

# 트위터 기반의 트렌드 뉴스 추천 기법<sup>1</sup>

김대용, 김대훈, 황인준  
고려대학교 전자전기전파공학부  
e-mail : {ritgd05, kdh812, ehwang04}@korea.ac.kr

## Trend-based Trend News Recommendation Scheme

Daeyong Kim, Daehoon Kim, Eenjun Hwang  
School of Electrical Engineering, Korea University

### 요 약

최근 스마트폰의 사용이 보편화되면서 많은 양의 온라인 뉴스가 다양한 경로를 통하여 서비스 되고 있다. 한편, 실시간으로 제공되는 뉴스의 양이 방대해지면서, 언론사에서 톱 뉴스로 제공하는 토픽과 달리, 실제 사용자들에게 화제가 되고 있는 토픽을 선별하는 데 어려움이 있다. 많은 사용자 들이 실생활에서 작성하고 공유하는 트위터는 실제 사람들 사이에 화제가 되고 있는 토픽을 담고 있는 경우가 많다. 이러한 트렌드를 뉴스와 연계시키면 화제가 되는 트렌드 뉴스를 사용자에게 제공할 수 있다. 본 논문에서는 클라이언트-서버 모델을 기반으로 실시간으로 사용자 트위터를 분석하여 추출된 트렌드를 기반으로 관련 뉴스를 검색하여 제공하는 시스템을 제안한다. 클라이언트를 통해 수집한 트위터 단문에서 서버는 화제가 되고 있는 트렌드를 추출하고, 이를 기반으로 Google 등을 통해 관련 뉴스를 검색하여 클라이언트에게 전달한다. 이 모든 과정을 실시간으로 제공하기 위한 알고리즘을 제안하고 프로토타입 시스템을 통하여 그 성능을 평가한다.

### 1. 서론

최근 스마트폰의 사용이 보편화되면서 온라인 뉴스 서비스가 크게 증가하였고 이에 따라 사용자들은 다양한 채널을 통해 뉴스를 접할 수 있게 되었다. 온라인 뉴스 서비스는 기존 신문과 달리 웹에 접속하여 클릭만으로 뉴스를 접할 수 있기 때문에 편의성에서 장점을 보이지만, 실시간으로 모든 분야의 뉴스가 업데이트 되기 때문에 언론사가 톱뉴스로 선택하여 제공하는 기사 외에 실제로 사용자들에게 화제가 되고 있는 토픽을 선택하는 것이 어렵다. 따라서, 사용자들이 자신의 관심사를 글로 남기는 SNS 서비스의 하나인 트위터를 활용하여 트렌드 뉴스를 검색한다면, 사용자에게 실제 화제가 되고 있는 뉴스를 추천하는 것이 가능할 것이다.

이를 위한 다양한 연구가 진행되었는데 [1]에서는 트위터 분석을 통한 트렌드 추출 기법을 제시하였다. 뉴스 추천에 있어서는 사용자의 과거 검색을 이용한 뉴스 추천[2] 및 사용자 방문 횟수 외의 뉴스 도메인의 온톨로지 기반[3] 혹은 사용자의 트위터 계정으로 접근한 뉴스 RSS 기반[4]으로 뉴스를 추천하는 방식이 제안되었다. 이러한 연구들은 사용자들이 자신의 정보를 입력하여야 뉴스를 추천 받을 수 있고, 여타 사용자들에 관심의 대상이었던 토픽은 추천 받기 힘들다. 따라서, 이를 해결하기 위해 본 논문에서는 트위터를 활용하여 실시간 트렌드 및 그와 연관된 키워

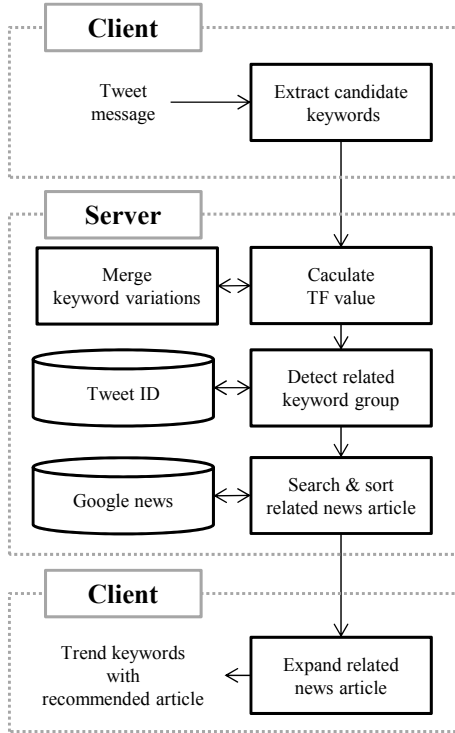
드를 추출하고 이를 기반으로 뉴스를 검색하는 클라이언트-서버 기반의 시스템을 제안한다. 먼저 클라이언트는 트윗에서 후보 키워드를 찾아내어 서버로 전달하고, 서버는 트렌드 키워드를 선출한 뒤 연관 키워드 그룹을 설정한다. 이를 기반으로 Google News를 통해 검색하고 해당 결과를 파싱하여 최신 트렌드와 연관된 뉴스 정보를 추천한다. 추천된 뉴스 정보를 클라이언트에 전송함으로써 사용자는 최근 화제와 관련된 뉴스를 제공받는다.

### 2. 시스템 구성

본 논문에서 구축한 시스템은 클라이언트와 서버로 구성되어 있으며 전체적인 구성은 (그림 1)과 같다. 트렌드 추출기는 기존 연구인 실시간 트위터 분석을 통한 트렌드 및 연관 키워드 추출[1] 기법을 사용하였다. 모바일 클라이언트는 트윗을 GPS 센서 등을 통한 메타 데이터와 함께 제공하며, 이는 사용자의 위치나 상황을 판단하는 데 활용된다. 한편, 서버는 해당 정보들을 이용해서 트렌드 키워드를 찾아내 뉴스를 추천해주는 역할을 한다. 이렇게 검색된 뉴스들은 사용자가 필요로 할 때 사용자의 클라이언트로 전달된다.

시스템에서는 우선 불특정 다수의 모바일 클라이언트에서 후보 키워드 추출 과정이 이루어진다. 후보 키워드 추출은 키워드 대부분이 고유명사라는 점을 활용하여 대문자로 시작되는 연속적인 단어 및 따옴

<sup>1</sup> 본 연구는 지식경제부 및 정보통신사업진흥원의 대학 IT 연구센터 지원사업(NIPA-2013-H0301-13-3006)과 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업(2012-0007202) 지원을 받아 수행된 것임



(그림 1) 시스템 구성도

표로 강조된 단어를 후보 키워드로 추출해 낸다. 다음으로, 서버에서는 각 후보 키워드 단위로 빈도수를 계산하는 키워드 수집 과정이 실행된다. 이 때 키워드의 결과는 지역 구분을 위해 2 차원 벡터의 형태로 저장된다. 이후에는 키워드 병합 과정이 진행된다. 줄임말, 이철 동의어, 오타자 병합이 진행되며, 연관 키워드 그룹 설정 및 Stopword 제거가 이루어진다. 추출 과정의 구체적인 방식은 선행 연구에서 설명하였다.

최종적으로, 획득한 트렌드 키워드를 Google News 를 통해 검색 결과를 파싱하여 뉴스 정보를 추천한다. 이 때, Google News 자체에서는 검색 결과의 정확성을 향상 시키기 위해 방문 횟수 등의 정보를 기반으로 사용자의 선호도에 따라 랭킹해준다. 본 시스템은 이를 이용하여 검색 결과상의 가장 랭킹이 높은 상위 10 개 뉴스를 선택한다. 또한, 트윗 내부에 연관 검색어 그룹으로 설정된 키워드들은 하나의 쿼리로 검색하여서 검색의 정확도를 향상시켰다. 검색된 결과는 뉴스가 작성된 시간 및 트렌드 키워드 순위 기반으로 점수를 부여한다. 뉴스가 작성된 시간은 10 분 단위로 검색 시간대와의 차이를 점수로 부여하며 트렌드 키워드는 순위별로 점수를 부여한다. 이때 연관 키워드 그룹 내 키워드끼리는 동일한 점수를 부여한다. 이를 통해 점수 기반으로 뉴스 검색 결과가 정렬되어 가장 최근에 이슈가 된 뉴스 순으로 보여준다. 한편, 뉴스 검색 결과 랭킹은 시간대별로 이뤄지며 특정 지역의 트렌드 기반 뉴스 검색 시에는 Google News API 를 통해서 해당 지역의 뉴스만을 검색하고 랭킹하여 사용자에게 제공한다. 또한, 중복된 결과를 제거하기 위해 뉴스 매칭 시 매칭된 결과를 분석하여 중복된 결과를 리스트에서 제거한다. 마지막으로, 사용자가 특정 뉴스를 선택할 시 정렬된 결과 외에 뉴스 내의 키워드와 연관된 모든 뉴스 결과를 보여주어 사용자가 특정

토픽과 연관된 뉴스에 관해서 확장된 정보를 접할 수 있게 하였다.

<표 1> 2013년 3월 13일 23시 뉴스 추천 결과

Trend keywords	Retrieved news
JUSTIN BIEBER	'Justin Bieber lashes out at media's 'countless lies'...'
DIA	
MADRID	'Lawmakers unmoved by President Obama's Capitol...'
NIALL DUBLIN	
TAYLOR SWIFT	
PRESIDENT OBAMA	'Justin Bieber makes it through his Madrid show...'
RAY LEWIS	'Vote For Ray Lewis, Ray Rice On Madden Cover'
ESPN NFL	
ENGLISH	'Taylor Swift romance rumors made ...'
LONDON	'Niall Horan crying: One Direction Dublin, Ireland...'
HARRY	

### 3. 실험결과

제한된 기법의 성능 평가를 위해 MATLAB 환경에서 트렌드 뉴스 추천을 위한 프로토타입 시스템을 구현하였다. 또한 실험에 사용된 트윗들은 Twitter Streaming API 를 통해 수집되었다.

표 1 은 2013년 3월 13일 23시의 261,200 개의 트윗에서 추출된 상위 10 개 트렌드 키워드와 연관 키워드 그룹을 이용하여 추천된 뉴스 결과를 보여준다. 결과적으로, 트렌드 키워드에 대한 직관적인 이해와 트윗 사용자에게 화제가 되었던 뉴스 리스트가 잘 추천되는 것을 알 수 있다.

### 4. 결과

본 논문에서는 실시간으로 사용자의 트윗을 분석하여 트렌드를 추출하고, 이와 연관된 뉴스를 검색하여 제공하는 클라이언트-서버 기반의 검색 시스템을 구축하였다. 이를 통해 사용자는 현재 이슈가 되고 있는 트렌드 뉴스를 손쉽게 제공 받을 수 있음을 실험을 통하여 보였다.

### 참고문헌

- [1] 김대용, 김대훈, 황인준 “실시간 트윗 분석을 통한 트렌드 및 연관키워드 추출” 한국정보처리학회 춘계학술발표대회, 2012, pp. 1710-1712.
- [2] Jianhui Liu, Peter Dolan and Elin R. Pedersen. “Personalized news recommendation based on click behavior,” In Proc. of the international conference on intelligent user interfaces, 2010, pp. 31-40.
- [3] Wouter Ijntema, Frank Goossen, Flavius Frasinca and Frederik Hogenboom. “Ontology-based news recommendation” In Proc. of the EDBT/ICDT workshops, 2010
- [4] Oewn Phelan, Kevin McCarthy and Barry Smyth “Using twitter to recommend real-time topical news” In Proc. of the ACM conference on Recommender systems, 2009, pp. 385-388.