

# 트위터 메시지의 감성 분류

## Classification of twitter messages into positive and negative emotions

박민우, 김진만, 이의철, 임좌상

상명대학교 컴퓨터과학부, 상명대학교 일반대학원, 국가수리과학연구소, 상명대학교 디지털미디어학부

**Key words:** 소셜네트워크서비스, 트위터, 감성분류

### 1. 서론

소셜네트워크서비스 (SNS: Social Network Services)가 많이 사용되면서, 여기서 의미 있는 정보를 분석하려는 요구가 증가하고 있다. 그러나 그 메시지가 비정형적인 특징을 띠고 있어 분석이 어렵다. 최근 트위터를 대상으로 사회 트렌드를 읽고 예측하는 연구가 수행되고 있으며 이미 서비스화 되고 있다. 다만 한글로 작성된 트위터는 아직 연구가 부족한 상황이다. 본 연구에서는 한글 트위터를 대상으로, 감성을 긍정, 부정으로 분류하는 인터페이스를 구현하였다.

### 2. 관련 연구

#### 2.1. 소셜네트워크 메시지의 감성

우리는 SNS 를 통하여 자신의 감성을 나타내는 것에 익숙하다. 한국(208 명)과 중국 (123 명)의 SNS 사용자를 대상으로 한 설문연구[1]는 이를 잘 설명하고 있다. 즉, 사회 관계에 있어서 자신의 감성이나 의견을 표현하는 것이 가장 중요한 것으로 나타났다. 또한 트위터와 페이스북 사용자 190 명을 대상으로 한 연구[2]에서도 자신의 의견이나 감성을 표현하기 위해 SNS 에 주로 접속하는 것으로 나타났다.

#### 2.2. 정서 분석 연구

정서 분석연구는 상품평이나 기사와 같은 정형적인 리뷰를 대상으로 정서를 분석하였다. 최근에는 트위터와 같은 비정형적 데이터를 대상으로 연구가 진행 중이다. Riloff and Wiebe 는 미국 해외방송 정보 서비스의 기사와 뉴스를 대상으로 Bootstrapping 통계 기법을 사용하여 문장 작성시 사람들이 쓰는 구문 표현을 주관관 객관으로 분류하였다[5]. Mao and Lebanon 은 249 개의 영화 감상 평을 대상으로 베이지안과 SVM 두 분류기법을 사용하여 영화의 감성을 5 가지 (아주 좋음, 좋음, 중립, 나쁨, 아주 나쁨)로 분류하였다[6]. 2440 개의 문서 (기사, 영화평,

상품평)를 대상으로 한 연구[3]에서는 SVM 을 사용하여 문서의 감성을 긍정, 부정으로 분류하였다. Huang and Bhayani 는 트위터를 대상으로 베이지안 분류기법과 NLP (Natural Language Processing)을 사용하여 트윗의 감성을 긍정, 부정으로 이분하였다[4].

### 3. 연구 방법

#### 3.1 시스템 프로세스

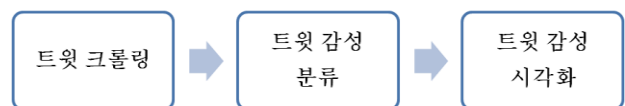


그림 1. 트위터 감성 분류

그림 1 은 트윗을 크롤링하여 감성을 분류하고 시각화하는 절차를 나타낸 것이다. 시스템은 웹으로 개발되었다. 서버는 Tomcat 6, MySQL 로 구축했다. 클라이언트는 HTML, JavaScript, CSS 로 구현하였다. 감성 분류를 위해 LIBSVM [7]을 사용하였으며 Twitter API 를 이용하여 트윗을 크롤링 하였다. 다음은 각 단계별 상세 설명이다.

(1) 트윗 크롤링: Twitter API 를 사용하여 트윗을 수집하여 DB 에 저장하였다. 임의의 트위터 사용자를 선택하여, 팔로워 id 를 무작위로 추출했다. id 는 screen name 으로 변환했다. screen name 정보를 사용하여 트윗을 크롤링하고 DB 에 저장했다.

(2) 트윗 감성 분류: 트윗의 형태소를 추출하고 감성단어를 사용하여 벡터로 변환한다. 감성단어는 본 연구의 사전연구에서 임의의 트윗 1000 개로부터 긍정, 부정 단어를 주관적으로 추출되었다. 이를 바탕으로 SVM 감성분류모델을 학습시키고 수정된 모델을 적용하였다. 트윗을 벡터모델로 만들어 SVM 모델에 입력하고 감성을 분류했다 (긍정, 부정, 중립 (1, -1, 0)).

(3) 트윗 시각화: DB 에서 트윗과 감성 값을 읽어와 웹브라우저에 감성에 따라 트윗의 배경화면 색을 다르게 표현했다 (긍정: 적색, 부정: 청색, 중립: 황색).

### 3.2 트위터 감성분류 사례



그림 2. 트위터 감성분류 인터페이스

본 연구에서는 트위터 메시지의 감성분류를 위해 시스템을 구현했다. 임의의 트위터 사용자 screen name 을 선정하여 시스템에 입력하였다. 시스템은 트위터 크롤링 단계를 수행했다. 100 명의 팔로워 id 가 무작위로 추출되어 screen name 으로 변환되었다. 이어서 변환된 screen name 정보를 사용하여 총 20,470 개의 트윗을 수집했다. 다음으로 시스템은 트윗 감성 분류 단계를 수행했다. SVM 기법을 사용하기 위해 형태소 분석기에서 트윗의 단어를 분리하고 감성단어를 이용하여 벡터데이터로 변환했다. 트윗 벡터데이터는 SVM 모델에 입력되어 감성 분류가 되었다. SVM 모델은 앞서 언급한 감성단어 추출에 사용된 트윗 1000 개로부터 학습을 거쳐 만들어 졌다. 마지막으로 시스템은 트윗의 감성을 시각화했다. 클라이언트는 서버에 트윗 텍스트와 감성 정보를 요청하고 웹 브라우저에 트윗의 텍스트와 감성을 출력했다. 그 결과, 그림 2 와 같이 화면에 트윗 각각이 가지고 있는 감성에 따라 배경화면이 다른 색으로 표현되었다.

## 4. 연구 결과

본 연구에서는 트위터 메시지인 트윗의 감성을 분류하였다. 100 명의 트위터 사용자로부터 20,470 트윗을 수집하고 SVM 분류기법을 사용하여 각 트윗의 감성을 긍정, 부정으로 이분화하여 색으로 표현 했다.

향후에는 단일 트윗의 감성 분류에서 나아가 연결된 트윗의 감성을 분석하여 트윗 작성자의 감성상태 유추하는 추가적인 연구가 필요하다.

### 감사의 글

This work was supported by the NAP (National Agenda Project) of the Korea Research Council of Fundamental Science & Technology.

### 참고문헌

- [1] 고상민, 황보환, 지용구 (2010). 소셜네트워크서비스와 온라인 사회적 자본 : 한국과 중국 사례를 중심으로. *한국전자거래학회지*, 15(1).
- [2] 홍삼열, 오재철 (2010). 트위터와 페이스북 사용자 접촉속인 비교분석. *한국인터넷정보학회 추계학술발표대회 논문집*, 11(2).
- [3] 황재원, 고영중 (2010). 감정 단어의 의미적 특성을 반영한 한국어 문서 감정분류 시스템. *정보과학회논문지, 소프트웨어 및 응용*, 27(4).
- [4] A. Go, L. Huang and R. Bhayani (2009). Twitter sentiment analysis. *Final Projects from CS224N for Spring 2008/2009 at The Stanford Natural Language Processing Group*.
- [5] E. Riloff and J. Wiebe (2003). Learning extraction patterns for subjective expressions. *Proceedings of the Empirical Methods in Natural Language Processing*, 10, 105-112.
- [6] Y. Mao and G. Lebanon (2007). Isotonic Conditional Random Fields and Local Sentiment Flow. *Advances in Neural Information Processing Systems*, 19, 961-968.
- [7] C.-C. Chang and C.-J. Lin (2001) LIBSVM: a library for support vector machines, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.