

COVID-19 확산 예측 모형에 관한 연구

윤석용
명지대학교 ICT 융합대학 빅데이터융합교육
icanibe@mju.ac.kr

A Study on the Diffusion Prediction Model of COVID-19

Seok-Yong Yun
Big Data Convergence Education, Myongji University

요 약

COVID-19(Coronavirus Disease 2019)는 RNA 형 바이러스로써 점막감염(粘膜感染)과 비말전파(飛沫傳播)로 전염되는 급성 호흡기성 질병이다. 2019년 12월 중국 후베이 우한에서 처음 감염이 보고된 후 빠르게 글로벌로 확산되었고, 현재 여러 국가와 지역이 Lockdown 상태에 있다.

COVID-19의 치사율은 국가별, 연령별 차이는 있으나 사스(SARS-CoV), 메르스(MERS-CoV) 등과 비교하여 높다고 할 수 없다. 그러나 COVID-19는 신종 코로나바이러스로써 아직 백신(Vaccine)과 항바이러스제가 개발되지 않았고 다른 질병과 비교하여 빠른 감염 속도때문에 의료 공백, 사회적 혼란, 경제적 손실을 크게 일으키고 있다.

따라서 바이러스의 확산 양상을 데이터 분석을 통하여 예측할 수 있다면 사회·경제적인 폐해를 줄일 수 있어 Bass 모델과 R 패키지를 이용하여 COVID-19 확산 예측 모형을 계량적으로 제시하였다.

1. 서론

코로나바이러스는 Group IV 군 니도바이러스목 코로나바이러스과에 속하는 RNA 형 바이러스로써 광륜의 모양을 하고 있어 코로나로 불리는 급성 호흡기성 질병이다. 그러나 COVID-19(SARS-CoV2)는 코로나바이러스의 변종으로 2019년 12월 감염이 처음 보고된 이후 전 세계 모든 국가로 빠르게 전파되어 의료 공백 사태를 일으키고 있고 아직도 증가세는 멈추지 않고 있다.

COVID-19의 이러한 현상으로 인하여 세계보건기구 WHO는 전염병 경보 최종 6단계인 팬데믹(Pandemic)을 3월 11일 선언하였다. 2009년 214개국에 감염되어 1만 8천여 명의 사망자를 가져온 신종 플루 이후 11년 만에 선언된 세계 수준의 전염병이다.

지구온난화 등의 이유로 팬데믹 수준의 전염병은 앞으로도 발생 가능성이 높고 이를 예방하기 위한 백신과 치료제인 항바이러스제 개발은 늦을 수밖에 없다.

팬데믹이 발생하면 의료 공백과 국가 및 지역 단위의 Lockdown으로 이어져 사회·경제적인 폐해는 예측할 수 없을 정도로 커진다. 따라서 이를 최소화하고 예측을 통한 사회적 혼란을 줄이기 위하여 데이터

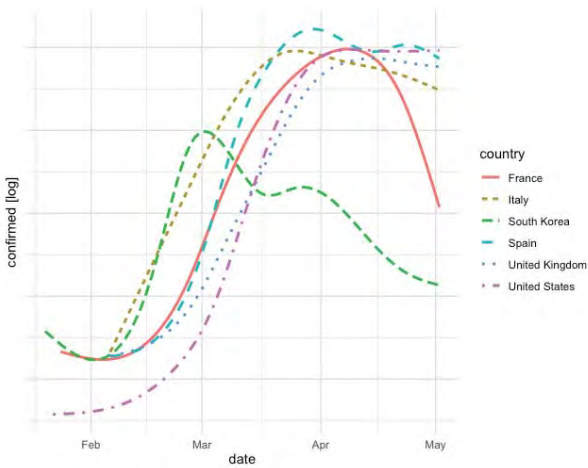
기반의 질병 확산 모형 개발이 무엇보다 중요하다.

본 연구는 COVID-19의 국가별 일자별 코로나바이러스의 확진자 등의 데이터를 이용하여 다양한 기술 분석(Descriptive Analysis)과 중장기 예측에 많이 사용되는 Bass 모델을 이용하여 COVID-19 확산 예측 모형을 제시하고자 한다.

2. COVID-19 데이터 분석과 연구 모형

COVID-19 데이터는 확진자, 완치자, 사망자 그리고 감염경로 등으로 구성되어 있다. 데이터 분석은 COVID-19 데이터에 더하여 검사자 수와 국가별 인구 수를 반영함으로써 비교 데이터의 객관성을 유지하였다.

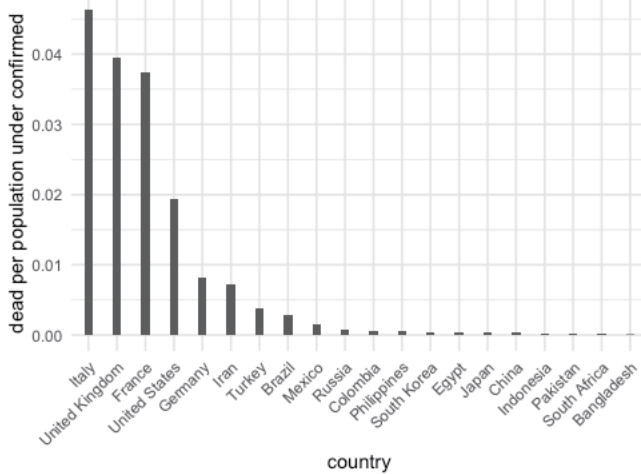
(그림 1)은 인구대비 일일 확진자 상위 5개국과 한국을 로그 스케일(Log Scale)로 비교한 그래프로써 한국 등 몇몇 국가를 제외하고는 일일 확진자 수가 줄지 않고 있다.



(그림 1) 인구대비 확진률

코로나바이러스 사망률은 다른 전염병에 비하여 높지 않으나 빠른 감염¹으로 인한 입원 환자 수의 증가와 단기간 사망자 수의 급증으로 의료 공백이 발생하고 있다.

(그림 2)는 국가별 인구대비 감염자의 사망률로써 인구 5 천만 명 이상의 국가를 대상으로 데이터를 분석하면 이탈리아, 영국, 프랑스, 미국, 독일 순으로 사망률이 높아 현재 의료에 어려움이 발생하고 있음을 알 수 있다.



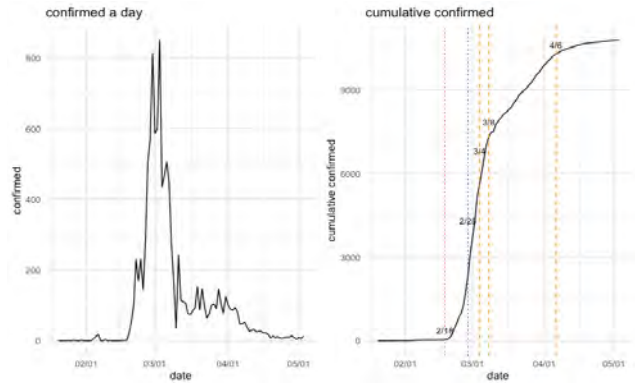
(그림 2) 인구대비 사망률

COVID-19 는 점막감염(粘膜感染)과 비말전파(飛沫傳播)로 전염되는 급성 호흡기성 질병으로 알려져 있기에 마스크 착용은 감염 속도를 낮추는 효과적인 수단이다.

(그림 3)은 2 월 18 일 31 번 확진자 발생으로 감염자가 급증하여 2 월 28 일 공적마스크를 공급함으로써

일일 확진자 발생 반감기가 약 1~2 주 단위로 나타나고 있음을 보여준다.

비교 데이터가 없어 반감기에 대한 통계적 유의성은 없으나 정책적 의사결정에는 참조가 가능할 수 있다.



(그림 3) 한국의 일자별 확진자

확산모형으로 마케팅을 포함한 여러 분야에서 활용되고 있는 것은 로지스틱과 Bass 모형 등을 들 수 있다.

로지스틱 모형은 1845 년 소개된 후 확산뿐만 아니라 S 형 성장모형에도[1] 널리 사용된다.

$$\hat{y} = \frac{m \cdot e^{(a+b \cdot t)}}{1 + e^{(a+b \cdot t)}}$$

$$t_{max_rate} = \frac{\ln 2 - a}{b}$$

$$t_{max_num} = -\frac{a}{b}$$

윗 식에서 a, b, m 은 OLS(Ordinary Least Squares)나 NLS(Nonlinear Least Squares)로 추정할 수 있는 모수이고 t_{max_rate} 와 t_{max_num} 은 증가율이 최대가 되는 시점과 예측값이 최대가 되는 시점이다.

Bass 모형도 Rogers[2]의 혁신이론에 기반한 확산 모형으로써 Bass[3]에 의해 수리적으로 증명된 후 수요 확산 예측에 많이 활용되고 있다.

$$\hat{y} = m \cdot f(t)$$

$$\frac{f(t)}{1 - F(t)} = p + q \cdot F(t)$$

¹ 사스의 기초감염재생산지수(R_0)는 2~5, 메르스는 0.4~0.9 로 COVID-19 는 사스 이상의 R_0 예상

$$\hat{y} = \frac{m \left[\frac{(p+q)^2}{p} \right] \cdot e^{-(p+q)t}}{\left[1 + \frac{q}{p} e^{-(p+q)t} \right]^2}$$

이 Bass 모형의 m, p, q 는 비선형 계획법을 이용하여 찾을 수 있다.

3. COVID-19 확산 예측 모형 결과

확산 예측 모형은 R 언어의 NLS 패키지를 이용하여 모수를 추정하였고, 데이터는 일일 확진자 수를 대상으로 데이터 최종 확진일 이후 120 일 동안을 일일 단위로 예측하였다. 그리고 예측 모형의 안정화를 위하여 일자별 확진자 수의 IQR 1.5 배 이상은 이상치(Outlier)로 처리하였다.

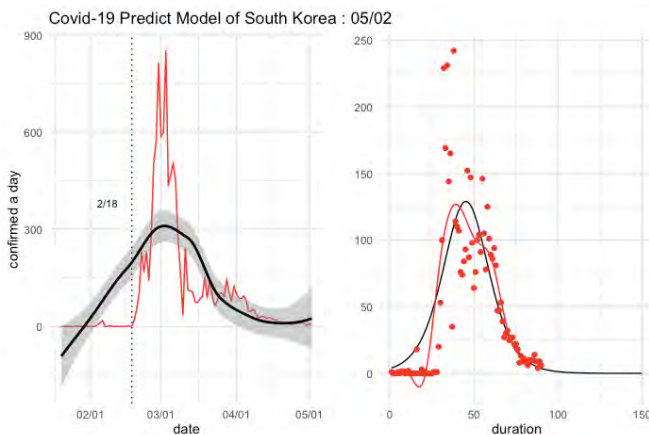
모형의 메트릭인 RMSE²는 Bass 354.76, 로지스틱 525.96 로 비교 우위가 있는 Bass 모형을 선택하였다.

<표 1>은 Bass 모형의 파라미터 계수와 통계량이다.

<표 1> Bass 모형 파라미터 통계량

	Estimate	Std. Error	t value	Pr(> t)
m	4678.81288	347.38501	13.4687	0.00000***
p	0.00077	0.00030	2.5455	0.01267**
q	0.10875	0.01062	10.2365	0.00000***

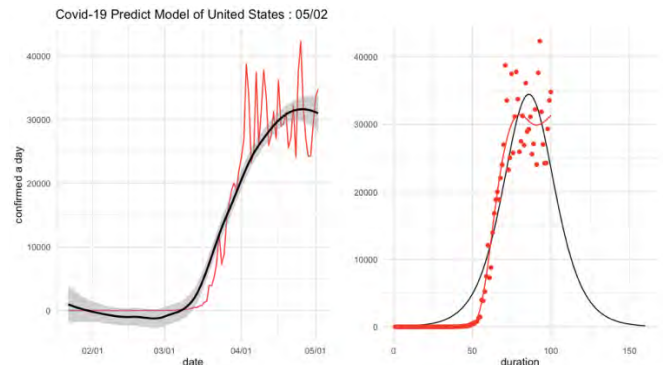
(그림 4)의 좌측 그래프는 한국의 일일 확진자 수와 흑색의 추세선을 보여주고 있고, 우측 그래프는 흑색의 COVID-19 확산 예측 모형과 일자별 확진자 수와 추세선이다.



(그림 4) COVID-19 예측 모형(한국)

예측 모형에서 한국은 5 월 초순에 확진자 영(Zero)

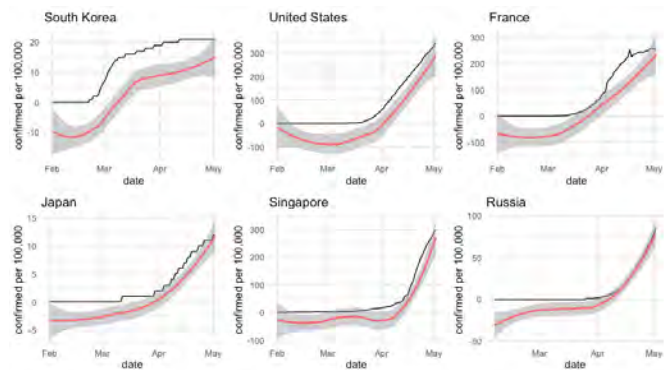
에 수렴하는 것으로 예측되었고 (그림 5)는 미국의 예측 모형으로 8 월하순에 확진자 발생이 영(Zero)에 수렴하는 것으로 예측하고 있다.



(그림 5) COVID-19 예측 모형(미국)

4. 연구결과 및 향후 연구과제

COVID-19 는 논문이 작성되고 있는 시점에서 현 재 진행형으로 (그림 6)은 10 만 명당 누적 확진자 수와 일일 확진자 수의 추세선으로 확산세가 아직 꺾이지 않고 있음을 보여주고 있다.



(그림 6) 인구 10 만 명당 확진자

팬데믹으로 한국을 포함한 많은 국가가 마이너스 성장을 예측하고 있어 Lockdown 조기 종료를 위한 바이러스 확산 예측 모형은 그 의미가 크다고 판단된다.

그러나 데이터에 기반한 예측 모형은 검진이나 확진 등과 관련한 데이터의 신뢰성이 무엇보다 중요하고 Bass 에 기반한 일반화된 단일 모형보다는 국가별 특성을 반영하여 세분화하고 로지스틱과 신경망 모델 등이 결합된 앙상블 모형으로의 추가적인 연구가 필요하다.

참고문헌

[1] Kingsland, S., “The Refractory Model : The Logistic Curve and the History of Population Ecology,” The

² Root Mean Square Error

- Quarterly Review of Biology, Vol.57, No.1(1982), pp.29-52.
- [2] Rogers, E.M., Diffusion of innovation, New York, 1962.
- [3] Bass, F.M., "A New Product Growth Model for Consumer Durables," Management Science, Vol.15, No.5(1969), pp.215-227.
- [4] Bass, F.M., "Comments on a new product growth for model consumer durables the bass model," Management Science, Vol.50 (2004), pp.1833-1840.
- [5] 홍정식, 김태구, 구훈영, "NLS 와 OLS 의 하이브리드 방법에 의한 Bass 확산모형의 모수추정", 대한산업공학회지, 제 37 권 제 1 호(2011), pp.74-82.
- [6] 양진아, 민대기, 최형석, "Bass 확산모형을 활용한 국내 주택연금의 중장기 수요예측", 한국경영과학회지, 제 42 권 제 1 회(2017), pp.29-41.
- [7] 이하늘, 김대회, 강지석, 이동환, 김윤배, "Bass 모형을 통한 WIPI 정책의 영향 분석:스마트 폰 시장을 중심으로", 정보통신정책연구, 제 22 권 제 4 호(2015), pp.1-18.
- [8] 홍정식, 김태구, 임달오, "확산 모형에 의한 고가 의료기기의 수요 확산의 특성분석 및 중장기 수요예측에 관한 연구", 보건행정학회지, 제 18 권 제 4 호(2008), pp.85-110.