

# 정치인의 지지율과 인터넷 뉴스 기사량의 상관관계 분석

이필수\* 이윤정\* 우균\*

\*부산대학교 컴퓨터공학과

e-mail: {skyz1104, leeyj01, woogyun}@pusan.ac.kr

## An Analysis of the Correlation Between Politicians Approval Rating and the Amount of Internet News Articles

Pil-Su Lee\* Yun-Jung Lee\* Gyun Woo\*

\*Department of CSE, Pusan National University

### 요 약

현재 인터넷 공간은 사람들의 관심사나 사회적인 이슈들을 반영하고 있다. 사회적으로 어떤 사건이 발생하면 그 사건에 관한 뉴스 기사나 관련된 다양한 콘텐츠들이 생성되어 여러 사람들에게 소비되고 공유된다. 뿐만 아니라 이와는 반대로 인터넷 공간에서 사람들에게 많은 관심을 받거나 이슈가 된 사건이 사회적인 관심거리가 되기도 한다. 최근에는 인터넷 공간에서 발생하는 정보 검색이나 콘텐츠 생성 패턴을 분석하여 실제 사회에서의 이슈나 트렌드를 예측하려는 연구가 활발히 진행되고 있다. 이 논문에서는 인터넷을 기반으로 분석한 자료와 전문 기관에서 분석한 자료의 상관관계를 분석하고자 한다. 그 중 최근 뉴스나 콘텐츠가 많이 생산되는 2012년 대통령 선거 후보에 관한 인터넷 뉴스 기사량과 전문조사 기관에서 발표한 각 후보의 지지율을 보이고 두 자료 간의 상관관계를 분석한다. 그리고 실험 결과로 대선 후보들의 기사 점유율과 발표된 지지율에 높은 상관관계가 있음을 보인다.

### 1. 서론

오늘날 인터넷 공간은 대중의 관심이나 사회적인 이슈를 반영하고 있다. 사회적으로 이슈가 되는 사건이 발생했을 때 인터넷 공간에서는 그 사건에 관한 많은 콘텐츠가 생성되어 여러 사람에게 공유된다. 뿐만 아니라 이와는 반대로 인터넷 공간에서 여러 사람의 관심을 끈 사건이 실제 사회에서 이슈가 되기도 한다. 최근 싸이의 '강남스타일'이 전 세계적으로 인기를 끌게 된 것도 유튜브 상에 공개한 뮤직 비디오의 인기가 실제 사회로 확산된 경우라고 할 수 있다.

이렇듯 인터넷 공간에서 나타나는 여러 현상들이 실제 사회 이슈나 사람들의 관심사를 반영함에 따라 인터넷 공간에서 나타나는 여러 현상들을 분석하여 이를 활용하려는 연구가 활발히 이루어지고 있다. 실제로 구글에서는 구글 트렌드를 통한 검색량을 분석하여 미국의 독감 유행지역과 시기를 예측하였다[1]. 또한 최근에는 SNS 데이터 분석을 통해 SNS에서 언급되는 대선후보에 관한 트윗 전량을 분석하여 대선 후보들의 특징 비교 및 분석하는 사이트도 개설되었다[8].

정치 분야에 있어서도 인터넷은 많은 영향을 미친다[9]. 2011년 10월 26일 서울시장 재보선에서도 파워 트위터안과 소셜테이너의 소신 발언 등으로 특정 후보의 지지성향이 선거여론에 지대한 영향을 미쳤다[2]. 2012년 대통령 선거를 앞두고 있는 상황에서 각 후보들의 지지율 변화는

많은 사람들의 관심사이다. 일반적으로 후보들의 지지율은 전문 조사 기관을 통해 전화나 이메일 등의 방법으로 조사된다. 이런 경우, 지지율에 영향을 미치는 특정 사건이 발생하더라도 즉각적인 지지율 변화를 파악하기 어렵고, 특히 개인들은 매체에서 발표된 지지율 정보 외에는 알 수 없다.

이 논문에서는 2012년 대선 후보자들에 관한 인터넷 뉴스 기사량을 기반으로 각 후보의 기사 점유율을 구하고, 설문 조사를 통한 각 후보의 지지율 사이의 상관관계를 분석하고, 실험을 통해서 두 자료의 유사함을 보인다.

이 논문의 구성은 다음과 같다. 2장에서는 인터넷의 정보나 변화가 실제 사회에 끼치는 영향에 관한 기존 연구들을 살펴본다. 3장에서는 설문조사 전문 기관의 지지율에 관해서 설명한다. 4장에서는 인터넷 뉴스 기사량과 전문 기관의 지지율 사이의 상관관계를 분석하고, 마지막으로 5장에서 결론을 맺는다.

### 2. 관련연구

사람들의 관심이나 사회적인 이슈를 반영하는 인터넷의 정보에 초점을 맞춘 연구로 검색어를 통한 유행성 독감 감지 및 예측 시스템을 들 수 있다[4]. 이 연구는 독감 관련 검색어에 대한 조건을 통해서 검색어를 추출하고, 해당 검색어의 검색 동향과 실제 환자의 데이터와의 상관관계를 분석하였다. 그 결과 검색 동향과 실제 데이터 사이

이 논문은 2010년도 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(NRF-2010-371-B00008).

에 높은 상관관계가 있음을 보였다. 이와 비슷한 연구로는 Jeremy의 연구를 들 수 있다[1]. 이 연구에서는 구글 웹 검색 기록을 이용하여 독감 유행의 시기와 시기를 예측하였고, 전문 기관의 자료와 비교한 결과 높은 상관관계를 보였다. 그리고 Hyunyoung Choi는 구글 트렌드를 통해 현재를 예측하는 시스템을 개발하였다[5]. 자동차의 월별 판매량 혹은 부동산 경기 등을 예측하였다.

### 3. 데이터 수집

설문조사 전문 기관의 지지율 조사는 전화 설문 혹은 인터넷 설문을 통해서 이루어지며 보통 백여 명에서 천 명 정도로 실행하게 된다. 하지만 이러한 조사는 지지율에 큰 영향을 주는 사건을 즉각적으로 반영하기 힘들며 설문조사 참가자들이 임의로 선정되기 때문에 지지율은 설문조사 기관마다 약간씩 다르게 나타난다. 또한 인터넷 설문조사의 경우 동일한 설문 참가자가 여러 개의 설문조사에 응답할 수도 있기 때문에 정확성이 떨어질 수도 있다[3]. 따라서 이 논문에서는 지지율의 신뢰성을 위이기 위해 서로 다른 조사기관에서 발표한 자료를 함께 사용한다. K기관, R기관에서 발표한 2012년 대선 후보자들의 지지율 데이터는 (표 1)과 같다.

후보 명	K기관	R기관	평균
P후보	41.0%	41.3%	41.1%
A후보	23.0%	34.4%	28.7%
M후보	15.5%	23.7%	19.6%

(표 1) 설문조사 전문기관의 지지율  
(조사기간:2012년 9월 10일 - 11일)

2012년 9월 10일부터 2012년 9월 11일까지 2일 동안 P후보의 지지율은 두 기관의 조사에서 각각 41.0%와 41.3%로 세 후보 중 가장 앞서는 것으로 나타났다. A후보의 지지율은 K기관에서는 23.0%, R기관에서는 34.4%로 나타났고, M후보의 경우도 두 기관에서 각각 15.5%와 23.7%로 약간의 차이를 보였다. 조사 기관에 따라 지지율은 차이가 있었으나 세 후보의 지지율 순위는 동일하게 조사되었다. 이 논문에서는 각 후보의 지지율 데이터의 신뢰성을 높이기 위해 두 기관에서 발표한 조사 결과의 평균을 각 후보의 지지율 데이터로 사용한다.

### 4. 실험

전문 조사 기관의 설문 조사 결과와 인터넷에서 생성되는 뉴스 기사량 사이의 상관관계를 조사하기 위하여 인터넷 포털 사이트 네이버에 등록된 뉴스 기사를 대상으로 실험 데이터를 생성하였다. 네이버의 뉴스 카테고리에 등록된 모든 뉴스를 대상으로 지정된 기간 동안 후보들과 연관된 뉴스 기사를 검색하였다. 이 논문에서는 기사의 제목과 본문에서 각 후보의 이름이 나타난 기사를 해당 후

보와 관련 있는 기사로 간주한다. 2012년 9월 10일부터 2012년 9월 11일까지 네이버 뉴스에 등록된 기사 중에서 각 후보와 연관된 뉴스 기사량은 아래 (표 2)와 같다.

P후보	A후보	M후보
2,294건	1,404건	815건

(표 2) 후보별 인터넷 뉴스 기사량  
(뉴스 등록 시간: 2012년 9월 10일 - 11일)

(표 2)에서 P후보에 관한 기사가 2,294건으로 가장 많이 생성되었고, A후보와 M후보는 각각 1,404건과 815건의 기사가 생성되었다.

각 후보의 기사 점유율은 인터넷 뉴스 기사량을 토대로 아래의 식 1과 같이 구할 수 있다.

$$x_i / \sum x_i \times 100, \quad x_i \in X \quad (1)$$

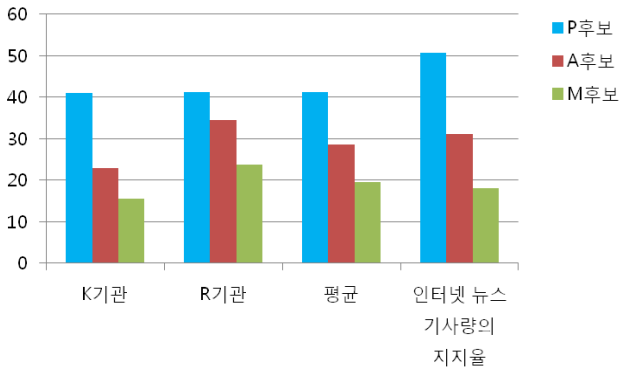
$x_i$ 는 후보  $i$ 의 뉴스 기사량이며,  $X$ 는 후보의 전체 집합을 나타낸다. 식 1을 통하여 구해진 기사 점유율은 아래의 (표 3)과 같다.

P후보	A후보	M후보
50.8%	31.1%	18.1%

(표 3) 인터넷 뉴스 기사량을 기반으로 구한 후보별 기사 점유율

(표 3)에서 P후보의 기사 점유율이 50.8%로 가장 높게 나타났으며, A후보와 M후보는 각각 31.1%와 18.1%로 나타났다.

[그림 1]은 전문조사 기관에서 발표한 후보별 지지율(표 1)과 인터넷 뉴스 기사량을 기반으로 계산한 후보별 기사 점유율(표 3)을 비교한 것이다. [그림 1]에서 전문조사 기관에서 발표한 세 후보의 지지율과 제안된 인터넷 기사 점유율의 실제 값에는 차이가 있지만 세 후보의 순위는 동일하게 나타남을 알 수 있다. 세 후보의 지지율과 기사 점유율 사이의 상관관계수는 0.9996으로 나타났다. 따라서 인터넷 기사량을 기반으로 한 기사 점유율이 설문조사를 통해 조사된 지지율과 상관관계가 있다고 할 수 있다. 즉, 인터넷 기사량을 통해 실제 각 후보들의 지지율을 추정해 볼 수 있을 것이다.



[그림 1] 각 후보별 전문기관에서 조사한 지지율과 인터넷 기사 점유율 비교

### 5. 결론 및 향후과제

이 논문에서는 대선 후보들에 관한 인터넷 뉴스 기사량과 실제 지지도 사이에 어떤 상관관계가 존재하는지 분석하였다. 분석을 위해 전문기관에 조사한 각 후보의 지지율을 대중의 선호도라고 간주하였다. 그리고 후보들의 인터넷 기사 점유율을 구하기 위해 네이버 뉴스 카테고리에 등록된 인터넷 기사들을 대상으로 각 후보의 이름이 거론된 뉴스 기사량을 수집하였다. 두 곳의 전문조사 기관에서 발표한 2012년 9월 10일부터 2012년 9월 11일까지 2일 동안의 대선 후보 각각에 대한 평균 지지율은 P후보가 41%, A후보가 28.7%, M후보가 19.6%로 나타났다. 그리고, 동일한 기간 동안 세 후보의 인터넷 뉴스 기사 점유율은 P후보가 50.8%, A후보가 31.1%, M후보가 18.1%로 조사되었다. 세 후보의 지지율과 기사 점유율의 실제 값은 차이가 있지만 후보간 순위는 동일함을 알 수 있다. 따라서 인터넷 뉴스 기사 점유율이 실제 대중들의 지지율을 반영한다고 할 수 있으며, 인터넷 뉴스 기사와 같은 콘텐츠의 검색량이나 생산량을 통해 누구나 손쉽게 어떤 대상에 대한 사람들의 선호도를 추정해 볼 수 있을 것이다.

향후에 인터넷 뉴스 기사 점유율뿐만 아니라 긍정 또는 부정과 같은 기사의 내용을 분석한다면 대중의 선호도를 좀 더 정확하게 예측할 수 있을 것이다.

### 참고문헌

[1] Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brtammer, Mark S. Smilinski & Larry Brilliant, "Detecting influenza epidemics using search engine query data", Nature Vol. 457, pp 1012-1014, 2009.

[2] 김용희, "트위터에서 선거여론 확산구조 예측", 정치커뮤니티 연구, 통원23호, pp 103-139, 2011.

[3] Kevin B. Wright "Researching Internet-Based Populations: Advantages and Disadvantages of Online

Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services" Journal of Computer Mediated Communication, Volume 10, Issue 3, page 00, April 2005.

[4] 박정은, 홍창범, 정진영, 송정길, "검색어를 통한 유행성 독감 감지 및 예측 시스템". 한국인터넷정보학회, 제20회, pp 205-210, 2009

[5] Hyunyoung Choi, Hal Varian, "Predicting the Present with Google Trends", Economic Record, Volume 88, Issue Supplements1, pp 2-9, 2012

[6]리서치뷰, 자료실, [http://rsview.kr/pdboard/bbs/board.php?bo\\_table=data](http://rsview.kr/pdboard/bbs/board.php?bo_table=data), 2012년 10월 10일.

[7]한국갤럽, 한국갤럽 조사 자료실 데일리정치치표 제 37호, <http://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=342>, 2012년 10월 8일.

[8]SNS민생닷컴, <http://www.snsminsim.com/>, 2012년 10월 10일.

[9]김경미, "미국 대통령 선거와 소셜 네트워크 사이트", 한국정보사회학회, 정보와사회 15, pp 1-26, 2009