

# 인스타그램 이미지와 텍스트 분석을 통한 사용자 감정 분류

## A User Sentiment Classification Using Instagram image and text Analysis

홍택은\*, 김정인\*\*, 신주현\*\*\*

(Taekeun Hong, Jeongin Kim, Juhyun Shin)

### 요약

최근 스마트폰과 태블릿 PC 등의 스마트 기기들의 발전으로 인해 SNS(Social Network Service) 사용자가 증가함에 따라 SNS 정보를 이용한 사용자 감정 분류 방법에 대한 기법들이 활발하게 연구되고 있다. 사용자 감정 분류는 SNS 게시글의 텍스트, 이미지 등을 이용하여 감정을 분류하는 것을 말한다. 본 논문에서는 텍스트에서 대표 형용사를 추출하고 이미지에서 Canny 알고리즘과 삼각함수를 이용해 대표 도형에 대한 값을 추출하여 사용자의 감정을 분류하는 방법을 제안한다. 텍스트에서 추출한 대표 형용사는 텍스트에서 추출한 형용사 중에 빈도수가 가장 높은 형용사로 선정하였으며, 영어 감정 어휘 사전인 SentiWordNet을 이용하여 긍정-부정의 수치를 측정했다. 이미지에서 추출되는 도형에서 삼각형, 사각형, 원 중에 추출되는 도형을 대표 도형으로 선정했으며, 대표 도형의 종류와 기울기에 따라 쾌-불쾌 수치를 측정하여 사용자의 감정을 분류했다. 최종적으로 Plutchik의 감정 바퀴를 긍정-부정과 쾌-불쾌의 수치를 나타내는 x축과 y축을 갖는 좌표평면으로 재정의하고 대표 형용사와 대표 도형의 값을 재정의한 Plutchik의 감정 바퀴의 좌표 평면에 나타내어 사용자의 감정 분류를 수행한다.

■ 중심어 : 소셜 네트워크 서비스; 인스타그램; 사용자 감정 분류; SentiWordNet.

### Abstract

According to increasing SNS users and developing smart devices like smart phone and tablet PC recently, many techniques to classify user emotions with social network information are researching briskly. The use emotion classification stands for distinguishing its emotion with text and images listed on his/her SNS. This paper suggests a method to classify user emotions through sampling a value of a representative figure on a trigonometrical function, a representative adjective on text, and a canny algorithm on images. The sampling representative adjective on text is selected as one of high frequency in the samplings and measured values of positive-negative by SentiWordNet. Figures sampled on images are selected as the representative in figures; triangle, quadrangle, and circle as well as classified user emotions by measuring pleasure-unpleased values as a type of figures and inclines. Finally, this is re-defined as x-y graph that represents pleasure-unpleased and positive-negative values with wheel of emotions by Plutchik. Also, we are anticipating for applying user-customized service through classifying user emotions on wheel of emotions by Plutchik that is redefined the representative adjectives and figures.

■ keywords : Social Network Service; Instagram, Emotions Classification of Users; SentiWordNet.

## I. 서론

최근 스마트폰과 태블릿 PC 등의 스마트 기기들의 발전으로 인해 SNS(Social Network Service) 사용자 또한 급증하였다. 이에 따라 SNS를 이용한 다양한 연구가 진행되고 있는데 그중 하나가 사용자 감정 분류이다. SNS 사용자들은 자신의 계정을 이용해 일상생활, 자신의 상태, 기분 등을 글이나 이미지, 동영상 등으로 게시하는데 이를 이용해서 감정 분류를 하고, 사용자의 감정에 맞는 다양한 서비스를 제공할 수 있다. 감정이란 어떤 현상이나 사건에 대하여 일어나는 마음이나 느끼는 기분을 뜻하

는 단어이다. 상황이나 환경 등의 외부 자극에 따라 사람마다 다른 감정을 느낄 수 있으므로, 사용자의 감정을 단순히 측정하기는 쉽지 않다. 본 논문에서는 SNS 게시글 내의 텍스트와 이미지를 사용하여 사용자의 감정을 측정하는 연구를 수행하고자 한다. 본 논문에서는 텍스트에서 대표 형용사를 추출한 후 영어 감정 어휘 사전인 SentiWordNet을 이용하여 긍정-부정의 수치로 나타내고, 이미지에서는 Canny 알고리즘을 이용해 대표 도형을 추출하여 도형의 종류를 식별하고 삼각함수를 이용해 도형의 기울기를 구하여 쾌-불쾌 수치를 측정한다. 사람의 8가지 기

\* 준회원, 조선대학교 소프트웨어융합공학과, \*\* 준회원, 조선대학교 컴퓨터공학과, \*\*\* 정회원, 조선대학교 제어계측로봇공학과

This research was financially supported by the Ministry of Trade, Industry and Energy(MOTIE) and Korea Institute for Advancement of Technology(KIAT) through the Promotion Regional Reading Industry.

접수일자 : 2015년 12월 01일

수정일자 : 2016년 01월 29일

게재확정일 : 2016년 02월 02일

교신저자 : 신주현 e-mail : jhshinkr@chosun.ac.kr

본 감정을 동심원 형태로 나타내는 Plutchik의 감정 바퀴를 긍정-부정의 축과 쾌-불쾌의 축을 가진 좌표평면으로 재정의하고 대표 형용사와 대표 도형의 값을 재정의한 Plutchik의 감정 바퀴의 좌표평면에 나타내어 사용자의 감정을 분류한다. 본 논문의 구성은 다음과 같다. 2장에서는 사용자 감정 분류와 관련된 기존 연구에 대하여 설명한다. 3장에서는 인스타그램의 사용자 게시글의 이미지 정보와 텍스트 정보를 이용하여 사용자의 감정을 분류하는 방법에 대해 기술하며, 4장에서는 실험한 결과를 보여준다. 마지막으로 5장에서는 결론 및 향후 연구를 제시하며 마무리한다.

## II. 관련연구

개인 취향에 따른 정보와 서비스를 제공하는 개인화 기술이 최근 높은 관심을 받고 있으며, 이를 위해 사용자의 요구사항 분석과 개인별 정보를 이용한 개인별 맞춤형 서비스 제공에 관한 연구가 활발히 진행되고 있다[1]. 개인별 맞춤형 서비스 제공에 필요한 개인별 정보는 사용자에게서 수집할 수 있는 자료들을 분석하여 사용하고 있으며, 그중 하나가 SNS 사용자의 게시글이다. SNS 사용자는 게시글을 통해 일상, 관심사, 성향 등을 나타내므로 SNS 게시글을 통해 SNS 사용자들의 자료를 수집할 수 있으며, 수집한 자료를 분석하여 개인별 맞춤형 서비스 제공에 필요한 개인별 정보를 추출할 수 있다. 추출한 다양한 개인별 정보 중에서 감정은 중요한 요소이며, 인간이 느끼는 다양한 감정을 토대로 개인별 맞춤형 서비스를 제공할 수 있다. 감정의 사전적 정의는 어떤 현상이나 일에 대하여 일어나는 마음이나 느끼는 기분이며, 심리학적 정의는 육체적 변화를 동반하는 유기체의 상태, 강렬한 느낌과 충동 때문에 나타나는 흥분 및 동요의 상태이다[2,3]. 쾌-불쾌 또한 이러한 감정 중의 하나이다. 쾌는 쾌감과 같은 뜻으로서 상쾌하고 즐거운 느낌을 말하며, 불쾌는 못마땅하여 기분이 좋지 않은 느낌을 표현한다. 인간이 감정을 인지하는 데 주로 사용하는 감각은 시각이며, 인간의 감정에 영향을 주는 시각적인 요인에는 텍스트, 이미지의 색채, 도형 등이 있는데 이러한 시각적인 요인을 이용하여 개인의 감정을 분석함으로써 사용자에게 개인별 감정기반의 서비스를 제공해 줄 수 있다. 인간의 감정을 분석하기 위한 기존 연구에는 텍스트 정보, 이미지의 정보를 이용한 연구가 있다[4-8]. [4]에서는 시각 감성 평가를 위해 제작된 표준 도형을 기반으로 도형감성, 방향감성, 비례감성, 색채감성의 평가 항목을 사용하여 각 항목에 대해 성별, 연령, 지역에 따른 집단이 느끼는 시각감정에 대해 연구했고, [5]에서는 삼각형, 사각형, 원을 기본도형으로 정의하여 기본도형을 중심으로 도형의 방향성, 정형성, 비례, 예리함에 따라 도형 간에 인간이 느끼는 감정에 대해 연구하였다. [6]에서는 영어 감정 어휘 사전인 SentiWordNet을 이용하여 트위터의 게시글

을 감정 분석하여 단어의 의미별 감정 분포도로부터 해당 단어의 감성 자질들을 추출하여 기계학습 기반의 감정 분석을 수행하는 연구를 하였고, [7]에서는 분노 감정을 중심으로 하여 영어 기반의 감정 사전인 SentiWordNet을 한국어 기반의 감정 사전을 구축하기 위한 기반을 연구했으며, [8]에서는 한국어 트윗을 대상으로 형태소와 음절을 기준으로 하여 어떤 기계학습이 트윗의 감정 분류에 적합한지에 대해 연구하였다. 이렇게 인간의 감정을 분류하기 위한 다양한 연구들이 활발하게 진행되고 있으며, 분류한 감정을 이용하여 개인별 맞춤형 서비스 제공을 위한 연구 또한 진행되고 있다[9,10]. [9]에서는 SNS 사용자의 감정, 정서 및 관심사에 대한 정보와 영화 정보를 분석하여 개인의 감정에 적합한 영화를 추천해주는 연구를 했고, [10]에서는 감정 분류 모델중 하나인 Thayer 모델을 기반으로 감정 분류 기준을 세우고, 트위터에서 수집한 트윗을 이용하여 각 영화에 대해 대중이 느끼는 감정을 분석하여 실제 영화 정보와 SNS 사용자가 느끼는 감정 정보가 유사한지 검증하는 연구를 했다. 하지만 기존 연구들에서는 대부분 SNS 사용자 게시글의 텍스트에 해당하는 정보만을 이용하여 사용자의 감정을 분류하는 것이 대부분이었으며, 감정을 분류할 때 텍스트 정보만 고려하는 것은 SNS 사용자의 감정을 정확하게 분류하기 어렵다는 문제점이 있다. 따라서 본 논문에서는 SNS 사용자의 게시글에서 이미지 정보와 텍스트 정보를 이용하여 감정을 분류한다. 이미지 정보와 텍스트 정보를 분석한 후 수치화 시켜서 SNS 사용자의 감정을 분류하는 방법을 제안한다.

## III. 사용자 감정 분류

### 1. 시스템 구성도

본 절에서는 인스타그램의 사용자 게시글 내의 텍스트와 이미지를 분석하여 사용자의 감정정보를 측정하기 위한 방법을 제안한다. 그림 1은 사용자 감정을 분류하기 위한 시스템 구성도이다. 사용자의 감정을 분류하기 위해서 이미지에서 Canny알고리즘으로 이미지의 Edge를 추출한 후 대표 도형을 추출한다. Canny알고리즘으로 추출하는 대표 도형은 삼각형, 사각형, 원이며, 삼각함수로 대표 도형의 기울기를 측정해서 도형의 종류와 도형의 기울기를 이용하여 도형의 최종 쾌-불쾌 수치를 구한다. 도형의 쾌-불쾌 수치와 기울기의 쾌-불쾌 수치를 더한 값이 도형의 최종적인 쾌-불쾌 값이 된다. 텍스트에서는 게시글의 본문, 해시태그, 댓글 등의 모든 텍스트를 이용하며, 전처리과정을 통해 형용사만을 추출한다. 형용사는 사람이나 사물의 성질이나 상태를 나타내는 품사이기 때문에 감정에 관련된 표현 또한 형

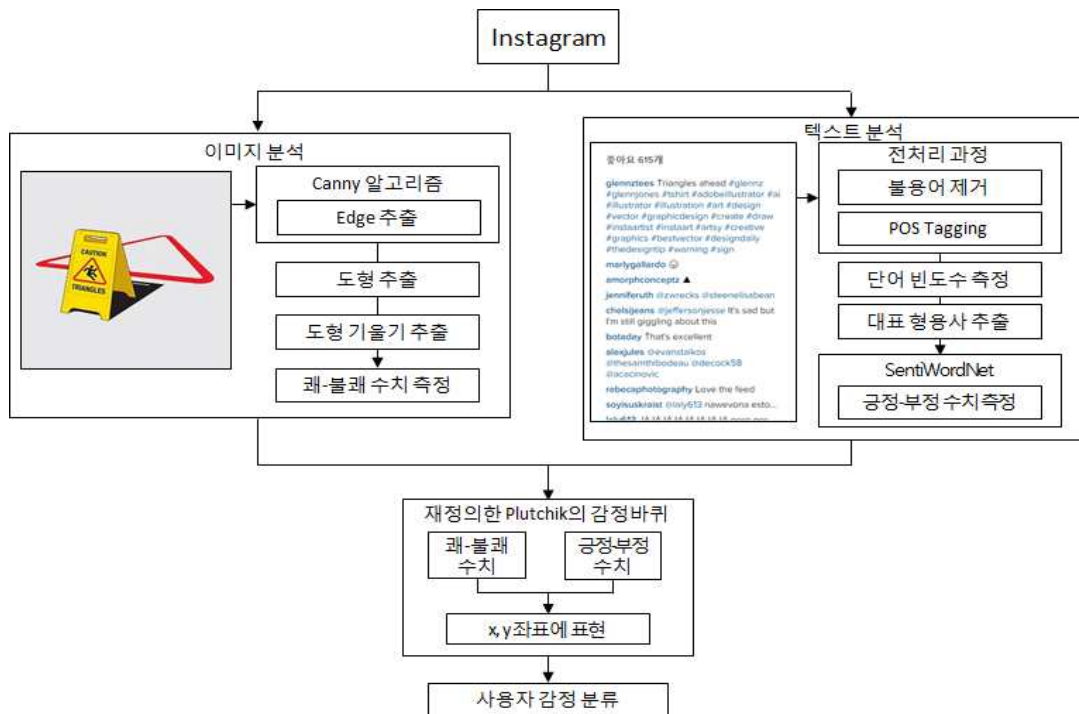


그림 1. 인스타그램 사용자 감정 분류 시스템 구성도


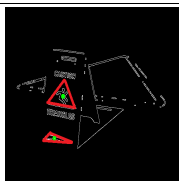


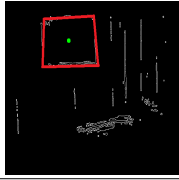
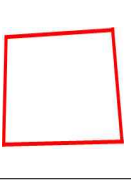

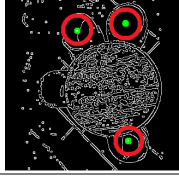
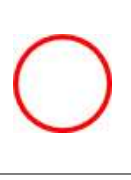
용사에 포함된 경우가 많으므로 텍스트에서 추출할 단어의 품사를 형용사로 선정했다. 추출한 형용사 중에 빈도수가 가장 높은 단어를 대표 형용사로 선정하고 SentiWordNet을 이용하여 대표 형용사의 긍정-부정 수치를 측정한다. 최종적으로 Plutchik의 감정 바퀴를 긍정-부정의 축과 쾌-불쾌의 축을 가진 좌표평면으로 재정의하며, 이미지에서 측정된 쾌-불쾌 값과 텍스트에서 측정된 긍정-부정 값을 재정의한 Plutchik의 감정 바퀴의 좌표 평면에 나타내어 사용자 감정을 분류한다.

## 2. Canny 알고리즘과 삼각함수를 이용한 이미지 분석

본 절에서는 인스타그램 사용자의 이미지를 이용하여 쾌-불쾌 수치를 측정하는 방법에 대해 기술한다. 실험에 사용된 데이터는 자바 기반의 인스타그램 Open API인 Jinstagram을 사용하여 수집하였다[11]. 수집된 이미지에서 쾌-불쾌 수치를 측정하기 위해서는 Canny 알고리즘을 이용해 도형의 Edge를 추출하고, Edge에서 도형을 추출한 후 삼각 함수를 이용해 도형의 기울기를 측정하여 쾌-불쾌 값을 구한다. Canny 알고리즘은 이미지의 윤곽을 추출하는 알고리즘 중 하나인데 뛰어난 윤곽 추출 능력을 자랑한다. Canny 알고리즘은 이미지의 잡음을 제거한 후 윤곽을 검출하고 최대 임계값과 최소 임계값을 이용하여 실제 윤곽과 가장 유사한 윤곽만을 추출한 후 최종적으로 유사한 윤곽들을 연결하는 단계를 거쳐서 이미지의 윤곽을 추출한다

[12]. 실시간 이미지 프로세싱에 중점을 둔 오픈 소스 컴퓨터 비전 C 라이브러리인 OpenCV를 이용하였으며, Canny알고리즘을 구현한 cvCanny함수를 이용하여 이미지의 윤곽을 추출한 후 윤곽에서 이미지의 대표 도형을 추출한다[13]. 이미지의 삼각형과 사각형을 추출하기 위해서 cvFindContours함수를 통해 윤곽선을 찾고 cvApporoxPoly함수를 통해 윤곽선을 다각형으로 근사화하여 그 꼭짓점 정보를 저장한다. 저장된 값이 3이면

표 1. Canny알고리즘을 이용한 이미지 내의 도형 추출

	원본 이미지	도형 검출	도형 추출
(가)			
(나)			
(다)			

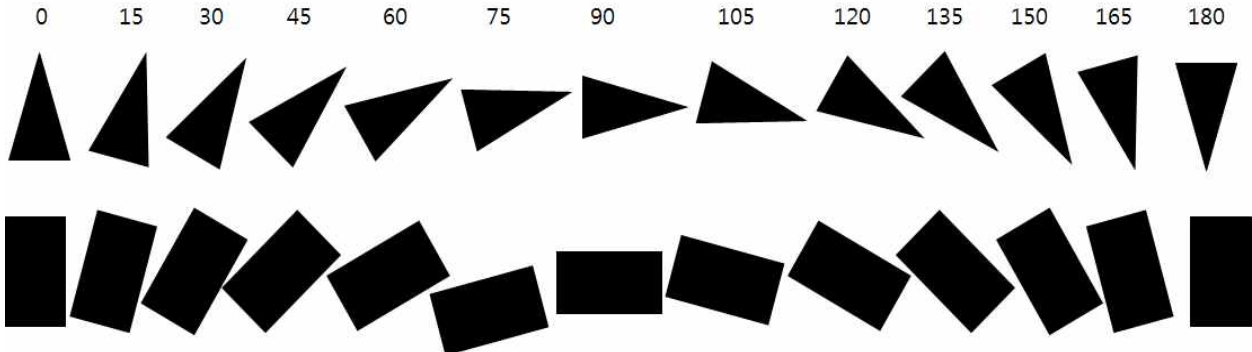


그림 2. 이미지에서 추출한 도형의 기울기 측정 기준

삼각형으로 판별하고, 4이면 사각형으로 판별하여 이미지에서 삼각형과 사각형을 추출한다. 이미지에서 원을 추출하기 위해서 HoughCircles 함수를 사용했다. 표 1은 이미지에서 도형을 추출하는 모습을 볼 수 있다. 표 1에서 보는 바와 같이 (가)에서는 삼각형, (나)에서는 사각형, (다)에서는 원을 추출한다. 추출된 도형의 기울기는 0도에서 180도 사이의 값을 이용하며, 도형이 기울어짐에 따라 발생하는 각도를 구한다. 추출된 도형의 기울기는 그림 2를 기준으로 측정하며, 본 논문에서는 도형의 정형성과 관련 없이 기본 도형의 기울기만을 측정하기 때문에 정형성을 갖는 타원은 실험 대상에 포함되지 않으므로 원의 경우에는 기울기가 항상 0의 값을 갖게 된다. 이미지에서 Canny알고리즘을 이용하여 도형을 추출할 때 각 도형의 꼭짓점의 좌표를 구할 수 있으므로 각 꼭짓점의 좌표 값을 이용하여 도형의 선분 길이를 구한 후 선분 길이를 이용하여 도형의 기울기를 구한다. 도형 빗변의 길이는 피타고라스의 정리를 이용하여 구한 후 삼각함수를 이용하여 도형의 기울기를 측정하였으며, (1)은 피타고라스 정리, (2)는 삼각함수의 공식이다[14, 15].

$$a^2 + b^2 = c^2 \tag{1}$$

$$\sin A = \frac{a}{h} \tag{2}$$

(1)에서 c는 직각삼각형의 빗변의 길이를 뜻하며, a와 b는 각각 나머지 두 변의 길이를 뜻한다. (2)에서 a는 A와 마주 보는 변의 길이이며, h는 직각삼각형의 빗변의 길이를 뜻한다. (1),(2)를 이용하여 이미지에서 추출한 도형의 기울기를 측정한다. 도형의 기울기를 측정하기 위해 밑변을 기준으로 하며, 길이가 가장 짧은 변을 밑변으로 선정한다. 그림 3은 표 1의 (가)에서 추출한 삼각형의 기울기를 측정하기 위해 x축과 y축을 갖는 보조선을 추가한 그림이다. 삼각형의 경우 그림2와 비교하였을 때 0도~15도 사이로 추측할 수 있으며, 밑변을 선정하기 위해 각 변

의 길이를 측정한다. 선분AB의 길이는 98이고, 선분 BC의 길이는 91, 선분 AC의 길이는 105이므로 길이가 가장 짧은 선분BC를 밑변으로 선정하여 도형의 기울기를 측정한다. 정확한 각도를 측정하기 위해 피타고라스의 정리를 이용하여 도형 빗변의 길이를 구한 후 삼각함수를 이용하여 도형의 기울기를 측정한다. 선분CD의 길이는 89이고 선분BD의 길이는 17이므로 보조선에 의해 생긴 직각 삼각형의 빗변인 선분BC는 (1)을 이용하여 계산하면 91의 값이 나온다. 최종적으로 (2)를 이용하여 x

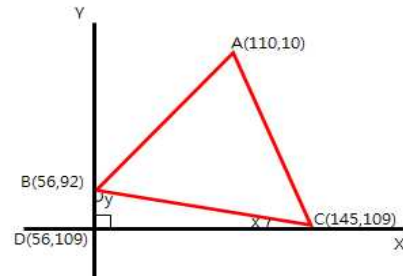


그림 3. 삼각형의 기울기 측정

의 각도를 구하면 삼각형의 기울기를 측정할 수 있다. x의 값은 y의 값을 구한 후에 삼각형의 내각의 합이 180도라는 것을 이용하여 계산하는데 y의 값이 78이므로 x의 값은 12도이며, 삼각형의 기울기는 12도이다. 삼각형의 기울기를 측정하는 방식과 같게 사각형의 기울기도 측정한다. 그림 4는 표1의 (나)에서 추출한 사각형의 기울기를 측정하기 위해 x축과 y축을 갖는 보조선을 추가한 것이다.

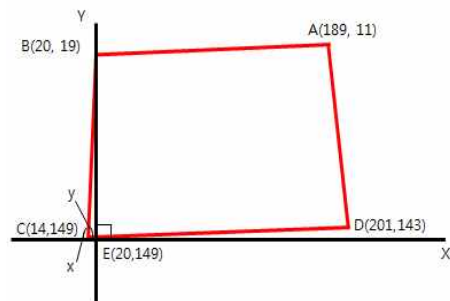


그림 4. 사각형의 기울기 측정

그림 2와 비교했을 때 사각형의 기울기는 76도~90도 사이로

추출할 수 있으며, 밑변을 선정하기 위해 각 변의 길이를 측정한다. 선분AB의 길이는 169.18이고, 선분 BC의 길이는 130.14, 선분 CD의 길이는 187.0, 선분AD의 길이는 132.45이므로 길이가 가장 짧은 선분BC를 밑변으로 선정하여 도형의 기울기를 측정한다. 정확한 각도를 측정하기 위해 (1)과 (2)를 이용하여 사각형의 기울기를 측정한다. 선분BE의 길이는 130이고 선분CE의 길이는 6이므로 보조선에 의해 생긴 직각 삼각형의 빗변인 선분 BC는 (1)을 이용하여 계산하면 130.14가 나온다. 사각형의 기울기인 x값을 구하기 위해서 (2)를 이용하여 y의 값을 구한 후 180도에서 y의 값을 빼준다. (2)를 이용한 y의 값은 89도가 나오며, x의 값은 91도이므로 사각형의 기울기는 91도가 된다. Canny알고리즘과 삼각함수를 이용하여 구한 도형과 도형의 기울기를 이용하여 궤-불괘 수치를 측정한다. 도형의 종류와 도형의 기울기에 따른 궤-불괘 수치 측정에 대한 연구를 참고하여 기존의 궤-불괘 수치를 재정의 하였다[16]. 궤-불괘 수치를 단순히 판별하는 것이 아니라 수치를 두어서 감정의 강함과 약함을 나타내었으며, 표 2와 표 3은 기존의 궤-불괘 수치를 재정의 한 내용이다.

표 2. 도형의 종류에 따른 궤-불괘 수치 재정의

삼각형		사각형		원	
기존	재정의	기존	재정의	기존	재정의
4.0	-1.0	4.5	-0.5	6.0	1.0

표 3. 기울기에 따른 궤-불괘 수치 재정의

기울기(°)	기존	재정의
0	5.7	0.7
1~15	5.6	0.6
16~30	5.3	0.3
31~45	5.1	0.1
46~60	4.9	-0.1
61~75	5.1	0.1
76~90	5.1	-0.1
91~105	5.4	0.4
106~120	4.9	-0.1
121~135	4.7	-0.3
136~150	4.5	-0.5
151~165	4.9	-0.1
166~179	4.9	-0.1
180	5.4	0.4

기존의 도형의 종류와 도형의 기울기에 대한 궤-불괘 수치는 4.0에서 6.0사이에 존재하는데 이 값들을 최댓값 1.0, 최솟값 -1.0을 갖는 좌표평면에 표현하기 위해서 기존의 각 값에 -5.0

을 더해서 -1.0에서 1.0사이의 값으로 재정의하였다. 도형의 종류에 따른 궤-불괘 수치와 도형의 기울기의 궤-불괘 수치를 더해서 최종적인 궤-불괘 수치를 측정한다. 표 1의 (가)에서 추출한 삼각형의 경우 도형의 궤-불괘 수치는 -1.0이고 기울기는 12도이므로 기울기의 궤-불괘 수치는 0.6이다. 따라서 표 1의 (가)이미지의 궤-불괘 수치는 -0.4이다. 표1의 (나)에서 추출한 사각형의 경우 도형의 궤-불괘 수치는 -0.5이며, 기울기는 91도이므로 기울기의 궤-불괘 수치는 0.4이다. 따라서 표 1의 (나) 이미지의 궤-불괘 수치는 -0.1이다. 표 1의 (다)이미지에서 추출한 도형은 원이므로 원의 궤-불괘 수치는 1.0이며, 원의 경우 기울기는 항상 0으로 정의하였기 때문에 표 1의(다)이미지의 궤-불괘 수치는 1.0이다.

### 3. SentiWordNet을 이용한 텍스트 분석

인스타그램 사용자들이 이미지와 함께 게시한 본문의 내용과 댓글을 이용하여 대표 형용사를 추출하는 방법에 대해 기술한다. SentiWordNet은 오피니언 마이닝을 위한 어휘 사전으로 WordNet의 동의어 집합에 긍정, 부정, 중립의 감정 점수를 부여한다[17]. 전 세계 300여 개 이상의 연구 그룹에서 활용되고 있으며, 현재는 WordNet3.0에 감정 점수를 부여한 SentiWordNet3.0이 가장 최근 버전이며, 명사, 동사, 형용사, 부사의 네 개의 품사에 속하는 총 147,278개 영어 단어들을 117,659개의 동의어 집합들로 분류해 놓은 어휘 사전이다[18]. 인스타그램 사용자 게시글에서 추출한 텍스트 데이터들은 전처리 과정을 통해 특수문자, 이모티콘, 숫자, 불용어 등을 제거하고 대문자와 소문자 간의 단어 중복을 피하고자 모든 텍스트를 소문자로 변환한다. 소문자로 변환된 텍스트들을 POS Tagging(Part-Of-Speech Tagging)을 통해 형용사를 추출한다. POS Tagging이란 단어가 가진 품사의 종류를 Tag 형식으로 나타내 주는 것이다. JAVA POS 라이브러리인 Stanford Log-linear POS Tagger를 이용하여 POS Tagging을 하였다[19]. 형용사의 경우 huge\_JJ, nice\_JJ 등의 형태로 나타나며, 추출된 형용사는 빈도수를 측정하여 빈도수가 높은 단어를 대표 형용사로 선정한다. 선정된 대표 형용사는 SentiWordNet을 이용하여 수치화하는데 SentiWordNet의 긍정-부정 수치는 각각 0.0에서 1.0사이의 값으로 정의되어 있다. 본 논문에서는 대표 형용사의 긍정-부정 수치를 -1.0에서 1.0사이의 값으로 사용하기 위해서 부정의 수치에 -1을 곱하여 음수의 값으로 사용한다. 표 4는 표1의 (가)이미지에 함께 게시되어 있는 텍스트에서 대표 형용사를 추출하는 과정을 보여준다. 원본 내용에서 특수문자, 이모티콘, 숫자, 불용어 등을 제거하고 소문자로 변환한 후 POS Tagging을 통해 각 단어의 품사를 추출하였다. 형용사형을 나타내는 JJ Tag를 표 4. 전처리 과정을 통한 대표 형용사 추출

원본 내용	Triangles ahead #glennz #glennjones #tshirt #adobeillustrator #ai #illustrator #illustration #art #design #vector #graphicdesign #create #draw #instaartist #instaart #artsy #creative #graphics #bestvector #designdaily #thedesigntip #warning #sign …….
	@zwrecks @steenelisabeam …….
불용어 제거	triangles ahead glennz glennjones tshirt adobeillustrator ai illustrator illustration art design vector graphicdesign create draw instaartist instaart artsy creative graphics bestvector designdaily thedesigntip warning sign …….
	zwrecks steenelisabeam …….
POS Tagging	triangles_NNS ahead_RB glennz_NN glennjones_NNS tshirt_NN adobeillustrator_NN ai_VB illustrator_NN illustration_NN art_NN design_NN vector_NN graphicdesign_NN create_VB draw_NN instaartist_NN instaart_NN artsy_JJ creative_JJ graphics_NNS bestvector_NN designdaily_RB thedesigntip_NN warning_VBG sign_NN …….
	zwrecks_NNS steenelisabeam_NN …….
추출된 형용사	(creative, 4) (artsy, 1) (excellent, 1) (sad, 1)
대표 형용사	creative

갖는 단어 중에서 가장 많은 빈도수를 보이는 단어가 최종 단어로 선정된다. 추출된 형용사는 creative, artsy, excellent, sad 이고 빈도수가 가장 높은 creative가 대표 형용사로 추출되었으며, SentiWordNet을 이용해 대표 형용사의 긍정-부정 수치를 측정한다. 최소값이 -1.0이고 최댓값이 1.0인 x축과 y축을 갖는 좌표평면에 표현하기 위해서 추출된 형용사의 부정 값에 -1을 곱하여 음수로 변환해준다. creative의 경우에는 긍정이 0.375로 나왔으므로 그대로 사용한다. 이와 같은 방법으로 표1의 (나)이미지에 함께 게시된 게시글의 대표 형용사를 선정하기 위해 추출된 형용사는 huge와 magnificent이며, 빈도수를 비교한 결과 magnificent가 더 높은 빈도수를 보였기 때문에 대표 형용사로 magnificent가 선정되었다. magnificent를 SentiWordNet을 이용하여 긍정-부정 수치를 측정하면 부정의 0.25가 측정되므로 -1을 곱한 -0.25가 표 1의 (나)이미지 게시글의 최종 긍정-부정 수치가 된다. 표 1의 (다)이미지 게시글에서 대표 형용사를 선정하기 위해 추출된 형용사는 abstract, geometry, minimal, nice이며, 가장 빈도수가 높은 형용사는 nice이기 때문에 nice가 대표 형용사로 선정된다. nice를 SentiWordNet을 이용하여 긍정-부정 수치를 측정하면 긍정의 0.88이 측정되므로 표 1의 (다)이미지 게시글의 긍정-부정 수치는 0.88이 된다.

IV. 실험 및 결과

인스타그램 게시글에서 210개의 이미지와 본문 텍스트, 2,940개의 댓글을 수집하여 실험 하였으며, 이미지와 텍스트를 이용하여 추출한 쾌-불쾌 수치와 긍정-부정 수치를 재정의한 Plutchik의 감정 바퀴에 표현하여 최종적인 사용자의 감정을 분류한다. 심리학이나 철학 등에서는 기본적으로 몇 가지 감정을 기본 감정으로 분류하고, 나머지 유사한 감정들을 기본 감정들의 조합으로 분류하고 있다. Plutchik의 감정 바퀴도 그중 하나인데 joy, trust, fear, surprise, sadness, disgust, anger, anticipation의 8가지를 기본 감정으로 정의하고 동심원 형태로 배치했다. 원의 안쪽으로 갈수록 더 강도가 강한 감정이며, 원의 바깥으로 갈수록 강도가 약한 감정인데 강도는 세 단계로 이루어져 있다. 그리고 각 감정 사이에는 두 감정의 조합 감정을 배치하고 있으며, 서로 반대인 감정은 동심원에서도 반대의 형태를 보인다. 예를 들어 joy와 sadness는 서로 반대 감정이므로 동심원에서도 반대의 형태를 보인다[20]. 본 논문에서는 Plutchik의 감정 바퀴의 조합 감정은 배제하고 기존 감정 강도를 반대로 배치하여 원의 안에서 바깥쪽으로 갈수록 더 강한 강도의 감정이 위치하도록 감정을 재배치하며, 긍정-부정의 축과 쾌-불쾌 축을 갖는 좌표평면으로 재정의한다. 재정의한 Plutchik의 감정 바퀴에 이미지에서 추출한 쾌-불쾌 수치와 텍스트에서 추출한 긍정-부정 수치를 좌표평면상에 나타내어 좌표평면상에 위치하는 감정을 사용자의 감정으로 분류한다.

표 5. 추출된 쾌-불쾌 값과 긍정-부정 값

NO	쾌-불쾌	긍정-부정	분류된 감정
1	0.6	0.375	fear
2	-0.1	-0.25	pensiveness
3	1.0	0.88	vigilance
4	-0.2	0.374	acceptance
5	-0.4	0.625	trust
6	1.0	0.125	vigilance
7	-0.3	0.25	apprehension
8	-0.6	0.75	admiration
9	-0.6	-0.25	rage
10	-0.3	-0.125	distraction
11	0.1	0.5	joy
12	1.0	0.75	vigilance
13	-0.9	-0.375	amazement
14	0.2	0.125	interest
15	1.0	0.625	vigilance
16	-0.6	0.875	admiration
17	1.0	0.375	vigilance
18	-0.6	-0.375	surprise
19	-0.9	0.125	terror
20	-0.6	0.5	fear
21	-0.1	0.25	acceptance
22	1.0	0.875	vigilance
∴	∴	∴	

표 5는 제안하는 방법으로 이미지에서 추출한 이미지의 도형



과 도형의 기술기에 따른 쾌-불쾌 값과 텍스트에서 추출한 대표 형용사에 따른 긍정-부정 값, 추출한 쾌-불쾌 값과 긍정-부정 값을 재정의한 Plutchik의 감정 바퀴에 표현하여 분류한 사용자 감정을 보여준다.

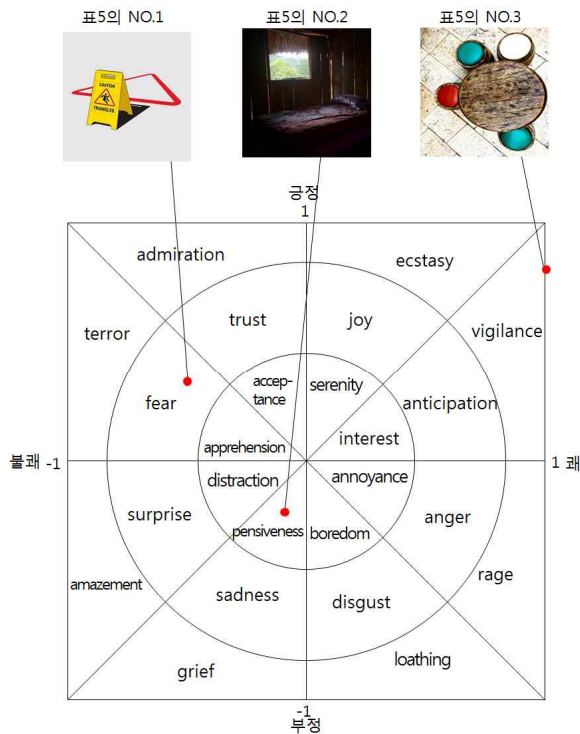


그림 5. 최종 사용자 감정 분류 예

그림 5는 최종적인 사용자 감정 분류를 하기 위해 쾌-불쾌 값과 긍정-부정 값을 재정의한 Plutchik의 감정 바퀴에 표현하는 것을 보여준다. 표 5의 NO.1, NO.2, NO.3을 재정의한 Plutchik의 감정 바퀴에 표현했을 때 NO.1의 쾌-불쾌 값은 0.6, 긍정-부정 값은 0.375이므로 사용자 감정은 fear로 분류되었고, NO.2의 쾌-불쾌 값은 -0.1, 긍정-부정 값은 -0.25이므로 사용자 감정은 pensiveness로 분류되었다. NO.3의 쾌-불쾌 값은 1.0, 긍정-부정 값은 0.89이므로 사용자 감정은 Vigilance로 분류되는 것을 알 수 있다.

### V. 결론 및 향후연구

기존에 연구된 SNS를 이용한 다양한 연구들은 대부분 텍스트를 활용하여 사용자의 감정을 분류하였으나 단순히 텍스트만으로는 사용자의 감정을 측정하기는 쉽지 않다. 이미지는 시각 정보이기 때문에 색상, 도형 등의 정보를 포함하고 있을 뿐만 아니라 텍스트보다 풍부한 정보를 내포하고 있다. 따라서 본 논문에서는 SNS 사용자 게시글의 텍스트뿐만 아니라 이미지를 활용하여 사용자의 감정을 분류한다. SNS 사용자의 게시글에서

대표 형용사를 추출하고, Canny 알고리즘과 삼각함수를 이용하여 사용자의 감정을 분류하는 방법을 제안하였다. Canny 알고리즘과 삼각함수를 이용해 이미지에서 추출한 쾌-불쾌 수치와 텍스트에서 추출한 대표 형용사의 긍정-부정 수치를 이용하였다. 쾌-불쾌 수치와 긍정-부정 수치를 사용하기 위해서 기존에 연구되어 있던 쾌-불쾌 수치와 SentiWordNet의 긍정-부정 수치를 재정의하였고 재정의한 쾌-불쾌 수치와 긍정-부정 수치를 이용하여 재정의한 Plutchik의 감정 바퀴의 좌표 평면에 표현함으로써 사용자의 감정을 분류하였다. 제안하는 방법으로 인스타그램 게시글에서 210개의 이미지와 본문 텍스트, 2,940개의 댓글을 이용하여 실험한 결과 73.33%의 사용자 감정을 분류할 수 있었다. 인스타그램 게시글에서 수집한 이미지와 텍스트에서 대표 도형과 대표 형용사가 추출되지 않는 경우 정확한 사용자의 감정 분류가 정확하게 되지 않았다. 따라서 본 논문에서 제안한 사용자 감정 분류 방법을 적용하기 위해서는 텍스트의 형용사 비중이 높아야 하며, 이미지에 도형이 포함되어야 한다는 한계점이 있다. SNS 사용자의 게시글을 통해 감정을 분류함으로써 감정기반의 다양한 연구에 적용이 가능할 것으로 보인다. 향후 연구로는 본 연구를 통해 분류된 사용자 감정을 기반으로 한 감정 마케팅 방법에 관한 연구, 사용자 감정기반의 영화, 드라마, 음악 등의 콘텐츠 추천에 관한 연구로 발전시킬 수 있을 것으로 보인다. 또한, 본 연구에서는 게시글의 이미지에서 추출한 도형의 정형성을 고려하여 쾌-불쾌 수치를 측정하지 않았지만, 이미지에서 도형의 정형성까지 고려하여 쾌-불쾌 수치를 측정한다면 이미지에서 더 정확한 도형을 추출할 수 있을 것으로 보이며, 더욱 정확한 감정 분류를 할 수 있을 것으로 보인다. 텍스트를 이용한 긍정-부정 수치를 측정하는데 있어서도 형용사만을 추출하는 것이 아니라 명사, 동사, 부사를 포함하여 긍정-부정 수치를 측정하는 연구로 확장되어야 한다.

### References

- [1] 박지수, 김무철, 노승민, “멀티미디어 콘텐츠의 맞춤형 정보 제공 연구”, 한국전자거래학회지, Vol.20, No3, August 2015, pp.79-87
- [2] “표준국어대사전”, 국립국어원
- [3] 이성식, 전신현, “감정사회학”, 한울아카데미, 1995, pp.13
- [4] 김묘향, 고한우, 윤종희, “표준 도형에 따른 지역별 시각 감성 평가 : 대전과 대구 지역의 대학생을 중심으로”, 한국감성과학회 학술지, February 2002, pp. 160-163
- [5] 정대현, 한광희, “형태를 통해서 느끼는 감성 - 기본 도형을 중심으로 -”, 한국HCI학회논문집, 2007, pp. 445-451

- [6] 강인수, “영어 트위터 감성 분석을 위한 SentiWord Net 활용 기법 비교”, 한국지능시스템학회논문지, Vol.23, No.4, August 2013, pp. 317-324
- [7] 최석재, 권오병, “빅데이터 분석을 위한 한국어 SentiWordNet 개발 방안 연구 : 분노 감정을 중심으로”, 한국전자거래학회논문지, Vol19, No.4, Nov 2014, pp. 1-19
- [8] 임좌상, 김진만, “한국어 트위터의 감정 분류를 위한 기계학습의 실증적 비교”, 멀티미디어학회논문지, Vol17, No.2, February 2014, pp.232-239
- [9] 노희용, 송기식, 이성주, “SNS정보를 활용한 감성 기반 서비스 추천-영화추천 서비스 사례”, 한국경영학회 통합학술발표논문집, August 2014, pp.380-391
- [10] 김경민, 김동윤, 이지형, “트위터를 활용한 감정 기반의 영화 유사도 측정”, 한국지능시스템학회 논문지, June 2014, pp.292-297
- [11] <https://github.com/sachin-handiekar/jInstagram>
- [12] John Canny, “A computational Approach to Edge Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, November 1986, pp. 679-698
- [13] itseez, “OpenCV Reference Manual”, V2.1, March 2010, pp.293
- [14] 최명기, 이지현, “수학사적 관점에서 본 피타고라스 정리의 증명”, 대한수학교육학회지, Vol.9, No.4, December 2007, pp.523-533
- [15] 유재근, “삼각함수 개념의 역사적 분석”, 대한수학교육학회지, Vol.24, No.4, November 2014, pp.607-622
- [16] 정대현, 한광희, “형태를 통해서 느끼는 감성 - 기본 도형을 중심으로-”, 한국HCI학회논문집, 2007, pp. 445-451
- [17] Andrea Esuli, and Fabrizio Sebastiani, “SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining”, In Proceedings of the 5th Conference on Language Resources and Evaluation, 2006
- [18] Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani, “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining”, Proceeding of the International Conference on Language Resources and Evaluation(LREC), 2010
- [19] <http://nlp.stanford.edu/software/tagger.shtml>
- [20] R. Plutchik, “Emotions and Life: Perspectives From Psychology, Biology, and Evolution” American psychological Association, 2003

---

 저자 소개
 

---

**홍택은(준회원)**

2015년 조선대학교 컴퓨터공학부 공학사

2015년~현재 조선대학교 소프트웨어 융합공학과 석사과정

<주관심분야 : 소셜 네트워크, 오피니 마이닝, 감성정보 분석>

**김정인(준회원)**

2011년 조선대학교 컴퓨터공학부 공학사

2014년 조선대학교 박사 수료

2014년~현재 조선대학교 석·박사 연계 과정

<주관심분야 : 소셜 네트워크, 정보처리, 시맨틱 웹과 온톨로지, 정보검색>

**신주현(정회원)**

1986년~2011년 ㈜청전정보 팀장, ㈜투루텍 기술이사

2007년 조선대학교 전자계산학과 이학박사

2011년~현재 조선대학교 제어계측로봇공학과 산학협력중점교수

<주관심분야 : 멀티미디어 데이터베이스, 빅 데이터 처리, 텍스트마이닝, 감성정보 처리 등>