한국컴퓨터정보학회 하계학술대회 논문집 제23권 제2호 (2015. 7)

소셜미디어상에서의 감성 전파 모델링 연구

우지영°, 최민석*, 이민정*
°고려대학교 정보보호대학원, *한국전자통신연구원, *세종사이버대학교 경영학과
e-mail : jywoo@korea.ac.kr°, cooldenny@etri.re.kr*, mjlee@sjcu.ac.kr*(교신저자)

An Epidemic Model for Sentiment Diffusion

Jiyoung Woo°, Minn Seok Choi*, Min Jung Lee*
°Korea University, Graduate School of Information Security,
*Electronics and Telecommunications Research Institute,
*Sejong Cyber University, Management Department

● Abstract ●

본 연구는 사용자의 감성이 온라인 소셜 미디어를 통해 감염이 된다는 사실을 감성 전파 모델링으로 보이고자 한다. 이를 위해 전염병 파생을 기술하고 이를 예측하는데 사용되었던 질병확산 모델을 기초로 소셜 미디어상의 감성 전파 모델을 제시한다. 제시한 모델의 타당성을 검증하기 위해 특정 리테일 산업에 대한 논의가 활발히 이루어지고 있는 웹포럼의 데이터를 수집한다. 수집된 데이터로부터 주요 주제어를 도출하고, 주제별 감성을 측정하고, 시간에 따른 감성 값을 도출하여, 제시한 모델을 추정한다. 실험 결과 사용자의 긍정적 감성과 부정적 감성이 서로 경쟁관계에 있다는 가정을 따른 제안한 모델이 타당함을 보였다.

키워드: 감성분석(sentiment analysis), 의료웹포럼(medical web forum), 질병확산 모델(epidemic model)

I. Introduction

온라인 커뮤니티는 정보를 생산하고, 공유하고, 서로간의 의견을 주고 받는 소통의 플랫폼이다. 과거 온라인 커뮤니티는 동일 주제에 대한 관심을 갖는 특정 이용자들로 구성되었지만, 소셜미디어의 등장으로 기업의 제품에서 더 나아가 사회 전반에 대한 이슈의 확대 재생산의 장으로 자리매김하고 있다. 특히, 관심 주제가 개인적 취향보다는 기업, 정치, 사회 일반의 성격을 띠면서, 사용자들의 의견은 집단 지성으로 나타나 여론 형성에까지 영향을 미치고 있다. 기업 입장에서는 이러한 사용자들의 집단 지성은 기업의 의사결정에 영향을 미치는 중요한 요소로 작용하고 있다. 소비자는 소셜 미디어 및 웹을 통해 제품과 기업에 관한 정보를 공유하고, 자신들의 의견을 토론하기 때문이다. 이에 따라 온라인상의 소비자 의견을 모니터링 하는 것은 기업에게 매우 중요한 일이 되었다. 과거처럼 몇 개의 미디어를 장악해서는 기업의 이미지를 관리할 수 없게 된 것이다. 웹에서 존재하는 소비자의 의견을 분석하고, 이의 감성 (Sentiment) 분석을 통해 소셜 미디어 상의 기업 이미지를 실시간으로 모니터링 하는 것은 기업의 중요 비즈니스로 자리 매김 될 것이다.

기업과 관련된 이벤트가 발생하였을 때, 좋은 뉴스를 어떻게 확산시킬 것이고, 나쁜 뉴스를 어떻게 효율적으로 관리할지에 대한 연구가 필요하다. 더 나아가 고객의 사회적/정치적 니즈를 파악하여, 기업과 관련된 잠재 니즈를 도출하고 이들의 동향을 예측하여, 기업의 이미지 전략 및 제품 개발 전략에 활용할 수 있다. 기업들이 온라인 커뮤니티를 전략적으로 활용하고자 한 예가 바이럴 마케팅(viral marketing)이다. 이는 기업의 제품 프로모션 정보가 웹상에서 빠르게 확산되도록 하는 것으로, 온라인 소셜 네트워크의 영향력 전파를 이용한 기본적인 전략이다. 일부 기업은 이를 악용해 특정 제품에 대한 좋은 리뷰나 광고성 글을 올려 돈을 지불하여 사용자들의 의견을 조작하려고도 한다. 하지만, 이런 기초 수준의 활용으로는 소셜 미디어의 빅 데이터가 가지는 잠재력을 제대로 활용하지 못하는 것이다. 이보다 좀 더 심층적인 분석을 통해 이를 전략적으로 활용할 수 있는 방안에 대한 연구가 필요하다.

본 연구는 기존 소셜 미디어에 대한 연구가 키워드 추출 및 이를 경영전략과 연결하여 분석하려는 시도에서 한발 더 나아가 최근 연구가 시작되고 있는 감성 분석을 주제로 한다. 기존의 감성과 관련된 연구를 살펴보면, 온라인상의 글로부터 사람의 감성을 유추하는 연구가 초기단계에 진행되었다. 다음 단계의 연구는 소셜 미디어의 감성이 제품의 판매 등의 기업의 성과 또는 주식 등의 금융시장에 미치는 영향을 분석하는 연구이다. 소셜 미디어로부터 얻어지는 빅데이터의 가치를 극대화하는 것은 온라인 사용자의 감성을 분석하고, 이의 중요성을 파악한 후의 다음 단계의 연구는 감성의 전파를 주제로 하는 연구이다. 이를 주제로 하는 연구는 아직 초기단계로 사용자들의 감성의 변화를 관찰하거나 감성의 변화와 이벤트를 연관 지어 관찰하는 것을 주요 주제로 하고 있다. 본 연구는 감성의 흐름을 관찰하는 것에서 더 나아가 감성이 형성되는 현상 및 전파되는 현상 안에

존재하는 메커니즘을 파악하고자하는 것이다. 이를 주제로 한 연구는 국내외를 통틀어 현재까지 없었다.

II. Preliminaries

웹과 소셜 미디어 분석에 대한 연구 동향을 살펴보면 과거에는 비정형 텍스트로부터 키워드를 추출하는 기술 위주의 연구가 주를 이루었고, 이를 기반 기술로 하여 트렌드를 파악하고 기업의 성과와 연관 짓는 연구가 주를 이루었다 [2]. 최근에는 소셜 미디어의 등장으로 웹상의 소셜 네트워크 분석이 이슈가 되면서, 네트워크 효과를 이용한 바이럴 마케팅에 대한 연구가 활발하게 진행되고 있다 [3, 4]. 소셜 네트워크 상에서의 영향력이 높은 유력자를 파악하는 연구 및 네트워크상에서 정보나 영향력이 확산되는 현상을 관찰하는 연구가 새로운 연구 주제로 떠오르고 있다. 하지만 기존의 연구는 키워드 추출 및 네트워크 분석에 한정되어 있고, 글의 감성을 분석하는 연구는 아직 초기 단계에 있다 [5].

국내에서도 최근 소셜미디어 상에서의 사용자 감성 분석에 대한 연구가 일부 소개되고 있다.

KAIST 차미영 교수팀이 트위터 데이터를 분석하여 사람들의 우울감을 측정한 연구를 소개하였다 [7]. 해당 연구는 우울감을 느끼는 사용자와 일반 사용자의 트위터 행태 분석에 그치고 있다.

최홍구와 황인준[8]은 트위터 문서 분석의 감성 분석을 통한 음악 추천 시스템을 개발하였다.

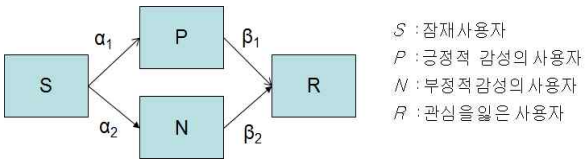
김영민 외 [9]의 연구에서는 소셜 미디어상의 데이터를 수집하여 주식시장 참여자들의 감성을 분석하고 이를 기반으로 주가 등락을 예측하는 절차를 제안하고자 소셜 미디어 감성분석을 통한 주가 등락 예측에 관한 연구를 수행하였다.

홍소라외 [9]의 연구에서는 대용량 소셜 미디어 감성분석을 위한 반감독 학습 기법을 제시하였다. 해당 연구는 "감성 레이블이 있는 데이터"와 함께 "감성 레이블이 없는 데이터"도 활용하기 위해서 반감독 학습기법인 self-training 알고리즘을 적용하여 감성분석 모델을 생성한다.

III. The Proposed Scheme

가설1. 긍정과 부정의 감성은 각각 독립적으로 타 사용자의 감성 형성에 영향을 준다.

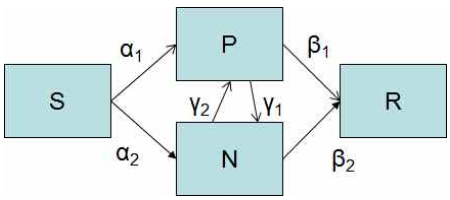
감정 전파 모델링을 위한 기초 모델은 긍정과 부정의 감성이 서로 경쟁하는 구도인 '경쟁 모델'이다. 특정 주제에 관심 있는 사람수는 정해져 있는 상황에서 일부는 긍정으로 일부는 부정의 감성으로 전염되며, 이때 두 감정이 전파되는 속도는 다르다고 가정한다. 처음 해당 주제에 관심이 있는 사용자 그룹을 S라고 하고, 긍정의 감성을 가지는 사용자를 P, 부정의 감성을 가지는 사용자 그룹을 N이라고 하면 아래의 그림과 같이 표현할 수 있다. R은 해당 주제에 대해 과거에는 관심을 가졌지만, 현재는 관심을 잃은 사용자를 말한다. α는 잠재 사용자에게 긍정과 부정의 감성이 영향을 주어 감성이 형성되는 비율을 나타내고, β는 해당 감성이 소멸되는 정도를 나타내는 비율이다.



Competitive Model for sentiment diffusion

가설 2. 긍정과 부정의 감성이 사용자간의 상호작용에 의해서 서로의 감성 변화에 영향을 준다.

가설 2에 따른 감성 전파 다이어그램은 그림 3과 같다. 긍정과 부정 감성간의 상호작용으로 인해 긍정적 감성의 일부가 부정적 감성으로, 반대로 부정적 감성의 일부가 긍정적 감성으로 전이된다. 가설 2에 따른 모델은 그림 1의 모델에 γ 부분을 추가하여, 일부 사용자의 감성이 변화하는 과정을 기술하게 한 모델이다. $\gamma 2$은 긍정과 부정의 상호작용으로 인해 부정에서 긍정으로 전이되는 정도, $\gamma 1$은 긍정에서 부정으로 전이되는 정도를 나타낸다.



Interaction Model for sentiment diffusion

IV. Experiments

제안하는 모델을 검증하기 위해 글로벌 금융 포럼을 선정하여 데이터 수집을 실시한다. 글로벌 금융 포럼 중 그 사용자 수가 많은 야휴 파이낸스(http://messages.finance.yahoo.com/)를 대상으로 하고, 그 중에서도 리테일 산업의 대표격인 Walmart 기업을 대상으로 한다.

웹 크롤러 개발해 데이터를 수집하고, 수집된 데이터로부터 필요한 필드를 추출하여 데이터베이스화하는 과정을 거친다.

글의 감성을 측정하기 위해서 SentiWordNet (http://sentiwordnet.isti.cnr.it/)을 이용한다. SentiWordNet은 워드넷 (http://wordnet.princeton.edu/)을 기반으로 각 단어를 객관성, 긍정적, 감성적 클래스로 구분한 어학 사전이다. 감성 값은 다음과 같이 계산된다.

사용한 데이터는 다음과 같다.

분석에 사용한 Yahoo-Finance 데이터의 통계는 다음과 같다.

Data statistics

Data source	Duration	# of messages	# of users
News	Aug. 1999 - Mar. 2007	4,081	657

모델의 파라미터 값을 측정하기 위해서 실제 값과 모델에서 나온 값의 차이를 최소화하는 최소자승법을 이용한다. 우선 P, N에 해당하는 사용자의 수를 측정하고, 나머지 클래스와 파라미터는 모델로 추정하도록 한다.

표2는 가장 중요한 주제어로 선정된 네 개의 주제어에 대한 감성 발생을 세가지 모델로 기술한 값이다. 이벤트의 센티멘트를 반영한 모델(Model #2)의 결과가 긍정과 부정 감성 사이의 상호작용을 고려한 모델(Model #1)의 결과보다 높은 것으로 나타났다. 이는 매스미디어의 감성을 고려하면, 소셜미디어 사용자의 감성을 더 잘 기술할 수 있음을 의미한다.

Parameter Estimation

	R^2 긍정 감성	R^2 부정 감성	$\alpha 1$	$\alpha 2$	$\beta 1$	$\beta 2$
healthcare	0.569		0.0017	0.0019	0.0811	0.1609
wage			0.001	0.0009	0.0353	0.0388
stock price			0.0031	0.0037	0.0474	0.0801
low price			0.0021	0.0021	0.0853	0.0865

V. Conclusions

본 연구는 사용자의 감성이 온라인 소셜 네트워크를 통해 감염이 된다는 사실을 밝히고, 여기서 한발 더 나아가 기존에 전염병 파생을 기술하고 이를 예측하는데 사용되었던 질병확산 모델이 소셜미디어상의 감성 전파에 활용 가능한지를 파악한다. 이는 기존 이론의 활용 범위를 확대하는 것이다. 또한 기존 모델을 적용하는 것에 그치는 것이 아니라 대립된 감성 사이의 상호작용을 포함한 감성전파를 기술할 수 있는 새로운 모델을 제시한다. 감성 전파 현상을 모델링할 수 있다면, 초기 상태의 감성 전파 현상으로부터 미래의 감성 형성을 예측할 수 있다. 또한 감성 전파 과정을 모델화하게 되면 이에 영향을 주는 요소를 파악할 수 있게 된다. 주요 요소 예를 들어, 서로 대립되는 감성 사이의 상호 작용이라던지, 사용자의 소셜 네트워크상의 사용자의 상호작용을 제외한 기타 외부 환경적인 요소의 영향도 파악할 수 있다. 이는 기존의 연구가 관찰을 통해 외부 사건의 발생이 소셜미디어상의 감성 변화에 미치는 영향을 파악하는 것에서 발전하여, 시스템적으로 영향 요소를 파악하고, 그 정도를 계량화할 수 있게 된다.

References

[1] J. Bollen, H. Mao, and X. J. Zeng, “Twitter mood predicts the stock market,” Journal of Computational Science, Vol. 2, No. 1, pp. 1-8, 2008.

[2] M. Bampo, M. T. Ewing, D. R. Mather, D. Stewart, and M. Wallace, “The effects of the social structure of digital networks on viral marketing performance,” Information Systems Research, Vol. 19, No. 3, pp. 273-290, 2008

[3] J. Leskovec, L. A. Admic, and B. A. Huberman, “The dynamics of viral marketing,” ACM Transactions on the Web, Vol. 1, No. 1, Article 5, 2007.

[4] B. Pang, and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends in Information Retrieval, Vol. 2, No. 1, pp. 1-135, 2008.

[5] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, “Emotions as infectious disease in a large social nework: The SIS model,” Proceedings of the Royal Society B. Vol. 277, pp. 3827-3835, 2010.

[6] M. Park, M. Cha, and C. Cha, “Depressive moods captured in Twitter, an online social network,” In Proc. of the 1st Congress for Qualitative Health Research, 2011.

[7] H. Choi and E. Hwang, “Emotion-based Music Recommendation System based on Twitter Document Analysis,” Journal of KIISE, Vol. 18, No. 11, pp. 762-767, 2012.

Y. M. Kim, S. J. Jeong and S. J. Lee, “A Study on the Stock Market Prediction Based on Sentiment Analysis of Social Media,” Entrue Journal of Information Technology, Vol. 13, No. 3, pp. 59-69, 2014.

S. Hong, Y. Chung, and J. Lee, “Semi-supervised learning for sentiment analysis in mass social media,” Journal of Korean Institute of Intelligent Systems, Vol. 24, No. 5, pp. 482-488, 2014.

[8] E. Katz, “The two-step flow of communication: An up-to-date report on a hypothesis,” The Public Opinion Quarterly, Vol. 21, No. 1, pp. 61-78, 1957.