

인공 신경망 분류기를 이용한 인간 행동의 성별 인식

류중원^U 조성배
연세대학교 컴퓨터학과 컴퓨터학과
rjungwon@candy.yonsei.ac.kr, sbcho@csai.yonsei.ac.kr

Gender Recognition of Human Behavior with Neural Network Classifier

Jungwon Ryu^U Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

인간과 기계가 효과적인 상호작용을 하기 위해서는 컴퓨터 시스템이 인간의 행동을 인식할 수 있어야 한다. 본 연구에서는 인공 신경망을 사용하여 컴퓨터 시스템이 인간의 움직임에 관찰한 후 행위자의 성별을 인식하도록 하는 시스템을 구현하였다. 두 가지 감정상태(보통 상태, 화난 상태) 하에서 일어난 인간의 세 가지 동작(문 두드리기, 손 흔들기, 물건 들어올리기)을 대상으로 하여 인간 동작 데이터를 통해 만들어진 학습 데이터를 통해 98.0%의 인식률을 보일 때까지 학습시키고 나서, 이전에 사용하지 않았던 새로운 데이터에 대해 얼마나 성별을 잘 구별해 내는지 실험하였다. 동작이 일어나는 동안 행위자의 몸 여섯 군데에서 속도 데이터를 얻어내서 신경망의 입력값으로 사용하였다. 그 결과 최저 62.3% 이상 최고 94.3%까지 인간 성별을 구분해 낼 수 있었고 이는 같은 데이터에 대하여 사람을 통해 실험한 것 보다 훨씬 나은 것이다.

1. 서론

컴퓨터가 처음으로 발명되었을 때, 컴퓨터 시스템은 단순히 수학적 연산을 수행하는데 주로 사용되었다. 그러나 과학의 눈부신 발전에 의하여 현재는 그 활용 영역을 정보 처리나 데이터 통신과 같은 복잡하고 다양한 분야로 넓혀 나가고 있다. 또한 원하는 결과를 얻기 위해 인간이 주어진 입력값으로 단순 연산만 반복하던 수동적인 정보처리의 시대에서 사용자에게 필요한 정보를 스스로 제공하는 능동적인 정보처리의 시대로 접어들고 있다.

이러한 시대적 흐름에 발맞추기 위해서는, 인간과 컴퓨터 사이에 자연스러운 상호작용이 가능해야 하는데, 이를 위한 가장 기본적인 접근 중의 하나는 컴퓨터 시스템이 인간의 행동을 인식할 수 있도록 하는 것이다.

2. 배경

2.1 심리학 분야에서의 성별 인식에 관한 연구
생물체의 움직임 인식에 관한 연구는 심리학 분야에서 흥미있는 연구로 최근 30년간 활발히 연구되어왔다. Wolff가 최초로 걸음걸이로 친구를 인식하는 연구를 수행한 이후로[1], 생물체의 움직임 인식분야의 개척자라고 할 수 있는 Johansson은 물체의 모양이나 크기가 아닌 점이나 선의 움직임에

초점을 맞추었다[3]. 실험에서 그는 행위자의 특정 신체 부위에 유리구슬을 달고 주변을 어둡게 한 후 유리구슬의 움직임을 점광화면(point light display)으로 기록하여 이를 보고 보행자를 인식하도록 하였다. 그 후 이 분야의 많은 연구들은 이 점광화면을 기초로 진행되어 왔는데 Cutting과 Kozlowski의 연구에서는 이 방법을 이용하여 60~70% 정도로 행위자의 성별을 구별해 낼 수 있었으며 상체의 움직임이 하체의 움직임보다 성별 인식에 효과적이라는 사실을 발견하였다.[2]

이러한 견지에서, 우리는 무엇이 성별 구별을 가능케 하는지를 알아보기 위해, '문 두드리기', '손 흔들기', '물건 들어올리기'의 세 가지 동작이 각각 '보통상태'와 '화난 상태'의 두 가지 정서 하에서 얼마만큼 인식될 수 있는지를 알아보았다.

2.2 인공 신경망과 역전파 알고리즘

인공 신경망은 음성이나 화상 인식 분야에서 널리 사용되어왔다. 이 신경망들을 분류해 보면 크게 교사 학습 방식과 비교사 학습 방식으로 나눌 수 있는데[3], 본 연구에서 수행하려는 성별 인식과 같은 패턴인식의 관점에서 보았을 때, 신경망의 올바른 출력값을 이미 알고 있는 상황에서 학습이 이루어지게 되므로 교사 학습 방식이 더 적합하다.

인공신경망이 적극적으로 사용되기 시작한 것은 기존의 입력층과 출력층으로 구성된 다층퍼셉트론의 한계점을 극복하면서부터이다. 과거 입력층과 출력층만으로 구성된 신경망구조는 선형분리(Linear Separable)가 가능한 문제에만 적용이 가능했고, 배타적인 논리합(Exclusive OR)문제는 해결하지 못했다. 그러나, 입력층(Input Layer)과 출력층(Output Layer) 사이에 은닉층을 두는 다층퍼셉트론의 역전파 알고리즘이 제시되면서 배타적인 논리합의 문제까지 해결하게 되었다. 역전파 알고리즘은 주어진 입력값으로 계산된 출력값과 실제값의 오차의 방향을 줄이는 방향으로 학습이 이루어지게 되는데 Δt 시간 후의 가중치값은 아래와 같이 계산될 수 있다.

$$w_i(t + \Delta t) = w_i(t) + \eta \delta x_i$$

여기서 Δt 는 학습 주기, η 는 학습률 조절을 위한 이득값이며, δ 는 에러값이다. 입력층에 N 개, 은닉층에 J 개, 출력층에 n 개의 노드를 갖는 신경망의 경우 은닉층의 결과값 V 와 출력층의 결과값 Y 는 각각 다음과 같이 쓸 수 있다.

$$V_j = F \left[\sum_{m=1}^N w_{jm} X_m \right] \quad (j=1, 2, \dots, J)$$

$$Y_i = F \left[\sum_{n=1}^J w_{in} X_n \right] \quad (i=1, 2, \dots, n)$$

$$F[.] = \text{sigmoid}(n, b) = \frac{1}{1 + e^{-(n+b)}}$$

에러는 $E[w] = \frac{1}{2} \sum_{i=1}^n [Y_i - Y_{i \text{ target}}]^2$ 로 계산한다.

3. 실험 방법

3.1 속도값 추출

동작이 일어나는 동안 각 지점 또는 관절의 속도값을 신경망의 입력값으로 사용하였다. 특정 동작이 일어나는 동안에 머리, 어깨, 팔꿈치, 손목, 손가락(두 군데)의 여섯 지점에 적외선 발사 장치를 붙여 Optotrak이라는 3D 좌표 분석시스템을 통해 x, y, z 좌표값을 얻어내었다. 샘플링 주기는 60Hz였다. 이렇게 얻어진 3D 좌표값을 통해 속도를 계산하였는데 위의 3D 좌표값들은 일정한 시간 간격으로 얻어진 것이기 때문에 현재 지점과 그 다음 지점의 거리를 구하여 속도값으로 사용하였다.

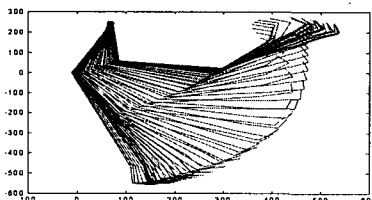


그림 1 행위자가 문을 두드리는 모습

3.3 데이터 정규화

이미 언급한 바와 같이 BP 네트워크의 입력값으로 각 지점들의 속도값을 입력하였다. 그런데 실제로 동작이 일어나는 동안의 시간 길이는 사람마다 혹은 측정할 때마다 일정하지 않으므로 일정한 비율로의 정규화가 필요하다. 또 모든 속도값의 크기가 일정하지 않으므로 신경망의 입력값이 0과 1사이여 있도록 조정해주는 과정이 필요하다.

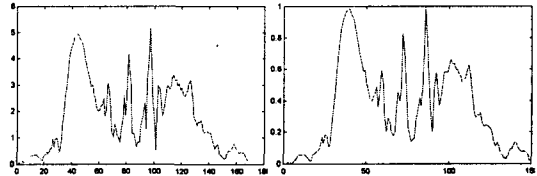


그림 2 정규화 전

그림 3 정규화 후

3.4 데이터 집합 생성

실험 데이터는 남·녀 각각 13명(총 26명)으로부터 얻어졌다. 6가지 서로 다른 동작(두 가지 감정)에 대해서 모두 10번씩 반복하여 그중 8개는 학습 데이터로, 나머지 2개는 테스트 데이터로 사용하였다. 데이터 셋 생성시 남자는 1, 여자는 0, 보통 상태는 [0 1], 화난 상태는 [1 0], 문 두드리기는 [0 0 1], 손 흔들기는 [0 1 0], 물건 들어올리기는 [1 0 0]로 표현하였다. 데이터 셋을 만드는 모든 과정은 MATLAB으로 수행되었다. 아래 그림 3, 4는 위 과정들을 거쳐 신경망 입력값으로 사용되는 문 두드리는 동작에 대한 예이다.

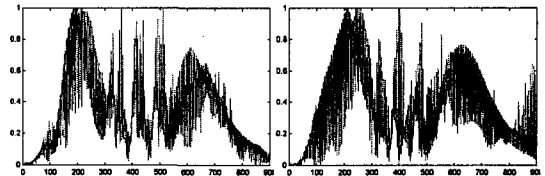


그림 4 Male knocking

그림 5 Female knocking

4. 실험 결과

이 실험에서 사용된 몇가지 파라미터들은 다음과 같다.

lrate : 학습률

Momentum : 모멘트 값

L_2_NODE : 은닉층 노드 수

L_3_NODE : 출력층 노드 수

srate : 샘플링률(sampling rate)

Max_iteration : 한계 반복 계산 횟수(여기서는 800번으로 제한)

실험 1에서는 파라미터별 인식률의 변화를 관찰하였다. 각 실험은 3회씩 실시 되었다. 실험 2에서는 실험 1에서 가장 좋은 결과를 나타낸 값들을 이용하여 각 감정 상태별, 동작 별로 인식률을 측정하였다.

4.1 실험 1

모멘트 값을 0.6에서 0.8까지 변화시켜가면서 인식률과 iteration 값을 측정해 보았다. 모멘트는 신경망이 얼마나 빨리 학습을 하느냐에 영향을 미치는데 표 2의 결과에서처럼 모멘트 값이 0.6일 때 인식률이 가장 높게 나왔으나 iteration 측면에서 보았을 때 0.8 이상이 되면 학습 속도가 갑자기 느려지게 되거나 수렴하지 않는 현상을 보이고 있다.

표 1 모멘트 값에 따른 인식률[%]

모멘트 값	평균인식률	평균 iteration
0.6	78.0	40.7
0.7	75.5	42.7
0.8	76.1	800.0

L₂_NODE 값을 변화시켜가면서 인식률을 측정하였더니 표3과 같은 결과를 보였다. 은닉층 노드 수가 50개 정도가 되었을 때 높은 인식률을 보이고 있으며 iteration 값도 비교적 안정적이었다. 하지만 노드 수를 60개 이상으로 늘리면 더 이상 학습이 이루어지지 않았다.

표 2 L₂_NODE 값에 따른 인식률[%]

L ₂ _NODE	평균인식률	평균 iteration
10	71.7	305.0
20	76.8	55.7
30	75.5	48.0
40	76.1	47.7
50	76.7	47.0
60	.	.
70	.	.

샘플링률 별로 실험한 결과는 표 4와 같다.

표 3 srate에 따른 인식률[%]

srate[Hz]	50	100	150	200	250
평균인식률	79.5	83.9	83.9	80.3	84.0

150Hz 정도로 샘플링 하였을 때 가장 높은 인식률을 보였다.

4.2 실험 2

Lrate 0.1, L₂_NODE 50, Momentum 0.7, srate 150에서 실험을 하였다.

표 4 여섯 조건하에서의 인식률[%]

	knocking	waving	lifting
neutral	79.2	79.2	86.8
angry	75.5	92.5	83.0
total	82.7		

표 5 동작별 인식률[%]

	knocking	waving	lifting
평균인식률	77.35	85.85	84.9

각 감정상태별 인식률은 표 6과 같다.

표 6 정서별 인식률[%]

	neutral	angry
평균인식률	81.7	83.7

그림 6에서 보는 것처럼 6지점의 데이터를 모두 사용할수록 인식률이 높았다.

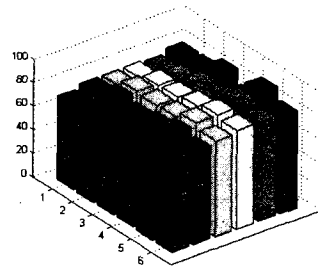


그림 6 사용된 joint 데이터 개수별 인식률

5. 결론 및 향후 연구

이 실험에서 우리는 행위자의 특정 부위의 속도값을 신경망의 입력값으로 사용하여 어느정도 성별을 구별해낼 수 있었다. 여러 변수들이 조건을 만족시킨다면 인식률이 최고 94.3%까지 가능했다. 실험 결과 보통 상태 보다 화난 상태에서 손을 흔드는 동작이 가장 잘 인식되는 결과를 보였다. 이 실험 결과는 사람이 수행했을 때 약 40~50% 정도 밖에 인식하지 못하는 것에 비하면[4] 훨씬 나은편이긴 하지만 이것은 단순히 속도의 변화를 이용하여 기계적으로 분류한 것에 지나지 않으므로 성별 인식에 대한 궁극적인 해답은 아니다. 특정 동작이나 감정이 성별을 구별하는데 어떤 영향을 미치는지에 대한 구체적인 연구가 필요하다.

6. 참고 문헌

- [1] W. Wolf, *The Expression of Personality : Experimental Depth Psychology*, New York : Harper & Brothers, 1943.
- [2] J. E. Cutting, and L. T. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, 9, pp. 353-356, 1977.
- [3] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, pp. 4-22, April 1987.
- [4] V. Lestou, F. E. Pollick, "Gender recognition of arm movements from dynamic point light displays. A pre-test, training, post-test design," 2000. (private communication)