



# 이미지와 해시태그를 이용한 인스타그램의 감정 분석 연구

정다혜\*, 김장원\*\*

## A Study on the Emotion Analysis of Instagram Using Images and Hashtags

Dahye Jeong\*, Jangwon Gim\*\*

---

이 성과는 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.  
(No. NRF-2018R1C1B6008624).

---

### 요 약

최근 소셜 네트워크 서비스 사용자들은 게시글을 통해 사회적 이슈 및 관심 콘텐츠들에 대한 자신의 감정을 적극적으로 표현하고 공유한다. 그 결과 소셜 네트워크에서의 개인 및 특정 집단의 감정 공유는 빠르게 확산된다. 그러므로 사용자들의 게시글에 대한 감정 분석 연구가 활발히 진행되고 있다. 그렇지만 다양한 감정이 포함된 게시글에 대한 감정 분석 연구가 미흡하다. 따라서 본 논문에서는 해시태그와 이미지를 이용한 인스타그램 게시글의 대표 감정 분석 방법을 제안한다. 이를 통해 사용자 게시글에 포함된 다종의 리소스를 활용하여 다종의 감정으로부터 대표 감정을 추출할 수 있으며 66.4%의 정확도와 81.7%의 재현율로 기존 방법보다 감정 분류 성능 향상을 보인다.

### Abstract

Social network service users actively express and share their feelings about social issues and content of interest through postings. As a result, the sharing of emotions among individuals and community members in social network is spreading rapidly. Therefore, resulting in active research of emotion analysis on posting of users. However, There is insufficient research on emotion analysis for postings containing various emotions. In this paper, we propose a method that analyzes the emotions of an Instagram posts using hashtags and images. This method extracts representative emotion from user posts containing multiple emotions with 66.4% accuracy and 81.7% recall, which improves the emotion classification performance compared to the previous method.

### Keywords

emotion analysis, social networking services, instagram, multi resources, hashtag

---

\* 군산대학교 소프트웨어융합공학과

- ORCID: <https://orcid.org/0000-0001-5263-3940>

\*\* 군산대학교 소프트웨어융합공학과 교수(교신저자)

- ORCID: <https://orcid.org/0000-0002-4480-7944>

· Received: Aug. 01, 2019, Revised: Sep. 06, 2019, Accepted: Sep. 09, 2019

· Corresponding Author: Jangwon Gim

Dept. of Software Convergence Engineering, Kunsan National University,  
Soutn Korea

Tel.: +82-469-8916, Email: [jwgim@kunsan.ac.kr](mailto:jwgim@kunsan.ac.kr)

## 1. 서 론

최근 스마트폰의 증가와 무선 인터넷 서비스의 확장으로 인해 소셜 네트워크 서비스(SNS, Social Network Service)가 생활 속으로 밀접하게 자리 잡았다. 소셜 네트워크 서비스를 통해 사용자들은 특정 이슈 또는 개별 관심사에 대한 정보를 서로 주고받거나 자신의 의견을 적극적으로 표현한다. 또한 최근 소셜 네트워크 서비스 사용자들은 게시글을 통해 감정을 적극적으로 표현하고 이를 온·오프라인 친구들과 공유하여 상호간의 공감대를 구축하기 위해 많은 시간과 노력을 기울이고 있다. 그 결과 정보의 전파 및 확산 속도가 점점 더 빨라지고 있으며 특히 감정의 빠른 확산은 사회적인 이슈를 야기하기도 한다[1]. 유명 연예인에 대한 팬덤을 구축하여 서로의 감정을 공유하거나 유명인의 상해·자살 등과 같은 극단적인 상황에 또한 빠르게 전파되어 추가적인 사건 사고를 발생하는 원인이 되기도 한다[2]. 이러한 상황을 예방하고 대책을 수립하기 위해서 소셜 네트워크 상에서 표출되는 감정의 전염 및 확산의 추이에 대한 연구가 필요하다.

‘인스타그램(Instagram)’은 대표적인 소셜 네트워크 서비스로서 2018년 통계에 따르면 매일 5천5백만 건의 이미지 및 비디오가 게시되고 있으며 2020년까지 미국에서 월간 사용자가 1억 2천만 명에 이를 것으로 보인다[3][4]. 인스타그램 사용자들은 게시글을 통해 이미지, 동영상을 공유하고 그에 대한 추가 설명을 위해 해시태그를 사용한다. 이러한 해시태그를 통해 자신의 감정이나 현재 상태에 대한 상세 정보를 표현한다[5]. 인스타그램에서 감정이 표현된 해시태그를 포함하는 게시글은 약 2천만건으로 인스타그램을 통해 사용자들의 감정 공유가 활발하게 일어남을 알 수 있다[6]. 해시태그에 표현된 감정 어휘를 기준으로 게시글의 감정은 ‘행복’(91%), ‘화남’(2%), ‘슬픔’(6%), ‘평온’(1%)로 다양한 감정이 표현됨을 알 수 있다. 또한 각 감정 별 ‘평균 좋아요 수’는 ‘행복’(14%), ‘화남’(23%), ‘슬픔’(28%), ‘평온’(35%)을 보인다. 사용자들 사이에서 ‘좋아요’ 기능은 자신의 게시글에 대한 빠른 공감을 빠르게 이끌어 낼 수 있어 10대와 20대 연령층

에서 높은 참여율을 보인다. 이를 통해 사용자들은 인스타그램의 게시글을 통해 감정을 공유하고 있으며 각 감정 별로 공감하는 정도가 다름을 알 수 있다. 또한 게시글에 포함된 다중 리소스 중에서 이미지, 해시태그 등을 통해 자신의 감정을 표현한다. 따라서 한 개의 게시글에는 다양한 감정 표현이 다중의 리소스에서 나타날 수 있다. 그러므로 게시글의 대표 감정추출을 위해 다중의 리소스에 표현된 감정을 추출하는 방법이 필요하다.

본 연구는 인스타그램 게시글에 포함된 이미지와 해시태그 기반으로 게시글의 대표 감정 추출 방법을 제안한다. 하나 이상의 감정을 포함하는 이미지와 해시태그에서 대표 감정 추출을 위해 먼저 이미지에 포함된 후보 감정 속성과 해당 속성별 수준값을 추출한다. 이때 추출된 감정 속성과 동일한 감정을 표현하는 감정 해시태그를 비교 쌍으로 생성하고 동일 감정 해시태그의 출현 빈도수를 가중치로 적용하여 게시글의 대표 감정을 추출한다. 2장에서는 관련 연구를 소개하고, 3장에서 제안하는 방법을 설명하고 제안 방법의 실험 결과를 4장에서 논한다. 결론 및 향후 연구는 5장에서 기술한다.

## II. 관련 연구

### 2.1 해시태그 기반 감정 분석 연구

SNS 사용자가 트위터에 게시한 텍스트 분석을 통해 유명 관광 장소나 축제 이벤트에 대한 사용자의 감정을 분석한 연구가 수행되었다[7]. 트위터 메시지로부터 사용자의 감정을 분석하기 위해 감정 어휘 추출 및 감정 어휘 사전을 구축하여 장소에 대한 긍정 및 부정의 감정 분석을 수행하였다. 또한 SNS 게시글에 포함된 해시태그를 이용한 감정 분석 연구가 수행되었다[8]. 사용자들은 공통의 어휘 형태인 해시태그를 통해 정보를 공유하거나 감정을 표현한다. 따라서 이러한 해시태그에 표현된 감정(‘행복’, ‘화남’, ‘슬픔’, ‘평온’ 등)을 기반으로 게시글의 감정을 분석하는 연구가 수행되었다[9][10]. 그렇지만 게시글에 포함된 이미지와 해시태그에서 표현하는 감정이 서로 다르거나, ‘#눈물, #행복, #짜

증' 등과 같이 다양한 감정 어휘가 포함된 게시글의 대표 감정 추출 및 분석을 위한 방법이 필요하다. 위의 연구들에서 사용된 4가지 대표 감정은 Thayer의 감정 모델 평면으로부터 도출되었다[11].

그림 1은 감정의 모호성 구분을 위해 4분면을 기준으로 감정을 분류한 Thayer 모델을 나타내며 음악, 영상 등의 다양한 분야에서 콘텐츠의 감정 분류 및 분석에 적용되고 있다[12][13].

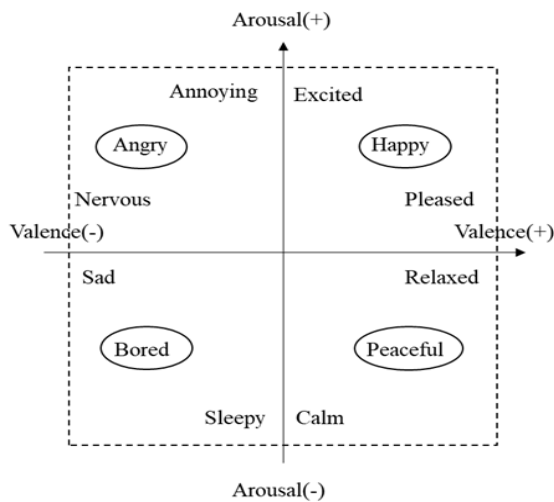


그림 1. Thayer의 감정 모델  
Fig. 1. Thayer's emotional model

## 2.2 이미지 기반 감정 분석 연구

감정 어휘 해시태그가 포함되어 있지 않거나 다중 어휘가 동시에 출현하는 경우 해시태그 이외의 리소스(이미지, 동영상 및 메타정보 등)를 활용한 감정 분석 방법이 필요하다. [14]은 인스타그램 게시글에 포함된 이미지의 RGB값을 추출하여 ('빨간색': '화남'), ('초록색': '행복'), ('파란색': '평온'), ('검은색': '슬픔')으로 게시글의 대표 감정을 추출한다. 그렇지만 사용자에게 의해 필터가 적용된 이미지, 자연광에 의해 특정 색상이 강조된 이미지의 경우 감정 분석 결과가 감소함을 보인다. 또한 감정 어휘 등의 다중 리소스가 반영되지 않았다. 이런 문제를 해결하기 위해 이미지의 대표 색상을 추출하고 텍스트와 함께 인스타그램 게시글의 사용자 감정을 분석한 연구가 수행되었다[15].

그러나 인물사진에서 추출 가능한 표정 등의 특징 정보가 반영되지 않고 RGB 색상만으로 감정 분

석을 수행한다. 그러므로 게시글의 정확한 감정 분석을 위해 다중 리소스를 고려한 감정 분석 방법이 필요하다. [16]은 인스타그램 게시글 중 해시태그로 감정이 기 분류된 이미지들에 대하여 자동화된 방법으로 이미지의 감정을 분류한다. 이를 위해 구글 클라우드 비전 API를 적용하여 이미지로부터 다양한 감정 어휘 및 각 감정 어휘에 대한 수준 정보를 확인할 수 있다. 그렇지만 사람 이미지에 대한 감정 분석만 가능하며, 다중의 감정 어휘를 결과로 제시하거나, 다수의 해시태그가 포함된 게시글에 대한 정확한 대표 감정 추출이 어렵다. 따라서 사람 뿐만 아니라 사물 이미지가 포함된 게시글과 다중의 감정 해시태그를 포함한 게시글에 대한 대표 감정 분류 방법이 필요하다.

## III. 제안 방법

본 논문에서 제안하는 게시글의 대표 감정 추출은 그림 2와 같이 3단계로 구성된다. 첫 번째 단계는 인스타그램 게시글에 포함된 이미지, 해시태그 및 메타정보를 수집하고 데이터 정제를 하여 실험 데이터를 구축한다. 이때 수집된 메타정보('alt'의 태그 값)는 웹 사이트에 사진이 불러오지 않을 경우를 대신하여 생성되는 텍스트로 사진에 등장하는 개체 식별 정보('사람' 또는 '사물' 등)를 포함하며 사람·사물 이미지 구분에 적용할 수 있다[17].

두 번째 단계는 수집된 데이터로부터 후보 감정을 추출하는 것으로 해시태그와 이미지를 분리하고 RGB 기반 분석[14]과 구글 비전 API[18] 기반의 이미지 객체 인식 방법으로 게시글의 대표 감정 추출에 사용될 후보 감정을 추출한다. 이때 사물 이미지는 RGB 기반 방법을 적용하고 사람 이미지는 구글 클라우드 비전 API를 각각 적용한다. RGB 기반 방법은 이미지의 각 픽셀 색상을 이용하여 대표 감정을 도출한다. 이를 위해 이미지의 크기는 가로, 세로 512픽셀로 고정하고 각 픽셀에서 추출된 RGB값에 대한 색상별 빈도수를 합하여 대표 색상을 추출한다. 이때 같이 출현한 해시태그와 비교하여 대표 감정을 추출한다. 구글 클라우드 비전 API는 사람 이미지로부터 개체 인식 및 감정 분류를 위해 사용된다[18][19].

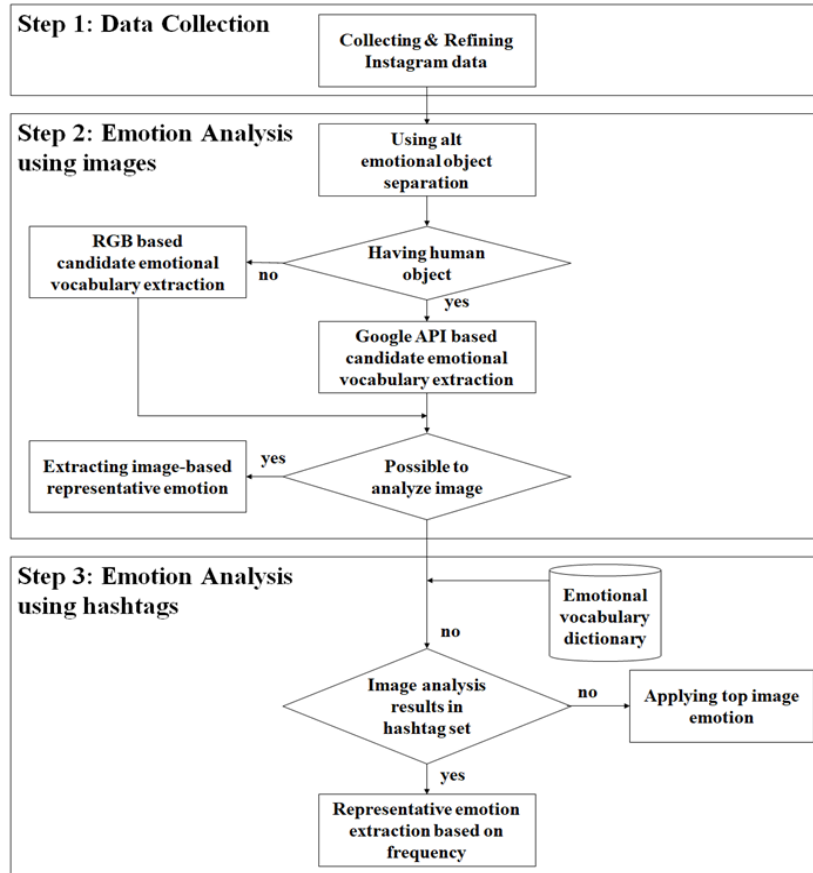


그림 2. 감정 분석 방법의 전체 개요  
Fig. 2. Overview of emotion analysis method

구글 클라우드 비전 API는 사람 개체가 표현할 수 있는 4가지 감정(‘행복’, ‘화남’, ‘슬픔’, ‘평온’)별로 ‘매우 좋지 않음(Very unliked)’, ‘좋지 않음(Unliked)’, ‘가능성 있음(Possible)’, ‘좋음(Liked)’, ‘매우 좋음(Very liked)’으로 감정의 수준 값을 제공한다. 따라서 게시글에 포함된 이미지로부터 다양한 감정 및 각 감정 별 수준 정보를 획득할 수 있으며 이를 게시글의 대표 감정 선정을 위한 후보 감정 집합으로 생성한다.

마지막 단계는 후보 감정 집합으로부터 게시글의 대표 감정을 추출하는 단계이다. 해시태그로부터 감정 어휘 추출을 위해 표 1과 같이 감정 어휘 사전을 구축 및 적용한다.

이때 Thayer 모델에서 ‘지루(Bored)’로 표현되는 감정은 실제 인스타그램의 해시태그에서 ‘슬픔(Sad)’ 또는 ‘슬픔(Sad)’과 관련된 어휘로 출현되는 빈도수가 높기 때문에 ‘지루’ 감정은 ‘슬픔’으로 대체하여 감정 어휘 사전을 구축한다[11].

표 1. 해시태그에서 추출된 감정 용어

Table 1. Emotion vocabularies extracted from hashtags

Emotions	Hashtag emotion vocabulary list
Happy	#행복, #행복해, #행복하자, #행복스타그램, #감동, #기쁨, #상쾌, #즐거움 ...
Angry	#화남, #짜증, #짜증나, #개짜증, #긴장, #불쾌, #불만, #원망 ...
Sad	#슬픔, #눈물, #눈물나, #속상, #우울, #울적, #외로움, #무기력 ...
Peaceful	#평온, #차분, #차분한, #여유, #여유로움, #한가, #자유, #편안 ...

$$\begin{array}{l}
 P_1(\text{happy: Possible}) \quad H_1(\text{happy: 3}) \\
 P_2(\text{sad: Unlikely}) \quad H_2(\text{sad: 2}) \\
 C_1(P_1, H_1) = 0.7 \times 3 + 0.3 \times 2 = 2.7 \\
 C_2(P_2, H_2) = 0.7 \times 2 + 0.3 \times 2 = 2.0 \\
 \therefore \text{Representative emotion} \\
 = e_1(\text{happy})
 \end{array}$$

그림 3. 대표 감정 추출 방법  
Fig. 3. Method for extracting representative emotion from a post

식 (1)과 그림 3은 대표 감정 선정을 위해 이미지로부터 추출된 후보 감정 집합과 해시태그 집합을 비교하는 방법을 나타낸다. 사람 이미지는 다수의 감정 속성( $P$ )으로 표현될 수 있기 때문에 각 감정 속성( $p_i$ )은 감정 어휘( $e_i$ ) 및 수준(level) 값의 쌍( $e_i:level$ )으로 표현된다. 수준 값은 가장 낮은 수준(1점)에서 가장 높은 수준(5점)으로 표현한다. 이때 감정 해시태그( $H$ ) 집합은 감정 속성( $h_i$ )으로 구성되며 각 감정 속성은 감정 어휘( $e_i$ ) 및 빈도 수( $freq$ )의 쌍( $e_i:freq$ )으로 구성된다.

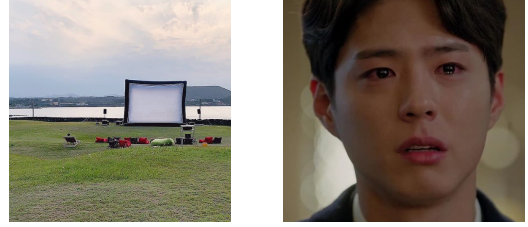
그러므로 이미지 정보와 해시태그 정보 비교를 위해 동일한 감정 어휘( $e_i$ )를 가진 비교 쌍( $C_i$ )을 생성하여 각 감정의 수준에 해시태그의 빈도수를 적용하였을 때  $C_i$ 의 값이 최대치가 되는 쌍에 포함된  $e_i$ 를 게시글의 대표 감정으로 추출한다. 또한 이미지를 주요 콘텐츠로 하는 인스타그램의 특징을 고려하여 이미지와 해시태그에 각각  $\alpha 0.7$ ,  $\beta 0.3$ 의 가중치를 부여한다.

$$\begin{aligned} level &= \{(VU:1), (U:2), (P:3), (L:4), (VL:5)\} \\ P &= \{p_1, p_2, \dots, p_i\}, H = \{h_1, h_2, \dots, h_i\} \\ p_i, h_i &= \{e_1, e_2, \dots, e_i\} \\ p_i &= (e_i : level_i), h_i = (e_i : freq_i) \\ C_i(p_i, h_i) &= (\alpha \times level_i + \beta \times freq_i) \\ \therefore Representative\ Emotion(e_i) &= Max(C_i(p_i, h_i)) \end{aligned} \quad (1)$$

## IV. 실험

### 4.1 데이터 구축

인스타그램 감정 분석을 위해 파이썬 기반 웹 크롤러를 개발하여 인스타그램 게시물 정보를 수집하고 제안 모델의 실험 평가를 위해 3가지 데이터 집합을 구축하였다. 신뢰성 있는 정답 집합 구축을 위해서 소셜 네트워크 서비스를 지속적으로 사용하고 있는 연구원(5명)들이 수작업으로 게시글에 대한 대표 감정을 추출, 분류 및 교차 검증하여 실험 데이터를 구축하였다. 또한 감정어휘 사전 어휘에 대한 검증을 실시하여 해시태그로부터 감정 어휘 추출에 대한 신뢰성을 확보하였다.



(a) 행복

(b) 슬픔

그림 4. 사물과 사람 이미지의 예시

Fig. 4. Examples of objects and people, (a) Happy, (b) Sad

첫 번째 데이터 집합은 기존 RGB 감정 분석과 제안 방법을 비교하기 위해 표 1에서 정의한 감정 어휘가 적어도 하나 이상 포함된 1,600개의 게시글의 이미지를 대상으로 한다. 사물과 사람 각 800개씩으로 구성되고 4가지 감정 별로 각 200개씩 구성된다. 그림 4는 실험 데이터의 일부 예시로 감정을 포함하고 있는 사물('행복'), 사람('슬픔') 이미지를 각각 보인다.

두 번째 데이터 집합은 다중 리소스를 감정 분석에 적용할 경우의 타당성을 확인하기 위한 것으로 게시글(이미지, 해시태그)에 사람이 포함되어있으며 이미지는 각각 4가지 감정으로 분류된다. 이를 위해 감정별 각 200개씩 총 800개로 정답 데이터를 구축한다. 두 번째 실험 데이터 집합에서 86건은 이미지와 해시태그에서 다중 감정을 포함한다. 이를 이용하여 다중 리소스 사용에 대한 감정 분석 정확도를 비교 분석한다. 또한 사용자 감정이 포함될 수 있는 키워드('제주도')를 이용하여 최근 공개된 게시글 5,000건 중 이미지(인물, 사물)와 다중 감정 어휘 해시태그를 동시에 포함하고 있는 300개의 게시글(행복 167건, 화남 10건, 슬픔 27건, 평온 95건; 사물 202건, 사람 98건)을 선정하고 다수(5명)의 연구원들을 통해 감정 레이블링 및 교차 검증을 실시하여 정답 데이터를 구축하였다.

### 4.2 다중 리소스 기반 게시글의 감정 분석

첫 번째 실험은 이미지를 대상으로 기존 RGB 방법과 제안방법의 정확도 및 재현율을 비교한다. 이를 위해 4.1에 정의된 첫 번째 감정 데이터 집합을 사용한다. 두 번째 실험은 다중 리소스(이미지, 해

시태그)가 대표 감정 추출에 어떠한 영향을 주는지 확인하기 위한 것으로 이미지 만을 대상으로 한 경우와 다중 리소스인 경우로 구분하여 제안 모델을 실험하여 단일 및 다수의 감정이 발생된 게시글의 대표 감정 추출 여부를 검증한다. 마지막 실험은 세 번째 실험 데이터 집합을 통해 이미지(사물, 사람) 만을 대상으로 한 실험과 제안방법(이미지, 해시태그)을 비교하여 제안 방법(다중 리소스)에 대한 성능을 검증한다. 제안 모델과 기존 모델의 평가를 위해 수식 2, 3의 감정 분류의 정확도와 재현율 지표를 정의한다. 정확도는 모델로부터 분류된 이미지 중 실제로 대표 감정을 정확하게 분류하였는지를 판단하고 재현율은 실제 대표 감정을 나타내는 이미지 중 모델이 얼마나 재현하는지를 나타낸다.

$$Precision = \frac{\text{images with classified emotion}}{\text{images categorized by the model}} \times 100 \quad (2)$$

$$Recall = \frac{\text{images categorized by the model}}{\text{images with representative emotion}} \times 100 \quad (3)$$

### 4.3 실험 결과

표 2는 사물 및 사람 이미지에 대하여 각 방법을 통해 정확하게 추출된 대표 감정 건수를 보인다.

첫 번째 데이터를 대상으로 한 기존 RGB 방법은 4가지 감정 모두에 대하여 평균 정확도(24.2%)와 평균 재현율(23.0%)로 나타났으며, 제안 방법(RGB 방법, 구글 API)은 평균 정확도 43.1%, 평균 재현율 39.2%로 나타났다. 제안 방법의 경우 ‘평온’과 ‘행복’의 감정들이 ‘화남’과 ‘슬픔’의 감정 보다 정확도와 재현율이 높음을 확인할 수 있다. 이는 구글 API의 성능이 행복한 이미지의 표정을 RGB 방법보다 정확하게 분류하기 때문이다. 또한, ‘화남’ 및 ‘슬픔’의 이미지는 사용자의 표정을 뚜렷하게 드러내지 않아 다른 감정보다 정확도가 낮다. 반면 RGB 실험의 경우 풍경 및 음식(총 70건) 사진에서 다수의 붉은 색상이 포함되어 ‘화남’의 감정 정확도가 높다.

표 3은 다중 리소스의 제안모델을 검증하기 위한

것으로 단일·다중 감정 및 다중 리소스(이미지, 해시태그)에 대한 실험 결과를 보인다. 4가지 모든 감정에 대한 실험 결과에서 다중 리소스를 적용한 제안 방법이 이미지만을 적용한 결과보다 단일감정 평균 정확도(28.3%), 평균재현율(9.7%)이 더 높게 나타난다.

표 2. 이미지에 대한 RGB 및 제안 방법 비교  
Table 2. Comparison results between the proposed model and RGB model

Methods \ Emotions		Happy (400)	Angry (400)	Sad (400)	Peaceful (400)
RGB	Objects	25	116	63	0
	People	27	102	52	1
	Precision	13.0	54.5	28.8	0.3
	Recall	23.0	24.4	24.2	20.0
RGB + G_API	objects	25	116	63	0
	People	141	2	55	171
	Precision	70.5	29.5	29.5	42.8
	Recall	61.5	26.2	36.2	32.8

표 3. 다중 리소스(이미지, 해시태그) 기반 감정 추출  
Fig. 3. Emotions extraction results using multiple resources(images, hashtags)

Methods		Emotions	Happy (200)	Angry (200)	Sad (200)	Peaceful (200)	Results
Images	Single	Number	135/192	2/176	49/179	145/167	331/714
		Precision	72.2	40.0	49.0	35.7	49.2
		Recall	70.3	1.1	27.4	86.8	46.4
	Multiple	Number	5/8	0/24	6/21	26/33	37/86
		Precision	55.6	0.0	50.0	51.8	39.4
		Recall	71.4	0.0	62.5	51.9	42.5
Hashtags	Single	Number	192/192	176/176	179/179	167/167	714/714
		Precision	100.0	100.0	100.0	100.0	100.0
		Recall	100.0	100.0	100.0	100.0	100.0
	Multiple	Number	0/8	0/24	0/21	0/33	0/86
		Precision	0.0	0.0	0.0	0.0	0.0
		Recall	0.0	0.0	0.0	0.0	0.0
Multiple resources	Single	Number	161/192	5/176	90/179	146/167	402/714
		Precision	86.1	100.0	87.3	36.4	77.45
		Recall	83.9	2.8	50.3	87.4	56.1
	Multiple	Number	7/8	15/24	18/21	30/33	70/86
		Precision	55.6	100.0	60.0	50.0	66.4
		Recall	87.5	62.5	85.7	90.9	81.7

또한, 다중 감정 평균 정확도와 평균 재현율은 각각 27.0%, 39.2% 향상된 결과를 보인다. 다중 감정이 포함되어 있는 게시글의 경우에는 해시태그만으로 대표 감정 추출이 어렵기 때문에 해시태그 방법의 다중 감정 분류 정확도와 재현율이 낮다.

그러므로 다중 감정이 포함된 게시글은 다중 리소스를 기반으로 대표 감정을 추출할 수 있다. 이때의 평균 정확도, 평균 재현율은 각각 66.4%, 81.7%를 보인다. 또한, 단일 감정의 경우에서도 평균 정확도 77.45%, 평균 재현율 56.1%로 제안방법의 우수함을 보인다.

표 4. 키워드(‘제주도’) 게시글에 대한 제안 모델의 성능 결과

Table 4. Performance results of the proposed model for posts including images and hashtags

Methods		Emotions		Happy (160)	Angry (10)	Sad (27)	Peaceful (95)
		Single	Multiple				
Images (RGB + G_API)	Objects+	Single		67/150	4/9	4/20	5/75
		Multiple		12/17	0/1	1/7	0/20
	Precision			81.6	4.1	11.9	20.8
	Recall			47.3	40.0	18.5	5.3
G_API + Hashtags	People	Single		58/67	0/0	0/6	5/6
		Multiple		13/13	0/0	2/3	2/2
	Precision			96.1	0.0	0.0	36.4
	Recall			88.9	0.0	22.2	87.5
Proposed model (Images + Hashtags)	Objects+	Single		69/150	4/9	4/20	5/75
		Multiple		15/17	0/1	4/7	4/20
	Precision			87.9	5.7	13.6	33.3
	Recall			50.3	40.0	29.6	9.5

표 4는 세 번째 데이터에 대한 제안 방법의 감정 분류 성능을 보이며 이미지 기반 방법(RGB, 구글 비전 API)과 제안 방법(이미지, 해시태그)의 실험 결과이다. 세 번째 실험에서 제안 방법은 모든 감정에 대해서 이미지만 적용한 방법보다 평균 정확도와 재현율이 각각 5.5%, 4.6% 높은 결과를 보여 새로운 데이터에서도 제안 방법이 상대적으로 정확하다는 것을 확인할 수 있다.

또한, 첫 번째 실험 데이터(각 감정별 동일한 건수를 사용)와 임의의 감정을 포함하는 세 번째 실험 데이터의 비교 결과 ‘이미지’에 대한 평균 정확도는 30%, 평균 재현율은 28%로 표 2의 RGB 기법의 실험 결과와 유사하다. 기존 이미지만을 대상으

로 한 실험의 정확도와 재현율이 낮은 이유는 RGB 기법으로 추출된 게시글의 대표 감정이 함께 등장한 해시태그에 포함된 감정과 다른 경우 때문이다. 구글 API 방법(사람)과 해시태그가 적용된 방법은 단일 감정(79.7%), 다중 감정(94.4%)의 평균 재현율을 보인다. 이 결과는 표 3의 다중 리소스에 대한 실험 결과와 비슷한 성능을 보인다. 따라서 제안 방법은 다양한 사물 및 사람 데이터와 다중 리소스에도 적용할 수 있다.

## V. 결론 및 향후 연구

최근 소셜 네트워크 서비스를 통해 사용자들은 자신의 의견이나 감정을 활발하게 공유 및 전파할 수 있게 되었다. 그 결과 ‘감정 확산’을 통해 극단적인 감정까지도 공유하는 사회적 현상이 나타나고 있다. 소셜 네트워크 서비스는 개인의 감정을 쉽게 표현할 수 있는 대표적인 방법의 하나이므로 이를 대상으로 한 감정 분석 연구가 중요한 역할을 한다.

그러므로 본 논문은 대표적 소셜 네트워크 서비스의 하나인 인스타그램 게시글에 대한 감정 분석 방법을 제안한다. 이를 위해 인스타그램에 포함된 이미지, 해시태그 등에서 추출되는 감정 정보를 이용하여 대표 감정을 분류하였다. 그 결과 기존의 RGB 방법과 비교하여 다중 리소스(이미지, 해시태그)를 대상으로 할 경우 다중 감정이 포함된 게시글의 대표 감정을 66.4%의 정확도와 81.7%의 재현율로 분류하였다. 그러므로 다중 리소스를 사용한 감정 분석 방법이 단일 리소스 사용 시 보다 우수한 성능을 보였다. 또한 사물과 사람이 빈번하게 나타나는 특정 장소 및 이벤트 등에 대한 사용자의 감정 분류가 가능함을 보였다. 그러므로 특정 이벤트에 대한 사용자들의 감정 발산 및 확산의 추이 분석과 특정 집단(관광객, 팬덤 등)을 대상으로 한 감정 분석을 향후 연구로 한다.

## References

[1] A. DI. Kramer, "The spread of emotion via Facebook", Proceedings of the SIGCHI Conference on Human Factors in Computing



- Systems, Austin, Texas, USA, pp. 767-770, May 2012.
- [2] Y. J. Jung, W. J. Seo, Y. R. Byeon, and T. W. Kang, "Classification of the suicidal signs in SNS using naive bayesian classifier", Proceedings of KIISE Conference, pp. 2129-2131, Jun. 2015.
- [3] 96 Instagram Statistics for Social Media Marketing Gurus, <https://99firms.com/blog/instagram-marketing-statistics>. [accessed: Jul. 30, 2019]
- [4] N. H. Chung, T. H. Um, and C. M. Koo, "Factors Influencing Photo Sharing for Creating Social Relationships on Instagram", Knowledge Management Research, Vol. 17, No. 4, pp. 129-145, Dec. 2016.
- [5] M. J. Kim, B. M. Kim, H. Y. Kim, S. Y. Baek, H. S. Shin, I. S. Lee, and J. W. Kim, "The Effect of Emotion Sharing with Hashtag on Perceived Attention Seeking and Social Interactivity", Proceedings of HCI Korea Conference, Jeongseon, Korea, pp. 460-467, Jan. 2016.
- [6] Star Tag, <https://startag.io>. [accessed: Jul. 30, 2019]
- [7] P. H. Oh and B. Y. Hwang, "Real-time Spatial Recommendation System based on Sentiment Analysis of Twitter", Journal of SEBS, Vol. 21, No. 3, pp. 15-28, Feb. 2016.
- [8] S. M. Kwon and J. H. Lee, "Feasibility of Hashtag of Former Posts on Instagram for using Contextual Information of Sentiment Analysis", Proceedings of KIISE Conference, pp. 649-651, Jun. 2018.
- [9] M. J. Nam, E. J. Lee, and J. H. Shin, "A Method for User Sentiment Classification using Instagram Hashtags", Journal of KMMS, Vol. 18, No. 11, pp. 1391-1399, Aug. 2015.
- [10] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, "Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach", Proceedings of the 20th ACM international conference on Information and Knowledge management(CIKM), Glasgow, United Kingdom, pp. 1031-1040, Oct. 2011.
- [11] R. E. Thayer, "The Biopsychology of Mood and Arousal", Oxford University Press, 1989.
- [12] Rachman, F. Hastarita, R. Sarno, and C. Fatichah, "Music emotion classification based on lyrics-audio using corpus based emotion" International Journal of ECE, Vol. 8, No. 3, pp. 1720-1730, Jun. 2018.
- [13] B. J. Kim and J. W. Lee, "A deep-learning based model for emotional evaluation of video clips", International Journal of Fuzzy Logic and Intelligent Systems, Vol. 18, No. 4, pp. 245-253, Dec. 2018.
- [14] S. B. Ki, J. W. Park, J. M. Park, G. Y. Lee, H. C. Lee, T. Saputri, and S. W. Lee, "Using Ontological Approach for Music Recommendation Based on Emotion Analysis on Instagram", Proceedings of KIISE Conference, pp. 1926-1928, Dec. 2017.
- [15] M. J. Nam, J. I. Kim, and J. H. Shin, "A User Emotion Information Measurement Using Image and Text on Instagram-Based", Journal of KMMS, Vol. 17, No. 9, pp. 1125-1133, Sep. 2014.
- [16] D. H. Jeong, S. B. Kim, C. R. Jeong, and J. W. Gim, "An Emotion Analysis Method Using Hash Tags and Images in Instagram", Proceedings of KIIT Conference, pp. 302-304, Nov. 2018.
- [17] ALT text on Instagram Photos, <https://thepreviewapp.com/alt-text-instagram-photos-use-like-professional/>. [accessed: Jul. 30, 2019]
- [18] Google cloud vision API, <https://cloud.google.com/vision/> [accessed: Jul. 30, 2019]
- [19] Z. Othman, N. A. Abullah, K. Y. Chin, F. F. W. Shahrin, S.S. Syed Ahmad, and F. Kasmin, "Comparison on Cloud Image Classification for Thrash Collecting LEGO Mindstorms EV3 Robot", International Journal of HaTI, Vol. 2, No. 1, pp. 29-34, Apr. 2018.



저자소개

정 다 혜 (Dahye Jeong)



2019년 9월 현재 : 군산대학교  
소프트웨어융합공학과(학부생)  
관심분야 : 텍스트 마이닝, 빅데이터  
분석, 소셜 네트워크 분석

김 장 원 (Jangwon Gim)



2005년 2월 : 상명대학교  
컴퓨터소프트웨어공학과(공학사)  
2008년 2월 : 고려대학교  
컴퓨터학과(이학석사)  
2012년 8월 : 고려대학교 컴퓨터·  
전파·통신공학과(공학박사)  
2013년 3월 ~ 2017년 3월 :

한국과학기술정보연구원 선임연구원

2017년 4월 ~ 현재 : 국립군산대학교 소프트웨어융합  
공학과 조교수

관심분야 : 텍스트 마이닝, 빅데이터 분석, 소셜 네트워크  
분석