroke characteristics, US Patent 5,557,686.
- [2] S. Z. Cho, B. H. Han, Apparatus for authentication an individual based on a typing pattern by using a neural network system, US Patent 6,151, 593.
- [3] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, Cambridge, M.I.T. Press, 1986.