

트위터 타임라인을 이용한 감정 상태 분석

김종인*, 김경록**, 문남미**

*호서대학교 컴퓨터공학과

**호서대학교 벤처전문대학원 융합공학과

e-mail:kji6252@naver.com

The Emotional State Analysis using Twitter Timeline

Jong-In Kim*, Kyung-rok Kim**, Nammee Moon**

*Dept of Computer Engineering, Hoseo University

**Dept of Convergence Engineering, GSV Hoseo University

요 약

최근 스마트 기기와 SNS가 대중화가 되면서, 이를 기반으로 사용자의 감정 분석을 통한 추천 연구가 활발하게 진행되어 지고 있다. 본 논문에서는 감정표현단어 500여개와 이모티콘을 활용하여 감정 범주를 9가지로 분류하고, 트위터광장의 여러 유명인사 트위터 타임라인의 텍스트를 가져온 후, 감정상태 분석으로 감정 범주를 카운트 한다. 이를 통해, 한사람의 감정상태의 수치를 나타내고, 추후 음악, 음식, 문화 활동 등을 다양하게 추천할 수 있는 감정상태분석모듈을 설계 및 구현한다.

1. 서론

최근 인터넷 뿐만 아니라 스마트 기기의 폭넓은 보급으로 언제 어디서나 트위터(Twitter), 페이스북(Facebook)과 같은 SNS(Social Network Service)에 접속할 수 있게 되었다. 이를 통해, 사람들의 개인 생활이나, 정치적 사건, 영화, 제품 등 다양한 분야에 대한 생각이나 느낌을 표현할 수 있게 되었다. 개인의 일상을 소소하게 남기며 지인들과 근황을 주고받는 매개체가 되기도 하며, SNS는 실시간으로 대중의 생각을 반영하게 되었다[1][2].

본 논문에서는 기존 연구[3][4]를 참조하여 500여개의 감정 표현단어들과 이모티콘(Emoticon)을 9가지의 감정 범주로 나눈 뒤 트위터광장에서 유명한 트위터사용자의 타임라인 텍스트를 끌어와 감정상태분석 모듈을 통하여 감정상태를 알아보고자 한다.

2. 관련연구

2-1. 빅데이터 활용

빅 데이터(Big Data)란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다[5].

대규모 글로벌 IT 기업들은 이에 따라 한국 시장에 빅데이터 솔루션을 출시하면서 기업 경영에 초점을 맞춰 제품 홍보와 교육을 진행해 왔다. 하지만 최근에는 빅데이터가 공공 서비스 부문에 도입되면서 기업 경영 뿐 아니라 전 사회적으로 활용될 수 있는 광의의 개념으로 인식되는 분위기다. 해외에서는 이미 빅데이터와 시스템통합이 결합된 새로운 공공 서비스 모델이 제시되며 시민들에게 양질

의 서비스를 제공하고 있다. 빅데이터 솔루션이 공공기관의 정보화 시스템의 두뇌역할을 하며 행정서비스의 질을 개선하는데 중요한 역할을 하고 있는 것이다[6].

2-2. 트위터기반 감정 분석

국내에서도 트위터를 기반으로 감정분석을 하는 연구가 다양해졌다. SNS를 분석 하여 사람의 감정 상태를 분류하는 연구들이 진행 되고 있다. 감정범주를 Happy, Sad, Angry, Excited로 정하고 영어로 구성된 트위터를 대상으로 타임라인을 끌어와 감정상태를 나타내 음악을 추천해 주었다[7]. 다른 연구에서는 기계학습을 기반으로 다양한 자질들(이모티콘 극성, 리트윗 극성, 사용자 극성, 대체 어휘)사용하고 실험하여 긍정 부정으로 나타내었다[4]. 대부분은 긍정 부정으로 감정을 극성으로 분류하는 연구가 많다.

본 연구에서는 기존의 긍정 부정을 넘어서 다양한 감정을 기반으로 한사람의 트위터 타임라인을 불러들여 텍스트를 분석하고 그 분석한 결과를 9가지 감정상태로 나타내고자 한다.

2-3. 감정상태 분류

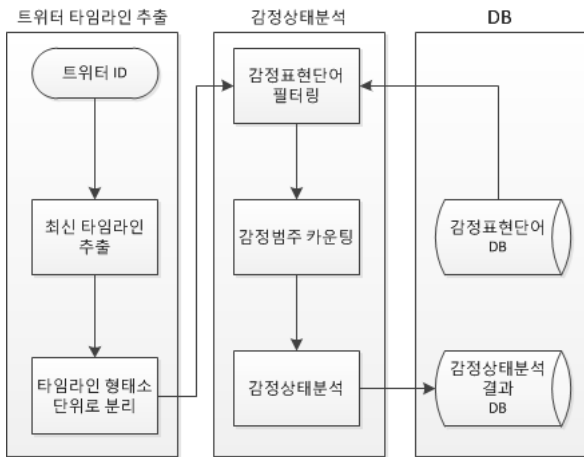
유명 트위터 사용자들의 감정을 분석하기 위해 먼저 일상생활에서 빈도 높게 사용하는 감정표현단어 500여개를 기쁨, 슬픔, 공포 등의 9가지 감정 표현 단어 범주로 분류한 후[3], 트위터에서 자주 사용되어지는 이모티콘들을 해당 분류 감정 범주에 포함 시켰다[4].

<표 1> 한국어 감정표현단어 예시 및 감정범주

극성	감정분류	키워드			
		가쁨	감격	감동	^^
긍정	기쁨	가쁨	감격	감동	^^
	흥미	갈구	갈망	감질	궁금
부정	슬픔	가련	간절	걱정	TT
	분노	모욕	미친	격분	무시
	통증	통증	끌치	아야	지끈
	혐오	가관	가소	경박	괴상
	공포	겁나	공포	긴장	무섭
중립	지루함	갑갑	귀찮	답답	따분
	놀람	갑작	경악	급작	기겁

3. 감정상태 분석 모듈 설계 및 구현

본 논문의 감정상태 분석 모듈은 (그림 1)과 같이 구성되어 있으며, 7단계로 통해 Twitter OpenAPI를 이용하여 원하는 대상의 트위터 타임라인을 끌어와 감정상태를 분석 한다.



(그림 1) 감정상태분석모듈

세부적으로, 트위터에 있는 타임라인을 끌어오기 위해 선 우선 Twitter OAuth인증을 받기 위해 PHP Twitter OAuth인증 라이브러리를 사용하여 Twitter OpenAPI사용 권한을 획득 한다. 감정분석이 필요한 대상의 ID를 입력 받아 GET/User_Timeline와GET/User_Show API에 ID를 넘겨주면 해당 ID의 타임라인과 개인정보를 가져와 PHP로 작성된 감정상태분석기를 통하여 분석을 하게 된다. 감정상태분석기는 최신타임라인 15일 이내와 200개로 한정하여 가져온 후, 한 개의 트위터 타임라인 텍스트를 형태소 단위로 분리한 후, DB(Data Base)에 있는 감정표현단어 들과 비교를 하게 된다. 비교된 값이 맞으면 해당 감정표현단어에 포함된 감정 범주에 카운트가 되어 +1이 되고 아니면 바로 다음 형태소를 비교하게 된다. 그렇게 마지막 타임라인까지 비교를 하게 되면 해당 ID의 감정상태와 개인정보를 DB에 저장하게 된다.

제안 모듈은 Windows7 x64 환경에서 Apache서버 및 Mysql연동을 하여 개발환경을 구성하고 주 소스 코드는

PHP를 활용하여 작성하였다.

3-1. 데이터 실험군

감정상태 분석을 수행하기 위해, 국내 트위터 서드-파티(Third-Party)사이트인 트위터광장(www.twtagora.com)에서 트위터 활동이 활발한 10명의 트위터 ID를 획득하여 실험을 진행하였다. 트위터광장에는 사용자를 연예인, 정치인, 기업인, 교수, 기자로 분류되어 있으며, 연예인은 코미디언, 배우, 가수, 아나운서로 분류되어 있다[8]. 본 논문에서는 연예인 그룹과 그 외의 지식인 그룹으로 분류하여, 감정상태 분석을 수행하였다.



(그림 2) 트위터 감정분석 실험군

3-2. 실험 결과

본 논문에서는 앞서 설계 및 구현한 감정상태 모듈을 가지고 트위터광장에서 유명인사들 각 분류(코미디언, 정치인 등등)에 10명씩을 실험대상으로 선정하여 분석하였으며, 그 결과는 <표 2><표 3>과 같다. 대체로 트위터 사용자들은 기쁨 감정을 표현하는데 트위터를 이용하고 있으며, 차이점으로는 지식인그룹이 연예인그룹보다 분노와 슬픔 감정이 더 크다는 것을 파악할 수 있다.

<표 2> 연예인그룹의 감정상태

직업 감정	연예인				
	코미디언	배우	가수	아나운서	연예인합
기쁨	555 (86.45%)	138 (73.80%)	813 (79.55%)	291 (79.51%)	1797 (81.06%)
흥미	4 (0.62%)	2 (1.07%)	5 (0.49%)	4 (1.09%)	15 (0.68%)
슬픔	43 (6.70%)	27 (14.44%)	58 (5.68%)	34 (9.29%)	162 (7.31%)
분노	20 (3.12%)	15 (8.02%)	133 (13.01%)	23 (6.28%)	191 (8.62%)
통증	1 (0.16%)	0	1 (0.10%)	3 (0.82%)	5 (0.23%)
혐오	0	1 (0.53%)	0	2 (0.55%)	3 (0.14%)
공포	6 (0.93%)	2 (1.07%)	5 (0.49%)	3 (0.82%)	16 (0.72%)
지루함	6 (0.93%)	0	2 (0.20%)	2 (0.55%)	10 (0.45%)
놀람	7 (1.09%)	2 (1.07%)	5 (0.49%)	4 (1.09%)	18 (0.81%)

<표 3> 지식인그룹의 감정상태

직업\감정	지식인				
	정치인	기업인	교수	기자	지식인합
기쁨	384 (55.41%)	493 (65.38%)	247 (43.33%)	247 (59.95%)	1371 (56.44%)
흥미	5 (0.72%)	24 (3.18%)	4 (0.70%)	9 (2.18%)	42 (1.73%)
슬픔	89 (12.84%)	86 (11.41%)	56 (9.82%)	32 (7.77%)	263 (10.83%)
분노	169 (24.39%)	122 (16.18%)	221 (38.77%)	95 (23.06%)	607 (24.99%)
통증	3 (0.43%)	3 (0.40%)	5 (0.88%)	3 (0.73%)	14 (0.58%)
혐오	10 (1.44%)	8 (1.06%)	2 (0.35%)	6 (1.46%)	26 (1.07%)
공포	17 (2.45%)	4 (0.53%)	17 (2.98%)	8 (1.94%)	46 (1.89%)
지루함	5 (0.72%)	1 (0.13%)	3 (0.53%)	1 (0.24%)	10 (0.41%)
놀람	11 (1.59%)	13 (1.72%)	15 (2.63%)	11 (2.67%)	50 (2.06%)

4. 결론

본 논문에서는 감정표현단어 500여개와 트위터에서 자주 사용하는 이모티콘을 9가지의 감정 범주에 묶어서 PHP로 감정상태분석기를 구현 하였고 Twitter Oauth인증을 통해 Twitter OpenAPI의 사용권한을 획득하여 원하는 사람의 트위터 타임라인을 끌어와 감정상태분석기를 통해 감정상태를 분석해 보았다.

트위터광장에서 분류한 그룹(연예인, 지식인) 분류(코미디언, 정치인 등등)의 분석을 통해 연예인그룹에 있는 트위터 사용자들은 대체로 기쁨과 분노 슬픔 순으로 감정상태를 파악할 수 있었으며 지식인그룹은 연예인그룹에 비해 기쁨이 조금 더 낮고 분노와 슬픔 감정상태가 많이 나타나 있는 결과를 얻었다.

그러나 활발한 트위터 활동이 있는 대상에 한해 감정상태를 파악할 수 있다는 한계가 있어, 이를 극복하기 위해서는 다른 소셜 데이터와의 결합이 요구되어진다. 다양한 감정표현단어와 이모티콘들로 감정표현단어DB에 업데이트하면서 감성분석모듈의 정확성을 높이면 음악, 영화, 음식 등을 다양한 추천 시스템에 적용하여 사용자의 편의성과 기업의 광고 효과에 크게 기여할 것이라고 기대된다.

참고문헌

- [1] 이철성, 최동희, 김성순, 강재우 “한글 마이크로블로그 텍스트의 감정 분류 및 분석”, 한국정보과학회논문지 : 데이터베이스 40(3), pp.159-167, 2013.6
- [2] 전민아, 전상훈, 노승민, 황인준 “트위터 해시태그와 트윗 분석을 통한 음악 추천 기법”, 2013년 한국컴퓨터종합학술대회 논문집, pp.1257-1259, 2013.6
- [3] 손선주, 박미숙, 박지은, 손진훈 “한국어 감정표현단어의 추출과 범주화”, 감성과학, Vol. 15, No. 1, pp.105-120, 2012.3

[4] 홍초희, 김학수 “트윗 감정 분류를 위한 다양한 기계학습 자질에 대한 비교 연구”, 한국콘텐츠학회논문지, Vol. 12 No. 12, pp471-478, 2012.12.12.

[5] James Manyika, Michael Chui, “Big data: The next frontier for innovation, competition, and productivity”, McKinsey Global Institute, 2011.05

[6] 이성훈 “빅데이터 활용 현황”, 한국정보기술학회지 10(3), pp.51-54, 2012.9

[7] 최홍구, 황인준 “트위터 문서 분석을 통한 감정 기반의 음악 추천 시스템”, 한국정보과학회논문지, 제18권, 제11호, pp.762-767, 2012.11.

[8] 트위터광장, “www.twttagora.com”, 2013.09.23.최종접근