

# Long-term Driving Data Analysis of Hybrid Electric Vehicle

Ji-Young Woo\*, In-Beom Yang\*\*

## Abstract

In this work, we analyze the relationship between the accumulated mileage of hybrid electric vehicle(HEV) and the data provided from vehicle parts. Data were collected while traveling over 70,000 Km in various paths. The data collected in seconds are aggregated for 10 minutes and characterized in terms of centrality, variability, normality, and so on. We examined whether the statistical properties of vehicle parts are different for each cumulative mileage interval of a hybrid car. When the cumulative mileage interval is categorized into  $\leq 30,000$ ,  $30,000 < \leq 50,000$ , and  $> 50,000$ , the statistical properties are classified by the mileage interval as 82.3% accuracy. This indicates that if the data of the vehicle parts is collected by operating the hybrid vehicle for 10 minutes, the cumulative mileage interval of the vehicle can be estimated. This makes it possible to detect the abnormality of the vehicle part relative to the accumulated mileage. It can be used to detect abnormal aging of vehicle parts and to inform maintenance necessity.

▶ Keyword: Hybrid electric vehicle, OBD-II, Driving data, Mileage, Machine learning

## I. Introduction

인공지능과 부분자율주행 기능을 탑재한 스마트자동차와 하이브리드자동차를 포함한 고성능 전기자동차의 지속적인 발전과 보급이 이루어지고 있다. 이는 반도체, 정보통신기술 등 다양한 분야에서의 기술혁신에서 기인한 것이다. 또한 이러한 기술혁신과 사회적 요구가 결합하여 초연결시대로의 발전이 진행되고 있다. 이 발전은 경제·사회·문화 전반의 인식변화와 변화를 가져오고 있다. 공유경제는 이러한 변화를 대표하고 있으며, 새로운 자동차공유시스템은 대도시를 중심으로 지속적인 비즈니스모델이 개발될 것으로 예상된다. 이러한 모든 것들은 향후 자동차산업, 관련 전후방산업 및 자동차문화에 있어서 많은 변화와 혁신을 가져올 것으로 예상된다.

이러한 변화와 혁신은 개인이 소유하고 관리하는 자동차뿐만 아니라, 법인 등이 소유하는 대규모의 자동차에 대한 효율적이고 안전한 유지보수 및 관리 체계를 포함할 것으로 예상된다. 즉, 현재의 주기적 정비 및 사후 수리에서 예방 정비와 예측 관리 체계로의 변환을 의미한다. Condition Based Maintenance(CBM),

Prognostics Health Monitoring(PHM) 등의 기법이 자동차에 점진적으로 적용되는 것은 이러한 이유에서다.

CBM과 PHM과 같은 기법들 외에도 실주행 데이터를 이용하여 차량의 상태를 예측하는 다양한 방식과 접근법이 논의되고 있다. 실주행 데이터를 이용하더라도, 모든 자동차의 데이터를 실시간으로 전송하거나 저장하여 분석하는 방법은 물리적인 한계가 분명하기 때문에 일차적으로 차량에서 자체적으로 판단한 후에 표준화된 분석 데이터를 전송하는 것이 타당하다. 따라서 실주행 데이터를 이용하여 자체적으로 판단할 수 있는 근거와 표준화된 분석 기법이 향후 중요한 이슈가 될 것으로 예상된다.

본 논문에서는 승용 하이브리드자동차를 대상으로 장기간의 주행데이터를 분석하여 마일리지에 따른 차량의 상태 변화를 파악할 수 있는 판단 기준 및 근거에 대한 기초 연구를 수행하였다. 이를 위해 별도의 OBD-II 진단 포트를 통해 제공되는 실주행 데이터를 장기간에 걸쳐 수집하는 플랫폼을 활용하였다. 실시간 주행모니터링 시스템을 활용하여 하이브리드자동차에

---

• First Author: Ji-Young Woo, Corresponding Author: In-Beom Yang  
\*Ji-Young Woo (jywoo@sch.ac.kr), Dept. of Big Data Engineering, Soonchunhyang University  
\*\*In-Beom Yang (ibyang@sch.ac.kr), Dept. of Smart Automobile, Soonchunhyang University  
• Received: 2018. 01. 29, Revised: 2018. 02. 02, Accepted: 2018. 02. 28.  
• This work was supported by the Soonchunhyang Univ. Research Fund (No. 20180033)

장착된 모터, 인버터, 고전압 배터리 등 주요 부품의 주행 데이터를 계절별, 주행도로별로 다양한 조건에서 수집하였다.

수집된 데이터를 이용하여 누적 마일리지에 따른 부품 및 모듈의 상태 변화를 분석하고 마일리지 변화에 따라 달라지는 변수를 도출하였다. 이러한 변수를 이용하여 마일리지 구간을 구분할 수 있는 모델을 구축하고, 마일리지 구간에 따라 특성변화가 두드러지는 부품 및 모듈의 유무를 파악하였다.

이러한 연구를 통하여 하이브리드자동차의 주행 마일리지 누적에 따른 부품 및 모듈의 특성 변화를 구분 및 확인하여 부품 및 모듈의 노화상태 추정, 기능이상 탐지 그리고 차량 및 부품의 수명예측을 위한 기초자료로써 활용가능 할 것으로 기대된다.

논문의 구성은 다음과 같다. 2장에서는 자동차 실주행 데이터 분석과 관련된 논문을 소개하고 본 연구의 차별점에 대하여 기술하였다. 3장에서는 대상 하이브리드자동차, 주행 환경, 데이터 수집 방법을 언급하였다. 4장에서는 수집된 데이터의 요약정보를 제공하였고, 기계학습을 기반으로 주행 데이터를 분석하여 하이브리드자동차의 누적마일리지에 따른 부품의 특성 차이가 있는지 확인하였다. 마지막으로 본 연구의 의의에 대해 기술하였다.

## II. Related Works

과거에는 차량의 주행 데이터를 수집하고 처리하는 것이 어려웠기 때문에 시뮬레이션을 활용하여 많은 연구가 진행되었으나 정보통신기술의 발전으로 실주행 데이터를 이용하는 연구가 많이 진행되고 있다. 실주행 데이터 기반의 연구는 실제 자동차를 대상으로 다양한 조건에서의 장기간 데이터를 이용하는 연구와 운전자의 성향을 판별하거나 행동을 분석하는 연구로 구분될 수 있다.

미국은 Advanced Vehicle Testing Activity 프로젝트에서 Plug-in Hybrid Electric Vehicle(PHEV)과 Electric Grid 간의 상호 연계 방법과 차량 모니터링을 통한 실증평가 및 PHEV 상용화 촉진 등을 수행하고 있다[1]. 미국 에너지성은 Hydrogen Program, HyFLEET-CUTE Project를 통하여 하이브리드자동차와 연료전지자동차의 실시간모니터링 및 데이터 분석을 추진하고 있으며, 독일 등 유럽도 유사한 연구를 추진하고 있다[2-3].

Wakita 외[4], Miyajima 외[5], Y. Nishiwaki[6], 광병일 외[7]의 연구에서는 차량 부품의 통신 데이터를 수집하여 운전자를 판별하는 연구를 수행하였다. 운전자의 운전 패턴에 따라 자동차 부품의 데이터 값이 달라지는 점에 착안하여 운전자를 판별하는 기계학습 모델을 구축하였다.

Meng[8]의 연구에서는 운전자의 행동 변화를 모델링하는 연구를 수행하였다. 가속기, 제동기, 스티어링휠 데이터를 이용하여 사용자가 앞으로 어떻게 운전할지를 확률 모델을 이용하여 모델링하였다. Choi[9]는 CAN 데이터를 이용하여 운전자 부주의를 탐지하는 모델을 제안하였다. Wahab[10]의 연구는 운전자 프로파일링에 가속 페달 밟는 패턴과 제동장치 밟는 압

력이 주요하다는 것을 밝혔다. Kedar-Dongarka[11]은 운전자의 가속기, 제동기, 기어로부터 추출되는 데이터를 이용하여 운전자 성향을 보수/중립/과격으로 나누었다. Enev 외[12]의 연구에서는 브레이크 페달 위치, 스티어링휠 각도, 장기간 가속도, 회전속도, 운행속도, 기어 변속, 가속기 페달 위치, 엔진 속도, 최대 엔진 토크, 연료 소비율, 연료조절판 위치 등을 이용하여 주행 도로 상황을 주행구간과 주차 구간으로 구분했다.

기존의 연구를 살펴보면 실주행 차량으로부터 데이터를 수집하기 어려웠기 때문에 시뮬레이션 기반으로 차량 데이터를 수집하여 현실성이 떨어졌다. 최근에는 차량 실주행 데이터를 수집하여 활용하는 연구가 다양하게 수행되고 있으나, 하이브리드자동차에 대한 연구는 시작 단계에 있다. 또한 운전자 인지, 주행 스타일 구분, 도로상황 인지 등이 주요 연구주제로써, 부품 데이터를 이용해 차량 마일리지를 구별하려는 시도는 없었다. 특히 하이브리드자동차는 배터리의 수명이 중요하기 때문에 배터리뿐만 아니라 이와 관련된 부품을 파악하고 마일리지별 부품의 차이를 분석하고 모니터링하는 것이 필요하다.

본 연구는 하이브리드자동차의 실주행 데이터를 이용하여 주행 마일리지 따라 자동차 부품 및 모듈의 특성변화를 모니터링하여, 차량의 효율적인 운용과 이상 진단 및 수명 예측 등을 위한 근거로 사용 가능하다는 점에서 의의를 가진다.

## III. Driving Data Collection

### 1.1 Data collection system

하이브리드자동차의 주행 데이터 확보를 위하여 전문연구기관에 구축되어 있는 실시간주행모니터링시스템을 활용하였다. 실시간주행모니터링시스템은 시험로가 아닌 일반 도로에서 주행하는 차량의 다양한 주행 데이터를 실시간으로 모니터링하고 저장할 수 있는 무선 통신 기반의 시스템이다. 이 시스템을 이용하여 주행 중인 차량 및 주행 환경에 대한 데이터를 실시간으로 수집할 수 있으며, OBD-II 진단 포트를 통해 차량의 제어 신호 및 세부 부품에 대한 데이터를 수집할 수 있다.

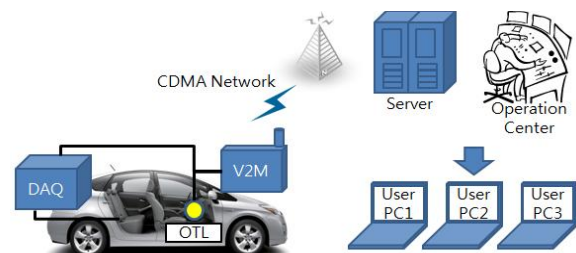


Fig. 1. Schematic Diagram of Real-Time Monitoring System

Fig. 1은 실시간주행모니터링시스템의 구성과 동작 개념을 보여준다. 차량 및 부품의 주행 데이터를 수집하는 차량정보 수

집장치, 차량정보 수집장치로부터 데이터를 전송받아 서버로 전송하고 서버로부터 전송받은 설정에 따라 모니터링 항목 및 장치를 설정하는 실시간 무선전송 단말기, 국내 이동통신망을 사용하여 전송된 데이터를 수집하고 표시하는 M2M 네트워크 플랫폼, 주행 데이터를 저장하고 시험 차량의 주행 상황을 모니터링하는 운영센터로 구성된다.

차량에서 수집된 데이터는 암호화 및 압축되어 M2M 이동통신망을 통하여 실시간으로 운영기관의 수집서버로 전송된다. 전송된 데이터는 차량별로 구분되어 각 프로젝트 데이터베이스에 체계적으로 저장되며, 사용자가 필요로 하는 차량 및 부품의 주행데이터와 분석 환경을 제공한다.

### 1.2 Test vehicle and data specification

본 연구의 대상 차량을 선정함에 있어서 몇 가지 조건이 고려되었다. 상용화된 하이브리드 자동차 중에서 여러 평가에 근거하여 제품의 신뢰성 및 안정성이 입증된 모델을 우선 선정하였다. 이는 첫 상용화 시기 이후 일정 기간이 지났음을 의미한다. 다음으로 의미 있는 판매 대수를 보이는 모델을 선정하였다. 향후 연구의 확장성을 고려한 기준이다. 마지막으로 표준 OBD-II 규격의 진단 데이터를 제공하는 차량을 선정하였다.

Table 1은 선정된 하이브리드자동차의 사양을 보여준다. 60kW 전기모터와 1.8L 가솔린 엔진을 탑재한 Full Hybrid 구동방식의 5인승 승용차이고, 선정 기준을 모두 만족한다.

Table 1. Specifications of Test Vehicles

Item		Value
Performance	Max. Speed	180 km/h
	Acc.(0→96km/h)	9.8 s
	Fuel Efficiency	29.2 km/l
Driving Unit	Motor	60 kW
	Generator	42 kW
	Engine	73 kW
	Type	Full hybrid
Motor	Max. Torque	207 Nm
Engine	Displacement	1,798 cc
	Max. Torque	143 Nm
Battery	Type	Ni-MH
	Capacity	1.3 kWh

대상 차량에서 수집하는 데이터는 별도의 센서나 DAQ를 추가 장착하지 않고 OBD-II에서 제공하는 진단데이터로 한정하였다. 총 21개 ECU에서 1148개 항목의 데이터가 제공된다. Table 2는 21개 ECU와 해당 데이터 항목 수를 보여준다.

1,148개의 데이터 항목 중에서 제어 신호, On/Off 등과 같은 단순 상태 신호로 구성된 데이터들은 모두 제외하였다. 실제로 본 연구에서 의미를 갖는 데이터는 속도, 토크, 온도, 전압, 전류 등 물리적 의미를 갖는 데이터들과 일부 제어 목표 값이기 때문이다. 또한 오류 코드 등 차량 진단 및 유지를 위한 정보 데이터들도 모두 제외 하였다.

Table 2. ECU in the HEV and Included Data

ECU	Number of Data
Engine and ECT	185
Hybrid Control	524
Cruise Control	11
ABS/VSC/TRAC	99
Tire Pressure Monitor	34
EMPS	22
Transmission Control	28
Air Conditioner	30
SRS Airbag	5
Main Body	27
PM1 Gateway	9
D-Door Motor	7
P-Door Motor	7
RL-Door Motor	7
RR-Door Motor	7
Master Switch	13
Combination Meter	25
Smart Key	46
Power Source Control	20
Occupant Detection	18
PM2 Gateway	3

### 1.3 Driving condition

100대 이상 다수의 차량을 대상으로 하는 Fleet 시험에서는 지역별 분배가 중요하지만, 본 연구에서는 1대의 하이브리드자동차를 대상으로 하기 때문에 적절한 주행 시나리오를 설정하는 것이 중요하다. 기본적인 주행 시나리오는 고속도로, 시내도로 및 혼합 구간을 반복하여 주행함으로써 하이브리드 효과가 다양하게 반영되도록 설정하였다. Fig. 2는 실시간주행모니터링시스템에 기록된 주행 경로를 나타낸다.

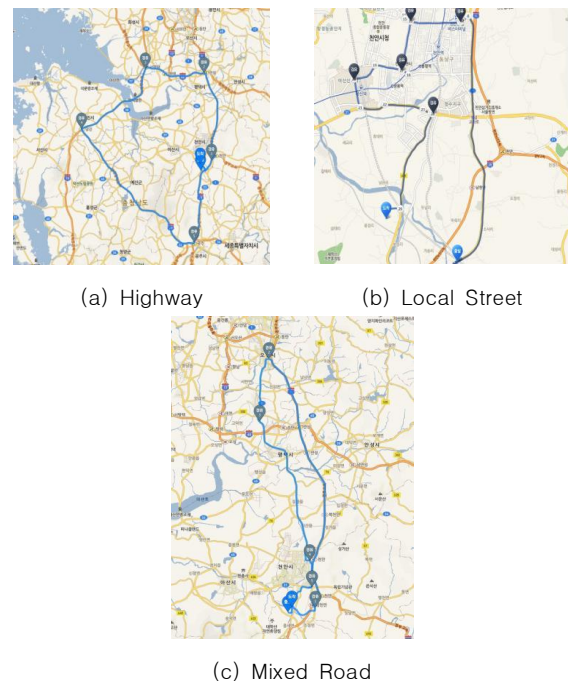


Fig. 2. Driving Road

하루 평균 주행거리, 주차시간 등은 각 국가별로 상이한 통

계를 보이기 때문에 본 연구에서는 이러한 통계 값에 근사한 구간거리를 설정하고 이를 반복 주행하였다. 주차 장소는 야외 주차로 통일하였다. 각 구간별 주행 시험 시, 하이브리드 모드는 일반적인 Normal Mode로 세팅하고 주행하였으며 차량 탑승인원은 1명으로 동일하게 하였다. 법정 속도를 준수하며 주행하였으며, 도로 정체 등 교통상황은 별도로 고려하지 않았다.

여름, 겨울과 같은 계절 효과와 눈, 비와 같은 날씨 효과가 반영되도록 5년에 거쳐 총 7만Km 이상을 주행하여 데이터를 수집하였다. 여름철의 평균 주행온도는 32~36℃, 겨울철 평균 주행온도는 -5~2℃를 나타낸다.

### IV. The Proposed Methodology

#### 1.1 Data exploration

마일리지별 주요 부품 값의 변화를 파악하기 위해서 누적 마일리지별 보조 배터리의 특성을 살펴보았다. Fig. 3과 같이 마일리지 증가할수록 평균 전압레벨은 낮아지고, 전압 변동 폭은 커지는 것으로 나타났다. 또한 단위 시간당 배터리 온도 상승량은 점진적으로 커진다.

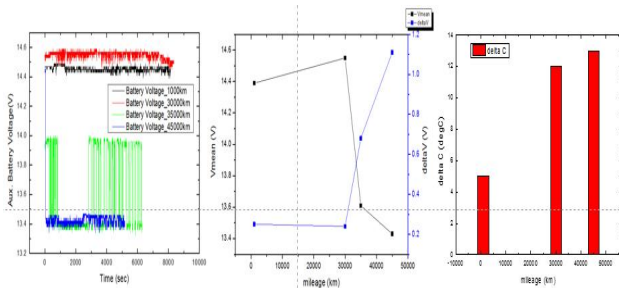


Fig. 3. Auxiliary Battery Characteristics by Mileage

이를 통해 하이브리드자동차의 부품 또는 모듈이 마일리지별로 서로 다른 특성 값을 갖는다는 것을 알 수 있다. 마일리지에 따라 평균값이 변하기도 하고, 변동성이 커지기도 한다. 이를 토대로 마일리지별로 부품 특성의 차이를 도출해낼 수 있는 특질을 도출한다.

#### 1.2 Features

데이터 분포의 특징은 중심성, 변동성, 정규성 등으로 측정한다. 중심성은 평균, 중위수, 최빈치 등으로 평가하며, 데이터의 변동성은 최대, 최소, 평균, 범위, 분산으로 측정한다. 또 하나의 특징은 정규성으로 왜도와 첨도로 데이터가 정규분포를 따르는지를 측정한다. 본 연구는 분포를 예측하거나 통계적 검증을 할 목적이 아니므로 데이터의 정규성을 제외하고, 중심성과 변동성으로 특질을 다각화한다. 중심성을 측정하는 변수로는 평균을 이용하였고, 변동성은 최소, 최대, 표준편차를 이용

하였다. 중심성중에 평균만을 이용한 이유는 차량 부품 데이터는 특정 범위 내에서 자유로이 값이 발생할 수 있기 때문에 발생 빈도가 높은 특정 값이 무엇인지보다는 평균이 더 의미를 가지기 때문이다. 또한 변동성 측면에서는 범위는 최대, 최소값으로부터 도출되고, 분산 또한 표준편차로부터 도출되기 때문에 변수 간 의존성을 줄이기 위해 이 두 변수를 제외하였다.

초 단위로 수집되는 데이터를 구간화하여, 해당 구간 내에서의 차량 부품 데이터의 변동성을 측정한다. 데이터의 특질을 추출하는 구간은 10분으로 설정하였다. 이는 의사결정자가 결정할 수 있는데, 본 연구에서는 구간의 길이와 데이터의 수를 고려하여 10분으로 설정하였다. 데이터 구간을 길게 설정하면, 데이터의 변동성을 충분한 기간 동안 측정할 수 있다는 장점이 있지만, 대신 한정된 주행데이터를 대상으로 하기 때문에 구간 내에서 추출할 수 있는 데이터가 줄어든다는 단점이 있다. 예를 들어, 한 시간 동안의 주행 데이터를 추출하였다면, 10분 간격으로 특질을 추출하였을 때는 총 6개 구간의 특질이 추출된다.

#### 1.3 Mileage interval detection

본 연구에서는 자동차 주행 데이터의 패턴이 누적 마일리지에 따라 변화하는 점을 이용하여 하이브리드자동차 마일리지 구간을 구분할 수 있는지 실험하고자 한다. 이를 통해 마일리지 구간별로 통계적 속성이 크게 변화하는 부품을 찾을 수 있고, 또한 부품의 특성 변화가 큰 마일리지 구간을 알 수 있기 때문이다. 이를 통해 하이브리드자동차의 부품의 이상 상태 등을 예측하거나 비교 분석하는 근거로 활용할 수 있다.

##### 1.3.1 Feature extraction

초단위 데이터를 10 분단위로 변경하여 추출한 결과 각 구간에서 도출된 데이터는 Table 3과 같다.

Table 3. Data Statistics by Mileage Intervals

mileage criteria	# of data	# of driving
Interval 1 =<30,000	32	3
Interval 2 =<50,000	92	7
Interval 3 >50,000	62	6

하이브리드 자동차에서 주요 부품인 배터리의 전압(VB), 전류(IB), 온도(Temp of Batt.), 충전량(State of Charge)의 통계 값 특질을 시각화하여 살펴보았다. 마일리지별 통계 값의 분포를 살펴보면 Fig. 4-7과 같은데, 마일리지 증가할수록 평균, 최대값은 특정 값으로 수렴하는데 비해, 편차는 점점 더 커지는 양상을 보인다. 분석결과와 같이 연구에서 사용한 통계치가 마일리지에 따라 다른 분포를 보이므로 마일리지별 부품특성의 변화를 살펴보는 적합한 변수임을 알 수 있다.

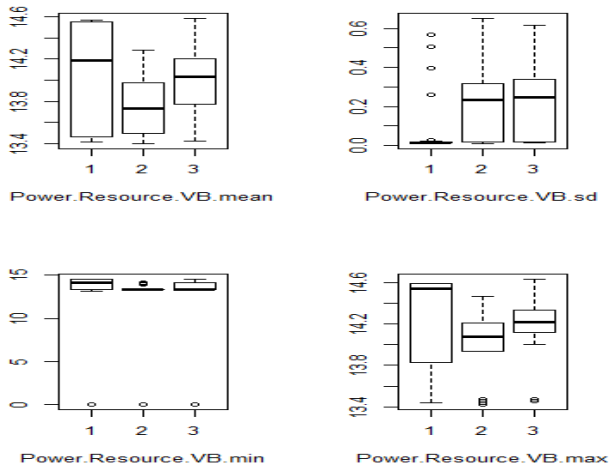


Fig. 4. Boxplot of VB Statistics according to Mileage Intervals

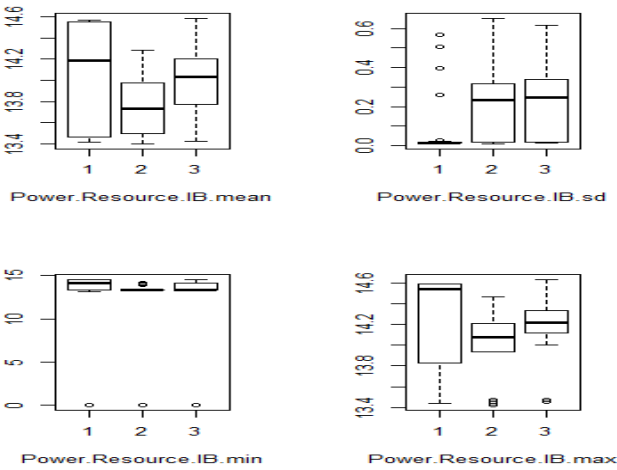


Fig. 5. Boxplot of IB Statistics according to Mileage Intervals

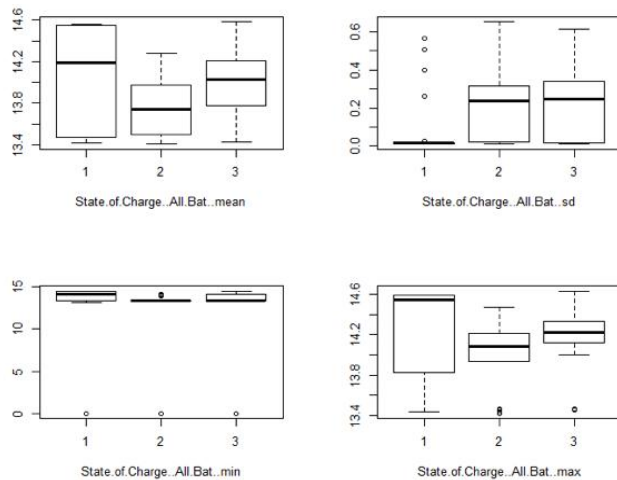


Fig. 6. Boxplot of State of Charge Statistics according to Mileage Intervals

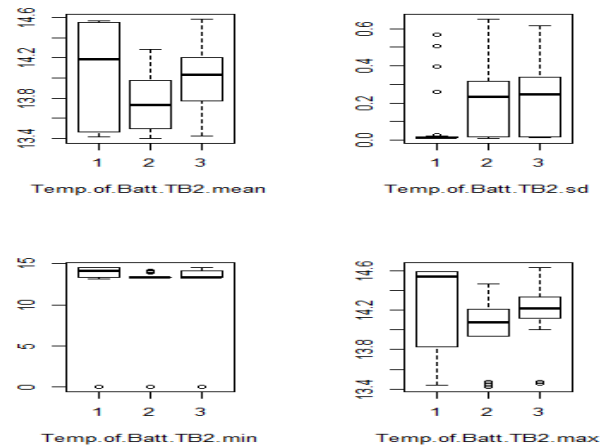


Fig. 7. Boxplot of Battery Temperature Statistics according to Mileage Intervals

1.3.2 Classifier

본 연구에서는 추출된 특질을 기반으로 하이브리드자동차 마일리지를 판별할 수 있도록 학습 알고리즘을 구축한다. 동작 방식이 크게 다른 알고리즘 중에서 대표적인 알고리즘을 선택하여 성능을 비교하였다. 트리(Tree)기반의 알고리즘 중에서는 랜덤 포레스트(Random Forest)를 선택하였고, 인공신경망에서는 오류역전과 알고리즘, 확률기반 모형에서는 나이브 베이즈(Naive Bayes) 모델을 선택하였다. 선택한 알고리즘은 기존 연구에서 그 성능이 입증된 알고리즘이다.

트리 기반의 알고리즘인 랜덤 포레스트는 앙상블 학습 방법의 일종으로, 의사결정트리를 확장한 모델이다. 인공신경망 모델 중에는 가장 기본적인 역전과 알고리즘을 사용하였고, 은닉 층은 2개 층으로 1개 층은 5개의 노드를, 2개 층은 3개의 노드로 설정하여 학습하였다. 나이브 베이즈 모델은 확률 기반의 원칙에 근거하여 확률 모델의 추정 값을 최대우도방법(Maximum Likelihood Estimation)을 이용하여 적은량의 데이터로도 학습 모델을 구축한다.

34개의 부품 데이터로부터 평균, 표준편차, 최대값, 최소값의 총 136개의 특질이 도출되어 학습 데이터로 사용되었다. 마일리지 구간은 3만 이하, 5만 이하, 5만 이상으로 설정하였다. 7만 이상의 구간을 별도로 구분하는 것도 고려하였는데, 5만과 7만 사이에는 부품 데이터의 차이가 크지 않아 학습이 되지 않았다.

1.3.3 Algorithm performance

세 개의 모델을 학습시킨 결과는 Table 4-6과 같다. 학습 모델 평가는 Confusion Matrix와 Accuracy 평가지표를 사용하였다. 세 알고리즘 중 나이브 베이즈 모델이 82.3% 정확도로 가장 좋은 성능을 냈으며, 인공신경망은 학습에 실패했다.

Table 4. Random Forest Performance

Reference \ Prediction	Prediction		
	1	2	3
1	16	0	0
2	16	83	35
3	0	8	27

• Accuracy : 0.68

Table 5. Neural Network Performance

Reference \ Prediction	1	2	3
1	0	0	0
2	32	91	62
3	0	0	0

• Accuracy : 0.49

Table 6. Naive Bayes Performance

Reference \ Prediction	1	2	3
1	26	0	7
2	3	81	9
3	3	10	46

• Accuracy : 0.83

나이브 베이즈의 구간별 정확도는 3만~5만 구간에서 정확도 87.1%로 가장 높다. 인공지능망은 신경망 구조를 복잡하게 설계하여도 학습 성능이 낮아지지 않았다.

마일리지 5만 이상의 구간을 5~7만, 7만 이상으로 나누어 학습한 결과는 Table 7과 같다. 나이브 베이즈 학습 알고리즘을 학습한 결과 5~7만 구간에서 학습이 이루어지지 않는 것을 알 수 있다. 즉, 하이브리드 자동차부품의 특성은 5만~7만, 7만 이상으로 구분되지 않는다. 본 연구의 실험에서는 7만1천까지 운행하였기 때문에 그 후의 마일리지에 대해 추가 주행데이터를 얻어 부품의 특성이 달라지는 구간을 찾는 것을 향후에 수행할 필요가 있다.

Table 7. Naive Bayes Performance with 4 Classes

Reference \ Prediction	1	2	3	4
1	15	0	0	0
2	14	71	15	3
3	3	20	26	18
4	0	0	0	0

• Accuracy : 0.61

### 1.4 Feature analysis

하이브리드자동차 부품의 특성 값으로 마일리지를 구분하는 알고리즘을 구축한 후 알고리즘을 구축하는데 중요하게 사용된 변수와 변수간의 관계에 대해 살펴보았다. 중요하게 도출된 부품은 마일리지 구간에 따라 통계 값이 명확하게 변하는 것으로, 하이브리드자동차 마일리지 구간별로 점검해야하는 부품이라고 할 수 있다.

#### 1.4.1 Feature importance

학습에 주요하게 작동한 변수를 살펴보면 다음과 같다. 랜덤 포레스트에서 사용한 변수 중요도는 incNodePurity로 구하는데, 이는 특정 변수로 트리를 분기했을 때 감소하는 노드의 불순도로 측정된다. 불순도는 각 노드의 잔차 곱의 합으로 구해진다. Fig. 8에 변수의 중요도를 시각화하였다.

가장 중요한 변수는 엔진 흡기 온도, 모터 인버터 온도, 보조 배터리 온도, 엔진 부하 계산 값, 배터리 흡기 온도이다. 마일리지

구간에 따라 특성이 달라지는 주요 부품을 살펴본 결과, 배터리의 전압, 전류, 온도, 충전량 보다는 배터리 흡기 온도, 보조 배터리 온도 등에 변화가 먼저 나타난다는 것을 발견하였다. 이는 대기 온도 등 주행 환경에 따른 물리 값으로써, 실주행 상황에서 나타나는 자명한 결과이다.

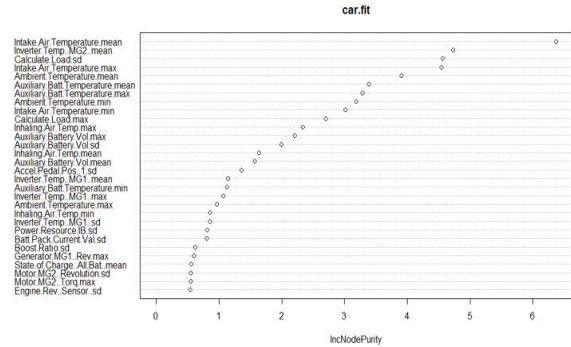


Fig. 8. Variable Importance Plot

#### 1.4.2 Correlation between features

엔진의 온도는 모터 부품의 온도와 상관관계가 높다. 엔진 흡기온도는 모터 온도와 상관관계가 높게 나타난다. 엔진 부하량의 편차는 부스팅 전후 전압 비율, 부스팅 이후 전압, 시동 직후 모터 토크, 제너레이터 토크의 값과 상관관계가 높게 나타난다. 또한 평균 통계치는 다른 변수의 평균값과 표준편차는 다른 변수의 표준편차 값과의 상관성이 높게 나타나며, 다른 주요 부품 값의 최소 값과 역의 상관관계를 가진다.

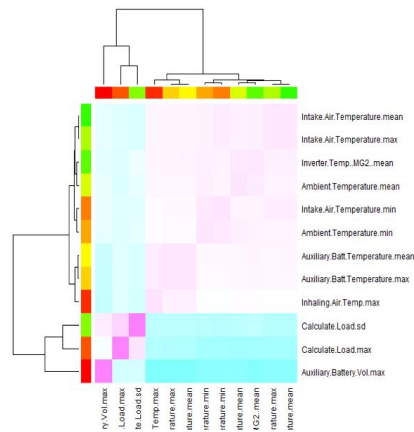


Fig. 9. Heatmap for Key Features

주요 특징간의 상관관계를 살펴보면 Fig. 9와 같다. Heatmap을 통해 이 관계를 알 수 있는데, 보조배터리 전압의 최대 값과 엔진 부하량과 관련 있고, 보조배터리 온도는 모터 인버터와 관련이 높다는 것을 알 수 있다. 이러한 상관관계는 자동차의 일반적인 주행환경에서 나타나는 물리적 특성과 부품 및 모듈에 대한 시스템제어에서 보이는 목표와 일치한다.

## IV. Conclusions

하이브리드 자동차의 마일리지와 OBD-II에서 제공하는 데이터와의 상관관계를 분석하기 위하여 다양한 공로에서 7만 Km 이상을 주행하면서 데이터를 수집하였다. 수집 데이터에서 물리적 의미를 갖는 유의미한 데이터를 선별하고 분석하였다. 초 단위로 수집된 CAN 데이터를 10분 단위로 집계하여, 중심성, 변동성, 정규성 등으로 측정하였다. 이렇게 측정한 통계 특질이 하이브리드 자동차의 마일리지 구간별로 달라지는지를 살펴보았다. 마일리지 구간을 3만이하, 5만이하, 5만초과로 구분하였을 때 10분 단위의 통계값으로 82.3%의 정확도로 구분이 되는 것으로 나타났다. 즉, 하이브리드 자동차를 10분 운행하여 부품의 데이터를 수집하면, 해당 차의 마일리지 구간을 추정할 수 있음을 의미한다. 이를 통해 차량의 마일리지 대비 부품의 이상 여부를 파악할 수 있을 것으로 기대된다. 또한 차량 부품의 노화 등을 예측하여 유지관리 및 정비 필요 여부를 알려 주거나 중고차 거래 시 부품의 노화를 검증하는데 활용할 수 있을 것이다.

분석 결과, 예상과 달리 마일리지 구간별로 부품별 데이터의 통계치가 일관되게 증가하거나 감소하지는 않고, 평균, 최대값은 특정 값으로 수렴하는데 비해, 편차는 점점 더 커지는 양상을 보인다. 평균적으로는 3만이하일 때보다 3~5만 구간에서 주요 부품인 배터리의 전압(VB), 전류(IB), 온도(Temp of Batt), 충전량(State of Charge) 등의 값이 감소하다가 5만 이상일 때 다시 증가하는 양상을 보인다. 하지만 값의 편차는 마일리지 증가할수록 점점 더 커지는 것으로 나타나, 부품의 노후 정도를 측정하려면 값의 평균보다는 편차를 이용하는 것이 적절함을 발견하였다.

본 연구에서는 1대의 차량에서 얻은 데이터를 분석하는 기초연구를 수행하였다. 주행 마일리지와 데이터간의 정확한 상관관계에 기초한 다양한 서비스모델을 개발하기 위해서는 다수의 차량 시험과 그 데이터에 기반한 참조모델의 개발이 필요하다. 지속적인 연구를 통하여 참조모델이 개발되면, 주행환경에 따른 차량별 맞춤 정비 등에 활용될 수 있을 것으로 기대된다.

Information and Systems, vol. 89, no. 3, pp. 1188-1194, 2006.

- [5] C. Miyajima, Y. Nishiwaki, K. Ozawa, T. Wakita, K. Itou, K. Takeda, and F. Itakura, "Driver modeling based on driving behavior and its evaluation in driver identification," *Proceedings of the IEEE*, vol. 95, no. 2, pp. 427-437, 2007.
- [6] Y. Nishiwaki, K. Ozawa, T. Wakita, C. Miyajima, K. Itou, and K. Takeda, "Driver identification based on spectral analysis of driving behavioral signals," in *Advances for In-Vehicle and Mobile Systems*. Springer, 2007, pp. 25-34.
- [7] B. I. Kwak, J. Woo, and H. K. Kim. "Know your master: Driver profiling-based anti-theft method." *Privacy, Security and Trust (PST), 2016 14th Annual Conference on. IEEE*, 2016.
- [8] X. Meng, K. K. Lee, and Y. Xu, "Human driving behavior recognition based on hidden markov models," in *Robotics and Biomimetics, 2006. ROBIO'06. IEEE International Conference on. IEEE*, 2006, pp. 274-279.
- [9] S. Choi, J. Kim, D. Kwak, P. Angkititakul, and J. H. Hansen, "Analysis and classification of driver behavior using in-vehicle can-bus information," in *Biennial Workshop on DSP for In-Vehicle and Mobile Systems, 2007*, pp. 17-19.
- [10] A. Wahab, C. Quek, C. K. Tan, and K. Takeda, "Driving profile modeling and recognition based on soft computing approach," *Neural Networks, IEEE Transactions on*, vol. 20, no. 4, pp. 563-582, 2009.
- [11] G. Kedar-Dongarkar and M. Das, "Driver classification for optimization of energy usage in a vehicle," *Procedia Computer Science*, vol. 8, pp. 388-393, 2012.
- [12] M. Enev, A. Takakuwa, K. Koscher, and T. Kohno, "Automobile driver fingerprinting," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 1, pp. 34-50, 2016.

## REFERENCES

- [1] Vehicle Technologies Program : Advanced Vehicle Testing Activity Site ; <http://www1.eere.energy.gov/vehiclesandfuels/avta/>
- [2] Hydrogen Program : <http://www.hydrogen.energy.gov>
- [3] HyFLEET CUTE Project : <http://www.h2moves.eu>
- [4] T. Wakita, K. Ozawa, C. Miyajima, K. Igarashi, I. Katunobu, K. Takeda, and F. Itakura, "Driver identification using driving behavior signals," *IEICE TRANSACTIONS on*

## Authors



Ji Young Woo received the B.S., M.S., Ph.D. degree in Industrial Engineering from KAIST in 2000, 2002, and 2006. She is currently a Professor in the Department of Big Data Engineering, Soonchunhyang University. From 2008 to 2010, She was

a researcher with AI lab of Arizona University, USA. She was a research professor at Graduate School of Information Security, Korea University from 2010 to 2016. Her research interest includes data mining and business intelligence.



In Beom Yang received the B.S. degree in Mechanical Engineering from Seoul National University and M.S. degree in Mechatronics Engineering from GIST, Korea in 1996 and 1998, respectively. He received the Ph.D. degree in Control

Instrumentation Engineering from Korea University, Korea, in 2008. From 1998 to 2016, Dr. Yang worked as a researcher at Korea Automotive Technology Institute (KATECH), Korea. He joined the faculty of the Department of Smart Automobile at Soonchunhyang University, Asan, Korea. He is currently a Professor in the Department of Smart Automobile, Soonchunhyang University. He is interested in ADAS, intelligent agent and human-x.