

소셜네트워크서비스 추이를 활용한 수정된 감염 예측 모델 제안: COVID-19 한국 사례를 중심으로

배수진*, 황경화*, 조유정*, 권오병*

*경희대학교 경영학과

Modified Infection Prediction Model Using Social Network Service Trends: Focusing on Korean COVID-19 Pandemic Case

Sujin Bae, Kyunghwa Hwang, Yujung Cho, Ohbyung Kwon

Kyung Hee University

E-mail : luckybsj@khu.ac.kr, hoi7i@khu.ac.kr, yujung251@khu.ac.kr, obkwon@khu.ac.kr

요약

COVID-19는 비말을 통해 불특정 다수에게 쉽게 전염되는 높은 감염력으로 인해 유래 없는 팬데믹을 야기했다. 발생 초기 전파력과 치명율 정도를 알 수 없어 감염 예측의 어려움이 있었다. 특히 COVID-19는 대면 접촉을 줄이는 것으로 감염 예방에 나섰는데, 대면접촉이 어려운 상황에서는 SNS를 매개로 의사소통을 통한 방역에 도움이 되기도 한다. 그러나 가짜뉴스 등 정보가 정확하지 않을 경우, 더 많은 감염자를 발생하는 문제를 야기한다. 따라서 본 연구의 목적은 기존 감염 예측 모델인 SEIR 모델을 기준으로 SNS의 효과를 반영한 수정 SEIR 모델을 제안하는 것이다. 본 연구는 SNS 효과를 감염 모델의 정확도 증가에 영향을 미치는 것을 실증적으로 분석한 최초의 연구이다. 실무적으로는 방역 기관이 SNS 발생 양상을 분석해보다 신속하게 감염 추이를 파악하고 대처하는데 유용할 것이며, 대유행의 시점에 SNS가 주는 영향이 있음을 주지할 필요가 있다는 시사점이 있다.

1. 서론

COVID-19는 높은 감염력으로 인해 전 세계에 팬데믹으로 선포되었다. 한국에서도 1월 20일 첫 확진자가 발표된 이후로 정부 차원에서 방역 대책을 내놓으며 대응을 하고 있다. 기존 메르스, 사스와는 다르게 무증상 상태에서도 강력하게 전염되는 등 그 정도가 심각한 상황에서 전염을 막고 확진자의 수를 일정 수준으로 유지하는 것이 매우 중요한 일이다. 이를 위해서는 감염의 추이를 사전에 정확하게 예측하고 대응 전략을 구사해야 하는 바 PIM, SIR, SEIR 등 다양한 감염 예측 모형들이 제안되어 사용되고 있다. 그런데 이들 모형만으로는 현재 확진자의 추이를 정확하게 추정하기 어렵다.

특히 대면접촉이 어려운 상황에서 대부분 사람이 SNS를 매개로 하여 의사소통을 하는데 이를

통해 방역에 도움이 되기도 하지만 가짜뉴스로 인한 전염, 비난으로 인한 감염자의 자가 은폐를 유발해 방역에 부정적인 영향을 줄 수도 있다. 그러나 이러한 사회 의사소통의 양상을 감염 예측 모형에 반영한 예는 거의 없었다. 따라서 본 연구의 목적은 기존 감염 예측 모델인 SEIR 모델을 기준으로 SNS의 효과를 반영한 수정 SEIR 모델을 제안하는 것이다. 그리고 팬데믹 위험 관리에 있어서 SNS 정책이 미치는 영향을 기반으로 정책적 시사점을 제시하고자 한다.

2. 본론

2.1. SNS 이용동기 및 보건분야에서의 SNS

SNS를 사용하는 사람들은 이용동기가 있다고 설명하는데, 이는 SNS에 따라라도 연구자들마다도 요인의 구분이 조금씩 차이가 있다. 본 연구에서는

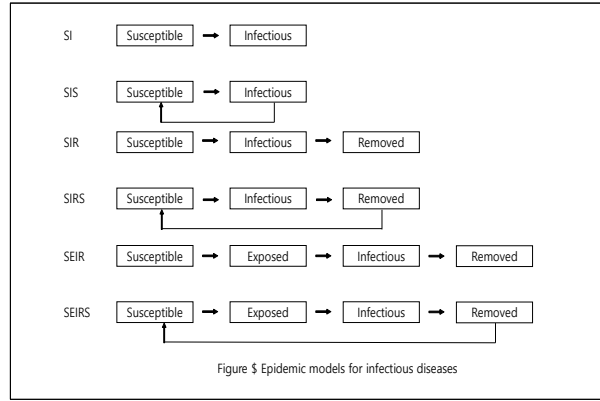
팬데믹 상황에서의 SNS 이용 동기는 크게 사회적 동기와 정보 동기로 구분한 연구[1]를 따라 알아보 고자 한다.

보건 분야에서는 SNS가 미치는 영향에 대한 연 구가 진행되어왔다. 일부는 SNS의 순기능 또는 역 기능으로 제시했다. 먼저 SNS는 지역주민, 재난 구호 단체, 정부가 재난에 대한 실시간 정보를 수 집하고 배포하는데 유용하며[2], 보건을 위한 홍보 캠페인이나 교육을 실시해 질병 재난 관리에 활용 하는 등[3] 재난 문제를 해결하는데 유용한 순기능 이 있다. 그런가 하면 보건적으로 문제가 되는 정 보를 공유하는 회합을 하거나[4] 가짜뉴스(예: 잘못된 약 사용법)를 유통하여 혼란을 초래하고 문제 를 더 악화시키는 등[5] 역기능의 경우도 있다.

2.2. COVID-19 팬데믹과 전염병 모형

코로나 바이러스(CoV)는 사람과 다양한 동물에 감염될 수 있는 바이러스로서 사람 간 전파는 비 말(호흡기 분비물) 전파 추정이 된다[6]. 전염력이 강해 잠복기에 있더라도 전염이 되는 것으로 알려 져서 본인이 미처 증상을 느끼기 전에 많은 사람 과 접촉하면서 전염이 되고 있고, 아직 마땅한 치 료제가 개발되지 않아 팬데믹으로 선포되었다. 전 염병의 대유행은 갑자기 발병해(epidemic) 높은 감 염율을 이야기하고, 대유행이 지난 이후에도 그 여 파로 인한 경제적 침체가 지속하는 등 사회적인 문제를 일으키므로 전염병의 역사와 함께 이를 규 명하고자 하는 질병의 수학적 모델이 발전해 왔다. 구획 모형(compartmental model) 유형의 전염병 모형으로는 일반적으로 SI, SIR, SIRS, SEIR 등이 있다(그림1 참고).

SEIR 모델은 잠복기를 고려한 SIR 모형보다 복 잡한 모형으로 SIR 모델 보다 현실적이다. SEIR 모형을 이용한 연구로는 메르스 데이터로 감염 기 간을 구축해 모형화했다[7]. 그리고 말라리아 질병 은 모기가 옮긴 전염병이라는 것을 알아내고 확산 모형을 제시함으로써 예방했다[8].



<그림 1> 전염병 예측 모형

2.3. 간접전파 감염경로 또는 비의약적 개입 방법(치료(방역)경로)으로서의 SNS

본 연구에서는 이용 동기에 따라 SNS를 정보 (제공) 동기와 사회 동기, 또한 순기능과 역기능으 로 2x2 분류를 제시했다(표1 참고). 그리고 이를 SEIR 모델에 반영한 수정 모델을 통해 한국의 확 진자 데이터와 비교했다. 여기에서 SNS의 빈도가 확진자 수에 영향을 주는 것은 정보동기 순기능은 확진자를 줄이는데, 그리고 정보동기 역기능은 확 진자를 증가시키는 데 영향을 준다고 가정했다. 또 한 사회동기 활동은 확진자 수의 추이에 영향을 받아 활동의 빈도가 영향을 받을 것이라 가정했다.

[표 1] 간접전파 감염경로로서의 SNS 네 가지 분류

동기	순기능	역기능
정보 동 기	정보제공, 홍보, 안내 등 [발생 시점] 감염 사건 발생 전후 [효과] 감염 정도를 저감함: 방역 및 치료 방법 검사자 수를 높임: 검사자 필요자에 대한 정보를 공유	가짜뉴스, 사기, 불안 감 조성 [발생 시점] 감염 사건 발생 전후 [효과] 감염 정도를 높임: 잘못된 방역 방법
사 회 동 기	의견제시, 캠페인, 공감과 격려, 집회 [발생 시점] 감염 사건 발생 후 [효과] 감염 정도를 저감함: 감염 원 접촉 감소 등의 사유	비난, 사기 [발생 시점] 감염 사건 발생 후 [효과] 감염 정도와는 무관함

3. 실험 방법 및 결과

3.1. 자료 수집

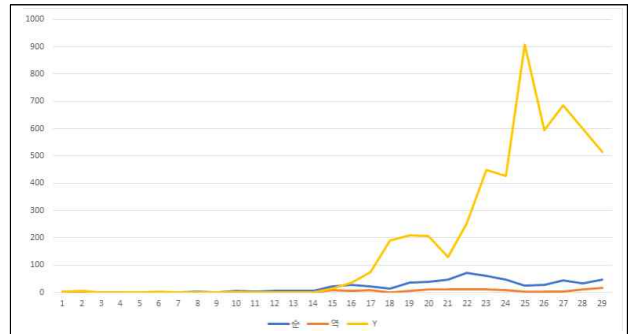
자료수집에 있어 코로나19 관련 통계는 질병관리본부(<http://ncov.mohw.go.kr>)에서 브리핑 자료로 배포한 보도 자료를 근거로 했다. 그리고 SNS 관련 자료수집은 뉴스 DB 검색 사이트인 빅카인즈(<https://www.bigkinds.or.kr/>)를 활용했다. SNS 데이터를 뉴스 기사 검색을 통해 한 이유는 언론에서 다루주는 SNS 활동 내용이 SNS 전체 빈도만을 반영한 것보다는 정제된 정보이기 때문이다. 즉, SNS의 모든 정보의 빈도를 분석 도구를 삼는 것이 아니라, 코로나19와 관련해 SNS에 활동한 내용을 뉴스기사에 반영된 것을 수집해 기사에서 언급한 기사의 내용이 SNS의 정보동기와 사회동기에 대한 분류하고, 그에 대한 순기능과 역기능을 분류하는 과정을 수행했다.

SNS 기능에 대한 분류를 위해 먼저 신문 발행 부수 순위와 정치적 성향을 토대로 조선일보, 중앙일보, 경향신문, 한겨레신문 총 4개 언론사를 대상으로 자료를 수집했다. 2월 14일 이전에는 뉴스 기사가 없으니 1월 20일 ~ 2월 4일 데이터는 필요 없었다. 키워드 검색으로 ‘코로나19 SNS 플랫폼명’을 검색했다. 예를 들어 ‘코로나19 유튜브’, ‘코로나19 페이스북’ 등으로 SNS 플랫폼명을 반영해 자료를 수집했다. 그리고 SNS를 통해 활동하는 단어인 ‘댓글, 단톡’ 등 SNS를 통해 활동하는 형태에 대한 키워드도 코로나19와 함께 반영해서 수집했다. 수집 기간은 2020년 1월~ 3월 31일로 수집했으며, 총 2,095건 중에, 실제 SNS 기능과 관련된 내용이 포함된 기사만을 반영해 총 2,067건을 분석 도구로 삼았다. 자료 수집된 SNS 플랫폼의 결과는 표2와 같다.

[표 2] 언론사 및 SNS 플랫폼 빈도

SNS 구분	언론사별 빈도 (n=2067)			
	경향신문 (n=423)	조선일보 (n=507)	중앙일보 (n=893)	한겨레 (n=245)
YouTube	104	115	219	71
Facebook	171	201	343	98
Instagram	20	26	81	14
Twitter	86	121	141	41
Weibo, Wechat	1	5	9	1
Kakao	3	8	25	1
댓글	38	30	75	19
온라인카페	0	1	0	0

수집된 자료들의 추이를 살펴보면 그림2와 같이 SNS의 순기능과 역기능이 확진자 수의 빈도 추이와 관련성을 살펴볼 수 있다.



<그림 2> SNS 순기능 및 역기능과 확진자 수 추이

3.2. 실험 진행

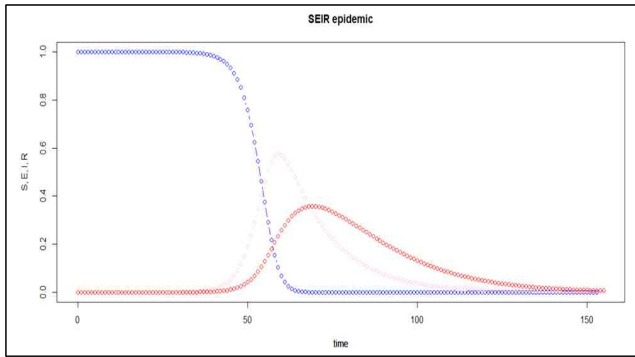
본 연구를 진행하기 위해서는 먼저, COVID-19 상황에 적합한 감염 잠복기를 고려한 예측 모델인 SEIR 모델을 한국 상황에 적합한 최적의 예측 결과를 도출한다. 또한, 앞서 수집한 뉴스 기사에서 언급된 SNS의 기능 두 가지 정보동기와 사회동기의 순기능 및 역기능으로 분류한 기사들의 일별 빈도 데이터 세트를 활용해 SNS의 영향 값을 도출하고자 했다. 즉, SEIR 모델을 통해 나온 결과에서 확진자 수 예측인 감염자(Infected)에 대한 값을 기준으로 SNS의 영향 값이 반영된 수정된 SEIR 모델의 Infectious 데이터를 기반으로 실제 확진자 수에 가장 가깝게 예측하는지 알아보하고자 했다.

그리고 SNS 영향 값은 발생 빈도에 따라 SNS에서도 잠복기를 거쳐 확진자 수에 영향을 미치는 데, 이때 SNS 빈도에 따라 분포를 이룬다고 할 수 있다. 따라서 정규분포와 포아송분포를 활용해 SNS의 영향 값을 찾고자 몬테카를로 방법론[9]을 통해 수행했다. 한편 사회동기 순기능 및 역기능을 종속변수로 확진자 수가 독립변수로 한 회귀분석을 통해 유의하게 영향을 미치는지 확인했다.

3.3. 실험 결과

먼저 SEIR모델을 R 코드를 활용해 한국 상황에 맞는 결과를 그림3과 같이 확인했다. 분석 결과 SEIR 모델 보다 확진자 추이를 더욱 정확하게 예측할 수 있었다. 정보동기의 SNS는 확진자 추이의

선행 효과로, 사회 동기는 확진자 추이의 선행 효과가 있는 것을 관찰할 수 있었다.



〈그림 3〉 한국 사례 SEIR모델 결과

4. 결론

본 연구는 SNS 효과를 감염 모델에 적용하고 실증적으로 분석한 최초의 연구이다. 실무적으로는 방역 기관이 SNS 발생 양상을 분석해보다 신속하게 감염 추이를 파악하고 대처하는데 유용할 것이며, 대유행의 시점에 SNS가 주는 영향이 있음을 주지할 필요가 있다는 시사점이 있다.

현재 한국 사례를 대상으로 정보제공 유형의 SNS 발생에 대한 시민 인지의 정도가 COVID-19 감염에 미치는 영향 정도를 최적 추정해 예측한 수치로 SEIR 모델에 의한 예측을 보정하는 연구들이 진행 중이다. 한편, 감염자 및 확진자 수에 대한 인지가 사회동기 유형의 SNS 발생에 미치는 영향에 대해서도 확인 중이다. 이 두 가지 연구가 진행되고 나면 COVID-19 팬데믹과 같은 환경적 요인 발생의 경우에 SNS 관련 정책을 어떻게 수립해야 하는지에 대한 가이드라인을 제공할 수 있을 것이다.

본 연구가 빅데이터와 보건 분야의 융합 연구로서 SNS 팬데믹을 빠르게 극복하는 도구가 되기를 기대한다.

Acknowledgement

이 논문 또는 저서는 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A3A2066740)

[참고문헌]

- [1] Johnson, P. R., & Yang, S. (2009, August). Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use. In *Communication Technology Division of the annual convention of the Association for Education in Journalism and Mass Communication in Boston, MA*.
- [2] Huang, C. M., Chan, E., & Hyder, A. A. (2010). Web 2.0 and Internet social networking: a new tool for disaster management?—Lessons from Taiwan. *BMC Medical Informatics and Decision Making*, 10 (1), 57.
- [3] Norman, C. D., & Yip, A. L. (2012). eHealth promotion and social innovation with youth: using social and visual media to engage diverse communities. *Studies in Health Technology and Informatics*, 172, 54-70.
- [4] Young, S. D., & Rice, E. (2011). Online social networking technologies, HIV knowledge, and sexual risk and testing behaviors among homeless youth. *AIDS and Behavior*, 15 (2), 253-260.
- [5] Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, 38 (3), 182-188.
- [6] Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity*, 102433.
- [7] Assiri, A., Al-Tawfiq, J. A., Al-Rabeeah, A. A., Al-Rabiah, F. A., Al-Hajjar, S., Al-Barrak, A., ... & Makhdoom, H. Q. (2013). Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *The Lancet Infectious Diseases*, 13 (9), 752-761.
- [8] ROSS, R. (1911). Some Quantitative Studies in Epidemiology. *Nature* 87, 466 - 467.
- [9] Mooney, C. Z. (1997). *Monte Carlo Simulation* (Vol. 116). Sage Publications.