논문 2011-4-31

# 감정요소를 이용한 SNS 메시지 분류기 구현에 대한 연구

# A Study on the Implementation of SNS Message Classification by Emotion Factors

김재영*, 김명관**

**Jae-Young Kim, Myung-Gwan Kim**

**요 약** 최근 SNS가 급격하게 성장하고 있고 많은 사용자들이 이 SNS를 하나의 다른 커뮤니케이션 매체로 사용하고 있다. SNS를 이용하는 개인 사용자들은 자신의 소식과 감정의 변화를 표현하는 수단으로 SNS를 이용하고 있다. 이에 본 연구에서는 감정을 나타내는 감정 요소를 이용하여 메시지를 분류하는 프로그램을 구현하였다. 감정 성분 추출은 OMLS(Ocean-Monmouth Legal Services)에 있는 감정 어휘를 이용하여 로젯(Roget)의 시소러스와 워드넷(WordNet)을 이용하여 이루어졌다.

**Abstract** SNS is growing by leaps and bounds, and many users of SNS are using by a medium of communication. Using SNS users are using means of their own news and the change of emotional expression. In this study using emotional elements to the program was implemented to classify the message. Extraction of emotional elements were used for emotional vocabulary in OMLS (Ocean-Monmouth Legal Services). Emotional elements were extended by The Roget of the thesaurus and WordNet.

**Key Words :** SNS, 감정분루, Twitter, 시소러스, Emotion, 감정

## Ⅰ. 서 론

Social Network 란 사람과 사람 사이의 연결망으로 둘만의 관계뿐만 아니라 여러 가능한 모든 관계를 일컫는 말이다. SNS(Social Network Service)라는 말은 인맥 서비스라고 할 수 있다. SNS가 2007년 이후급성장하는 추세를 보이고 있다. eMarketer보고서[1] 에서는 2007년 미국 성인 인터넷 사용자의 37%가 최소 한 달에 한번 이용하고 있으며 2011년에는 약 50%가 사용할 것이라고 전망하고 있다. 이러한 SNS의 급성장의 배경에는 SNS기업들의 Open API를 제공함으로서 개발자들이 쉽게 프로그램을 개발 할 수 있도록 하였다. 대표적인 SNS로는 Twitter, Facebook, Cyworld, Me2day등등이 있다. 최근 급격하게 성장한 SNS통하여 많은 사용자들이 다른 하나의 커뮤니케이션 매체로 사용하고 있다. 많은 기업들이 기업의 마케팅에 SNS를 사용하고 있을 뿐만 아니라, 많은 개인 사용자들은 SNS를 통하여 자신의 소식과 감정의 변화를 표현 하는 수단으로 사용하고 있다.

텍스트로부터 추출할 수 있는 유용한 정보 중에 하나는 작가가 해당 문서에서 표현 한 감정 혹은 의견이다.[13] 방송통신위원회와 한국 인터넷 진흥원에서 3000명을 대상으로 조사한 통계보고서에[15] 따르면 인터넷이용자의 SNS 이용률은 61.3%이고 SNS 이용자의 주된 목적으로 친교・교제를 위해93%, 재미와 즐거움을 위해 63.4%, 개인의 일상생활이나 관심사 공유 81% 사업이나 업무, 학업등에 필요한 정보습득 55.3%, 취미・여가 활동 78.3%, 시사・현안문제 등에 대한 의견 표현 및 공유 14.9%, 일

*을지대학교 의료IT마케팅학과
**을지대학교 의료IT마케팅학과교수(교신저자)
접수일자 2011.6.8, 수정일자 2011.7.13
게재확정일자 2011.8.12

상생활에 관한 정보 습득 65.7%, 기타 2.5%로 나타나고 있다. 항목 중 재미와 즐거움을 위해서는 감정과 밀접한 관계를 가진다. 감정 분류는 주제에 대한 긍정이과 부정적인 면을 분류할 수 있고, 이로써 주제에 대한 의견의 호(好), 불(不)을 나눌 수 있고, 기업에서도 댓글을 통한 고객 평가, 고객 성향을 분석하는 등의 응용영역을 가질 수 있다.[12]

감정요소를 사용하여 분류 방법에 대한 연구로는 정신병환자의 치료를 위한 Colby의 Parry, Dyer의 BORIS, OdEd, Reeves의 THUNDER, CMU의 Oz등이 있다. 전통적인 문서 분류는 주제(Topic)에 초점을 맞추어 있었다.

본 연구에서는 Twitter의 메시지를 추출하여 문서로 사용하였고, 추출된 문서를 감정 요소를 이용하여 분류하는 프로그램을 구현하여 보았다.

## II. 본 론

### 1. 관련연구들

정보처리에 있어서 감정에 대한 연구는 인공지능 분야의 중요한 과제였다. 대표적인 감정 처리를 포함한 시스템은 Colby의 PARRY[2], Dyer의 BORIS[3]와 OdEd,[4] Reeves의 THUNDER[5], ACRES, CMU의 Oz, Wright의 감정행위자(Emotion Agents) 등이 있다.

PARRY는 '공포', '분노', '불신' 세 가지 감정 요소를 지원하는 시스템으로서 영어와 비슷한 입력형식을 가졌다. 주로 정신병 환자와의 대화를 목적으로 설계되었으며 악의적인 질문이 발견되면 물리적인 공포와 심리적인 분노로부터 불신에 관한 변수 값을 변경하여서 대응하는 결과 값을 출력하는 구조였다. BORIS는 유명한 인공지능 사례인 '이혼'에 관한 에피소드를 인식하는 심층인식(In depth understanding) 소프트웨어이다. 이 프로그램은 감정과 관련된 이야기를 분석하여 문맥을 이해한다. 감정을 위하여 6개의 스롯을 두어 변화를 처리한다. 이프로그램은 목표, 계획, 스크립트, 물리적 객체, 설정, 대인관계, 사회적 역할, 감정적 반응 등이 상호작용하며 관리한다. OdEd는 신문의 사설을 이해하기 위한 시스템으로서 Shank의 CD(Conecptual Dependency)이론을 적용하여 구현하였다. 지식베이스를 가지고 일종의 목표 기반 추론을 수행하였다. 즉, 실망을 만나면 지금의 목표는 취소하고 정지되어 있는 목표들 중에 다른 것을 선택하게 된다. THUNDER는 문서 구조를 이해하기 위한 시스템으로써 좋은 것과 악한 것, 옳은 것과 잘못된 것에 대한 판단을 지식을 기반으로 판단한다. 그리고 이야기 안에 유용한 것을 어떻게 평가되는지 보여준다. Oz 프로젝트는 Carnegie Mellon 대학의 감성추론에 대한 연구로 가상 인물 간 상호작용을 통하여 감성을 생성한다. 가상인물 간 상호작용은 텍스트 기반의 인터페이스를 통한 상호작용 소설(Interactive Fiction)과 몇 가지 경로 중 하나를 선택하는 상호작용 희곡(interactive Play)로 이루어진다. 각각 가상인물은 서로 행위를 통하여 환경을 변화 시키고 적절한 감정 변수를 갖는다. 이 프로젝트는 OCC (Ortony, Collins, Clore)의 모델을 기반으로 하고 있다.

### 2. 감정성분 추출

본 연구의 첫 번째 목표는 감정표현 단어들에 대한 시소러스를 구축하는 것이다. 시소러스의 대표적인 결과물은 영국에서 나온 유의어, 반의어를 모아노은 로젯(Roget)의 시소러스[8]와 유의어, 반의어 상위어 하위어를 검색 할 수 있는 미국 프리스턴 대학의 워드넷(WordNet)[7]을 들 수 있다. 밀러와 존슨에 의해 만들어진 워드넷은 대규모 영어 어휘 데이터베이스이다. 워드넷은 영어 어휘를 명사(noun), 동사(verb), 형용사(adjective), 부사(adverb)로 크게 나누고 이들 어휘의 동의어(synonym) 집합을 정의한 후 이들 동의어 집합간의 의미적 상관관계를 컴퓨터로 처리 가능하도록 체계적으로 정리하고 있다. 워드넷 에는 약 20만여개의 단어-의미 쌍이 저장 되어있다. 많은 단어 들이 의미 관계로 연결되어있고 이러한 관계는 단어의 타입에 종류에 따라 검색할 수 있다. 본 연구에서는 감정 단어들을 행복, 슬픔, 성난, 혼란스러운, 두려운(Happy, Sad, Angry, Confused, Scared)등 다섯가지 범주의 단어들로 구성 하였다.

**표 1. 타입에 따른 관계**
Table 1. Depending on the type relationships

| 종류 | 관계 |
|---|---|
| 명사 | 상위어, 하위어, 등위어, 전체어, 부분어 |
| 동사 | 상위어, 양태어, 수반, 등위어 |
| 형용사 | 관계있는 명사, 동사의 분사 |
| 부사 | 파생된 영용사 |

OMLS(Ocean-Monmouth Legal Services)에서 나온 5가지 감정어휘(Vocabulary of Feelings)들을 토대로 로젯의 시소러스와 워드넷을 이용하여 유사어, 상위어, 하위어를 검색하여 OMLS의 감정단어 들을 확장하여 시소러스를 구성 하였다. 다음 표 2. 는 OMLS의 구성 중 일부이다.

**표 2. OMLS 감정단어**
**Table 2. OMLS Emotion Words**

| | |
|---|---|
| Happy | serene, complacent, promise, trust, desire, fortunate, favored, golden rosy, pleas, delight, wish, care, like, flatter.... |
| Sad | devastate, destroy, ruin, waste, desolate, ravage, scourge, overwhelm, sweep over, whelm, overcome..... |
| Angry | angry, furious, raging, tempestuous, wild, strangle, strangulate, throttle, smother, stifle, muffle |
| Scared | fright, affright, intimidate, restrain, awful, dire, direful, dread, dreaded, dreadful, fearful, fearsome, frightening |
| Confused | throw, fox, befuddle, fuddle, bedevil, contract, press, constringe, narrow, trouble, confound, flurry |

## 3. 프로그램 설계

본 프로그램에서는 SNS(Social Network Service)중에서 Twitter의 Open API 사용하여 프로그램을 구성하였다. Twitter Open API를 사용하기 위해서는 인증 절차가 필요하며 인증에 통과한 사람만이 모든 API를 사용할 수 있다. 인증방식의 표준이 없기 때문에 제 각각의 방법으로 개발되고 있고 Twitter에서는 OAuth, xAuth인증이 준비 되어 있다. OAuth는 유저이름과 패스워드 대신 규약에 따라 정해진 순서로 입수한 토큰을 사용해서 인증하는 방식이다.[18] xAuth는 OAuth의 축소판이라고 볼 수 있다. OAuth와의 가장 큰 차이점은 유저에 의한 브라우저 조작이 필요 없다는 것이다. 본 프로그램에서는 OAuth 인증 방식을 사용하여 프로그램을 구성 하였다. OAuth 인증 순서는 다음과 같다.

1) 애플리케이션을 Twitter에 등록하고 Consumer key와 Consumer secret을 취득하기
2) Request Token 취득 API를 사용해서 Request Token을 취득하기
3) Request Token의 token을 사용해서 유저에게 접속 허가를 요구하기 위한 URL 생성하기
4) 유저에 의한 접속 허가로 Twitter에 접속
5) PIN(인증번호) 취득하기
6) Access Token 취득 API를 실행해서 Access Token을 취득하기

본 연구에서 Twitter의 메시지를 가져오기 위해 Timeline의 API중 비공개 설정을 하지 않은 유저의 메시지를 수집 할 수 있는 public_timeline을 사용하여 메시지를 가져왔다. public_timeline API를 취득하게 되면 tweet 중에서 최신 tweet 20건을 취득하며 API 실행이 성공한 경우 타임라인 정보의 형식으로 public_timeline의 내용이 반환된다.

Twitter로부터 가져온 메시지로부터 하나의 단어를 추출 하여 Poter의 Stemming 알고리즘을 사용하여 어근을 뺀 표준형으로 바꾼다. 표준형으로 바꾼 단어와 만들어진 감정 시소러스와 비교 하여 성분 벡터를 추출한다. 추출한 벡터들을 비교하여 가장 큰값을 찾아 감정에 따라 분류한다. 전체적인 흐름은 그림 1. 과 같다.

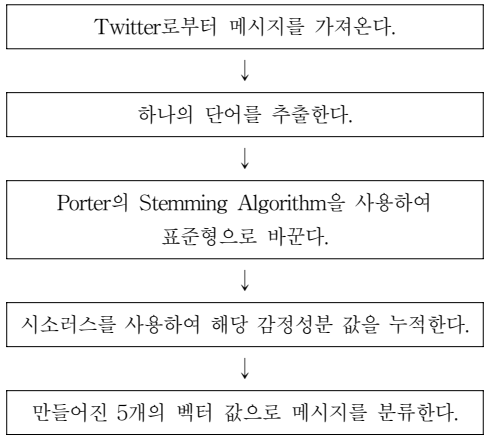



**그림 1. 프로그램 흐름도**
**Fig. 1. Program Flow Chart**

## 4. 프로그램 구현

감정에 따른 SNS(Twitter)검색 프로그램을 시험적으로 구현하였다. 본 프로그램은 SNS중 하나인 Twitter의

Open API를 이용하여 Twitter 메시지에서 문서를 추출하였다. 추출된 문서에서 하나의 단어를 추출하여 Poter의 Stemming 알고리즘을 사용하여 단어의 어근을 추출 복수형, 진행형 등을 표준형으로 바꾼다. Poter의 Stemming 알고리즘은[9] 그림 2. 와 같이 6단계에 걸쳐 표준형으로 변형된다. 바꾼 표준형을 저장하고 위 감정 시소러스와 비교 감정 벡터를 구성한다. 벡터는 <happy, angry, scare, sad, confuse> 다섯 가지로표현한다. 문서 i의 추출된 성분의 벡터는 EVi = <vi1, vi2, vi3, vi4, vi5> 로 표현한다. 해당 벡터에 해당되지 않으면 normal문서로 분류한다. 구성된 감정벡터들을 비교하여 가장 큰 벡터를 가지는 감정 영역으로 추출된 문장을 감정에 따라 분류한다. 예를 들어 “I think it must be a lot of hooey just like the Fatima letter in 1960”란 메시지가 있다면 프로그램을 실행 했을 때 백터의 값은 <1,1,1,2,4>가 나오고 이중 가장 높은 벡터인 5번째 감정벡터가 이 메시지의 감정이 된다. 5번째 감정은 confuse이므로 이 메시지는 confuse로 분류되어 나오게 된다.

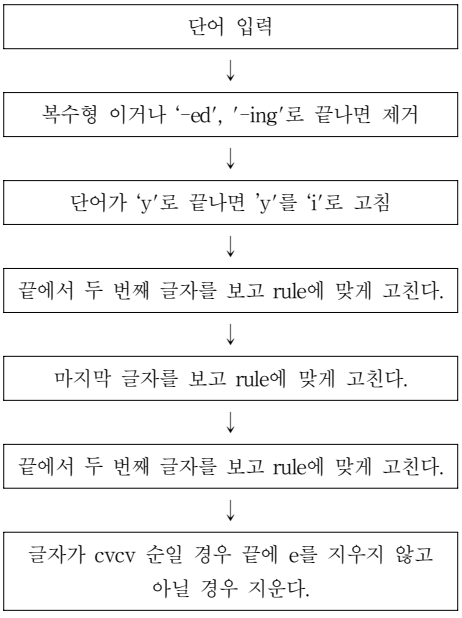



**그림 2. Porter의 Stemming 알고리즘**
**Fig. 2. Porter's Stemming Algorithm**

프로그램 실행 하게 되면 아래 그림 3. 과같이 실행이 된다.

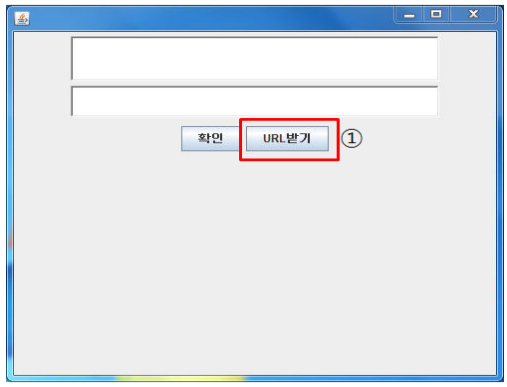



**그림 3. 실행화면1**
**Fig. 3. Implementation1**

그림 3.에서 URL받기 버튼을 누르게 되면 트위터 API를 사용할 수 있는 PIN번호를 받는 주소가 그림 4.와 같이 나오게 되며 그 URL을 통하여 인터넷에 접속하게 되면 그림 5. 와 같이 PIN번호를 취득 할 수 있다. 취득한 핀번호를 그림 4. 같이 textbox에 입력하고 확인을 누르면 Twitter 메시지를 추출하여 그림 6.과 같이 화면에 뿌려지게 된다.


**그림 4. 실행화면2**
**Fig. 4. Implementation2**


**그림 5. Pin 번호**
**Fig. 5. Pin Number**

다음 실행화면의 인터페이스 구성은 그림 5.와 같다.

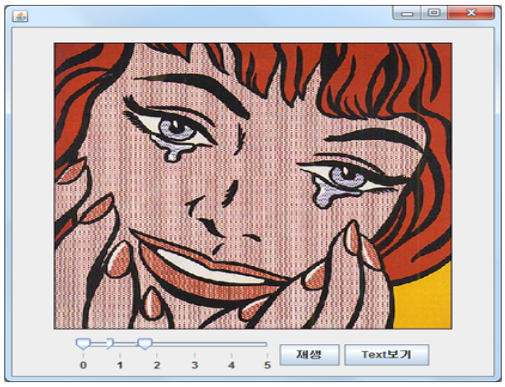



**그림 6. 실행화면3**
**Fig. 6. Implementation3**

아래 있는 슬라이드 버튼으로 채널을 맞추면 각 채널에 해당하는 감정과 관련된 그림이 나오게 된다. 0 ~ 5까지 각각 happy, angry, scare, sad, confuse, normal감정에 해당한다. 그림 6.은 0번 즉 happy 감정을 선택했을 때 나타는 화면이다. 그림 6.에서 재생 버튼을 누르게 되면 해당 감정으로 분류된 Twitter 메시지를 읽어 주게 된다.

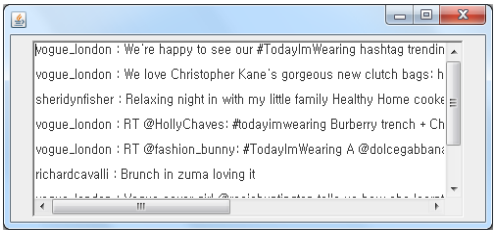



**그림 7. 택스트 출력**
**Fig. 7. Text generating**

그림 6.에서 Text보기 버튼을 누르게 되면 그림 7.에서와 같이 해당하는 감정의 메시지를 Text로 출력하여 보여주게 된다.

## Ⅲ. 결 론

기존 SNS 프로그램들은 여행, 정치, 경제와 같은 분류는 존재 하였으나 감정에 따른 분류는 사용하지 않았다. 텍스트로부터 추출할 수 있는 유용한 정보 중에 하나는 작가가 해당 문서에서 표현 한 감정 혹은 의견이다. 문서의 감정에 따른 분류는 그 문서에 대한 호(好), 불(不)을 나눌 수 있고, 기업측면에서 마케팅을 할 때도 댓글이나 메시지를 분류하여 고객의 평가나 성향을 분석할 수 있다.

본 논문에서는 감정 성분을 해당 문서로부터 추출하여 감정 성분을 기반으로 한 SNS프로그램을 구현하여 보았다. OMLS의 감정 어휘를 기반으로 언어학자인 로젯의 시소러스와 프리스턴 대학의 워드넷을 사용하여 감정 시소러스를 구축 하였고, 많이 사용하고 있는 SNS중 하나인 Twitter를 이용하여 문서를 추출 하여 프로그램을 구성하여 보았다.

## 참 고 문 헌

[1] Debra Aho Williamson, “Social Network Marketing : Ad Spending and Usage”, eMarketer, 2007

[2] K. Colby, “Artificial Paranoia: A computer simulation of paranoid process”, Pergamon Press, 1975

[3] Dyer M. G, “In depth understanding”, MIT Press, 1983

[4] Dyer M. G, “Emotions and their computations: Three computer models”, Cognition & Emotion 1(3), pp.323-347, 1987

[5] Reeves J. F, “Computational morality : A process model of belief conflict and resolution for story understanding”, Technical Report UCLA-AI-91-05, 1991

[6] Myung-Gwan Kim, “Information Retrieval Agents Using Emotional Features”, 정보처리학회지, 제10-B권, 제6호, pp.579-586, 2003

[7] Miller G. A, "WordNet : An On-line Lexical Data Base", Hillsdale, 1993

[8] Roget P. M, “Roget's Thesaurus”, Gramercy Books, 1979

[9] Porter M. F, “An algorithm for suffix stripping”, Program 14(3), pp.130 - 137, 1980

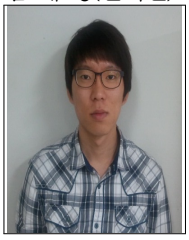
[10] 츠지무라 히로시, 이규홍역 “twitter API 개발자 레퍼런스”, Youngjin, 2010

[11] 유훈식, “커뮤니케이션 유형에 따른 SNS의 인터랙션 특성에 따른 연구,” 국민대학교 테크노디자인전문대학원 인터랙션디자인 전공, 2009

[12] 황재원, “감정 분류를 위한 한국어 감정 자질 추출기법과 감정 자질의 유용성 평가”, 한국인지과학회, 제19권, 제4호, pp.299-517, 2008
[13] M. Rimon, "Sentiment Classfication : Linguistic and Non-Linguistic Issues", Hebrew University
[14] Hong Jiang, “EBDI: An Architecture for Emotional Agents”, AAMAS, May.2007
[15] 서재철 외 5명, “인터넷이용자의 SNS이용실태조사”, 방송통신위원회, 한국인터넷진흥원, 2009
[16] 황현수, “Social Network Service”, SK Comunication, 2007

## 저자 소개

**김 재 영(준회원)**



• 을지대학교 의료IT마케팅학과 재학생

**김 명 관(정회원)**

• 1981년 3월~1985년 2월 숭실대학교 전자계산학과 학사
• 1985년 3월~1987년 2월 숭실대학원 전자계산학과 석사
• 1996년 9월~2004년 2월 숭실대학원 컴퓨터학과 박사
• 1989년 8월~1993년 2월 한국전자통신연구소 인공지능연구실 연구원
• 1993년 3월~2007년 2월 서울보건대학 컴퓨터정보과 부교수
• 2007년 3월~현재 을지대학교 의료IT마케팅학과 부교수

<주관심분야 : 인공지능, 자연어처리, 질의응답시스템, 시멘틱웹>