

DTMC기반 감염병 확산 모델링을 통한 숨은 감염자 수 추정 연구: 대한민국 COVID-19데이터를 중심으로

안수진*, 권민혜^o

DTMC-Based Epidemic Model for Hidden Patients Estimation: Using COVID-19 Dataset of the Republic of Korea

Sujin Ahn*, Minhae Kwon^o

요약

코로나바이러스(SARS-CoV-2; COVID-19)와 같은 감염병의 전파 과정을 이해하기 위해서는 수학적 모델링을 활용하여 감염병의 진행 양상을 파악하는 것이 중요하다. 본 논문에서는 discrete-time Markov chain(DTMC; 이산시간 마르코프 체인)을 활용한 감염병 모델을 제시하고, 이를 바탕으로 숨은 감염자 수를 데이터 기반 방식으로 추정하고자 한다. 이를 위해, 감염병 전파 양상을 결정하는 핵심 파라미터 중 하나인 감염 전파율을 시변성을 가지도록 설정하여 감염병 확산을 억제하는 정부의 방역 정책과 감염 시작일로부터 전파 진행 기간에 따라 변화하는 감염병 확산 정도를 정량적으로 분석 가능하도록 하였다. 또한, 기존의 감염병 전파 모델에서 격리 상태와 백신 접종 상태를 추가하여 현실 정책을 잘 반영하는 모델을 제안하였다. 제안한 모델을 통해 잠복기 상태의 인구수, 병에 감염되었지만 격리되지 않아 병을 전파시키는 숨은 감염자 수 등과 같이 공식적으로 집계될 수 없는 데이터지만 전파양상 파악에 중요한 잠재 상태(latent state)의 인구수를 추정할 수 있다. 제안된 모델의 우수성을 확인하기 위해 최신 모델 두가지와 비교하였고, 그 결과 제안된 모델이 확진자 수 급증의 흐름을 가장 잘 반영함을 확인하였다.

Key Words : Markov chain, Epidemic model, COVID-19, Compartmental models, Latent state estimation

ABSTRACT

Recently, the importance of mathematical modeling for infectious diseases has been highlighted. It can investigate the transmission by capturing the ongoing pattern of infectious diseases. This paper proposes a data-driven approach based on discrete-time Markov chain model for the epidemic by estimating the model's parameters and latent state values. We introduce the key model parameters such as transmission rate as time-varying so that the parameters can quantitatively evaluate the effectiveness of the national prevention policies. Furthermore, our epidemic model contains two additional states complying with national intervention: vaccinated and isolated states. To verify the proposed model, we use the COVID-19 dataset of the Republic of Korea and analyze the data by using estimated parameters and the number of people in the latent states. Through the results, we confirm that the proposed solution performs the best by capturing important details.

* 본 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(IITP-2021-0-00739, 분산/협력 AI기반 5G+네트워크 데이터 분석 기능 및 제어 기술 개발)과 한국연구재단(NRF-2020R1F1A1069182, 모바일 네트워크에서의 강화학습기반 분산적 의사결정방법 연구)의 지원을 받아 수행된 연구임.

• First Author : Soongsil University Department of Information and Telecommunication Engineering, asujin331@soongsil.ac.kr, 학생회원

^o Corresponding Author : Soongsil University School of Electronic Engineering, minhae@ssu.ac.kr, 종신회원

논문번호 : 202205-094-C-RN, Received May 6, 2022; Revised July 13, 2022; Accepted July 14, 2022

1. 서 론

2020년 3월, 세계보건기구는 COVID-19로 인한 팬데믹을 선언하였다. 많은 국가가 마스크 착용, 여행 제한, 격리 정책 등의 구체적인 지침(guideline)을 결정하는 데 많은 노력을 쏟았다. 더욱이, 몇몇 국가들은 감염병 발생 초기 의료 병상 부족, 의료 인력 배치 불균형 등의 어려움을 겪으며 방역정책의 중요성을 인식하게 되었다^[1]. 변이 바이러스와 돌파 감염(breakthrough infection)으로 인해 감염병이 지속되는 상황에서 각 국가의 정부는 장기적이고 효율적인 대처 방안을 강구하게 되었다. 방역 정책을 결정하는데 있어 근거가 되는 감염병의 수학적 모델링의 중요성이 높아지고 있다. 이는, 감염병의 수학적 모델이 방역 정책의 감염병 확산 억제 정도를 객관적으로 분석하고, 감염병 유행 확산 예측을 위한 도구로 사용될 수 있기 때문이다.

감염병 모델을 수학적으로 접근하는 이유는 다음과 같다^[2]. 첫째로, 윤리적인 문제로 인한 현실에서의 실험적 검증이 불가능하기 때문에 수학적으로 재현함으로써 감염병의 감염 양상을 파악하는 과정이 필요하다. 둘째로, 갑작스러운 신종 감염병의 발생으로 인해 양질의 경험적 데이터가 충분하지 않거나 앞으로의 상황이 예측되지 않는 불확실성이 높은 상황에서 수학적 모델링은 미래의 수치 예측에 기여할 수 있다. 그러므로, 감염병의 수학적 모델링은 다양한 질병의 경향성을 이해하고 감염병의 역학 예측을 위한 시스템을 구축할 수 있다는 점에서 유용하다.

감염병 모델링의 대표적인 접근으로는 시계열 분석(time series analysis), 딥러닝(deep learning), 구획 모델(compartmental models) 등이 있다. 시계열 분석은 추세 요인, 계절 요인, 순환 요인 등의 변동 요인들을 파악하여 시계열 자료를 분석하는 방법이다^[3]. 하지만 시계열 분석은 자료가 가지는 통계적 분포에 대한 가정이 위배되지 않아야하며, 한정된 수의 변동 요인만을 고려하여 추정 값과의 복잡한 연관 관계를 파악하기에 변동요인 선택에 성능이 크게 의존된다는 한계점이 있다. 딥러닝 모델^[4]은 신경망의 layer를 사용하여 출력 값과 특성(feature)과의 깊은 관계성을 근사화 할 수 있다. 딥러닝 모델은 학습과정을 통해 중요한 변동 요인에는 높은 가중치를 부여하거나 활성화 함수를 통해 비선형성까지 고려하여 출력 값에 반영할 수 있다는 장점이 있다. 시계열 데이터에 관한 대표적인 딥러닝 기반의 방법에는 RNN(Recurrent Neural Network)이 있다^[5]. 입력 값을 바탕으로, 새로

운 출력 값에 대한 예측이 가능하다. 하지만, 딥러닝 모델 학습에는 추론 결과에 대한 평가를 위해 실제 데이터가 필요하다. 따라서 실제 데이터가 부재하는 상황에서는 해당 데이터에 대한 예측이 불가능하다. 이에 반해, 구획 모델은 감염 진행 상황과 같이 순서가 있는 상태를 추정할 수 있기 때문에 실제 데이터가 부재하는 잠재 상태를 추정하기에 용이하다. 여기서 구획 모델이란 전체 인구수를 감염 상태에 따라 여러 구획으로 나눈 것을 말한다. 구획 모델은 감염자의 비율을 고려하여 접촉 수를 계산하여 감염병이 확산하는 현상을 분석적으로 설명한다. 이로 인해, 구획 모델은 숨은 감염자 수와 같은 정부가 공식적으로 제공하지 않는 데이터에 대해서도 추정 가능하다.

대표적인 구획 모델로는 Susceptible-Infectious-Recovered (SIR)^[6]와 Susceptible-Exposed-Infectious-Recovered (SEIR)^[7] 모델이 있다. SIR 모델은 감수성(susceptible; S) 상태에서 감염 상태의 개체와의 접촉을 통해 감염(infectious; I) 상태로 변화될 수 있으며, 이후에는 회복기를 거쳐 회복(recovered; R) 상태로 도달한다. 이와 같은 구획 모델은 감염병이 종료될 때까지의 전반적인 감염 양상을 이해하는 데 도움이 된다. 하지만, 구획 모델은 다음과 같은 한계점이 있다. 첫 번째로 기존의 구획 모델은 결국 모든 인구가 마지막 단계에서는 회복 상태에 도달한다는 이론을 이끌며 감염병의 초기 지점부터 종료 지점까지의 거시적인 진행 과정만을 파악할 수 있다. 이는, 기존의 구획 모델에서의 파라미터들의 종류를 시간 독립적인 상수로 가정하면서 처음과 끝이 상수인 파라미터에 의해 결정되는 결정론적인(deterministic) 모델로 접근을 하였기 때문이다. 거시적인 관점보다는 미시적인 관점에서 감염병이 진행되는 모델을 구축하기 위해서는 상수였던 파라미터를 시간에 의해 변하는 변수(variable)로 바꾸어 작동하는 것이 현실에 더 유의미할 수 있다. 두번째로, SIR모델과 같은 고전적인 구획 모델은 S, I, R과 같은 제한된 구획만 설명 가능하다. 하지만, 현재와 같이 백신이나 격리를 정책적으로 도입한 복잡한 상황에서는 기존의 구획만으로는 상황을 구체적으로 설명할 수 없기 때문에 구획을 추가하여 현 상황을 이해하는 것이 필요하다. 마지막으로, 기존의 구획 모델은 결정론적 모델의 특성에 의해 잠복기 상태, 감수성 상태 등의 잠재 상태의 각 인구수를 그래프로 표현하면 연속적인 곡선으로 재현될 뿐 이산적인 형태로 추정하기 어려웠다. 즉, 매일매일 이산적인 형태로 증감하는 현실 세계의 데이터와 이론에 기반한 감염병 모델 사이의 괴리는 불가피했다. 따라서,

일일 확진자 수, 사망자 수 등 실제 데이터에 기반하여 잠재 상태의 인구수를 이산적인 형태로 파악한다면 기존의 감염병 모델보다 현실적인 세계를 잘 재현해낼 수 있을 것이다. 추가적으로, 잠재 상태의 수가 파악 가능하다면 통계적으로 집계되지 않은 숨은 감염자의 수를 추정 가능하므로 정책적으로 중요한 정보를 제공함과 동시에 정부는 앞으로의 상황을 미리 대비할 수 있다는 장점이 있다.

본 논문에서는 이산 시간 마르코프 체인을 활용하여 데이터 기반의 감염병 모델을 제안한다. 국내의 정책을 잘 반영할 수 있도록 백신 접종 상태, 격리 상태 등을 포함한 7가지 상태를 포함하여 수학적 모델링을 진행하였다. 이후, 한국의 COVID-19 데이터를 제안한 모델에 적합하여 잠복기 상태, 숨은 감염자 상태 등의 잠재 상태에 속해 있는 각 상태의 수를 추정하였다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 본 연구와 관련된 선행 연구를 서술한다. III장에서는 본 논문에서 제안한 감염병 모델에 대해 설명하고, IV장에서는 파라미터 및 숨은 감염자 수 추정 방법을 제안한다. V장에서는 시뮬레이션 결과, 다른 모델과의 비교 결과에 대해 분석하고, VI장에서는 본 논문에서 제안한 감염병 모델에 대한 결론을 내린다.

II. 선행 연구

COVID-19가 발생하면서 다양한 구획 모델은 여러 분야에서 중요한 역할을 하였다. 감염병 발생 초기에는 다양한 구획을 나눌 만한 구체적인 체계가 잡혀 있지 않았기 때문에 SIR 모델이나 SEIR 모델을 변형 없이 감염자 수 추정을 위해 이용한 경우가 많았다. SEIR 모델을 토대로 우한의 감염자 수를 추정하고, 이를 바탕으로 다른 나라의 미래의 감염병 확산 정도를 추정한 연구가 있다⁸⁾. 이를 좀 더 확장하여, SEIR 모델에 AI (Artificial Intelligence) 모델을 활용하여 우한 지역의 감염병 추세를 분석한 뒤 우한 외의 지역의 감염병 추세를 예측한 연구도 있다⁹⁾. 하지만 SIR, SEIR 모델은 한정된 구획의 수만을 설명하기 때문에 인구의 감염 상태가 세분화된 경우에는 다양한 감염 상태를 반영하지 못한다는 단점이 있다. 이러한 한계점을 극복하기 위해 고전적인 SEIR 모델을 확장하는 연구가 진행되고 있다.

백신 접종이 시작되고, 감염자의 접촉 빈도를 감소시키기 위한 격리 정책이 시행되면서 구획을 세분화하여 설명할 수 있도록 격리 상태나 백신 접종 상태를 포함한 모델들이 제안되었다. 격리 상태를 포함한

Susceptible - Exposed - Symptomatic infectious - Hospitalized - Recovered (SEIHR) 모델에서는 격리를 파라미터를 상수의 형태로 추가하여 방역 조치 효과에 따른 시나리오별 확진자 수를 제시하였다¹⁰⁾. 백신 상태가 포함된 SEIR 모델에서는 MTL (Metric Temporal Logic) 개념을 이용하여 모델을 제시하였다¹¹⁾. 더 나아가, 병의 중증도를 포함하거나 감염되었지만 진단되지 않은 경우도 고려하는 등 다양한 접근 방식도 나타났다. Susceptible - Infected - Diagnosed - Ailing - Recognized - Threatened - Healed - Extinct (SIDARTHE) 모델은 병의 중증도, 진단 여부에 따라 세분화하여 구획을 추가하였다¹²⁾.

확률론적 과정은 파라미터의 타입과 모델 구조에 따라 수학적 모델에 적용될 수 있다. 확률론적 과정은 현상을 현실과 비슷하게 재현해준다는 장점이 있다. SIR 모델에 확률론적 과정인 마르코프 체인을 적용하여 결핵, HIV 등의 질병의 상태 전이 행렬을 나타낸 제시한 연구¹³⁾와 입원 상태, 무증상 상태를 추가하고 연속 시간 마르코프 체인 (continuous-time Markov chain) 을 적용한 Susceptible - Exposed - Infected - Hospitalized - Asymptomatic - Cured - Death (SEIHACD) 모델에서는 시뮬레이션을 통해 COVID-19의 상태 값을 추정한 연구가 존재한다¹⁴⁾.

세분화된 상태를 고려한 모델 및 확률론적 과정을 적용한 모델들이 출현했지만, 현실의 감염병 전파 양상을 정확하게 파악하기 위해서는 감염병 모델의 파라미터의 특성을 고려하여 설계하는 것이 필요하다. 예를 들어, 구획 모델의 경우에는 감염전파율이 감염자 증가에 큰 영향을 미치므로 정확하게 추정해야 할 파라미터이다. 감염 전파율이란, 감수성자가 감염 상태의 개체에 의해 감염되는 비율(rate)로, 감염병의 확산 정도를 측정할 수 있는 수치이다. 감염전파율을 정확하게 추정하기 위해서는 잠복기에 속하는 감염자의 수, 감수성자의 수 등의 잠재 상태의 수가 필요하다. 하지만, 이러한 데이터는 공식적으로 기록될 수 없는 데이터이기 때문에 감염전파율을 추정하기는 쉽지 않다. 따라서 감염전파율을 상수로 가정한 채로 연구를 진행하고 있다¹⁵⁾. 하지만 현실 세계에서 감염 전파율은 접촉 횟수에 의해 영향을 받으므로 접촉 수를 제한하는 사회적 거리두기와 같은 방역 정책들이 감염전파율에 영향을 미친다¹⁶⁾. 사회적 거리두기와 같은 정책들은 당시의 확진자의 수와 증감 정도에 따라 그 강도가 달라지므로, 감염전파율 또한 시변성을 갖는 파라미터라는 가정이 현실적이다. 관련된 연구로, 감염전파율을 시간과 위치에 따라 변하도록 설계하였지만,

SEIR모형을 사용하여 한정된 구획만을 설명하였다^[17]. Dynamic - Susceptible - exposed - Infective - Quarantined (D-SEIQ)모형에서는 감염전파율을 시간에 따라 변하는 감염 재생산지수 (reproduction number)에 잠복기로 나눈 값으로 설정하였다^[18]. 하지만, 격리자 수, 격리 해제 수 등의 다양한 데이터를 활용하지 못하고, 확진자 수만을 반영하여 모델 식을 적은 수의 파라미터로 단순화하였다는 한계점이 있다.

고전적인 SEIR모형을 확장하는 다양한 선행연구가 존재한다. 그러나 시간에 따라 변화하는 감염병의 확산 정도를 정량적으로 분석하기 위해서는 감염전파율의 시변성을 고려하여 추정하는 것이 중요하다. 따라서, 본 논문에서는 방역 정책을 고려하여 모델을 세분화하고, 감염병 모델 파라미터를 데이터 기반으로 추정하고자 한다. 이때, 감염전파율을 비롯한 중요 파라미터에 시변성을 부여하고, 확진자 수, 격리자 수, 격리해제 수 등의 활용 가능한 데이터를 반영하여 기존의 결정론적 감염병 모델보다 현실적인 모델을 수립하고자 한다. 수학적 재현 과정을 통해 감염병 확산 현상을 수학적으로 설명가능한 영역으로 전환함으로써, 정책 의사 결정에 중요한 근거가 될 수 있는 증거 기반 (evidence-based)의 감염병 모델을 제시하고자 한다.

III. DTMC 기반의 감염병 모델 제안

본 논문에서는 DTMC기반으로 상태와 상태 전이 확률을 제안한다. 한 명의 인구는 특정 시점에 오직 한 가지의 상태에만 속할 수 있으며, 각 상태의 인구 수는 음이 아닌 정수이다. 본 모델은 다음과 같은 가정을 포함한다. 첫째로, 모든 인구수는 동등하게 섞여 있다. 즉, 임의의 한 개체가 감염된 상태의 개체를 만났을 때(encounter)의 확률은 장소와 관계없이 항상 동일하다. 이는 개체 간의 상호작용을 균질적으로 가정하였기 때문이다. 두번째로, 재감염이나 돌파 감염의 경우는 고려하지 않는다. 세번째로, 국가나 장소에 따라 크게 영향을 받는 역학적(epidemiological) 파라미터를 항상 일정하게 설정한다. 역학적 파라미터란 감염병의 고유의 특성을 표현하는 파라미터이다. 네번째로, 해외 유입을 통한 감염은 제외하고, 국가 내에서의 감염 발생 경우만을 고려한다.

3.1 상태 정의

- **S (Susceptible)** 미래에 감염될 확률이 있는 감수성자 상태를 의미한다. 감염되거나 백신을 맞은 경험이 없는 상태이다.

- **V (Vaccinated)** 백신 접종을 1번이라도 실시한 상태이다. 백신 상태가 되고 일정 기간이 지나면 면역성을 갖추게 된다.
- **E (Exposed)** 감염되었지만 잠복 기간(latent period)에 속하므로 바이러스를 노출할 수 없어 다른 사람을 감염시킬 수 없는 상태이다.
- **I (Infectious)** 감염이 된 후 잠복기 상태가 지나 다른 사람에게 바이러스를 전파할 수 있는 상태이다. 감염되었지만 아직 확진이 되지 않아 격리되지 않은 숨은 감염자 상태이다.
- **H (Hospitalized)** 확진이 된 후 격리 중인 상태이다. 격리 중인 상태이기 때문에 환경적으로 다른 사람에게 전파할 수 없는 상태이다.
- **R (Recovered)** 면역성을 갖춘 상태이다. **I**상태나 **H** 상태가 면역성을 갖춘 뒤에 도달할 수 있는 상태이다.
- **D (Dead)** 감염으로 인해 사망한 상태이다.

t 시점에서의 **K** 상태에 속하는 값인 $K(t)$ 는 **K** 상태에 머무르는 인구 수로 상태 값을 의미한다. $t = 0, 1, 2, \dots, T$ 로, 설정한 초기일로부터 t 번째 시점을 의미하고, $K \in \{S, V, E, I, H, R, D\}$ 을 만족한다. 전체 인구수 N 은 K 에 속하는 7가지 상태 값의 합으로 표현된다. 전체 인구수 N 은 인구의 출생과 자연 사망을 고려하지 않았기 때문에 항상 일정한 상수임을 가정한다.

3.2 상태 전이 확률

마르코프 체인은 $t-1$ 시점까지의 과거 관측값과 t 시점의 현재의 관측값이 주어진 상황에서 $t+1$ 시점의 미래의 예측 값의 조건부 확률 분포는 현재의 관측 값에만 의존한다고 가정한다^[19]. $J, J' \in \{S, V, E, I, H, R, D\}$ 을 만족하고 $p_{JJ'}$ 는 t 시점의 **J** 상태에서 $t+1$ 시점에서 J' 상태로 전이될 확률을 말한다. $J' = J$ 일 때의 $p_{JJ'}$ 는 $t+1$ 시점에서도 **J** 상태에 머무르는 확률을 말한다. 역방향 상태 전이 확률은 0으로 고려한다.

$X_{JJ'}(t)$ 는 t 시점에서의 **J** 상태에서 J' 상태로 전이될 때의 인구수를 나타내는 확률 변수로 $p_{JJ'}(t)$ 의 확률을 갖는 이항 분포(binomial distribution)를 따른다. $X_{JJ'}(t)$ 는 J' 상태 대신에 이전 상태인 **J**에 속한다고 가정하므로 $X_{JJ'}(t) \in J(t)$ 로 표현할 수 있다. 이항 분포 성질에 의해 기대되는 인구수는 $\mathbb{E}[X_{JJ'}(t) | J(t)] = J(t)p_{JJ'}(t)$ 이다. 본 모델에서 사용한 파라미터는 표 1에 정리하였다. 본 모델에서 설정한 각 상태 간 이동 과정과 확률을 나타낸 것은 그림 1과 같다.

그림 1에 따르면 **S** 상태에서는 **V** 상태와 **E** 상태로

전이될 수 있다. **V** 상태로 전이될 때의 확률은 $p_{SV}(t) = \theta(t)$ 로, 여기서 $\theta(t)$ 는 t 시점에서의 일일 백신 접종 비율을 나타낸다. **S** 상태에서는 **I**상태의 개체와의 접촉을 통해 **E** 상태로 전이되는데 이때의 확률은 $p_{SE}(t) = (1 - \theta(t)) \beta(t; z) \frac{I(t)}{N^*(t)}$ 이다. 여기서 $\beta(t; z)$ 는 t 시점의 감염전파율을 의미하고 $N^*(t) = N - H(t) - D(t)$ 로 전체 인구수에서 격리되거나 죽은 인구수를 제외하여 t 시점에서 **I** 상태가 만날 수 있는 남아 있는 전체 인구수를 의미한다. t 시점에서 백신을 맞지 않아 여전히 감수성자의 상태로 남아 있는 비율은 $1 - \theta(t)$ 이고, $\beta(t; z) \frac{I(t)}{N^*(t)}$ 는 감수성자에게 병이 전파될 가능성도 (likelihood)를 의미한다. 종합해보면, $p_{SE}(t)$ 는 백신을 맞지 않은 나머지 사람이 **I** 상태의 개체와의 접촉을 통해서 감염될 확률을 뜻한다. 식 (1) 은 **S** 상태 값을 나타내는 수식이다.

$$S(t + 1) = S(t) - X_{SV}(t) - X_{SE}(t) = S(t) - \theta(t)S(t) - (1 - \theta(t)) \frac{I(t)}{N^*(t)} \beta(t; z) S(t) \quad (1)$$

E상태에서는 잠복기가 지나면 전파 가능한 상태인 **I** 상태로 도달할 수 있다. $p_{SE}(t) = \epsilon$ 에서 ϵ 는 감염된 이후 전염력이 생기기까지의 기간인 평균 잠복 기간의 역수이다. **E**상태 값은 식 (2)에 따라 결정된다.

표 1. 제안한 모델의 파라미터
Table 1. Parameters of proposed model

Notation	Description
N	Total population
$\beta(t; z)$	Transmission rate
$1/\epsilon$	Mean latent period
$1/\gamma$	Mean infectious period
$1/\xi$	Mean duration of build immunity
$\delta(t)$	Fatality Rate
$\rho(t)$	Recovery rate for hospitalized individuals
$\varphi(t)$	Detected case ratio
$\theta(t)$	Vaccination ratio of the first dose

$$E(t + 1) = E(t) + X_{SE}(t) - X_{EI}(t) = E(t) + (1 - \theta(t)) \frac{I(t)}{N^*(t)} \beta(t; z) S(t) - \epsilon E(t) \quad (2)$$

I 상태에서는 **H** 상태와 **R** 상태로 이동할 수 있다. **H** 상태에 도달하기위한 상태전이 확률 $p_{IH}(t) = \varphi(t)$ 이고, $\varphi(t)$ 는 t 시점에서의 일일 확진 비율 (detected case ratio)을 나타낸다. 이는 확진이 되고 나면 격리가 진행되어 전파를 막도록 하는 국내 정책에 따르도록 하였다. 무증상 감염으로 인해 확진이 어려운 상태에서는, 그 증상의 정도가 가벼우므로 자연적으로 회복할 가능성이 있다. 이를 수학적으로 설명하기 위해

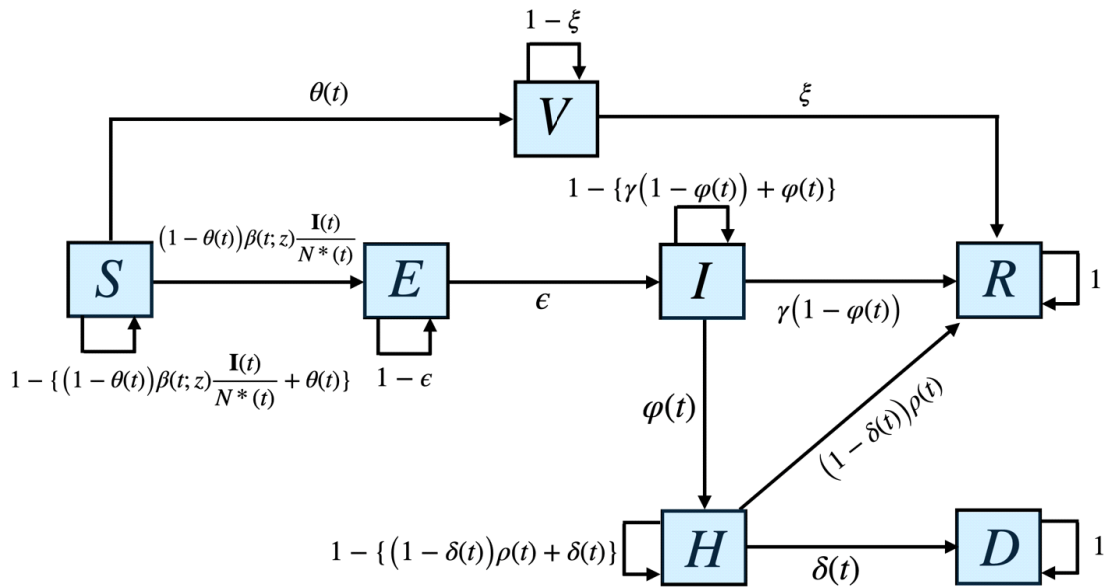


그림 1. 제안한 모델의 State Transition Diagram
Fig. 1. State Transition Diagram of proposed model

R 상태에 도달하기위한 상태전이 확률 $p_{IR}(t) = \gamma(1 - \varphi(t))$ 로 설정하였고, 확진 판정을 받지 않은 비율에 회복률을 뜻하는 γ 를 곱하였다. **I**상태 값은 식 (3)을 따른다.

$$\begin{aligned} I(t+1) &= I(t) + X_{EI}(t) - X_{IH}(t) - X_{IR}(t) \\ &= I(t) + \epsilon E(t) - \varphi(t)I(t) - \gamma(1 - \varphi(t))I(t) \end{aligned} \quad (3)$$

H 상태에서는 **D** 상태와 **R** 상태로 이동할 수 있다. **D** 상태로 가는 전이 확률은 $p_{HD}(t) = \delta(t)$ 이고, $\delta(t)$ 는 t 시점에서의 격리한 사람의 사망률을 의미한다. **R** 상태로 가는 전이 확률은 $p_{HR}(t) = \rho(t)(1 - \delta(t))$ 이고, 사망하지 않는 사람이 회복한 확률을 의미한다. 여기서, $\rho(t)$ 는 t 시점에서의 격리한 사람의 회복률을 나타내므로, **H** 상태의 값을 계산하는 식은 (4)와 같다.

$$\begin{aligned} H(t+1) &= H(t) + X_{IH}(t) - X_{HR}(t) - X_{HD}(t) \\ &= H(t) + \varphi(t)I(t) - \rho(t)(1 - \delta(t))H(t) \\ &\quad - \delta(t)H(t) \end{aligned} \quad (4)$$

V상태에서 **R** 상태로 가는 확률은 $p_{VR}(t) = \xi$ 이고, ξ 는 최소 한번의 백신 접종을 한 이후 완전한 면역력이 생기기까지 걸리는 평균 기간으로 정의한다. 백신 종류에 따라 접종 횟수가 달라지고, 따라서 면역력이 생기기까지의 기간도 달라질 수 있다. 국내에서 가장 보편적으로 사용되고 있는 백신을 기준으로 ξ 는 1차 백신접종을 기준으로 2차 백신 접종을 맞는 기간과 면역력이 생기는 기간까지 포함하는 기간이다. 이후 **R** 상태로 이동하게 되고 **V** 상태의 식은 (5)와 같다.

$$\begin{aligned} V(t+1) &= V(t) + X_{SV}(t) - X_{VR}(t) \\ &= V(t) + \theta(t)S(t) - \xi V(t) \end{aligned} \quad (5)$$

마지막 단계인 **D** 상태와 **R** 상태와는 흡수 상태 (absorbing state)로 다른 상태로 이동할 수 없는 상태이다. **D** 상태와 **R** 상태의 수식은 각각 (6), (7)으로 표현하였다.

$$\begin{aligned} D(t+1) &= D(t) + X_{HD}(t) \\ &= D(t) + \delta(t)H(t) \end{aligned} \quad (6)$$

$$\begin{aligned} R(t+1) &= R(t) + X_{IR}(t) + X_{VR}(t) + X_{HR}(t) \\ &= R(t) + \gamma(1 - \varphi(t))I(t) + \xi V(t) \\ &\quad + \rho(t)(1 - \delta(t))H(t) \end{aligned} \quad (7)$$

IV. 제안한 모델의 파라미터 및 숨은 감염자 수 추정

제안한 모델의 각 상태를 추정하기 위해서는 $\beta(t; z)$ 를 파악하는 것이 중요하다. 특히, $\beta(t; z)$ 추정을 위하여 실시간 감염재생산지수 (instantaneous reproduction number)를 도입하였다. 감염재생산지수란 한 명의 감염자가 평균적으로 감염시킬 수 있는 2차 감염자의 수로 정의한다. 실시간 감염재생산지수는 특정 시간의 감염재생산지수를 추정하여 정책, 행동 패턴 등의 외부적 요인의 효과를 분석하는 데에 적합하다. 본 논문에서는 다양한 접근을 포함하고 있는 실시간 감염재생산지수 추정법 중 Cori의 방법론을 채택하였다. Cori의 실시간 감염재생산지수는 식 (8)과 같다²⁰⁾. τ 의 경우, 감염병 시작일로부터 τ 번째 시점으로 정의한다.

$$R_\tau = \sum_{s=1}^{\tau} \frac{X_{IH}(\tau)}{X_{IH}(\tau-s) w(s)} \quad (8)$$

$w(s)$ 는 감염된 후 s 일이 지난 후 다른 사람에게 전염시킬 수 있는 확률 수치로 현재의 감염성을 의미한다. 즉, τ 시점에서의 감염재생산지수는 τ 시점에서의 확진자 수를 τ 시점에서의 감염성의 값을 기반으로 산출된 추정 확진자 수를 나누어 산출한다.

본 연구에서 동일한 날짜를 바탕으로 τ 와 t 를 산출하여 구한 $\beta(t; z)$ 는 식 (9)와 같이 정의하였다.

$$\beta(t; z) = \frac{R_\tau}{z} \quad (9)$$

z 는 $\beta(t; z)$ 를 역 스케일링(inverse scaling) 역할을 한다. **I** 상태는 전방 추론을 통한 $\hat{I}_f(t; z)$ 와 후방 추론을 통한 $\hat{I}_b(t)$, 두가지 방향에서 추론될 수 있다. 이들의 차이를 최소화하는 z 를 찾는 것이 목표이다. 따라서, 최적의 z 를 찾기 위해 설정한 Loss function은 식 (10)과 같다.

$$z_{t+1} \leftarrow \arg \min_z = \frac{1}{T+1} \sum_t \frac{(\hat{I}_f(t; z) - \hat{I}_b(t))^2}{\sigma} + \lambda \quad (10)$$

이때, λ 는 상태 값이 음수가 되지 않도록 하는 처벌 항을 의미한다. σ 는 식 (3)을 통해 추정된 $\hat{I}_f(t; z)$ 와 $\hat{I}_b(t)$ 의 MSE (Mean Squared Error)를 조절한다. $\hat{I}_b(t)$ 는 확진자 수 X_{IH} 를 바탕으로 후방 추론을 통해

설계한 식이고, $\hat{I}_f(t; z)$ 는 식 (3)을 확장하여 $\mathbf{S}(t)$ 와 z 를 바탕으로 전방 추론으로 표현한 식이다. $\hat{I}_b(t)$ 의 식은 (11)과 같다.

$$\hat{I}_b(t) = \varphi^{-1} X_{IH}(t) + \sum_{m=2}^t (\varphi^{-1} - 1)^m X_{IH}(t - m + 1) (1 - \gamma)^m \quad (11)$$

여기서, φ 는 임의로 가정한 확진 비율을 나타낸다. 즉, φ^{-1} 는 확진 판정을 받기 전의 총 감염자 수를 추정하는 역할을 한다. 따라서, $\varphi^{-1} X_{IH}(t)$ 는 t 시점의 총 감염자 수를 뜻한다. $\sum_{m=2}^t (\varphi^{-1} - 1)^m X_{IH}(t - m + 1) (1 - \gamma)^m$ 는 t 시점까지 여전히 확진되지 않고, 회복되지 않은 상태의 수를 말한다. 이는, 감염된 이후 면역력이 형성되지 않아 \mathbf{R} 상태가 아닌 \mathbf{I} 상태에 머무르게 된 수이다. 즉, $\hat{I}_b(t)$ 는 확진된 감염수와 확진되지 않은 감염수를 포함한 t 시점의 총 감염 수와 t 이전 시점을 바탕으로 추정된 t 시점에서 여전히 \mathbf{I} 상태에 머무르는 수의 합을 나타낸다. $\hat{I}_f(t; z)$ 는 전방 추론을 활용하여 구한 숨은 감염자 수식으로 식 (12)과 같이 표현이 가능하다.

$$\hat{I}_f(t; z) = \epsilon(1 - \epsilon)\mathbf{E}(t - 2) + \epsilon(1 - \theta(t - 2)) \frac{\mathbf{I}(t - 2)}{N^*(t - 2)} \beta(t - 2; z) \mathbf{S}(t - 2) + (1 - \gamma) \mathbf{I}(t - 1) - (1 - \gamma) X_{IH}(t - 1) \quad (12)$$

최종적으로, 식 (12)를 통하여 본 모델에서 제안한 숨은 감염자 상태인 \mathbf{I} 상태를 추정할 수 있다.

V. 제안한 모델의 검증

본 연구에서 사용한 데이터는 격리 중인 인원, 격리 해제 인원, 확진자 수, 사망자 수가 모두 공개된 2021년 4월 15일부터 2021년 10월 7일까지의 대한민국 데이터셋에 적합하여 잠재 상태의 수를 추정하였다¹⁾. 표 1에서 가정한 ϵ, γ, ξ 는 역학적 파라미터로서 각각 1/3, 1/9, 1/49로 설정하였다^[21-23]. 모델의 파라미터인 $\beta(t)$ 를 추정하기 위해 필요한 $w(s)$ 는 WHO-PAHO (Pan American Health Organization)에 의해 평균 4.8, 표

준편차 2.3을 따르는 감마 분포를 이용하였다^[24].

대한민국 COVID-19데이터 적합에 있어 \mathbf{I} 상태의 마지막 값은 0으로 수렴하는 경우를 피하기 위해 초기점과 끝점에 5일의 margin을 주어 구간을 설정하였다. 정책의 변화에 따라 구간을 끊어 각 구간 내의 최적의 z 를 찾도록 하였다. \mathbf{I} 상태의 초기값은 식 (11)과 같다. 구간별로 각 상태의 초기값은 초기화된다. 본 실험의 Loss function에 사용된 값은 표 2와 같다.

표 2. Loss function에 사용된 상수 값
Table 2. Summary of constants used in Loss function

Notation	Description	Value
N	Total population	51821669
σ	Inverse scaling coefficient	10000
λ	Penalty coefficient	1000

5.1 사용한 데이터셋

본 연구에서 적용한 데이터는 보건복지부의 ‘코로나나 감염 현황 데이터’와 질병 관리청의 ‘코로나 19 예방 접종 통계 데이터 조회 서비스’를 사용하였다^[25]. ‘코로나 감염 현황 데이터’는 해당 날짜의 격리 중인 인원, 격리 해제 인원, 확진자 수, 사망자 수 등을 담고 있다. ‘코로나 19 예방 접종 통계 데이터 조회 서비스’는 해당 날짜의 1차 백신 접종자 수, 2차 백신 접종자 수에 대한 데이터가 기록되었다. 보건복지부가 주관하는 COVID-19 사이트를 활용하여 정책의 변화를 조사하였다^[26]. COVID-19 사이트를 활용하여 요약한 policy timeline은 표 3과 같다.

표 3. 대한민국의 정책 흐름표
Table 3. Summary of Policy timeline in the Republic of Korea

Date	Description
2021.03.29	Social distancing level 2
2021.07.01	Reorganization of social distancing
2021.07.07	The 4th wave of COVID-19
2021.07.11	Social distancing level 4
2021.09.18	National holiday (~2021.09.22.)
2021.09.25	The highest number of confirmed cases (3273 cases)

5.2 추정된 숨은 감염자 수 및 신규 감염자 수

대한민국 확진자 수, 격리 중인 수, 격리 해제 수, 사망자 수 등을 적합하여 상태 값을 추정된 결과는 그림 2(a) ~ 그림 2(c)와 같다. 제안한 모델을 통해 추정된 값은 점선으로, 실제 데이터는 실선으로 표현하였다. 표3의 정책들을 바탕으로 구간을 설정하였고, 그

1) 2021년 10월 7일 이후 대한민국 COVID-19데이터는 확진자 수의 급증으로 인하여 격리 해제 인구 수, 격리 중인 인구 수 등의 구획을 분리하지 못한 채 일일 신규 확진자 수를 중심으로 공개되고 있다. 따라서 본 논문에서는 공개 데이터 내 구획 구분이 잘 되어있는 2021년 10월 이전의 데이터를 이용하여 모델을 검증하였다.

림 2(b)와 그림 2(c)에서는 배경색으로 동일한 z 의 설정 구간을 표현하였다.

그림 2(a)는 과거 대한민국의 전체적인 감염 양상을 보여준다. **S** 상태는 시간이 지날수록 감소하는 양상을 보이고 있다. 이는 **S** 상태에 머무르는 인구는 결국 **R** 상태와 **D**상태로 도달했기 때문이다. **V**상태 값은 백신 공급 현황, 백신 접종 현황에 따라 증감한다. **E, I** 상태 값은 감염병의 전파 양상에 따라 증감한다. 흡수 상태인 **R** 상태와 **D** 상태는 증가하는 양상을 보인다. 또한, 동일한 흡수 상태이더라도, **R** 상태가 **D** 상태보다 가파르게 증가하고 있음을 확인할 수 있다.

그림 2(b)는 X_{SE} 와 **I** 상태 값, 일일 확진자 수(Daily confirmed)를 비교하였다. **I** 상태의 양상은 일일 확진자 수의 양상과 비슷함을 확인할 수 있다. 즉, 숨은 감염자 수의 양상은 일일 확진자의 양상을 따르고 있을 확인할 수 있다. X_{SE} 는 **S** 상태에서 **E** 상태로 전이되는 인구의 수로, 이후 **I** 상태로 들어가 잠재적인 감염자가 될 수 있기 때문에 중요한 의미가 있다.

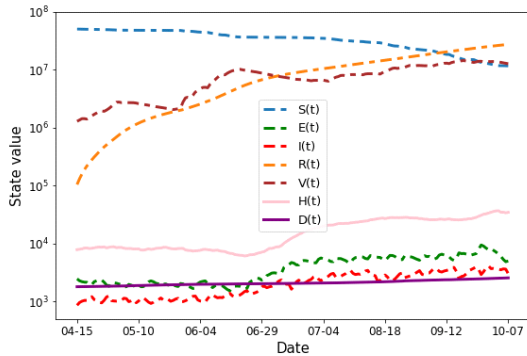


그림 2. (a) 제안한 모델의 추정한 상태값
Fig. 2. (a) The state value of proposed model

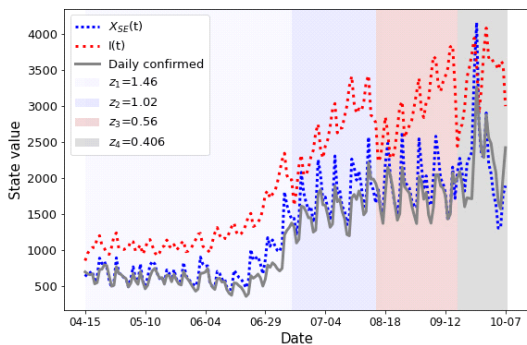


그림 2. (b) X_{SE} , **I** 상태 값과 일일 확진자 수 비교
Fig. 2. (b) The comparison of X_{SE} , state value of **I** and daily confirmed cases

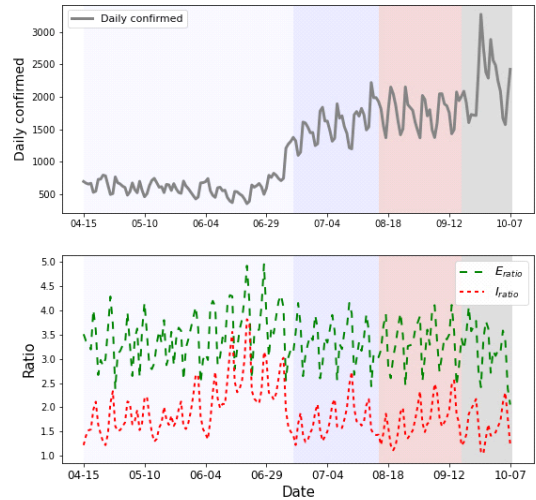


그림 2. (c) 일일 확진자 수와 확진자 수 대비 **E, I** 상태 값의 비율
Fig. 2. (c) The daily confirmed cases and the ratio of the daily confirmed cases to the state value of **E, I**

따라서, X_{SE} 가 일일 확진자 수와 비슷한 패턴을 갖는다면, 이후 X_{SE} 의 흐름을 기반으로 확진자 수 예측에도 도움이 될 수 있다. 6월 14일부터 7월 4일은 X_{SE} 가 먼저 증가한 후 일일 확진자 수가 증가하는 양상을 확인할 수 있다.

그림 2(c)는 미시적 관점에서의 **E, I** 상태의 양상을 보여준다. 전체 기간의 **E**상태의 인구수는 평균적으로 일일 확진자 수의 약 3.38배, **I** 상태의 인구수는 확진자 수의 약 1.86배 정도로 추정되었다. 특히, z 중 6월 14일부터 7월 6일까지 **E** 상태의 인구수는 평균적으로 일일 확진자 수의 3.8배, **I** 상태의 인구수는 일일 확진자 수의 2.65배까지 증가하는 것을 확인할 수 있다. 이는 확진자 수와 숨은 감염자 수 간의 차이가 큰 시기로, 많은 수의 숨은 감염자들이 감염병의 확산이 가속화하고 있음을 의미한다. 정부는 7월 7일 4차 대유행을 선언하였고, 이후 z 구간에서는 이전 구간의 급격한 확산의 영향으로 확진자 수는 평균적으로 1500명대의 높은 수를 유지하고 있음을 확인할 수 있다.

5.3 제안 모델 기반 추정 파라미터를 활용한 결과 분석

2021년 4월 15일부터 2021년 10월 7일 사이의 데이터를 이용하여 제안 모델에 기반해 추정된 파라미터들을 분석한 결과는 그림 3(a) ~ 그림 3(c)와 같다. 추정된 파라미터는 실선으로 표현하였다.

그림 3(a)는 추정된 **I** 상태의 수와 **H** 상태의 수를 바탕으로 일일 확진 비율인 $\varphi(t)$ 를 추정된 결과이다.

그래프 관찰 결과, z_1 구간 중 6월 14일부터 7월 6일까지의 추정된 평균 확진 비율은 40%로 다른 구간보다 낮았다. 추정 확진 비율이 80%보다 큰 경우에는 7월 11일 이후, 8월 15일 이후, 9월 17일 이후이다. 공휴일이 끝난 직후, 대유행 선언 이후에는 높은 검사율을 보이고 있다고 해석할 수 있다.

그림 3(b)는 **H** 상태의 수와 **D** 상태에 기반하여 병원에 입원한 이후의 일일 사망률인 $\delta(t)$ 를 추정한 결과이다. 그림 2(c)를 통해 z_2 구간 이후, 확진자의 수는 증가하는 것을 확인하였다. 하지만, 사망률을 추정해 본 결과, z_1 구간의 사망률이 0.3%로 다른 구간들보다 높은 것으로 추정되었다. 즉, z_2 구간 이후, **H** 상태에서 **R** 상태로 이동하는 인구 비율이 z_1 구간보다 많아지고 있음을 알 수 있다.

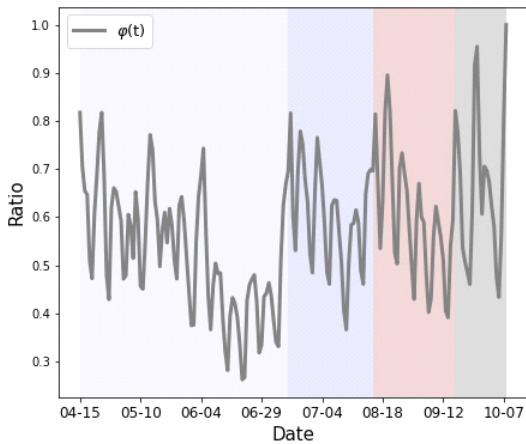


그림 3. (a) $\varphi(t)$ 값 추정
Fig. 3. (a) The estimation of $\varphi(t)$

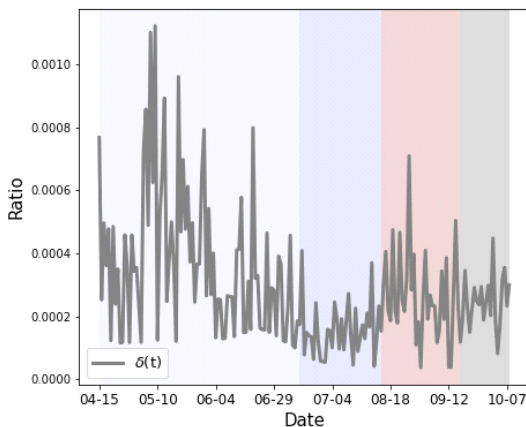


그림 3. (b) $\delta(t)$ 값 추정
Fig. 3. (b) The estimation of $\delta(t)$

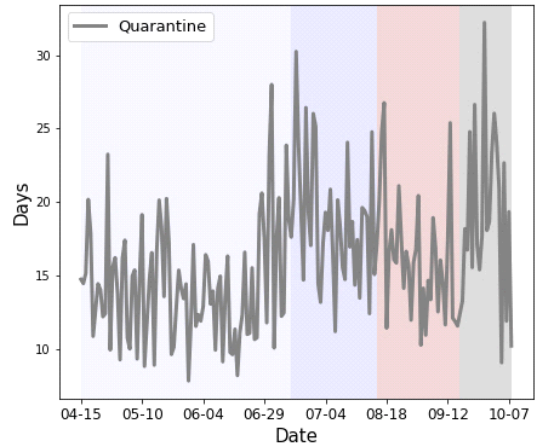


그림 3. (c) 완치까지의 격리 기간 추정
Fig. 3. (c) The estimation of quarantine period

그림 3(c)는 **H** 상태의 수, 추정한 **R** 상태의 수, 그리고 앞서 추정한 $\delta(t)$ 를 바탕으로 병원에 입원한 이후 완치될 때까지의 격리 기간인 $1/\rho(t)$ 를 추정한 결과이다. 입원 환자의 평균 격리 기간은 구간별로 달랐다. z_2 구간과 z_4 구간에서 완치 기간은 평균 17일로 추정되었다. 이에 반해, z_1 구간의 경우에는 평균 13일, z_3 구간의 경우에는 평균 15일로 z_2 구간, z_4 구간은 다른 구간보다 긴 완치 기간을 보였다. 그림 3(b)를 종합해서 볼 때 z_1 구간 이후 격리 기간이 늘어나지만, 사망률은 감소함을 알 수 있다.

5.4 결과 비교

제안한 모델의 우수성을 입증하기 위하여 숨은 감염자 수 추정 결과를 최신 구획 모델, 덤퍼닝기반 모델과 비교하였다²⁾. 본 논문에서 사용된 비교 모델은 다음과 같다.

- SIDARTHE model^[12]: 병의 진행 과정을 8가지의 구획으로 나눠 숨은 감염 상태를 설명한 최신 모델이다. 확진자, 격리자, 회복자, 사망자, ICU 환자수 등의 데이터를 사용하여 미분방정식을 기반으로 숨은 감염자 수를 추정하는 모델이다.
- Deep LSTM model^[27]: 과거 확진자, 사망자, 이동량 데이터를 사용하여 미래 시점에서의 확진자, 사망자, 이동량을 예측하는 LSTM기반의 예측모델이

2) 제안한 모델은 숨은 감염자 수를 추정하는 모델로, 추정 정확도 평가를 위해서는 실제 숨은 감염자 수 데이터가 필요하다. 그러나 이는 공식적으로 집계될 수 없는 데이터이기에 정확도 평가는 불가능하다. 대신, 제안 모델의 우수성 입증 위해 기존 연구와 추정 값 비교를 진행하였다.

다. 다변량 LSTM 모델로, 2개의 LSTM 층과 드롭아웃 층, FC 층으로 이루어져 있다.

그림 4는 비교 모델들과 제안한 모델의 I 상태 값을 비교한 결과이다. 제안한 모델의 I 상태 값은 수식에 의해 명확한 점으로 계산되기 때문에 직관적인 숨은 감염자 수를 제공한다는 장점이 있다. 또한, 그림 4를 통해 제안한 모델의 I 상태 값의 변화 양상은 일일 확진자 수의 변화 양상과 유사한 양상을 보임을 확인할 수 있다. 즉, 제안한 모델의 I 상태 값은 확진자 급증 구간(6월 4일부터 6월 29일까지의 구간)에서도 급격한 변화의 흐름을 반영할 수 있다.

반면, SIDARTHE 모델의 I 상태 값은 확진자 수의 변화를 즉각적으로 따라가지 못함을 관찰할 수 있다. 이는 SIDARTHE 모델이 거시적인 관점에서 감염병 진행 과정을 모델링하였기 때문이다. 그 결과, SIDARTHE 모델은 일일 확진자 수 증감과 같은 미시적인 변화를 포착하지 못하고 큰 흐름의 단순한 곡선 형태로 표현된다. 게다가, 특정 구간에서의 SIDARTHE 모델의 I 상태 값은 확진자 수 대비 6배까지 과대추정되는 경향이 있음을 확인하였다.

Deep LSTM 모델은 숨은 감염자 수 추정이 불가능한 모델이다. 딥러닝 모델 학습에는 추론 결과에 대한 평가를 위해 실제 데이터가 요구되기 때문이다. 숨은 감염자에 관한 실제 데이터는 존재하지 않으므로 숨은 감염자 수는 확진자 수의 2~3배라는 [28]의 연구 결과를 통해, Deep LSTM 모델의 숨은 감염자 수를 설정하였다. 확진자 수의 3배는 검정색 점선, 확진자 수의 2배는 파랑색 점선을 의미하며, 해당 범위는 음영으로 표기하였다. Deep LSTM 모델 추정 결과, 초반

에는 학습을 위한 데이터가 충분하지 않기 때문에 5월 10일 전후로 추정된 숨은 감염자 수는 확진자 수 증가 폭보다 크게 관찰된다. 또한, Deep LSTM 모델은 과거 데이터의 경향성에 기반하여 미래를 예측하는데, 6월 4일 전에는 증가 현상이 존재하지 않아 6월 24일부터 6월 29일까지의 증가 현상에 대해서는 예측을 하지 못하였다. 이에 따라 Deep LSTM 모델이 추정된 숨은 감염자 수는 확진자 수 급증의 흐름을 반영하지 못함을 확인할 수 있다.

VI. 결론

본 연구는 국내의 방역 정책에 따라 기존의 구획 모델을 7가지 상태로 확장하였다. 제안한 모델에 실제 대한민국 COVID-19 데이터를 적용하여 공식적으로 집계될 수 없는 숨은 감염자의 수를 추정하였고, 모델의 타당성을 검증하기 위해서 최신 모델들과 비교를 진행하였다. 또한, 매일 변화하는 전파 양상을 $\beta(t)$, $\delta(t)$, $\rho(t)$ 와 같은 시변성을 갖는 파라미터를 이용하여 정량적으로 분석할 수 있다. 제안한 감염병 모델은 각 상태에 해당하는 인구수를 추정 가능하다는 점에서 전염병의 유행 양상을 정량적으로 설명 가능하다는 장점이 있다. 제안한 모델은 앞으로 새로 나올 신종 감염병에도 적용 가능하고, 한국뿐만 아니라 다른 나라의 정책들을 고려하여 변형이 가능하다는 점에서 유연성이 크다.

References

- [1] P. Block, M. Hoffman, I. J. Raabe, J. B. Dowd, C. Rahal, R. Kashyap, and M. C. Mills, "Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world," *Nature Human Behaviour*, vol. 4, no. 6, pp. 588-596, Jun. 2020. (<https://doi.org/10.1038/s41562-020-0898-6>)
- [2] K. Kim, "Sociological study of the covid-19 modelling: Why the mathematical model failed to implement infection control during the early stage of the covid-19 pandemic in britain?," *Korean Soc. for Social Theory*, pp. 263-302, 2020. (<https://doi.org/10.17209/st.2020.11.37.263>)

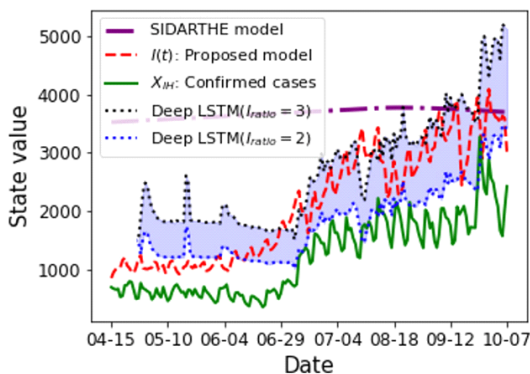


그림 4. SIDARTHE 모델 및 Deep LSTM 모델의 I 상태 값 추정 결과 비교
Fig. 4. Comparison of estimation results with SIDARTHE model and Deep LSTM model

- [3] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, Springer International Publishing, 2017. (<https://doi.org/10.1007/978-3-319-52452-8>)
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT press, 2016.
- [5] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, Mar. 2020. (<https://doi.org/10.1016/j.physd.2019.132306>)
- [6] A. L. Barabási, *Network science*, Cambridge University Press, 2019.
- [7] F. Brauer, *Mathematical Models in Epidemiology*, Springer, 2019.
- [8] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," *The Lancet*, vol. 395, no. 10225, pp. 689-697, Feb. 2020. ([https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9))
- [9] S. Feng, Z. Feng, C. Ling, C. Chang, and Z. Feng, "Prediction of the COVID-19 epidemic trends based on SEIR and AI models," *PLOS ONE*, vol. 16, Jan. 2021. (<https://doi.org/10.1371/journal.pone.0245101>)
- [10] S. Choi and M. Ki, "Estimating the reproductive number and the outbreak size of COVID-19 in Korea," *Epidemiology and Health*, vol. 42, Mar. 2020. (<https://doi.org/10.4178/epih.e2020011>)
- [11] Z. Xu, B. Wu, and U. Topcu, "Control strategies for COVID-19 epidemic with vaccination, shield immunity and quarantine: A metric temporal logic approach," *PLOS ONE*, vol. 16, no. 3, Mar. 2021. (<https://doi.org/10.1371/journal.pone.0247660>)
- [12] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy," *Nature Med.*, vol. 26, no. 6, pp. 855-860, Apr. 2020. (<https://doi.org/10.1038/s41591-020-0883-7>)
- [13] C. Twumasi, L. Asiedu, and E. N. N. Nortey, "Markov chain modeling of HIV, tuberculosis, and hepatitis B transmission in Ghana," *Interdisciplinary Perspectives on Infectious Diseases*, pp. 1-8, Nov. 2019. (<https://doi.org/10.1155/2019/9362492>)
- [14] A. Raheeriririna, T. S. Fandresena, A. R. Hajalalaina, H. Rabetafika, R. A. Rakotoarivelo, and F. Rafamatanantsoa, "Probabilistic modelling of covid-19 dynamic in the context of madagascar," *Open J. Modelling and Simulation*, vol. 9, no. 03, pp. 211-230, 2021. (<https://doi.org/10.4236/ojmsi.2021.93014>)
- [15] M. Jahangiri, M. Jahangiri, and M. Najafgholipour, "The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (COVID-19) in different provinces of Iran," *Sci. The Total Environ.*, vol. 728, Aug. 2020. (<https://doi.org/10.1016/j.scitotenv.2020.138872>)
- [16] S. Ahn and M. Kwon, "SI model-based analysis of social distancing strategies on covid-19 spread in South Korea," in *Proc. 31st JCCI*, Apr. 2021.
- [17] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, "Mobility network models of COVID-19 explain inequities and inform reopening," *Nature*, vol. 589, no. 7840, pp. 82-87, Nov. 2020. (<https://doi.org/10.1038/s41586-020-2923-3>)
- [18] J. Sun, X. Chen, Z. Zhang, S. Lai, B. Zhao, H. Liu, S. Wang, W. Huan, R. Zhao, M. T. A. Ng, and Y. Zheng, "Forecasting the long-term trend of COVID-19 epidemic using a dynamic model," *Scientific Reports*, vol. 10, no. 1, Dec. 2020. (<https://doi.org/10.1038/s41598-020-78084-w>)
- [19] H. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, 3rd Ed., Academic Press, 1998.
- [20] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, "A new framework and software

to estimate time-varying reproduction numbers during epidemics,” *Am. J. Epidemiology*, vol. 178, no. 9, pp. 1505-1512, Sep. 2013. (<https://doi.org/10.1093/aje/kwt133>)

- [21] S. Ma, J. Zhang, M. Zeng, Q. Yun, W. Guo, Y. Zheng, S. Zhao, M.H. Wang, and Z. Yang, “Epidemiological parameters of covid-19: Case series study,” *J. Med. Internet Res.*, vol. 22, no. 10, Oct. 2020. (<https://doi.org/10.2196/19994>)
- [22] X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, and G. M. Leung, “Temporal dynamics in viral shedding and transmissibility of COVID-19,” *Nature Med.*, vol. 26, no. 5, pp. 672-675, Apr. 2020. (<https://doi.org/10.1038/s41591-020-0869-5>)
- [23] J. Arzt, M. A. Branan, A. H. Delgado, S. Yadav, K. I. Moreno-Torres, M. J. Tildesley, and C. Stenfeldt, “Quantitative impacts of incubation phase transmission of foot-and-mouth disease virus,” *Scientific Reports*, vol. 9, no. 1, Feb. 2019. (<https://doi.org/10.1038/s41598-019-39029-0>)
- [24] <https://www.paho.org/coronavirus>
- [25] <https://www.kdata.or.kr>
- [26] <https://ncov.mohw.go.kr>
- [27] B. Nikparvar, M. Rahman, F. Hatami, and J. C. Thill, “Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network,” *Scientific reports*, vol. 11, no. 1, pp. 1-12, 2021. (<https://doi.org/10.1038/s41598-021-01119-3>)
- [28] C. Zhan, K. T. Chi, Y. Gao, and T.g Hao, “Comparative study of COVID-19 Pandemic progressions in 175 regions in Australia, Canada, Italy, Japan, Spain, UK and USA using a novel model that considers testing capacity and deficiency in confirming infected cases,” *IEEE J. Biomed. and Health Informatics*, vol. 25, no. 8, pp. 2836-2847,

2021.

(<https://doi.org/10.1109/JBHI.2021.3089577>)

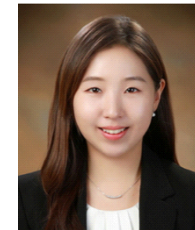
안 수 진 (Sujin Ahn)



2021년 2월: 숭실대학교 정보통계·보험수리학과 학사
2021년 3월~현재: 숭실대학교 정보통신공학과 석사 과정
<관심분야> 네트워크과학, 수학적 모델링, 인공지능, 빅데이터 분석

[ORCID:0000-0003-3805-711X]

권 민 혜 (Minhae Kwon)



2011년 8월: 이화여자대학교 전자정보통신공학과 학사
2013년 8월: 이화여자대학교 전자공학과 석사
2017년 8월: 이화여자대학교 전자전기공학과 박사
2017년 9월~2018년 8월: 이화여자대학교 전자전기공학과 박사 후 연구원
2018년 9월~2020년 2월: 미국 Rice University, Electrical and Computer Engineering, Postdoctoral Researcher
2018년 9월~2020년 2월: 미국 Baylor College of Medicine, Center for Neuroscience and Artificial Intelligence, Postdoctoral Researcher
2020년 3월~현재: 숭실대학교 전자정보공학부 IT융합 전공 조교수
<관심분야> 네트워크과학, 수학적 모델링, 인공지능, 강화학습, 이상탐지기술, 자율주행, 모바일네트워크
[ORCID:0000-0002-8807-3719]