

Machine Learning Methods to Predict Vehicle Fuel Consumption

Kwangho Ko*

*Professor, Dept. of Smart Mobility, Pyeongtaek University, Gyeonggido, Korea

[Abstract]

It's proposed and analyzed ML(Machine Learning) models to predict vehicle FC(Fuel Consumption) in real-time. The test driving was done for a car to measure vehicle speed, acceleration, road gradient and FC for training dataset. The various ML models were trained with feature data of speed, acceleration and road-gradient for target FC. There are two kind of ML models and one is regression type of linear regression and k-nearest neighbors regression and the other is classification type of k-nearest neighbors classifier, logistic regression, decision tree, random forest and gradient boosting in the study. The prediction accuracy is low in range of 0.5 ~ 0.6 for real-time FC and the classification type is more accurate than the regression ones. The prediction error for total FC has very low value of about 0.2 ~ 2.0% and regression models are more accurate than classification ones. It's for the coefficient of determination (R^2) of accuracy score distributing predicted values along mean of targets as the coefficient decreases. Therefore regression models are good for total FC and classification ones are proper for real-time FC prediction.

▶ **Key words:** Fuel Consumption, Fuel Economy, Regression, Classification, Machine Learning

[요 약]

본 연구에서는 주행 차량의 실시간 연료소모량을 예측할 수 있는 머신러닝 기법을 제안하고 그 특성을 분석하였다. 머신러닝 학습을 위해 실도로 주행을 실시하여 주행 속도, 가속도, 도로 구배와 함께 연료소모량을 측정하였다. 특성 데이터로 속도, 가속도, 도로구배를, 타깃으로 연료소모량을 지정하여 다양한 머신러닝 모델을 학습시켰다. 회귀법에 해당하는 K-최근접이웃회귀 및 선형회귀와 함께, 분류법에 해당하는 K-최근접이웃분류, 로지스틱회귀, 결정트리, 랜덤포레스트, 그래디언트부스팅을 사용하였다. 실시간 연료소모량에 대한 예측 정확도는 0.5 ~ 0.6 수준으로 전반적으로 낮았고, 회귀법의 경우 분류법보다 정확도가 떨어졌다. 총연료소모량에 대한 예측 오차는 0.2 ~ 2.0% 수준으로 상당히 정확했고, 분류법보다 회귀법의 오차가 더 낮았다. 이는 예측 정확도의 기준으로 결정계수(R^2)를 사용했기 때문인데, 이 값이 작을수록 타깃의 평균 부근에 예측치가 좁게 분포하기 때문이다. 따라서 실시간 연료소모량 예측에는 분류법이, 총연료소모량 예측에는 회귀법이 적합하다고 할 수 있다.

▶ **주제어:** 연료소모량, 연비, 회귀, 분류, 머신러닝

-
- First Author: Kwangho Ko, Corresponding Author: Kwangho Ko
 - Kwangho Ko (kwangho@ptu.ac.kr), Dept. of Smart Mobility, Pyeongtaek University
 - Received: 2022. 07. 26, Revised: 2022. 09. 02, Accepted: 2022. 09. 05.

I. Introduction

주행하는 차량의 연료소모량을 실시간으로 측정할 수 있다면 유용하게 사용할 수 있다. 속도나 가속도와 같은 차량의 주행 조건 및 도로의 경사도와 같은 환경 조건에 따라 달라지는 연료소모량을 확인하거나 분석하여 경제적이고 친환경적인 차량 운영을 도모할 수 있기 때문이다[1]. 운송 체계의 전체적인 에너지 소모량을 모니터링하고 다양하게 분석하여 효율적인 운영 및 관리가 가능해지기도 한다[2].

최근에 출시된 대부분의 차량 계기판에 실시간 연료소모량에 해당하는 그래프나 연비 등이 표시되고 있지만 이를 분석하거나 모니터링 용도로 사용하기는 불편하다. 데이터를 수기로 기록하여 별도의 모니터링/분석 장치에 옮기거나 차량 센서/액추에이터 신호를 측정할 수 있는 차량용 스캐너와 같은 전용 장비를 이용하여 데이터를 저장/이동/전송시켜야 하기 때문이다[3].

본 연구에서는 저렴한 GPS 센서에서 쉽게 측정할 수 있는 차량의 주행 속도, 가속도 및 도로의 구배(경사도)를 이용하여 실시간으로 차량의 연료소모량을 예측할 수 있는 머신러닝 모델을 제안하고 그 성능과 특성을 분석하였다. 실도로에서 차량 실험을 수행하여 모델 학습에 필요한 데이터를 측정하였고, 확보된 데이터에 대해 다양한 머신러닝 기법을 적용하기 위한 데이터 전처리 기법과 함께 학습된 모델의 예측 성능과 오차를 다양하게 비교 분석하였다. 특히 회귀법(Regression)과 분류법(Classification)에 따라 전처리 과정, 그 예측 성능 및 특성이 달라지는 것을 알 수 있었다. 이를 통해, 실시간 연료소모량, 총연료소모량 혹은 평균연비의 도시/연산/분석과 같이 다양한 사용 목적에 적합한 머신러닝 모델을 제안하고자 한다.

II. Related Works

차량의 실시간 연료소모량을 연산 혹은 예측하는 기법은 크게 두 가지로 나눌 수 있다. 우선 차량의 주행저항력을 기반으로 하여 필요한 차량 출력 및 엔진 출력을 역학적으로 계산한 후 연료-출력 변환 모델로 연료소모량을 연산할 수 있다[4-6]. 이러한 방식에서는 차량의 주행저항력을 결정짓는 차량 정보가 필요하다. 차량의 공기저항계수, 전면투영면적, 타이어회전마찰계수 등과 같은 데이터가 필요하고, 차량의 총출력, 총중량, 공인연비와 같은 데이터도 필요하다. 그 계산과정이 상당히 복잡하고 차량의 저항

계수와 같은 정보는 구하기가 어려워 다양한 차량에 대해 이러한 기법을 적용하기는 곤란하다고 할 수 있다. 하지만 정확한 차량 계수값들만 확보되면 계산과정이나 연산 방식이 동일하므로 확장 가능성이 높다고 할 수 있다.

두 번째로 회귀방식의 머신러닝/딥러닝(ML/DL)을 이용한 평균연비(Average Fuel Economy) 예측법이 있다 [7-9]. 입력 뉴런(Neuron)의 개수가 5 ~ 10개, 은닉층(Hidden Layer)이 하나인 단순한 구조로, 일반적인 선형 회귀(Linear Regression) 기법과 동일한 방식이다. 이러한 기법에서는 연료소모량 자체를 타깃으로 하지 않고 일정한 주행거리에 대한 평균연비를 타깃으로 하여 회귀법을 적용한다. 입력 뉴런도 주행 거리/시간에 대한 평균 속도/가속도/도로구배/정지횟수/정차시간과 함께 이러한 값들의 조합으로 만든 변수들의 평균값을 사용한다. 이러한 기법들은 실시간 연료소모량을 예측할 만큼 정확도가 높지 않기 때문에 평균치의 예측을 목표로 회귀법을 적용한 것으로 판단된다. 역학적 기법과 달리 차량과 관련된 계수값들을 확보하지 못한 상태에서 접근하는 방식이기 때문에 쉽게 측정할 수 있는 입력 데이터와 그 조합을 입력 뉴런으로 사용한 특성도 보인다.

이상과 같이 역학적인 연산법이나 ML/DL 기법은 선행 연구에서 찾아볼 수 있으나, 대표적인 ML 기법인 선형회귀, 로지스틱회귀, 결정트리, 앙상블러닝 등을 적용하여 개별 타깃에 대한 예측 정확도와 전체 주행 데이터에 대한 예측 오차 등을 종합적으로 비교 분석한 연구는 없다. 이에 본 연구에서는 이러한 대표적인 ML 기법을 실도로 주행 실험 데이터에 적용하여 실시간 연료소모량 및 총연료소모량, 평균연비 등에 대한 예측 성능, 특징 및 그 오차 등을 비교 분석했다.

이러한 분석을 위해 우선 도로 주행을 통해 차량의 속도, 가속도, 연료소모량 및 도로구배 등을 측정하여 특성 데이터 및 타깃 데이터로 사용하였다. 또한 회귀와 분류 기법을 모두 적용하여 그 성능 및 특성을 분석하기 위해 연료소모량을 일정 구간으로 나누어 레이블을 부여하는 기법도 도입하였다.

이후 K-최근접이웃과 같은 기본적인 ML 기법으로 전반적인 예측 성능 및 특성을 살펴본 후 대표적인 회귀 및 분류 기법의 ML을 적용하여 분석하였다. 이러한 분석을 통해 연료소모량을 사용하는 분야별로 적절한 예측 기법을 제안하였다.

III. Result and Analysis

1. Vehicle Test Results

ML 적용을 위한 차량 실험 데이터를 확보하기 위해 가솔린 연료를 사용하는 중형 승용차를 실도로에서 주행하였다. 주행속도, 가속도, 도로구배 등을 얻기 위한 GPS 센서와 차량의 실시간 연료소모량을 측정하기 위한 OBD(On Board Diagnostics : 차량 정비용 표준 입출력 단자)가 적용된 VBOX라는 장비를 사용하였다. 영국 Racelogics사의 DAS(Data Acquisition System) 장비인 VBOX는 모터스포츠 분야에서 널리 사용되는 장비로 10Hz의 측정속도로 다양한 차량 데이터를 수집한다. 국내에서는 VBOX Korea라는 회사에서 구입이 가능하다(vboxkorea.com 참조). 시험 차량 및 측정 장비의 사양을 Table 1과 2에 정리하였고, 실험 장비가 연결된 차량을 Fig. 1에 도시하였다.

Table 1. Test Vehicle Specification

Property	Spec.
Vehicle Type	Mid Size Sedan
Total Weight	1875kg
Certified Fuel Economy	11.0km/liter
Model Year	2010
Engine Volume	2656cc
Fuel	Gasoline

Table 2. Test Equipment Specification (VBOX)

Property	Accuracy	Resolution
Heading	0.1deg	0.01deg
Position	3.0m	0.1m
Distance	0.05%	1.0cm
Speed	0.1km/h	0.01km/h
Acceleration	1.0%	0.01G

약 81km 정도의 실도로 주행을 통해 약 7만 라인의 데이터를 확보하였는데, 평균 주행 속도 40km/h, 총연료소모량 6.5liter, 평균연비 12.5km/liter 정도이다. Table3에 주행 실험에서 측정된 데이터의 전반적인 특성을 정리하였다. 주행속도(v), 가속도(a), 도로구배(sin)를 입력(특성) 데이터로 사용하고, 타깃을 실시간 연료소모량(FC : Fuel Consumption)으로 설정하였다. 구배가 큰 도로에서 높은 속도/가속도로 주행할 때 연료소모량이 높을 것으로 예상된다. 하지만 연료차단관성주행(fuel-cut)이나 각종 연료보정(가속보정, 냉각수온보정, 고도보정 등)에 의해 연료소모량은 이러한 주행 데이터에 선형적으로 비례하지 않는다[10].

Test vehicle
: Mid passenger sedan



Fig. 1. Test Vehicle & Equipment

일반적으로 연료차단관성주행은 엔진회전수 1200rpm 이상이고 변속기 기어가 주행(D) 모드 상태에서 가속페달을 방치하는 경우 작동되는 주행 모드로 연료소모량이 제로가 되는 영역이다. 이는 감속/정지가 예상되는 주행 조건에서 연비 향상을 도모하는 기법으로 대부분의 전자제어엔진이 장착된 차량에 적용되어 있는 기술이다[11]. 하이브리드/전기 차량의 경우에는 이러한 조건에서 회생제동(Regenerative Braking) 모드가 작동되어 에너지소모량(-)의 값을 가지기도 한다(전력의 회수). 이렇게 비선형적인 변화를 보이는 FC를 예측하기 위해 다양한 ML 기법을 적용하여 그 예측 정확도와 특성 등을 비교, 분석하였다.

Table 3. Test Driving Data Measured

Property	Measured Data
Data Number	72,119
Driven Distance	81.34km
Total Fuel Consumption	6.493liter
Average Fuel Economy	12.528km/liter
Speed (Max. / Avg.)	124.3 / 40.6 km/h

2. Prediction Method

본 연구에서는 ML 기법을 적용하여 FC를 예측한다. ML 기법은 크게 회귀와 분류로 나눌 수 있다. FC가 입력 데이터인 속도/가속도/도로구배에 따라 실시간으로 변화하기 때문에 회귀법으로 예측하는 것이 적절하다. 일반적인 회귀 기법인 선형회귀(Linear Regression), K-최근접 이웃회귀(KNN-Reg : K - Nearest Neighbors Regression)을 적용하여 그 성능을 확인했다. 이때 표준

점수(Standard Scale)로 변환하는 경우의 성능 변화도 확인하였고, 3차의 특성공학(3-order Feature Engineering)도 별도로 적용하였다. 표준점수변환은 사용하는 입력 데이터의 스케일에 차이가 있는 경우 평균과 표준편차를 이용하여 정규화하는 기법이고, 특성공학은 입력 데이터의 조합(곱과 제곱 등)을 입력 데이터로 사용하여 특성 데이터를 풍부하게 활용할 수 있는 기법으로 ML에서 널리 사용된다.

앞에서 언급했듯이 연료소모량의 경우 비선형적인 성질이 있어 분류법을 적용할 필요도 있다. 본 연구에서는 연속적으로 변화하는 연료소모량 타격을 분류법으로 예측하기 위해 Table 4와 같이 연료소모량을 일정한 구간으로 나누어 레이블(Label)을 부여하였다. 연료차단관성주행 영역을 의미하는 레이블 0에서, 고부하영역을 의미하는 레이블 4까지 총 5개의 영역으로 분류하였다. 특히 레이블 4는 연료소모량이 높은 영역으로 실도로에서는 자주 발생하기 어려운 조건에 해당한다.

Table 4. Labelling Scheme

Label	FC (mcc)	Data Number	Data Percent	Load Domain
0	0.0~0.1	5723	7.9%	Fuel-cut
1	~ 8.04	34218	47.4%	Low
2	~ 15.98	18848	26.1%	Mid-Low
3	~ 23.92	9528	13.2%	Mid-High
4	~ Max.	3802	5.3%	High

본 연구에서 측정된 실도로 주행 데이터에서 FC의 최대 값은 46.38mcc 인데, 차량에 따라 다르겠지만 이 정도의 값이 발생하는 주행 조건은 전부하(WOT : Wide Open Throttle)에 해당한다. 이러한 주행 조건은 실도로에서는 드물게 발생하므로 얻을 수 있는 데이터의 개수에 제한을 받을 수밖에 없다. 데이터 수가 불충분한 경우 학습이 어려우므로 본 연구에서는 약 24mcc 이상의 넓은 범위의 FC 영역에 모두 레이블 4를 부여하여 학습이 이루어지도록 도모하였다. 이러한 레이블 부여 방식은 실도로 차량 실험을 통해 학습 데이터를 확보하는 경우 유용할 것으로 판단된다. 즉, 실도로 주행시험에서 고부하 영역에 해당하는 주행 데이터를 충분히 얻기가 어려우므로 상당히 넓은 범위의 연료소모량 데이터에 이 레이블(고부하 영역)을 부여하는 것이다.

이상의 FC 레이블 5개를 부여 설명하자면, 레이블 0는 FC가 제로가 되는 연료차단관성주행 영역(Fuel-cut Domain), 레이블 1은 무부하정차(Idle) 조건을 포함한

저부하 영역(Low Load Domain), 레이블 2는 실도로 주행에서 자주 사용하는 중저부하 영역(Mid-Low Load Domain), 레이블 3은 상당한 가속/고속 주행에 해당하는 중고부하 영역(Mid-High Load Domain), 레이블 4는 WOT 조건을 포함한 고부하 영역(High Load Domain)에 해당한다.

실도로 주행의 경우 교통 정체나 정지 신호 대기과 같은 무부하정차를 하는 경우가 빈번하게 발생한다. 레이블 1이 이런 영역에 해당하는데, 약 47% 정도의 데이터가 이 영역에 속한다. 레이블 4에 해당하는 고부하 영역에 상당히 넓은 범위의 FC를 할당했지만 실도로 주행에서는 드물게 발생하므로 전체 데이터의 5% 정도만 해당되는 것을 알 수 있다. 이러한 데이터 분포는 실도로 주행 조건을 적절히 반영한 결과로 판단된다. 이렇게 총 5개의 영역에 레이블을 지정하여 로지스틱회귀(Logistic Regression), 결정트리(Decision Tree), 랜덤포레스트(Random Forest), 그라디언트부스팅(Gradient Boosting)과 같은 대표적인 분류 방식의 ML 기법을 적용하였다.

3. Machine Learning Result

우선 K-최근접이웃 회귀/분류 기법을 통해 기본적인 예측 성능을 평가해 보았다. 전체 데이터의 20%를 시험용 데이터로 무작위 선정해 이웃의 개수(n_neighbors)를 변화시키면서 예측 정확도를 비교해보았다. 본 연구에서는 사이킷런(Scikit-Learn) 라이브러리의 ML 클래스를 사용하였는데, 예측 정확도(Score)는 회귀의 경우 결정계수(R^2 , Coefficient of Determination)이고, 분류의 경우 맞게 분류한 타깃 개수의 비율로 표현된다.

$$R^2 = 1 - \frac{\sum (Target - Predict)^2}{\sum (Target - Mean of Target)^2} \quad (1)$$

결정계수는 식(1)로 계산되는데, 예측값(Predict)의 분포가 타깃(Target)의 평균(Mean)과 가깝게 분포할수록 그 값이 0에 근접하여 낮은 예측 정확도를 지시하는 성질이 있다. 결정계수가 1에 가까울수록 예측 정확도가 높은 것으로, 이 경우 예측치가 타깃의 변화 범위를 충분히 추종한다고 볼 수 있다. 모델의 예측치가 타깃의 변화 범위를 충분히 표현할 수 있을 때 결정계수의 값이 1에 가까워지고 모델의 예측 정확도가 높은 것이다.

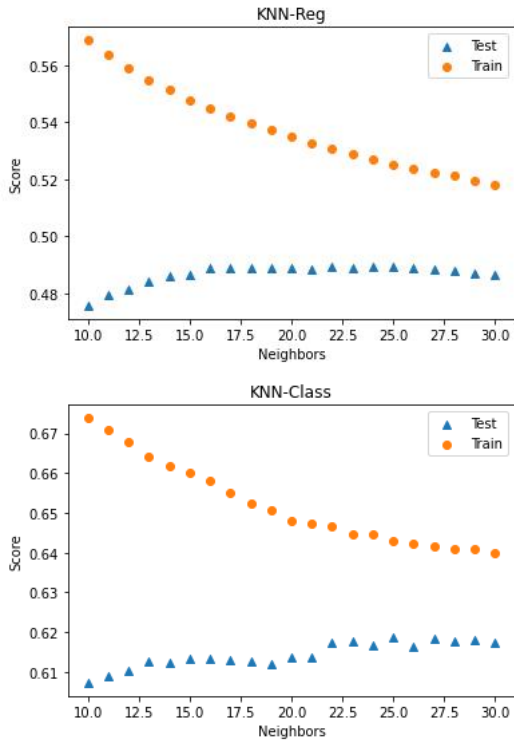


Fig. 2. Prediction Score by Number of Neighbors

이웃의 개수를 10 ~ 30개까지 변화시키면서 학습용 데이터 및 시험용 데이터에 대한 K-최근접이웃 회귀/분류 기법의 예측 정확도 그래프를 Fig. 2에 도시하였다. K-최근접이웃 회귀(KNN-Reg)의 경우 이웃개수 22일 때 예측 정확도가 학습/시험 데이터에 대해 각각 0.53/0.49로 가장 높다. K-최근접이웃 분류(KNN-Class)의 경우 이웃개수 25일 때 학습/시험 데이터에 대한 예측 정확도가 각각 0.64/0.62 수준으로 가장 높다. 두 경우 모두 예측 정확도가 0.5 ~ 0.6 수준으로 낮은 편이고, 회귀에 비해 분류 기법의 예측 정확도가 높다는 것을 알 수 있다. 따라서 타깃 FC의 값 자체를 예측하는 회귀법보다 일정 구간으로 나누어 부여한 타깃 레이블을 예측하는 분류법이 차량의 실시간 연료소모량 예측에는 좀더 적합하다고 판단된다.

Fig. 3과 같이 동일 구간에서 50개의 속도에 대한 K-최근접이웃 회귀/분류에 대한 FC 타깃과 예측 결과를 도시한 그래프에서도 이를 확인할 수 있다. 즉, 레이블을 예측하는 분류법에서 예측치가 좀더 넓은 범위에서 타깃을 추종함을 알 수 있다. 회귀법의 경우 분류법에 비해 좀더 좁은 범위에서 타깃을 예측하고 있다. 정리하자면, 분류법은 타깃 레이블을 좀더 넓은 범위에서 예측하고, 회귀법은 타깃 연료소모량을 좀더 좁은 범위(혹은 평균에 가까운 범위) 내에서 예측하는 특성을 보이기 때문에 예측 정확도가 분류법에서 다소 높은 것으로 나타난다.

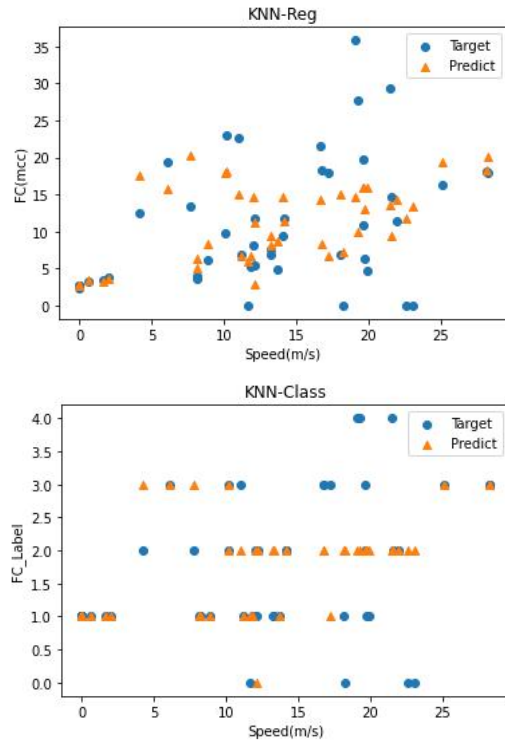


Fig. 3. Prediction Characteristics by Regression/Classification

Table 5. Prediction Results by KNN - Reg / Class

Model		Score	
		Train	Test
KNN - Reg	In	0.531	0.489
	SS	0.560	0.520
	FE	0.558	0.521
KNN - Class	In	0.643	0.619
	SS	0.647	0.613
	FE	0.646	0.616

앞에서 구한 최적의 이웃개수(22(회귀)/25(분류))로 고정하고 입력 데이터인 속도/가속도/도로구배에 정규화(표준점수변환)와 3차의 특성공학을 적용하였다. Table5에 그 결과를 정리하였는데, 원래의 입력 데이터를 그대로 사용하는 경우(In), 표준점수로 변환한 경우(SS : Standard Scaled), 특성공학을 적용한 경우(FE : Feature Engineered by 3-order)의 학습/시험 데이터에 대한 예측 정확도를 확인할 수 있다. 시험 데이터에 대한 예측 정확도를 기준으로 비교하면, 회귀의 경우 정규화/특성공학을 적용하는 경우 좀더 정확해지고, 분류의 경우에는 그 효과가 미미함을 알 수 있다.

Table 6. Characteristics of ML

Model	Characteristics
Lin Reg	Bias = False
Log Reg	Bias = False
DT	Max Depth = 7
RF	Max Depth =7 Estimators = 100
GB	Estimators = 500 Learning Rate = 0.1

Table 7. Prediction Results by Machine Learning

Model		Score		TFC (Liter)	FE (km/L)	Error (%)
		Train	Test			
Lin Reg	In	0.415	0.419	6.482	12.55	-0.17
	SS	0.415	0.419	6.482	12.55	-0.17
	FE	0.499	0.503	6.481	12.55	-0.18
Log Reg	In	0.586	0.584	6.001	13.56	-7.58
	SS	0.586	0.584	6.002	12.55	-7.56
	FE	0.604	0.605	6.471	12.57	-0.34
DT	In	0.612	0.617	6.447	12.62	-0.71
	SS	0.612	0.617	6.447	12.62	-0.71
	FE	0.614	0.621	6.395	12.72	-1.53
RF	In	0.623	0.621	6.418	12.67	-1.16
	SS	0.623	0.621	6.417	12.68	-1.17
	FE	0.629	0.623	6.478	12.56	-0.23
GB	In	0.714	0.636	6.615	12.30	1.88
	SS	0.713	0.637	6.609	12.31	1.79
	FE	0.740	0.632	6.609	12.31	1.79

다음으로 선형회귀(Lin-Reg), 로지스틱회귀(Log-Reg), 결정트리(DT, 최대분기수(max_depth) 7), 랜덤포레스트(RF, 최대분기수 7, 가지수(n_estimator) 100개), 그래디언트부스팅(GB, 가지수 1000개, 학습율(learning rate) 0.1) 기법을 적용하여 학습/시험 데이터에 대한 예측 정확도를 연산하였다. Table6에 각 ML 기법별로 적용한 특성을, Table7에 그 결과를 정리하였는데, 각 기법별로 정규화(SS) 및 특성공학(FE) 적용에 따른 결과도 함께 정리하였다. 또한 이렇게 학습된 모델을 본 연구에 사용된 연료소모량 전체 데이터에 적용하여 예측 총연료소모량(Total FC : TFC [liter]), 예측 평균연비(FE [km/liter]) 및 실측된 총연료소모량과의 차이 백분율(Error [%])도 정리하였다. 로지스틱회귀의 경우 정규화 및 특성공학의 적용 효과가 좋았지만, 다른 기법의 경우에는 정규화/특성공학 적용 여부에 따라 오차(Error)가 크게 달라지지 않았다. 대부분의 경우 예측 총연료소모량과 실측 총연료소모량의 차이가 1% ~ 2% 내외로 충분히 정확한 것을 알 수 있다.

4. Analysis of Result

앞에서 정리한 ML 기법별로 실시간 연료소모량(FC) 예측 정확도(Score) 및 총연료소모량(TFC) 예측 오차(Error)를 비교 분석하였다. 선형회귀의 경우 예측 정확도(R^2)는 낮지만 총연료소모량 오차는 0.2% 수준으로 가장 낮다. 예측 정확도가 낮을수록 타깃의 평균 부근에 예측값이 집중되기 때문에, 회귀법에 있어서 총연료소모량이나 평균연비와 같이 전체 데이터에 대한 총합 혹은 평균 예측은 더욱 정확해질 수 있다. 이는 앞에서 살펴본 K-최근접 이웃 회귀/분류법의 비교 결과와 일치한다. 본 연구에서처럼 차량의 연료소모량 예측 분야에서는 넓은 범위에서 타깃 레이블을 추종하는 분류법인, 평균 부근의 좁은 범위에서 타깃을 추종하는 회귀법에 비해 전체 혹은 평균 예측에는 다소 불리한 것이다.

이를 확인하기 위해 같은 회귀법에 속하지만 FC 예측 정확도와 TFC 예측 오차에 있어 약간의 차이를 보이는 K-최근접이웃 회귀(KNN-Reg)와 선형회귀(LinReg) 모델의 결과를 비교 분석하였다. 두 모델의 동일한 50개 데이터(실시간 연료소모량)에 해당하는 속도별 예측 결과 그래프를 Fig. 4에 도시하였다. 선형회귀(Error = 0.2% 수준)보다 높은 예측 정확도를 보이지만 총연료소모량 오차가 높은 K-최근접이웃 회귀(Error = 0.5% 수준)의 경우 타깃별 예측값이 선형회귀에 비해 좀더 넓게 분포되어 있음을 알 수 있다. 즉, 최근접이웃 회귀는 선형회귀에 비해 실시간 타깃 예측 표현력 자체는 좋지만 전체 타깃의 총합이나 평균에 해당하는 예측 결과에서는 다소 불리해지는 특성을 보인다고 할 수 있다. 이는 K-최근접이웃 회귀 모델보다 그 예측치가 타깃의 평균(중앙) 부근에 밀집해 있는 선형회귀 모델의 특성(낮은 결정계수) 때문이다. 따라서 실시간으로 연비나 연료소모량을 도시/연산하는 용도로 사용할 때는 예측 정확도(결정계수)가 높은 분류법 위주로 사용하고, 총연료소모량이나 평균연비 등을 도시/연산하는 것이 중요한 경우에는 총연료소모량 오차율(Error)이 낮은 회귀법 위주로 사용하는 것이 유리하다고 판단된다. 또한 연료소모량을 일정 구간으로 나누어서 레이블을 부여하여 분류법을 적용하면 회귀법 대비 예측 표현력도 우수하고, 총연료소모량 예측 오차율도 1% ~ 2% 수준으로 충분히 정확하므로 좀더 범용적으로 사용 가능한 것으로 판단된다. 본 연구에서 결정트리, 랜덤포레스트, 그래디언트부스팅과 같은 분류 기법을 적용했을 때 타깃별 연료소모량 예측 정확도가 회귀 방식보다 높고, 총연료소모량 오차율도 충분히 낮았기 때문이다.

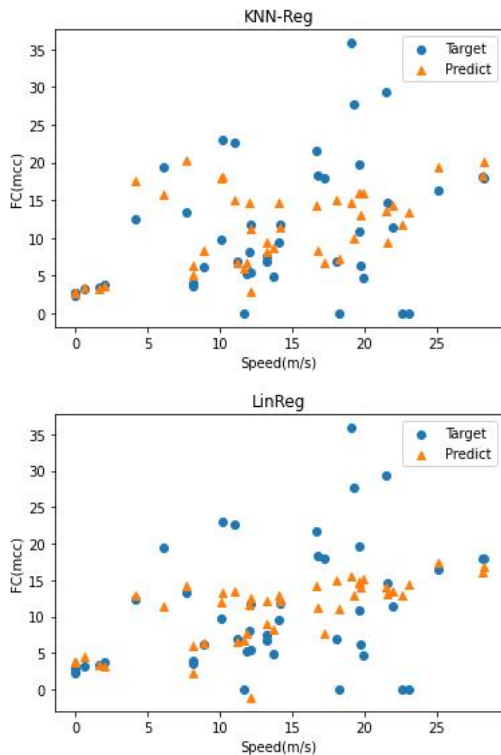


Fig. 4. Prediction Characteristics by Regression/Classification

IV. Conclusions

실도로 주행 실험을 통해 얻은 속도, 가속도, 도로구배 및 실시간 연료소모량을 이용해 차량의 연료소모량 예측을 위한 ML 모델을 학습시켜 그 성능과 특성을 비교하였다. 입력 데이터로 속도/가속도/도로구배를, 타깃으로 실시간 연료소모량(FC)을 설정하여 학습시켰다. 예측 정확도(Score)를 비교하기 위해 결정계수(R^2 : 0 ~ 1)를, 전체적인 예측 성능을 평가하기 위해 예측 총연료소모량(TFC)과 실측 총연료소모량의 차이인 오차(Error(%))를 이용하였다.

실시간으로 변화하는 FC를 예측하기 위해서는 회귀법(Regression)이 적당하나, 연료소모량의 비선형성을 고려하기 위해 FC의 크기에 따라 균일하게 구간을 나누어 5개의 레이블을 부여하여 분류법(Classification)도 적용하였다. 연료차단관성주행 영역, 무부하정차 조건을 포함한 저부하 영역, 중저부하 영역, 중고부하 영역, 전부하 조건을 포함한 고부하 영역에 각각 0 ~ 4의 레이블을 부여하였다. 특히 고부하 영역은 실도로 주행에서 데이터를 확보하기 어려워 넓은 범위의 FC 구간을 포함하는 것이 필요했다.

측정된 데이터셋에 K-최근접 이웃, 선형회귀, 로지스틱회귀, 결정트리, 랜덤포레스트, 그래디언트부스팅과 같

은 ML 기법을 적용하여 그 예측 성능을 평가하였다. FC의 예측 정확도는 0.5 ~ 0.6 정도로, 분류법이 회귀법에 비해 높은 수준이지만 전반적으로 낮은 값으로 볼 수 있다. 총연료소모량 예측 오차는 회귀법의 경우 0.2% 수준이었고, 분류법의 경우 1% ~ 2% 정도로 분류법의 오차가 다소 높았지만 전반적으로 오차는 충분히 낮은 수준으로 볼 수 있다.

이상의 분석 결과를 통해, 실시간 연료소모량 예측 정확도는 분류법이 우수하고, 총연료소모량 예측 오차에 있어서는 회귀법이 우수한 결과를 얻었는데, 이는 실시간 연료소모량을 타깃으로 하는 회귀법의 경우 타깃의 변화 범위를 충분히 추종하지 못하고 평균 부근의 좁은 영역에 예측치가 몰려 있기 때문이다. 따라서 차량의 실시간 연료소모량이나 순간연비를 도시/연산하는 용도에 대해서는 회귀법을, 총연료소모량이나 평균연비를 도시/연산하는 목적에 대해서는 분류법을 적용하는 것이 유리하다.

REFERENCES

- [1] S. Tanaka, "When Tax Incentives Drive Illicit Behavior: The Manipulation of Fuel Economy in the Automobile Industry," *J. Environ. Econ. Manag.*, Vol.104, 2020. DOI: 10.1016/j.jeem.2020.102367
- [2] W. Huang, Y. Guo and X. Xu, "Evaluation of Real-time Vehicle Energy Consumption and Related Emissions in China: A Case Study of the Guangdong-Hong Kong-Macao Greater Bay Area," *J. Clean. Prod.*, Vol.263, 2020. DOI: 10.1016/j.jclepro.2020.121583
- [3] S. Baek and J. Jang, "Implementation of Integrated OBD-II Connector with External Network," *Inf. Syst.*, Vol.50, pp.69-75, 2015. DOI: 10.1016/j.is.2014.06.011
- [4] S. Lee, B. Lee, H. Zhang and C. Sze, "Development of Greenhouse Gas Emissions Model for 2014-2017 Heavy- and Medium-duty Vehicle Compliance," *SAE Tech. Pap.*, 2011. DOI: 10.4271/2011-01-2188
- [5] G. Fontaras, R. Luz, K. Anagnostopoulos, D. Savvidis, S. Hausberger and M. Rexeis, "Monitoring CO2 Emissions from HDV in Europe—An Experimental Proof of Concept of the Proposed Methodological Approach," *Proceedings of the Conference on Transport and Air Pollution, Austria, September, 2014*. JRC91415
- [6] K. Ko, B. Moon and T. Lee, "An Economic Calibration Method for Fuel Consumption Model in HDM4," *Wirel. Pers. Commun.*, Vol.89, pp.959-975, 2016. DOI: 10.1007/s11277-016-3353-2
- [7] S. Alexander, A. Byerly, B. Hendrix, R. Bagwe, E. Santos and

- Z. Ben-Miled, "A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles," *IEEE Trans. Veh. Technol.*, pp.1-10, 2019. DOI: 10.1109/TVT.2019.2916299
- [8] S. Wickramanayake and B. Dilum, "Fuel Consumption Prediction of Fleet Vehicles Using Machine Learning: A Comparative Study," *Proceedings of the Moratuwa Engineering Research Conference*, pp.90-95, April 2016. DOI: 10.1109/MERCon.2016.7480121
- [9] J. Ziółkowski, M. Oszcypała, J. Małachowski and J. Szkutnik-Rogoz, "Use of Artificial Neural Networks to Predict Fuel Consumption on the Basis of Technical Parameters of Vehicles," *Energies*, Vol.14, No.9, 2021. DOI: 10.3390/en14092639
- [10] H. Vdovic, J. Babic and V. Podobnik, "Eco-efficient Driving Pattern Evaluation for Sustainable Road Transport Based on Contextually Enriched Automotive Data," *J. Clean. Prod.*, Vol.311, 2021. DOI: 10.1016/j.jclepro.2021.127564
- [11] E. González, J. Flórez and S. Arab, "Development of the Management Strategies of the ECU for an Internal Combustion Engine: Computer Simulation" *Mech. Syst. Signal Process.*, Vol.22, pp.1356-1373, 2008. DOI: 10.1016/j.ymsp.2007.11.030

Authors



Kwangho Ko received the B.S., M.S. and Ph.D. degrees in Automotive Engineering from Seoul National University, Korea, in 1991, 1993 and 2012, respectively. Dr. Ko joined the faculty of the Department of

Smart Mobility at Pyeongtaek University, Gyeonggido, Korea, in 2018. He is currently a Professor in the Department of Smart Mobility at Pyeongtaek University. He is interested in deep learning, machine learning, and autonomous driving technology.