

기계학습을 이용한 회화 감성 예측 모델에 관한 분석 연구

A Study on the Emotional Prediction Model from Paintings by Machine Learning

이태민¹

Taemin Lee¹

요 약

이미지에서 감성을 예측하는 기술들은 많이 연구되어 지고 있다. 기계학습 및 딥러닝 기술들이 발전함에 따라서, 더 많은 연구들이 진행되었다. 이미지중에서도 특히 예술작품들은 감성과의 연관이 매우 크다. 일반적으로 예술가들이 자신의 감성을 작품에 넣는 경우가 많기 때문이다. 이런 감성들은 색상, 질감 등의 물리적 요소들이 결합된 대칭성, 구도 등의 예술적 요소들로 제어가 된다. 본 연구에서는 이런 특징들을 회화로부터 추출 및 분석한다. 회화에서 감성에 영향을 미칠 것으로 예상되는 특징들을 추출하여 이를 감성 예측에 활용한다. 주어진 회화로부터 색상, 선, 질감등의 물리적 특징과, 대칭성, 색상조합 등과 같은 예술적 특징을 추출하여, 다양한 기계학습 모델을 제작한다. 제작된 기계학습 모델들을 통해 회화-감성 예측에서 가장 관련이 깊은 특징들 및 감성 추출에 어울리는 기계학습 모델이 무엇인지 분석한다. 최종적으로 딥러닝 기반의 예측 모델과의 비교를 통해 기계학습 모델의 정당성 및 정확도에 대해 검증한다.

핵심어 : 회화 분석, 기계학습, 감성 컴퓨팅, 감성 추출, 감성 예측 모델

Abstract

Techniques for predicting emotions in images have been studied a lot. As machine learning and deep learning technologies developed, more studies were conducted. Among the images, artworks in particular are very related to emotions. In general, artists often put their emotions into their works. Emotions are controlled by artistic features such as symmetry and composition, which combine physical elements such as color and texture. In this study, these features are extracted and analyzed from paintings. Features that are expected to affect emotions in paintings are extracted and used to predict emotions. Various machine learning models are built by extracted physical features such as color, line, texture, etc. and artistic features such as symmetry and color combination from a given painting. Through the built machine learning models, this paper analyze which machine learning models are suitable for the most relevant characteristics and emotional extraction in conversation-emotional predictions. Finally, we verify the legitimacy and accuracy of machine learning models by comparing them with predictive models based on deep learning.

Keyword : Painting Analysis, Machine Learning, Affective Computing, Emotion Extraction, Emotion Predicting Model

¹ Davinci SoftWare Institute, ChungAng University, Seoul, Korea [Special Affair Professor]
e-mail: kevinlee@cglab.cau.ac.kr

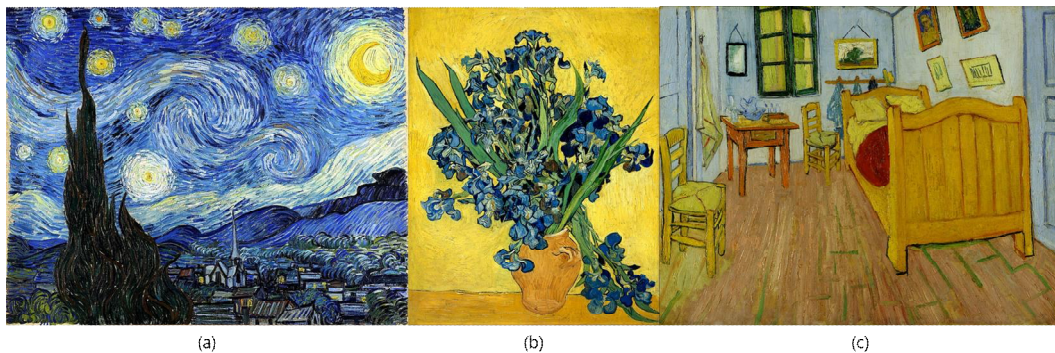
Received(July 31, 2021), Review Result(1st: August 18, 2021), Accepted(September 10, 2021), Published(September 30, 2021)



© 2021 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

1. 서론

미술 작품을 관찰할 때 사람들은 다양한 감정을 느낄 수 있다. [그림 1(a)]를 보면 대부분의 사람들이 감상하게 되면 격렬한 감정과 활동적인 느낌을 느낀다. 이는 비연속적이고 동적으로 꺾이는 붓 터치가 그런 감정들을 관찰자들에게 제공한다. 또한 어두운 색상의 객체와 밝은 색상의 객체가 대비되는 효과 등을 통해서 관찰자들에게 동적인 느낌과 우울한 느낌을 동시에 제공한다. 고흐는 이렇게 붓터치, 선, 구도, 색상, 색상의 대비 등의 다양한 특징들을 사용하여, 자신의 감정을 표현하고자 하였다 [1]. 이 그림은 다양한 예술적 특징들이 선명하고, 그 당시의 고흐의 감정들을 그대로 담고 있다는 평가들이 많아서 다양한 인공지능 연구에서도 사용되어 진다 [2][3]. 이렇게 감정이 쉽게 눈이 띄는 그림들이 있는 반면에, [그림 1(b)]의 그림처럼 이유가 쉽게 분석이 되지 않지만 감정이 느껴지는 그림과 [그림 1(c)]와 같이 감정도 쉽게 느껴지지 않고 분석도 어려운 그림들도 존재한다. [그림 1]은 모두 빈센트 반 고흐의 이미지이고, 화풍 스타일이 서로 비슷한 부분들이 존재하지만, 고흐의 그림을 그리는 정신 상태에 따라서 그림에서 느껴지는 감정이나 분석이 가능한 특징들이 달라진 것을 관찰할 수 있다. 즉, 그의 배경을 알고 있어야 분석이 쉬워지고, 단순히 그림만을 관찰하게 되었을 때 어려운 부분들이 존재한다.



[그림 1] 빈센트 반 고흐의 그림 (a : “별이 빛나는 밤”, b : “붓꽃”, c : “아를의 빈 고흐의 방”)

[Fig. 1] Vincent Van Gogh's paintings (a : “The Starry Night”, b : “Iris”, c : “Vincent's Bedroom in Arles”)

최근 정보 통신이 발전하면서 우리는 유명한 화가들의 회화작품을 직접 미술관에 찾아가서 감상할 필요 없이 PC나 모바일 등으로 감상할 수 있게 되었다. 회화작품에 쉽게 접근하게 되면서 자신의 기분에 따른 회화를 감상하거나, 회화를 검색할 때 감성 키워드를 기반으로 검색을 하기 시작하였다. 이를 위해서 임의의 회화가 어떤 감성을 가지고 있는지, 혹은 관찰자가 회화를 감상할 때 어떤 기분을 느끼는지를 자동적으로 감상하는 연구들이 중요하게 연구되어졌다. 머신 러닝과

딥러닝 기술들이 발전하면서, 감성을 예측하는 기술들이 발전하였으나, 일반적으로 회화작품이 아닌 사진에서 감성을 예측하고자 하는 연구들이 많았다. 사진에서 보여지는 정보들은 회화에서 얻는 것보다 훨씬 습득에 용이하다. [그림 2]는 고흐의 그림과 실제 테라스 사진을 비교해 놓은 것이다. 이미지의 왼쪽 문 부분을 비교해 보면, 사진은 정확한 직선과 함께 색상이 하나로 일치하지만, 회화의 경우 다양한 색상을 사용하여 음영감을 제공한다. 또한 바닥 타일의 경우 일체감 없이 많은 곡선들을 사용함으로써 엣지를 추출하거나 영역을 나누는데 어려움이 존재할 수밖에 없다.



[그림 2] 빈센트 반 고흐의 회화와 실제 사진 비교 (좌 : 회화 “Cafe Terrace”, 1888, 우 : 실제 모습)

[Fig. 2] Comparing Vincent van Gogh's paintings with real photographs

본 연구에서는 주어진 회화에 대해서 감성공간[4]을 기반으로 감성과 연관이 될 수 있는 여러 가지 특징들을 추출하고, 이 값을 기반으로 감성을 예측하는 모델을 제작한다. 높은 성능의 예측 모델을 제작하기 위해서, 다양한 기계학습 모델을 기반으로 감성 예측 모델을 제작한다. 제작된 감성 예측 모델들을 서로 비교 분석하여, 어떤 특징들이 감성과 연관이 있고, 어떤 기계학습 기술이 감성 예측에 효과적인지를 분석한다. 또한 딥러닝을 통한 결과와 같이 비교하면서 기계학습 모델의 정확도에 대해서 분석한다.

본 연구의 진행은 다음과 같다. 2장에서 이미지에서 감성을 예측하는 관련 연구들과 본 연구에서 사용할 감성 공간에 대해서 설명한다. 감성 분석의 연구를 크게 사진에 대한 것과 회화에 대한 것으로 나누어 설명한다. 3장에서는 감성을 예측 모델을 제작하는 과정에 대해서 설명한다. 주어진 회화에 대해서 어떤 특징들을 감성을 예측하는 데에 사용하는지 정의하고 그들을 추출하는 방법에 대해서 서술한 후, 다양한 기계학습 분석을 진행하는 방법에 대해서 서술한다. 4장에서는 우리의 예측 모델에 대해서 비교 분석을 서술한다. 어떤 특징들이 감성예측 분석에 중요하고, 관련성에 대해서 설명하고, 최종적으로 딥러닝 모델과의 비교를 통해 정당성을 검증한다. 마지막으로, 5장에서는 연구의 성과에 대해서 설명하고 부족한 부분에 대한 향후 연구를 설명하며 결론을 맺는다.

2. 관련연구

2.1 사진 분석에 관한 연구

사진에서 정보를 얻고 그 것을 기반으로 분석하는 연구들은 많이 진행되었다. [5][6]은 감성을 예측하기 위해 분류법을 사용하였다. [5]은 디시전 트리를 이용하여, 같은 성격을 같은 사진들끼리 분류하였다. [6]은 여기서 더 발전하여 연구를 진행하였다. 먼저 감성을 크게 8가지로 분류하였고, 주어진 사진들의 색상과 질감 특징을 이용하여 분류된 8가지 감성과 매핑시켰다. 두 연구 모두 분류법을 통해서 감성을 예측했다는 장점이 있지만, 비연속적인 단어들로 정의를 하기 때문에 다른 연구들에 응용하기 어렵다는 단점이 존재한다. [7][8]은 분류법이 아닌 기계학습을 통해 감성을 예측하는 연구를 제안하였다. IAPS [9]라는 사진-A.V. 정보가 저장되어 있는 데이터베이스를 기반으로 학습하여 감성 예측 모델을 제작하였다. 이 연구들 역시 학습은 A.V.값을 기반으로 하지만 예측된 결과들이 감성 형용사로 나온다는 단점이 그대로 존재하였다.

[10]은 사진에서 A.V.값으로 수치화 된 데이터를 예측해 내는 연구를 진행하였다. [10]은 사진과 음악을 A.V. 감성 좌표계를 기반으로 매칭하는 연구를 진행하였다. 이를 위해서 사진에서 색상정보를 기반으로 A.V. 감성 값을 예측하는 모델을 제작하였다. 주어진 사진에 대해서 수치화한 연구이지만, 사용된 정보가 색상 정보뿐이기 때문에 높은 정확도를 기대하기는 어려웠다. 사진에서 나아가서 비디오 영상에서 감성을 예측하는 연구도 진행되었다. [11]은 비디오 영상과 음악을 감성 기반으로 싱크로 하는 연구를 진행하였다. 비디오를 하나의 스틸샷들로 분리한 후에 각각의 사진에서 특징들을 추출하고 이를 기반으로 A.V.값을 예측하였다. 사진들에서 감성을 추출하기 위해서 사용된 정보들은 사진에서 직접 추출할 수 있는 물리적 특징들이고, 이는 회화에 비해서 선명한 데이터를 갖는다. 이를 회화에 적용시키면 높은 정확도를 갖기는 어렵고 뽑은 물리적 정보들이 감성을 예측하는데에 완벽하지 않다. 따라서 회화의 경우에는 직접적으로 뽑아진 데이터와 이를 기반으로 가공한 데이터를 모두 사용하여 감성을 예측하는 데 사용한다.

2.2 회화 분석에 관한 연구

예술작품이나 회화에 대한 분석들은 초창기에는 장르를 분석하는 연구들이 많이 진행되었다 [12-14]. [12]은 회화작품들을 사조별로 분류하는 방법을 제안하였다. 그라디언트, 어두운 색 사용 비율, 밝기 히스토그램의 범위 등 6가지 정보값을 정의하고 이 특징들을 기반으로 k-최근접 이웃 알고리즘(k-NN classification), 베이지안 분류법(Bayesian parametric classifier), 서포트 벡터머신(Support Vector Machines: SVMs)을 이용하여 회화를 3가지의 장르로 분류하였다. [13]는 가버 필토

를 통해 회색 정보(grayscale), 엣지 정보, 색상 정보등을 통해서 5가지의 장르로 분류하였다. 이런 장르들은 시각적 정보를 기반으로 쉽게 분류가 가능하기 때문에 70% 이상의 분류 정확도로 예측이 가능하였다.

[15]는 예술작품에서 감성을 추출하는 연구를 진행하였다. Lab색상 모델과 SIFT(Scale Invariant Feature Transform) 특징을 기반으로 예술작품에서 감성을 예측하였다. [15]에서는 회화작품보다는 추상화를 위주로 분석을 진행하였다. 또한 Valence 축 기준으로만 분류를 하였기 때문에 A.V. 좌표계에 직접적으로 응용하기 어렵다.

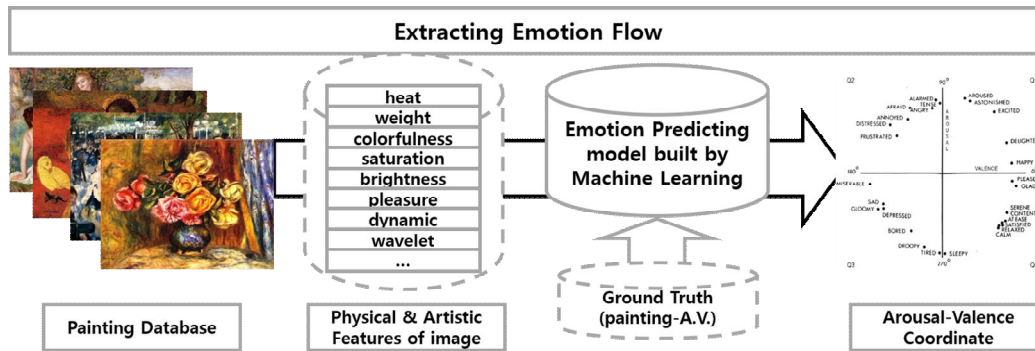
회화에서 감성을 추출하는 연구는 색상을 중심으로 연구되어 졌다 [16][17]. [16]은 Munsell 색상 좌표계를 130개로 분류하고 각 색상을 자신이 만든 좌표계에 매핑시켜 감성을 표현하였다. 또한 색상끼리의 조합을 통해서 생기는 감성에 대해서도 연구를 진행하고 자신이 만든 감성 공간을 컬러 이미지 스케일이라고 정의하였다. [16]의 감성 공간을 이용하여 [17]에서는 색상과 감성 단어들과의 관계에 대해 연구를 진행하였다.

[16]의 감성 좌표계를 값에 다른 특징들을 이용하여 [18][19]의 연구들이 진행되었다. [18]은 컬러 이미지 스케일상에 존재하는 3색 배색을 찾아 진행하였다. 비율과 유사도를 기준으로 가장 많은 3색 배색을 찾은 후에 가장 어울리는 감성 형용사를 찾아 매핑하였다. 나아가서 [19]의 연구는 음악과 매칭을 위해서 주어진 회화를 A.V. 감성좌표계에 매핑 시키는 연구를 하였다. [20]은 [18][19]에서 사용되는 특징들에 더 많은 특징들을 이용하여 회화로부터 감성을 예측하였는데, 단순히 선형 회귀 분석만을 이용하여 감성을 예측하고 각 특징들간의 연관성에 대해서는 분석하지 않았다.

본 연구에서는 주어진 회화에서 다양한 특징들을 추출하고 추출된 특징들간의 상관관계에 대해서 분석하고, 감성 예측에 어울리는 특징들이 무엇인지, 감성을 예측하기 위해서 유용한 기계학습 모델이 무엇인지에 대해서 연구한다.

3. 시스템

회화에서 감성 값을 예측하기 위한 본 연구는 두 단계로 나눌 수 있다. 하나는 특징 추출이고, 다른 하나는 특징 분석을 통한 학습 단계이다. 특징 추출 단계에서는 회화와 감성과 연관이 될 수 있는 특징들을 선별하고 추출한다. 색상과 질감(texture)와 같은 물리적 특징과, 색상 조합, 대칭성 등과 같은 예술적 특징들을 정의하고 추출한다. 학습 단계에서는 분석에 사용되는 회화-A.V.데이터 베이스셋 구축과 학습 모델 제작으로 나누어진다. 사용자 조사를 통해 A.V.값을 수집 받은 정보를 기반으로 회화의 정답 A.V.값을 구축하고 이를 기반으로 A.V.값을 예측하는 모델을 제작한다. [그림 3]는 본 연구의 시스템 개요를 도식화 한 것이다.



[그림 3] 시스템 개요도

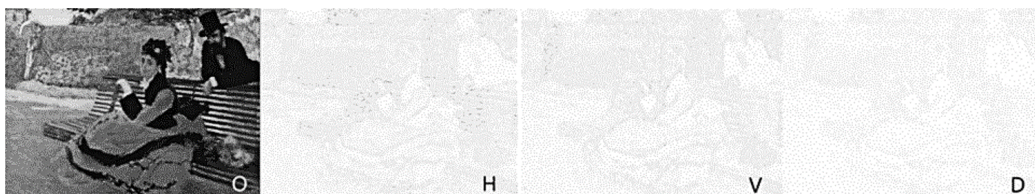
[Fig. 3] System overview

3.1 특징 추출

3.1.1 물리적 특징

회화에서 추출할 수 있는 특징들은 크게 2가지로 나눌 수 있다. 이미지에서 바로 추출할 수 있는 저차원적 특징(물리적 특징)과 이 특징들을 조합하여 만들어내는 고차원적 특징(예술적 특징)이 있다. 회화에서 색상은 중요한 특징중 하나이다. 화가들이 회화에서 감성을 표현하기 위해서 색상(Hue)을 사용하고, 채도를 낮추거나 밝기를 올리는 식의 표현을 이용하여 회화에서 감성 표현을 가능하도록 하였다. 본 연구에서는 주어진 이미지의 색상을 HSV값의 평균을 계산하여 감성 추출에 사용하였다.

회화의 다른 시각적 정보로 질감이 있다. 질감은 화가가 사용하는 붓의 형태나 스타일에 따라 서로 다른 그림으로부터 생겨나는 속성이다. 화가의 붓터치가 부드러우면, 그에서 느껴지는 감정은 차분하거나 따듯하게 느껴지고, 두껍고 딱딱 끊겨지는 붓터치를 사용한다면 회화로부터 역동적이거나 거친 느낌이 느껴진다. 본 연구에서는 [21]에서 사용하는 wavelet을 사용한다. [그림 4]는 주어진 회화로부터 Wavelet을 출력한 결과 예시이다. 추출된 정보에서 Horizontal(H), Vertical(V), Diagonal(D) 값을 사용하였다. 각 방향에서의 저주파, 고주파, 평균 값을 계산하여 사용하였다.



[그림 4] 회화로부터 추출한 질감 정보

[Fig. 4] Texture information from paintings

회화의 세 번째 물리적 특징으로는 선이 있다. 선의 방향성에 따라서 구도를 특정할 수 있기 때문이다. 선들의 방향이 일정하지 않거나 대각방향의 선들이 주로 이루어진 그림에서 혼란, 혹은 동적인 느낌을 얻을 수 있다. 반대로 수평적인 선들이 많은 경우에는 안정적이고 차분한 느낌, 수직 선이 많은 경우에는 무거운 느낌을 그림에서 얻을 수 있다. 우리는 회화를 구성하고 있는 선들의 방향성을 추출하기 위해 Progressive Probabilistic Hough Transform[22](PPHT)를 수정하여 사용하였고 추출된 선들의 경사도를 정의하여 수평, 수직, 그리고 대각선 강도를 계산하였다. 이후 검출된 직선들을 방향이 0~30도는 수직선, 30~60도는 대각선, 그리고 60~90도는 수평선으로 정의하고, 이를 수치화하였다. 수식 (1)을 통해 각 직선을 이루고 있는 픽셀 강도의 총합, Line Score를 특징값으로 계산하였다. k 는 세 가지 방향성을 의미한다. 이미지 전체에서 해당 픽셀의 변화량 정도를 강도로 정의하기 위해, 전체 변화량의 평균(μ_{total})과 표준편차(σ_{total}^2)를 기반으로 한 누적분포함수 값을 픽셀의 강도(PixelScore)로 정의하였다. 각 PixelScore의 값의 합은 LineScore가 되고 주어진 회화의 모든 선의 강도의 합을 MaxRadius로 정규화함으로써 각 선의 강도를 정의하게 된다. 이는 이미지의 변화량 평균과 유사한 변화량을 갖는 라인은 낮은 값을 갖게 되고, 다르게 되면 높은 값을 갖게 되는 것을 의미한다.

$$S_k = \frac{\sum LineScore_k}{MaxRadius} \quad (k = V, D, H) \quad (1)$$

$$LineScore_k = \sum PixelScore$$

$$MaxRadius = \sqrt{width^2 + height^2}$$

$$PixelScore_i = CDF(\mu_{total}, \sigma_{total}^2, PixelGradient_i)$$

3.1.2 예술적 특징

회화의 예술적 특징으로는 본 연구에서는 2가지를 사용하였다. 하나는 색상을 이용한 것이고, 다른 하나는 구도를 이용한 것이다. 색상을 이용한 예술적특징은 Ou의 색상정서 공간 [23]을 활용하였다. 회화의 모든 색상을 Ou의 정서공간에 매핑하고, 그 매핑된 공간의 좌표값을 특징으로 사용하였다. 색상을 이용하여 정서공간에 매핑한 방법은 수식 (2)와 같다.

$$color\ activity = -2.1 + 0.06[(L^* - 50)^2 + (a^* - 3)^2 + (\frac{b^* - 17}{1.4})^2]^{\frac{1}{2}} \quad (2)$$

$$color\ weight = -1.8 + 0.04(100 - L^*) + 0.45\cos(h - 100^\circ)$$

$$color\ heat = -0.5 + 0.02(C^*)^{\frac{1}{2}}\cos(h - 50^\circ)$$

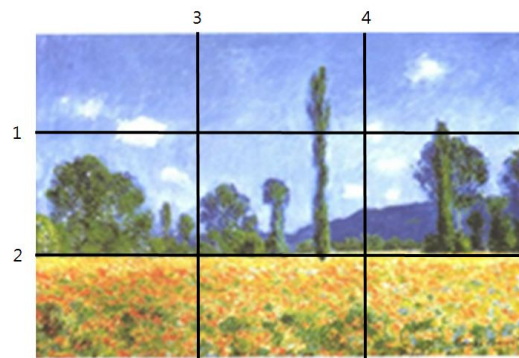
대칭성은 어떤 기준을 중심으로 얼마나 같은지를 의미한다. 회화에서 이 대칭성은 시각적으로 비슷한 물체들 혹은 색들이 배치되는가를 의미한다. 대칭성이 클수록, 관찰자들은 회화에서 안정감

을 느낄 수 있다. 반대로 대칭성이 적을수록 복잡한 느낌을 받거나, 불안정한 느낌을 받을 수 있다. 본 연구에서는 엔트로피를 기반으로 대칭성을 계산하여 회화의 한 특징으로 사용하였다. M. Endres et al. [24]의 Jensen-Shannon Diversity를 이용하였다. Entropy가 유사할수록 대칭성이 높은 것이고, 값이 낮을수록 대칭성이 낮은 것을 의미한다. 수식 (3)의 $M(n)$ 값은 구조적 대칭성의 계산법을 보여준다. $H(x)$ 는 각 구역의 히스토그램을 색상의 확률에 따라 엔트로피를 구한 것이다. 그리고 각 구역의 엔트로피 값과 그 구역의 비율(Π)의 곱의 합을 통해 $M(n)$ 을 구함으로써, 두 구역의 엔트로피 값이 유사할수록, $M(n)$ 값은 커지게 된다.

$$M_j(n) = \frac{\sum_{i=1}^n \Pi_i H(p_i)}{H_p} \quad (3)$$

$$H(x) = - \sum_{x \in X} p(x) \log p(x)$$

영상에서 위치(position)는 이미지에서 느껴지는 의도와 주제를 강조할 수 있는 요소이다. 그런 위치를 강조할 수 있는 방법으로 Rule-of-Third가 있다. 이는 사진이나 디자인에서 가장 광범위하게 사용되고 있는 관찰자들에게 안정감을 줄 수 있는 기법중 하나이다. 우리는 이 기법을 회화의 특징으로 적용하여 안정감을 줄 수 있는 요소로써 정의하였다. Rule-of-Third는 [그림 5]에서처럼 주어진 이미지를 4개의 선으로 분할하여, 검정색 선에 물체가 걸쳐있게 하는 기법이다. 각 영역의 1/3 지점으로서, 이 부분에 물체가 있게 되면, 구도적으로 안정적인 그림이 된다. 따라서, 우리는 주어진 회화를 [20]의 연구에서 응용하여 [그림 5]와 같이 4개의 선을 기준으로 분해한 후에 각 부분의 구조적 유사성을 비교하였다. 구조적 유사성을 계산하는 방법은 수식 (3)을 이용하였다. 나뉜 부분들이 서로 다를수록, 두 공간이 잘 나누어져 있다는 의미이고, 이는 Rule-of-Third를 잘 따른다고 판단할 수 있기 때문이다.



[그림 5] Rule-of-Third의 개념

[Fig. 5] Concept of Rule-of-Third

3.2 기계학습

3.2.1 데이터 수집

기계 학습을 위한 회화 이미지로는 Kobayashi가 사용한 16장의 회화를 포함하여 총 65장의 회화를 사용하였다. 주로 인상파의 그림을 사용자 조사에 사용되었으며, 회화의 전체적인 느낌보다 물체에 영향을 많이 미칠 수 있는 고전주의 회화들은 배제하였다. 피카소의 그림이나 그림 자체가 난해한 것들도 몇 개 포함은 하였는데, 이는 결과의 비교하기 위해서 데이터를 수집하였다.

비전문가들의 회화 감상 평균을 구하기 때문에 비전문가들 50명에게 총 65장의 이미지를 보여주고 SD법을 활용하여 Arousal Valence 강도를 5단계로 평가하도록 하였다. 설문조사 확장이 가능하도록 영어로 질문을 진행하였고, 감성축의 강도를 이해하지 못하는 경우를 위해 사전적인 설명도 포함시켰으며, 완벽한 이해를 위해서 구두적 설명 또한 포함 하여 진행하였다. 한국인 20대~30대의 사용자들에게 설문조사를 진행하였다. 주어진 65장의 이미지중 설문조사의 값의 편차가 너무 큰 이미지들을 제외하고 총 57장의 이미지에 대해서만 데이터베이스를 구축하였다.

3.2.2 실험 환경

AMD Ryzen 5 3600 6코어 프로세서, 16GB RAM 과 Nvidia GTX 1660 SUPER 6GB 그래픽 카드 환경에서 실험을 진행하였다. 특징 추출에 사용된 프로그래밍 언어는 C++이며, 실험을 위한 프로그래밍은 Python으로 수행하였다. 기계학습에 대한 프로그래밍은 Scikit-learn, Pycaret, Pytorch 등에서 수행되었다.

[표 1] 기계학습에 사용된 회화의 특징들

[Table 1] Features of paintings for machine learning

	특징 주제	특징 이름	특징 설명
물리적 특징	색상	colorfulness, saturation, brightness	이미지의 색상, 채도, 밝기의 평균
	질감	wavelet_{LH, LV, LD, AH, AV, AD, BH, BV, BD}	wavelet의 방향과 최대, 평균, 최소를 이용한 값
	선	{horizontal, vertical, dynamic}_intensity	엣지 정보의 방향성에 따른 값
예술적 특징	색상조합	heat, weight	색상의 heat/weight 값 [23]
	대칭성	{vertical, horizontal}_similarity	엔트로피 기반의 수평과 수직 대칭성
	구도	rule-of-third	엔트로피 기반의 삼단 구도

3.2.3 학습 모델을 통한 A.V.예측

표준 정규화 전처리가 이루어진 각 이미지와 [표 1]의 특징들을 레이블링하고, AutoML을 활용하여 감성을 예측하였다. AutoML을 활용하여 감성 분류 성능이 가장 좋은 모델을 찾은 후 하이퍼 파라미터 최적화 과정을 거쳤다. 평가 지표로 AU-ROC Score를 활용하여 각각의 감성 Label에 대한 분류 성능을 측정하고자 하였다. 총 학습 데이터의 70%를 Train Set, 나머지 30%를 Test Set으로 하여 10회의 교차 검증(K-Fold Cross Validation)을 거쳤다.

4. 실험결과 및 분석

일반적으로 많이 사용되는 기계학습 모델을 통해 감성 분류 모델을 만든 후 그 성능을 AUC 평가 지표를 기준으로 비교 분석하였다. 총 학습 데이터의 70%를 Train Set, 나머지 30%를 Test Set으로 하여 10회의 교차 검증(K-Fold Cross Validation)을 거쳤다. 본 연구에서 사용된 학습 모델은 [표 2]와 같다.

[표 2] 사용된 기계학습 모델들에 대한 설명

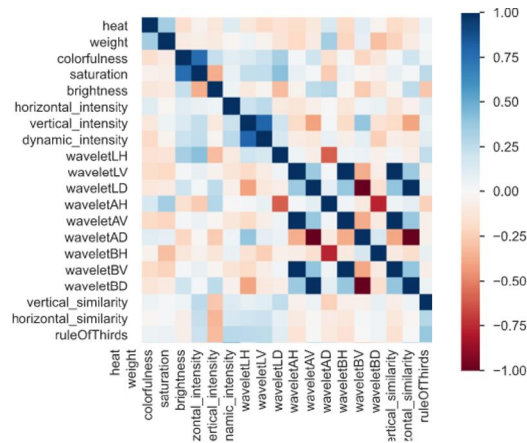
[Table 2] Explanation of used machine learning model

모델명	모델 설명
Gradient Boosting	순차적 tree 기반의 약한 학습기를 만들고 이 학습기의 잔차에 대해 가중치를 부여하는 방식으로 진행되는 Boosting 알고리즘을 적용
Random Forest	학습 데이터의 다차레 샘플링을 통해 여러개의 Decision Tree를 형성하고 이 분류기들의 투표를 통해 분류를 진행하는 앙상블 기법
Ada Boost	Boosting 알고리즘의 한 종류로, 가중치를 부여한 약한 분류기들을 모아 강한 분류기를 생성하는 기법
Bayesian Ridge	오버피팅이 발생할 가능성이 높은 MLE의 단점을 보완하기 위해 로그 한계 우도를 최대화하는 방향으로 규제 파라미터를 조정하는 릿지 회귀 방식
Linear Regression	주어진 한 개 이상의 독립변수와 종속변수 간의 관계를 선형 수식을 통해 분석하는 기법
Huber	L1과 L2의 장점을 합쳐 이상치에 강인하게 제작된 예측 모델, 델타값을 기준으로 에러의 크기에 따라 L1과 L2를 따로 사용함.
K-Neighbors	새로운 데이터가 주어졌을 때 기존 데이터 가운데 가장 가까운 k 개 이웃의 정보로 새로운 데이터를 예측하는 기법
Decision Tree	일련의 분류 규칙에 따라 데이터를 분류 혹은 회귀하는 트리 모델을 이용한 분석 기법

[그림 6]은 heat map을 이용하여 각 특징들간의 상관관계(correlation)를 보여준다. 그림에서 볼 수 있듯이, 특징들의 주제가 같은 경우에는 서로 연관성이 있지만, 일반적으로 서로 다른 특징들의 경

우에는 크게 상관관계가 없는 것을 볼 수 있다. 즉 우리가 사용한 특징들이 서로 간에 영향을 주지 않는 독립적인 특징들로서 사용될 수 있다는 것을 보여준다.

[그림 7]은 8개의 기계학습 모델들을 이용하여, 수집한 57개의 이미지를 분석한 결과이다. x축은 57개의 이미지를 순서대로 나열한 것이며, 선으로 각각을 연결하여 선형으로 보여줄 수 있도록 하였다. y축은 예측된/수집된 arousal값과 valence값을 의미한다. 수집된 값은 1~9까지이지만, 대부분의 값이 극값이 없기 때문에 2~8까지만 그래프에 표현되었다. 다양한 모델을 비교하기 위해서 에러율을 기반으로 평가를 하였다. [표 3]이 각 모델들을 비교할 수 있도록 정리한 에러율이다.



[그림 6] 각 특징들간의 상관 관계를 보여주는 히트맵

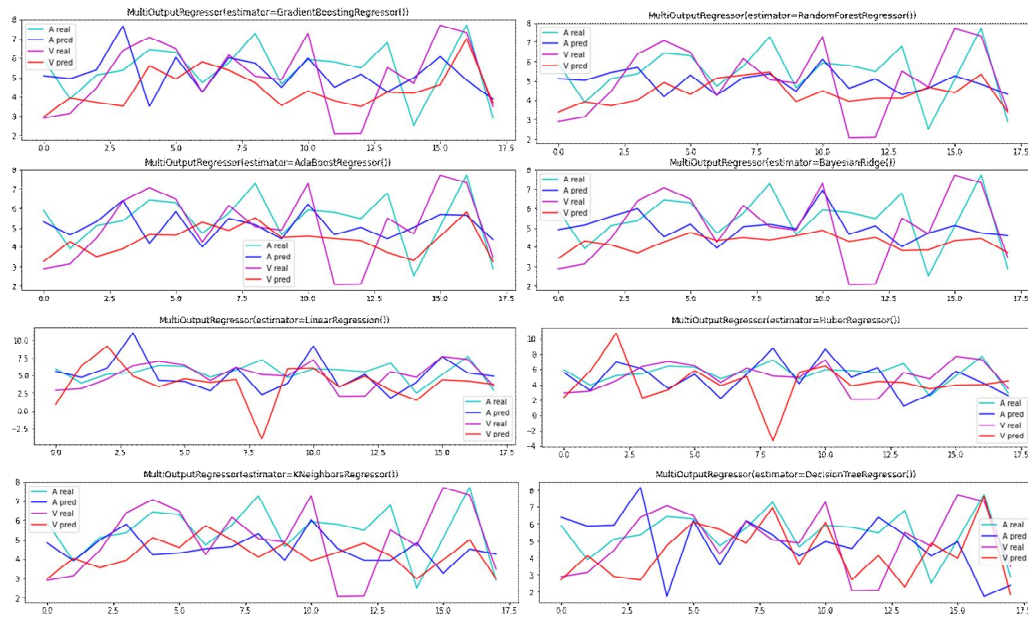
[Fig. 6] Heat map that shows the correlation between each features

[표 3] 기계 학습 모델들의 에러율 비교

[Table 3] Comparison error of each machine learning

모델	A MAE	V MAE	A MSE	V MSE	A RMSE	V RMSE	A R2	V R2
GBR	1.1965	1.2760	2.3751	2.4515	1.5411	1.5657	-0.3593	0.2108
RFR	1.0952	1.4010	1.9357	2.8215	1.3913	1.6797	-0.1078	0.0917
ADR	1.084	1.5396	1.8131	3.1069	1.3465	1.7626	-0.0375	-0.0002
BR	1.243	1.5581	2.231	3.4869	1.4947	1.8673	-0.2785	-0.1225
LR	2.2029	2.6134	7.3345	10.6349	2.7082	3.2611	-3.1976	-2.4237
HR	1.5092	2.4538	4.2251	10.5626	2.0555	3.25	-1.418	-2.4004
KNN	1.3461	1.6545	2.6946	3.7587	1.6415	1.9387	-0.5422	-0.21
DT	1.5321	1.5539	4.662	3.5914	2.1592	1.8651	-136681	-0.1562

GBR : Gradient Boosting, RFR : Random Forest, ADR : Ada Boost, BR : Bayesian Rdige, LR : Linear Regression, HR : Huber, KNN : K-Neighbors, DT : Decision Tree, A : Arousal, V : Valence, MAE : Mean Absolute Error, MSE : Mean Square Error, RMSE : Root Mean Square Error, R2 : R-squared, Coefficient of detemination



[그림 7] Arousal/Valence 예측 결과

[Fig. 7] Result of estimating Arousal and Valence

Linear Regression과 Huber Regression을 제외하고는 Arousal과 Valence의 MAE(Mean Absolute Error)가 1.0~1.5사이에 분포되어 있다. Arousal과 Valence의 값이 1부터 9까지의 범위를 가지고 있기 때문에 81~87.5% 정도의 정확도를 가지고 있다고 말할 수 있다. Linear Regression의 경우 다른 예측모델에 비해 가장 기본적인 선형 함수만을 이용하기 때문에 예측값에 많은 오류가 존재할 수밖에 없다. 특히 데이터를 정제하는 작업을 거치지 않았기 때문에, 단순한 선형 함수만으로 예측의 정확도를 높이는 것이 어려울 수밖에 없다. 정확도가 높은 모델로 Ada Boost Regression과 Random Forest를 대표적으로 볼 수 있다. 그 이유는 분류기들의 성능을 분산시켜 과적합현상을 감소시키고, 여러 개의 약한 분류기에 가중치를 줘서 계산하기 때문에 만들어진 분류기는 높은 정확도를 갖기 쉬운 것을 볼 수 있다. 약한 분류기들을 취합하는 부분에서 대표적인 Random Forest와 약한 분류기들을 연결하는 부스팅의 대표적인 Ada Boost Regression이 기존의 분류기보다 높은 정확도를 갖는 향상된 모델이라는 것을 알 수 있다.

5. 결론

본 논문은 회화로부터 감성에 영향을 미칠 수 있는 특징들을 추출하고 이 특징들을 이용하여 다양한 기계학습 모델을 통해 감성 예측 모델을 제작하고 분석하였다. 이를 토대로 추출한 특징들 중에서 감성에 영향을 미칠 수 있는 것들을 추리고, 여러 기계학습 모델중에서 감성 추출에 어울

리는 기계학습 모델을 분석 및 선별하였다. 특징들 간의 관계와, 특징과 감성간의 관계를 분석하였기 때문에, 앞으로 진행할 연구에서 어떤 특징들을 더 자세히 분석할 수 있는지에 대한 진행성을 찾을 수 있다. 학습의 결과에서 볼 수 있듯이 높은 정확도를 같은 예측 모델을 제작하였고, 이는 기존의 결과보다 향상되었고, 사진보다 분석이 어려운 회화에도 적합한 것을 알 수 있다.

본 연구에서는 제안한 특징들을 가지고 기계학습을 통해 감성을 예측하지만 약간의 한계를 가지고 있다. 기존의 데이터들을 기준으로 학습을 하는 모델이기 때문에 학습되는 데이터가 한쪽으로 치우치는 경우 다른 한쪽을 처리하지 못하는 경우가 생긴다. 이를 위해 데이터의 분포를 비교하며 넓게 분포되는 데이터셋을 수집하지만, 회화는 부정적인 축에서 극단적인 값을 갖기 힘들다. 또한 사진은 객체의 영향을 크게 받지만 회화는 주된 객체가 없는 이미지를 사용하기 때문에, 그 영향은 적다. 따라서 하나의 객체가 존재하는 회화이거나 중요도가 높은 객체가 있으면 그로 인하여 감성의 결과가 바뀌어야 한다. 현재 본 연구의 모델은 자동화를 통한 특징 추출을 주로 하고 있는데, 객체를 고려하기 위해서는 현재 태깅정보를 넣어주어야 한다. 자동적으로 회화에서 객체를 추출해 낼 수 있다면 더 높은 정확도를 갖는 예측 모델을 제작할 수 있을 것이다.

References

- [1] J. Greenberg, S. Jordan, Vincent van Gogh: Portrait of an artist, LibraryThing, 2001.
- [2] S. Jafapour, G. Polatkan, E. Brevdo, S.Hughes, A.Brasoveanu, I. Daubechies, "Stylistic analysis of paintings using wavelets and machine learning", 17th European Signal Processing Conference, August 24-28, 2009, Glasgow, UK.
- [3] W. R. Tan, C. S. Chan, H. E. Aguirre, K. Tanaka, "ArtGAN: Artwork synthesis with conditional categorical GANs", 2017 IEEE International Conference on Image Processing, September 17-20, 2017, Beijing, China, doi: 10.1109/ICIP.2017.8296985.
- [4] J. Russell, "Circumplex model of affect", Journal of Personality and Social Psychology, vol. 39, no. 6, December 1980, pp. 1161-1178, doi: 10.1037/h0077714.
- [5] R. Datta, D. Joshi, J. Li, J. Z. Wang, "Studying aesthetics in photographic images using a computational approach", 2006 European Conference on Computer Vision, May 7-13, 2006, Graz, Austria, pp. 288-301.
- [6] J. Machajdik, A. Hanbury, "Affective Image Classification using Features inspired by Psychology and ArtTheroy", International Conference on Multimedia, October 25-29, 2010, Firenze, Italy, pp. 92-93, doi: 10.1145/1873951.1873965.
- [7] D. Joshi, R. Datta, E. Fadorovskaya, Q. Luong, J. Wang, J. Luo, "Aesthetics and emotions in images", IEEE Signal Processing Magazine, vol. 28, Iss. 5, September 2011, pp. 94-115, doi: 10.1109/MSP.2011.941851.
- [8] S. Zhao, Y. Gau, X. Jiang, H. Yao, T.Chua, X. Sun, "Exploring principles of art features for image emotion recognition", MM '14: 2014 ACM Multimedia Conference, November 3-7, 2014, Orlando Florida, USA, pp. 47-56, doi: 10.1145/2647868.2654930.

- [9] P. Lang, M. Bradley, B. Cuthbert, "International affective picture system(IAPS) : Affective ratings of pictures and instruction manual", Technical report A-8, 2008.
- [10] S. Sasaki, T. Hirai, H. Ohya, S. Morishima, "Affective music recommendation system reflecting the mood of input image", 2013 International Conference on Culture and Computing, September 16-18, 2013, Kyoto, Japan, pp. 153-154, doi: 10.1109/CultureComputing.2013.42.
- [11] K. Shin, I. Lee, "Music synchronization with video using emotion similarity", IEEE International Conference on Big Data and Smart Computing, February 13-16, 2017, Jeju, Korea, pp. 47-50, doi: 10.1109/BIGCOMP.2017.7881714.
- [12] O. Icoğlu, B. Gunşel, S. Sariel, "Classification and indexing of paintings based on art movements", Signal Processing Conference 2004 12th European, September 6-10, 2004, Vienna, Austria, pp. 749-752.
- [13] J. Zujovic, L. Gandy, S. Friedman, B. Pardo, T. N. Pappas, "Classifying paintings by artistic genre: An analysis of features and classifiers", IEEE International Workshop on Multimedia Signal Processing, October 5-7, 2009, Rio de Janeiro, Brazil, pp. 1-5, doi: 10.1109/MMSP.2009.5293271.
- [14] L. Shamir, J. Tarakhorsky, "Computer Analysis of Art", ACM Journal on Computing and Cultural Heritage, vol. 5, no. 2, July 2012, pp. 1-11, doi:10.1145/2307723.2307726.
- [15] V. Yanulevskaya, J. Uijlings, E. Bruni, A. Sartori, E. Zamboni, F. Bacci, D. Melcher, N. Sebe, "In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings", 20th ACM international conference on Multimedia, October 29-November 2, 2012, Nara Japan, pp. 349-358 doi: 10.1145/2393347.2393399.
- [16] S. Kobayashi, "The aim and method of the color image scale", Color Research and Application, vol. 6, no. 2, Summer 1981, pp. 93-107, doi: 10.1002/col.5080060210.
- [17] S. Kobayashi, Color Image Scale, Kodensha International, 1992.
- [18] D. Kang, H. Shim, K. Yoon, "A method for extracting emotion using colors comprise the painting image", Multimedia Tools and Applications, vol. 7, no. 4, February 2018, pp. 4985-5002, doi: 10.1007/s11042-017-4667-0.
- [19] T. Lee, H. Lim, D. Kim, S. Hwang, K. Yoon, "System for matching paintings with music based on emotions", In Technical Brief of the Siggraph Asia, no. 31, November 2016, pp. 1-4, doi: 10.1145/3005358.3005366.
- [20] T. Lee, D. Kang, S. Seo, K. Yoon, "Emotion-Based Painting Image Display System", Intelligent Automation and Soft Computing, vol. 26, no. 1, 2020, pp. 181-192, doi: 10.31209/2019.100000139.
- [21] A. Samuel, "Some Studies in Machine Learning Using the Game of Checkers", IBM Journal of Research and Development, vol. 3, Iss. 3, July 1959, pp. 210-229, doi: 10.1147/rd.33.0210.
- [22] J. Matas, C. Galambos, J. Kittler, "Robust Detection of Lines Using the Progressive Probabilistic Hough Transform", Computer Vision and Image Understanding, vol. 78, no. 1, April 2000, pp. 119-137, doi: 10.1006/cviu.1999.0831.
- [23] L. Ou, M. Luo, A. Woodcock, A. Wright. "A study of colour emotion and colour preference. Part 1: Colour Emotions for Single Colours", Color Research and Application, vol. 29, no. 5, April 2004, pp. 381-389, doi: 10.1002/col.20010.
- [24] D. M. Endres, J. E. Schindelin, "A new metric for probability distributions", IEEE Transactions on Information Theory, vol. 49, no. 7, June 2003, pp. 1858-1860, doi: 10.1109/TIT.2003.813506.