

코로나 확진자 수 예측을 위한 BI-LSTM과 GRU 알고리즘의 성능 비교 분석

김재호¹ · 김장영^{2*}

Comparative analysis of performance of BI-LSTM and GRU algorithm for predicting the number of Covid-19 confirmed cases

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

위드 코로나의 예정 발표일이 결정되었고, 위드 코로나에 가장 중요한 조건인 백신 접종을 아직 부작용 걱정 때문에 완료하지 않은 사람들이 있다. 또한 위드 코로나로 경제는 회복될 수 있지만 감염자 수는 급증할 수 있다. 본 논문은 위드 코로나에 앞서 코로나19에 대한 경각심을 깨우고자, 코로나19를 비선형 확률과정으로 예측한다. 여기서 딥러닝의 RNN중 양방향 LSTM인 BI-LSTM와 LSTM보다 gate수를 줄인 GRU를 사용하고 이것을 train set, test set, 손실함수, 잔차분석, 정규분포, 자기 상관을 통해서 비교 분석하여 어떠한 성능이 더 좋은지 비교하고 예측한다.

ABSTRACT

Even the announcing date for the starting date of "With Corona" has been decided, still many people have not completed vaccination, the most important condition for starting the With Corona, because of concerns for its side effects. In addition, although the economy may can be recovered by the With Corona, but the number of infected people may can be surged. In this paper, in order to awaken the people for the awareness of Corona 19 in advance of the With Corona, the Corona 19 is predicted through a non-linear probability process. Here, among the deep learning RNN, BI-LSTM, which is a bidirectional LSTM, and GRU, gates decreased than LSTM have been used. And this has been compared and analyzed through train set, test set, loss function, residual analysis, normal distribution, and autocorrelation, and compared and predicted for which has a better performance.

키워드 : 코로나19, RNN, LSTM, BI-LSTM, GRU

Keywords : Covid-19, RNN, LSTM, BI-LSTM, GRU

Received 4 November 2021, Revised 11 November 2021, Accepted 18 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.2022.26.2.187>

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

1.1. 개발동기

계속해서 코로나19 바이러스는 계속해서 일일 감염자 수는 증가하고 있고, 계속되는 변이 바이러스 발생으로 인해 사태가 점점 심각해지고, 해외유입 감염으로 인해 우리나라에도 변이 바이러스가 퍼지고 있다. 변이 바이러스에 따라서 치사율과 전염성은 모두 다르고, 신종 코로나 바이러스 같은 경우에는 정보가 상대적으로 적기 때문에 더 조심해야한다. 또한 정부는 위드코로나도 염두해 두고 있기 때문에 감염자 수는 더욱 증가할 것이라고 예측된다. 본 논문은 일일 감염자수를 예측함으로써 코로나19의 경각심을 강조한다. 또한 여태까지 선형 확률과정으로 예측을 하였지만, 비선형 확률과정중 LSTM과 BI-LSTM모형을 통해 예측하고, 비교한다. 이는 적분 선형과정보다 더 좋은 예측 결과값을 가지고 급증하거나 급감하는 패턴에서 장점을 보인다.

1.2. 백신의 효과

현재 사용하는 모든 COVID-19 백신은 임상 시험에서 확인되었듯이 COVID-19 예방에 효과적이다. mRNA COVID-19 백신이 실제 환경에서 유사한 보호 기능을 제공한다는 연구 증거가 늘어나고 있다. 백신 접종은 코로나의 팬데믹 종식에 도움이 되는 중요한 수단중 하나다. 백신 접종은 코로나19 감염이나 중증을 방지하고 전염 방지에 도움이 된다. 100% 효과적인 백신은 없기 때문에 백신 접종을 완료했어도 감염되는 사람이 있을 수 있다. 백신 접종은 코로나19에 감염되지 않도록 면역력을 갖추는 데 도움이 된다. 신체가 코로나에 대한 면역력을 구축하고 있다는 징후로써 부작용이 나타날 수도 있다. 부작용은 일상 활동이나 생활에 영향을 미치기도 하지만, 며칠 내에 사라진다. 일반적인 부작용으로 주사 맞은 팔에는 통증, 조홍, 부어오름이 있고, 몸 전체에는 두통, 피로감, 근육통, 발열, 메스꺼움, 오한 등이 나타날 수 있다. mRNA의 코로나19 백신 1차 접종 후 심각한 반응이나 또는 즉각적인 알레르기 반응이 나타났다면, mRNA 백신 중 어떠한 것이라도 2차 접종을 받지 않아야 한다. 대부분의 열이나 통증으로 인해서 불편함을 겪는다면 신체가 면역력을 구축하고 있다는 정상적인 신호다. 부작용을 겪지 않는 사람들도 존재한다. 무수한 사람들이 코로나 백신을 접종 받았으며, 장기적인 부작용은 발견되지 않았다.

백신의 안전성을 계속해서 모니터링하고 있고, 안전성 문제와 백신 사이의 연관성을 발견되면 식품의약품과 백신 제조업체는 안전성 문제 해결을 위한 해결책을 모색하고 있다[1],[2].

코로나19 확진자 수 예측 논문 사례로는 FB Prophet 알고리즘을 이용한 예측이있다. 이 논문에서는 결정계수와 평균 절대 오차, 평균 백분율 오차로 논문의 설명력을 더한다[3].

II. 분석 알고리즘

LSTM 알고리즘에는 다양한 변칙 패턴들이 존재한다. 예를 들어 Depth Gated RNNs나 완전히 다른 방식으로 장기적 의존성 문제를 해결한 Clockwork RNNs도 있다. 어떠한 패턴이 최고인지 판단하기는 쉽지않다. 각각 비슷하다는 하긴 하지만, 차이점들과 어떤 의미를 가지는지 따져봐야한다. Greff, et al은 RNN의 변칙패턴들에 대해 비교분석했고, Jozefowicz, et al은 대략 만 가지 이상의 RNN의 아키텍처를 실험했는데, 몇몇 RNN 변칙 패턴들은 LSTM보다 성능이 우수한 경우가 있다는 것을 발견했다[4]. 따라서 어떠한 알고리즘이 정답이라고는 정할 수 없다. 상황과 데이터에 맞게 알고리즘을 적용하며 최고의 결과를 도출해야 한다. 그에따라 본 논문은 BI-LSTM과 GRU를 비교하고, 예측한다.

2.1. LSTM(Long Short-Term Memory)

RNN알고리즘의 단기 기억 문제를 해결책으로 고안된 LSTM알고리즘은 원인과 결과의 관계를 장시간 확장하고 반영한다. 일정한 오류를 유지함으로써 여러 네트워크가 장시간에 걸쳐 학습을 가능하도록 하며, cell의 개념을 도입하여 열리고 닫히는 데이트를 통해 어떤 데이터를 저장/읽기/쓰기/삭제 할지 결정한다. 기존 RNN에 Cell State를 추가하여 얼마나 과거의 데이터를 결정할지 제안한다. 이 구조 덕분에 정보는 큰 변함없이 다음 단계에 전달하게 된다. LSTM은 cell state에 정제된 구조를 가진 gate 요소를 활용하여 정보를 더하거나 제거하는 기능을 포함하고 있다. 장기기억셀(C)은 더하기와 곱하기 노드에만 전달되기에 기울기 변화(감소)는 발생하지 않는다. 곱하기 노드에서는 행렬곱이 아닌 원소별 곱을 하고 매 시각 다른 게이트 값을 이용하기에

누적되지 않는다[5].

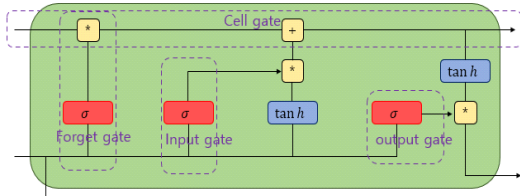


Fig. 1 LSTM

위 그림1은 LSTM의 구조이다.

2.2. BI-LSTM (BI - Long Short-Term Memory)

Manning은 task가 무엇이든지 간에 attention이 적용된 BiLSTM 모델은 자연어 처리 분야에서 우수한 성능을 낸다고 한다. Manning이 소개한 stanford NLP group의 최고 연구성과는 BiLSTM이 적용된 모델이었다. RNN 구조는 gradient vanishing문제에 취약하다. h1의 gradient를 구하려면 체인룰에 의해 해당하는 그래디언트를 계속 곱해줘야 하기 때문이다. LSTM은 cell state를 도입해 그래디언트 문제를 해결한다. LSTM은 직전 시점 데이터와 현 시점 데이터를 더해줌으로써 그래디언트가 효과적으로 전달 될 수 있다[6].

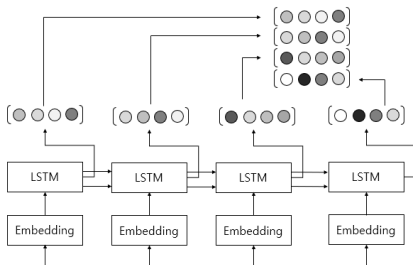


Fig. 2 BI-LSTM

위 그림2는 BI-LSTM의 구조이다. RNN이나 LSTM은 입력한 순서를 시간 순서대로 입력하기 때문에 결과물이 직전 패턴을 기반으로 수렴하는 경향을 있다는 한계가 보인다. 이러한 단점을 해결하는 방법으로 양방향 순환신경망(Bi-RNN)이 제안되었다. Bi-RNN은 기존 RNN의 순방향에 역방향을 은닉층에 추가하여 성능을 향상시켰다. 그러나 데이터 길이가 길고 layer가 깊으면, 과거 데이터가 손실되는 단점이 있다. 이를 극복하기 위해서 제안된 알고리즘이 양방향 LSTM이다. 양방향

LSTM(BI-LSTM)의 핵심 개념은 LSTM을 첫 번째 인스턴스부터 시작하여 정방향 모드로만 실행하는 대신 뒤에서 앞으로 실행되는 마지막 모드에서 다른 하나를 시작한다. BI-LSTM은 정보를 역방향으로 전달하는 히든 레이어를 추가하여 이러한 정보를 보다 유연하게 처리한다. 이는 각 시점에서 은닉 상태가 이전 시점과 미래 시점의 정보를 모두 갖는 효과가 있다. BI-LSTM계층은 시계열데이터 또는 시퀀스 데이터의 시간 간격 간의 양방향 장기적 종속성을 학습한다. 이러한 장기적 종속성은 신경망이 각 시간 간격에서 전체 시계열로부터 학습할 수 있도록 하려는 경우 유용할 수 있다. LSTM의 각각 시각의 은닉 상태의 벡터는 h_s 로 모아진다. LSTM의 시각별 은닉 상태 벡터를 모은 h_s 의 각 행에는 그 행에 대응하는 정보들을 많이 포함하고 있다. 그러나 여기에서 짚어야 할 점은 정보를 왼쪽에서 오른쪽으로 읽는다는 점이다. 그렇기에 대응하는 벡터에는 총 세가지 정보가 인코딩 된다. 정보가 많아질 수록 한 벡터에 포함하고 있는 정보는 많아질 것이다. 대응하는 정보의 주변 정보를 균형 있게 담기 위한 방법이 BI-LSTM 즉, 양방향 LSTM이다. BI-LSTM은 현재까지의 LSTM 계층에 역방향으로 처리하는 또다른 LSTM 계층을 추가하여, 최종 은닉 상태에서는 이 두개의 LSTM 계층의 은닉 상태를 연결한 벡터를 출력한다. 연결 이외에도 더하거나 평균을 내는 방법 등 다양하게 적용할 수 있다[7].

2.3. GRU (Gated Recurrent Unit)

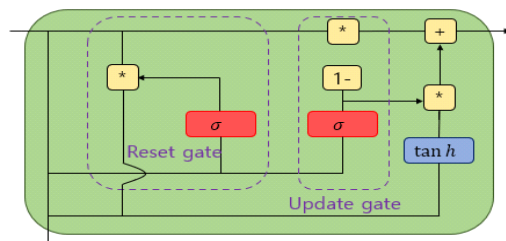


Fig.3 GRU

위 그림3은 GRU 구조이다. GRU 알고리즘 역시RNN의 단기 기억문제 해결책으로 고안되었다. Gate 구조가 적용된 RNN 일종으로 LSTM보다 학습에 사용될 게이트와 파라미터가 적은 구조이므로 보다 간결함에도 불구하고 빠른 속도와 유사한 성능을 가진다. 가중치를 줄여 컴퓨팅 시간은 줄이고 정보의 양은 과거의 정보를 잘

조합해서 멀리까지 내보낼 때 많은 정보를 가진다. LSTM은 Gate가 3개지만 GRU는 2개이며, Reset Gate & Update Gate(:LSTM의 Forget+Input과 유사)이다. output으로 나오는 hidden unit(정보)는 outputgate로 나오지 않아 정보량이 훨씬 많다. 즉 Output Gate가 없는 LSTM이기에 메모리셀에 담기는 정보 양 증가한다. LSTM의 C_t (장기기억 셀)와 H_t 가 하나의 벡터 H_t 로 통합한다. GRU가 LSTM보다 학습할 가중치가 적어지는 것이 이점이다. 주제별로 LSTM과 GRU의 성능은 차이가 있다. 무조건 LSTM대신 GRU를 사용하는 것은 좋지 않다. LSTM과 GRU중 어떤 것이 모델 성능이 낫다고 확신할 수 없고, 기존 LSTM을 사용하면서 최적의 파라미터를 찾아낸 상황이라면, 굳이 GRU로 바뀌서 적용할 필요는 없다. 데이터 크기가 작을 때는, 파라미터의 수가 적은 GRU가 조금 더 낫고, 데이터 양이 많으면 LSTM이 좋다고 알려져 있다. 하지만 직접 사용해보고 비교해보는게 더 좋은 결과를 얻을 수 있다. GRU보다 LSTM에 대한 연구 내용이나 사용량이 상대적으로 더 많은데, LSTM이 GRU보다 먼저 나온 구조이기 때문이다[8][9].

III. 실험결과 및 분석

Train data set과 test dataset을 8:2비율로 지정하였고, 딥러닝에서 중요한 파라미터 설정은 모델 내부 파라미터를 업데이트 하기 전 샘플의 개수인 batch_size는 32로 설정했으며, 각각의 epoch와 뉴런의 layer와 개수를 달리하여 실험하였다. 본 연구에서는 뉴런의 layer는 5개로, dense를 128, 256, 138, 64, 1개로 각각 지정하였다. sequence는 7, drop out값은 0이다. 활성화함수로는 relu알고리즘을, 모델 compile 과정에서 손실함수의 값을 낮추는 optimizer 함수는 adam알고리즘을, 오차평가 측정은 MSE로 설정하였다. 또한 train set과 test set의 검증 지표로는 MAE와 RMSE를 사용했다.

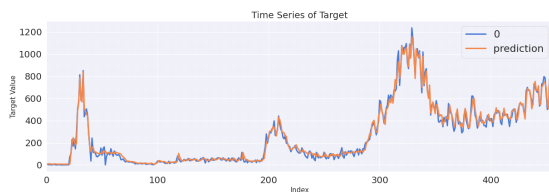


Fig. 4 GRU train set

위 그림4는 train set이 얼마나 훈련을 잘 했는지 보여주는 그래프다. 여기서 train의 MAE값은 29.3으로 잘 훈련된 모습을 보여준다.



Fig. 5 GRU test set

위 그림5는 test의 정확성을 보여주는 그래프다. 실제 값을 따라가는 그래프를 보여준다. GRU 알고리즘의 Test set의 MAE는 215.29, RMSE값은 301.06이다. 반면, BI-LSTM의 MAE값은 350.08, RMSE값은 450.91이 나왔다. Test set의 검증 결과 GRU가 더욱 더 좋다는 결과가 나왔다. 따라서 본 연구에서는 GRU를 중점으로 하여 예측한다.

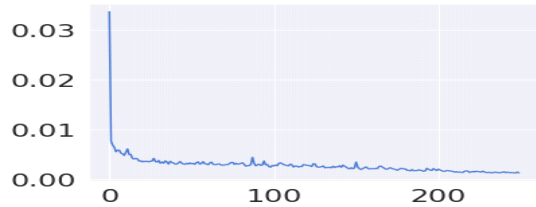


Fig. 6 GRU Loss Function

위 그림6은 각각 epoch에 따라 손실함수 값을 보여주는 값이다. epoch란 훈련데이터의 알고리즘 수행 횟수이며, 손실함수란 train을 통해 생성된 모델의 예측 정도와 실제값의 차이를 나타내는 함수다. 즉 0에 가까울수록 모델의 정확도가 더 높다고 할 수 있다. 신경망 학습에서 최적의 매개변수(가중치, 편향등) 값을 선택할 때 손실함수의 값을 가능한 최대한 작게 하는 파라미터 값을 선택한다.

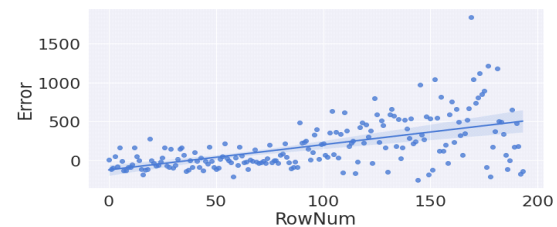


Fig. 7 GRU Residual

위 그림7은 GRU의 Test set의 잔차값을 보여주는 데이터다.

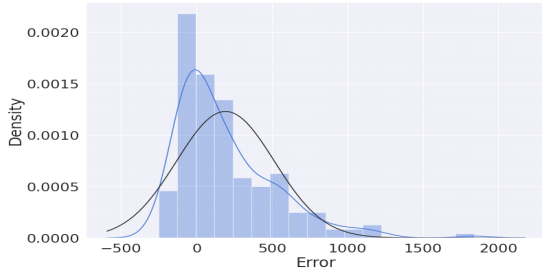


Fig. 8 GRU Normal Distribution

위 그림8은 train과 test의 잔차에 대한 전체적인 분포를 보여준다. 그림과 같이 0에 몰릴수록 예측이 잘 됐다고 할 수 있다.

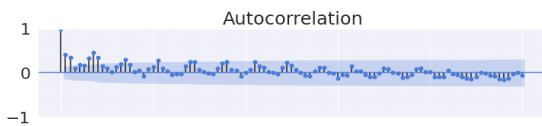


Fig. 9 GRU Autocorrelation

위 그림9는 GRU 알고리즘의 자기상관 그래프이다. 데이터에 자기 회귀 향이 보인다. 시간이 지날수록 정상성을 가진 데이터를 가진다.

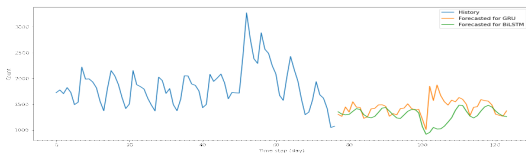


Fig. 10 Comparison of GRU ans LSTM

위 그림10은 GRU와 BI-LSTM의 단기적인 예측값 비교이다. 흥미로운 점은, 비슷한 패턴을 가지면서도, GRU와 같은 경우 예측 그래프를 보면 전에 있던 대유행을 다시 고려하는 패턴을 가지고 예측을 했고, LSTM 같은 경우는 점차 감소하고 있고, 과거에 대유행을 고려하지않는 패턴을 예시하고 있다는 것이다. 이것은 파라미터를 다르게 했을 경우에도 비슷한 그래프를 보인다.

Table. 1 Pratical infectee & Prediction infectee

Date	COUNT	BI-LSTM	GRU
2021-10-12	1579	1369	1585
2021-10-13	1937	1587	1636

Date	COUNT	BI-LSTM	GRU
2021-10-14	1682	1506	1681
2021-10-15	1617	1443	1587
2021-10-16	1418	1330	1353
2021-10-17	1049	1106	1230
2021-10-18	1073	931	1236
2021-10-19	1570	1353	1584

위 표1은 실제 코로나 감염자 수와 BI-LSTM과 GRU의 예측값이다. 결과는 10월 13일과같이 급증하거나, 17,18일과 같이 급감하지 않는 이상 준수한 예측값을 가진다. 대유행과 마찬가지로 현재 백신으로 인해 감염자 수가 줄어드는 패턴이 보이면서 예측값은 많이 벗어난 걸 볼 수 있다. 하지만 데이터가 많이 쌓인다면, 다시 좋은 예측값을 가질 수 있다. 또한 SARIMA와 이전 연구했던 비교했을 경우 주말이나 공휴일에 검사횟수가 적어진 점을 고려해 데이터에 Seasonality이 보였는데, 딥러닝을 이용한 경우 계절성 파라미터를 따로 설정해주지 않았기 때문에, Seasonality를 덜 고려하는 것으로 보인다.

IV. 결론 및 향후연구

코로나 감염자 수 데이터는 Kaggle에서 John Hopkins 대학 COVID-19 데이터셋에서 Raw_global데이터를 사용했다.[10] 코로나 19예측 수 예측 방법에서 직분선형 확률과정보다 딥러닝의 비선형 확률과정이 더 좋은 예측 성능을 가졌다. 그에따라 본 논문은 비선형 확률과정을 통해 코로나 확진자 수를 예측했다. 비선형 확률과정은 높은 예측률과 급증하거나 급감하는 패턴에서 장점을 보인다.

현재까지의 데이터에서는 GRU가 우수한 성능을 가졌다. 하지만 BI-LSTM보다 GRU가 항상 좋은 것은 아니다. 앞으로의 결과는 위드코로나, 새로운 변이바이러스 등 여러 가지 원인에 따라서 어떻게 될지 모른다. GRU에서 계절성 부분에서 따로 파라미터를 지정한 SARIMA 보단 덜 반영된 모습을 보이지만, 이것 또한 train으로 반영을 어느 정도 하면서 점점 좋은 값을 보여준다. 데이터가 쌓임에 따라서 어떻게 변할지는 모르지만, n차 대유행이나, 위드 코로나로 인해서 크게 증가할 수 있고 백신

에 대한 효과로 예측값이 크게 벗어날 수 있음에 유의해야 한다. 외인성에는 취약할 수 있지만, 이것 또한 훈련을 통해 일반적인 값을 예상할 수 있고, 또 다시 n차 유행이 올수 있음을 인식함으로써 사회적 거리두기 방침이나 마스크 착용을 잘 지킴으로써 코로나 감염자 수를 줄이는데 경각심을 일깨우는데 이의가 있다.

향후 연구로는 의료적인 측면에서 환자들의 뇌파, 심장박동수 등을 예측해 환자들이 어느 질병에 취약할지 미리 예상해 사전에 예방하는 기술로 이어나갈 수 있다.

REFERENCES

[1] Centers for Disease Control and Prevention. Ensuring COVID-19 Vaccines Work [Internet]. Available: <https://korean.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness.html>.

[2] Centers for Disease Control and Prevention. Effectiveness of Pfizer-BioNTech and Moderna Vaccines Against COVID-19 [Internet]. Available: https://www.cdc.gov/mmwr/volumes/70/wr/mm7018e1.htm?s_cid=mm7018e1_w.

[3] 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.

[4] R. Józefowicz, W. Zaremba, and I. Sutskever, "An Empirical Exploration of Recurrent Network Architectures," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, pp. 2342-2350, Jul. 2015.

[5] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451-2471, Oct. 2000.

[6] K. Clark, M. T. Luong, C. D. Manning, and Q. V. Le "Semi-Supervised Sequence Modeling with Cross-View Training," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1914-1925, Oct. 2018.

[7] S. J. Ko, H. Y. Yun, and D. M. Shin, "Electronic Demand Data Prediction using Bidirectional Long Short Term Memory Networks," *Journal of Korea Software Appraisal Association*, vol. 14, no. 1, pp. 33-40, 2018.

[8] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," in *NIPS 2014 Deep Learning and*

Representation Learning Workshop, Dec. 2014.

[9] D. Britz, A. Goldie, M. T. Luong, and Q. Le, "Massive Exploration of Neural Machine Translation Architectures," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1442-1451, Sep. 2017.

[10] 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