

# 인간행동 인식의 신경망적 접근

류중원\*, 조성배

연세대학교 컴퓨터과학과

## A Neural Network Approach to Recognition of Human Behaviors

Jungwon Ryu and Sung-Bae Cho

Dept. of Computer Science, Yonsei University

### 요 약

인공 신경망은 체계적인 알고리즘으로 풀기 어려운 문제들을 해결하는데 사용되어오고 있다. 이는 인간의 뇌세포가 외부자극에 대해 반응하는 과정을 컴퓨터 시스템 상에서 구현한 것으로써 인간과 컴퓨터의 상호작용을 연구하는데 흥미로운 접근방식이다. 본 논문에서는 신경망의 접근방법을 이용하여 인간행위 인식시스템을 구현하였다. 신경망을 이용해 구현된 컴퓨터 인식 시스템이 인간의 두 가지 정서 하에서 일어난 세가지 서로 다른 행동을 보고 행위자의 성별이나 감정상태를 얼마나 인식해낼 수 있는지 실험해 보았다. 특히, 성별 인식 실험에서는 신호탐지 이론에서 사용하는 민감도(discriminability)를 이용해 사람에 대한 이 시스템의 효율도를 계산하였다.

### 1. 서론

어떤 아바타의 움직임을 보고 '사람같다(human-like)'라고 말한다면 그는 아바타의 무엇을 보고 그렇게 느끼는 걸까? 예전에는 컴퓨터가 단순한 수학적 계산만 반복하는게 고작이었지만, 요즘은 '좀더 빨리' 계산하는 시스템 보다는 '좀더 사람처럼' 동작하는 시스템을 선호하는 추세이다. 이렇게 사람처럼 행동하는 시스템을 구현하기 위해서는 인간의 행동 인식에 대한 제반 연구가 필요하다.

이와 같은 맥락에서 본 논문에서는 신경망 접근 방식을 통해 인간의 행동을 인식하는 시스템을 구현하고, 인간의 인식능력과 비교하여 그 가능성을 보이고자 한다.

### 2. 관련연구

생물학적 운동의 인식(the perception of biological motion)에 대한 연구는 최근 30년 동안 심리학 분야에서 흥미로운 과제가 되어왔다. 1943년 Wolff의 연구가 있었고[1], 생물학적 운동 분야의 개척자라고 할 수 있는 Johansson은 Wolff의 연구에서 식별의 단서로 사용된 물체의 크기나 모양 등의 한계를 어느

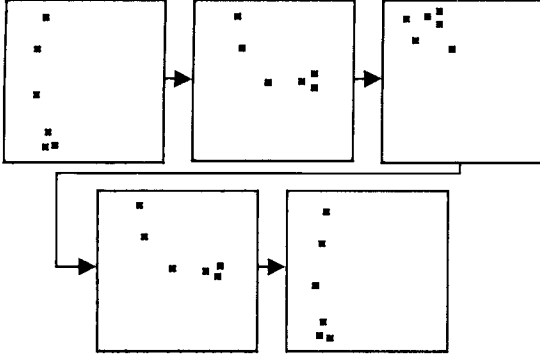
정도 극복할 수 있는 대안을 제시했다[2]. 그는 인간 운동의 패턴을 연구함에 있어서 대상물의 크기나 모양보다는 움직이는 점이나 선에 관심을 두었는데, 특히 사람의 10개 관절에 반짝이는 유리구슬을 부착하고 어두운 곳에서 이것의 움직임을 관찰함으로써 사람 식별능력을 테스트하였다.

이 실험에서 Johansson은 정지해있는 점광 화면(point light display)으로부터는 어떠한 특이점도 찾아낼 수 없었지만, 이 점광들이 움직이기 시작하면 어느 정도 사람들이 행위자를 인식할 수 있다는 사실을 밝혀내었다.

그 이후로 생물학적 운동의 인식에 관한 많은 연구들이 움직이는 점광 화면을 사용하여 진행되었는데, 1977년 Cutting과 Kozlowski의 연구에서는 사람의 주요 관절에 부착된 점광의 움직임을 통해서 보행자의 성별을 60% 70% 정도 구별해 낼 수 있었다[3]. 게다가 그들은 상체의 움직임을 관찰하는 것이 하체 보다 성별 구분에 있어서 더욱 효과적이라는 것도 알아내었다.

이와 같은 맥락에서 본 논문에서는 행위자의 운동을 점광화면으로 나타내어 각 점이 움직이는 속도의 변화를 인식의 특징점으로 사용하였다.

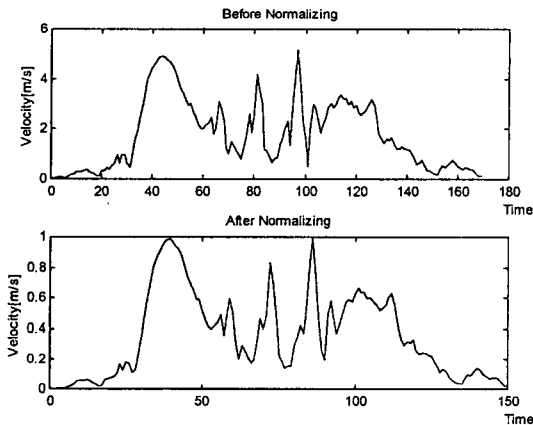
[그림 1] 점광화면의 예: 행위자가 물을 마시는 동작



### 3. 연구 방법

#### 3.1 전처리

행위자의 동작이 일어나는 동안 Optotrak (Northern Digital Co.)이라는 3D 좌표 분석시스템을 통해 머리, 어깨, 팔꿈치, 손목, 손가락1, 손가락2의 여섯 지점들의 좌표 데이터를 수집하였다. 이로부터 각 지점의 속도값을 계산하였고, 사람마다 동작의 길이가 다르므로 일정한 수의 입력 노드를 가져야 하는 신경망에 입력하기 위해 다시 일정한 길이로 정규화하였다. [그림 2]는 원래의 패턴을 150으로 정규화한 예를 보여주고 있다.



[그림 2] 패턴 정규화

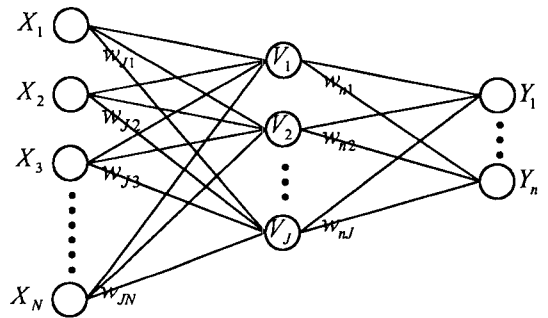
#### 3.2 신경망 인식기

인공 신경망 모델은 음성이나 영상 인식과 같은 분야에서 오랫동안 연구, 개발되어왔다. 이는 인간의 뇌세포인 뉴런이 외부 자극에 대해서 반응하는 과정을 흉내낸 것으로, 뉴런의 출력값은 자신에게 연결된 뉴런으로부터 오는 자극들에 일정한 가중치를 적용하여 그 합을 구하는 방식으로 결정된다. 신경망적 접근에 따르면 한가지 정보가 한 기억 단위에 저

장되는 것이 아니라 정보의 조각들이 메시지들을 교환함으로써 의사소통하는 매우 단순한 계산 요소들에 의해서 분산 표상된다.

이러한 신경망 모델들 중 본 논문에서 인간 행동의 인식을 위해서 사용한 것은 역전파 (backpropagation) 알고리즘으로 학습된 다층 신경망이다. 이는 현재 주어진 연결강도로 생성되는 오차 값을 구하여 이를 감소시키는 방향으로 연결강도의 값을 조정하는 델타 학습 알고리즘을 사용한 것이다

역전파 알고리즘은 현재 신경망 분야에서 가장 널리 사용되고 있다. 맨 처음 1974년 Paul Werbos가 다층 신경망에 대한 아이디어를 내놓은 이후, David Rumelhart, Geoffrey Hinton, Ronald Williams, David Parker, Yann Le Cun 등에 의해서 1980년대 중반에 각각 독립적으로 연구가 진행되어갔다. 이 논문에서는 [그림 3]와 같은 하나의 은닉층을 가진 3층 역전파 신경망을 사용하였다.



[그림 3] 3층 신경망

일련의 입력패턴 ( $X_1, X_2, \dots, X_N$ ) 이 신경망으로 들어오면 은닉층 노드  $V_j$  와 출력층 노드  $Y_i$  의 출력값은 각각 다음과 같이 계산된다.

$$V_j = F \left[ \sum_{k=1}^N w_{jk} X_k \right] \quad j = 1, 2, 3, \dots, J$$

$$Y_i = F \left[ \sum_{m=1}^J w_{im} X_m \right] \quad i = 1, 2, 3, \dots, n$$

그리고 아래와 같은 로지스틱 시그모이드 활성화 전이 함수를 사용하여 신경망 노드들의 최종 출력값을 결정한다. [4]

$$F[.] = \frac{1}{1 + e^{-x}}$$

### 4. 실험 결과

4.1 실험 환경 및 방법

행위자의 성별 및 감성인식 실험을 위해 사용된 정서 상태와 동작은 '보통상태(neutral),' '화난상태(angry)'의 '문 두드리기(knocking),' '손 흔들기(waving),' '물건 들어올리기(lifting)' 행동이다. 이러한 동작은 짚고 쉬울 뿐만 아니라, Cutting과 Kozlowski의 연구에서처럼 상체의 움직임이 하체보다 패턴 인식이 쉽다는 점 때문에 실험의 대상으로 채택되었다. 동작이 일어나는 동안 사람의 머리, 어깨, 팔꿈치, 손목, 손가락1, 손가락2의 여섯 지점의 속도변화를 추출하였다. 데이터는 남녀 각각 13명으로부터 얻어졌으며 각 동작은 10회씩 반복하였다. 신경망은 C언어로 작성되었으며 학습률은 0.1, 모멘텀 값은 0.7, 입력 패턴의 정규화율(normalizing rate)은 150, 은닉층 노드수는 50을 사용했다. 여기에서 모든 매개변수값들은 여러 차례의 실험을 통하여 결정되었다.

4.2 인간 성별 인식

행위자의 속도 데이터를 위에 언급한 매개변수값들과 함께 신경망에 입력하여 표 1과 같은 실험 결과를 얻었다.

표 1. 인간 성별 인식 실험

감성	행동	계산횟수	인식률[%]	사람의 실험결과[%]
Angry	Knocking	760	75.5	49.2
	Waving	42	77.4	55.1
	Lifting	67	83.0	51.3
Neutral	Knocking	34	83.0	50.9
	Waving	102	92.5	53.6
	Lifting	43	86.8	47.4
전체			83.0	51.3

또한 같은 데이터 집합을 사용해서 사람에게 인식 실험을 수행한 결과도 기록하였다[5]. 수치가 말해주듯이 사람이 행위자의 성별을 구별하는 데 사용할만한 중요한 단서가 별달리 없는 듯 싶다. 하지만 신경망으로 분류했을 때는 전체 평균 약 83.0%로 사람보다 약 31.7% 정도 나은 인식률을 보였다. 신경망 실험 결과 인식률이 가장 높은 것은 '화난상태 손 흔들기'였으며 가장 낮게 기록된 것은 '보통상태 문 두드리기'였다.

민감도(discriminability)와 효율도

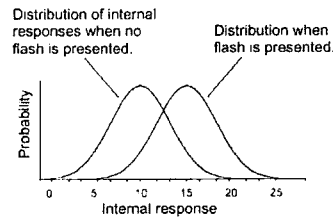
신호탐지 이론에서는 쉽게 구별되지 않는 두 가지 비연속 신호에 대한 반응 기준을 설정하는데 민감도(discriminability 혹은 sensitivity)라는 개념을 사용한다. 신호탐지 이론에서는 주어진 신호에 대한 사람의 반응을 표 2와 같이 네가지로 분류한다.

표 2. 신호탐지 이론의 네가지 가능한 반응 결과

	실상태
--	-----

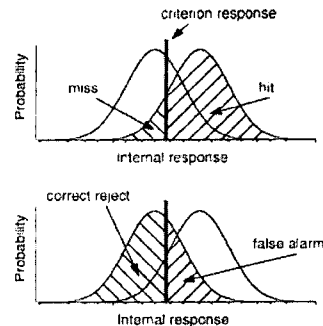
		신호(S)	잡음(N)
반응	예	적중(H)	오경보(FA)
	아니오	탈루(M)	정기각(CR)

입력 신호에 대한 피험자의 내부 반응 확률 분포를 그리면 [그림 4]와 같이 되는데, 민감도란 이 두 분포 곡선의 꼭지점간 거리를 말한다.



[그림 4] 사람의 내부 반응 확률 분포

입력 신호에 대하여 사람이 '신호'인지 '잡음'인지를 결정하게 되면 그에 따라 [그림 5]와 같이 적중(hit)과 오경보(false alarm) 확률이 정해지게 된다. 민감도가 커질수록 오경보에 대한 적중 확률이 증가하게 된다.



[그림 5] 사람의 반응에 따른 적중률과 오경보를

식별도  $d'$ 과 사람에게 대한 효율도  $E$ 는 다음과 같이 계산된다.

$$d' = \frac{|\mu_0 - \mu_1|}{\sigma}, \quad E = \frac{(d'_{NN})^2}{(d'_{human})^2}$$

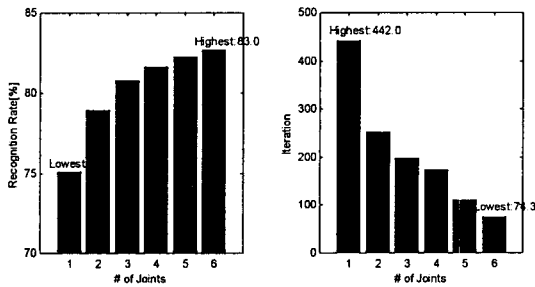
여기에서  $\mu_0, \mu_1$ 은 두 분포 곡선의 중심,  $\sigma$ 는 표준편차,  $d'_{NN}$ 은 신경망의 민감도,  $d'_{human}$ 은 사람의 민감도이다. 이에 따라 신경망 인식기와 사람의 민감도와 사람에게 대한 신경망 인식기의 효율도를 계산하면 표 3과 같은데, 신경망이 사람보다 훨씬 나은 성능을 보임을 알 수 있다.

표 3. 신경망 인식기의 민감도와 효율도

감성	행동	$d'_{NN}$	$d'_{human}$	$E$
Angry	Knocking	1.03	0.08	165.9
	Waving	0.83	0.04	426.5

	Lifting	1.28	0.11	135.6
Neutral	Knocking	1.01	0.20	25.4
	Waving	2.90	0.28	107.0
	Lifting	1.21	0.07	297.4

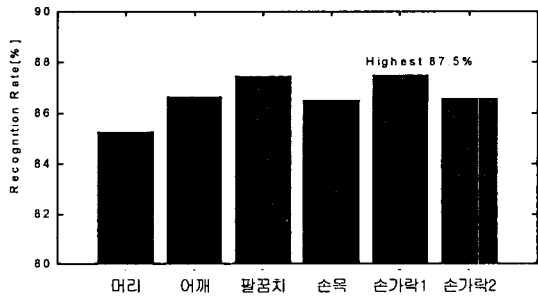
사용된 지점 개수별 인식률  
 학습을 위해 사용된 지점의 개수를 달리 하면서 실험해 보았는데 각각의 경우에 대한 성별 인식률을 살펴보면 [그림 6]과 같다.



[그림 6] 사용된 지점 개수별 인식 실험 결과

여섯개의 사용가능한 지점들 중에서 하나만 학습데이터로 사용하였을 때가 가장 학습 속도도 느리고 인식률도 낮았다. 하지만 많은 지점을 고려할수록 성능이 향상되어 여섯 지점을 모두 고려하는 경우 83.0% 정도까지 인식 가능하였다.

사용된 입력점별 인식률  
 여섯개 입력점들 중에서 어느 것이 성별 인식에 가장 효과적인지를 살펴보았다. [그림 7]에서 나타난



바와 같이 손가락1이 사용되었을 때 가장 높은 인식률을 보였다.

[그림 7] 사용된 입력점별 인식률

#### 4.3 인간 감성 인식

인공 신경망을 이용한 인간 감성의 인식 실험의 결과는 표 3와 같다.

표 3. 인간 감성 인식 실험

성별	행동	계산횟수	인식률[%]
남자	Knocking	43.0	86.8
	Waving	61.3	86.8
	Lifting	30.0	86.5
여자	Knocking	49.0	82.4
	Waving	46.3	92.5
	Lifting	49.7	84.3
전체			86.6

전체적으로 평균 86.6%의 인식률을 보였는데 이는 성별 인식 실험 보다 약 3.6%가량 높은 것이다. 표 1, 3를 비교해보면 '손 흔들기' 동작에 대한 인식률이 높게 나온 것을 볼 수 있다.

#### 5. 결론

본 논문에서 생물학적 운동의 인식 문제를 신경망 방법으로 해결해보고자 하였다. 실험 결과, 신경망을 이용한 인간 행동의 성별 인식 실험에서는 83.0% 정도의 인식률을 보였고, 감성 인식 실험에서는 약 86.6%의 인식률을 보였다. 이는 인간의 생물학적 움직임 인식 문제에 있어서 연결주의적 접근이 어느 정도 적절한 계산적 패러다임을 제공한다는 것을 보여주는 것이다.

앞으로 속도 데이터가 아닌 2D 혹은 3D 좌표를 이용한 실험도 계획하고 있으며 신경망 성능향상을 위해서 여러 알고리즘을 결합하는 방법도 사용할 것이다.

#### 참고문헌

- [1] W. Wolff, *The Expression of Personality: Experimental Depth Psychology*, New York: Harper & Brothers, 1943.
- [2] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, pp. 201-211, 1973.
- [3] D. E. Cutting and L. T. Kozlowski, "Recognising friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, pp. 353-356, 1977.
- [4] R. P. Lippman, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, pp. 4-22, April 1987.
- [5] V. Lestou and F. E. Pollick, "Gender recognition of arm movements from dynamic point light displays. A pre-test, training, post-test design," 2000. (private communication)