

ARIMA모형을 이용한 코로나19 확진자수 예측

김재호¹ · 김장영^{2*}

Prediction of Covid-19 confirmed number of cases using ARIMA 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

요 약

2019년 12월경 후베이 우한시에서 발생한 코로나19 바이러스가 점차 줄어드는 듯 보였으나, 2020년 11월, 2021년 6월 기준으로 점차 늘어나고 있으며, 전세계적으로 총 1억 9천 2백만명, 대한민국 기준 총 확진자는 대략 18만4천 명으로 추정된다. 이에 따른 대책으로 중앙재난안전대책본부는 사회적 거리두기 4단계를 시행하면서 강력한 대응책을 내고있지만, 델타바이러스등 전염성이 강한 코로나 변이 바이러스가 기승을 부리면서 국내 일일 확진자 수는 1800명대 까지 증가하게 되었다. 그에따라 코로나바이러스의 심각성을 강조하고자 코로나 누적 확진자 수를 ARIMA 알고리즘을 이용해 예측한다. 그 과정에서 추세와 계절성을 제거하기 위해서 차분을 이용하고, MA, AR, 자기상관함수와 편자기상관함수를 이용해 ARIMA에서 p,d,q값을 결정하고 예측한다. 마지막으로 예측값과 실제값을 비교해 얼마나 잘 예측되었는지 평가한다.

ABSTRACT

Although the COVID-19 outbreak that occurred in Wuhan, Hubei around December 2019, seemed to be gradually decreasing, it was gradually increasing as of November 2020 and June 2021, and estimated confirmed cases were 192 million worldwide and approximately 184 thousand in South Korea. The Central Disaster and Safety Countermeasures Headquarters have been taking strong countermeasures by implementing level 4 social distancing. However, as the highly infectious COVID-19 variants, such as Delta mutation, have been on the rise, the number of daily confirmed cases in Korea has increased to 1,800. Therefore, the number of cumulative confirmed COVID-19 cases is predicted using ARIMA algorithms to emphasize the severity of COVID-19. In the process, differences are used to remove trends and seasonality, and p, d, and q values are determined and forecasted in ARIMA using MA, AR, autocorrelation functions, and partial autocorrelation functions. Finally, forecast and actual values are compared to evaluate how well it was forecasted.

키워드 : 코로나19, 이동평균, 자기회귀모델, 자기회귀이동평균, 자기회귀누적이동평균

Keywords : Covid-19, Moving average(MA), Auto-regressive(AR), Auto regressive moving average(ARMA), Auto regressive integrated moving average(ARIMA)

Received 4 November 2021, Revised 11 November 2021, Accepted 27 November 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.2021.25.12.1756>

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. 서 론

1.1. 개발동기

2019년 12월 12일에 처음 발생해, 2020년 1월 20일에 처음 국내로 유입된 코로나19는 백신개발과 국민들의 노력으로 인해서 조금씩 나아지는 듯 하였으나, 코로나 3차 유행이후 500~600명대 확진자와 방역완화, 델타변이 바이러스의 확산, 자가검사 키트로 인한 ‘가짜 음성자’의 전파 가능성등 ‘조용한 전파’가 원인이 되어 2021년 7월 14일 기준 1600명의 감염자를 돌파하며 상황은 견잡을 수 없이 되고있다. 중앙재난안전대책본부는 항상 조기적으로 방역에 성공할 것이라고 예측하고 있지만, 델타바이러스의 발생과 엄청난 전파력, 백신접종에도 불구하고 코로나 바이러스의 감염, 백신확보 실패, 예상치 못한 집단감염과 해외유입으로 인해 감염자수가 계속 늘어나고 있다. 따라서 중앙재난안전대책본부는 2021년 7월 12일부터 수도권은 사회적 거리두기 4단계, 비수도권은 사회적 거리두기 개편안 1단계, 부산과 대전은 2단계를 2021년 7월 25일까지 2주간 적용하면서 코로나19에 대한 강력한 대응책을 내고있다[1].

1.2. 델타변이바이러스

코로나 바이러스 변이는 현재까지 알파, 베타, 감마, 델타, 람다 총 5가지로 변이되어 확산되고 있다. 알파는 영국, 베타는 남미지역, 델타는 인도, 감마는 일본과 브라질, 람다는 페루에서 발생한 변이 바이러스다. 현재 전세계 감염의 약60%를 차지하고, 가장 전염성이 강하다고 알려진 것이 델타변이 바이러스다. 전염성은 일반 코로나 바이러스보다 2.8배 강력하다. 기존 코로나 바이러스의 같은 경우에는 고온, 기침, 미각상실, 후각 상실 등의 증상이 주류를 이었지만, 델타변이 증상으로는 두통, 인후염, 콧물 등으로 독감과 유사한 형태를 띠고 있다. 델타 변이 바이러스의 경우 백신의 효과는 현저히 감소되지만, 치사율은 현저히 낮춘다[2]. 현재 우리나라는 알파 변이가 우세종이지만, 전 세계적으로는 델타 변이가 우세종이다. 우리나라도 2021년 6월 27일부터 7월 3일까지 주요 변이 바이러스를 보면, 델타 변이가 크게 늘고 있다. 델타바이러스는 강력한 전염성과 폐세포 수용체에 대해서 강력하게 결합하며 항체 반응을 감소시킨다.

1.3. 백신 현황

코로나19의 대유행을 종식시키려면 전 세계의 많은 사람들이 바이러스에 면역이 되어야 한다. 이를 달성하는 가장 안전한 방법은 백신을 사용하는 것이다. 코로나 19의 대유행이 시작된지 1년도 채 되지않아 여러 연구팀이 백신개발을 개발했다. 과제는 이 백신을 전 세계 사람들이 사용할 수 있도록 하는 것이다. 부유한 국가뿐만 아니라 모든 국가의 사람들이 백신으로 인해 코로나에 보호받는 것이 중요하다. 현재까지 전 세계적 인구의 약26.1%가 코로나 백신 접종을 한 것으로 확인되었다. 세계적으로 약36억 1천만회 투여되고 있으며, 매일 약3046만회가 투여되고 있다. 우리나라의 경우 완전히 예방접종을 받은 사람들의 비율은 12.34%, 부분적으로만 예방접종을 받은 사람들의 비율은 18.79%이다[3]. 안정성이나 다방면의 문제가 있겠지만, 현재까지 우리나라의 경우 백신 확보를 우선적으로 하지않고, 백신 접종을 하지 못한 채 선부르게 방역 완화를 거론한 것이 최근이 되어 코로나 4차 유행이 진행되었다.

1.4. 사회적 거리두기 4단계

중앙재난안전대책본부는 2021년 7월 14일부터 코로나19의 확산을 막기위해서 수도권을 중심으로 사회적 거리두기 4단계를 적용하면서 상황에 따라 연장이 될 가능성이 크다. 몇몇 상황을 제외하고 18시 이전에는 4인까지, 이후에는 2명까지 사적모임이 가능하며, 스포츠 관람의 경우 무관중, 종교활동의 경우에도 비대면 정교활동을 진행해야하며 행사, 식사, 숙박등을 금지하고 학교 수업의 경우에는 원격수업으로 전환한다. 직장근무 같은 경우에도 시차 출퇴근제, 점심시간 시차제, 재택근무 30%권고하고 있다[4].

II. 기존연구

2.1. ARIMA 모델을 이용한 데이터 흐름 예측 기법

기존 데이터의 패턴 예측에는 통계를 기반으로 한 수학적 모델이 주로 사용되었으나 새로운 데이터에 대한 피드백이 부족하기 때문에 장기간의 데이터 예측에 한계가 있다. 또한 데이터의 특성이 다양하고 복잡한 경우에는 수학적 모델의 결합 및 계산과정 이 어려워진다. 따라서 본 논문에서는 데이터의 학습 및 예측에 기존 정

적 모델이 아닌 기계학습 중 시계열 데이터 분석 (Time Series Analysis) 을 기반으로 연구를 진행하였다. 기계 학습은 복잡한 특성을 가진 데이터를 학습하여 미래의 데이터 값을 예측 하거나 분류하는데 있어서 정확도 및 처리시간 측면에서의 성능을 향상시킬 수 있다[5].

III. 분석 알고리즘

최근 딥러닝, 인공지능 등 트렌드와 함께 시계열 데이터 분석에도 인공지능망이 널리 사용되고 있다. 특히 시계열분석에는 RNN(Recurrent Neural Networks), LSTM (Long Short Term Memory), GRU(Gated Recurrent Unit)등을 이용할 수 있지만, 가장 큰 단점은 설명력이 부족해 제대로 된 모형해석이 거의 안된다는 점이다. 알고리즘적인 측면에서야 성능이 잘나오면 되지만 데이터과학, 특히 의사결정 지원 시스템에서는 예측한 근거가 필요하다. 이러한 관점에서 AR(AutoCorrelation Function), MA(Moving Average) 등과 같은 통계기반의 시계열분석법인 ARIMA(Auto Regressive Integrated Moving Average)모형을 사용한다. ARIMA모형은 시계열 데이터를 이용해 과거의 데이터를 기반으로 한 행동을 기반으로 한다. ARMA(Auto Regressive Moving Average)모형을 일반화한 것으로 분기별, 반기별, 주간, 월간, 연간, 단위로 단기적 예측에 많이 사용된다. ARIMA모형은 비시계열 데이터에도 적용이 가능하다. 본 논문은 차분을 통해 정상성 확보를 통해 안정성이 높아지고 예측력이 향상시킨다. 또한 적은 모수로 모델링이 가능하기 때문에 과적합을 줄인다. 본 논문은 차분을 통해 정상성을 확보하고 자기상관함수와 편자기상관함수로 ARIMA에서 p,d,q를 정하고 데이터를 ARIMA모형에 적용시킨다[6].

3.1. 차분(Difference)

특정 시점 또는 시점들의 데이터가 발산할 경우 시점 간 차분(변화량)으로 정상성을 변환한다. 정상성 확보를 위해서 차분을 이용해 계절성(Seasonality)과 추세(Trend)를 제거한다.

$$(1 - L^d) Y_t \tag{1}$$

$$\begin{aligned} \text{Main Equation of } d = 1 \\ Y_t = (1 - L^1) Y_t = (1 - Lag^1)_t Y_t \\ = Y_t - Lag^1(Y_t) = Y_t - Y_{t-1} \end{aligned}$$

3.2. 자기상관함수(ACF, AutoCorrelation Function)

관찰과 관찰 사이의 시차 함수로서의 관찰 간의 유사도이다. 즉 시차에 따른 일련의 자기상관을 뜻한다. 자기 상관 분석은 노이즈에 의해 가려진 주기적인 신호의 반복 패턴을 찾는다. 시간영역, 신호와 같은 기능 또는 일련의 값을 분석하기 위해 많이 쓰인다. 시차가 커질수록 ACF(AutoCorrelation Function)는 0에 가까워진다. 정상 시계열 데이터는 상대적으로 빠르게 0에 수렴하고, 비정상 시계열 데이터는 천천히 감소하며 종종 큰 양의 값을 가진다[7].

3.3. 편자기상관함수(PACF, Partial AutoCorrelation Function)

시계열 분석에서 편자기상관 함수(PACF)는 고정 시계열의 부분 상관을 자체 시차 값과 함께 제공하고 모든 더 짧은 시차에서 시계열 값을 회귀한다. 시차에 따른 일련의 편자기상관이며, 시차가 다른 두 시계열 데이터 간의 순수 상호 연관성이다. 어떤 시점과 또다른 시점 사이의 포함된 다른 값들의 영향은 무시한다. 다른 시차를 제어하지 않는 자기상관 함수와 대조된다. 부분 자기상관 함수를 잘못하면 AR(p) 모델 또는 확장된 ARIMA (p,d, q) 모델에서 p를 결정할 수 있다[7].

3.4. 이동평균(MA, Moving Average)

알고리즘 차수(q)가 가우시안 백색잡음 과정의 선형 조합이다. 지수 평활법 내에있는 이동평균 평활법은 과거의 Trend-Cycle추정, MA는 미래 값을 예측한다.

$$MA(q) : \tag{2}$$

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

where $\epsilon_i \sim i.i.d. WN(0, \sigma_\epsilon^2)$ and $\sum_{i=1}^{\infty} \theta_i^2 < \infty$

$$\therefore MA(q) : \tag{3}$$

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

여기서 μ 는 계열의 평균이고 θ 는 모델의 매개변수, ϵ 는 백색잡음을, q의 값은 MA모델의 다항식 차수를 의미한다.

3.5. 자기회귀모델(AR, Auto-Regressive)

알고리즘의 차수(p)가 유한한 자기자신의 과거값들의 선형조합이다. ACF모델은 시차(Lag)가 증가해도 0이 되지 않는다. 자기회귀 모델은 출력 변수가 과거 값과 불안정하게 예측한 항에 선형적으로 의존함을 지정한다. 따라서 모델은 확률적 차이 방정식의 형태다. MA 모델과 함께 시계열의 보다 ARMA와 ARIMA모델의 특수 경우이자 핵심 구성요소다.

$$AR(p) : Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (4)$$

where $\epsilon_t \sim i.i.d. WN(0, \sigma_{\epsilon}^2)$ and $\sum_{i=1}^{\infty} \phi_i^2 < \infty$

$$\therefore AR(p) : Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t \approx MA(\infty) \quad (5)$$

여기서 ϕ 은 매개변수, ϵ 는 백색잡음, c 는 상수, p 의 값은 AR모델의 다항식 차수를 의미한다.

3.6. 자기회귀이동평균(ARMA, Auto Regressive Moving Average)

알고리즘의 차수(p~q)가 AR(p)와 MA(q)의 선형조합이며, AR과 MA두가지 다항식이다. 약한 정상성을 가진 확률적 시계열을 표현한다.

$$Y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (6)$$

where $\epsilon_t \sim i.i.d. WN(0, \sigma_{\epsilon}^2)$

and $\sum_{i=1}^{\infty} \phi_i^2 < \infty, \sum_{i=1}^{\infty} \theta_i^2 < \infty$

여기서 ϕ 은 자기회귀 모형의 모수, θ 은 이동평균 모형의 모수, ϵ 은 백색잡음을 의미한다.

3.7. 자기회귀누적이동평균(ARIMA, Auto Regressive Integrated Moving Average)

1이상의 차분이 적용된 ΔY 는 알고리즘 차수가 유일한 AR(p)와 MA(q)의 선형조합이다. ARMA와 ARIMA는 자기회귀 모형인 AR(p)와 이동평균 모형인 MA(q)를 포함하는 것은 같다. AR(p)모형은 종속변수의 과거 값을 이용해 예측하고, MA(q)모형은 시계열의 평균과 과거의 오차를 이용해 예측 한다. ARMA와 ARIMA의 차

이는 차분이다. ARMA모형은 정상성을 가진 모형이다. 모형에 정상성이 없다면, 시계열 데이터를 차분함으로써 정상 시계열로 만들 수 있다. ARIMA에서 'I'는 누적(integrated)을 뜻하며, 비정상시계열 데이터를 정상성을 가진 시계열 데이터로 만들기 위한 필요한 차분의 횟수를 뜻한다. 차분이 필요없는 정상성 모형이면 ARIMA모형은 ARMA모형이 된다. ARMA(p,q)모형에 d차 차분을하면 차수가 (p,d,q)인 ARIMA과정이다. 자기상관계수가 천천히 감소하는것이 ARIMA와 같은 적분과정의 특징이다[8].

$$\Delta^d Y_t = Y_t - Y_{t-1} = (1-L)^d Y_t \quad (7)$$

$$\Delta^d Y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (8)$$

where $\epsilon_t \sim i.i.d. WN(0, \sigma_{\epsilon}^2)$

and $\sum_{i=1}^{\infty} \phi_i^2 < \infty, \sum_{i=1}^{\infty} \theta_i^2 < \infty$

여기서 ϕ 은 자기회귀 모형의 모수, θ 은 이동평균 모형의 모수, ϵ 은 백색잡음을 의미한다.

IV. 실험결과 및 분석

데이터는 Kaggle에서 John Hopkins 대학 COVID-19 데이터셋에서 Raw_global데이터를 사용한다[9].

4.1. 코로나19의 확진자수와 트렌드

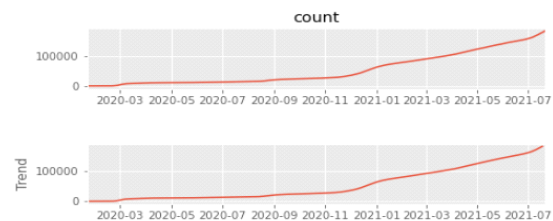


Fig. 1 South Korea, Covid-19 Case & Trend

위 그림1은 코로나19의 누적 확진자수와 트렌드를 보여준다. 2020년 11월부터 그래프가 가팔라지며 현재는 더 심해지고 있는 경향을 띤다. 별다른 대응책이 없다면 이러한 경향을 계속 될 것이다.

4.2. 1차분 한 데이터

일반적으로 차분의 횟수 d 가 높아질수록 정상성의 모습에 가까워지기 때문에 d 를 높게 잡을수록 ARIMA로 fitting될 가능성이 높다. 하지만 그럴 경우 ARIMA 모델식이 너무 복잡해져 데이터에 과적합된 모델이 나올 경향이 크기 때문에, 되도록 d 가 2보다 큰 ARIMA 모델을 사용할 경우엔 다른 모델을 찾는 것을 고려해야한다. 또한 두 개 이상의 ARIMA 모델로 fitting이 가능할 때는, 되도록 d 가 작은 ARIMA 모델을 사용하는 것이 좋다. 이것은 차분을 많이 하는 경우 오히려 fitting에 필요한 매개변수가 늘어나기 때문이다.

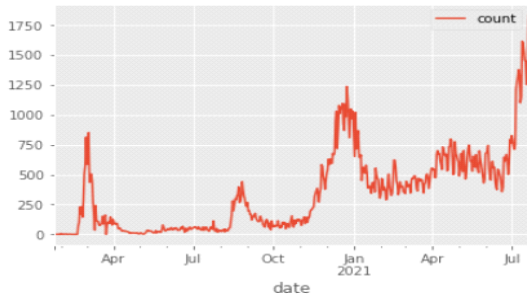


Fig. 2 1-Difference Data

위 그림2는 원래 데이터를 1차분한 데이터이다. 일일 확진자 수 그래프라고 볼 수도 있다. 차분을 통해서 데이터의 정상성을 확보하고, 추세와 계절성을 줄인다.

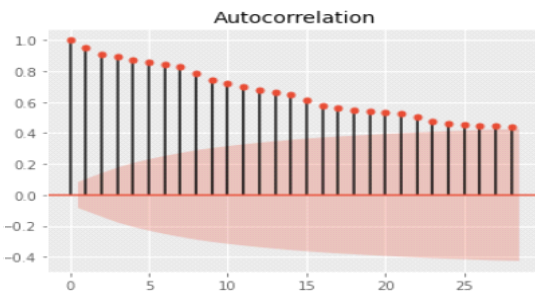


Fig. 3 1-Difference Auto Correlation

위 그림3은 1차분한 ACF 그래프다. 일반적으로 현실에 존재하는 시계열 데이터는 비정상성이고 이러한 비정상성은 크게 추세나 주기성으로 나뉜다. 추세나 주기성은 보통 plotting이나 ACF를 통해서 존재 여부를 확인할 수 있다. 시계열 데이터에 추세가 존재하는 경우는 보통 ACF가 중단되거나 아주 느리게 감소하고, 주기성

이 존재하는 데이터의 경우는 ACF도 같은 주기를 갖기 때문에 함수값이 요동치게 된다. 사실상 코로나 감염자가 계속 증가하는 상황에서 추세를 제거하긴 어렵다.

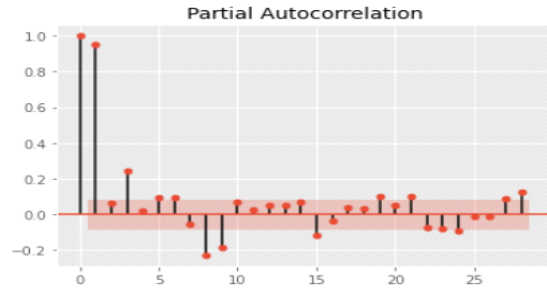


Fig. 4 1 Difference partial Auto Correlation

1차분한 PACF 그래프이다. 시차가 1에서 cut off되고 2부터 지수적인 감소를 하고 진동하는 sin형태다. 위 그림 3, 4에서 ACF는 빠르게 감소하고, PACF는 처음의 시차를 제외하고 급격히 감소한다. 이것은 AR의 특성을 띄며, $p=0, q=1$ 이 적당하다고 추측 할 수 있다. 그림2에서는 뒤에 값이 조금씩 유의값에서 튀는 것을 보이므로 조금의 계절성이 있는 것으로 판단된다. 그리고 파이썬에서 제공하는 Auto_arima 코드를 이용해 쉽게 p, d, q 값을 구할 수 있다. 여기서 p, d, q 값을 (0,1,1)로 지정했다. ARMA 과정의 특징(ACF, PACF)을 보이고 잔차 분석이 잘 되었는지 알기위해 사용한다.

4.3. ARIMA를 이용한 예측

Table. 1 Pratical infectee & Pred infectee

Date	Count	Pred Count	Count-Pred
2021-07-12	170,296	170,450	-154
2021-07-13	171,911	171,541	370
2021-07-14	173,511	173,162	349
2021-07-15	175,046	174,887	159
2021-07-16	176,500	176,522	-22
2021-07-17	177,951	178,014	-63
2021-07-18	179,203	179,456	-253
2021-07-19	180,481	180,670	-189

위 표1은 실제 데이터와 ARIMA를 이용한 예측값에 대한 비교로, 가장 직관적으로 알 수 있는 데이터이다. 하지만 시계열 데이터므로 날마다 쌓이는 데이터에 따라서 예측값은 바뀔 수 있음에 유의해야한다. 즉

최근 동향에 의해 예측값은 좌지우지 된다.

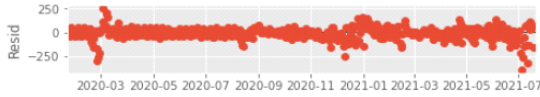


Fig. 5 Residual

위 그림5은 실제데이터와 ARIMA모형으로 예측값의 잔차값을 보여준다. 코로나 감염자 수가 급증될 때마다 0으로부터 멀리 퍼져있다. 7월 20일경 일일 확진자가 500명 가량 급증했는데, 이 경우 잔차가 400대 까지 증가함을 볼 수 있다.

V. 결론 및 향후연구

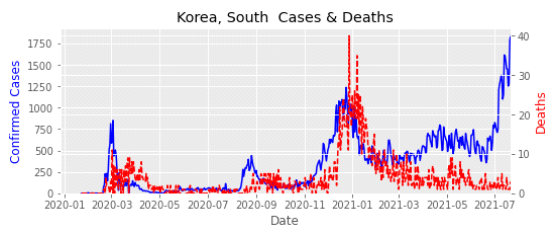


Fig. 6 South Korea, Case & Death

위 그림6의 그래프는 코로나 바이러스의 확진자 수, 사망자 수다. 사망자 수가 눈에 띄게 줄어든 것은 긍정적이지만 확진자수가 눈에 띄게 증가하는 것은 심각한 문제라고 볼 수 있다. 본 논문에서는 ARIMA모형을 이용해 코로나 확진자 수를 예측한다. 단기적인 측면(최대 2~3년)에서 예측하기 좋은 알고리즘으로 갑자기 급증하지 않는 이상 큰 잔차는 보이지 않는다. 하지만 데이터가 쌓인다면 이 잔차도 줄어들 것으로 보인다. 향후 연구로는 혈압 예측, 당뇨수치 예측, 응급의료 예측, 의료 예측 부분 등 많은 분야에서 사용될 수 있다.

REFERENCES

[1] Korea Centers for Disease Control and Prevention, COVID pademic [Internet]. Available: http://ncov.mohw.go.kr/tcmBoardView.do?brdId=&brdGubun=&dataGubun=&ncvContSeq=366394&contSeq=366394&board_id=140&gubun=BDJ.

[2] Centers for Disease Control and Provention. SARS-CoV-2 Variant Classifications and Definitions [Internet]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>.

[3] Our World in data, Statistics and ResearchCoronavirus (COVID-19) Vaccinations [Internet]. Available: <https://ourworldindata.org/covid-vaccinations>.

[4] Coronavirus Infectious Disease-19 (COVID-19), social distancing step 4 [Internet]. Available: <http://ncov.mohw.go.kr/shBoardView.do?brdId=6&brdGubun=64&ncvContSeq=5619>.

[5] D. H. Kim, M. W. Kim, B. J. Lee, K. T. Kim, and H Y Youn, "Data Flow Prediction Scheme using ARIMA Model," *Proceedings of the 2018 Korean Society for Computer and Information Sciences Summer Conference*, vol. 26, no. 2, pp. 141-142, 2018.

[6] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1394-1401, Dec. 2018.

[7] D. Serge and L. Sophie, "Characterization of the partial autocorrelation function of nonstationary time series," *Journal of Multivariate Analysis*, vol. 87, no. 1, pp. 46-59, Oct. 2003.

[8] P. Giovanni, P. Sonia, and C. Patrizia, "Dynamic Linear Models with R," *Springer Science & Business Media*, pp. 31-84, May. 2009.

[9] Kaggle data set. covid19-data-from-john-hopkins-university [Internet]. Available: <https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university>.



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