

트위터의 감정 분석을 통한 실시간 장소 추천 시스템

Real-time Spatial Recommendation System based on Sentiment Analysis of Twitter

오평화(Pyeonghwa Oh)*, 황병연(Byung-Yeon Hwang)**

초 록

본 논문에서는 모바일에서 획득한 GPS(Global Positioning System)를 활용하여 사용자의 위치 주변에서 발생한 SNS 데이터를 수집하고 분석을 통해 사용자가 원하는 장소를 추천하는 시스템을 제안한다. 이를 위해 트위터에서 위치정보를 포함하는 게시글을 표본 집합으로 정하고 모바일의 위치정보와 함께 활용했을 때, 사용자의 검색의도에 부합하는 양질의 정보를 제공할 수 있음을 실험을 통해 증명하였다. 이를 위해 2015년 11월부터 12월까지 수집한 트윗(Tweet)을 대상으로 임의의 위치정보와 검색어로 구성된 질의를 구성하고 형태소 분석을 거쳐 분석에 적합한 형태의 데이터로 변환하였다. 또한 장소 추천을 위해 감정사전을 구축하여 긍정 및 부정을 의미하는 극성 키워드들을 정의하고 레이블을 구성한 후, 감정사전과 극성키워드를 이용해 개별 트윗의 추천 점수를 도출하였다. 논문은 추천 점수와 사용자의 현재 위치, 트윗이 작성된 위치와 사용자 위치 사이의 거리 계산을 통해 가까운 거리 순으로 10개의 장소 정보를 정렬하여 결과를 보인다. 또한 성능평가를 위해 감정 분석된 트윗에 대한 정밀도와 재현율을 도출하여 시스템의 성능을 확인한다. 실험은 ‘맛집’, ‘공연’ 2개의 키워드와 10개 지역을 기준으로 수행하였다. 실험 결과 키워드 1개당 수집된 트윗은 평균 10.5개였으며, 총 10번의 실험에 사용된 평균 210개의 트윗 중 긍정 또는 부정의 단어를 포함한 트윗의 개수는 평균 122개였다. 또한 감정 분석을 통해 긍정 또는 부정으로 분류된 트윗은 평균 65개였으며 그 중 실제로 긍정 또는 부정의 의미를 담은 트윗은 평균 46개였다. 이를 통해 시스템은 38%의 재현율로 감정 요소를 담은 트윗을 탐지하고, 71%의 정밀도로 감정 분석을 수행했음을 확인했다.

ABSTRACT

This paper proposes a system recommending spatial information what user wants with collecting and analyzing tweets around the user's location by using the GPS information acquired in mobile. This system has built an emotion dictionary and then derive the recommendation score of morphological analyzed tweets to provide not just simple information but recommendation through the emotion analysis information. The system also calculates

본 연구는 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단 기초연구사업(No. 2011-0009407)의 연구비 지원으로 수행되었음.

* Department of Computer Science and Engineering, The Catholic University of Korea

** Corresponding Author, Department of Computer Science and Engineering, The Catholic University of Korea(byhwang@catholic.ac.kr)

Received: 2016-03-28, Review completed: 2016-08-03, Accepted: 2016-08-10

distance between the recommended tweets and user's latitude-longitude coordinates and the results showed the close order. This paper evaluates the result of the emotion analysis in a total of 10 areas with two keyword 'Restaurants' and 'Performance.' In the result, the number of tweets containing the words positive or negative are 122 of the total 210. In addition, 65 tweets classified as positive or negative by analyzing emotions after a morphological analysis and only 46 tweets contained the meaning of the positive or negative actually. This result shows the system detected tweets containing the emotional element with recall of 38% and performed emotion analysis with precision of 71%.

키워드 : 소셜 네트워크 서비스, 트위터, 빅데이터, 형태소 분석, 감정 분석, 위치기반 서비스, 장소 추천
 SNS, Twitter, Big Data, Morphological Analysis, Sentiment Analysis, Location Based Service, Spatial Recommendation

1. 서 론

무선인터넷 환경이 보편화되면서 언제 어디서나 인터넷에 접속할 수 있는 모바일은 현대인의 삶에서 가장 중요한 정보 검색 수단이 되었다. 특히 PC와 비견될 만큼의 성능 개선과 더불어 소형화 및 경량화를 실현함으로써 모바일은 PC 기반의 온라인 생태계를 위협하는 강력한 사용자 장악력을 보유하게 되었다. KT경제연구소의 발표 자료에 따르면 2015년 3월 기준 56개 국가의 평균 스마트폰 보급률은 60%에 달했으며, 우리나라는 83%로 상위권을 차지하고 있다[5]. 또한 Wikipedia의 '2015년 전 세계 평균 인터넷 접속 속도 순위'에 따르면 인터넷 평균 속도에서 우리나라는 8분기 연속 세계 1위를 차지했다[13]. 높은 스마트폰 보급률과 세계 최고 수준의 인터넷 인프라는 인터넷 환경을 PC에서 모바일로 빠르게 이동시켰다.

모바일의 성장으로 사용자는 시간과 장소에 구애받지 않고 필요한 정보를 즉시 취득하여 활용할 수 있게 되었다. 직접 은행이나 백화점에 가지 않아도 금융서비스를 이용하고 쇼핑

을 즐기는 모습은 이미 익숙한 모습이 되었다. 사용자의 위치정보를 활용해 터치 한 번으로 택시를 호출하고 차량 이동경로와 예상시간, 기사와 요금정보까지 실시간으로 조회할 수 있는 카카오택시나 현 위치에서 사용할 수 있는 할인쿠폰을 자동으로 보내주는 Yap, 모바일에 저장된 계좌정보로 결제와 송금이 가능한 카카오뱅크의 사례에서 보듯 모바일은 기존의 오프라인 시장을 대체하거나 새로운 시장을 개척하는 서비스들의 등장을 가속화시키고 있다.

반면 PC 기반의 강세를 모바일로 이어가지 못하는 사례도 존재한다. 국내에서 가장 많은 접속이 발생하는 도메인의 대부분은 검색 서비스로 나타났다[1]. 그러나 구글이나 애플 앱스토어 다운로드 순위에서는 대부분 10위권 밖에 있으며, SNS나 메신저와 같은 네트워크 기반 서비스가 상위권을 대체하고 있다[2]. 이러한 비대칭 현상은 인터넷 환경변화에 빠르게 대응하지 못했다는 점과 모바일의 다양한 기능을 제대로 활용하지 못했다는 점 등을 원인으로 꼽을 수 있다. 모바일 서비스에서 가장 두드러

지는 특징은 개인화다. 모바일은 사용자의 연락처, 메일, 메시지 등은 물론 경우에 따라 행정이나 금융 정보까지 보관하고 이용하며 위치정보, 이용패턴 등의 개인정보를 활용해 진보된 형태의 사용자 경험을 제공하고 있다. 이러한 모바일의 장점을 적절히 활용한 검색 서비스를 개발한다면 사용자들의 편리함을 더해줄 서비스로 성장할 것이다.

본 연구의 바탕이 되는 소셜 네트워크와 감정분석에 대한 다양한 연구들이 진행되고 있다. Lim 등은 형태소분석을 기반으로 한 사전분류 방법 대신 기계학습을 통한 트위터의 감정분류를 시도하였으며 평균 76% 이상의 정확도를 보였다[8]. 또한 Lee 등의 연구에서는 사회의 최근 동향에 대한 여론의 반응을 관찰하기 위해, 최근 동향을 나타내는 키워드를 신문기사로부터 추출하고, 추출된 키워드를 이용하여 수집된 트윗의 감성 분석을 통해 최근 동향에 대한 여론을 분석하여 70% 중반의 정확도를 보였다[6]. Oh 등은 소셜 네트워크에서 시멘틱 웹기반의 유용한 정보를 찾기 위해 구성원간의 잠재되어있는 관계를 인식하고 랭킹을 매긴 연구를 했다[9].

이에 본 논문에서는 사용자의 위치정보를 포함한 검색 질의 결과에 감정 분석을 더해 신뢰성 높은 장소를 추천하는 시스템을 제안한다. 시스템은 사용자로부터 입력받는 검색 키워드와 모바일의 위치정보 조합을 매개로 트위터 API를 이용해 질의하고, 수집된 트윗의 형태소 분석을 통해 추천할만한 트윗을 선별하여 구글 지도를 통해 현 위치로부터 근거리 순의 추천 리스트를 출력한다. 본 논문에서 제안하는 시스템과 기존 시스템의 차이점 및 의의는 다음과 같다.

- 1) 기존 시스템이 정확한 지명과 명칭을 포함한 질의에 대해서만 검색 결과를 보장하는 반면, 제안하는 시스템은 사용자의 모바일 위치정보를 활용해 ‘맛집’, ‘데이트코스’ 등의 단순검색으로도 가장 가까운 장소 정보를 제공한다.
- 2) 사용자가 검색한 장소를 거리나 시간 순으로 단순 정렬하는 기존 시스템과 달리 제안하는 시스템은 감정 분석을 통해 추천할만한 신뢰성 높은 검색 결과를 제공한다.
- 3) 본 시스템은 ‘길거리 공연’, ‘깜짝 이벤트’ 등 현 위치 주변에서 산발적으로 발생하는 실시간 정보도 검색 결과로 제안할 수 있다.

본 논문은 다음과 같은 순서로 구성되어있다. 제2장에서는 모바일 환경 및 자연어처리에 관련된 연구들과 제안하는 시스템과 관련된 API를 살펴본다. 제3장에서 제안하는 시스템의 구성과 함께 실제로 취득한 트윗에 감정 키워드를 적용하여 레이블을 부여한 후 개별 트윗에서 분류한 극성과 트윗의 위치를 토대로 장소를 추천하는 과정을 보인다. 제4장에서는 제3장의 결과를 검증하기 위한 성능 평가를 수행하고 제5장에서 결론을 맺는다.

2. 관련 연구

Lee[7]의 논문에서 소셜 미디어 데이터로부터 검색을 통해 원하는 자료를 찾고, 그 데이터들이 생성된 위치정보를 웹 지도상에 표현하는 모바일 시스템을 소개했다. 이주용의 연구

에서는 소셜 네트워크 서비스의 위치 정보를 활용하는 방법 중 하나로 사용자가 원하는 키워드에 대한 위치정보를 구글 지도 API를 통해서 보여주었다는 점에서 본 연구와 유사점을 찾을 수 있다. 하지만 사용자의 위치와 무관한 장소검색의 한계를 보였으며, 사용자의 주변에서 가까운 정보를 감정 분석을 통한 추천으로 사용자의 판단에 기여한다는 점에서 본 연구와의 차이점을 찾을 수 있다.

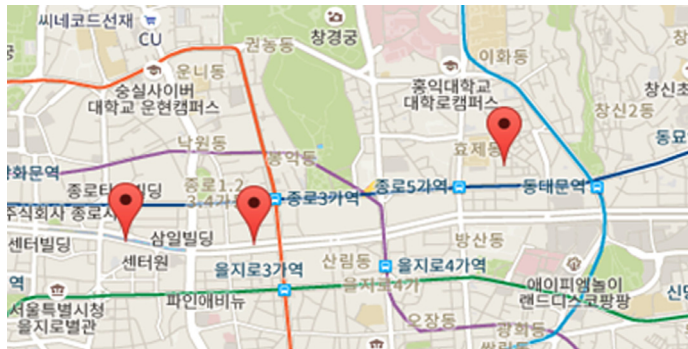
Yoo[14]의 논문에서는 웹이 아닌 모바일 환경에서 장소를 검색할 때 행동을 파악하여 사용자의 장소 찾기 프로세스의 형성과정을 연구했다. 실험 분석 결과, 모바일 사용자의 장소 검색 행동에 대한 이해를 기반으로 ‘장소 정보 검색 행동’ 모형을 구조화할 수 있었으며, 모바일 사용자가 어떻게 장소를 찾는지, 또 그 과정에서 어떤 부분들이 주목되어야 하는지를 언급함으로써 모바일 환경에서 그 중요성이 증대된 장소 정보라는 디지털 콘텐츠의 질적 이해를 높였다. 이는 본 연구에서 제안하는 실시간 장소 추천 시스템이 그 자체로 충분한 의미가 있음을 보여준다. 사용자가 모바일에서 장소를 검색할 때 현 위치에서 가까운 장소에 대한 정보를 원하는 경향은 ‘길거리 공연’, ‘깜짝 이벤트’ 등의 이벤트 장소까지 추천받을 수 있는 본 논문의 시스템의 타당성을 높인다.

국외 사례로 Hul[3] 등은 장소를 추천하기 위해 STT(Spatio-Temporal Topic) 모델을 제안하였다. STT 모델은 사용자의 체크인(Check-ins)을 통해 시간, 공간에 맞는 공간을 추천하고자 한다. 이들은 실제 주거 지역과 사용자들의 장소, 시간에 따른 활동과 지역 사이의 상호 의존 관계가 있음을 밝혔다. 이는 국외에서도 위치를 기반으로 한 서비스에 관련한 연구가 진행되

고 있고 위치 기반 소셜 네트워크가 활용됨을 증명한다. 본 연구는 개별 사용자 패턴에 따른 추천 방식이 아닌 위치 정보를 기반으로 최근 장소 정보를 결과로 보여주는 방식으로 차이점을 볼 수 있다.

자연언어 처리는 문장에 포함된 각각의 단어가 갖는 품사를 분석하고 이를 태깅(Tagging)하는 과정을 의미한다[11]. 언어는 이러한 품사 태깅 과정(Part-Of-Speech Tagging)을 거쳐 문장에 사용된 형태소들의 품사와 문장의 구조를 파악함으로써 기계적으로 분류되고 시스템에 활용될 수 있다. 본 논문에서는 KAIST의 한나눔 형태소 분석기[4]를 사용한다. 한나눔 형태소 분석기는 GPLv3 라이선스를 따르기 때문에 연구 목적으로 사용하는데 문제가 없고 자바 버전으로 개발되어 배포되기 때문에 이용이 편리하다. 또한 플러그인 컴포넌트 아키텍처를 적용하여 사용자에게 보다 유연하게 사용될 수 있도록 개선되었다. 개발자는 한국어 처리 목적에 따라 개발된 Plug-in인 뿐만 아니라 필요한 기능을 개발하여 함께 Workflow를 구성하여 사용할 수 있다.

대부분의 SNS는 각자의 플랫폼을 이용한 서드파티(Third Party)들의 다양한 서비스 개발을 독려하기 위한 오픈 API(Open API)를 제공한다. 트위터는 API의 이용을 편리하게 이용하도록 자바(Java) 언어에 최적화된 Twitter4j 라이브러리(Library)[12]를 제공하고 있다. 트위터는 REST(Representational State Transfer) 기반의 Search API와 Streaming API 2가지 방식의 API를 제공한다. Search API는 요청 시각 이전에 발생한 트윗을 검색하는 반면 Streaming API는 요청 시각 이후로 발생하는 트윗을 실시간으로 수집한다. 이렇듯 서로 다른 기준 시점은



〈Figure 1〉 Making Location Information of Tweet Using Google Map

검색 조건의 차이를 가져온다. 본 논문에서는 Streaming API에 비해 정의할 수 있는 검색 조건의 종류가 다양하고 대상 범위가 넓은 Search API를 이용한다.

온라인에서 포털 서비스 기업들이 제공하는 다양한 서비스에는 뉴스, 음악, 사진뿐만 아니라 본 논문에서 제안하는 시스템의 추천 결과를 표시하는 지도 서비스도 포함되어 있다. 각각의 포털 서비스들은 개발자 참여공간을 제공하고 있으며, 자사 서비스를 쉽게 이용할 수 있는 매시업(Mash Up) API를 제공하고 있다. 지도 서비스의 경우 원하는 위치의 키워드 또는 좌표 정보를 매개변수로 서비스를 호출하면 자체 지도 위에 마커(Marker)로서 해당 위치를 표시해준다. 개발자는 이를 이용해 다중 마커, 위치간 거리, 경로 등 다양한 정보를 획득할 수 있으며 개발 중인 서비스에 활용할 수 있다. 본 논문은 국내와 해외를 별개 서비스로 제공하는 네이버와 다음 대신 하나의 서비스에서 세계 지도를 제공하는 구글 지도 API를 이용하였다. 〈Figure 1〉은 구글 지도 API를 이용해 ‘서울 빙초롱축제’와 관련된 임의의 트윗에서 발췌한 위치정보를 구글 지도에 표시한 예이다.

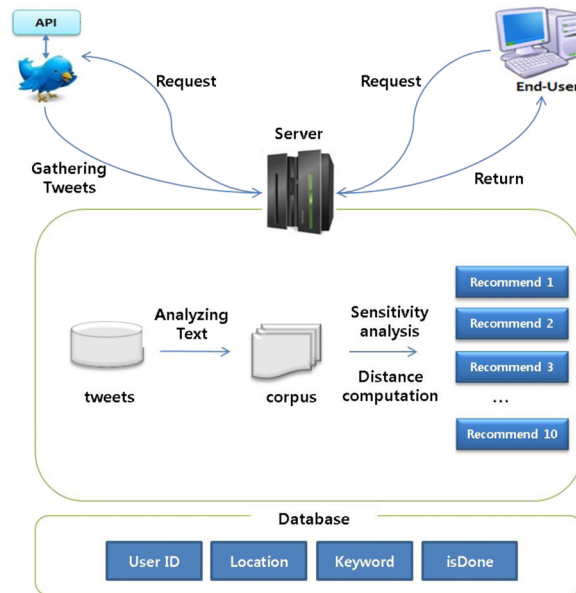
3. 시스템 구성 및 설계

3.1 시스템 구성

이 절에서는 제안하는 시스템의 아키텍처와 시스템 흐름에 대해 설명한다. 제안하는 시스템은 크게 4가지 과정을 거친다.

- 1) 사용자의 위치 정보와 장소 키워드 수집과 이에 따른 트윗 수집 단계
- 2) 수집한 트윗을 형태소 분석 후 감정단어가 속한 태그가 붙은 단어 추출 단계
- 3) 감정단어로 분류된 단어를 감정사전에 비교하여 긍정이면 ‘+1’, 부정이면 ‘-1’, 중립이면 ‘0’으로 계산하여 개별 트윗의 추천 점수를 보이는 단계
- 4) 추천 점수를 갖는 개별 트윗의 위치와 사용자의 거리계산을 통해 감정 분석 점수가 긍정으로 분류되고 사용자와의 거리가 가까운 10개의 트윗을 구글 지도상에서 추천하는 단계

제안하는 시스템은 〈Figure 2〉와 같이 구성된다. 위의 각 단계의 자세한 내용은 다음 절에서 보인다.



<Figure 2> System Configuration

3.2 트윗 수집

트윗 수집은 사용자가 요구하는 검색 키워드와 사용자의 위치 정보를 활용한다. 모바일 환경에서는 장비에서 직접 위치 정보를 제공받기 때문에 사용자에게 요구되는 정보는 오직 검색 키워드이다. 시스템은 사용자에게서 검색 키워드와 위치 정보를 통해 트위터 API에 활용할 질의를 구성한다.

트윗에 포함된 위치 정보는 구글 지도 API에 트윗의 위치를 결과로 보여줄 때 활용되며 이를 통해 사용자는 현재 위치로부터 추천된 위치를 확인할 수 있도록 한다. 입력받은 데이터 중 사용자, 위치, 키워드는 데이터베이스에 저장하여 동적으로 시스템을 사용할 수 있도록 구현하였다. 시스템은 트윗 수집을 위해 사용자의 검색어와 위치 정보를 조합하여 질의를 구성한다.

3.3 형태소 분석

검색과 추천의 차이점은 무의미한 결과의 단순 나열과 검색 의도에 적합한 결과를 정렬한 리스트에 있다. 장소 추천에 필요한 과정 중 핵심은 바로 검색된 트윗에 존재하는 감정 단어의 극성을 판별하여 해당 트윗을 게시한 사용자의 의도를 파악하는 것이다. 이를 위해서는 수집한 트윗에 대한 기계적인 분류가 선행되어야 형태소 분석을 통해 수집된 트윗을 품사단위로 구분하고, 시스템은 필요한 키워드들만 취사선택하여 장소 추천에 활용할 수 있다. 논문에서는 GPLv3 라이선스를 따르기 때문에 연구 목적으로 사용하는데 문제가 없고 자바 버전으로 배포되기 때문에 이용에 용이한 한나눔 형태소 분석기를 이용해 형태소를 분석한다. <Table 1>은 한나눔 형태소 분석기를 통해 형태소 별 태그가 부여된 결과의 예이다.

〈Table 1〉 Morphological Analysis and Tagging

ex) 신당동 맛집 1. #바오쯔 깔끔한 국물의 수제 만둣국, 먹기에 가까운 군교자 충무 일찍오 시는날 요기로~~	
morphological analysis	
신당/ncn+동/ncn	먹기/ncn+에/jca
맛집1/ncn	아깝/paa+은/etm
#바오쯔/ncn	군교/ncn+자/ncn
깔끔/ncps+한/ncn	충무/ncn
국물/ncn+의/jcm	일찍오시는날/ncn
수제만둣국/ncn+/,sp	요기/ncn+로/jca+~/sd+ ~/sy+~/sy

3.4 감정 분석

본 논문에서는 감정 분석을 위해 감정사전

〈Table 2〉 Morphological Analysis and Sentiment Tag of Sentiment Dictionary

word	tag	positive/ negative
갈등하	ncn	N
감동	ncps	P
감탄	ncpa	P
거부	ncpa	N
거부감	ncn	N
경멸	ncps	N
고맙	paa	P
공감	ncpa	P
쾌짹	ncps	N
괴롭	paa	N
귀찮	paa	N
그립	paa	N
근심	ncpa	N
근심걱정	ncpa	N
기쁘	paa	P
낮뜨겁	paa	N
놀라	pvg	P
답답	ncps	N
덤덤	ncps	N
만족	ncps	P

을 구축했다. 감정사전은 Park[10] 등의 연구에서 작성된 한국어 감정단어 434개를 바탕으로 트윗에서 추출한 감정단어 100여 개를 추가하여 구축했다. 사전은 형태소 분석된 형태에 극성의 태그를 추가하여 긍정을 나타내면 ‘P’, 부정을 나타내면 ‘N’을 태깅했다. 〈Table 2〉는 Park[10] 등의 연구의 434개 감정단어 중 20개의 단어에 감정 태그를 붙여 나타낸 예이다.

트윗에서 추출한 감정단어들 중에는 한글 맞춤법에는 맞지 않지만 SNS에서 흔히 사용하는어들도 감정사전에 추가했다. 표준어인 ‘예쁘다’ 외에도 보편적으로 쓰이는 ‘이쁘다’의 ‘이쁘/paa’를 사전에 등록한 것이 그 예다. 또한 ‘남남/ncn’, ‘대박/ncn’, ‘먹방/ncn’ 등도 SNS 상에서 흔히 쓰이는 명사로 분류하여 긍정을 나타내는 ‘P’ 태그를 붙여 사전에 등록했다. 〈Table 3〉은 Park[10] 등의 연구의 감정사전에 추가하여 등록한 단어의 예이다.

〈Table 3〉 The Words Added to the Sentiment Dictionary

word	tag	positive/ negative
꿀맛	ncn	P
남남	ncn	P
대박	ncn	P
맛나	paa	P
종당	ncn	P
척오	ncn	P
이쁘	paa	P
개꿀맛	ncn	P
먹방	ncn	P
망	ncps	N

복수의 형태소 분석 태그가 존재하는 경우 상태성명사와 상태성형용사를 우선하여 적용했다. 그 예로 ‘ncn’, ‘ncpa’, ‘ncps’ 등 복수의 태

그가 붙는 ‘평화’와 같은 단어는 상태성명사를 나타내는 ‘ncps’만을 사전에 포함했다. ‘paa’와 ‘pvg’의 태그가 붙는 ‘못하’의 경우 정상형용사를 나타내는 ‘paa’만을 사전에 포함했다.

구축된 감정사전을 통해 트윗에서 추출된 감정단어들의 점수를 도출해 개별 트윗에 대한 추천 점수를 부여했다. 제안하는 시스템은 실시간을 지향하므로 사용자가 질의하는 당시의 데이터를 트위터에서 수집하고 형태소분석, 감정분석을 수행하는 동안의 상당 시간이 소요된다. 사전에 수집한 데이터를 근거로 분석하는 방법이 대안이 될 수 있지만 이는 성능상의 이슈는 해소할 수 있지만 실시간이라는 본래의 목적에는 부합하지 않는다. 소셜 네트워크 서비스의 집단지성에 기반한 실시간 추천 시스템의 가능성을 타진함에 있어 감정분류 결과의 카운팅은 합리적인 대안이 될 수 있다. 따라서 긍정의 단어를 포함하면 ‘+1’, 부정의 단어를 포함하면 ‘-1’의 점수를 부여했으며, 긍정과 부정단어의 수가 같으면 추천 대상에서 제외하였다. <Table 4>는 ‘맛있/paa’와 ‘흡입/ncpa’의 총 2개의 긍정 단어를 포함하고 있는 트윗에 2점의 추천점수를 부여한 예이다.

<Table 4> Recommendation Grade of Tweet Samples

ex) lunch time 다들 맛있는 점심 드셨죠? 이제부터 흡입해보겠습니다.				
ncn	ncps	ncpa	pvg	paa
다들/ncn		흡입/ncpa	들/pvg	맛있/paa
맛/ncn			하/pvg	
점심/ncn				
드/ncn				
추천점수	흡입/ncpa/P + 맛있/paa/P = 1 + 1 = 2			

수집된 트윗 집합 $T = \{t_1, t_2, \dots, t_n\}$ 이고 임의의 트윗 t_i 에 존재하는 감정단어의 집합 $W_{t_i} = \{w_1, w_2, \dots, w_n\}$ 이라 할 때, 긍정점수 PG 와 부정점수 NG 의 합으로 이루어진 임의의 트윗 t_i 의 추천 점수 $RecommendGrade(t_i)$ 는 아래 식 (1)과 같이 표현할 수 있다.

$$Recommendation Grade(W_{t_i}) = \sum_{i=1}^n PG(w_i) + (-NG(w_i)) \quad (1)$$

3.5 추천 장소 리스트 출력

제3.4절을 통해 도출한 추천 점수를 가진 트윗을 거리 순으로 추천해 주기 위해 사용자의 현재 위치를 기준으로 트윗이 작성된 위도, 경도 좌표와의 거리를 계산했다. <Table 5>는 사용자의 현재 위치로부터 트윗이 작성된 위도, 경도 좌표에 대해 계산된 거리와 트윗의 추천 점수로 정렬된 예이다. 2015년 11월 27일 종로를 중심으로 트윗을 수집하여 ‘점심’이라는 키워드로 추천한 내용이다. 추천 점수가 높고 거리가 가까운 순으로 10까지 나타냈으며 추천 점수, 현재 위치와의 거리, 트윗이 남겨진 GeoLocation 정보, 트윗 내용을 확인할 수 있다.

트윗의 위치정보는 트윗의 작성위치로서 실제 이벤트가 작성된 위치와는 다를 수 있다. 따라서, 제안하는 시스템은 트윗의 작성위치를 지도상에 마킹으로 표시하되 해당 트윗의 내용을 말풍선으로 표시하여 정보로서의 인식의 최종 판단은 사용자의 몫으로 남겨두었다.

〈Table 5〉 The Result of Recommend Locations from the Suggested System

rank	recommend score	distance	latitude, longitude coordinate	tweet contents
1	2	7.0610	37.55524574 126.92703	lunch time 다들 맛있는 점심 드셨죠? 이제부터 흡입해보겠습니다. 하하 #점심 #먹방 #lunch #eat #foodpic #pic #dailycook #cook...
2	1	2.3709	37.56826 126.97783	#피자알블로 #100% #오늘은 좋은 일이 가득할것입니다 #점심 #피자 #맛난피자 #먹고보자 #대한항공 #마일리지 #199000miles #pizza #행복하세요 @ 아이카샵
3	1	3.7265	37.53853101 127.0010210	맛있는 점심 ~ @ 비체나
4	1	7.3882	37.55515285 126.923225	이 밥이 얼마나 맛있는지 알아요? #점심 #구동 @ 흥대어던가
5	1	7.6223	37.503393 127.0049737	Minions macaron! 하하 너무 귀엽네 미니언즈 마카롱! 율회사 스태프들과 행복한 점심시간! #일상 #점심 #마카롱 #미니언즈 #맛스타그램 #먹스타그램 #디저트 #테일리...
6	0	1.5240	37.57754976 126.9885474	문어 는 못먹어도 문어 #카르파치오 는 먹지요 #점심 #종로 #종로맛집 #맛스타그램 #먹스타그램 #foodstagram #instafood @ Dining in SPACE
7	0	1.8056	37.56969652 126.984036	눈여겨봤던 #팬스테이크 집에 와봤음! 사촌동생 칭찬 #점심~ ^^ 점심부터 지글지글 꼬기 남남 팬스테이크집 한 세군대쯤 가봤던것 같은데 여기가 젤 맛나넵 스테이크 라이스는 좀...
8	0	2.1332	37.5636473 126.9825027	점심 #seoulsearching @ Today
9	0	2.3709	37.56826 126.97783	표정이.. #먼치킨 #munchkin #catstagram #cat #고양이 #야미 #배고프다 #점심 @ 아이카샵
10	0	2.4147	37.5604964 126.9810446	점심2차 □ @ 신세계 본점 Shinsegae

4. 결과 및 성능 평가

본 장에서는 실제 구현된 시스템을 소개하고 성능평가를 수행한다. 본 논문에서 제안하는 시스템의 개발환경은 <Table 6>과 같다. 시스템은 이클립스에 Spring IDE를 추가하여 MVC 패턴으로 구현했으며, '축제', '공연' 두 개의 키워드를 시스템에 적용하여 트위터 API를 통해 수집한 트윗을 형태소 분석 한 후 감정 분석 과정을 거친 추천 트윗을 구글 지도 API에 마커로 표시하였다. 성능평가는 '맛집', '공

〈Table 6〉 Experimental Environment

division	detailed contents
CPU	Intel Core(TM) i5-460 @ 3.20GHz
RAM	8GB
HDD	1TB
OS	MS Window7 64bit
DEV.TOOL	Eclipse Mars.1 Release(4.5.1) + Spring IDE 3.7.1.201510041102-RELEASE
WAS	Apache Tomcat 7.0.65
OS	Oracle 11g

연' 두 개의 키워드를 시스템에 적용하여 나온 감정 분석 결과에 대한 정밀도와 재현율을 산출했다.

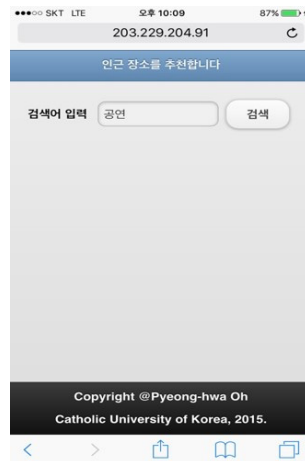
4.1 시스템 구현 화면

<Figure 3>은 2015년 11월 6일부터 2015년 11월 22일까지 '서울 빛초롱축제'가 개최됐던 중

로 인근에서 '축제' 키워드로 추천된 결과 화면이다. 시스템은 인근에서 발생한 축제와 관련된 트윗을 수집하여 사용자에게서 가깝고 추천점수가 높은 10개 트윗을 마커로 표시한다. 또한 infoWindow를 사용하여 트윗 내용과 해당 트윗에 대한 추천 점수를 출력한다. <Figure 4>는 동일한 방법으로 홍대입구역 인근에서 발생한 '공연' 관련 트윗을 추천하는 결과이다.



<Figure 3> Example of Recommend Locations about '축제' at Jongno



<Figure 4> Example of Recommend Locations about '공연' at Hongdae

4.2 트윗 감정 분석 성능 평가

빅데이터 분석 전문조직인 제일기획의 DnA 센터가 2015년 하반기 6개월 동안 디지털 패널 3,200명이 입력한 ‘서울 시내 핫플레이스’ 관련 검색어 25만 8,000건을 분석한 결과 가장 많은 검색 횟수를 기록한 최고의 핫플레이스는 홍대인 것으로 조사됐다. 홍대는 모든 연령대에서 핫 플레이스 검색순위 1위에 올랐다(29.8%). 강남역(12.9%), 신촌(11.8%), 이태원(10.3%), 논현동(7.0%) 등이 2~5위를 차지했다. 이에 본 논문에서는 상위 10개 지역에 ‘맛집’과 ‘공연’ 두 개의 키워드를 적용해 수집한 트윗을 이용해 시스템의 추천결과를 분석하였다. ‘맛집’은 Yoo [14]의 연구에서 모바일로 가장 많이 검색하는 장소가 음식과 관련된 장소였기 때문에 선정했다. 또한 ‘공연’은 ‘취미’의 분류에 속하고 ‘길거리 공연’, ‘라이브 공연’ 등 갑작스런 이벤트를 검색할 수 있는 키워드이므로 선정했다. <Table 7>은 구글 지도에서 상위 10개 지역 위도와 경도 좌표를 검색한 결과를 정리한 표로, 검색된 각 지역의 위도와 경도 좌표를 사용자의 현재위치로 설정하고 실험을 진행했다.

<Table 7> Hot Place at Seoul

rank	region	latitude, longitude coordinate
1	홍대	37.5568422, 126.9215853
2	강남역	37.4979462, 127.0254323
3	신촌	37.5597722, 126.9401193
4	이태원	37.5388618, 126.9834483
5	논현	37.5112102, 127.0195393
6	가로수길	37.5205788, 127.0207381
7	삼청동	37.5879444, 126.9748583
8	인사동	37.5720491, 126.9838173
9	북촌	37.5816943, 126.9815909
10	신사동	37.5227618, 127.0120044

실험은 수집-분석-분류 각각의 단계에서 산출된 산출물을 식 (2)와 식 (3)을 적용하여 개별 트윗 내의 감정단어로 구분된 긍정, 부정 성향의 트윗에 대한 정밀도(Precision)와 재현율(Recall)을 구했다. 식 (2)는 정밀도로 시스템에 의해 감정 분석된 트윗 개수 중 정확히 긍정, 부정으로 분류된 트윗 수를 의미한다. 식 (3)은 재현율로 긍정 및 부정의 단어를 포함하는 트윗 개수 중 긍정 및 부정으로 정확히 분류된 트윗 수를 의미한다.

$$\text{정밀도} = \frac{\text{정확히분류된트윗개수}}{\text{감정분석을통해감정이분류된트윗개수}} \quad (2)$$

$$\text{재현율} = \frac{\text{정확히분류된트윗개수}}{\text{긍정및부정의단어를포함한트윗개수}} \quad (3)$$

10번의 실험결과 10개 지역에서 총 210개의 트윗을 수집했으며 그 중 실제 긍정, 부정의 단어를 포함한 트윗은 122개였고, 감정 분석을 통해 감정이 분류된 트윗의 개수는 65개, 그 중 정확히 분류된 트윗의 개수는 46개였다. 이를 통해 계산한 시스템의 정밀도와 재현율은 <Table 8>과 같다. 실험 결과, 정밀도는 46/65로 0.71이며, 재현율은 46/122로 0.38이다.

<Table 8> Evaluation Criteria of Sentiment Analysis, Precision and Recall

item	result
number of average experimental data	210
number of tweet containing sentiment word	122
number of tweet classified positive or negative	65
number of tweet classified accurately	46
precision	0.71
recall	0.38

0.38의 재현율은 이미 구축된 감정사전을 기반으로 개별 키워드를 분석하는 기법이 신조어나 은어 등을 포함하지 못하는 감정사전의 대응성능의 문제라고 볼 수 있다. 반면 낮은 재현율에도 0.78의 정밀도를 도출했다는 것은 분석 단위가 키워드만으로 이루어졌어도 일정 수준의 장소추천이 가능함을 의미한다. 최근 사용되는 표제어를 포함하는 감정사전의 도입과 키워드에서 문맥으로 감정분석 단위 확대, 감정패턴을 갖는 데이터 셋 확보를 통한 패턴분석을 적용한다면 향상된 정밀도와 재현율을 기대할 수 있음을 본 실험을 통해 확인할 수 있다.

다만, 실험 과정 중 작성위치와 실제 언급된 장소가 달라 위치정보와 관계없는 트윗이 분석되는 사례가 발견되기도 하였다. 실험에서는 해당 트윗을 오류라고 판단했으며 전체 트윗 210개 중 23개에 대한 내용, 전체의 약 11%의 트윗이 사용자의 위치와 관련 없는 장소에 대한 내용을 언급하고 있었다. 이는 추후 감정 분석에 문맥 판단이 추가되어야 보다 정확한 결과를 얻을 수 있음을 말해준다.

5. 결론 및 향후 연구

본 연구는 사용자가 원하는 장소에 대한 키워드를 입력받아 사용자 주변에 위치하는 장소 정보를 추천하는 시스템을 제안했다. ‘맛집’, ‘축제’, ‘공연’ 등의 키워드를 통한 성능 평가는 시스템의 발전가능성을 보여준다. 성능 평가 결과 71% 정밀도와 38%의 재현율을 보였으며, 이는 추천된 결과가 실제 상황에서 추천 시스템으로 활용될 수 있음을 의미한다. 이에 제

안하는 시스템은 모바일 사용량이 늘어나고 그에 맞게 다양한 서비스들이 제공되는 가운데 공간 데이터를 활용한 검색 시스템으로써 의의가 있다. 또한 단순한 거리, 시간 계산으로 추천하는 시스템이 아닌 본 논문에서 구축한 감정사전으로 감정 분석된 점수를 통해 사용자에게 더 유용한 장소 정보를 추천했다는 차이점을 볼 수 있다. 앞으로 검색에 대한 질의 결과 품질을 높이기 위해 트위터 데이터 외에 위치 정보를 제공하는 다른 소셜 네트워크 서비스인 페이스북, 인스타그램 등을 추가한다면 더욱 효율성 높은 시스템으로 활용될 수 있을 것이다.

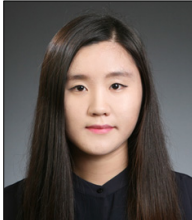
제안하는 시스템은 소셜 데이터가 포함하고 있는 위치정보를 활용해 공간추천이 가능함을 보였다. 그러나 실험결과 감정사전 기반의 키워드 매칭만으로는 추천에 한계가 있음도 함께 보이고 있다. 이에 향후 연구 과제로 분석 단위를 키워드에서 문맥으로 넓히고 이를 통해 분석 데이터 셋을 확보한 후, 패턴분석 알고리즘을 적용하여 분석의 정확도 향상을 꾀할 것이다. 또한 미리 입력한 키워드를 저장하고 이동하는 위치에 따라 장소를 추천하거나 사용자가 원하는 시간에 또는 주기적으로 결과를 푸쉬(Push) 하는 추천 시스템으로 발전시킬 것이다.

References

- [1] Alexa, "Top Sites in South Korea," <http://www.alexa.com/topsites/countries/KR>, 2015.

- [2] Google Play, "The popular App," <https://play.google.com/store/apps/top>. 2015.
- [3] Hu, B., Jamali, M., and Ester, M., "Spatio-Temporal Topic Modeling in Mobile Social Media for Location Recommendation," *IEEE International Conference on Data Mining*, pp. 1073-1078, 2013.
- [4] KLDP.net, "Hanannum morpheme Analyser 0.8.3 manual for user," <http://kldp.net/projects/hannanum/forum/316173>, 2011.
- [5] KT economy management lab, "The first half 2015 mobile trend report," <http://www.digieco.co.kr/KTData/Report/FILE/PDF/2015%EB%85%84%20%EC%83%81%EB%B0%98%EA%B8%B0%20%EB%AA%A8%EB%B0%94%EC%9D%BC%20%ED%8A%B8%EB%A0%8C%EB%93%9C201507091436405462845.pdf?>, 2015.
- [6] Lee, G. H. and Lee, K. J., "Twitter sentiment Analysis for the Recent Trend Extracted from the Newspaper Article," *Korea Information Processing Society*, Vol. 2, No. 10, pp. 731-738, 2013.
- [7] Lee, J. Y., "Development of a spatial information extraction and representation systems utilizing social media as spatial big data," Master's thesis, The Graduate School University of Seoul, 2014.
- [8] Lim, J.-S. and Kim, J. M., "An Empirical Comparison of Machine Learning Models for Classifying Emotions in Korean Twitter," *Journal of Korea Multimedia Society*, Vol. 17, No. 2, pp. 232-239, 2014.
- [9] Oh, S. J., "A Model for Ranking Semantic Associations in a Social Network," *The Journal of Society for e-Business Studies*, Vol. 18, No. 3, pp. 93-105, 2013.
- [10] Park, I. J. and Min, K. H., "Making a List of Korean Emotion Terms and Exploring Dimensions Underlying Them," *Korean Journal of Social and Personality Psychology*, Vol. 19, No.1, pp. 109-129, 2005.
- [11] Shin, J. H., Han, Y. S., Park, Y. C., and Choi, K. S., "An HMM Part-of-Speech Tagger for Korean Based on Wordphrase," *Proceedings of the 6th Annual Conference on Human and Cognitive Language Technology*, pp. 389-394, 1994.
- [12] Twitter4j standard library, <http://twitter4j.org>, 2015.
- [13] Wikipedia, "List of countries by Internet connection speeds," https://en.wikipedia.org/wiki/List_of_countries_by_Internet_connection_speeds, 2015.
- [14] Yoo, J. H., "A Study on Place Search Behaviors of Mobile Users : Focusing on analysis of place search log of mobile users," Master's thesis, The Graduate School Seoul National University, 2014.

저 자 소 개



오평화

2013년

2013년~2016년

관심분야

(E-mail: oph312@nate.com)

가톨릭대학교 컴퓨터공학과 (학사)

가톨릭대학교 컴퓨터공학과 (석사)

소셜 네트워크분석, 데이터마이닝, XML, 데이터베이스



황병연

1986년

1989년

1994년

1994년~현재

1999년~2000년

2007년~2008년

관심분야

(E-mail: byhwang@catholic.ac.kr)

서울대학교 컴퓨터공학과 (학사)

KAIST 전산학과 (석사)

KAIST 전산학과 (박사)

가톨릭대학교 컴퓨터정보공학부 교수

(美) 미네소타대학교 방문교수

(美) 캘리포니아주립대학교 방문교수

소셜 네트워크분석, XML 데이터베이스, 정보검색,
데이터마이닝, 지리정보시스템