

하천 홍수 예측을 위한 CNN 기반의 수위 예측 모델 구현

조민우¹ · 김수진¹ · 정희경^{2*}

Implementation of CNN-based water level prediction model for river flood prediction

Minwoo Cho¹ · Sujin Kim¹ · Hoekyung Jung^{2*}

¹Graduate Student, Department of Computer Engineering, Paichai University, Daejeon, 35345 Korea

^{2*}Professor, Department of Computer Engineering, Paichai University, Daejeon, 35345 Korea

요 약

수해는 홍수나 해일을 유발하여 막대한 인명과 재산의 피해를 초래할 수 있다. 이에 대해 홍수 예측을 통한 빠른 대피 결정으로 피해를 줄일 수 있으며, 해당 분야에서는 시계열 데이터를 활용하여 홍수를 예측하려는 연구들도 많이 진행되고 있다. 본 논문에서는 CNN 기반의 시계열 예측 모델을 제안한다. 하천의 수위와 강수량을 사용하여 CNN 기반의 수위 예측 모델을 구현하였고, 시계열 예측에 많이 사용되는 LSTM, GRU 모델과 비교하여 성능을 확인하였다. 또한 입력 데이터의 크기에 따른 성능 차이를 확인하여 보완해야 할 점을 찾을 수 있었고, LSTM과 GRU보다 더 좋은 성능을 낼 수 있다는 것을 확인하였다. 이를 통해 홍수 예측을 위한 초기 연구로서 활용할 수 있을 것으로 사료된다.

ABSTRACT

Flood damage can cause floods or tsunamis, which can result in enormous loss of life and property. In this regard, damage can be reduced by making a quick evacuation decision through flood prediction, and many studies are underway in this field to predict floods using time series data. In this paper, we propose a CNN-based time series prediction model. A CNN-based water level prediction model was implemented using the river level and precipitation, and the performance was confirmed by comparing it with the LSTM and GRU models, which are often used for time series prediction. In addition, by checking the performance difference according to the size of the input data, it was possible to find the points to be supplemented, and it was confirmed that better performance than LSTM and GRU could be obtained. Through this, it is thought that it can be utilized as an initial study for flood prediction.

키워드 : CNN, GRU, LSTM, 시계열 데이터, 수위 예측

Keywords : CNN, GRU, LSTM, Time series data, Water level prediction

Received 19 August 2021, Revised 3 September 2021, Accepted 23 September 2021

* Corresponding Author Hoekyung Jung(E-mail:hkjung@pcu.ac.kr, Tel:+82-42-520-5640)

Professor, Department of Computer Engineering, Paichai University, Daejeon, 35345 Korea

Open Access <http://doi.org/10.6109/jkiice.2021.25.11.1471>

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

우리나라 및 세계 각지에서 태풍, 호우 등으로 인한 수해 피해가 매우 크다. 2019년에 행정안전부에서 발간한 재해 연보에 따르면 호우, 태풍과 같은 수해로 인한 피해 금액이 전체 95% 이상을 차지하고 있는 것을 확인할 수 있다[1]. 또한 수해로 인해 침수, 풍랑, 홍수 등을 유발하여 큰 손해를 끼치기도 한다.

홍수를 예측하기 위한 많은 연구가 세계 각지에서 이루어지고 있으며, 연구에서 사용된 핵심 파라미터는 수위와 강수량을 사용하는 것으로 나타났다[2]. 시계열 데이터로, 시계열 분석을 위해 RNN(Recurrent Neural Network) 기반의 LSTM(Long Short Term Memory), CNN(Convolutional Neural Networks), DNN(Deep Neural Network) 등 다양한 기술이 사용되고 있다.

LSTM은 유전자, 음성인식, 센서 데이터, 주가 등 시계열 데이터를 활용하는 많은 분야에서 사용되고 있으며, 최근에는 LSTM의 구조를 변형시켜 성능을 개선시키는 연구들이 활발하게 진행되고 있다[3,4]. CNN은 시계열 데이터의 특징을 자동으로 생성하기 위해 적합한 내부 구조를 검색하고, 추출할 수 있는 장점을 가지고 있어 CNN 기반으로 시계열 데이터를 이용하여 예측하는 연구 또한 많이 진행되고 있다[5].

이에 따라 본 논문에서는 수위와 강수량을 파라미터로 사용한 CNN 기반의 수위 예측 모델을 제안한다. 제안한 모델과 LSTM, GRU(Gated Recurrent Units)와 같은 모델과 실험을 통해 비교하여 성능을 확인하였다. 본 논문은 다음과 같이 구성된다. 2장은 홍수를 예측하기 위한 관련 연구와 비교 모델에 대해 설명하고, 3장은 데이터와 모델 학습에 관해 설명한다. 4장에선 모델을 비교 분석하고 5장의 결론을 기술한다.

II. 관련 연구 및 모델 선정

본 장에서는 수위 예측을 위한 관련 연구와 본 논문에서 제안한 모델의 성능 비교를 위한 모델에 대해 기술한다.

2.1. 관련 연구

홍수 피해 방지를 위한 관련 연구를 살펴보면, 현재 정부에서도 기후 변화에 따른 집중 호우에 대비하여 인

공지능·강우 레이더·위성과 같은 첨단 기술을 활용하여 2025년까지 홍수 예보 시스템을 도입하려 한다[6].

또한 세계적으로도 많은 연구가 진행되고 있는데, 홍수 예측을 위해 사용되는 기술들은 MLP(Multi Layer Perceptron), Bayesian Random Forest, LSTM 등 다양한 기법들이 활용되고 있다[7,8].

2.2. LSTM

LSTM은 시계열 데이터를 처리하기 위해 많이 사용되는 RNN 아키텍처 중 하나이다. RNN의 경우 신호가 순환하는 구조를 가지면서 과거의 데이터를 통해 미래의 값을 예측한다. 하지만 과거 데이터를 오래 기억하지 못하는 문제점이 있는데, 이 문제점을 Forget gate, Input gate, Output gate의 형태로 변형하여 과거 시계열 데이터를 현재 출력층에 반영시킴으로써 해결하였다[9,10]. 그림 1은 LSTM의 구조를 모식화한 것이다.

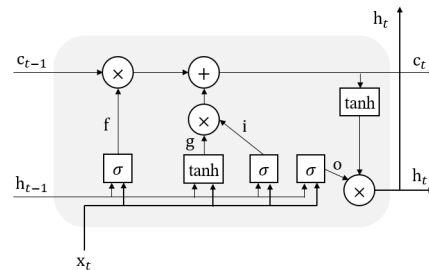


Fig. 1 Structure of LSTM

2.3. GRU

GRU는 LSTM을 간소화한 버전으로 그림 2와 같은 구조를 갖는다. LSTM에 비해 간소화된 모델로 학습 속도가 빠르고, 정확도 측면에서도 LSTM에 비해 크게 뒤처지지 않는 장점을 지닌다. GRU는 2개의 게이트를 가지며, 기존 LSTM의 Forget gate와 Input gate를 합친 Reset gate와 Update gate로 구성된다[11].

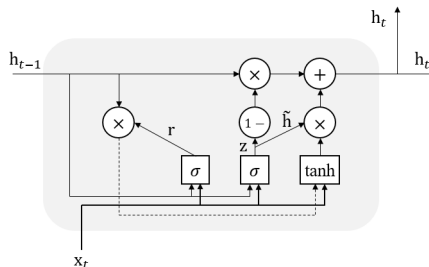


Fig. 2 Structure of GRU

III. 학습 데이터 및 모델 설계

본 장에서는 학습 데이터셋을 설명하고 CNN 기반의 모델과 비교 모델의 하이퍼파라미터에 관해 기술한다.

3.1. 학습 데이터 및 데이터 전처리

수위 예측을 위한 데이터는 하루 단위로 측정된 2015년 1월 1일부터 2021년 6월 24일까지의 울산 지역의 구주교의 수위 및 강수량 데이터를 사용하였다. 구주교는 2016년에 홍수 피해량이 가장 많은 지역이며, 그 이후로도 꾸준히 홍수 피해 및 호우가 발생한 지역이므로 해당 위치의 데이터를 사용하여 실험을 진행하였다. 전체 데이터는 물 환경 정보 시스템을 참조하였다. 그림 3, 4는 훈련 데이터로써 사용된 수위 및 강수량의 데이터를 시각화 한 것이다.

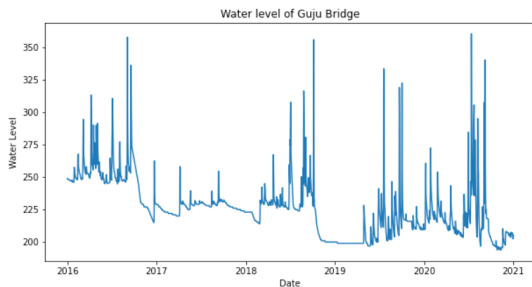


Fig. 3 Water level of Guju Bridge

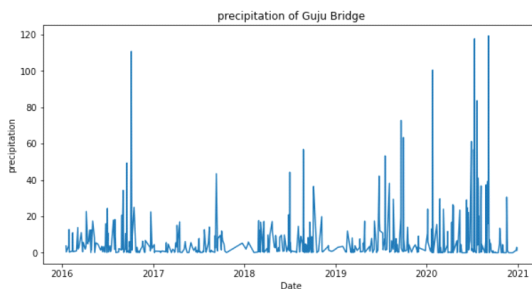


Fig. 4 Precipitation of Guju Bridge

2015년 1월 1일부터 2020년 12월 31일까지의 데이터 1,783개를 훈련 데이터로 사용하였고, 2021년 1월 1일부터 2021년 6월 24일까지의 데이터 171개를 검증에 위해 사용하였다.

강수량 데이터의 경우 결측값이 존재하여 결측값은 0으로 대체하여 진행하였고, 수위 데이터의 경우엔 결측값을 제거하였다. 또한 MinMaxScaler를 통해 전체 데이

터를 0~1 사이로 맞춰주는 데이터 정규화를 진행하였다. 이를 통해 데이터의 최솟값과 최댓값의 차이가 작아짐으로써 모델 훈련 시 더 좋은 성능을 얻는 것을 확인하였다. 이후 초기 실험을 위한 학습으로는 과거 5일의 값을 통해 6일째의 수위를 예측할 수 있도록 데이터를 6개씩 묶어서 진행하였다.

3.2. CNN 기반의 수위 예측 모델

본 연구에서 제안하는 홍수 예측을 위한 CNN 모델의 구성은 그림 5와 같다.

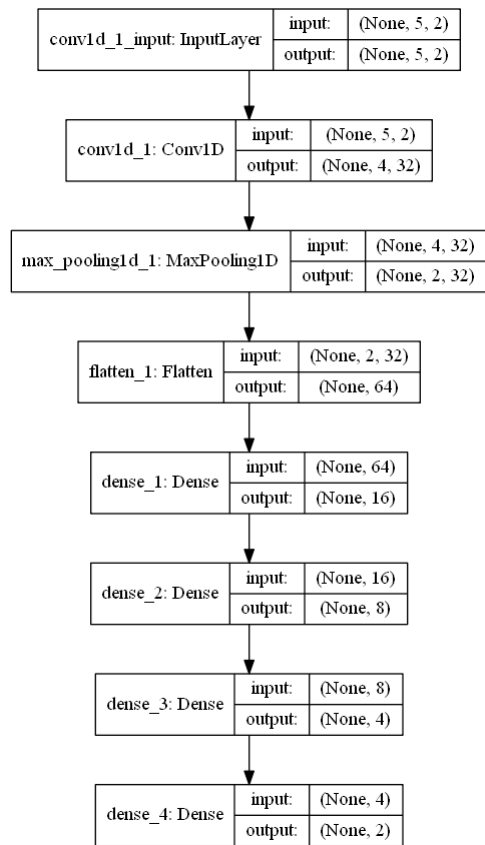


Fig. 5 Structure of CNN Model

입력 계층은 수위와 강수량의 5일치 데이터를 입력 데이터로 사용할 수 있도록 구성되어 있으며, 1개의 Convolution Layer, 1개의 Max pooling Layer, 1개의 Flatten Layer, 4개의 Dense Layer로 구성하였다. 전체 하이퍼파라미터의 수는 1,382개가 사용되었다. 여러 실험을 진행하였을 때 위와 같은 구조를 구성하였을 때 가

장 좋은 결과를 얻을 수 있었다.

3.3. 비교 모델 하이퍼파라미터

그림 6, 7을 통해 LSTM과 GRU의 구성을 확인할 수 있다. 두 개의 모델도 CNN 모델과 동일한 입력 데이터를 받을 수 있도록 입력층을 구성하고 있고, Keras에서 제공하는 LSTM 레이어와 GRU 레이어를 사용하였다.

사용된 하이퍼파라미터의 수는 LSTM은 265,730개를 사용하였고, GRU의 경우 200,194개를 사용하였다.

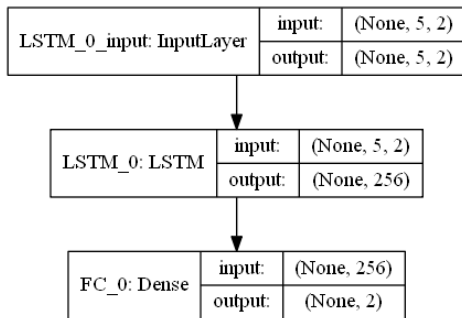


Fig. 6 Structure of LSTM Model

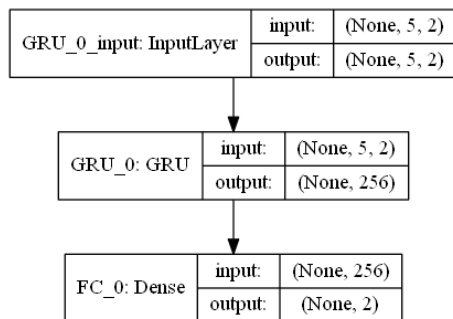


Fig. 7 Structure of GRU Model

IV. 실험 및 결과

4.1. 모델 비교

본 장에서는 LSTM, GRU, CNN 모델에 대해 실험을 진행한다. 예측을 위해 확인하는 데이터의 길이에 따른 성능을 비교하기 위한 실험을 진행하였다. 3가지 모델 모두 3장과 동일한 데이터를 사용하였고, 최적화 기법으로는 Adam을 사용하였으며, 손실 함수로는 mse (Mean Squared Error)를 사용하였다. 3가지 모델 모두

동일한 하드웨어에서 실험을 진행하였다. 표 1은 학습한 하드웨어의 사양이다.

Table. 1 System specification

OS	Windows 10
CPU	Intel i7-10700
GPU	Nvidia Geforce RTX 3070
RAM	64GB
Storage	SSD 500GB & HDD 2TB

그림 8, 9는 회귀 문제에서 성능 지표로 가장 많이 사용하는 MSE 값을 비교한 그래프이다.

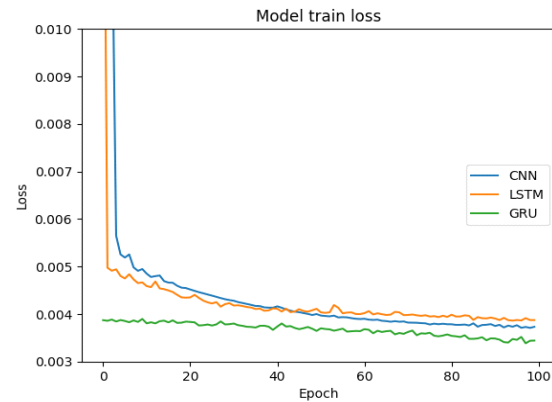


Fig. 8 Model train loss graph

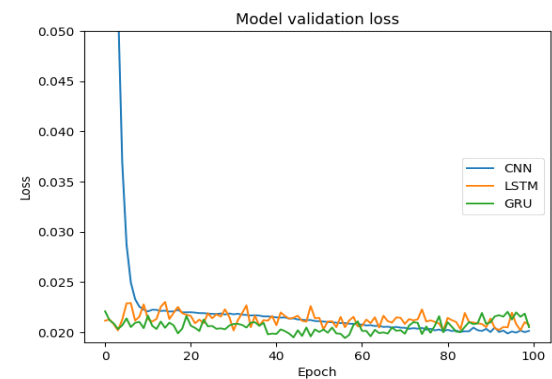


Fig. 9 Model validation loss graph

LSTM의 경우 훈련을 완료하기까지 32초, 최종 훈련 손실 값은 0.0039, 최종 검증 손실 값은 0.0207의 결과를 얻었다. GRU의 경우 훈련을 완료하기까지 17.07초, 최종 훈련 손실 값은 0.0032, 최종 검증 손실 값은 0.0205

의 결과를 얻을 수 있었다.

본 논문에서 제안한 CNN 모델의 경우 훈련을 완료하기까지 18.02초, 최종 훈련 손실 값 0.0034, 최종 검증 손실 값 0.0201의 결과를 얻었다. CNN 모델의 경우 1,382개의 적은 하이퍼파라미터를 가지고, 검증 시 더 좋은 손실 값을 확인할 수 있었다. 다만, GRU 모델보다 훈련 시엔 더 좋은 성능을 보이지 못하였지만, LSTM과 비교하였을 때 모든 부분에서 좋은 성능을 보인 것을 확인할 수 있었다. 아래 표 2를 통해 각 모델의 하이퍼파라미터 및 훈련 시간, 훈련 및 검증 손실 값을 확인할 수 있다.

Table. 2 Model comparison

	CNN	LSTM	GRU
Hyperparameter	1.382	265.730	200.194
Traning time	18.02	30	17.07
Train loss	0.0034	0.0039	0.0032
Validation loss	0.0201	0.0207	0.0205

4.2. 추가 실험

추가로 입력 데이터의 차원에 따른 모델 비교 실험을 진행하였다. 기존 실험에선 입력 데이터의 차원을 (5,2)로 구성했지만, 본 실험에선 입력 데이터의 차원을 (15,2)로 진행하였다. 해당 실험에선 기존의 훈련 반복 횟수를 제외한 나머지 조건은 이전 실험과 동일하게 진행하였으나, 훈련 반복 횟수를 100으로 진행하였을 때 과적합이 발생하여 훈련 반복 횟수를 50으로 변경하여 진행하였다. 그림 10과 11은 3가지 모델에 대해 추가 실험에 대한 손실 값을 비교한 그래프이다.

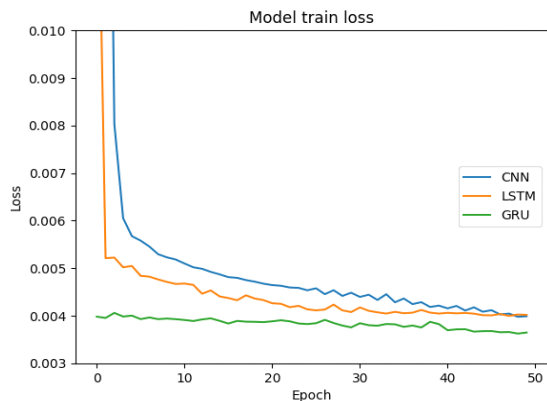


Fig. 10 Model train loss graph when the input data dimension is (15,2)

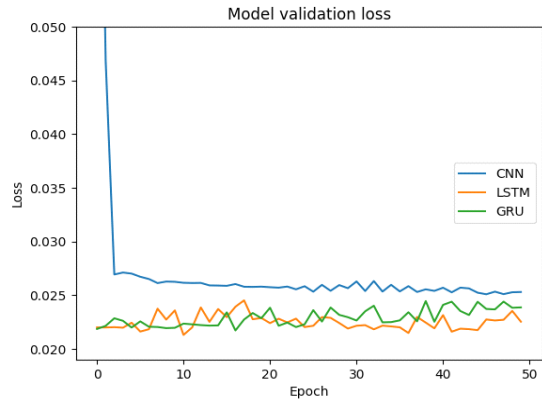


Fig. 11 Model validation loss graph when the input data dimension is (15,2)

입력 데이터의 길이를 길게 진행하였을 때 각 모델 별 하이퍼파라미터 수, 훈련 시간, 훈련 및 검증 손실 값은 표 3을 통해 확인할 수 있다. 3가지 모델 전부 입력 데이터의 차원을 (5,2)로 진행하였을 때 좋은 결과를 얻을 수 있었으며, 입력 데이터 차원이 (15,2)일 때 LSTM의 검증 시 손실 값이 가장 좋은 성능을 보이는 것을 확인할 수 있었다.

Table. 3 Model comparison when input data dimension is (15, 2)

	CNN	LSTM	GRU
Hyperparameter	3,942	265.730	200.194
Traning time	10.2	24.43	9.75
Train loss	0.0036	0.0040	0.0036
Validation loss	0.0253	0.0225	0.0239

V. 결론

본 논문에서 홍수 예측을 위한 중요한 변수 중 강수량과 수위 데이터를 사용하여 수위를 예측하는 실험을 진행하였다. CNN을 활용하여 기존 시계열 분석에 주로 사용되는 LSTM과 GRU 모델과 비교하였고, 적은 하이퍼파라미터로 의미있는 성능을 확인할 수 있었다.

하지만 시계열 데이터 예측에 뛰어난 LSTM과 GRU에 비해 입력 데이터의 길이가 길수록 손실 값이 GRU나 LSTM에 비해 떨어지는 문제점을 확인할 수 있었다. 현재 제안한 모델은 단기간의 데이터를 입력 데이터로

사용한다면 충분히 사용할 수 있지만, 과거 값을 더 기억하려면 성능이 조금 떨어지는 문제가 있어 이러한 내용을 보완하는 연구를 향후 연구로 진행할 필요성이 있다. 이를 통해 실제로 수위 예측을 진행 후 홍수 예측을 위한 초기 연구로 활용될 수 있을 것으로 사료된다.

ACKNOWLEDGEMENT

This study was carried out with the support of 'R&D Program for Forest Science Technology (Project No. 2021340A00-2123-CD01) provided by Korea Forest Service(Korea Forestry Promotion Institute).

REFERENCES

[1] Ministry of Public Administration and Security. 2019 Disaster Yearbook [Internet]. Available: https://www.mois.go.kr/ft/bbs/type001/commonSelectBoardArticle.do;jsessionid=9q+z+-8qP6PFH1L9NfdGfxr.node20?bbsId=BBSTR_00000000014&nttId=81886.

[2] N. A. Maspo, A. N. B. Harun, M. Goto, F. Cheros, N. A. Haron, and M. N. M. Nawı, "Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review," in *IOP Conference Series: Earth and Environmental Science*, 2020.

[3] K. A. Althelaya, E. M. Alfı, and S. Mohammed, "Evalutaion of bidirectional LSTM for short-and long-term stock market prediction," in *Proc. of the 2018 9th International Conference on Information and Communication Systems(ICICS)*, Irbid, Jordan, pp. 151-156, 2018.

[4] D. Niu, Z. Xia, Y. Liu, T. Cai, T. Liu, and Y. Zhan, "Alstm: Adaptiv LSTM for durative Sequential data," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 151-157, 2018

[5] C. H. Hwang and K. W. Shin, "CNN-LSTM Combination Method for Improving Particular Matter Contamination (PM2.5) Prediction Accuracy," in *Journal of the Korea Institute of Information and Communication Engineering*, vol. 24, no. 1, pp. 57-64, 2020.

[6] Ministry of Environment. Promoting flood forecasting using artificial intelligence [Internet]. Available: <http://www.me.go.kr/home/web/board/read.do?boardMasterId=1&boardId>

=1393640&menuId=286.

[7] I. R. Widiyari, L. E. Nugroho, and Widyawan, "Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining," in *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, pp. 1-5, 2017.

[8] F. C. C. Garcia, A. E. Retamar, and J. C. Javier, "A real time urban flood monitoring system for metro Manila," in *TENCON 2015 - 2015 IEEE Region 10 Conference*, pp. 1-5, 2015.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," In *Neural computation*, vol. 9, no. 8, pp. 1735-1780, Sep. 1997.

[10] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, "Detection and Correction Method of Erroneous Data Using Quantile Pattern and LSTM," in *Journal of information and communication convergence engineering*, vol. 16, no. 4, pp. 242-247, 2018.

[11] K. Cho, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.



조민우(Minwoo Cho)

2021년 배재대학교 컴퓨터공학(공학사)
2021년~현재, 배재대학교 컴퓨터공학 석사과정
※ 관심분야 : Deep Learning,
Machine Learning, Big data



김수진(Sujin Kim)

1994년 계명대학교 전자계산학과(공학사)
1996년 계명대학교 전자계산학과(공학석사)
2020년~현재 배재대학교 컴퓨터공학과(박사과정)
2021년~현재 베스핀글로벌 재직
※ 관심분야 : 정보보안, 딥러닝, 빅데이터



정희경(Hoekyung Jung)

1985년 광운대학교 컴퓨터공학과(공학사)
1987년 광운대학교 컴퓨터공학과(공학석사)
1993년 광운대학교 컴퓨터공학과(공학박사)
1994년~현재 배재대학교 컴퓨터공학과 교수
※ 관심분야 : Machine learning, Big data,
Embedded system, U-Healthcare,
IoT