

Facebook 뉴스피드를 이용한 사용자 성향 분석

정윤상*, 김경록**, 문남미**

*호서대학교 컴퓨터공학과

**호서대학교 벤처전문대학원 융합공학과

e-mail: lospl@naver.com

The User Inclination Analysis Using Facebook Newsfeed

Yoon-Sang Jeong*, Kyung-rog Kim**, Nammee Moon**

*Dept of Computer Engineering, Hoseo University

**Dept of Convergence Engineering, GSV Hoseo University

요 약

최근 페이스북(Facebook), 트위터(Twitter) 등의 SNS(Social Network Service)의 사용자가 급격하게 증가하고 있다. SNS가 발달하면서 언제 어디서나 쉽게 자신의 위치, 현재의 감정들을 온라인상에서 공유한다. 이에 따라 사람의 감정을 표현하는 단어 100여개를 7가지 감정(기쁨, 흥미, 슬픔, 분노, 놀람, 지루함, 통증)으로 분류하였으며[1]. 이를 분석하기 위한 감정 표현 분석기 모듈을 설계하였다. 설계한 모듈을 사용하여 페이스북의 사용자 뉴스피드(News-Feed)를 분석하여 사용자의 성향을 분석하였다.

1. 서론

오늘날 SNS 사용자가 증가함에 따라 페이스북, 트위터 등을 이용하여 쉽게 자신의 현재 감정들을 표현하고 공유할 수 있게 되었다. 동시에 오프라인(Off-line)에서 구축한 관계가 온라인(On-line)에서도 구축이 되고 있다 [2][3]. SNS에서 소통되는 데이터는 개인적인 데이터부터 사회적인 데이터까지 다양하며, 이러한 관계 혹은 활동 데이터는 개인의 감정, 인식, 경험 표출 등 다양한 성격을 포함한다[3][4].

사용자들의 성향 및 관심분야가 다르므로 이러한 SNS에 있는 사용자의 데이터를 분석함으로써 사용자의 성향을 파악할 수 있다[3].

본 논문에서는 개방형 SNS 페이스북을 활용하여[5] 사용자들의 뉴스피드를 바탕으로 감정 표현 단어를 분석해 보고자 한다.

2. 관련연구

2-1. 감정표현단어의 추출과 범주화

사람의 감정은 사건에 대한 개인의 평가로 시작되며 이러한 감정은 주관적 경험, 생리반응과 행동, 감정표현 언어와 어조 등에서 나타난다. 사람의 감정을 나타내는 단어 표현 또한 왜곡이 가능하기에 감정 이해가 어려우나 사람간의 커뮤니케이션에 있어서는 가장 발달된 감정표현 수단이다[1].

Ekman은 감정표현의 단어를 ‘기쁨’, ‘슬픔’, ‘공포’, ‘분노’, ‘혐오’, ‘놀람’의 6가지로 감정표현 단어를 분류하였다. 또한 국내 연구자는 Ekman의 6가지 분류에 HCI에서 활용

용도가 높은 ‘흥미’, ‘지루함’, ‘통증’의 3가지 감정과 ‘중성’, ‘기타’를 추가하여, 총 11개 감정으로 분류를 하였다[5].

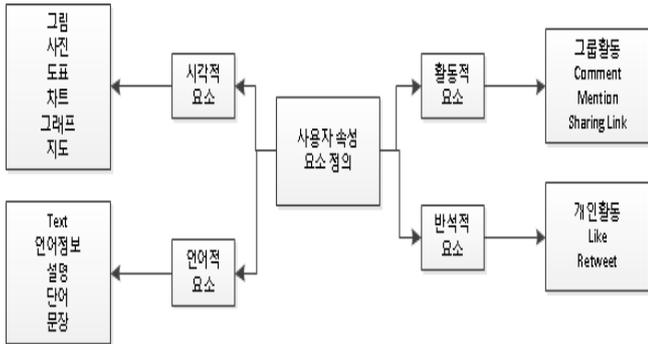
본 논문에서는 이러한 감정 분류를 기반으로 하였으나, ‘공포’라는 감정을 나타내는 단어는 ‘놀람’과 ‘혐오’와 유사한 성향을 가지고 있으므로 ‘공포’로 통합하여, 아래의 <표 1>과 같이 7가지의 감정으로 분류하였다.

<표 1> 감정표현 단어 분류

		키워드
긍정	기쁨	가뽀, 감격, 감동, 감미, 감사, 감흥, 경쾌, 근사, 기뻐, 깔깔, 깔깔, 키득, 하하, 호호, 히히, 따뜻, 방긋, 뿌듯, 사랑, 이쁘, 예쁘, 재미있다, 즐겁, 친숙, 편안, 행복, 께, 흥
	흥미	흥미, 궁금, 섹시, 탐나, 신비, 갈망, 감질, 충동
부정	분노	격노, 패썹, 분통, 복수, 개새끼, 개쌍년, 개자식, 격노, 격분, 그년, 그놈, 그자식, 나쁘, 니미, 무시, 버럭, 복수, 분개, 빌어먹을, 빼졌, 빼치, 싸가지, 쌍년, 쌍놈, 씹새끼, 병신, 짜증, 질투
	슬픔	슬퍼, 울먹, 침울, 걱정, 고독, 공허, 괴롭, 구슬, 그리워, 근심, 기구, 기운없다, 낙담, 낭패, 눈물, 망연, 비참, 서럽, 서운, 설섭, 속상, 씩씩, 슬퍼, 아련, 아휴, 애석, 외롭, 우울, 울상, 울적, 으앙, 절규, 절망, 절박, 좌절, T, TT
	공포	오싹, 으스스, 공포, 섬뜩, 겁쟁이, 긴장, 두렵, 무섭, 무시무시, 불안, 살벌, 섬뜩, 소름, 스산, 아찔, 깜짝, 아뽀싸, 쇼킹, 갑작, 경악, 급작, 기겁, 기막, 기묘, 기이, 기절초풍, 놀람, 당혹, 당황, 뜨끔, 아이쿠, 아차, 앓, 얼떨떨, 으악, 헉, 흥측, 구역질, 가관, 거북, 경박, 피상, 피딱, 꼴불견, 느끼, 더럽, 망측, 부담, 역겹다, 정그럽다, 찝찝
	통증	지끈, 기진맥진, 골치, 아아, 아프
	지루함	갑갑, 귀찮다, 답답, 따분, 아우, 무료하다, 심심, 어색, 지긋지긋

2-2. SNS 데이터 기반의 사용자 유형 모델링

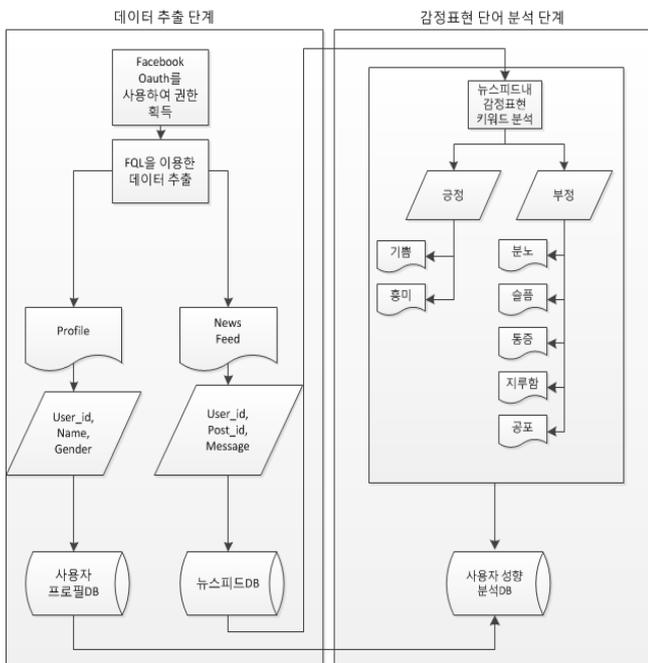
최근 SNS는 개인적인 데이터와 사회적인 데이터 등 여러 가지 데이터를 가지고 있다. 대표적 페이스북, 트위터에서 사용되는 SNS 데이터는 시각적인 요소, 언어적인 요소, 활동적 요소 그리고 반성적 요소로 나눌 수 있다.[4] 세부항목은 (그림 1)과 같다. 기존 관련연구를 바탕으로 사람의 감정표현의 방법중 가장 발달된 요소인 언어적 요소를 바탕으로 사용자의 성향을 분석하고자 한다.



(그림 1) 사용자 속성 요소 유형 맵

3. 감정표현 분석기 설계 및 구현

감정 표현 분석기 구성 및 처리 절차는 아래 (그림 2)와 같다.



(그림 2) 사용자 데이터 구축

먼저, 데이터 추출 단계에서는 OAuth를 통한 사용자 데이터 추출 권한을 획득한다. 허가받은 데이터를 페이스북 API인 FQL(Facebook Query Language)을 이용하여

사용자 프로필 데이터(User_id, Name, Gender)와 뉴스피드 데이터(User_id, Post_id, Message)를 추출하고 각각 사용자 프로필 DB(DataBase), 뉴스피드 DB에 저장한다. 이때 뉴스피드 데이터는 최근 작성한 50개의 뉴스피드를 기준으로 추출한다.

감정표현 단어 분석단계에서는 뉴스피드 DB에 있는 Message를 <표 1>의 감정표현 단어와 비교 분석하여, 감정표현 키워드와 일치하는 단어의 사용 횟수를 감정표현 분류별로 카운팅하여 사용자 프로필과 함께 사용자 성향 분석 DB에 저장한다.

위에서 설계한 감정표현 분석기는 i5-2500, 8GB RAM으로 구성된 시스템에 Window7 64비트 환경에서 구현되었다. 또한 Apache, PHP, Mysql를 이용하기 위해서 APMsetup7을 활용하였으며 모듈은 이클립스(Eclipse)를 이용하여 PHP와 JavaScript로 구현하였다.

실험은 페이스북 사용자들 중, 임의의 10명을 선정하여 진행하였고, 그 결과는 아래 <표 2>와 같다.

<표 2> 사용자의 감정 표현 단어 사용횟수

	긍정			부정			
	기쁨	흥미	슬픔	분노	공포	통증	지루함
1	134 (82.1%)	0	25 (15.2%)	2 (1.2%)	4 (2.4%)	0	0
2	121 (65.7%)	0	58 (31.5%)	2 (1.0%)	0	2 (1.0%)	1 (0.5%)
3	3 (7.1%)	1 (2.3%)	26 (61.9%)	3 (7.1%)	6 (14.2%)	0	3 (7.1%)
4	293 (57.0%)	3 (0.5%)	211 (41.0%)	11 (2.1%)	5 (0.9%)	0	1 (0.1%)
5	67 (80.7%)	0	8 (9.6%)	2 (2.4%)	3 (3.6%)	2 (2.4%)	1 (1.2%)
6	89 (33.5%)	2 (0.7%)	160 (60.3%)	11 (4.1%)	0	0	3 (1.1%)
7	9 (50.0%)	0	5 (27.7%)	0	3 (16.6%)	1 (5.5%)	0
8	151 (77.8%)	0	37 (19.0%)	0	4 (2.0%)	0	2 (1.0%)
9	212 (66.4%)	5 (1.5%)	85 (26.6%)	3 (0.9%)	7 (2.1%)	2 (0.6%)	5 (1.5%)
10	175 (79.9%)	10 (4.5%)	23 (10.5%)	3 (1.3%)	3 (1.3%)	2 (0.9%)	3 (1.3%)

실험에서 얻은 결과 값을 바탕으로 언어 사용 성향을 살펴보면, 기쁨을 표현하는 키워드의 사용횟수가 가장 많으며, 그 다음으로는 슬픔에 대한 키워드가 많이 사용되었다.

4. 결론

본 논문에서는 감정표현 분석기를 구현하기 위하여 PHP, JavaScript 및 페이스북 API를 활용하여 사용자 프로파일과 가장 최근 작성한 50개의 뉴스피드를 추출하였다. 또한, 감정을 표현하는 단어를 7가지로 분류를 한 뒤, 추출한 뉴스피드의 메시지를 감정 표현 단어들과 비교하여, 사용자의 감정 키워드 사용 횟수를 감정 분류별로 사용횟수를 DB에 누적하였다.

사용자들의 뉴스피드를 감정 표현 분석기를 통하여 분석을 한 결과, 대다수의 사용자들이 기쁨을 표현하는 단어를 가장 많이 사용하였으나, 사용자 3과 사용자 6의 경우에는 슬픔을 표현하는 단어를 많이 사용하였다. 이는 곧 사용자 3, 6은 최근 슬픈 감정을 느끼고 있음을 추측하는 근거가 될 수 있다.

추후, 다수의 실험 대상을 선정하여, 뉴스피드의 메시지뿐만 아니라, 코멘트(Comment), 좋아요(Like)와 페이스북에 남긴 이미지(Image), 동영상(Video) 등을 추가하여 분석한다면, 사용자의 최근 감정과 성향을 파악하고, 이를 기반으로 한 다양한 맞춤형 서비스 및 추천 시스템에 활용할 수 있을 것으로 기대된다.

참고문헌

- [1] 손선주, 박미숙, 박지은, 손진훈, “한국어 감정표현단어의 추출과 범주화”, 감성과학, Vol. 15, No. 1, pp.105-120, 2012.3
- [2] 설광수, 김정동, 심형남, 백두권, “소셜 네트워크 서비스 사용자 간의 친밀도 측정 기법 및 실험”, 정보과학회논문지, 제39권 제4호, pp. 335-341, 2012.8.
- [3] 박건우, 임선열, 이태양, 김정훈, 최웅철, “감성 단어 분석 모듈을 이용한 모바일 소셜 네트워크 서비스에서의 성향 판단 시스템 구현”, 한국정보과학회 학술발표논문집, 38(2D), pp.97-100, 2011.11.
- [4] 김경록, 문남미, “개인화 서비스를 위한 SNC 데이터 기반의 사용자 유형 모델링”, 2013년 한국방송공학회 하계 학술대회, 2013
- [5] 김준호, 은혜정, “인맥연결 형 네트워크 유형에 대한 사례분석”, 사이버커뮤니케이션 학보 28(4), pp.5-47, 2011.12,
- [6] 박지범, 임혜원, 김형주, “온라인 소셜 네트워크에서 사용자 간의 관계 강도 변화 추정 기법”, 정보과학회논문지, 데이터베이스 제40권 제1호, pp44-53, 2013.2