

Catboost 알고리즘을 통한 교통흐름 예측에 관한 연구

전민중, 최혜진, 박지웅, 최하영, 이동희, 이옥*
한양대학교 정보시스템학과

A Study on the traffic flow prediction through Catboost algorithm

Min Jong Cheon, Hye Jin Choi, Ji Woong Park, HaYoung Choi, Dong Hee Lee, Ook Lee*
Division of Information System, Hanyang University

요약 자동차 등록대수와 비례하여 증가하는 교통 혼잡은 도시의 사회경제 발전의 저해 요소로 작용하고 있다. 본 논문은 VDS(Vehicle Detection System)을 통한 데이터를 입력 변수로 사용한다. 본 연구의 목적은 교통 흐름을 단순히 2단계(원활, 정체)가 아닌 5단계(원활, 다소 지체, 지체, 다소 정체, 정체)로 더 정교하게 예측하고, 이 예측에서 가장 정확도가 높은 모델인 Catboost 모델과 다른 모델들을 비교하는 것이다. 이를 위해 본 논문에서는 머신러닝 알고리즘인 Catboost 모델을 통해 5가지 단계를 예측하고 정확도를 다른 머신러닝 알고리즘들과 비교, 분석한다. 또한, 하이퍼 파라미터(Hyper Parameter) 튜닝 및 원-핫 인코딩(One-Hot Encoding) 전처리를 거치지 않은 Catboost 모델과 랜덤 선택(RandomizedSearchcv)을 통해 튜닝 및 데이터 전처리를 거친 모델을 비교, 분석한다. 분석 결과 하이퍼 파라미터 튜닝을 하지 않은 초기 Catboost 모델이 정확도 93%를 보이며 가장 높은 정확도를 기록하였다. 따라서 본 연구는 두가지 의미를 가진다. 첫째로, 초기 세팅된 파라미터들이 적용된 Catboost 모델이 다수의 범주형 변수를 포함하는 교통 흐름 예측에서 다른 머신러닝, 딥러닝 모델들보다 성능이 높다는 결론을 도출했다는 점에서 의의가 있다. 둘째로, 기존 2단계로 예측하던 교통 흐름을 5단계로 예측함으로써 더욱 정교한 교통 흐름 예측 모델을 제안한다는 점에서 의의를 가진다.

Abstract As the number of registered vehicles increases, traffic congestion will worsen worse, which may act as an inhibitory factor for urban social and economic development. Through accurate traffic flow prediction, various AI techniques have been used to prevent traffic congestion. This paper uses the data from a VDS (Vehicle Detection System) as input variables. This study predicted traffic flow in five levels (free flow, somewhat delayed, delayed, somewhat congested, and congested), rather than predicting traffic flow in two levels (free flow and congested). The Catboost model, which is a machine-learning algorithm, was used in this study. This model predicts traffic flow in five levels and compares and analyzes the accuracy of the prediction with other algorithms. In addition, the preprocessed model that went through RandomizedSerachCv and One-Hot Encoding was compared with the naive one. As a result, the Catboost model without any hyper-parameter showed the highest accuracy of 93%. Overall, the Catboost model analyzes and predicts a large number of categorical traffic data better than any other machine learning and deep learning models, and the initial set parameters are optimized for Catboost.

Keywords : Machine Learning, Artificial Intelligence, Catboost, Deep Learning, LSTM

이 논문은 한양대학교 교내연구지원사업으로 연구되었음(HY-2020년도).

*Corresponding Author : Ook Lee(Hanyang Univ.)

email : ooklee@hanyang.ac.kr

Received December 2, 2020

Revised December 31, 2020

Accepted March 5, 2021

Published March 31, 2021

1. 서론

국가통계포털(KOSIS)에서 제공하는 통계표를 통해 자동차 등록 대수가 전년 대비 0.3% 증가하였음을 확인할 수 있다[1]. 이에 따라 꾸준히 증가하는 교통량으로 인한 교통혼잡은 도시의 사회경제 발전의 저해 요소로 작용하고 있다. 이러한 문제들을 예측 및 해결하기 위해 도심지 교통 데이터 이외에 통행 속도, 구간별 이동시간 등의 시계열 데이터를 추출한다[2]. 교통 흐름의 비선형적인 패턴을 예측하여 머신러닝 기반 AI지능형 교통 기술을 통해 해마다 줄어들지 않는 도심지 교통 혼잡을 절감시킬 수 있다[3]. 본 논문은 CatBoost 알고리즘을 통한 정확한 교통 흐름 예측을 목적으로 한다. 실증 분석을 통해 CatBoost 알고리즘의 성능을 다른 머신러닝 예측 모델과 비교한다. CatBoost 알고리즘은 범주형 변수를 사용할 수 있다. 또한 GridSearchcv와 RandomizedSearchcv로 파라미터 조정이 필수적인 다른 부스팅 기법들과 달리, 초기 파라미터 설정값만으로도 효율적인 예측이 가능하다는 장점이 있다[4]. 본 연구에서는 다수의 범주형 변수가 포함된 교통흐름 예측을 위해서 학습 데이터를 5단계(원활, 다소 지체, 지체, 다소 정체, 정체)로 Labeling 한다.

본 연구는 두 가지 절차로 진행된다. 첫 번째로, 하이퍼 파라미터를 조정할 수 있는 머신러닝 알고리즘들과 범주형 변수를 처리할 수 있는 Catboost 알고리즘을 사용하여 얻은 정확도 비교를 통해 최적의 알고리즘을 찾는다. 두 번째로, Catboost 알고리즘의 초기 파라미터로 학습시킨 결과와 하이퍼 파라미터 튜닝을 거친 결과를 비교한다.

본 논문은 다음과 같이 진행된다. 2장에서는 이론적 배경인 교통흐름 예측 관련 연구와 Catboost 알고리즘을 설명한다. 3장에서는 Catboost를 활용한 모델을 제안하고 Catboost와 다른 머신러닝 알고리즘의 파이프라인(Pipeline)을 비교한다. 4장에서는 본 논문에서 제안하는 교통흐름 예측 모델을 실험 및 분석하고 다른 머신러닝을 사용한 모델과 비교한다. 마지막으로 5장에서 본 논문의 결론을 맺고 향후 연구에 대해 제시한다.

2. 관련 연구

2.1 Catboost 알고리즘

Catboost 알고리즘은 범주형 변수 전처리와 오버피

팅 문제 해결에 초점을 둔 Ordered Boosting 기법이다. Ordered Boosting은 모든 잔여 오차(Residual Error)를 차례로 학습하는 기존의 부스팅 모델과 달리 일부 데이터의 잔여 오차를 계산하여 모델을 만들며, 이 모델을 통해 남은 데이터의 잔여 오차를 계산하는 기법이다. 또한, Ordered Boosting에 Random Permutation을 통해 데이터 순서를 섞어줌으로써 오버피팅(Overfitting)을 방지한다[4].

범주형 변수 전처리를 위해서 Catboost 알고리즘은 Random Permutation을 거친 데이터셋에서 같은 범주를 가진 변수들의 평균 표본 값을 계산한다. 이는 식 (1)과 같다.

$$\hat{x}_k^i = \frac{\sum_{j=1}^n [x_j^i = x_k^j] \circ y_j + \alpha P}{\sum_{j=1}^n [x_j^i = x_k^j] + \alpha} \quad (1)$$

Where α is corresponding weight, P denotes a prior value, $x_k = (x_k^1, \dots, x_k^m)$ denotes random vector of m features and $y_k \text{ INR}$ denotes corresponding label.

Catboost 알고리즘은 동일한 information gain을 가진 변수들을 하나로 묶는 변수 조합(feature combinations)을 통해 훈련 속도를 향상시킨다[4]. 또한, 최적의 하이퍼 파라미터를 찾기 위해 GridSearchcv나 RandomizedSearchcv를 사용하는 다른 앙상블 알고리즘과 달리 초기 하이퍼 파라미터값이 최적화 되어있어서 파라미터 튜닝 절차가 요구되지 않는다

2.2 인공지능을 통한 기존의 교통 흐름 예측 모델

조지호(2019)는 대도시나 혼잡도로에 차량검지기 VDS(Vehicle Detection System)을 설치하여 실시간으로 교통 데이터를 활용하였다. VDS를 입력변수로 사용한 LSTM(Long-Short Term Memory)를 설계하였다[5]. 이흥석(2016)은 5분마다 가공된 kilometers per hour (KPH) 평균 통행 속도 데이터를 독립변수로 사용한 DNN(Deep Neural Network) 모델을 설계하였다. 이를 통해 정체면 1, 원활하면 0을 예측하는 이진 분류 연구를 실행하였다. 반면 교통공학에서는 교통상황을 보편적으로 5단계로 구분한다[6].

따라서 본 연구에서는 VDS를 통해 얻은 데이터를 5단계로 예측한다

3. Catboost를 활용한 교통 흐름 예측

3.1 교통 흐름 예측에서 Catboost 모델의 장점

Catboost 알고리즘은 다른 머신러닝 모델들과 달리 범주형 변수를 전처리 과정을 거치지 않고 사용할 수 있다는 장점이 있다. 원-핫 인코딩을 하거나 selectKbest를 통해 변수를 추출하지 않아도 원데이터를 그대로 다룰 수 있다. 또한 다른 부스팅 기법들은 GridSearchcv와 RandomizedSearchcv로 파라미터 조정이 필수적이지만, Catboost 모델은 초기 파라미터 설정값으로도 효율적인 예측이 가능하다는 장점이 있다. 따라서 본 연구에서는 교통 흐름 예측 모델로 Catboost 알고리즘을 사용한다.

3.2 TPI 값을 통한 Labelling

본 연구에서는 공공데이터포털에서 제공 하는 고속도로 차량검지기(VDS) 정체 길이 데이터셋을 사용하였다. 데이터셋은 집계 일자, 집계 시, 콘존ID, 콘존명, 차로 유형, 평균 속도로 구성되어 있고, 이 중 2020년 6월 5일, 12일, 19일, 26일의 데이터를 사용하였다. 교통공학에서는 소통상황을 크게 5단계 (원활, 다소 지체, 지체, 다소 정체, 정체)로 분류하므로 본 연구는 학습데이터를 5가지로 Labelling 하였다. Labelling을 위해 교통 정체와 원활을 구분하는 지표인 TPI(Traffic Performance Index)를 사용하였다. 이홍석(2016)의 연구에 따르면 TPI는 식(2)와 같이 정의할 수 있다[6].

$$TPI = \frac{(v_{max} - v_i)}{v_{max}} \quad (2)$$

TPI 값을 임계값과 비교하여 Labelling한 결과는 다음과 같다.

Table 1. Result of labelling based on TPI values.

0.8 < TPI	5(정체)
0.6 ≤ TPI < 0.8	4(다소 정체)
0.4 ≤ TPI < 0.6	3(지체)
0.2 ≤ TPI < 0.4	2(다소 지체)
0 ≤ TPI < 0.2	1(원활)

3.3 훈련용 데이터셋과 테스트용 데이터셋

본 연구에서는 데이터셋에서 집계 시, 콘 존ID, 콘존명, 차로유형, 집계 일자를 포함한 5가지 독립변수의 데이터

셋으로 추려서 사용하고, 총 개수는 130288개이다. 이 중 70%는 훈련셋, 나머지 30%는 테스트셋으로 사용한다.

Table 2. Variables in dataset

Time
Conzone ID
Conzone name
Lane type
Date

3.4 Catboost 파이프라인(Pipeline)

Catboost 알고리즘은 알고리즘 자체로 범주형 변수를 다룰 수 있고, 기본 파라미터들이 최적화 되어있어서 일반적인 머신러닝 분석보다 적은 단계를 거치며, 파이프라인은 그림[1]과 같다[4].

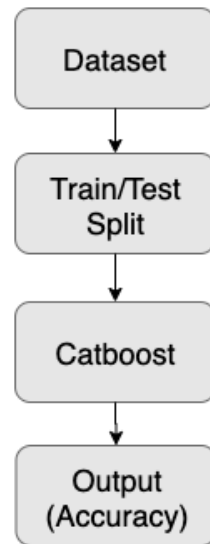


Fig. 1. Pipeline of Catboost

3.5 일반적인 머신러닝 파이프라인

부스팅 알고리즘뿐만 아니라 다른 머신 러닝 알고리즘들은 이진형 변수를 직접 처리 할 수 없다[7]. 따라서 원-핫 인코딩을 통해서 수치형 자료(Numerical data)로 변환시켜야 한다. 훈련 데이터와 테스트 데이터로 분류하고, 식(3)과 같은 StandardScaler를 통해 각 특성의 평균을 0, 분산을 1로 변경해 모든 특성들이 같은 스케일을 갖도록 조정한다. 일반적으로 StandardScaler가 Min-Max Scaler보다 일반 데이터 분석에서 더 효과가 좋다. 또한

본 연구 데이터를 StandardScaler와 Min-Max Scaler 전처리 과정을 각각 거친 후 머신러닝 모델에 적용했을 때 StandardScaler를 거친 데이터의 정확도가 더 높았기 때문에 Min-Max Scaler 대신 StandardScaler를 선택한다. 이후 랜덤 탐색(RandomizedSearchCV)을 통해 하이퍼 파라미터 튜닝을 진행하고, 모든 조합의 파라미터를 시도하는 대신 임의의 수를 대입하여 지정한 횟수만큼 시도한다. 랜덤 탐색은 본 연구에서 사용되는 데이터셋처럼 탐색 공간이 클 경우에 적합하다[8].

$$z = (x - \text{mean}(x)) / \text{std}(x) \quad (3)$$

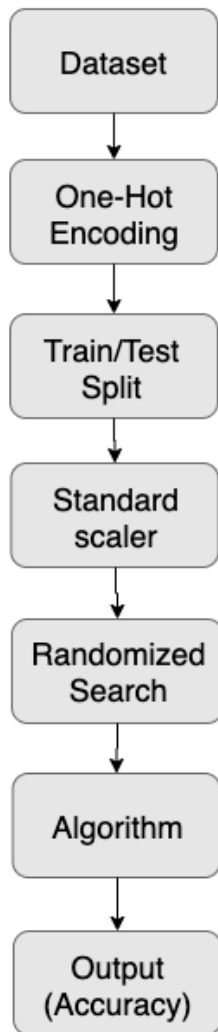


Fig. 2. Pipeline of Machine Learning Models.

4. 실험결과

4.1 Catboost 알고리즘 결과

Catboost 알고리즘 파이프라인을 따라서 실험 한 결과, 모델을 학습하는 데 소요된 시간은 총 4분 50초이다. 최대 최소 정규화나 원-핫 인코딩등의 절차 없이, 훈련 데이터와 테스트 데이터를 분류만 하기 때문에 총 5분 이내로 걸렸다. 그 결과 93%의 정확도가 나왔다.

CatBoost 알고리즘은 변수들이 결과에 얼마나 영향을 주는지를 파악할 수 있다. 변수 중요도를 확인해 본 결과, 콘존 ID, 집계시가 가장 큰 영향을 주었고, 차로 유형이 가장 적은 영향을 미쳤다.

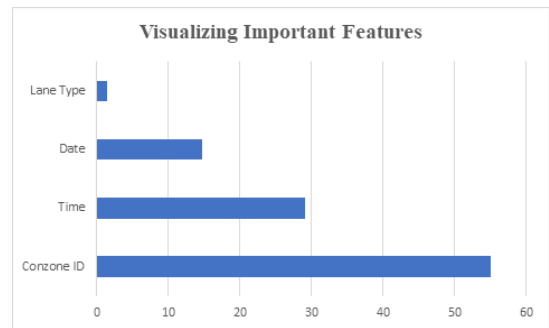


Fig. 3. Feature importance of Catboost

4.2 머신러닝 모델들과의 비교

그림[2]에서 제시한 파이프라인을 따라 머신러닝 모델을 적용했을 때, 전부 같은 C 환경에서 학습시켰음에도 불구하고 XG부스팅 알고리즘의 경우 3시간 이상이 필요하였다. 이는 원-핫 인코딩으로 인해 기존에 5개였던 변수들이 2679개까지 늘어났기 때문이다. 하이퍼 파라미터 튜닝이 필수적인 랜덤포레스트(Random Forest), 그라디언트 부스팅(Gradient Boosting), xg부스팅(XG Boosting), LightGBM(Light Gradient Boosting Machine)은 그림[2]와 같은 절차를 거쳤고, 앙상블 학습이 아닌 나이브 베이즈 분류(Naive Bayes Classification), 선형 회귀(Linear Regression), 결정 트리(Decision Tree) 모델은 랜덤 탐색을 생략하였다.

실험결과, 랜덤포레스트가 85.47%, 결정트리가 84.03% 순으로 높았고, 나이브 베이즈 분류가 44.23%로 정확도가 가장 낮았다.

4.3 Catboost 하이퍼 파라미터 튜닝

Catboost 알고리즘의 가장 큰 장점으로는 기본 하이퍼

파라미터들이 최적화 되어있어 다른 부스팅 알고리즘처럼 최적의 하이퍼 파라미터 탐색을 하지 않아도 된다는 것이다[4]. Catboost 파이프라인에 랜덤 탐색을 추가한 경우 90.52%, 원-핫 인코딩 후에 학습시킨 경우 87.4%의 정확도를 보였다. Catboost 알고리즘을 기본 설정값으로 훈련했을 때 정확도 93%로 성능이 가장 뛰어났다.

Table 3. Accuracy of Catboost and other models.

Catboost	93%
Random Forest	85.47%
Decision Tree	84.03%
LightGBM	83.57%
Logistic Regressor	83.80%
Gradient Boosting	83.97%
XG Boosting	83.90%
Naive Bayes Classification	44.23%
Catboost(Randomized Searchcv)	90.52%
Catboost(One-Hot Encoding)	87.4%

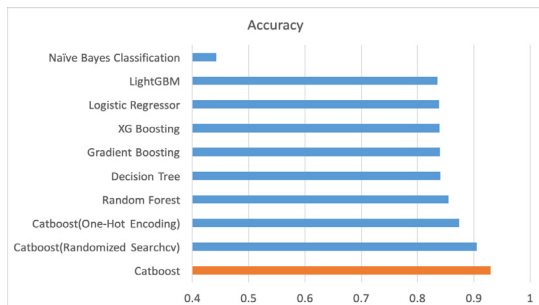


Fig 4. Accuracy of machine learning models

4.4 딥러닝 모델들과의 비교

본 연구에서는 딥러닝 학습 모델로 총 세가지 모델, DNN, LSTM, Bi-LSTM 을 적용하였다. 먼저 데이터 전처리 과정에서 범주형 데이터를 다루기 위해 원-핫 인코딩(One-Hot Encoding) 을 통해 범주형 데이터의 각 범주를 컬럼 레벨로 변경해줬다. 이로 인해 늘어난 변수들을 다루기 위해 selectKbest를 이용해 변수 10개를 추출하였다. 이후 StandardScaler를 이용해 값을 0-1 사이로 변환하고, Label-encoding을 통해 라벨링 하였다.

첫번째로 DNN 모델을 적용하였고, 활성화 함수는 시그모이드(Sigmoid) 함수, 출력층에서는 소프트맥스(Softmax) 함수를 사용하였다. 마지막 출력층에서는 다중 클래스 분류 문제에서 출력층에 주로 쓰이는 소프트맥스 함수를 활성화 함수로 사용하였다. 손실 함수로는

sparse categorical crossentropy가 사용되었다.

두번째 모델로 LSTM(Long Short Term Memory) 모델을 사용하였는데, 이는 데이터의 정보 중 시계열과 관련된 속성(Time, Date) 이 존재하고 RNN(Recurrent Neural Network) 기반의 모델이 시계열 데이터에 적합한 신경 네트워크 유형이기 때문이다[9]. LSTM의 입력층은 3차원이어야 하기 때문에 모델을 적용하기 이전 reshape 함수를 통해 입력층을 3차원 데이터로 만들어 주었다. DNN 모델과 동일하게 활성화 함수는 시그모이드 함수, 출력층 활성화 함수는 소프트맥스 함수, 손실 함수는 sparse categorical crossentropy 함수를 사용했다.

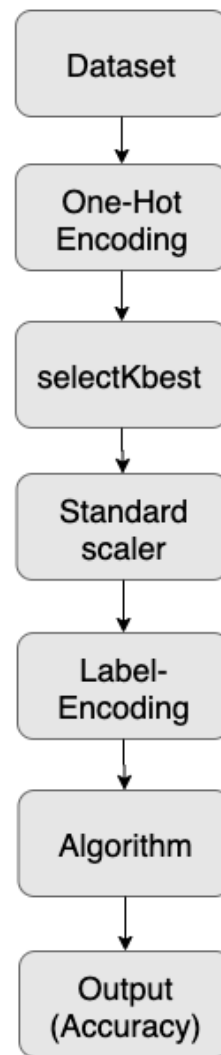


Fig. 5. Pipeline of Deep Learning Models

마지막으로 Bi-LSTM (Bidirectional Long Short Term Memory) 모델을 적용하였다. 본 연구 데이터의 시계열 속성을 양방향으로 고려하기 위해 Bi-LSTM 모델을 사용했으며, 활성화 함수와 손실 함수는 위 두 모델과 동일하게 사용되었다[10].

위와 같은 과정을 거쳤을 때, DNN 모델의 경우 80.05%, LSTM 모델의 경우 80.02%, Bi-LSTM 모델의 경우 80.03% 의 정확도를 보였다.

Table 4. Accuracy of Deep Learning models.

MODEL	Accuracy(%)
DNN	80.05
LSTM	80.02
Bi-LSTM	80.03

4.5 본 논문에서 제시하는 교통 흐름 예측 모델

본 연구에서 제시하는 교통 혼잡 예측 모델은 Catboost를 활용한 모델이다. 하이퍼 파라미터 튜닝을 하면 정확도가 오히려 낮아지기 때문에 기본 Catboost 모델을 제시한다.

교통 혼잡 예측의 경우 데이터셋에 콘존 id나 콘존명과 같은 범주형 변수가 많이 포함된다. Catboost 모델은 이 범주형 변수들을 특별한 전처리 과정 없이 처리하면서, 다른 머신러닝 모델들보다 높은 정확도를 갖고 있다. 5분 이내의 전체 훈련 과정 소요 시간과 원-핫 인코딩으로 인한 메모리 사용량 증가 방지 또한 Catboost 모델의 장점이다.

5. 결론 및 향후 과제

본 논문에서 제시한 Catboost 교통 혼잡 예측 모델은 하이퍼 파라미터 튜닝이 없을 때 정확도가 가장 높으며, 범주형 변수가 많이 포함된 데이터를 분석하는데 유용하다는 결과를 도출했다. 이는 이진 분류로 교통 혼잡을 예측했던 기존 연구와 달리 5가지 단계로 예측하며, 하이퍼 파라미터 튜닝과 원-핫 인코딩을 거치지 않은 Catboost 모델이 다른 머신러닝 모델들 및 딥러닝 모델들보다 성능이 높다는 점에서 학술적으로 의의가 있다. 본 연구에서 제안하는 모델을 통해 교통흐름뿐만 아니라 다수의 범주형 변수를 포함하는 다른 데이터 예측도 가능할 것으로 판단한다. 하지만 이번 연구의 경우 교통흐름에 날씨, 공휴일, 사고 발생과 같은 다른 변수들을 고려하지 않

고 도로 교통 관련 데이터만 활용하였다. 이처럼 다양한 데이터들을 활용하지 못 한 점은 본 연구의 한계점이다.

따라서 향후 연구에서는 날씨, 공휴일, 사고 발생 등과 같은 다른 데이터들을 활용하여 최적의 교통흐름 예측을 위한 모델 개발을 고안해 볼 것이다.

References

- [1] A Reum Oh, "The number of cars registered exceeded 23.44 million...One per 2.2 persons", Auto Times, Available From : <http://autotimes.hankyung.com/apps/news?mode=subscribe&popup=0&nid=03&c1=03&c2=02&c3=&nkey=201907151405391> (accessed November, 14, 2020)
- [2] H. G. Kim, C. Y. Park, C. S. Shin, "Time Series Analysis for Traffic Flow Using Dynamic Linear Model", *KIPS Transactions on Computer and Communication Systems*, Vol. 6, No. 4, pp. 179-188, 2017. DOI: <https://doi.org/10.3745/KTCCS.2017.6.4.179>
- [3] H. S. Yi, K. H. N. Bui, C. N. Seon, "A Deep Learning LSTM Framework for Urban Traffic Flow and Fine Dust Prediction", *Journal of KIISE*, Vol 47, No. 3, pp. 292-297, 2020.
- [4] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin, "CatBoost: unbiased boosting with categorical features", *Advanced in Neural Information Processing Systems* 31, pp. 6639-6649, 2018.
- [5] J. H. Cho, K. H. N. Bui, H. Yi, "An Automated Hyperparameter Search-based Deep LSTM Network for Urban Traffic Congestion Prediction", *Proceedings of The Korean Institute of Information Scientists and Engineers*, pp. 724-726, 2019.
- [6] H. S. Yi, H. J. Jeong, B. Y. Ryu, "Traffic Flow Prediction Using Deep Learning Neural Networks", *Proceedings of The Korean Institute of Information Scientists and Engineers*, pp. 558-560, 2016.
- [7] K. H. Kim, B. H. Chang, H. K. Choi, "Deep Learning Based Short-Term Electric Load Forecasting Models using One-Hot Encoding", *Journal of IKEEE*, Vol. 23, No. 3, , pp. 852-857, 2019. DOI: <https://doi.org/10.7471/ikeee.2019.23.3.852>
- [8] J. Bergstra, Y. Bengio, "Randomized Search for Hyper-Parameter Optimization", *Journal of Machine Learning Research*, Vol. 13, pp. 281-305, 2012.
- [9] D. H. Shin, K. H. Choi, C. B. Kim, "Deep Learning Model for Prediction Rate Improvement of Stock Price Using RNN and LSTM", *The Journal of Korean Institute of Information Technology*, Vol. 15, No. 10, pp. 9-16, 2017. DOI: <http://dx.doi.org/10.14801/jkiit.2017.15.10.9>
- [10] A. L. Lee, *A Study on Author Profiling of Web-based*

Korean Web Text of Elementary and Middle School Students using Word2Vec and Bi-LSTM for Binary Classification, Master's thesis, Ewha Womans University, pp. 15-16, 2020.

전 민 중(Min Jong Cheon)

[정회원]



- 2020년 2월 : 한양대학교 정보시스템학과 (정보시스템학사)
- 2020년 3월 ~ 현재 : 한양대학교 정보시스템과(석박사 통합과정)

<관심분야>

머신러닝, 딥러닝, 정보시스템

최 혜 진(Hye Jin Choi)

[준회원]



- 2019년 3월 ~ 현재 : 한양대학교 정보시스템과(석사과정)

<관심분야>

머신러닝, 딥러닝, 정보시스템

박 지 웅(Ji Woong Park)

[준회원]



- 2020년 3월 ~ 현재 : 한양대학교 정보시스템과(석사과정)

<관심분야>

머신러닝, 딥러닝, 정보시스템

최 하 영(Ha Young Choi)

[준회원]



- 2019년 3월 ~ 현재 : 한양대학교 정보시스템과(학사과정)

<관심분야>

머신러닝, 딥러닝, 정보시스템

이 동 희(Dong Hee Lee)

[준회원]



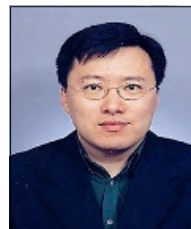
- 2016년 3월 ~ 현재 : 한양대학교 정보시스템과(학사과정)

<관심분야>

머신러닝, 딥러닝, 정보시스템

이 옥(Ook Lee)

[정회원]



- 1989년 6월 : Northwestern 대학교 전산학과(전산학석사)
- 1997년 1월 : Claremont대학교 경영정보학과(경영정보학박사)
- 2002년 3월 ~ 현재 : 한양대학교 정보시스템학과 교수

<관심분야>

정보시스템, IT 분야 철학/형태/응용