



## 감정 분석에서의 문맥적 정보로써 인스타그램 과거 게시글의 해시태그 활용 가능성 제안

Feasibility of Hashtag of Former Posts on Instagram for using Contextual Information of Sentiment Analysis

---

저자 (Authors)	권순모, 이종혁 Soonmo Kwon, Jong-Hyeok Lee
출처 (Source)	<a href="#">한국정보과학회 학술발표논문집</a> , 2018.6, 649-651(3 pages)
발행처 (Publisher)	<a href="#">한국정보과학회</a> KOREA INFORMATION SCIENCE SOCIETY
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07503105">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07503105</a>
APA Style	권순모, 이종혁 (2018). 감정 분석에서의 문맥적 정보로써 인스타그램 과거 게시글의 해시태그 활용 가능성 제안. 한국정보과학회 학술발표논문집, 649-651
이용정보 (Accessed)	가천대학교 203.249.***.201 2019/09/29 19:00 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# 감정 분석에서의 문맥적 정보로써 인스타그램 과거 게시글의 해시태그 활용 가능성 제안

권순모<sup>+</sup>, 이종혁<sup>++</sup>

포항공과대학교 창의/IT융합공학과<sup>+</sup>, 포항공과대학교 컴퓨터공학과<sup>++</sup>  
{soonmo.kwon<sup>+</sup>, jhlee<sup>++</sup>}@postech.ac.kr

## Feasibility of Hashtag of Former Posts on Instagram for using Contextual Information of Sentiment Analysis

Soonmo Kwon<sup>+</sup>, Jong-Hyeok Lee<sup>++</sup>

Department of Creative IT Engineering, POSTECH<sup>+</sup>,  
Department of Computer Science & Engineering, POSTECH<sup>++</sup>

### 요 약

최근 자연어 처리를 활용한 감정 분석을 소셜 네트워크 서비스에 적용하여 개인의 의견이나 감정 등을 이해하는 데 활용하고 있다. 기존에는 사용자가 올린 콘텐츠에서 감정을 표현하는 어휘를 활용하여 어떤 감정을 나타내는지 분석하였지만, 문맥적 정보가 더 정확한 분석을 할 수 있게 도와준다. 본 연구에서는 사용자들이 비슷한 감정과 분위기를 나타내는 추세가 실제로 그러한지 검증함으로써 기존 콘텐츠의 해시태그가 문맥적 정보로 활용할 수 있는지 연구해보았다. 먼저 분석을 위해서 활용할 감정 프레임워크를 설명하고 10명의 사용자들의 각 사용자마다 개인의 유사성과 사용자 간의 유사성을 통해서 기존 콘텐츠들이 감정 분석에 문맥적 정보로 활용할 수 있음을 증명하였다.

### 1. 서 론

Web 2.0의 발달을 통해 사용자는 웹에서 정보 및 콘텐츠를 소비하는 것 뿐만 아니라 생산까지 가능하게 되었다. 이러한 변화를 가장 크게 이끈 것 중 하나가 페이스북(Facebook), 트위터(Twitter), 인스타그램(Instagram) 등과 같은 소셜 네트워크 서비스들(SNS)이다. 이 SNS들을 기반으로 사용자들은 정보를 올리고 공유하는 것 뿐 아니라 자신의 의견을 표현하거나 감정을 표출하는 데 활용한다. 이렇게 올린 사용자들이 생산한 콘텐츠는 그들의 의견과 감정이 많이 담겨있으며, 이를 더 이해하기 위해 SNS기반 다양한 감정 분석 및 분류 연구가 진행이 되고 있다 [1-6].

연구자들은 보다 면밀한 이해를 위해 사용자들이 올린 콘텐츠를 중심으로 연구를 진행하였다 [1-4]. 또한, 기존의 극성 분류를 통해 나타나는 긍정, 부정의 분류 뿐만 아니라 다양한 감정들을 분류한 연구도 진행이 되었다 [5]. 그러나 제공되는 콘텐츠만으로는 사용자들을 완전히 이해하기엔 한계가 있다. 연구자들은 문맥적 정보를 통해서 분류하려는 시도를 하였고, 이는 유의하게 활용이 되었다 [6,7]. 결국 문맥적 정보로 활용할 수 있는 요소나 특징을 데이터화 시킨다면 이는 감정 분석에 유의미한 지표로 활용이 가능하다.

본 연구에서는 사용자들의 과거 콘텐츠가 문맥적 정보로 활용이 가능한지 살펴보았다. 인스타그램에서 사용자들이 일관된 감정과 분위기를 연출하기 위해 비슷한 콘텐츠를 게시하는 경향이 있다. 이러한 경향이

실제로 존재하는지 검증해봄으로써 다른 콘텐츠들이 문맥적 정보로 활용할 수 있는지 알아볼 수 있다. 이를 위하여 사용자들은 자신의 게시글들에 비슷한 감정을 표현할 것이라는 가정을 하여 한 사용자 개인의 감정의 유사성과 사용자 간의 감정의 유사성을 비교하는 실험을 진행하였다. 먼저 4개의 감정에 따른 사용자들을 임의로 선별하여 각각의 유사성을 탐구해보았고, 이 결과는 사용자의 기존 콘텐츠가 감정 분석의 문맥적 정보로 활용할 수 있음을 제시한다.

### 2. 관련 연구

SNS에서 사용자들의 감정을 분석하기 위한 많은 방법론들이 제시되고 연구들이 진행되어 왔다. 기존에는 n-gram [1], distant supervision [2], 극성 분류에 나타나는 어휘들을 활용 [3]하여 감정을 분석 및 분류를 해왔다. 하지만 SNS 상에서는 인터넷 은어를 사용한다거나 비문 등을 많이 사용할 뿐만 아니라 짧게 올리는 경우도 많아 분석에 어려움이 있었다. 그래서 Wang X. 등은 해시태그를 활용하여 분석을 하였고, 이는 트위터 상에서 사용자들이 감정표현을 위해 주로 사용한다는 결과가 있다 [4].

한편으로 더 정확한 감정 분석을 수행하기 위하여 문맥적 정보를 활용하기도 한다. Saif H. 등은 SentiCircle이라는 문맥적 시맨틱을 만들어 트위터에서 더 효과적인 감정분석 방식을 제안했다 [5]. Agarwal B. 등은 상식적 지식을 활용하여 특정 분야의 고유한 특징을 추출한 후 문맥적 극성 어휘를 정의하여 감정

분석을 수행하였다 [6].

본 연구에서는 이전 연구들을 활용하여 SNS에서 문맥적 정보가 감정 분석에 긍정적으로 작용함을 알고 해시태그가 문맥적 정보로써 활용될 수 있는지 분석해보았다.

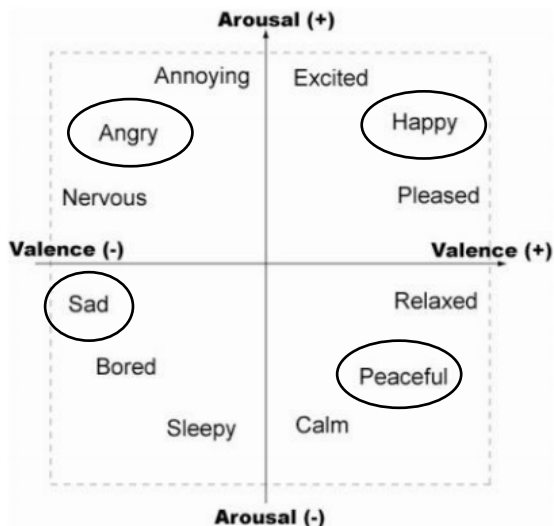


그림 1. Thayer의 감정 모델과 대표 감정

### 3. 실험 방법

#### 3.1 감정 분류 프레임워크

본 연구에서는 좀 더 깊이있는 탐구를 위해 기존 감정 분석과 달리 여러 형태의 감정을 분류해서 실험을 진행하였다. 감정과 키워드 분류는 [7]의 연구를 활용하였다. 이 프레임워크에서는 Tyayer의 감정 모델을 활용하여, 감정 분류의 정확성을 높이고 각 항목들의 독립성을 위하여 대표 4가지 감정인 행복(Happy), 분노(Angry), 슬픔(Sad), 평화로움(Peaceful)을 추출하였다. 각 감정들의 차원은 [그림 1]에 나타나있다. 각 감정들의 키워드는 전처리 과정을 수행하여 콘텐츠 내에 빈도수가 높은 키워드를 수집하였고 이는 [표 1]에 나타나있다.

#### 3.2. 데이터 수집 및 분석 방법

본 실험에서는 각 감정마다 대표 사용자를 뽑아 해당 사용자 개인의 감정 유사도(Intra-clustering sentiment similarity)와 사용자간의 감정 유사도(Inter-clustering

표 1. 감정 항목 [7]

주 감정	감정 키워드
Happy	cute, beautiful, adorable, funny, healthy, live, best, sweet, good, fit, hot
Angry	sad, cute, black, funny, honest, happy, mad, suicidal, dark, skinny, upset, bad, white, unhappy, anorexic, thin, fat, dead, hungry, bulimic, sweet, stupid, young, bloody
Peaceful	beautiful, happy, green, blue, sunny, quiet, natural, colorful, cute, gorgeous, square, grateful, normal, sad, thankful, live, scenic, good, health
Sad	suicidal, fat, happy, anorexic, black, alternative, cute, dead, dark, pale, skinny, bulimic, rad, white, thin, true, unhappy, empty, sorry, emotional, bipolar, bad, aesthetic, stupid, funny, pathetic, anxious, mad, soft

sentiment similarity)를 비교해보았다. 개인의 감정 유사도를 통해 맥락적 정보로 활용가능한지 살펴보고, 이 유사성이 감정이 아닌 다른 요소들 때문에 생겼을 수도 있기에 사용자간의 감정 유사도를 통해 이 요인이 감정에 의해서 발생했는지 알아보았다.

먼저 [표 1]에서 각 감정마다 대표 키워드(happy, angry, peaceful, sad)와 임의로 2개의 키워드를 선정하여 해당 키워드를 해시태그로 동시에 활용하여 콘텐츠를 올린 사용자 총 10명을 뽑았다. 행복을 나타내는 사용자 3명, 분노를 나타내는 사용자 2명, 평화로움을 나타내는 사용자 2명, 슬픔을 나타내는 사용자 3명을 추출하였다.

감정 형용사 벡터를 만들기 위해서 각 사용자별 최근 10개의 콘텐츠의 해시태그를 추출하였다. 추출한 해시태그를 POS tagging하여 형용사의 정보로만 이루어진 집합을 구성하고, [표 1]에서 제시한 키워드가 아니면 전부 제거하여 감정 형용사 벡터를 만들었다.

개인의 감정 유사도(Intra)를 측정하기 위해 앞서 만든 사용자별 감정 형용사 벡터와 해당 사용자의 가장 최근 콘텐츠의 해시태그를 이용하여 만든 벡터를 활용하였다.

사용자간의 감정 유사도(Inter)를 비교하기 위해서 각 사용자간의 감정 형용사 벡터 간의 유사도를 측정하였다. 같은 감정 그룹에서 도출된 사용자들의 유사도와 다른 감정 그룹에서 도출된 사용자들의 유사도를 비교해보았다. 유사도를 측정하는 지표는 코사인 유사도를 활용하였다.

### 4. 실험 결과

#### 4.1. Intra-clustering sentiment similarity

앞서 뽑았던 사용자들의 각각의 감정의 유사도가 얼마인지 보기 위해서 가장 최근의 콘텐츠들과 나머지 콘텐츠들의 감정의 유사도를 비교하여 [표 2]에 나타내었다. 최근 콘텐츠는 정보가 biased 될 위험성과 outlier의 위험성을 고려하여 최근 5개의 콘텐츠를 골랐다.

유사도가 전반적으로 높은 값을 나타내었다. 4번째 사용자와 6번째 사용자가 유사도가 낮게 나왔는데, 각 사용자를 들어가보니 4번째 사용자는 최근 5개의 게시물에 태그를 거의 달지 않았고 6번째 사용자는 인터넷 은어로 태그를 많이 달아서 POS tagging이

원활히 이루어지지 않아 분석시 데이터가 많이 누락이 되었다는 점을 확인할 수 있었다. 이 결과를 통해서 전반적으로 사용자는 자신의 인스타그램 계정에 비슷한 감정을 지속적으로 표현하고 있다는 것을 유추할 수 있다.

표 2. 사용자 10명의 개별 감정 유사도

사용자	유사도	사용자	유사도
User 1	0.99828256	User 6	0.05571821
User 2	0.92203409	User 7	0.98716738
User 3	0.90270757	User 8	0.92676696
User 4	0.11952286	User 9	0.89325960
User 5	0.99998937	User 10	0.95546348

#### 4.2. Inter-clustering sentiment similarity

사용자 간의 감정 유사도를 보기 위해 각 사용자별 감정 유사도를 비교하여 [표 3]에 나타내었다.

표 3. 사용자간 감정 유사도

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10
U1	1	.65	.64	.08	.00	.23	.34	.16	.27	.00
U2	.65	1	.56	.16	.00	.11	.62	.29	.00	.00
U3	.63	.56	1	.24	.00	.17	.30	.31	.39	.04
U4	.08	.16	.24	1	.35	.14	.10	.70	.53	.73
U5	.00	.00	.00	.35	1	.02	.00	.51	.24	.48
U6	.23	.11	.17	.14	.02	1	.51	.09	.10	.07
U7	.34	.62	.30	.10	.00	.51	1	.20	.00	.00
U8	.16	.29	.31	.70	.51	.09	.20	1	.55	.71
U9	.27	.00	.39	.53	.24	.10	.00	.55	1	.54
U10	.00	.00	.04	.73	.48	.07	.00	.71	.54	1

음영이 들어간 부분이 같은 그룹간의 감정 유사도고, 그렇지 않은 부분이 다른 그룹 간의 감정 유사도이다. 수치에서 확인할 수 있듯이, 같은 그룹 간에는 유사성이 높게 나오는데 반해 다른 그룹 간의 유사성이 상대적으로 낮게 나오는 것을 확인할 수 있다. 그 중 4번째 사용자는 분노의 항목으로 검색하였지만 슬픔 위주로 콘텐츠를 게시했던 것으로 확인했다. 전반적으로 다른 감정그룹이라도 긍정적인 감정인 행복(U1~U3)과 평화로움(U6, U7), 부정적인 감정인 분노(U4, U5)와 슬픔(U8~U10)이 상대적으로 높은 유사성을 나타냈다. 이런 결과들을 종합하면 4.1의 결과가 감정에 의해서 유사성이 생긴다고 유추할 수 있다.

#### 5. 결 론

본 연구에서는 인스타그램에서 사용자들이 일관된 감정과 분위기를 연출하기 위해 비슷한 콘텐츠를

게시하는 경향이 문맥적 정보로써 활용 가능할지 실험해보았다. 개개인의 사용자의 감정 유사도를 통해서 서로 관련이 있음을 발견하였고, 이 유사성이 사용자간 감정 유사도를 통해 사용자가 기본적으로 나타내고자 하는 감정에 표현에 의해서 생기는 것임을 확인할 수 있었다. 아직 연구 초기 단계라 많은 표본을 확보하지 못하였기에, 추후에는 더 많은 표본을 확보하여 본 제안의 신뢰도를 높일 예정이다. 또한, 해시태그가 아닌 다른 SNS의 다른 특징을 활용하여 다양한 SNS에서 과거 콘텐츠를 통한 문맥적 정보를 얻을 수 있는 방법을 추후 연구해 볼 예정이다.

#### 6. 참고 문헌

- [1] Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, 40(16), 6266–6282.
- [2] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12).
- [3] Speriosu, M., Sudan, N., Upadhyay, S., & Baldridge, J. (2011, July). Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First workshop on Unsupervised Learning in NLP* (pp. 53–63). Association for Computational Linguistics.
- [4] Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011, October). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 1031–1040). ACM.
- [5] Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, 52(1), 5–19.
- [6] Agarwal, B., Mittal, N., Bansal, P., & Garg, S. (2015). Sentiment analysis using common-sense and context information. *Computational intelligence and neuroscience*, 30.
- [7] Nam, M., Lee, E., & Shin, J. (2015). A Method for User Sentiment Classification using Instagram Hashtags. *Journal of Korea Multimedia Society*, 18(11), 1391–1399.