소셜네트워크 메시지에서 분류한 감성의 지도 표현

Map display of emotion classified from social networks

김진만, 임좌상

상명대학교 일반대학원 컴퓨터과학과, 상명대학교 소프트웨어대학 디지털미디어학부

Key words: *Social Network, Twitter, Data Visualization, Emotion Analysis*

1. 서론

소셜네트워크의 사용이 급증하면서, 이를 통해 의견/정서를 분석하는 연구가 발표되고 있다. 이런 새로운 미디어 이전에는 비교적 정형적인 리뷰를 대상으로 정서를 분석하였다. 즉, 리뷰에 포함된 제품이나 서비스에 대한 의견/정서를 이분법적으로 긍정-부정으로 나누어 요약하는 방식이었다. 반면, 소셜네트워크 메시지는 짧은 까닭에 줄임 말이나 이모티콘을 포함하는 비정형적인 특성을 보이고 있다. 그러나 소셜네트워크는 개인 간의 관계에서 교환되는 특성을 반영하고 있어 사회적 의견/정서를 이해하기 위한 연구가 필요하다.

최근에는 위치인식기능이 보편적으로 제공되고 있어, 정서를 장소에 따라 이해하려는 요구가 생겨나고 있다. 이러한 요구는 이미 오래 전부터 있었다. 예를 들어, 심리학에서는 사람의 행복을 지도에 나라별로 색으로 높고 낮음을 표현하는 연구가 있었다. 또한 사회, 경제, 문화 또는 정치적 관계를 지도 (Cartography)를 통해 시각화하는 연구가 있어왔다. 요즘에는 소셜네트워크에 위치기능이 추가되면서 이러한 데이터를 시각화하는 연구가 활성화 되고 있다. 본 연구에서는 소셜네트워크 메시지의 감성을 각성, 이완으로 분류하여, 지역별로 지도에 시각화하였다.

2. 관련연구

트윗은 보통 짧고 빠르게 쓰여지면서 비정형적인 특징을 보이지만, 많은 사람의 의견을 대량으로 포함하고 있다는 점에서 학문적으로 분석할 가치가 있다. O' Connor 는 미국 내 트위터 사용자를 대상으로 2008~2009 년까지의 트윗 약 10 억 개를 수집하였다. 수집된 트윗에서 소비심리와 정치적 성향을 분석하여 실제 설문조사기관에서 조사한 결과와 비교하였다[1]. Asur 와 Huberman 은 약 1.2 백만 트위터 사용자를 대상으로 2.89 백만의 트윗을 수집하고 이를 분석하여 2009~2010 년 사이 개봉된 24 개의 영화에 대한 박스오피스 수입을 예측하였다[2]. Hughes 와 Palen 은 2008 년 8~9 월 사이의 미국의 정당대회와 허리케인 재해와 관련하여 트위터에 어떤 정보가 확산되는지를 약 5.3 만 명의 사용자를 대상으로 약 14.7 만의 트윗을 수집하여 분석하였다[3].

지도를 이용한 데이터의 시각적 표현은 정보의 효과적인 전달 뿐만 아니라 사용자의 사고 프로세스에 가장 적합한 도구라 할 수 있다[4]. 따라서 소셜미디어 데이터와 같은 대량의 데이터를 지도에 나타냄으로 정보 자체의 시각적 형태보다는 대상의 상태와 패턴을 인식하여 지식을 형성할 수 있다. Nold 는 사람의 생리신호 데이터를 지도에 시각화 하여 사회, 경제, 문화의 관계를 표현하는 프로젝트를 진행하였다 [5]. Terraz 는 트위터 트윗을 지도에 3D 모델로 표현해주는 서비스를 제공하고 있다[6].

3. 시스템구현

3.1 시스템 환경과 프로세스

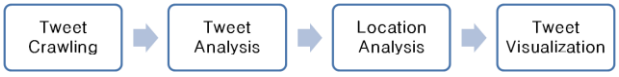

그림 1. 트윗의 분석 및 지도시각화 프로세스

그림 1 은 트윗에서부터 감성을 분류하고 지도에 표현하는 절차를 도식화하였다. 시스템은 웹으로 개발되었다. 즉, 클라이언트는 웹브라우저에서 구글맵을 구현하였다 (Google Map API V3). 서버는 리눅스 (Ubuntu 11.04), Python (Django 1.2.1 웹 프레임워크), CouchDB 로 구축했다. 데이터는 트위터에서 제공하는 API 를 이용하여 크롤링하였다. 다음은 각 단계별 설명이다.

(1) Tweet Crawling: 트윗을 수집하는 단계로서, 우선 트위터 id 를 너비우선으로 추출하였다. 즉 임의의 사용자를 선택하고, 그 팔로워를 대상으로 무작위로 id 를 추출하였다. 추출된 id 정보는 screen name 으로 변환되며 필터링을 수행해 트윗을 비공개로 설정한 사용자의 screen name 은 제거하였다. 다음으로 수집된 사용자들의 screen name 을 입력으로 사용자 정보와 트윗을 DB 에 저장했다.

(2) Tweet Analysis: 트윗의 감성을 분류하기 위해 러셀의 감성단어(28 개)를 사용하였다. 이 단어는 2 차원

감성모델로서 감성을 4 가지로 (각성, 이완, 쾌, 불쾌)분류할 수 있지만, 우리는 2 가지로 분류하였다. 즉, 각성에 흥분, 즐거움, 고통을, 이완에는 불행, 우울, 만족을 포함하였다. 각 트윗에 포함된 감성단어의 빈도에 따라 각성과 이완으로 분류되고 빈도가 동일한 경우, 중립으로 분류하였다.

(3) Location Analysis: 사용자 위치를 추출하여 위도, 경도 데이터로 변환한다. 위치정보는 트위터 사용자의 프로필 정보에서 읽어왔다. DB 로부터 사용자의 위치정보를 읽어 서울을 포함한 6 개 광역시명과 제주도를 포함한 전국 8 도에 속한 도 및 군 명까지를 한글과 영문명으로 인식하여 GEO 데이터 (예: 서울 126.9779, 37.5665)로 변환했다.

(4) Tweet Visualization: 이 단계는 트윗의 감성을 지도에 표현했다. Google Map API 를 사용하여 지도를 브라우저에 출력하고 트윗의 위치를 지도에 출력하기 위해 DB 로 부터 위/경도 데이터를 읽었다. 트윗의 위치를 heatmap 기법을 사용하여 지도 위에 오버레이 하였다.

3.2 사 례

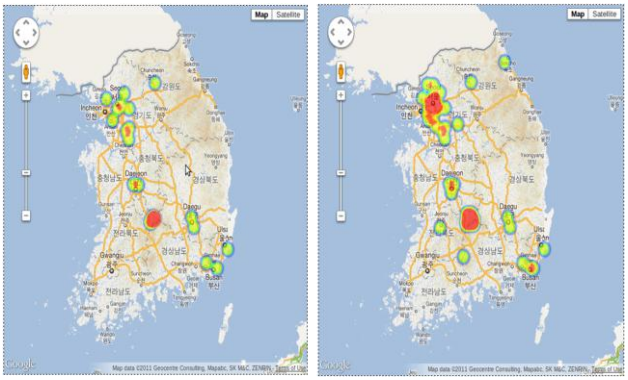

그림 2. 트윗 감성의 지역별 시각화
(좌: 이완, 우: 각성)

본 연구에서는 트윗의 감성을 분류하여 자동으로 지도에 시각화 하는 시스템을 구현하였다. 임의의 트위터 사용자를 선정하고, 이를 시작으로 시스템을 구동했다. 시스템은 그 사용자의 팔로워 중 1000 명을 무작위로 추출하고 트윗 비공개 설정한 사용자들을 필터링하여 최대 200 개 트윗을 크롤링 했다 (2011 년 11 월 이전). 그 결과 필터링을 통해 984 명의 팔로워로부터 총 110,819 개의 트윗이 수집 되었다 (트위터 API 사용제한으로 인해 수집에 약 240 분 소요). 다음으로 시스템은 트윗 내 포함되어 있는 감성단어의 빈도를 계산하여 트윗의 감성을 각성, 이완, 중립 (1, −1, 0)으로 분류했다. 사용자 위치정보는 위도, 경도 데이터로 변환했다. 위치 데이터는 ±0.09 도의 차이로 위/경도 값이 랜덤 하게 추가되었다. 그 이유는 트위터 사용자의 위치정보로 사용자 트윗을 지도에 나타내면 한 위치에 트윗이 겹쳐 보이기 때문에 시각화를 위해 트윗들을 분산시켰다. 마지막으로 트윗의 감성과 위치를 직관적으로 보여주기 위해 heatmap 기법을 사용하여 구글 지도에 오버레이 시켰다. 그 결과, 그림 2 와 같이 트윗의 감성이 분류되어 지도 위에 지역별로 표현되었다.

4. 결 론

본 연구는 소셜네트워크 메시지의 감성을 분류하여 지도에 시각화하였다. 110,819 개의 트윗을 수집하고 러셀의 감성 단어를 중심으로 감성을 분류하였고, 이를 heatmap 기법을 이용하여 지도에 시각화 하였다. 따라서 본 연구는 비정형적인 소셜네트워크의 메시지의 감성을 분류하고 이를 지도에 표현함으로 국내의 지역별 감성을 파악할 수 있어 사회 지표 측정의 도구로 활용될 수 있다.

향후에는 소셜네트워크 메시지의 감성분석에 효과적인 분석방법과 이의 표현에 알맞은 데이터 시각화 기법에 대해 추가적인 연구가 필요하다.

감사의 글

이 연구는 한국문화콘텐츠진흥원의 2011 년도 문화콘텐츠 산업기술지원사업의 연구결과로 수행되었음.

참고문헌

[1] B. O'Connor, R. Balasubramanyan, B. R. Routledge and N. A. Smith (2010). From tweets to polls: Linking text sentiment to public opinion time series. *International AAAI Conference on Weblogs and Social Media*, 122−129.

[2] S. Asur and B. A. Huberman (2010). Predicting the Future with Social Media. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*.

[3] A. L. Hughes and L. Palen (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management, 6(3−4)*, 248−260.

[4] A. M. MacEachren and M. J. Kraak (2001). Research challenges in geovisualization. *Cartography and Geographic Information Science, 28(1)*.

[5] C. Nold (2006). *Greenwich Emotion Map*. available at http://www.emotionmap.net

[6] J. Terraz (2010). *GeocodEarth*. available at http://www.geocodearth.com