The Korean Journal of Applied Statistics (2018) DOI: http://dx.doi.org/10.5351/KJAS.2018.31.5.539
31(5), 539–554

Analysis of facial expression recognition

Nayeong Son[a] · Hyunsun Cho[a] · Sohyun Lee[a] · Jongwoo Song[a,1]

[a]Department of Statistics, Ewha Womans University

(Received November 2, 2017; Revised August 31, 2018; Accepted September 4, 2018)

Abstract

Effective interaction between user and device is considered an important ability of IoT devices. For some applications, it is necessary to recognize human facial expressions in real time and make accurate judgments in order to respond to situations correctly. Therefore, many researches on facial image analysis have been preceded in order to construct a more accurate and faster recognition system. In this study, we constructed an automatic recognition system for facial expressions through two steps - a facial recognition step and a classification step. We compared various models with different sets of data with pixel information, landmark coordinates, Euclidean distances among landmark points, and arctangent angles. We found a fast and efficient prediction model with only 30 principal components of face landmark information. We applied several prediction models, that included linear discriminant analysis (LDA), random forests, support vector machine (SVM), and bagging; consequently, an SVM model gives the best result. The LDA model gives the second best prediction accuracy but it can fit and predict data faster than SVM and other methods. Finally, we compared our method to Microsoft Azure Emotion API and Convolution Neural Network (CNN). Our method gives a very competitive result.

Keywords: image classification, Haar cascade, face landmark, data mining

1. 서론

이미지자료의 분석이란 색상정보(RGB)와 광도정보(luminosity)를 이용하여 이미지의 특징을 파악하고, 이를 통해 피사체의 형태를 식별하거나 분류하는 것을 의미한다. 최근 등장하는 다양한 사물인터넷 기기 혹은 상황인식 기반의 인공지능에서는 사용자와 기기의 상호작용(interface)이 중요시되는데, 특히 인간의 표정을 실시간으로 인식하여 상황에 맞는 대응을 하기 위해서는 고도의 관찰력과 빠른 반응속도를 필요로 한다. 따라서, 안면의 특징을 보다 빠르고 쉽게 파악하여 감정을 정확히 예측하는 알고리즘을 개발하기 위해 많은 노력이 선행되어 왔다 (Lyons 등, 1999; Shan 등, 2009).

자동화된 얼굴 표정 인식 시스템을 구축하기 위해서는, 이미지에 포함된 많은 형태 중에서 사람의 얼굴을 정확히 식별해내는 인식작업과 특징을 추출하여 감정을 파악하는 분류작업이 유기적으로 이루어져야 한다. 이러한 요구에 맞추어 기존의 표정인식 연구는 다양한 특징정보(feature information)를 획득하

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017R1D1A1B03036078).

[1]Corresponding author: Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: josong@ewha.ac.kr

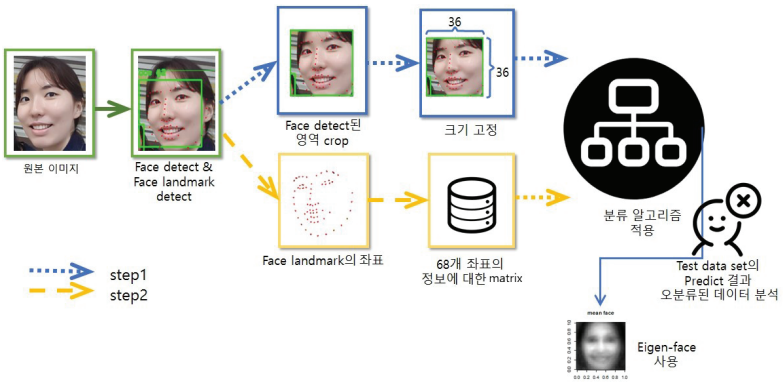

Figure 2.1. Facial expression recognition algorithm.

는 특징추출(feature extraction) 방법론을 바탕으로 꾸준히 진행되어 왔으며, 주로 기계학습(machine-learning) 분야의 과제로 여겨진다. 이미지의 특징 정보는 크게 피사체의 모양에 대한 shape information과 질감에 대한 texture information으로 구분되며, active shape model (ASM) (Cootes 등, 1955)이나 active appearance model (AAM) (Cootes 등, 2001) 등이 shape와 texture information을 종합하는 모델링 방법으로 잘 알려져 있다. Shape information은 피사체의 형태적 특성을 나타내는데, 특히 기하학적 성질을 나타낼 때는 기하정보(geometric information)으로도 이해된다. 사람의 얼굴에 관련된 대표적인 shape information으로는 얼굴 랜드마크(face landmark)를 사용할 수 있다. 각 얼굴에서 얼굴 랜드마크를 추출한 후, 이들간의 상대적 위치, 거리, 및 각도 등의 정보를 생성하면 얼굴의 형태 정보를 수치화하는 것이 가능하다. 한편 texture information은 이미지의 패턴을 이용하는 방법으로, Gabor filter (Daugman, 1988; Fogel과 Sagi, 1989; Jain과 Farrokhnia, 1991)나 locally binary pattern (LBP) (Ojala, 2002) 등을 적용하여 얻을 수 있다.

본 연구에서는 웹사이트 Kaggle (Goodfellow 등, 2013)에서 제공한 48×48픽셀 8-bit grayscale 이미지 데이터셋을 사용하여 얼굴인식과 표정분류로 구분된 두 단계를 거치는 얼굴표정 자동 인식 시스템을 구축하였고, 이를 기존의 연구와 비교하여 자료 및 방법론의 특징을 고찰하였다. 2장에서는 이미지에서 사람의 얼굴을 인식하는 Haar-cascade 알고리즘과 face ladnmark의 방법론 및 구현을 기술하였고, 4장에서 최종적으로 비교될 자료들의 구성을 소개하였다. 3장에서는 자료의 전처리 과정을 비롯하여 랜드마크 좌표를 이용한 표정 관련 변수 생성 과정을 서술하여 자료 구성 과정을 밝혔다. 다음으로, 4장에서는 각 자료에 linear discriminant analysis (LDA), random forest (Breiman, 2001), support vector machine (SVM) (Cortes와 Vapnik, 1995), bagging (Brieman, 1996) 등의 통계적 분류 방법론을 적용한 결과를 비교하였고, 얼굴 이미지 중 표정이 잘 분류되는 경우와 그렇지 않은 경우를 선별하여 특징을 분석하였다. 마지막으로 5장에서는 Microsoft Azure API를 적용하였을 때의 분류결과와 최종 모형에 의한 분류결과를 비교하였고, 본 연구의 결론 및 시사점을 서술하였다.

2. 방법론

본 연구에서는 사람의 얼굴 이미지를 이용해 표정을 분류하는 자동 분류기를 구현하고자 하였다. 분류과정은 이미지의 특징을 추출하는 과정과 통계적 예측모형을 적용하는 과정으로 나뉜다. 이번 장에서는 얼굴 이미지의 특징 추출 기법들에 대해 설명하고 분류 모델을 적용할 자료 구성을 간략하게 소개할 것이다. Figure 2.1은 본 장에서 설명하고자 할 알고리즘들을 도식화한 그림이다.

Figure 2.2. Haar features; (left) Edge feature, (middle) Line feature, (right) Center surrounded feature and special diagonal line feature.

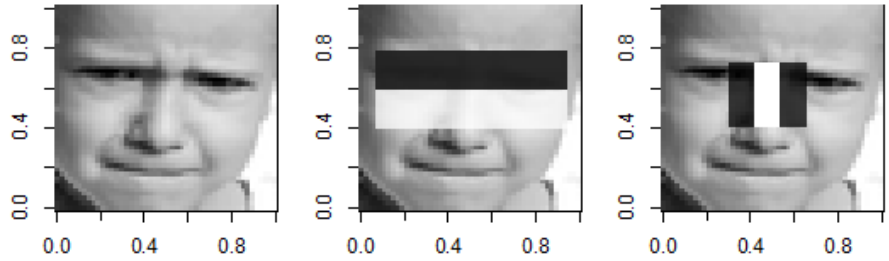

Figure 2.3. Haar features example for face detection.

2.1. 얼굴 식별과 정보 추출

2.1.1. Haar cascade Viola와 Jones에 의해 처음 제시된 Haar cascade (Viola와 Jones, 2001) 기반 형태인식 아이디어는 크게 Haar features의 사용과 cascade algorithm의 구현으로 구성된다. Haar feature란 단순한 의미에서 Figure 2.2와 같은 데이터의 사각형 패턴(rectangular pattern)이라고 할 수 있다.

각각의 사각형 feature에 convolution kernel을 적용하면 밝은 부분과 어두운 부분에 해당하는 픽셀 값 총합의 차이에 해당하는 하나의 값이 얻어지며, 이것은 상대적 명암과 관련되어 있다. 얼굴 인식에서 Haar feature의 적용이란 사람의 얼굴이 공통적으로 갖는 명암적 패턴에 Haar like feature를 적용하여 얼굴인지 여부를 판단하는 임계값(threshold)을 정의하는 것을 의미한다. 다음 그림은 얼굴에 Haar feature를 적용하는 예시이다.

Figure 2.3을 보면, 사람의 얼굴 구조상 눈 부근은 어둡게 표현되는 반면 광대 부분은 튀어나와 밝게 표현되는 경향이 있음을 알 수 있다. 따라서, Figure 2.3의 가운데 이미지와 같이 edge feature를 적용하는 것이 적절하다. 한편, 일반적으로 콧대는 밝게 표현되며 코 주변은 상대적으로 어둡게 표시되므로 Figure 2.3의 오른쪽 이미지와 같이 line feautre를 적용할 수 있다. Haar cascade는 이렇게 다양한 크기와 모양의 Haar feature와 그에 따른 임계치 수준의 조합을 이용해 사람의 얼굴 부위를 학습시킨다. 그러나 전체 이미지에서 얼굴을 찾기 위해 다양한 종류와 크기의 Haar feature를 모두 고려하는 것은 매우 비효율적이다. 따라서, Adaboost와 Cascade algorithm 등을 도입하여 training data에서 얼굴지역을 구분하는 최적의 Haar feature 및 임계치를 찾고, 얼굴이 위치할 가능성이 높은 지역에서만 탐색하는 효율적 알고리즘이 고안되었다 (Viola와 Jones, 2004).

위 알고리즘 (Viola와 Jones, 2004)을 이용하여 얼굴 객체를 탐지하기 위해서는 수많은 이미지들이 학습되어야 한다. 본 연구에서는 Python(https://www.python.org/, Python 3.6.0 version) API인 OpenCV-Python 라이브러리(http://sourceforge.net/projects/opencvlibrary/) (Bradski와 Kaehler, 2008)를 이용하였으며, cascade기반으로 미리 학습된 정면 얼굴 데이터(Haarcascade_frontalface_de

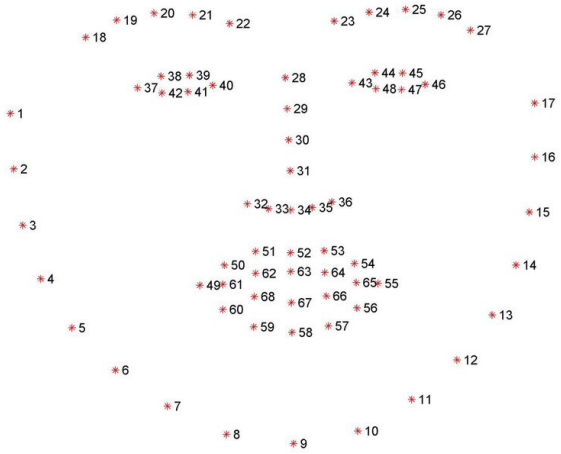

Figure 2.4. Face landmark.

fault.xml) (Steven Puttemans, 2013)를 제공받아 정면 얼굴을 식별하였다.

2.1.2. Face landmark 이미지에서 사람 얼굴의 영역을 찾은 후에는 표정을 구분할 수 있도록 얼굴의 구성 요소들을 찾아야 한다. 본 연구에서는 68개의 랜드마크로 얼굴의 구조를 데이터화하는 방법을 사용하였다. 이 방법을 이용하면 Figure 2.4와 같이 각 얼굴의 특정 포인트를 기계학습 알고리즘을 통해 훈련시키고, 새로운 이미지에서 공통된 68개의 랜드마크를 찾을 수 있다. 이의 구현에는 Python과 dlib (King, 2009) 라이브러리를 이용하였고, 68개의 점이 학습된 모델 파일(shape_predictor_68_face_landmarks.dat)을 다운받아 사용하였다. 이 얼굴 구성 요소 검출기는 선형 분류기, 이미지 피라미드(image pyramid) 및 슬라이딩 윈도우 감지 체계(sliding window detection scheme)와 결합된 histogram of oriented gradients (HOG) 기능을 사용하여 만들어졌다. 이는 Kazemi와 Sullivan (2014) 논문에 기반하였다.

2.2. 데이터 프레임과 분류 모델

이상에서 설명한 얼굴 특징 추출 방법론을 이용하면 최종적으로 픽셀에 관한 정보와 랜드마크에 관한 정보를 얻을 수 있다. 본 연구에서 비교하고자 하는 최종 데이터셋은 Table 2.1과 같다.

Table 2.1에서 각 자료는 다음과 같이 구성되었다. 자료1은 36×36픽셀 이미지에서 추출한 1,296개 픽셀을 독립적인 변수로 정의한 자료이며 픽셀의 밝기를 나타내는 8-bit grayscale 값을 각 변수의 값으로 사용하였다. 자료2는 픽셀정보를 사용하지 않으며 추정된 랜드마크의 좌표와 중심점으로부터의 거리·각도 정보를 모두 이용하는 자료이다. 이 때 얼굴형에 해당하는 1–17번의 점은 제외하였으므로 18–68번 점의 x좌표, y좌표, 중심점으로부터의 거리, 각도가 각각 51개씩 포함되었다. 자료3은 픽셀, 랜드마크 좌표, 거리·각도 정보를 모두 사용한 자료이다. 한편, 자료4에서는 자료3을 이루는 변수항목에 각각 PCA를 적용하여 픽셀정보에서 50개, 랜드마크 좌표정보에서 20개, 거리·각도 정보에서 30개씩의 PC를 추출하였다. 자료5는 자료4와 마찬가지로 픽셀에 PCA를 적용하여 50개의 PC를 추출하였지만, 랜드마크 좌표 정보와 거리·각도 정보를 함께 주성분 분석하여 총 30개의 PC를 추출했다는 차이점이 있다. 따라서, 자료5는 자료4에 비해 변수의 개수가 20개 적다. 자료6는 자료5와 같이 랜드마크 좌표정보와 거리·각도 정보를 함께 주성분 분석하여 총 30개의 PC를 추출하였으나 픽셀정보를 사용하지 않

Table 2.1. Structure of data

자료	총 변수 개수	변수 상세		개별 변수 개수
자료1	1296	픽셀	Raw	1296
		랜드마크좌표	X	0
		거리각도	X	0
자료2	204	픽셀	X	0
		랜드마크좌표	Raw	102
		거리각도	Raw	102
자료3	1500	픽셀	Raw	1296
		랜드마크좌표	Raw	102
		거리각도	Raw	102
자료4	100	픽셀	PCA	50
		랜드마크좌표	PCA	20
		거리각도	PCA	30
자료5	80	픽셀	PCA	50
		랜드마크좌표 & 거리각도	PCA	30
자료6	30	픽셀	X	0
		랜드마크좌표 & 거리각도	PCA	30
자료7	41	픽셀	PCA	21(변수선택)
		랜드마크좌표 & 거리각도	Raw	20(변수선택)
자료8	20	픽셀	X	0
		랜드마크좌표 & 거리각도	Raw	20(변수선택)
자료9	18	픽셀	X	0
		랜드마크좌표 & 거리각도	PCA	18(변수선택)

는다.

자료7–9는 R 패키지 klaR (Weihs 등, 2005)의 stepclass함수에 의해 변수를 선택하여 구성한 자료이다. 변수선택은 stepwise selection으로 진행하였으며 LDA모형의 정확도가 0.01% 이하로 개선되면 변수선택을 중지하였다. 픽셀정보에서는 1,296개의 픽셀에 PCA를 적용하하여 21개의 변수가 선택되었다. 랜드마크 정보에서는 좌표, 거리·각도를 모두 포함한 원 자료를 후보 변수로 지정하였을 때 20개의 변수가 선택되었고, 이에 PCA를 적용하였을 때는 18개의 변수가 선택되었다. 변수선택 결과에 따라 자료7–9를 다음과 같이 구성하였다. 자료7은 PCA를 적용한 픽셀정보와 랜드마크 정보로부터 각각 선택한 변수를 포함한다. 반면 자료 8은 픽셀정보를 일체 사용하지 않고 랜드마크 정보로부터 선택한 변수들만을 포함하였다. 나아가, 자료 9에서는 랜드마크정보에 PCA를 적용하여 선택한 18개 변수만을 이용하였다.

3. 데이터 소개

3.1. 분석 자료 출처

본 연구에서는 빅데이터 애널리틱스 플랫폼인 Kaggle(https://www.kaggle.com/)의 Challenges in Representation Learining: Facial Expression Recognition Challenge 자료 (Goodfellow 등, 2013)를 분석하였다. 제공받은 데이터는 사람의 얼굴을 포함한 48∗48 크기의 8-bit 흑백 이미지이며, 각 이미지의 표정을 나타내는 7가지 범주 라벨(총 33,298개 이미지)을 포함하고 있다. 본 논문에서는 P.Ekman (Ekman, 1992)의 6가지 기본 감정 중 주로 연구되어 온 4가지 기본 감정인 화남, 행복함, 슬픔, 중립(총

18,245개 이미지)만을 사용하였다 (Saatci와 Town, 2006).

3.2. 데이터 전처리

이미지에서는 동일한 피사체를 촬영하였더라도 촬영 각도와 장소 등의 외부 조건에 따라 대상의 형태 및 밝기가 다르게 나타나기 때문에, 이미지를 식별하여 가용한 자료로 만드는 데에는 전처리과정이 매우 중요한 것으로 여겨진다. 일반적으로 사람의 얼굴 이미지 분석에 필요한 전처리 작업으로는 얼굴의 크기 및 각도 조정과 사진의 밝기 조정이 요구된다. 다음의 과정은 3.1에서 제시한 데이터셋에 적용한 데이터 전처리 과정이다.

1. 자동 레벨

   자료(3.1)는 흑백 이미지 자료로서, 각 픽셀이 갖는 값은 빛의 세기(광도)를 의미한다. 따라서 피사체를 비추는 빛의 방향과 각도에 의한 효과를 고려하여 결과의 왜곡을 방지하기 위해 사진 편집기인 포토스케이프(Photoscape®software)의 자동 레벨 기능으로 각 사진의 밝기분포를 조정하였다.

2. 얼굴 식별

   Kaggle으로부터 제공받은 최초의 자료는 다수의 정면 얼굴로 구성되어 있지만, 일부 이미지는 표정을 알아볼 수 없는 옆모습이거나 사람의 얼굴이 아닌 이미지이다. 얼굴이 바라보는 방향이나 위치가 정면과 크게 다른 이미지는 얼굴 형태나 표정을 식별하기에 어려움이 있다. 따라서 표정을 잘 식별할 수 있는 정면 얼굴 이미지만을 선택하여 분석하였는데, 이를 위해 자동 레벨로 밝기가 보정된 48×48픽셀 이미지에 Haar Cascade 알고리즘을 적용하였다. 이 과정에서 정면 얼굴이 아니거나 사람의 얼굴이 아닌 이미지는 인식되지 않아 분석 자료에서 제외하였다. 또한 원본 이미지에서 얼굴이 차지하는 면적이 10×10픽셀보다 작은 경우는 없었기 때문에, 감지된 이미지의 크기가 10×10픽셀보다 작은 경우 얼굴이 아닌 것으로 판단하여 일괄 제외하였다. 식별된 정면 얼굴은 각기 다른 크기와 모양으로 추출되어 저장된다.

3. 크기 조정

   이상의 과정을 거친 결과 얼굴의 위치는 조정되었으나 추출된 얼굴 이미지의 크기가 상이하였다. 따라서 포토스케이프 프로그램의 내장 함수로 모든 이미지의 크기를 동일하게 조정하여 자료의 차원을 동일하게 하였다. 이 때 사용한 함수는 nearest neighbors (NN) 알고리즘에 기반한 것으로, 채우고자 하는 픽셀의 값을 주위 픽셀 정보를 이용해 대체하는 방법이다. 그 결과, 얼굴이 인식된 모든 이미지는 얼굴이 중앙에 위치하는 36×36크기 이미지로 변형되었다.

4. 랜드마크 추출

   2장에서 서술한 face landmark 기법을 이용하면 얼굴의 특정 위치에 해당하는 점의 좌표가 추출된다. 이 때 각 랜드마크 좌표는 식별된 얼굴의 중심점을 원점으로 하는 좌표공간상에서 정의되었다. 각 랜드마크의 X, Y 좌표 값 자체를 설명변수로 사용할 수도 있지만, 본 연구에서는 좌표 간의 기하적 관계를 추가로 도입하여 보다 효율적으로 표정을 설명하고자 하였다. 좌표 간의 기하학적인 관계로는 얼굴의 중심점(랜드마크 좌표의 원점)으로부터 각 좌표까지의 유클리디언 거리와 아크탄젠트 각도를 고려하였다. 따라서, 본 연구에서 사용한 랜드마크 정보란 각 랜드마크의 좌표 정보와 중심으로부터의 거리, 각도를 의미한다. 랜드마크 정보 변수의 값을 계산하는 알고리즘은 다음과 같다.

   Step 1. 68개의 랜드마크 좌표 추출

   Step 2. 랜드마크 좌표의 중심(mean) 계산 및 18–68번 랜드마크 좌표 sclae 조정

   Step 3. 랜드마크 좌표의 중심으로부터 각 좌표의 유클리디안 거리 계산

Table 3.1. Number of images in each step

Data step	Angry	Happy	Sad	Neutral	Total
Kaggle 제공 최초 데이터(4개라벨)	3,245	6,977	3,418	4,605	18,245
중복이미지, 라벨투표결과 제거 후	2,812	6,521	2,983	4,204	16,520
Haar cascade, Face landmark 적용 후	1,450	4,100	1,379	2,534	9,464

Step 4. 콧대의 각도(nose angle) 계산

Step 5. Nose angle로 보정한 각 좌표의 아크탄젠트 각도 계산

이상의 알고리즘을 Python으로 구현하는 데에는 Paul van Gent의 웹 페이지를 참고하였음을 밝힌다 (Paul van Gent, 2016).

3.3. 최종 자료

본 논문에서는 3.2장의 전처리 과정을 거친 픽셀자료와 face landmrak 자료를 이용하여 여러 종류의 데이터셋을 구성하고 이들을 비교하고자 한다. 우선, 픽셀자료와 face landmark자료의 변수 구성은 다음과 같다. 모든 이미지는 3.2장의 전처리 과정을 거친 후 36×36개의 픽셀을 가진 이미지로 변형되므로, 모든 픽셀을 독립된 개별 변수로 취급한다면 총 1,296개의 변수를 갖는 픽셀자료가 생성된다. 본 연구에서 사용한 이미지는 8-bit greyscale 이미지이므로 각 픽셀에 해당하는 변수는 0–255의 값을 갖는다. 다음으로, Face landmark자료의 경우 본래 68개의 랜드마크에 각각 X좌표, Y좌표, 중심점으로부터의 거리, 각도 값이 부여되므로 총 $68 \times 4 = 272$개의 변수를 갖는다. 그러나 본 연구에서는 얼굴형을 나타내는 1–17번 랜드마크가 표정 예측에 불필요하다고 가정하여 분석 데이터에서 제외하였다. 따라서, 랜드마크 자료로부터 사용되는 변수의 수는 $(68 - 17) \times 4 = 204$개이다.

다음은 전처리 전후 이미지의 개수 변화에 대한 설명이다. Kaggle에서 제공받은 원본 자료에는 관심 라벨(angry, happy, sad, neutral)에 해당하는 총 18,245개의 이미지가 있었지만, 이 중 다수가 중복된 이미지이거나 불명확한 라벨을 가진 것으로 드러났다. 따라서, 원본 자료를 그대로 사용하지 않고 다음의 기준에 의해 일부 데이터를 먼저 제거하였다. 첫째, 원본의 픽셀 자료 값을 기준으로 모든 픽셀 값이 같은 경우 중복된 이미지로 판단하였다. 중복된 이미지가 존재하면 분류기의 성능을 공정하게 평가할 수 없으므로 분석에서 제외하였다. 둘째, 실제 이미지와 그에 해당하는 종속변수(라벨)이 다른 경우, 혹은 얼굴 표정이 한 가지 감정에 귀속되기에 애매한 모양인 경우를 고려하였다. Figure 3.1과 같이 주어진 라벨이 실제 이미지가 나타내는 감정과 다를 경우 학습기의 성능을 저해할 수 있으므로 분석에서 제외하는 것이 옳다. 그러나 감정에 대한 판단은 주관적이라는 점을 고려하여 본 논문의 연구자 3인을 대상으로 비밀 투표를 실시하였고, 그 결과에 따라 3인 중 2인 이상이 라벨에 동의하지 않는 이미지를 제거하였다.

마지막으로, 위 과정을 거친 이미지 중 Haar cadcade 방법과 face landmark 추출 알고리즘을 모두 통과하여 식별 가능한 자료로 분류된 이미지만을 가지고 최종 자료를 구성하였다. 이상의 모든 과정을 거친 관측치의 개수는 9,464개(Table 3.1의 가장 하단 참고)이며, 각 자료 구성 단계에서 각 범주의 개수는 다음과 같다.

3.4. 자료의 한계

기존의 얼굴표정 분류 연구에서는 흔히 실험적으로 제작된 데이터셋이 사용되었다. 이러한 데이터 셋에 포함된 이미지는 주로 고정된 장소와 위치에서 촬영되었으며, 각 인물의 여러 표정을 여러 각도에서

Figure 3.1. Examples incorrect label; 0 = Angry, 3 = Happy, 4 = Sad, 6 = Neutral.

Figure 3.2. Ambiguity in facial expressions.

반복 촬영한 고화질 자료이다. 따라서 이러한 데이터셋이 포함한 이미지들은 각 표정이 갖는 특징을 적나라하게 드러낸다. 반면 본 연구에서 사용한 자료는 인터넷으로 수집된 다양한 얼굴사진으로 구성되어 있으며 인간의 얼굴이 아닌 2D, 3D 애니메이션 캐릭터의 얼굴 등을 포함한다. 이외에도 안경을 쓰거나 워터마크가 포함된 사진, 어두운 곳에서 촬영되어 배경과 인물의 구분이 어려운 사진 등 분류가 어려운 이미지들이 포함되어 있다.

또한, 인터넷으로 수집한 이미지 자료는 미리 표정을 설정하여 촬영한 사진이 아니므로 감정이 드러나는 정도가 다양하다. 따라서 Figure 3.2에서 예로 제시한 바와 같이 이미지를 단순히 어느 한 표정 범주에 귀속시키기에 어려움이 있을 수 있다. 예를 들어, 울고 있는 사람이 동시에 화가 난 경우 정확하게 범주를 분류하기 어려울 것이다. 또한 사람의 생김새에 따라 화가 나지 않았는데도 불구하고 원래의 인상이 화난 것처럼 보이는 사람도 있을 것이다.

Figure 3.2의 이미지는 모두 Neutral의 라벨을 갖고 있는 이미지들이다. 그러나 보는 사람의 관점에 따라 (왼쪽부터) 첫 번째 이미지는 화남, 두 번째 이미지는 행복함, 세 번째와 네 번째의 이미지는 슬픔에 속한다고 판단할 수 있다. 기존에 원활한 연구를 위하여 사진에서 3.3절의 비밀 투표 과정을 거쳤으나, 여전히 Figure 3.2와 같이 두 가지 이상 범주에 속할 수 있는 이미지들이 존재한다.

4. 분석 결과

이번 장에서는 3장에서 제시한 과정을 거쳐서 구성한 자료에 통계적 분류 방법을 적용한 결과를 비교하고, 잘 분류되는 이미지와 오분류되는 이미지의 특징을 파악하고자 한다. 이 장에서 제시하는 내용은

Table 4.1. Structure of data

자료 구성					CV accuracy			
자료	총 변수 개수	변수 상세		개별 변수 개수	LDA	RF	SVM	Baggging
자료1	1296	픽셀	Raw	1296	0.539	0.603	0.640	0.609
		랜드마크좌표	X	0				
		거리각도	X	0				
자료2	204	픽셀	X	0	0.708	0.698	0.712	0.699
		랜드마크좌표	Raw	102				
		거리각도	Raw	102				
자료3	1500	픽셀	Raw	1296	0.664	0.692	0.706	0.693
		랜드마크좌표	Raw	102				
		거리각도	Raw	102				
자료4	100	픽셀	PCA	50	0.715	0.693	0.739	0.698
		랜드마크좌표	PCA	20				
		거리각도	PCA	30				
자료5	80	픽셀	PCA	50	0.712	0.677	0.734	0.678
		랜드마크좌표 & 거리각도	PCA	30				
자료6	30	픽셀	X	0	0.705	0.692	0.721	0.686
		랜드마크좌표 & 거리각도	PCA	30				
자료7	41	픽셀	PCA	21(변수선택)	0.703	0.695	0.710	0.695
		랜드마크좌표 & 거리각도	Raw	20(변수선택)				
자료8	20	픽셀	X	0	0.705	0.696	0.703	0.693
		랜드마크좌표 & 거리각도	Raw	20(변수선택)				
자료9	18	픽셀	X	0	0.692	0.681	0.694	0.675
		랜드마크좌표 & 거리각도	PCA	18(변수선택)				

LDA = linear discriminant analysis; RF = random forests; SVM = support vector machine.

R프로그램(R version 3.3.2)으로 구현한 결과이며, MASS (Venables와 Ripley, 2002), randomForest (Liaw와 Wiener, 2002), e1071 (Meyer 등, 2017) 패키지를 이용하였다. Table 4.1은 예측모형에 포함되는 변수의 종류와 형식, 개수에 따라 다양하게 구성된 자료에 LDA, Random Forest, SVM, Bagging을 적용한 결과를 제시하고 있다. Table 4.1의 CV정확도는 최종 데이터에 10-fold CV를 10회 적용하여 정확도의 평균을 구한 값이다.

4.1. 모형 성능 비교

Table 4.1의 모든 자료에서 SVM이 가장 높은 정확도를 보였으며, LDA가 SVM 다음으로 높은 정확도를 기록하였다. 대부분의 자료에서 Random Forest나 Bagging은 비슷한 정도의 정확도를 보이며 자료1을 제외하고는 네 방법론 중 가장 성능이 낮다. 모형 적합 시간은 LDA < Random Forest < SVM < Bagging 순으로 더 긴 시간을 요한다. 자료5 기준 전체 데이터의 90%를 train set으로 하여 모형을 적합하고 10%를 test set으로 하여 예측값을 계산하는 데 걸린 총 시간은 LDA 1.040초, Random Forest 12.210초, SVM 12.874초, Bagging 20.636초이다. 이상의 구동시간은 i7-4790 CPU, 16GB 메

모리와 64bit 운영체제를 갖춘 PC에서 측정하였다. 물론 새로운 데이터에서 예측값의 계산은 즉각적으로 이루어진다. SVM 정확도가 가장 높은 자료는 자료4로, 픽셀과 랜드마크 좌표, 거리·각도에 각각 주성분분석을 적용하여 $50, 20, 30$개의 PC를 추출한 자료이다. 자료4에서 100개의 변수를 사용하여 73.9%의 SVM정확도를 보이는 한편, 자료5에서는 변수의 개수가 20개 더 적으면서도 SVM정확도가 73.4%를 기록한다. 두 자료에서 SVM 다음으로 성능이 좋은 LDA 정확도는 단 0.3% 차이를 보인다. 자료5와 자료6의 차이는 픽셀 주성분 정보를 사용했는지의 여부이며, 자료6에서는 픽셀 주성분 정보를 사용하지 않으므로 단 30개의 변수만으로 예측모형이 구성된다. 그럼에도 불구하고 자료6의 SVM정확도는 72.1%로 자료 4에 비하여 크게 떨어지지 않는다.

변수선택을 통하여 구성한 자료7–9에서는 랜드마크 정보에 PCA를 적용하지 않을 때가 더 나은 정확도를 보이는 것으로 나타난다. 픽셀정보에 PCA를 적용하여 21개의 변수를 추가한 자료7은 픽셀정보를 사용하지 않은 자료8보다 조금 높은 SVM 정확도를 갖지만, 그 차이는 약 0.6%로 매우 미미하다. 따라서, 자료4–6의 비교 및 자료7–8의 비교를 통하여 이미 랜드마크 정보가 투입된 자료에 픽셀정보를 추가하는 것이 정확도를 높이는 데 큰 기여를 하지는 않는다는 것을 알 수 있다.

Table 4.1에서 사용한 거리·각도는 모든 랜드마크의 중심점으로부터 계산한 거리와 각도 정보이다. 이들 정보 외에 눈 중심에서의 거리, 각도 등을 추가로 고려하여 모형을 개선하려 하였으나, 본 연구의 데이터에서는 이들 정보가 추가적인 도움이 되지 않는 것으로 밝혀졌다. 우리는 그 이유를 다음과 같이 제시한다. 양쪽 눈의 중심으로부터 각 눈썹의 랜드마크까지 계산한 거리와 각도는 눈썹의 위치와 모양을 잘 드러내는 변수이며, 얼굴 표정을 분류하는데 매우 중요한 요소이다. 그러나 본 연구에서 사용한 특징추출 기법인 face landmark는 해상도가 매우 낮은 사진에서 일부 랜드마크 위치를 제대로 추출하지 못하는데, 특히 눈썹의 위치를 잘 못 추정하는 경우가 많다. 따라서 양쪽 눈의 중심으로부터 거리와 각도를 계산하더라도 눈썹의 모양이 제대로 추정되지 않았으면 표정을 정확히 예측하는 데 큰 도움이 되지 못한다.

4.2. 최종 모형

Table 4.1에 제시한 결과에 따라 사용 변수 개수가 적고 모델 적합이 빠르며 비교적 CV 정확도가 높게 나타난 자료6-SVM모형을 본 논문의 최종 모형으로 선택하였다. 자료5(자료6)은 단 30개 PC만을 이용해 랜드마크좌표정보와 거리각도 변수의 총변동 중 약 95.16%를 설명하여 매우 효율적인 자료모형이라고 할 수 있다. 변수 100개를 사용하는 자료4-SVM모형의 정확도가 약 1.7% 더 높은데도 불구하고 자료6-SVM모형을 선택한 이유는, 모형의 간결성과 적합·예측 속도 면에서 변수를 적게 사용하는 자료6-SVM모형이 우수하기 때문이다. 실제로 이미지 표정 인식을 적용하는 분야에서는 연속적인 프레임에서 빠르게 표정을 식별하는 것이 중요하므로, 데이터셋을 구성하고 예측값을 계산하는 데 있어서 모형의 간결성은 매우 큰 장점이라고 할 수 있다.

최종 모형인 자료6-SVM모형을 이용하여 10-fold CV를 1회 실행하였을 때, 그 예측값과 실제값을 비교한 confusion matrix는 Table 4.2와 같다.

본 연구에서는 최종 선택 모형인 자료6-SVM모형을 이용하여 이미지에서 여러 개의 얼굴을 동시에 탐지하고 표정을 분류하는 Python 알고리즘으로 구현하였다. 구현된 알고리즘은 https://github.com/sunsmiling/facial-emotion-detector에서 다운로드 및 사용 가능하다.

4.3. 잘 분류되는 사진과 오분류되는 사진의 특징

이번 절에서는 최종 모형으로 선택된 자료6-SVM 결과로부터 잘 분류되는 사진과 잘 분류되지 않는 사진의 특징을 분석하고자 한다. 다음의 내용에서 잘 분류되는 사진이란 예측값과 실제 값이 일치하며, 예

Table 4.2. Final model: confusion matrix

True	Predicted				정확도
	Angry	Happy	Sad	Neutral	
Angry	666	89	158	137	0.721
Happy	209	3722	161	180	
Sad	117	42	393	165	
Neutral	458	247	667	2053	

Table 4.3. Characteristics of anger facial expression

	Angry로 분류	Happy로 분류	Neutral로 분류
눈썹	눈썹과 눈 사이 거리가 가까움	눈썹이 완만한 모양	눈썹이 완만한 모양
미간	미간 사이가 좁음	미간 사이가 좁지 않음	미간 사이가 좁지 않음
입	크게 벌린 모양	다소 벌린 모양	다문 모양

Figure 4.1. Eigenfaces of angry face images; (left) images classified as 'angry', (right) misclassified images.

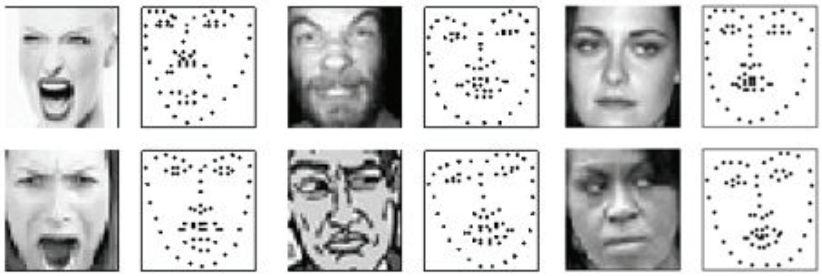

Figure 4.2. Landmark images of angry face images.

측확률(pseudo-probability)이 0.7 이상인 것으로 정의하였다. 한편, 잘 분류되지 않는 사진이란 예측값과 실제 값이 일치하지 않으며, 예측확률이 0.7 이상인 것으로 정의하였다. 예외적으로 실제 값이 슬픔이며 예측 값이 슬픔인 경우에는 예측확률이 0.7 이상인 경우가 매우 적어, 잘 분류되는 슬픔 사진이란 예측확률이 0.65 이상인 것으로 정의한다. 본 논문의 최종 모형은 행복(happy)와 중립(neutral)은 비교적 잘 분류하지만, 화남(angry)와 슬픔(sad)는 비교적 오분류율이 높은 특징을 보인다. 따라서 화남(angry) 이미지를 대표로 분석하여 잘 분류되는 이미지와 잘 분류되지 않는 이미지의 특징을 알아보고자 한다. 우선 각 감정에서 실제 값과 예측 값이 일치하는 이미지와 그렇지 않은 이미지를 구분하여 eigenface를 추출하고, 그 차이점을 비교하였다. 다음으로, 잘 분류되는 사진과 잘 분류되지 않는 사진의 예를 추출된 랜드마크와 함께 나타내어 랜드마크의 특징을 분석하였다 (Table 4.3).

실제 값이 화남(angry)인 이미지 중 다수는 행복(happy)와 중립(neutral)으로 분류되었다. Figure 4.1의 좌측 8개 이미지는 잘 분류된 화남(angry)사진의 eigenface를 나타내고, 우측 8개 이미지는 오

Table 5.1. Microsoft API: confusion matrix

True	Predicted					정확도	
	Anger	Happiness	Sadness	Neutral	Fail	Fail 포함	Fail제외
Anger	476	74	81	504	315	0.595	0.770
Happiness	8	2937	5	88	1062		
Sadness	38	20	373	689	259		
Neutral	3	118	50	1845	519		

분류된 화남(angry)사진의 eigenface를 나타낸다. Figure 4.2은 좌측부터 잘 분류된 화남(angry)이미지, 행복(happy)으로 분류된 이미지, 중립(neutral)으로 분류된 이미지의 원본과 랜드마크를 보여준다. Figure 4.2의 좌측에 위치한 이미지와 같이 잘 분류된 화남(angry)이미지의 랜드마크는 대부분 미간과 눈 사이가 가깝게 추정되었고, 입을 벌리고 있는 모습을 드러낸다. 반면 Figure 4.2의 가운데 위치한 이미지와 같이 행복으로 분류된 이미지에서는 입을 벌리고 있지만 눈썹과 눈 사이의 거리가 비교적 실제보다 멀게 추정되었으며, 눈썹의 모양이 사진과는 달리 둥글과 완만한 모양을 띠고 있다. 따라서, 입을 벌리고 있는 사진이라도 랜드마크에서 눈썹의 모양이 완만하게 추정되면 행복으로 분류될 가능성이 크다는 것을 알 수 있다. 마지막으로, Figure 4.2의 우측에 위치한 이미지와 같이 실제 값이 화남이지만 중립으로 분류된 사진은 대부분 입을 다물고 있으며, 눈썹에 해당하는 랜드마크의 모양이 둥글고 완만하게 표현되었다.

5. 결론 및 시사점

5.1. Microsoft Azure API 및 CNN과의 비교

5.1.1. Microsoft Azure API의 적용 Microsoft Azure(http://www.microsoft.com/azure/)에서 제공하는 Emotion API (Project Oxford, https://www.projectoxford.ai/emotion)은 1인 이상의 얼굴이 담긴 사진 혹은 비디오에서 얼굴 영역을 찾아내고, 표정을 감지하여 각 표정 별 점수로 나타낸다. Azure Emotion API가 감지하는 표정 범주는 총 8개이며, 본 논문에서 연구한 Angry, Happy, Sad, Neutral의 범주가 포함되어 있다. 우리는 Kaggle 자료에 이 API를 적용하여 4.2절의 최종 모형 결과와 비교하고자 한다. 현재 https://westus.api.cognitive.microsoft.com/emotion/v1.0/recognize에서 Microsoft Azure emotion API의 무료 평가판 및 유료 서비스 구입이 가능하다.

본 논문에서 이용한 방법과 Microsoft API는 모두 전체 데이터셋 중 일부에서 얼굴을 탐지하지 못한다. 또한 본 논문에서 이용한 방법으로 탐지하지 못하는 이미지 중 Microsoft Emotion API는 탐지 가능한 이미지가 있으며, 반대로 Microsoft Emotion API는 탐지하지 못하지만 본 논문에서 이용한 방법으로는 탐지 가능한 이미지가 있다. 따라서, 공정한 비교를 위해 두 방법 모두에서 탐지 가능한 이미지만을 이용하여 정확도를 비교하고자 한다. Microsoft API의 분류 결과는 API에서 출력하는 8개의 감정 점수 중 본 논문에서 사용한 감정 범주인 Angry, Happy, Sad, Neutral의 네 가지 범주의 점수만을 비교하여 지정하였다. Table 5.1은 Table 2.1의 최종 데이터셋에 Microsoft Azure Emotion API를 적용한 결과이다. 비교 결과 두 방법 모두에서 탐지 가능한 이미지 중 Microsoft Azure Emotion API의 분류 정확도는 0.770이며, 본 논문의 최종 모형(자료5-SVM)의 분류 정확도는 0.719이다. 분류 정확도가 Table 4.1보다 낮아진 이유는 본 논문의 방법으로 탐지 가능한 데이터 중 Microsoft API가 탐지하지 못하는 데이터의 상당수가 Happy에 속하는 이미지이기 때문이다. 최종 모형(자료5-SVM)에서는 다른 라벨에 비해 Happy 이미지의 분류 정확도가 훨씬 높으므로 분류 정확도가 Table 4.1에 비해 낮아짐이 자명하다.

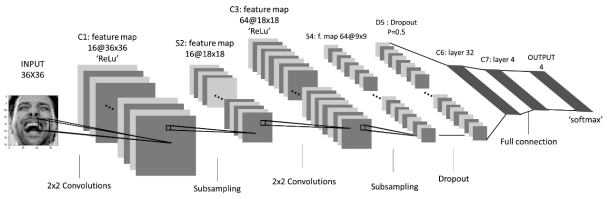

Figure 5.1. Structure of CNN.

Table 5.2. CNN: confusion matrix

True	Predicted				정확도
	Angry	Happy	Sad	Neutral	
Angry	563	285	165	437	0.647
Happy	115	3524	109	350	
Sad	199	288	375	516	
Neutral	206	409	262	1656	

5.1.2. Convolutional neural network의 적용 이 절에서는 픽셀 정보만을 이용(Table 4.1의 자료1 참고)하여 컨볼루션 신경망(convolutional neural network, CNN)을 적용한 결과를 소개하고, 앞서 구한 최종 모형과 비교하였다. CNN은 2차원 구조의 입력 데이터를 처리하는 데 적절한 특수한 형태의 특징 추출기가 신경망에 포함되어 있는 형태로, 최근 영상 및 이미지 인식분야에서 활발히 사용되고 있는 딥러닝 기법 중 하나이다. CNN은 일반적으로 컨볼루션 연산을 하는 합성곱 계층(convolution layer)과 집단화 작업을 하는 풀링(pooling)계층, 과적합을 방지하기 위한 드롭아웃(drop out)계층 등으로 이루어진다. 이 연구에서는 입력 계층 이후 총 7개의 계층으로 이루어진 CNN 구조를 사용 5.1하였으며, 그 과정은 다음과 같다. 첫 번째 계층은 2×2픽셀 크기의 16개 필터로 이루어진 합성곱 계층(convolution layer, C1)으로, zero-padding을 주어 합성곱 계층을 거치면 36×36픽셀 크기의 16개 특징맵(feature map)이 만들어진다. 특징맵은 ReLu 액티베이션 함수를 거친 뒤 풀링 계층(subsampling layer, S2)에서 max pooling에 의해 절반 크기로 서브 샘플링 되었다. 입력값은 이렇게 서브 샘플링 된 16개의 특징맵을 거쳐 다음 계층으로 전달된다. 다음 단계에서는 다시 2×2 크기의 합성곱 계층, ReLu, 풀링 계층을 거친 후 네트워크 과적합이 발생하지 않도록 드롭 아웃 계층(D5)을 거치도록 하였다. 마지막으로, 위 과정에서 만들어진 모든 특징맵을 이용해서 완전히 연결된 32개, 4개 길이의 계층과 softmax 계층을 거쳐 4개의 output(4가지 표정) 분류 학습이 이루어진다.

Table 5.2에서 제시한 CNN 모델링 결과는 Tensorflow1.5.0 버전과 Keras2.1.3 버전 (Chollet, https://pypi.python.org/pypi/Keras/2.1.3, 2015)을 이용한 것이며, Intel i7-4790 CPU(3.60GHz), RAM 16GB, Windows 7의 환경 하에서 수행되었다. 딥러닝을 위한 오픈 소스 라이브러리인 Keras는 다른 딥러닝 라이브러리인 Theano와 Tensorflow가 내부적으로 구동되는 형태(back-end)로 사용되고 있으며, 모듈화를 직관적으로 구현할 수 있고 python API를 지원한다는 것이 주된 장점이다.

Table 5.2의 정확도 값은 10-fold CV에 의해 계산되었다. Test data에서의 평균 정확도는 0.647으로, 픽셀정보만으로 이루어진 Table 4.1-자료1의 결과에 비하여 높지만 최종 모형인 자료 6-SVM의 성능보다는 낮음을 확인하였다.

5.1.3. 최종 모형(자료6-SVM모형)과의 비교 이상에서 제시한 결과를 각 표정 별 정분류율의 관점에서 본문의 최종 모형과 비교하였다. 아래의 Table 5.3는 최종 모형과 Azure API, CNN 기법 적용

Table 5.3. Comparison of accurracy between final model, Microsoft API (Azure) and CNN

	Emotions			
	Anger	Happiness	Sadness	Neutral
Final model(자료6-SVM)	63.43%	87.13%	54.81%	59.94%
Azure (fail 포함)	32.83%	71.63%	27.05%	72.78%
Azure (fail 제외)	41.94%	96.68%	33.30%	91.52%
CNN	38.83%	85.99%	27.21%	65.38%

CNN = Convolutional neural network.

시의 분류결과에서 각 표정별로 사진이 잘 분류된 비율을 나타낸 것이다.

Azure가 탐지에 성공한 사례(fail 제외)를 기준으로 각 표정별 분류성능을 비교하였을 때, 본 연구의 최종 모형은 Anger와 Sadness 부문의 사진을 더 잘 분류하였다. 한편 Azure는 이들을 제외한 Happiness와 Neutral 부문의 사진을 90% 이상의 정확도로 분류해내었다. 그러나 Azure가 탐지에 실패한 사례(fail 포함)를 사실상 표정 분류에 실패한 것으로 가정하여 정분류율을 계산했을 때, Neutral을 제외한 모든 표정 부문에서 본문의 최종 모형보다 정분류율이 낮게 나타난다. 따라서 실질적으로 본 연구의 최종 모형이 Azure API에 비해 Anger나 Sadness에 해당하는 표정 분류 능력이 더 뛰어나며, Azure API는 일단 탐지에 성공할 경우 Happiness와 Neutral에 해당하는 사진들을 더 잘 분류한다고 할 수 있다. CNN기법을 적용한 결과는 무표정(neutral)에서만 최종 모형보다 나은 분류정확도를 보였고, 나머지 표정에서는 정확도가 낮은 것으로 나타난다.

5.2. 결론

이상으로, 최종 모형 검토 및 오분류되는 사진에 대한 탐구를 통하여 본 연구에서 사용한 여러 방법론의 장·단점 및 한계를 알아보았다. Face landmark를 이용하여 표정을 인식하는 기법은 주입되는 이미지의 크기와 모양에 관계없이 사람의 얼굴을 식별할 수 있으며, 식별한 landmark의 위치정보를 이용하여 거리, 각도와 같은 다양한 변수를 생성하고 활용할 수 있다는 장점이 있다. 또한 성별과 인종, 연령에 관계없이 적용 가능하며 실제 사람의 얼굴이 아닌 애니메이션 캐릭터의 표정도 인식해내었다. 그러나 본 연구에서는 정면 얼굴만을 기준으로 하여 얼굴을 식별하였기 때문에 다수 이미지에서 랜드마크를 추정할 수 없다는 한계점이 있다. 더불어, 해상도가 매우 낮거나 사람의 생김새에 따라 눈썹 등의 특정 부위를 식별하기 어려운 경우 랜드마크를 추정하더라도 분류가 정확하지 않을 수 있다.

여러 데이터마이닝 방법들과 Microsoft Azure API, CNN 적용 결과를 비교하였을 때, 최종 모형은 30개의 주성분 정보만을 이용하여 빠른 시간 내에 비교적 높은 정확도를 보여주었다. 또한, Microsoft Azure API나 CNN은 행복(happy) 및 무표정(neutral)만을 잘 분류한 반면 최종 모형은 화남(anger)이나 슬픔(sadness) 표정에서 다른 기법들에 비해 훨씬 정확히 분류한다는 것을 확인하였다.

최종 모형에서 선택된 SVM방법은 본 연구에서 실험한 데이터마이닝 기법 중 가장 정확도가 높다는 장점이 있지만 LDA에 비해서는 더 긴 모형적합시간을 요한다. LDA는 SVM 다음으로 높은 정확도를 보이며 다른 방법에 비해 월등히 적은 시간을 소모하므로 빠른 모형적합과 예측의 면에서 유용하다. 따라서, 추후 연구에서 랜드마크 이외에 추가적인 정보를 도입할 시 design matrix의 크기가 커진다면 SVM과 함께 LDA를 사용할 것을 적극 고려해 볼 만 하다.

References

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*, O'Reilly Media, Sebastopol.

Brieman, L. (1996). Bagging predictors, *Machine Learning*, **24**, 123–140 .

Breiman, L. (2001). Random forests, *Machine Learning* , **45**, 5–32.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**, 681–685 .

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application, *Computer Vision and Image Understanding*, **61**, 38–59.

Cortes, C. and Vapnik, V. (1995). Support-vector networks, *Machine Learning*, **20**, 273–297 .

Daugman, J. G. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36**, 1169–1179.

Ekman, P. (1992). An argument for basic emotions, *Cognition & Emotion*, **6**, 169–200.

Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminator, *Biological Cybernetics*, **61**, 103–113 .

Goodfellow, I. J., Erhan, D., Carrier, P. L., *et al.* (2013). Challenges in representation learning: a report on three machine learning contests, arXiv:1307.0414, from: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data

Jain, A. K. and Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters, *Pattern Recognition*, **24** , 1167–1186.

Kazemi, V. and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1867–1874.

King, D. E. (2009). Dlib-ml: a machine learning toolkit, *Journal of Machine Learning Research*, **10**, 1755–1758 .

Liaw, A. and Wiener, M. (2002). Classification and Regression by randomforest, *R News*, **2**, 18–22.

Lyons, M. J., Budynek, J., and Akamatsu, S. (1999). Automatic classification of single facial images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**, 1357–1362.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., and Lin, C. C. (2017). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, R package version 1.6-8. from: https://CRAN.R-project.org/package=e1071

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 971–987.

Paul van Gent (2016). http://www.paulvangent.com/2016/08/05/emotion-recognition-using-facial-landmarks/

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, from: https://www.R-project.org/

Saatci, Y. and Town, C. (2006). Cascaded classification of gender and facial expression using active appearance models, In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, IEEE, 393–398.

Saragih, J. M., Lucey, S., and Cohn, J. F. (2009). Face alignment through subspace constrained mean-shifts, in *Proceedings / IEEE International Conference on Computer Vision*, 1034–1041.

Shan, C., Gong, S., and McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: a comprehensive study, *Image and Vision Computing*, **27**, 803–816.

Steven Puttemans (2013). https://github.com/opencv/opencv/tree/master/data/Haarcascades.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed), Springer, New York.

Viola, P. and Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai.

Viola, P. and Jones, M. J. (2004). Robust real-time face detection, *International Journal of Computer Vision*, **57**, 137–154.

Weihs, C., Ligges, U., Luebke, K., and Raabe, N. (2005). klaR analyzing German business cycles, In *Data Analysis and Decision Support*, 335–343, Springer, Heidelberg.

표정 분류 연구

손나영[a] · 조현선[a] · 이소현[a] · 송종우[a,1]

[a]이화여자대학교 통계학과

(2017년 11월 2일 접수, 2018년 8월 31일 수정, 2018년 9월 4일 채택)

요 약

최근 등장하는 다양한 사물인터넷 기기 혹은 상황인식 기반의 인공지능에서는 사용자와 기기의 상호작용이 중요시 된다. 특히 인간을 대상으로 상황에 맞는 대응을 하기 위해서는 인간의 표정을 실시간으로 인식하여 빠르고 정확한 판단을 내리는 것이 필요하다. 따라서, 보다 빠르고 정확하게 표정을 인식하는 시스템을 구축하기 위해 얼굴 이미지 분석에 대한 많은 연구들이 선행되어 왔다. 본 연구에서는 웹사이트 Kaggle에서 제공한 48*48 8-bit grayscale 이미지 데이터셋을 사용하여 얼굴인식과 표정분류로 구분된 두 단계를 거치는 얼굴표정 자동 인식 시스템을 구축하였고, 이를 기존의 연구와 비교하여 자료 및 방법론의 특징을 고찰하였다. 분석 결과, Face landmark 정보에 주성분분석을 적용하여 단 30개의 주성분만으로도 빠르고 효율적인 예측모형을 얻을 수 있음이 밝혀졌다. LDA, Random forest, SVM, Bagging 중 SVM방법을 적용했을 때 가장 높은 정확도를 보이며, LDA방법을 적용하는 경우는 SVM 다음으로 높은 정확도를 보이며, 매우 빠르게 적합하고 예측하는 것이 가능하다.

주요용어: 이미지 분류, Haar cascade, 페이스 랜드마크, 데이터 마이닝

이 논문 또는 저서는 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017R1D1A1B03036078).

[1]교신저자: (03760) 서울시 서대문구 대현동 이화여대길 52, 이화여자대학교 통계학과.
E-mail: josong@ewha.ac.kr