

신경망 앙상블을 이용한 인간 성별 인식

류중원, 조성배
연세대학교 컴퓨터과학과
e-mail : rjungwon@candy.yonsei.ac.kr, sbcho@csai.yonsei.ac.kr

Human Gender Recognition Using Neural Network Ensembles

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

요 약

본 논문에서는 인간 행동의 성별 인식문제를 해결하기 위해 여러 개의 전문가(expert) 신경망의 앙상블로 이루어진 결합 신경망 분류기를 제안한다. 하나는 여러 개의 modular 다층퍼셉트론을 계층적으로 결합한 모델이고, 다른 하나는 modular 다층퍼셉트론들의 출력값을 의사결정트리로 결합하는 모델이다. 데이터 베이스는 남녀 각 13 명의 데이터로 이루어져 있고, 문 두드리기, 손 흔들기, 물건 들어올리기의 세 가지 동작을, 보통 상태 혹은 화난 상태에서 10 회씩 반복 수행하여 저장하였다. 행위자의 움직임은 몸에 부착된 6 개의 적외선 센서를 사용하여 기록 되었으며, 2 차원 혹은 3 차원 속도 및 좌표가 그 특징값으로 사용되었다. 앙상블 분류기의 성능을 비교하기 위하여 단일 다층퍼셉트론, 의사결정트리, 자기구성지도 및 support vector machine 을 사용한 실험 결과를 보였다. 실험 결과, 신경망 앙상블 모델이 다른 전통적인 분류기 및 사람에 비하여 훨씬 우수한 성능을 보였음을 알 수 있었다.

1. 서 론

인간 행동에 대한 인식의 문제는 좀더 지능적인 시스템을 구현하려고 하는 컴퓨터 과학 분야에서도 매우 흥미로운 연구토픽 중 하나이다. 심리학 분야에서도 최근 30년간 이 문제에 대한 다양한 연구들이 이루어져 오고 있으며, Johanssen이 최초로 제안한 점광 화면(point light display)은 [1] 생물학적 움직임을 대상으로 하는 연구에 널리 이용되어 오고 있다.

한편, 1980년대 이후 인공 신경망이 패턴 인식 분야에서 널리 사용되면서 여러가지 용도로 일상 생활에 편리함을 제공하고 있다. 하지만, 여전히 실세계의 문제를 해결하고자 할 때 여러 종류의 학습 알고리즘과, 신경망 중에서 문제에 대한 최적의 분류기 모델을 찾는 것은 쉽지 않다. 따라서, 최근에는 패턴 분류기로서 단일 신경망을 사용하기보다, 다수의 특성화된 신경망들을 결합하여 분류기의 성능을 향상하는 연구가 진행되고 있다[2].

본 논문에서는 점광화면으로 표현된 인간의 여섯가지 행동으로부터 행위자의 성별을 인식하기 위하여 신경망 분류기 앙상블을 사용하였다. 앙상블 분류기의 성능을 객관적으로 검증하기 위하여 기존의 전통적인 인공지능 방법들 중 단일 다층 퍼셉트론, 의사결정트리, 자기구성지도, support vector machine이 사용되었다. 또한 같은 데이터에 대하여 사람을 대상으로 한 실험 결과와의 비교도 제시하고자 한다.

2. 인간 행동의 성별 인식

Wolf는 보행자의 걸음걸이를 통하여 성별을 인식하는 문제를 최초로 다루었는데, 이후 Cutting과 Kozlowski (1977년), Barclay(1978년), Runeson과 Frykholm (1983년), Mather와 Murdoch(1994년) 등에 의하여 관련 연구가 진행되었다. 특히, 생물학적 운동을 연구하는데 있어서 물체의 크기나 모양 보다는 움직임의 선이나 점이 더욱 중요한 정보를 가지고 있음이

알려지면서, 1973년에 점광화면이 최초로 제안되었다. Johanssen은 보행자의 움직임을 관찰하기 위해 어두운 곳에서 주요 부위 10곳에 유리구슬을 부착하였다. 점광들이 움직이지 않는 상태에서는 보행자를 인식할 수 없었지만, 이들이 움직이기 시작하면서부터 어느정도 인식이 가능했음을 발견하였다. Cutting과 Kozlowski는 사람의 상반신 및 하반신 동작을 이러한 점광화면으로 표현하여 행위자의 성별을 구분하는 실험을 수행하였는데, 사람을 대상으로한 실험에서 63%의 인식률을 얻었다[3]. 또 상반신의 움직임이 하반신의 움직임보다 성별인식에 효과적임을 알아 내었다.

하나의 복잡한 문제는 여러 개의 계산 가능한 문제들로 나누어 풀고, 이를 결합함으로써 해결될 수 있다 [4]. Osherson의 정의에 따르면, 서로 의사소통 없는 구별되는 입력 공간상에서 작동하는 둘 이상의 하위 시스템에 의하여 계산이 이루어 진다면, 이 신경망은 modular하다고 말한다[5]. 인공 신경망을 사용한 패턴 분류 문제에서도, 전체 학습 작업을 데이터의 locality를 최대한 반영하도록 위와 같은 다수의 modular 신경망(expert)에 분산하여 수행하도록 함으로써 성능을 향상시킬 수 있다. Gutta 등은 FERET 얼굴 이미지 데이터베이스를 사용하여 사람의 성별, 인종 및 얼굴 포즈 인식을 위하여 RBF와 의사결정트리의 결합 모델을 제시하였다. RBF를 학습하기 위하여 원래이미지에 5° 각도로 변형을 가하거나, 가우시안 잡음을 섞어 데이터의 특징으로 사용하였고, 의사결정트리는 RBF의 출력을 결합하였다. 성별 인식 결과 93.3%의 인식률을 얻었다[6].

본 논문에서는 ‘보통상태’와 ‘화난상태’의 두 정서 상태 하에서 일어나는 ‘문 두드리기,’ ‘손 흔들기,’ ‘물건 들어올리기’의 총 여섯 가지 상반신의 행동을 대상으로 하였다. 입력 특징으로는 2차원 및 3차원 속도와 좌표값을 사용하였는데, 각 행동 및 특징 별로 성별을 인식하는 전문가 신경망을 구축하였다. 이 전문가 신경망을 행동 클래스 인식기와 의사결정트리를 사용하여 결합하였으며, 이를 단일 신경망 분류기의 실험 결과와 비교였다.

3. 인간 행동 데이터

행위자의 동작은 3D 좌표 분석 시스템(Optrak, Northern Digital Co.)을 사용하여 수치로 저장되었다. 동작이 일어나는 동안 행위자의 머리, 오른쪽 어깨, 팔꿈치, 손목, 엄지 손가락, 가운데 손가락에 적외선 센서를 부착하여 60Hz로 각 센서로부터의 x, y, z 좌표값을 저장하였다. 이로부터 여섯 지점별로 2차원 및 3차원 평면상 속도(2Dvel, 3Dvel)를 각각 계산하였다. 행위자의 동작은 측면 혹은 정면에서 관찰 되기 때문에, 신경망의 입력 특징으로는 2Dvel, 2Dpos(y, z좌표), 3Dvel과 3Dpos(x, y, z축 좌표)가 함께 사용되었다. 남자 13명과 여자 13명이 ‘보통상태에서 문 두드리기,’ ‘보통상태에서 손 흔들기,’ ‘보통상태에서 물건 들어올리기,’ ‘화난상태에서 문 두드리기,’ ‘화난상태에서 손 흔들기,’ ‘화난상태에서 물건 들어올리기’의 여섯 가지 행동을 각각 10회씩 반복 수행하였다. 이들 중 8회분

은 학습데이터로, 나머지 2회분은 테스트데이터로 사용되었다. 그림 1은 전체 학습 데이터의 모습이다. 모든 패턴은 각 지점 별로 150으로 정규화 하였다.

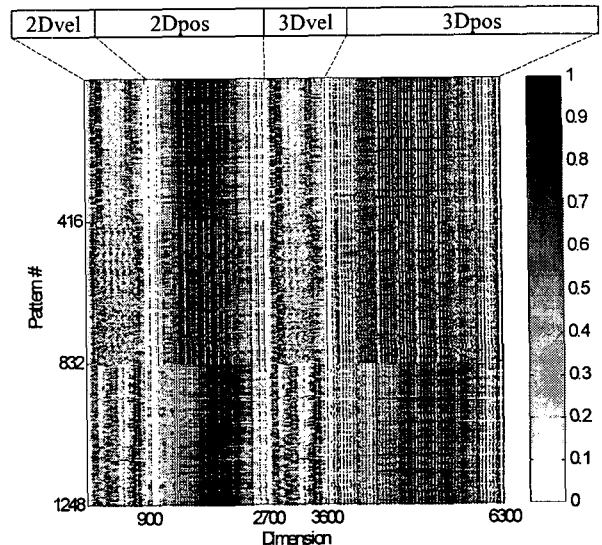


그림 1. 인간 행동 데이터

4. 성별 인식기

4.1 계층적 Modular MLP 앙상블

다층 퍼셉트론(MLP)은 안정적인 성능과 학습 능력 때문에 가장 널리 사용되고 있는 신경망 중의 하나이다. 또, 오류 역전파 알고리즘은 델타 학습 규칙의 하나로, 최종 출력층의 오류를 이전 층으로 반복적으로 퍼드백함으로써 노드들 간의 연결 가중치를 업데이트한다.

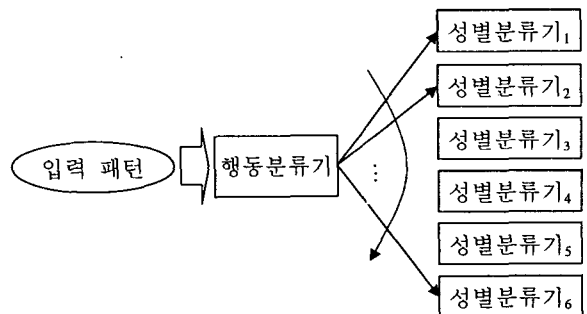


그림 2. Modular MLP 앙상블(EMMLP)

행동 데이터에는 ‘보통상태에서 문 두드리기,’ ‘보통상태에서 손 흔들기,’ ‘보통상태에서 물건 들어올리기,’ ‘화난상태에서 문 두드리기,’ ‘화난상태에서 손 흔들기,’ ‘화난상태에서 물건 들어올리기’의 여섯 종류의 행동과 남자 혹은 여자의 성별이 존재한다. 이에 따라 입력 패턴이 어떤 행동인지를 맞추는 행동분류기와, 각 행동 클래스별 성별인식 분류기를 그림 2와 같이

구성하였다. 성별분류기₁은 ‘보통상태에서 문두드리기’ 행동만으로 학습된 신경망, 성별분류기₂는 ‘보통상태에서 물건들어올리기’만으로 학습된 신경망이다. 성별분류기는 해당 클래스 내에서만 학습이 이루어지므로 단일 신경망을 사용하는 경우보다 학습도 쉽고, 클래스 내에서의 분류 성능을 최대화 할 수 있다. 나머지 성별분류기 4개도 이와 같이 학습된 modular MLP들이다. 입력 패턴이 들어오면 앞서 언급한 여섯 행동 클래스 중 어느 클래스에 해당하는지를 분류하는 행동인식기가 출력값을 내고, 해당 클래스의 성별분류기가 최종 답을 결정하게 된다. 행동분류기가 클래스를 정확히 인식할수록 성별 인식률도 높아지겠지만, 행동분류기가 오류를 범하는 경우라 하더라도 해당 성별 인식기가 올바른 답을 낼 수 있다.

행동분류기 및 성별분류기 3계층 오류 역전파 MLP를 사용하였다. 행동분류기는 6개의 출력노드를, 성별분류기는 1~2개의 출력노드를 갖는다. 모든 신경망은 학습데이터에 대한 인식률이 98.0%가 될 때까지 학습하였다.

4.2 Modular MLP/의사결정트리 앙상블

의사결정트리(decision tree)와 같은 개념 학습 추론 시스템의 목적은 클래스 레이블이 알려져 있는 오브젝트들의 집합으로부터 귀납적으로 분류 규칙을 유도하는 방법이다. Quinlan의 C4.5 알고리즘은 데이터의 정보량을 기반으로 하여 divide-and-conquer 방식으로 의사결정트리를 구축한다. X가 학습 데이터 T를 n개의 하위 집합으로 나누면, 나누어지기 전후의 데이터 집합의 엔트로피 변화는 아래와 같이 계산된다.

$$split_info(X, T) = -\sum \left(\frac{|T_i|}{|T|} \right) \log_2 \left(\frac{|T_i|}{|T|} \right)$$

$$gain_ratio(X) = \frac{gain(X)}{split_info(X)}$$

이때, gain_ratio(X)가 최대가 되도록 X를 선택하면서 트리를 구축하게 된다.

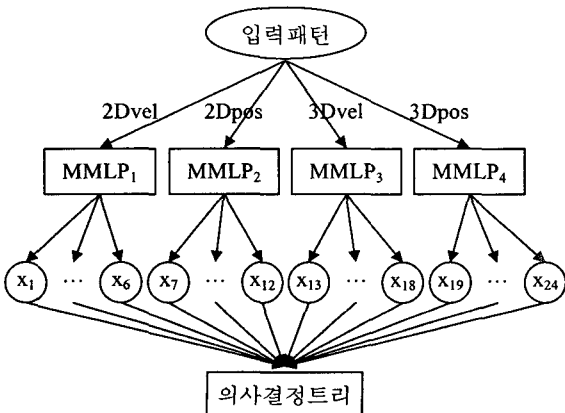


그림 3. Modular MLP/의사결정트리 앙상블 (EMMLP/DT)

그림 3은 EMMLP에서 사용되었던 여섯 개의 modular MLP의 출력값을 결합하는 모듈에 의사결정트리를 사용한 분류기 모델이다. 하나의 입력 패턴을 네 가지 특징으로 변환하고, 각각의 경우에 대하여 EMMLP에서 사용된 6개의 modular성별분류기가 각각 답을 내게 된다. 4종류의 특징이 각각 6개의 답을 내므로 총 24개의 출력값이 나오게 되는데, 이 값을 결합 모듈의 입력으로 사용하게 된다. 이 모델이 EMMLP와 다른 점은, 네 가지 데이터 특징(2Dvel, ..., 3Dpos) 정보를 모두 이용할 수 있고, EMMLP에서는 최종 선택된 한 개의 분류기가 인식 결과를 내는 반면, EMMLP/DT는 모든 분류기가 의사결정과정에 참여할 수 있다는 점이다. 또한 의사결정트리의 학습 결과 생성되는 트리 구조를 통하여 각 특징이나, 행동클래스가 서로 어떻게 관련되어있는지를 시각적으로 알아볼 수 있다.

5. 실험 결과

신경망 앙상블 분류기와 전통적 분류기와의 성능 비교를 위하여 단일 MLP, self-organizing map (SOM), support vector machine (SVM), 단일 의사결정트리(DT)가 사용되었다. 데이터는 남자 [1 0], 여자 [0 1]로 인코딩되었다. MLP는 입력노드 900~2700, 은닉노드 50~70, 출력노드 2개, 학습률 0.001~0.2가 사용되었다. SOM은 10 by 10 지도를 사용하였고, 모든 실험은 5회 실시하였다. 성능 평가에는 인식률과 식별도(discriminability) 및 효율도(efficiency)가 사용되었다. 1,248개의 학습데이터와 312개의 테스트데이터를 사용한 실험 결과는 다음과 같다.

표 1. 분류기별 인식률 [%]

분류기	데이터				평균
	2Dvel	2Dpos	3Dvel	3Dpos	
EMMLP	80.0	84.6	80.8	86.5	82.9
MLP	75.2	59.6	81.4	84.6	75.2
SVM	68.6	75.0	71.8	73.2	72.2
SOM	65.7	76.6	60.6	76.9	70.0
DT	70.2	69.2	67.0	72.8	69.8
분류기	데이터				평균
	복합 특징				
EMMLP/DT	81.4				81.4
MLP	77.2				77.2
SOM	55.9				55.9
사람	51.3				51.3

표 1은 개별 데이터 특징을 사용한 경우와 복합 특징을 사용한 경우의 인식률 결과이다. 개별 특징을 사용한 경우는 EMMLP가 평균 82.9%, 복합 특징을 사용한 경우는 EMMLP/DT가 81.4%로 가장 높은 인식률을 보였다. 이는 단일 MLP나 SVM, SOM, DT 등의 방법을 사용한 경우 보다 훨씬 높은 수치이며, 다수의 modular 신경망들을 결합함으로써 인식률의 향상을 가져왔음을 볼 수 있다. 가장 인식률이 높았던 경우는 3Dpos 특징을 사용하여 행동분류기 및 modular 성별

분류기를 사용한 경우(EMMLP)로 86.5%였다. 데이터 특징 별로 인식률의 표준편차가 EMMLP의 경우 3.1로써, MLP 11.1보다 훨씬 낮은 수치를 보였으며, 이는 분류기의 안정된 성능을 입증하는 것이다. 데이터 특징 중에서는 3Dpos가 가장 좋은 결과를 보였다. 한편, 본 논문에서 사용한 동일한 학습 및 테스트데이터를 사용하여 사람을 대상으로 한 실험 결과[7] 51.3%의 인식률을 보였다. 신경망 분류기가 사람보다 훨씬 높은 인식률을 보였다.

그림 4는 EMMLP/DT결과 생성된 의사결정트리의 모습이다. F는 여자, M은 남자이다. 트리의 최상위 노드로부터, EMMLP/DT모델의 노드 출력값이 화살표 가운데의 역치값 이상일 경우 왼쪽화살표로 진행하고 이하인 경우는 오른쪽으로 진행하게 된다. 예를 들어, 어떤 테스트데이터에 대한 x_{18} 의 값이 0.48이상이고, x_{17} 값이 0.003 이하이면 여자로 분류된다.

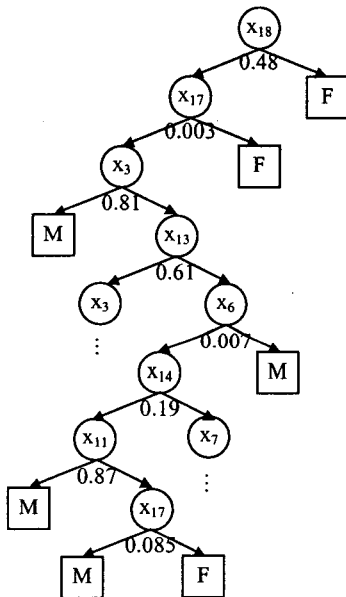


그림 4. EMMLP/DT 결과

식별도(d')는 입력 신호에 대한 관찰자의 식별력을 confusion matrix를 사용하여 계산하는 방법이다. 식별도는 탐지(hit) 및 오경보(false alarm)율을 기반으로, 외부의 입력 신호가 들어왔을 때와 들어오지 않았을 때의 내부반응 확률분포 곡선 간의 거리로 계산된다. 따라서 식별도가 큰 값일수록 신호에 대한 식별력이 높음을 의미한다.

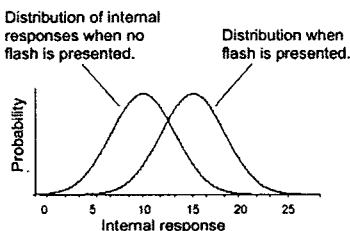


그림 5. 내부반응 확률분포 곡선

또, 입력 신호에 대하여 관찰자 A와 B가 각각 d_A' 과 d_B' 의 식별도를 가질 때 B의 A에 대한 효율도는

$$E_{B \rightarrow A} = \left(\frac{d_A'}{d_B'} \right)^2$$

이다. 이를 적용하여 분류기별 식별도(표 2)와 사람에 대한 EMMLP와 EMMLP/DT의 효율도를 계산한 결과(표 3), 이 경우도 역시 신경망 앙상블 분류기가 가장 높은 식별도를 가짐을 알 수 있었다. EMMLP가 가장 높은 수치를 나타내었는데, 이번에는 2Dpos를 사용한 결과가 1.30으로 가장 높은 식별도를 기록하였다.

표 2. 분류기별 식별도

분류기	데이터				평균
	2Dvel	2Dpos	3Dvel	3Dpos	
EMMLP	1.20	1.30	1.07	1.22	1.20
EMMLP/DT	1.03				1.03
MLP	0.73	0.27	0.98	0.75	0.68
SVM	0.62	0.77	0.70	0.66	0.69
SOM	0.54	0.91	0.37	1.01	0.71
DT	0.71	0.66	0.62	0.73	0.68
사람	0.13				0.13

표 3. 분류기별 효율도

분류기	데이터				평균
	2Dvel	2Dpos	3Dvel	3Dpos	
EMMLP	85.2	100.0	67.7	88.1	85.3
EMMLP/DT	62.8				62.8

6. 결 론

본 논문에서는, 인간 행동의 성별 인식을 위하여 다수의 전문화된 신경망을 결합하여 분류성능을 향상 하였음을 보였다. 인식률, 식별도 및 효율도를 사용하여 EMMLP 및 EMMLP/DT가 MLP, SVM, SOM, DT 및 사람보다 훨씬 좋은 성능을 보임을 입증하였다.

참고문헌

- [1] G. Johanssen, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, pp. 201~211, 1973.
- [2] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Patt. Anal. Mach. Inte.*, vol. 12, pp. 993~1001, 1990.
- [3] J. E. Cutting, et al., "Recognising friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, pp. 353~356, 1977.
- [4] S. Haykin, *Neural networks: A comprehensive foundation*, 2nd Eds., Chapter 7, pp. 351~391, Prentice Hall, 1999.
- [5] D. N. Osherson, S. Weinstein and M. Stoli, "Modular learning," *Computational Neuroscience*, E. L. Schwartz, Eds., pp. 369~377, Cambridge, MA : MIT Press, 1990.
- [6] S. Gutta, et al., "Mixture of experts for classification for gender, ethnic origin, and pose of human faces," *IEEE Trans. Neural Networks*, vol. 11, no. 4, July 2000.
- [7] F. E. Pollick, V. Lestou, J. Ryu and S. B. Cho, "Estimating efficiency in the categorization of biological motion," *Proc. of Vision Science Society*, pp. 31, May 2001.