

SNS의 관심도가 선거결과에 미치는 영향 분석*

최은정, 최세원, 이시연, 김명주
서울여자대학교 정보보호학과

Analysis of the effect of the mention in SNS on the result of election

Eun-Jung Choi, Sea-Won Choi, Se-Yeon Lee, Myhung-Joo Kim
Dept. of Information Security, Seoul Women's University

요 약 SNS(Social Network Service)를 통해 개인의 의견을 표현하거나 논의가 이루어짐에 따라 SNS는 여론의 방향을 판단할 수 있는 새로운 근거로 사용되고 있다. 이러한 변화는 선거에서도 나타나고 있는데, 다수의 유권자들의 SNS를 통해 다양한 의견을 개진하기 때문에 후보자도 새로운 소통의 장으로 활용하고 있다. 본 논문에서는 선거에 대한 SNS 상의 관심 정도가 실제 선거결과에 어떠한 영향을 미치는지 알기 위해서 우리나라 20대 국회의원 선거과정에서의 SNS 관심도를 수집하고 분석하였다. 이러한 결과를 전통적인 여론조사 결과와 비교하여 어느 것이 더 결과를 잘 예측했는지 비교해 보았다. 결론적으로 전통적인 여론조사보다는 SNS관심도가 선거결과에 더 일치하였음을 알 수 있었다.

주제어 : SNS, 빅데이터, 선거예측, 여론조사, SNS 분석

Abstract As individual opinions are expressed and discussed through SNS, SNS is used as a new basis to estimate the direction of public opinion. This change also appears in election. So many voters state their views through SNS, so that candidates utilize it as a new space for communication. In this paper, positive mention in SNS were collected and analysed in the course of the election of Korean 20th Congressman, to understand how the mention on election in SNS affects the result of election. This result was compared with the traditional survey on public opinion, to find out which one more corresponds to the result. In conclusion, mention in SNS coincide more with the result of election than the traditional survey.

Key Words : SNS, Big Data, Big Data Analysis, SNS Analysis, Predict election results

1. 서론

SNS(Social Network Service)는 서비스 다각화를 통해 사용자간의 자유로운 의사소통과 정보공유의 차원을 넘어 새로운 여론 형성의 장으로 발전하고 있다[1,2,3,4]. 이러한 변화는 기존의 신문 방송과 같이 매체 중심의 의

사 전달이 아닌 다양한 개인들의 의견으로부터 시작되는 여론을 형성하고 있다. 2015년 전체 응답자 9,873명 중에서 4,250명이 SNS를 이용하고 있다고 응답하여 이용률은 43.1%로[5] SNS 이용률의 증가는 이를 기반으로 생성되는 데이터 즉 여론 규모의 급속한 팽창으로 나타나고 있다[6]. 이러한 현상은 우리 사회에 여론으로써 미치

* This work was supported by a research grant from Seoul Women's University(2016).

Received 30 December 2016, Revised 31 January 2017

Accepted 20 February 2017, Published 28 February 2017

Corresponding Author: Eun-Jung Choi

(Seoul Women's University)

Email: chej@swu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

는 영향력이 더욱 확대되는 것을 의미하고 있다. SNS 이용자들은 개인의 일상을 공유하고 사진을 올리는데 그치지 않고 특정 제품에 대한 의견, 유명인에 대한 관심도나 선호도를 피력하기도 하고 사회나 정치 현상에 대해서도 의견을 피력한다. 이러한 현상은 선거의 과정에서도 동일하게 나타나고 있다. 예를 들어 특정 후보 관련 동영상수와 블로그에 후보를 언급한 수가 선거 결과에도 동일하게 나타나는 영향을 보여주었다[7].

우리나라도 선거기간 중에 후보들이 SNS를 통해 자신의 공약을 내세우고 지지를 호소하는 것이 일반적인 표현수단이 되었다. 유권자들 또한 SNS를 통하여 자신이 원하는 후보자를 지지하기도 하고, 후보자들의 정책에 대한 활발한 의견을 나누기도 한다. 이를 기반으로 각 정당과 후보자들은 SNS상에서 이슈를 선점하는 것에 대한 중요성을 인식하고 있다.

본 논문에서는 유권자들의 SNS를 기반으로 한 후보자들에 대한 관심이 선거 결과에 어떠한 영향을 미치는지를 분석해 보고자 한다. 이를 위해 SNS 상에서의 후보에 대한 언급, 즉 특정 후보에 대한 관심 정도를 분석하고 실제 선거 결과에 어떠한 영향을 미치는지 분석하였다. 또한 선거 전까지 지속적으로 발표되었던 기존 여론조사와 SNS 상의 후보자 관심도를 비교하여 어느 방법이 더 정확한 예측을 하였는지 비교해 보도록 한다.

2. SNS 여론 분석

소셜미디어(Social Network Service, SNS)가 발달함에 따라 수많은 사람들이 이를 활용하여 정보를 유통시키고, 의견을 공유하고 네트워크를 형성한다. 소셜미디어를 통해서 아는 사람, 모르는 사람에 관계없이 자유로운 의사소통을 할 수 있고, 관심사나 의견에 따라 다양한 사회적 관계를 생성할 수 있다. 상호간에 수많은 데이터를 주고받으면서 타인에게 영향을 미치고 영향을 받기도 하며, 개인적으로 피력한 의견이 SNS 공간 내에서 여론의 기능을 하기도 한다.

수많은 SNS 채널중에서도 '트위터(twitter)'의 역할과 특징은 주목할 만하다. 시장조사업체 스태티스타의 조사에 따르면 2016년 트위터 월간 실질 이용자 수는 3억 1천 만명에 달한다. 국내 이용자는 2012년에 이미 600만명을

돌파했다. 트위터는 익명성이라는 고유의 속성을 바탕으로 몇 가지 특성을 보이고 있다. 우선 팔로잉, 즉 관계 맺기의 일방향성이다. 팔로잉을 하고자하는 해당 트위터 사용자의 허락이 필요하지 않고 본인이 관심 있는, 혹은 의견이나 관심사가 같은 사용자를 팔로우하여 트윗을 자유롭게 받아볼 수 있다. 때문에 특정 트윗의 이용자수(좋아요 수, 리트윗 수, 댓글 수)는 기하급수적으로 늘어날 수 있다. 또한 트위터가 제공하는 리트윗 기능은 특정 트윗을 널리 전달되게 하는데 일조한다[7]. 그리고 트위터는 '트렌드'라는 기능을 제공하여 트위터 상에서 실시간으로 많이 언급되고 있는 단어나 트윗을 실시간 검색어 순위처럼 사용자들에게 제공한다. 이것은 많이 언급되는 트윗들이 트렌드로 나타나는 것을 보여주고 있다. 따라서 트윗을 분석하여 다양한 사회적 흐름을 예측하는 많은 연구[8,9,10]가 이루어지고 있다.

이러한 SNS에서 이슈를 선점한 제품은 흥행을 하고, 선거에서 승리를 거머쥐기도 한다. 일례로 트위터에서 언급된 트윗 수와 영화의 흥행관계를 분석한 논문에 따르면 2010년 9월 13일부터 19일까지 개봉된 영화 가운데 367개의 관련 트윗과 함께 트위터 상에서 336,527번 노출되었던 <시라노 : 연애조작단>이 105개의 관련 트윗과 50,440번 노출된 <레지던트 이블 4 : 끝나지 않은 전쟁>과 비교하여 268만명 대 118만명으로 흥행 상의 우위를 점하였다[11]. 트위터 즉, SNS 상에서의 관심도가 제품이나 영화의 흥행을 결정할 수 있듯, 선거 과정에서 대중의 참여를 분석[12]하거나 선거 결과 예측의 주요한 수단인 여론조사[13, 14]의 대안으로 활용되는 방안에 대한 연구가 이루어지고 있다.

3. 연구가설 및 검증

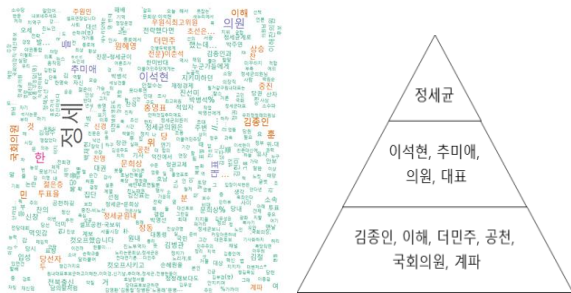
3.1 연구가설

본 연구에서는 SNS 중 트위터 상에서의 데이터양과 내용을 분석하였다. 특정 선거구의 후보자 A와 후보자 B의 데이터를 추출하였을 때, 그 데이터의 양이 후보자 B에 비해 A가 우세하였을 경우 후보자 A가 당선될 것이라고 가정하였다. 즉, 트위터 상에서 이용자들에 의해 언급되는 양이 많을수록, 관심도가 높을수록 선거에서의 승리를 가져갈 확률이 높다고 가정하였다.

3.2.3 종로 선거구

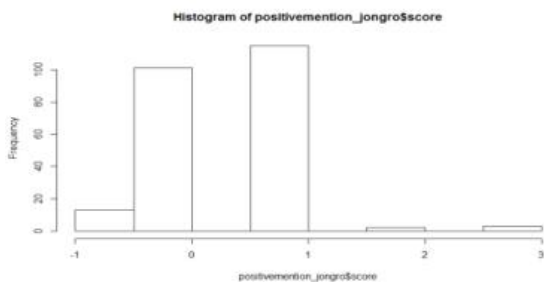
특정 선거구를 입력하였을 때 후보들의 관심도를 측정해보기 위해 선거 전부터 주목 받는 선거구로 언급되었던 서울 '종로'의 트위터 데이터를 같은 기간 동안 수집했다. 여론조사 결과에서는 '새누리당'의 '오세훈' 후보가 우세할 것으로 예상되었다. 하지만 데이터 추출 결과 '더불어민주당'의 '정세균' 후보가 더 많은 빈도로 언급되어 워드클라우드상에 크게 나타남을 확인했다. ('정세균' 후보의 이름에서 '균'이 의미 단어로 추출되어, 워드클라우드 상의 후보자의 이름에서 '균'이 제외되었다.) 다른 빈도수가 높은 단어들 또한 '정세균' 후보자와 관련된 이슈가 다수 포함되어 있음을 확인 할 수 있었다.

실제 개표 결과, SNS상에서의 관심도와 마찬가지로 '정세균' 후보가 승리했다.



[Fig. 4] Data analysis III-1

데이터를 보다 심층적으로 분석해보기 위해서, '정세균' 후보자의 이름을 트위터 상에서 검색했을 때 추출된 데이터에 대해 긍정 부정 언급 분석을 하였다. 다음의 [Fig. 5]에서 보면 0 이상이 긍정을 의미하고 있기 때문에 정세균 후보에 대한 긍정적 언급이 많은 것을 알 수 있다.



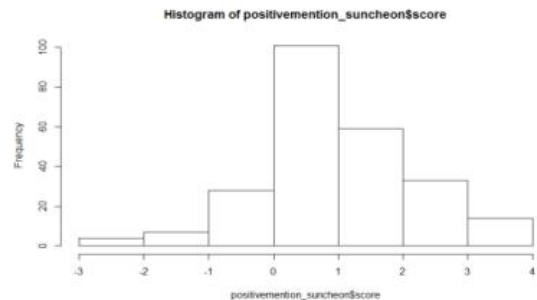
[Fig. 5] Data analysis III-2

3.2.4 순천 선거구

순천 선거구의 경우도 같은 방법으로 분석을 진행하였다. '이정현' 후보와 '노관규' 후보에 대해 여론조사 상에서 우세했던 더불어 민주당의 '노관규' 후보와 비교했을 때 새누리당의 '이정현' 후보의 데이터 양이 압도적으로 많음을 확인할 수 있다. 결과적으로 여론조사와 달리 SNS 상에서의 관심도가 높았던 새누리당의 '이정현' 후보가 당선되었다. 또한 당선된 이정현 후보의 워드클라우드에서 추출된 단어에 대한 긍정 부정 단어 언급 빈도를 그래프로 확인해본 결과 0점 이상의 점수를 얻은 긍정 언급이 많음을 확인할 수 있다.



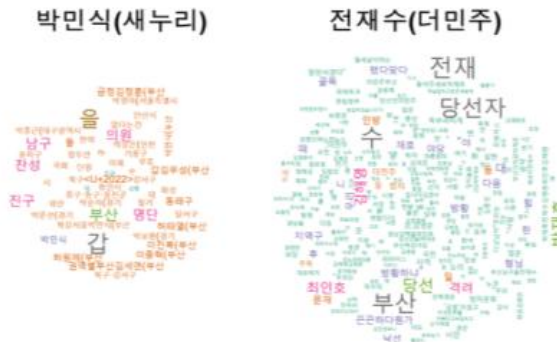
[Fig. 6] Data analysis IV-1



[Fig. 7] Data analysis IV-2

3.2.5 부산 북강서갑 선거구

부산 북강서갑의 데이터 결과를 보면 '전재수' 후보가 '박민식' 후보에 비해 현저히 많은 데이터 양을 가지고 있음을 알 수 있다. 부산 북강서갑 선거구를 대상으로 시행된 여론조사에서는 새누리 '박민식' 후보가 더민주 '전재수' 후보를 누르고 승리할 것으로 예상되었다. 하지만 결과적으로 SNS상에서 언급된 데이터양이 많은 더민주의 '전재수' 후보가 당선되었다.



[Fig. 8] Data analysis V

3.2.6 부산 진구갑 선거구

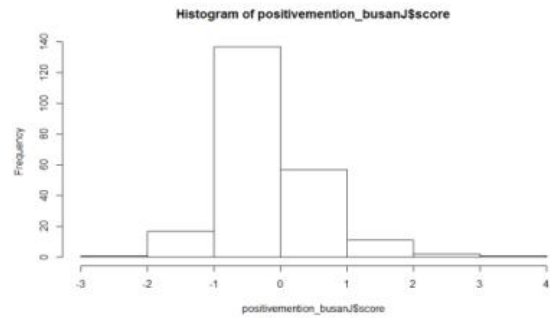
부산 진구갑의 두 후보를 키워드로 트위터 데이터를 검색한 결과를 보면, 더불어민주당의 '김영춘' 후보에 대한 SNS 이용자들의 관심도가 더 높았다. 이에 따라 워드클라우드를 비교해보면 '김영춘' 후보가 현저하게 많은 양과 보다 더 촘촘한 양의 단어 데이터가 추출됨을 확인할 수 있다. 결과적으로 선거 전 실시되었던 여론조사에서 승리가 예상되었던 새누리당의 '나성린' 후보를 제치고 '김영춘' 후보가 승리를 거머쥐었다.



[Fig. 9] Data analysis VI-1

부산 진구갑의 당선자 김영춘 후보자의 데이터의 긍정 부정 언급 빈도수를 확인해본 결과, 중립성을 띄는 구간의 단어를 제외하면 긍정적인 언급의 수가 더 많음을 확인할 수 있다.

부산 진구갑의 데이터 또한 특정 지역구를 검색했을 때 더 많은 수로 언급된 후보가 승리를 가져간다는 본 연구의 가설을 입증하는 사례임을 확인할 수 있었다. 또한 앞선 사례들은 SNS의 데이터를 통해 특정 지역구에서는 특정 정당이 절대적으로 유리하다는 사회적 통념을 벗어난 선거 결과 또한 예측할 수 있음을 확인하였다.



[Fig. 10] Data analysis VI

3.3 결과 요약

본 논문에서 사례로 제시된 양천을, 전주을, 종로, 순천, 부산 북강서갑, 부산 진구갑의 데이터를 요약하여 정리한 결과이다. 각 지역구는 로마숫자로 표시했다. 데이터는 여론조사와 SNS데이터로 구분하였으며 각 부문에서 후보들은 A와 B로 구별하여 유리(△)한 후보자와 불리(▽)한 후보자를 표시하고 최종적으로 당선된 사람을 표기하였다. 위의 사례에서 확인할 수 있듯 SNS 데이터가 우세한 후보자가 당선을 거머쥘 가능성이 높다는 것을 확인할 수 있다.

<Table 1> Summary of election results

Election district	I	
Candidate	A	B
Survey	△	▽
SNS	△	▽
Election results	○	×
Election district	II	
Candidate	A	B
Survey	▽	△
SNS	△	▽
Election results	○	×
Election district	III	
Candidate	A	B
Survey	△	▽
survey	▽	△
Election results	×	○
Election district	IV	
Candidate	A	B
Survey	▽	△
SNS	△	▽
Election results	○	×
Election district	V	
Candidate	A	B
Survey	△	▽
SNS	▽	△
Election results	×	○
Election district	VI	
Candidate	A	B
Survey	△	▽
SNS	▽	△
Election results	×	○

4. 결론

본 논문에서는 우리나라 20대 국회의원 선거 과정에서 후보자들이 SNS 이용자들의 관심을 많이 받을수록, SNS 이용자들에게 더 많이 언급될수록 선거에서 승리를 거머쥌 가능성이 크다는 것을 보여주었다. 선거구에 대해 트위터 상의 관심도에서 우위를 점하는 후보, 경쟁 후보의 이름을 검색했을 때 관심도에서 우위를 점하는 후보가 당선될 확률이 컸다. 또한, 이러한 과정에서 추출된 트윗 중 당선된 후보자의 데이터에 대한 긍정 부정 언급 빈도를 조사해본 결과 긍정적인 단어를 포함하는 언급이 비교적 많았음을 알 수 있었다. 그리고 여론조사에서 우세하거나 사회 통념상 후보자의 지역구에서 한 정당이 절대적으로 우세하다고 예상되더라도, SNS 상에서 관심을 더 받지 못한다면 선거에서 당선되지 못하는 결과를 보여주었다.

추후에는 데이터 분석 과정에서 단순히 관심도를 측정하는 것을 넘어, 여론 몰이나 여론 조작 여부를 판단하여 좀 더 세밀한 분석이 가능하도록 할 것이다. 더불어 분석 대상을 트위터 외에 블로그나 뉴스 기사의 댓글, 또 다른 SNS 채널의 의견 등으로 확장하여 보다 고도화된 분석이 가능하도록 확장할 것이다.

ACKNOWLEDGMENTS

This research was supported by the MISP(Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW (R7116-16-1018) supervised by the IITP(Institute for Information & communications Technology Promotion)

This work was supported by a research grant from Seoul Women's University(2016).

REFERENCES

- [1] <http://facebook>
- [2] <http://www.twitter.com>
- [3] <http://blog.naver.com>
- [4] <http://www.instagram.com>
- [5] Y. H. KIM, "Analysis of trend and behavior on SNS (Social Network Service)" KISDI STAT REPORT, Vol. 2, No. 9, pp. 151-170, 1999.
- [6] Sunil, W., David, S., Munir, M., and Christine, B. Willians, "Web2.0 and Politics : The 2008 U.S. Presidential Election and an E-Politics Research Agenda", MIS Quarterly, Vol. 16-07, pp. 669-688, 2016. 4. 15.
- [7] W. S. Song, J. H. Lee. "Predicting News Story Virality in Social Media." Conference of Korean Institute of Information Scientists and Engineers, pp. 474-476, 2014.12.
- [8] J. S. Min, "Study on Twitter users' political participation", KRCRA No.12, Vol 2, pp. 274-303, 2012. 6.
- [9] Y. W. No, D. Y. Kim, J. E. Han, J. T. Lim, K. S. Boek, J. S. Yoo, "Hot Topic Prediction Scheme Considering User Influences in Social Networks", Journal of Korea Contents Association Research, Vol.15, No. 8, pp24-36, 2015.8.
- [10] J. H. Na, Y. S. Kim, "Forecasting Unemployment Rate using Social Media Information", Journal of the Korea Industrial Information Systems Research, Vol. 18, No. 6, pp95-101, 2013.12
- [11] S. G. Yoon, S. H. Kim, K. S. Shin. "Development of the Accident Prediction Model for Enlisted Men through an Integrated Approach to Datamining and Textmining." Journal of Intelligence and Information Systems, Vol. 21.3 pp. 1-17, 2015.9.
- [12] J. B. Lee, C. K. Lee, K. J. Cha, "An Analysis of IT Trends Using Tweet Data", Journal of Intelligence and Information Systems, Vol. 21, No. 1, pp143-159, 2015.3
- [13] D. Y. Kim, H. E. Hong, "The Limited Meanings and Roles of the Reply as the Instrument of Political Participation", Cybercommunication Academic Society Vol. 32, No. 1, pp51-86, 2015.3
- [14] J. Y. Kim, C. K. Kang, "They Representative Enough? : Comparative Modes of Public Opinion Surveys", Peace Strides, Vol.22, No. 2, pp360-397, 2014.10
- [15] <http://www.realmeter.net/>

최 은 정(Choi, Eun Jung)



- 1997년 2월 : 서울여자대학교 컴퓨터학과(이학사)
- 2000년 2월 : 서울여자대학교 대학원 컴퓨터학(이학석사)
- 2005년 8월 : 서울여자대학교 대학원 컴퓨터학(이학박사)
- 2006년 3월 ~ 현재 : 서울여자대학교 정보보호학과 교수

- 관심분야 : 시스템보안, 암호, 빅데이터
- E-Mail : chej@swu.ac.kr

최 세 원(Choi, Se Won)



- 2013년 2월 : 서울여대 자율전공학부(인문사회계열) 입학
- 2013년 2월 ~ 현재 : 서울여자대학교 경영학과 재학
- 2013년 2월 ~ 현재 : 서울여자대학교 정보보호학과 재학(복수전공)
- 관심분야 : 빅데이터, MIS
- E-Mail : bisa1018@hanmail.net

이 시 연(Lee, Si Yeon)



- 2013년 2월 : 서울여대 자율전공학부(인문사회계열) 입학
- 2013년 2월 ~ 현재 : 서울여자대학교 언론영상학부 재학
- 2013년 2월 ~ 현재 : 서울여자대학교 정보보호학과 재학(복수전공)
- 관심분야 : 빅데이터, 네트워크 보안
- E-Mail : siyeon92557@gmail.com

김 명 주(Kim, Myuhng Joo)



- 1986년 2월 : 서울대학교 컴퓨터공학과(공학사)
- 1988년 2월 : 서울대학교 컴퓨터공학과 (공학석사)
- 1993년 8월 : 서울대학교 컴퓨터공학과 (공학박사)
- 1993년 9월 ~ 1995년 8월 : 서울대학교 컴퓨터신기술공동연구소 특별

연구원

- 2003년, 2010년 : 미국 펜실바니아대학교(UPenn) 객원 연구원
- 관심분야 : 정보보호, 인터넷윤리, 창의성교육
- E-Mail : mjkim@swu.ac.kr