

휴먼 타이핑 패턴 스케일링의 신경망 인식성능에의 영향

권희주, 배중기, 김병환
 세종대학교, 전자공학과

Influence of Human Typing Pattern Scaling on Neural Network Recognition Performance

Hee Ju Kwon, Jung Ki Bae, Byungwhan Kim
 Sejong University, Electronic Engineering

Abstract - 개인 사용자의 정보보호를 위한 키보드 타이핑 패턴 인식을 개발한 바 있었다. 키보드 타이핑 패턴의 스케일링 방식에 따라 신경망 인식기의 성능이 차이가 있을 것이라 기대되어 본 연구에서 이를 수행하였다. 총 3 종류의 방식을 이용하여 스케일링을 하였으며, 그 영향을 인식기의 예측에러, 제 1종과 2정의 인식에러측면에서 분석하고 평가하여 최적의 스케일링 방식을 결정하였다.

1. 서 론

컴퓨터 등의 IT 기기에서 사용자 인식을 위해서 ID 와 PASS 등의 개인 정보가 사용되고 있다. IT가 발달하고 개인정보의 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. 최근, 단일 출력을 가지는 신경망을 이용한 타이핑 패턴인식기가 개발된 바 있으며, 그 성능을 타이핑 패턴 유형의 함수로 평가한 바 있다 [3].

본 연구에서는 타이핑 패턴의 스케일링 방식에 따른 신경망 인식성능의 변화를 살펴본다. 이를 위해 타이핑 패턴을 3 종류의 스케일링 방법으로 변환시켰다. 신경망 인식기의 인식 오류는 접근 권한이 있는 사용자를 침입자로 판단하는 1종 인식 오류와 접근 권한이 없는 사용자를 접근 권한이 있는 사용자로 판단하는 2종 인식 오류로 구분하여 인식 성능을 평가하였다.

2. 본 론

2.1 타이핑 패턴의 수집

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

표 1. 타이핑 패턴의 구성요소

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

사용자 타이핑 패턴을 10회 수집하였으며, 이를 다시 각 5개씩 나누어 학습과 테스트에 이용하였다. 한편, 비특이 패턴은 이미 보고된 방식을 이용하여 앞서의 특이패턴으로부터 각 5개씩 발생하였다. 결국, 학습과 테스트 패턴은 각각 10개로 구성되었으며 표 2와 3에 나타내었다. 표 2와 3에서와 같이 각 입력 패턴은 총 11개의 타이핑 특성변수로 구성되었다.

표 2 학습 데이터

특성인자	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
press / release Time Gap	94	94	62	62	62	119.04	129.17	140.14	39.01	127.53
	78	63	63	62	62	33.457	80.009	41.217	20.484	41.422
	78	78	78	93	78	95.744	120.06	128.4	123.4	54.766
	63	47	63	78	62	4.0758	110.76	111.38	28.882	18.946
	141	157	156	141	141	288.78	278.79	2.6371	97.351	260.2
	94	94	63	110	94	149.27	144.19	43.939	168.4	161.59
press / press Time Gap	172	141	156	125	141	166.6761	155.3295	131.4968	104.2232	165.3778
	157	172	172	172	203	165.456	67.7716	85.3127	112.8631	117.3743
	187	203	203	156	156	152.7787	208.551	155.5877	167.9763	134.369
	234	281	203	203	187	181.5762	179.5189	248.205	197.2437	182.895
	110	125	125	109	110	93.5807	131.1882	65.2052	95.0099	94.8878
합계	860	922	859	765	797	760.0677	742.3592	685.8074	677.3162	694.9039

표 3. 테스트 데이터

특성인자	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
press / release Time Gap	94	62	94	63	78	141.33	139.19	23.621	146.3	4.227
	78	93	62	78	62	13.091	3.5252	30.056	106.65	16.951
	63	63	62	78	78	21.476	30.975	20.417	9.4609	134.53
	62	47	47	78	63	97.113	107.64	12.088	102.63	109.8
	141	156	140	140	157	128.68	190.12	44.082	69.11	1.5189
	94	109	110	79	94	35.205	163.88	24.532	128.59	128.09
press / press Time Gap	125	125	141	141	156	68.417	160.815	63.5385	109.9475	111.1998
	187	187	172	156	156	146.8416	137.5757	75.3751	140.6362	119.1155
	157	156	172	141	156	160.6034	128.9802	111.3872	156.6061	89.7724
	218	188	203	203	172	211.5632	228.2532	245.3688	225.1523	165.6366
	110	125	109	140	125	69.8289	71.145	89.4338	66.1753	49.2794
합계	797	781	797	781	765	657.2541	726.7691	585.1034	698.5174	535.0037

2.2 타이핑 패턴의 스케일링

스케일링에 이용된 공식은 다음과 같이 표현된다.

$$X \text{의 스케일링} = [X - \text{Max}(X)] / [\text{Max}(X) - \text{Min}(X)] \quad (1)$$

(1)에서 Max와 Min은 모든 X 값에서 결정되는 최대와 최소값을 지칭한다. 최대와 최소값은 각 행 벡터에서 찾을 수도 있으며, 이에 기초하여 스케일링한 입력패턴을 #1 패턴이라 정의한다. 최대와 최소값은 전체 행 벡터에 대해서도 찾을 수 있으며, 이에 기초하여 스케일링한 입력패턴을 #2 방식이라 정의한다. #3 방식은 #2에서 결정된 최대값은 그대로 사용하지만 최소값은 0으로 설정해서 정의된 유형이다.

보였지만, #1 방식은 제 2종 오류 값이 2였고, #2, #3방식은 0였다. 결국 #2과 #3 방식은 서로 비슷한 인식성능을 보였으며, #1 방식에 비해서는 매우 증진된 인식성능을 보였다.

3. 결 론

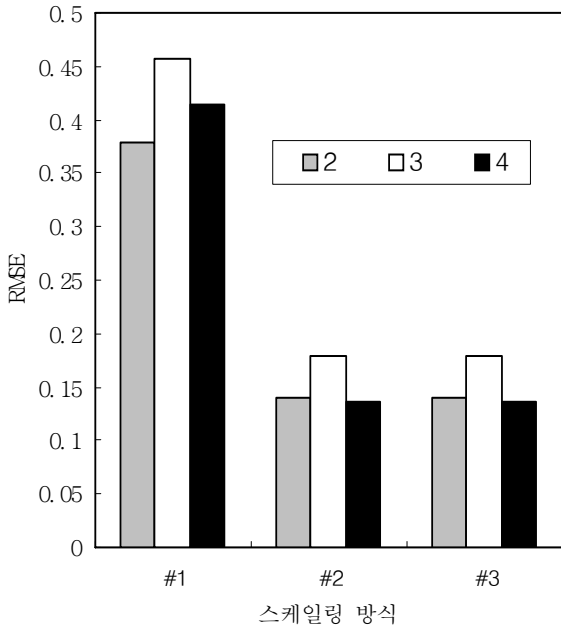
본 연구에서는 타이핑 패턴의 신경망 인식기에의 스케일링 영향을 고찰하였다. 신경망의 인식성능은 스케일링 방식과 뉴런수의 함수로 평가하였다. 연구결과가 보여주듯이, 유형 2와 3 방식이 유형 1의 방식에 비해 더 나은 인식성능을 보였다. 즉, 전체 입력패턴의 최대값과 최소값 (또는 "0")을 이용해 스케일링하는 것이 타이핑 패턴 인식성능에 더 효과적임을 확인할 수 있었다.

감 사 의 글

이 논문은 중소기업기술혁신개발사업(전략과제)의 지원으로 수행되었습니다.

[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] 배중기, 김병환, 이상규, "신경망을 이용한 휴먼 타이핑 패턴 인식," 06정보 및 제어학술대회논문집, 449-451, 2006.
- [4] D. E. Rummelhart and J. L. McClelland, *Parallel Distributed Processing*, Cambridge, M.I.T. Press, 1986.



<그림 1> 스케일링 방식에 따른 신경망 인식 시스템 성능

표 4. 스케일링 방식에 따른 인식 어려 평가

뉴런수	스케일링 방식					
	#1		#2		#3	
	1종 오류	2종 오류	1종 오류	2종 오류	1종 오류	2종 오류
2	0	2	0	0	0	0
3	0	2	0	0	0	0
4	0	2	0	0	0	0

2.3 역전파 신경망 인식기

인식 모델 개발을 위해서 역전파 신경망 [4]을 이용하였다. 은닉층은 1개였다. 학습 규칙으로는 일반화된 델타 규칙을 적용하였으며, 뉴런간 웨이트 조정은 (1)과 (2)에 의해 이루어졌다.

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

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

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

인식 성능은 은닉층 뉴런수의 함수로 평가하였다. 각 각의 스케일링 방식에 대하여 Hidden neuron 수 2, 3, 4에 대하여 3번에 걸쳐 신경망 인식 시스템을 개발하였다. 인식성능은 우선 테스트 데이터에 대한 Root Mean Square Error (RMSE)로 평가하였으며, 그 결과가 그림 1에 각 스케일링 방식과 은닉층 뉴런의 함수로 도시되어 있다. 그림 1에서와 같이, #3 방식의 뉴런수 4에서 0.136808의 가장 낮은 RMSE 값을 보였다. #2와 #3의 스케일링 방식은 0.001 미만의 아주 작은 RMSE 차이를 보이거나 두 방식은 #1과는 0.2 이상의 큰 RMSE 차이를 보이고 있다. 그 원인을 찾기 위해 각 방식으로 변환된 학습과 테스트 데이터를 비교해 보았다. 그 결과 #2와 #3 방식의 경우, 두 데이터가 매우 유사했던 반면에, #1 방식의 경우 큰 차이를 보였다. 결국 이러한 데이터간의 유사성의 차이에 의해 앞서의 RMSE의 차이가 발생한 것으로 해석된다.

표 4는 뉴런 수와 스케일링 방식에 따른 신경망 인식 시스템의 제 1종 오류와 제 2종 오류 값이다. 세 방식 모두 0의 제 1종 오류 값을