

SARIMA모형을 이용한 코로나19 확진자수 예측

김재호¹ · 김장영^{2*}

Prediction of Covid-19 confirmed number of cases using SARIMA model

Jae-Ho Kim¹ · Jang-Young Kim^{2*}

¹Graduate Student, Department of Computer Science, The University of Suwon, Hwaseong, 18323 Korea

^{2*}Associate Professor, Department of Computer Science, The University of Suwon, Hwaseong, 18323 Korea

요약

코로나19의 일일 확진자 수는 천명 후반대에서 2천명대를 유지하고 있으며, 백신접종률이 증가함에도 불구하고 확진자수가 쉽게 줄어들지 않는 상황이다. 변이바이러스는 계속해서 등장하고, 현재는 무 변이 바이러스까지 국내에 유입되었다. 본 논문은 코로나 예방전략을 위해 SARIMA 모델을 통해 코로나19 국내 확진자 수를 예측한다. ADF Test와 KPSS Test를 통해 데이터에 추세와 계절성이 있음을 확인한다. SARIMA(p,d,q)(P,D,Q,S)의 p, d, q, P, D, Q의 값은 모형 차수결정 정리로 파라미터를 추출한다. ACF와 PACF를 통해 p, q 파라미터를 추론한다. 차분, 로그변환, 계절성제거 등을 통해 데이터를 정상성 형태로 변환하고, 도식화 하여 파라미터를 도출하고, 계절성이 있다면 S를 정하고, SARIMA P,D,Q를 정하고, 계절성을 제외한 차수에 대해 ACF와 PACF를 보고 ARIMA p,d,q를 정한다.

ABSTRACT

The daily number of confirmed cases of Coronavirus disease 2019(COVID-19) ranges between 1,000 and 2,000. Despite higher vaccination rates, the number of confirmed cases continues to increase. The Mu variant of COVID-19 reported in some countries by WHO has been identified in Korea. In this study, we predicted the number of confirmed COVID-19 cases in Korea using the SARIMA for the Covid-19 prevention strategy. Trends and seasonality were observed in the data, and the ADF Test and KPSS Test was used accordingly. Order determination of the SARIMA(p,d,q)(P, D, Q, S) model helped in extracting the values of p, d, q, P, D, and Q parameters. After deducing the p and q parameters using ACF and PACF, the data were transformed and schematized into stationary forms through difference, log transformation, and seasonality removal. If seasonality appears, first determine S, then SARIMA P, D, Q, and finally determine ARIMA p, d, q using ACF and PACF for the order excluding seasonality.

키워드 : 계절성 자기회귀누적이동평균, 코로나19, ADF Test, KPSS Test

Keywords : Seasonal auto regressive integrated moving average(SARIMA), Covid-19, Augmented dickey-fuller(ADF) test, Kwiatkowski-phillips-schmidt-shin(KPSS) test.

Received 4 November 2021, Revised 11 November 2021, Accepted 20 December 2021

* Corresponding Author Jang-Young Kim(E-mail: jykim77@suwon.ac.kr, Tel:+82-31-229-8345)

Associate Professor, Department of Computer Science, The University of Suwon, Hwasung, 18323 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.1.58>

print ISSN: 2234-4772 online ISSN: 2288-4165

© 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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

코로나19 바이러스 일일 감염자 수는 계속해서 증가하고 있고, 변이 바이러스 발생으로 인해 사태가 점점 심각해지고, 해외유입 감염으로 인해 우리나라에도 변이 바이러스가 퍼지고 있다. 변이 바이러스에 따라서 치사율과 전염성은 모두 다르고, 신종 변이 코로나 바이러스 같은 경우에는 정보가 상대적으로 적기 때문에 더 조심해야 한다. 코로나 변이 바이러스는 계속 새로운 종류가 나타나고 있다. 현재 WHO(세계보건기구)는 코로나 변이를 증상, 전파력, 백신 효과 등을 고려하여, 우려 변이와 관심 변이로 분류한다. WHO(세계보건기구)는 2021년 1월 콜롬비아에서 보고된 B.1.621 변이 바이러스를 '뮤(Mu)' 변이 바이러스로 부르고, 관심 변이로 분류했다. 뮤 변이는 WHO(세계보건기구)가 관심을 갖고 지켜보는 다섯 번째 변이이다. 뮤 변이 바이러스의 면역 정도, 전파력 등을 확인하기 위해서는 추가적 연구가 필요하다는 입장이다[1].

지역 사회에서 코로나19 전파는 발생하는 감염의 양과 상관관계가 있다. 코로나19 감염은 지역 사회 비율이 높을 때, 잠재적으로 전파 가능성이 증가한다. 현재까지 예방 전략을 이행하고, 구현될 때 지역 사회 내 전파가 제한될 수 있다. 예방 전략으로는 백신 접종 장려, 사회적 거리두기, 환기, 손 씻기, 아플 경우 격리 및 검사받기, 접촉자 추적과 격리 및 예방격리, 소독이 있다. 이러한 전략을 구현하는 것은 실질적 또는 높은 전파율을 예방하고 코로나19로부터 사람들을 보호한다[2].

현재 승인된 코로나19 백신은 무증상 및 증상 감염, 중증환자 및 사망을 포함한 코로나19 감염에 대해 효과적인 것으로 나타났다. 예방 접종을 받은 사람들은 전염병 사슬을 방해함으로써 질병의 부담을 실질적으로 줄일 수 있는 잠재력을 가지고 있다. 하지만 예방 접종된 사람들의 코로나 감염의 위험은 바이러스의 지역 사회 감염으로부터 완전한 예방을 하는 것은 아니다. 여전히 감염되어 바이러스를 다른 사람들에게 전파할 수 있다[3].

현재 코로나 확진자 수는 계속해서 증가하고 있다. 코로나19 확진자 수를 예측 한다면, 방역에 더욱 신경쓸 수 있고 예방전략을 세울 수 있다. 본 논문에서는 SARIMA(Seasonal Auto Regressive Integrated Moving Average)모형을 이용해서 코로나19 누적 확진자수를 예측하고, 기존 ARIMA(Autoregressive Integrated Moving

Average)모형을 이용해 예측한 것과 비교한다. 또한 ACF Test (Augmented Dickey-Fuller Test)와 KPSS Test (Kwiatkowski-Phillips-Schmidt-Shin Test)를 통해 SARIMA 모형을 사용하는 것이 높은 예측값을 가진다는 것을 설명한다.

II. 기존연구

2.1. 시계열 데이터를 활용한 코로나19 동향 예측

국내 코로나19의 감염자 수가 백신과 사회적 거리 두기, 백신 등 여러 가지 노력 덕분에 차츰 줄어드는 듯 보였으나 2020년 2월 20일 특정한 사건 이후 감염자 수가 증가한 것처럼, 2020년 12월부터 또다시 급격히 감염자 수가 증가하는 추세이며 꾸준히 일일 500명가량의 감염자 수가 이어지고 있다. 본 논문은 Kaggle의 데이터셋을 이용해서 Prophet 알고리즘을 통해 미래 코로나19를 예측하고 사이킷런을 통해 결정계수, 평균 절대 오차, 평균 백분율 오차, 평균 제곱 차, 평균 제곱근 편차를 통해 이 예측에 대한 설명력을 더한다. 또한 코로나19 감염자 수를 특정한 사건이 없었을 경우 국내 감염자 수를 예측해 앞으로 우리가 미래의 질병에 대해서 방역과 방역 수칙 실천의 중요함을 강조한다[4].

III. 분석 알고리즘

ARIMA(Autoregressive Integrated Moving Average) 모형을 통해서 예측을 했을 경우, ADF Test(Augmented Dickey-Fuller Test)에서는 p-value 0.85로 강한 추세성이 있는 비정상 데이터로 나타났으며, 실제로도 꾸준히 코로나19 확진자 수가 증가하고 있다. 또한 KPSS Test (Kwiatkowski-Phillips-Schmidt-Shin Test)에서 p-value가 0.01로 계절성이 있다는 Test 결과가 나왔다. ARIMA 모델은 추세차분만 고려하지만, SARIMA 모델은 계절성 차분까지 고려한다. 즉, SARIMA는 Seasonal+ARIMA 모델이다. 따라서 본 논문은 SARIMA 모델을 이용해 코로나 19 감염자 수를 예측한다.

3.1. ADF Test (Augmented Dickey-Fuller Test)

Augmented Dickey Fuller Test는 간단한 Dickey Fuller

테스트의 확장 버전이다. Dickey Fuller같은 경우 오차항에 자기 상관관계가 존재하지 않는다는 가정을 하지만, Augmented Dickey Fuller Test는 자기 상관관계가 있을 경우를 고려한다. 오차항 때문에 백색잡음일 가능성이 낮고, 자기 상관관계의 문제를 제거하기 위해 종속 변수 측면에서 추가 지연항을 포함하여 테스트를 확장한다. 간단히 말해 기존 모델에 종속변수의 시차 값을 포함하고 자기상관계가 제거될 때까지 이 절차를 계속한다.

$$\Delta Y_t = a_0 + a_1 t + \pi Y_{t-1} + \sum_{i=1}^p \gamma_i \Delta Y_{t-i} + \epsilon_t \quad (1)$$

$$H_0 : \gamma = 0$$

위 식에서 a_0 는 상수항, a_1 은 시차 변수의 개수, t 는 선형추세 변동을 나타내는 시차변수, γ 은 자기회귀모수 추정계수, ϵ 은 오차항을 의미한다. 이때 귀무가설 H_0 이 $\gamma=0$ 이면 단위근이 존재한다. 따라서 해당 시계열이 비정상이다. ADF Test검정을 시행하기 위해서는 잔차들이 자기 상관관계가 가지지 않도록 시차의 길이를 선택해야 한다. 시차를 정하는 데에는 AIC값 또는 BIC값을 최소로 하는 값을 선택하거나, 시차가 통계적으로 유의할 때까지 시차를 변경한다[5].

3.2. KPSS Test (Kwiatkowski-Phillips-Schmidt-Shin)

추세 정상성, 수준 정상성 및 0인 평균 정상성의 귀무가설에 대해 이러한 검정을 사용할 수 있다. 점근적 null 분포를 도입하고 관련 비정상 대안에 대한 일관성을 결정하고, 테스트의 속성을 정상성에 대해 제안된 다른 테스트의 속성과 비교한다[5].

$$S_t = \sum_{i=1}^t \xi_i, \eta = \frac{\sum_{t=1}^T S_t^2}{\hat{\sigma}_t^2}, \hat{\eta}_\mu = T^{-2} \frac{\sum_{t=1}^T S_t^2}{\hat{\sigma}_t^2(q)} \quad (2)$$

KPSS 검정은 u_t 가 임의확률(r_t)과 정상오차(ϵ_t)의 합으로 가정한다. 여기서 u_t 을 절편에 회귀시킨 잔차를 ξ_t 라 한다. $\hat{\sigma}^2$ 은 모집단의 분산을 의미한다.

3.3. 계절성 자기회귀누적이동평균 (SARIMA, Seasonal Auto Regressive Integrated Moving Average)

ARIMA(p,d,q)은 1이상의 차분이 적용된 $\Delta d Y_t = (1-L)dY_t$ 가 알고리즘의 차수(p and q)가 유한한 AR(p)와

MA(q)의 선형조합이다. 비정상 시계열 데이터 Y_t 를 차분한 결과로 만들어진 $\Delta Y_t = Y_t - Y_{t-1} = (1-L)Y_t$ 가 정상성인 데이터이고 ARMA 모델을 따르면 원래의 Y_t 를 ARIMA 모형이라고 한다. ARIMA는 추세 차분만 고려한다. 계절성을 고려하기 위해서는 SARIMA모형을 사용해야 한다. ARIMA모형의 업그레이드버전이 SARIMA모형이라고 할 수 있다.

SARIMA(p,d,q)(P,D,Q,S)의 p, d, q, P, D, Q의 값은 모형 차수결정 정리로 추출한다. 추정 및 예측을 하기 전에 파라미터에 따라 모형이 어떠한 결과를 도출할지 이해가 필요하고 결과이해는 기계의 실수를 방지하고 결과의 확신을 증가시킨다. ACF와 PACF를 통해 p, q 파라미터 추론은 정상성 형태 변환, 즉 차분/로그변환/계절성제거 등을 통해 데이터를 정상성 형태로 변환하고 ACF, PACF를 도식화 하여 ARIMA의 파라미터 차수를 추론한다. 차분이나 로그를 이용해 ACF와 PACF를 정상성 데이터로 전처리하고, 계절성이 있다면 Seasonality S를 정하고, SARIMA P,D,Q를 정하고, Seasonality를 제외한 lag(시점의 차이)에 대해 ACF와 PACF를 보고 ARIMA p,d,q를 정한다. 여기서 시차인 S는 직접 정해야 한다. 차분한 함수에서 lag인자를 추가 해주어서 계절적 차분을 수행한다. 데이터를 차분한 모델의 ACF, PACF에서 lag가 7마다 미세하게 반복되는 패턴을 보이기 때문에 S는 7로 정했다. 다음은 ARIMA (Auto Regressive Integrated Moving Average)와 SARIMA (Seasonal Auto Regressive Integrated Moving Average)의 수식이다.

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1-L)^d Y_t = (1 + \theta_1 L + \dots + \theta_q L^q) \epsilon_t \quad (3)$$

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - \phi_1 L^m - \dots - \phi_{Pm} L^{Pm})(1-L)^d (1-L^m)^D Y_t = (1 + \theta_1 L + \dots + \theta_q L^q)(1 + \Theta_1 L^m + \dots + \Theta_Q L^{mQ}) \epsilon_t \quad (4)$$

여기서 ϕ 은 자기회귀 모형의 모수, θ 은 이동평균 모형의 모수, ϵ 은 백색잡음을 의미한다. ARIMA에서는 AR(p), 추세 차분과 MA(q)만 존재했지만, SARIMA에서는 AR(P), 계절성 차분, MA(Q)데이터가 추가되었다.

IV. 실험결과 및 분석

데이터는 Kaggle에서 John Hopkins 대학 COVID-19 데이터셋에서 Raw_global 데이터를 사용한다.[6]

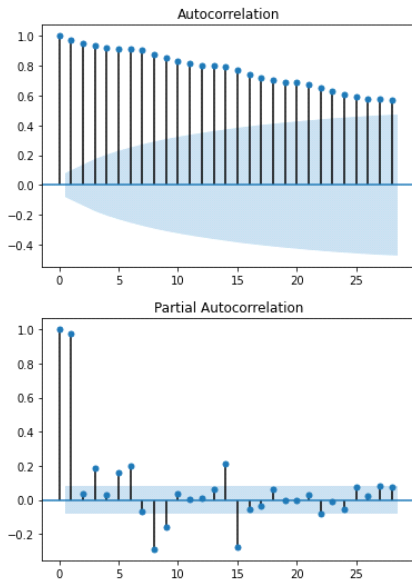


Fig. 1 ACF, PACF of diff data

그림 1은 데이터를 차분한 데이터의 ACF와 PACF이다. 데이터를 차분함으로써 원래의 데이터보다 ACF값이 좋아지긴 했지만, 누적 확진자수는 계속해서 증가하는 추세이기 때문에 천천히 감소하는 값을 가진다. AR(Auto-Regression)적인 특성을 가지지만 lag(시점의 시차)가 1이후에도 유의수준에서 벗어나는 값들을 가지는 것을 확인할 수 있다. 이것은 여러 가지 경우가 있지만 sclae(데이터 크기)이 크기 때문에 발생하는 경우와 overfitting(과적합)이 발생하는 경우가 있다.

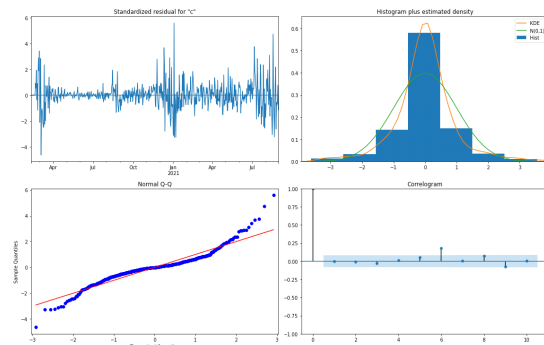


Fig. 2 Residual, Normal Distrubstion, Q-Q plot, Correlogram

그림2는 모델의 잔차, 정규분포, Q-Q plot, 상관도표를 보여준다. Kline(2005)은 왜도의 기준을 절대값 3을 초과하지 않고, 첨도는 절대값 8 또는 10을 초과하지 않으면 정규분포로 볼 수 있다고 제시한다. 이 모델은 Skew(왜도)는 0.37, Kurtosis(첨도)는 7.89 이므로 정규분포 형태라고 볼 수 있다[7].

p-value에서는 ma.L2에서만 유의하지 않다는 결과가 나왔고, Jarque-Bera Test(정규분포 테스트)에서 p-value 값이 0 이므로 데이터는 정규분포를 가지고 있다. 실제로 히스토그램에서 SARIA를 적용한 이 시계열 데이터는 정규분포를 가지고 있다. Ljung-box Test(자기상관 테스트)에서 p-value 값은 0.98 이므로 자기상관을 가지고 있지 않다. 따라서 독립성을 가지고 있다고 볼 수 있다. Homoscedasticity Test(등분산성 테스트)에서 p-value 값이 0이므로 등분산성 데이터이다. q-q plot은 정규성을 보이는 듯 하지만, 양쪽 끝을 보면 큰 noise를 보인다. 이것은 n차 감염으로 급격히 늘어난 감염자 수가 있다는 것을 보여준다고 할 수 있다. 시계열 데이터의 성질을 분석하는데 중요하게 활용되고 있는 것이 상관도표(Correlogram)이다. Correlogram은 ACF(Autocorrelation Function)와 PACF(Partial Autocorrelation Function) 중 하나를 그래프로 표현한 것이다. Correlogram을 활용하면 시점의 자료와 다른 시점의 차이(lag)를 가지는 자료를 비교하여 어떤 관계를 가지는지 있는지 분석할 수 있다. 따라서 Correlogram을 통해서 시점의 차이와 영향력을 알 수 있다.

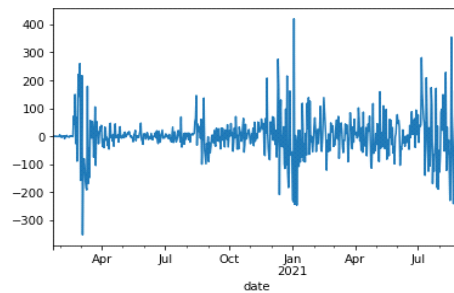


Fig. 3 Residual

그림3은 실제 데이터 값과 예측 데이터의 잔차(Residual)값이다. 갑자기 큰 폭으로 진동하는 경우가 나타나는데 이것은 예상치 못한 대유행이 발생시기와 겹친다. 2020년 2~3월 경에는 신천지 종교 집단감염 사태로 1차 대유행의 시발점이 되었다. 2020년 8월에는

광화문 시위 횡수가 증가했고, 2차 대유행이라 부른다. 2020년 11월부터 대학가 중심으로 코로나가 확산되며 2021년 2월 설날 즈음까지 계속 이어지며 이것을 3차 유행 이라고 부른다. 2021년 7월부터 또다시 급격히 증가 하는데, 직장, 학교, 백화점 등 일상 공간 곳곳에서 집단 감염이 속출하고 전파력이 더 센 인도에서 나타난 ‘델타’ 변이 바이러스까지 확산되고 있어 유행 규모가 커졌다. 4차 유행이 대유행까지 퍼지게 되었다. 급증 시기와 잔차가 급증하는 시점이 들어맞는다[8].

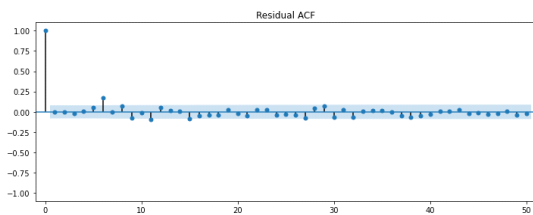


Fig. 4 Residual ACF

위 그림4는 잔차분석 값의 자기상관함수(ACF)이다. 대부분의 값들이 화이트 노이즈임을 보인다.

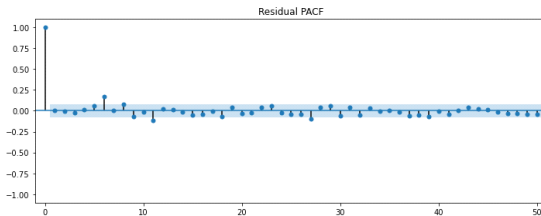


Fig. 5 Residual PACF

위 그림5는 잔차분석 값의 편자기상관함수(PACF)이다. 조금씩 유의수준을 벗어나는 값들이 있으나 대부분의 값들이 화이트 노이즈이며, 자기상관이 없는 시계열 데이터라고 말할 수 있다. 화이트 노이즈이기 때문에 정상성을 가진 데이터 이다.

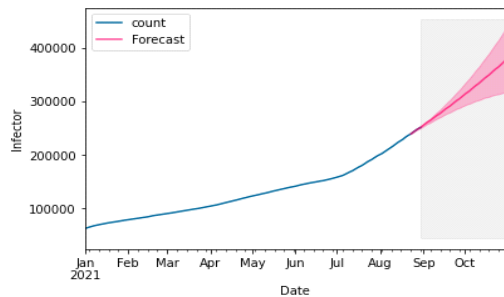


Fig. 6 South Korea, Prediction data

위 그림6은 코로나19 감염자 수의 신뢰수준과 예측 데이터이다. 9월 4일 데이터가 누락되어 있으므로 데이터를 drop하고 진행했다.

Table. 1 Pratical infectee & Prediction infectee

Date	Count	SARIMA-Pred	ARIMA-Pred
2021-08-29	250,051	250,062	250,312
2021-08-30	251,421	251,507	251,708
2021-08-31	253,445	253,417	252,983
2021-09-01	255,401	255,231	255,148
2021-09-02	257,110	257,083	257,184
2021-09-03	258,913	258,874	258,868
2021-09-05	261,778	261,963	260,119
2021-09-06	263,374	263,380	263,496

위 표1은 ARIMA와 SARIMA의 예측을 서로 비교한다. ARIMA보다 SARIMA가 더 정확한 예측을 한다. 한계는 시계열 알고리즘 특성상 Auto-Regression이기 때문에 전날 데이터의 영향 받아, 데이터가 날마다 쌓이면서 Training Set과 Test Set이 바뀌기 때문에 매일매일 예측값이 바뀐다는 것이다. Test set과 Training Set을 고정해서 예측할 수도 있지만, 전날에 예측한 데이터가 다음날 한 예측을 가장 잘 맞추는 결과를 가진다. 하지만 예상치 못한 집단감염으로 인한 급격한 변화가 있는 사례의 경우에는 일시적으로 예측력이 급격히 줄어든다. 집단감염 이후에 데이터가 쌓이면 다시 좋은 예측값을 가진다. 또한, 기존 논문인 ‘시계열 데이터를 활용한 코로나19 동향 예측’ 논문에서 사용한 prophet 알고리즘보다 MAE(Mean Absolute Error, 평균 절대 오차)측면에서 좋은 결과를 가져왔다. 9일간 MAE비교결과 prophet 모델의 결과 59, SARIMA 모델의 경우 7이라는 값을 가졌다.

V. 결론 및 향후연구

위 그래프들에서 알 수 있듯이 현재 코로나 감염자수는 날이 지날수록 최고점을 기록한다. 예측에서도 코로나 감염자 수는 나날이 증가할 것이라고 예측한다. 백신 접종자가 증가하면 감염자 수가 줄어들어야 하는데, 감염자 수는 계속해서 늘어날 것으로 예측된다. 변이바이러스와 집단감염이 우려되기 때문에, 백신접종의 무무와 별

개로 방역에 힘을 써야한다.

본 논문에서는 SARIMA모형을 통해 적분 선형확률 과정을 통해서 코로나 감염자 수 데이터를 예측을 했다. 기존 ARIMA모형에서와 달리 계절성을 고려했기 때문에 더 정확한 예측값을 가질 수 있었다. 한계점으로는 집단감염 등 급격한 변화예측에 약하며, 매일매일 데이터가 쌓이기 때문에 Training Set과 Test Set이 매일 달라지기 때문에 예측값이 매일매일 달라진다.

향후, 변이 바이러스 감염자 수에 대한 데이터 셋을 이용해 추가적인 변이 바이러스에 대한 감염자 수를 예측할 수 있다. 또한 비선형 확률과정인 딥러닝 모델중 BI-LSTM과 GRU모형을 통해 코로나 감염자 수를 예측하고 각 모델의 결과값을 비교한다. 질병 방면에서 어떠한 예측이 더 효과적인지 분석하고자 한다. 향후 코로나 19의 예방전략에 도움이 될 것으로 기대된다.

REFERENCES

- [1] K. Laiton-Donato, C. Franco-Muñoz, D. A. Álvarez-Díaz, H. A. Ruiz-Moreno, J. A. Usme-Ciro, D. A. Prada, J. Reales-González, S. Corchuelo, M. T. Herrera-Sepúlveda, J. Naizaque, G. Santamaría, J. Rivera, P. Rojas, J. H. Ortiz, A. Cardona, D. Malo, F. Prieto-Alvarado, F. R. Gómez, M. Wiesner, M. L. O. Martínez, and M. Mercado-Reyes, "Characterization of the emerging B.1.621 variant of interest of SARS- CoV-2," *Infection, Genetics and Evolution*, vol. 95, pp. 105038, Nov. 2021.
- [2] National Center for Immunization and Respiratory Diseases (NCIRD). Science Brief: Transmission of SARS-CoV-2 in K-12 Schools and Early Care and Education Programs [Internet]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK570438/>.
- [3] CDC Public Health Science Agenda for COVID-19 [Internet]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html>.
- [4] J. H. Kim and J. Y. Kim, "Covid19 trends predictions using time series data," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 7, pp. 884-889, Dec. 2021.
- [5] J. S. Yoo and J. G. Choo, "A Study on the Test and Visualization of Change in Trends associated with the Occurrence of Non-stationary of Long-term Time Series Data based on Unit Root Test," *KIPS Transactions on Software and Data Engineering*, vol. 8, no. 7, pp. 289-302, Jul. 2019.
- [6] Kaggle data set, covid19-data-from-john-hopkins-university [Internet]. Available: <https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university>.
- [7] T. J. Kline, "Psychological testing: A practical approach to design and evaluation," *Sage Publications*, 2005.
- [8] The progress of the COVID-19 pandemic in Korea [Internet]. Available: <https://www.joongang.co.kr/article/25010254#home>.



김재호(Jae-Ho Kim)

수원대학교 컴퓨터학부 학사
수원대학교 컴퓨터학부 석사과정
※관심분야 : 인공지능



김장영(Jang-Young Kim)

연세대학교 컴퓨터과학 공학사
Pennsylvania State Univ. 공학석사
State University of New York 공학박사
University of South Carolina 교수
수원대학교 컴퓨터학부 교수
※관심분야 : Big data, AI, Cloud computing, Networks