

A Machine Learning Approach for Mechanical Motor Fault Diagnosis

Hoon Jung*[†] · Ju-Won Kim**

*Hyper-connected Communication Research Lab., Postal Technology Research Center, ETRI

**Korail Research Institute, Korea Railroad Corp.

기계적 모터 고장진단을 위한 머신러닝 기법

정 훈*[†] · 김주원**

*한국전자통신연구원 초연결통신연구소 우정기술연구센터

**한국철도공사 연구원

In order to reduce damages to major railroad components, which have the potential to cause interruptions to railroad services and safety accidents and to generate unnecessary maintenance costs, the development of rolling stock maintenance technology is switching from preventive maintenance based on the inspection period to predictive maintenance technology, led by advanced countries. Furthermore, to enhance trust in accordance with the speedup of system and reduce maintenances cost simultaneously, the demand for fault diagnosis and prognostic health management technology is increasing.

The objective of this paper is to propose a highly reliable learning model using various machine learning algorithms that can be applied to critical rolling stock components. This paper presents a model for railway rolling stock component fault diagnosis and conducts a mechanical failure diagnosis of motor components by applying the machine learning technique in order to ensure efficient maintenance support along with a data preprocessing plan for component fault diagnosis. This paper first defines a failure diagnosis model for rolling stock components. Function-based algorithms ANFIS and SMO were used as machine learning techniques for generating the failure diagnosis model. Two tree-based algorithms, RadomForest and CART, were also employed. In order to evaluate the performance of the algorithms to be used for diagnosing failures in motors as a critical railroad component, an experiment was carried out on 2 data sets with different classes (includes 6 classes and 3 class levels). According to the results of the experiment, the random forest algorithm, a tree-based machine learning technique, showed the best performance.

Keywords : Motor Failure, Fault Diagnosis, Classification Method, Machine Learning

1. 서 론

2004년 KTX의 도입 후 300km/h의 KTX로의 급격한 기술 발전을 이뤘으나 국내 운영기술의 발전은 더디게

진행되어 고속철도 개통 초기 기존철도와의 혼용에 따른 기술상·운영상 난관을 해결해 오고 있으며, 현재 철도 운영 분야의 IoT(Internet of Thing) 및 빅 데이터 분석 기술의 미흡으로 철도차량 관리 및 운영 등 다양한 분야에서 수작업이 이루어지고 있다.

철도 유지보수 분야는 KTX의 도입에 따른 급격한 철도기술의 발전에 비해서 원천기술의 부재, 이전 기술획득 미흡에 따른 컨설팅 비용 등 철도운영사인 한국철도

공사의 경영에도 악영향을 미치고 있다. 따라서 철도차량 유지보수 기술 분야에 위와 같은 비효율을 제거하여 생산성을 증대시키고, 운영 및 유지보수 기술력을 확보를 통해 국내 철도운영과 철도산업의 역량을 강화하여 철도 운영 및 유지보수 기술 수출국으로서의 입지를 강화할 수 있는 방향 제시가 필요하다.

철도차량 분야에서의 유지보수 시스템의 핵심은 현재 운행 중인 차량의 부품에 대한 상태(condition) 진단을 위해 실시간 모니터링 기술 및 상태 정보를 차량의 검수 및 정비에 보다 효율적으로 적용할 수 있는 방안이 필요하다. 이를 위해서 철도차량의 주요장치들은 센서 네트워크 기술을 접목하여 유지보수 데이터를 실시간으로 확인하고, 네트워크화 하여 유지보수 DB로 관리함으로써 한국철도공사의 유지보수 관리 효율성을 대폭적으로 향상시킬 수 있는 기술 개발이 요구 된다[5, 8].

철도의 사용중단 상태를 야기하는 철도차량의 부품 손상과 불필요하게 발생하는 유지보수 비용을 감소하기 위해 철도 차량의 유지보수 기술은 선진국의 주도로 검사주기 기반 예방정비로부터 예측 적시정비 기술로 발전하고 있다. 또한 시스템 고속화에 따른 신뢰성 향상과 유지보수 비용 절감이라는 두 가지 측면을 모두 충족시키기 위해 고장예지 및 PHM(Prognostic Health Management, 건전성관리) 기술 수요가 증가하고 있다. 핀란드 교통기구에서는 수동형 RFID 솔루션이 적용된 유럽 최대 규모의 철도 인식 파일럿 시스템 구축하여 2012년부터 모든 철도 차량의 실시간 모니터링을 목적으로 시스템 구축 중이다. 유럽의 철도차량 기술 사양인 TSI(Technical Specification for Interoperability)에서는 철도차량에 대한 실시간 차상 모니터링 방안에 대해 규정함으로써 철도차량의 탈선감지 및 오작동 검출을 강화하고 있으며, TSI의 Directive 96/48/EC 규격(2008/232/CE)에서는 기존 지상설비 윤축베어링 검지장치의 오작동에 의한 한계 및 주행 안전성을 위하여 250km/h 이상의 고속열차의 경우 의무적으로 윤축 베어링에 대하여 실시간 차상 모니터링 시스템 장착 규정하고 있다. 국내의 경우에는 서울 메트로 2호선에 전동차 이상검지장치를 설치하여 열 및 진동검지방식으로 운용 중에 있으며, 철도차량 하부 핵심부품에 대한 비파괴진단 기술개발을 위한 연구 및 전동차의 상태기반 유지보수를 위한 핵심기술 개발 연구를 진행 중에 있으나 아직 상용화 이전인 연구진행 단계에 있다[1, 3, 9, 10].

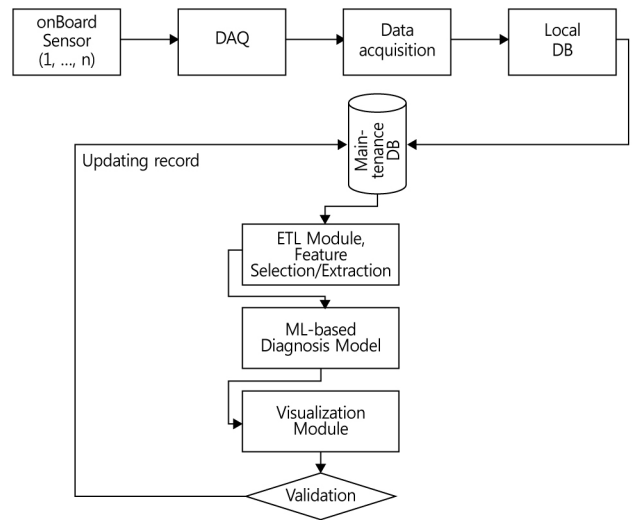
본 논문에서는 철도차량 부품 관리의 효율적인 유지보수 지원 시스템 구축을 위한 방법을 제시한다. 고속열차 차량에 부착된 센서 들을 통해 수집된 상태 정보를 분석하여 부품의 결함여부를 자동으로 판단하는 기술 적용된 머신러닝 기반의 철도차량 고장진단 모델을 개발한다.

논문의 효율적인 이해를 위해서 제 2장에서는 유지보

수지원 시스템의 서비스 흐름도에 기반한 고장진단 서비스 모델을 정의 한다. 제 3장에서는 모터(motor)부품의 베어링(bearing) 및 조인트(joint) 문제, 비균형(unbalance) 등의 부품 고장모드 설정 등의 전처리 단계를 기술한다. 제 4장에서는 머신러닝 기법을 활용한 고장진단 모델의 성능평가 결과를 제시하고 제 5장에서 결론을 맺는다.

2. 철도부품 고장진단 모델 정의

철도차량 부착 센서들로부터 데이터 수집 및 전처리, 고장진단에 대한 시스템 수행 단계는 <Figure 1>과 같다.



<Figure 1> A Flowchart Summarizing Individual Steps for Diagnosing Bearing Fault

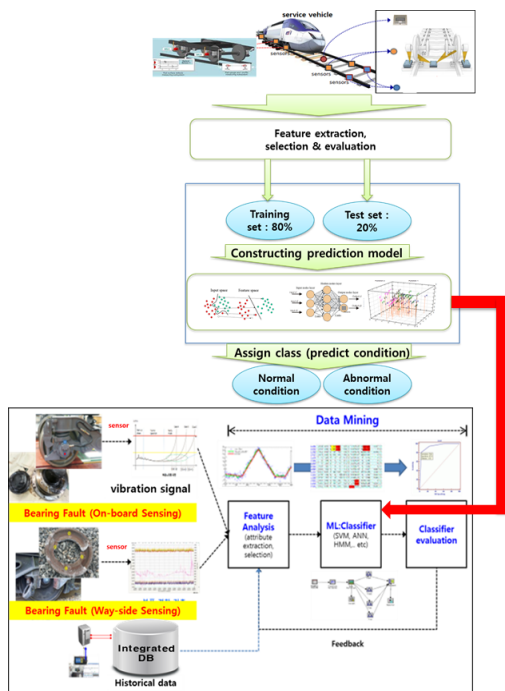
유지보수 DB의 핵심 기능은 대용량의 센싱된 원시데이터의 저장/관리 기능이다. 또한 본 논문의 핵심인 고장진단 모듈은 DB로부터의 분석대상 데이터 추출 단계인 ETL(Extract, Transformation, Load) 및 데이터 전처리 단계를 거친 고장진단 인자들을 활용한 학습된 예측 모델이며, 새로운 입력 데이터에 대해, 고장 여부를 판단한다. 예측 결과의 유효성이 검증된 경우, 그 결과는 유지보수 DB에 저장되는 순환구조이다. <Figure 1>의 간략한 고장진단 단계 정의에 대한 상세 기능 정의는 다음과 같다.

- 차상 로컬 DB에 저장된 센서 측정정보는 주기적으로 유지보수 통합 DB에 전송 가능해야 함
- 차상 DB에 저장된 결함여부 상태정보는 이벤트 발생 시점에 유지보수 통합 DB에 전송 가능해야 함
- 유지보수 통합 DB에 저장된 차상 관련 데이터는 유지보수지원 시스템에 실시간으로 연동되어야 함

- 유지보수 통합 DB의 모든 데이터는 긴급 상황 발생 시, 복구가 가능하도록 주기적 백업을 위해 백업서버에 저장됨
- 시스템은 유지보수 통합 DB로부터 분석 대상 데이터를 위해 ETL(Extract, Transformation, Load) 기능을 제공함
- 시스템은 고장진단 및 수명예지를 위해서 특징 추출기능을 제공함
- 차/지상 부품 고장진단을 위해 분류모델을 제공함
- 시스템은 유지보수 현장 사용자에게 상태 데이터 분석 결과를 제공해야 함
- 시스템은 유지보수 현장 사용자에게 분석 결과에 따른 부품정비 반영 GUI를 제공해야 함
- 시스템은 고장 진단 내역을 유지보수 통합 DB에 갱신할 수 있는 기능을 제공해야 함

유지보수지원 시스템의 부품 고장진단 서비스 모델은 <Figure 2>와 같고 그 내용은 다음과 같다.

- 고속열차 차량에 부착된 센서들을 이용한 부품의 상태를 측정(진폭, 주파수, 회전수)
- 학습을 위한 훈련데이터 셋 생성 및 클래스 라벨 정의 (5가지 고장유형 및 정상 클래스 할당)
- 머신러닝 분류/예측 모델 생성(지도도벡터, 신경망, 베이저안 네트워크, 랜덤포레스트 등)
- 고장 및 정상 데이터(테스트 셋)를 활용한 진단 모델 검증



<Figure 2> Fault Diagnosis Model in Predicting Framework

3. 부품 고장진단을 위한 데이터 전처리

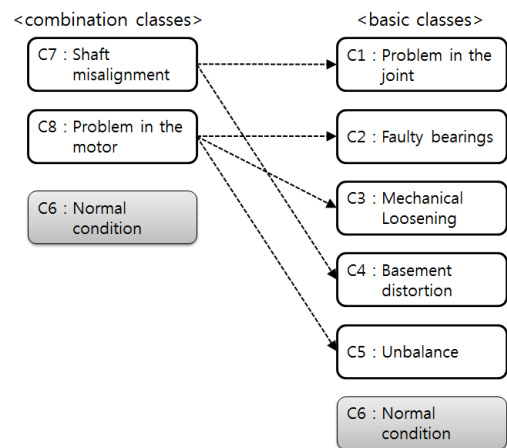
고장진단을 위한 센싱 데이터는 KTX 시험용 차량과 유사한 조건에서 시뮬레이션된 진동(Vibration) 데이터를 활용한다. 특징벡터 선택 및 추출을 위한 단계로서 원시 진동 데이터는 푸리에 분석(Fourier analysis)을 통해 진폭(Amplitude), 주파수(Frequency), 속도(Velocity)에 대한 특징벡터를 추출한다. 주파수에 대한 측정값은 CPM(Cycle Per Minutes)이며, 속도는 분당 회전수를 나타내는 RPM (Revolutions Per Minutes)으로 표현한다. 추출된 특징벡터의 총 수는 9,254개이다.

진단 모델 생성을 위한 데이터 셋 생성의 다음 단계로는 모터의 고장 유형별 정의, 즉 클래스 라벨 생성 및 할당이다. 클래스 라벨은 <Table 1>과 같이 총 6가지로 구성된다.

<Table 1> Basic Classes for Motor Faults

Class Label	Fault Name
C1	Problems in the joint
C2	Faulty Bearings
C3	Mechanical Loosening
C4	Basement Distortion
C5	Unbalance
C6	Normal conditions

기계적 고장유형은 계층적 구조를 형성하며, 유지보수 전문가에 의해 포함관계 형성이 가능하며, 이를 분류트리(taxonomy tree)라 한다. 본 논문에서는 가장 낮은 레벨에서의 고장유형을 기본 클래스라 정의하고 <Figure 3>의 분류트리에 따라 조합클래스를 구성하여 두 가지 데이터 셋을 구성한다. 즉, 기본클래스 구조의 데이터 셋은 고장유형 C1-C5의 5개이며, 정상 상태인 C6로 구성한다.



<Figure 3> Diagnostic Taxonomy of the Motor Faults

조합 클래스 구조의 경우, 총 3개의 클래스를 할당하며, C7, C8 고장유형과 C6인 정상 상태 라벨로 구성한다. 두 가지 데이터셋의 속성 테이블은 <Table 2>와 같다.

<Table 2> Extracted Features From Raw Data

Feature	Description
Class label	Class label (C1, C2, ..., C6), (C6, ..., C8)
CPM	frequency of the measure
Amplitude(Horizontal)	Amplitude measure for direction(Horizontal)
Amplitude(Vertical)	Amplitude measure for direction(Vertical)
Amplitude(Axial)	Amplitude measure for direction(Axial)
RPM	rpm of the machine

4. 부품 고장진단을 위한 머신러닝 기법

본 연구에서는 다양한 머신러닝 알고리즘을 적용하여 철도차량 주요부품의 최적 및 고신뢰도의 학습모델을 제시하는 것이 최종목표이다. 따라서 머신러닝 기법의 함수 기반 및 트리 기반의 대표적인 알고리즘을 선정하여 3장에서 생성한 데이터 셋을 시험한다. 고장진단 모델 생성을 위한 선택 알고리즘은 <Table 3>과 같다.

<Table 3> Classification Methods and Algorithms

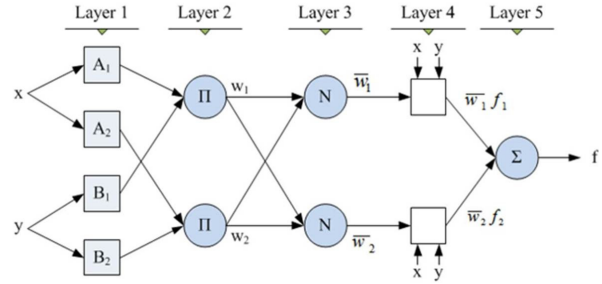
Method Category	Algorithm
function-based	- ANFIS(Adaptive Neuro Fuzzy Inference System) - SVM(Support Vector Machine)
Tree-based	- DT(Decision Tree) - Random Forests

4.1 함수 기반 머신러닝 기법

기계적 장치의 고장 예측은 입출력 변수의 불확실성을 포함하므로 학습에 의해서 특정 고장유형을 결정한다. 따라서 예측시스템을 수리적으로 모형화하기가 어렵고 비선형성을 내재하고 있어 ANFIS(Adaptive Neuro Fuzzy Inference System)[7] 기법을 활용하여 예측의 정확도를 높일 수 있다.

이 논문에서 적용된 ANFIS 구조는 <Figure 4>와 같고 각 layer의 동작 특성과 학습 절차는 다음과 같다.

- ① Layer 1 : 각 노드 i 는 함수 $O_{1,i} = \mu_{A_i}(x)$, $i = 1, 2$ 와 $O_{1,i} = \mu_{B_{i-2}}(y)$, $i = 3, 4$ 를 가진다. x 는 노드 i 의 입력 값이며, A_i 는 노드 함수와 연관



<Figure 4> Structure of ANFIS[12]

된 fuzzy 집합을 의미한다. $O_{1,i}$ 는 입력 값 x 가 A_i 에 소속된 정도를 나타내는 membership 함수이다. μ_{A_i} 는 다양한 형태로 식 (1)과 같이 표현되고, 파라미터를 조정하여 식 (2)과 같이 표현할 수 있다. 본 연구에서는 gaussian 변수 함수를 이용하여 정성적인 자료를 정량화 시켰으며 식 (1)은 변수 함수를 나타낸 것이다 또한 이 층에서 사용한 변수를 premise 매개변수이다.

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^b} \tag{1}$$

$$\mu_{A_i}(x) = \exp - \left\{ \left(\frac{x - c_i}{a_i} \right)^2 \right\} \tag{2}$$

- ② Layer 2 : T-norm 연산을 수행하여 각 membership 함수를 곱하여 다음 식 (3)와 같이 표현한다.

$$O_{2,i} = \omega_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i = 1, 2$$

- ③ Layer 3 : i 번째 규칙을 정규화하여 식 (4)와 같은 연산을 수행한다.

$$O_{3,i} = \bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2}, \quad i = 1, 2$$

- ④ Layer 4 : 각 규칙의 출력함수에 layer 3에서 구한 적합도(compatibility)를 곱하는 연산을 식 (5)과 같이 수행한다. 파라미터 p_i, q_i, r_i 는 오차를 최소화 하도록 값을 결정한다.

$$O_{4,i} = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x + q_i y + r_i), \quad i = 1, 2$$

- ⑤ Layer 5 : 앞의 과정을 통하여 출력값을 계산한다.

$$O_{5,i} = \sum \bar{\omega}_i f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i}$$

분류/예측(classification/prediction) 분야에서 사용되는 대부분의 분류 기법들은 훈련 데이터의 클래스별 데이터 크기가 동일하다는 전제 하에 디자인되었기 때문에 클래스별 불균형 데이터 크기 문제에서 자유롭지 못하다. 따라서 각 클래스 간의 데이터 불균형 문제는 분류모델의 성능의 저하를 가져온다. 모터의 기계 고장진단 문제에서 이러한 클래스 불균형은 예외 일 수 없다. 센서로부터 수집된 데이터에 대한 클래스 분포는 불균형을 이룬다. 이러한 불균형 학습을 위해 클래스에 서로 다른 정규화 값이 부여되는 SVM(Support Vector Machine)의 최적화 문제의 구현에 SMO(Sequential Minimal Optimization)[11] 알고리즘이 적합하다. SMO 알고리즘의 학습단계는 훈련 데이터를 이용하여 최적의 고차원 평면(hyper-plane)을 찾게 되고 테스트 데이터를 이용하여 분류를 하게 된다. 여기서 커널 함수 및 최적 파라미터 선택은 분류의 중요한 역할을 한다. SMO 알고리즘은 두 가지 Poly Kernel와 RBF(Radial Basis Function) Kernel을 제공하지만 일반적으로 다음과 같은 이유로 인해서 RBF을 커널로 많이 사용한다. RBF kernel은 클래스 라벨과 속성(attributes) 사이의 비선형 관계(nonlinear relationships)의 조작(handle)이 가능하고, Poly 커널에 비해 더 적은 파라미터를 갖는다. 이는 모델 선택에 대한 복잡도를 줄여 주는 효과가 있다. RBF 커널 함수의 정의는 다음과 같다.

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right) \quad (7)$$

실험에서는 커널함수로 RBF 커널을 선택하였고, 파라미터인 C , γ 의 각 데이터집합에 대한 최적 값 결정을 위해서 a grid-search 접근법을 사용하였으며, 그 결과 파라미터 범위를 $C \in \{2^{-5}, 2^{-3}, \dots, 2^{15}\}$ 와 $\gamma \in \{2^{-15}, 2^{-13}, \dots, 2^3\}$ 로 결정한다.

4.2 트리 기반 머신러닝 기법

RF(Random Forest)[6]는 배깅 접근방식을 사용하는 대표적인 머신러닝의 트리 구조 알고리즘이다. 배깅이란 bias-variance trade off를 극복하기 위해 사용되는 방법이며, Bias와 Variance는 학습 오류를 구성하는 두 요소인데 Bias가 높으면 예측결과가 실제결과와 비교해서 부정확한 것이고 Variance가 높으면 예측결과가 어떤 훈련 데이터 셋에서는 잘 맞다가 테스트 데이터 셋에서는 높은 오류를 보이는 즉 과잉적합(overfitting)의 경우이다. 배깅(Bagging)은 Bootstrap Aggregation의 약자인데 주어진 훈련 데이터 셋에서 랜덤하게 부분집합(subset)을 N 번 샘플링해서(더 정확하게는 observations과 features들을 random

하게 sampling) N 개의 예측모형을 만들어 개별 예측모형이 voting하는 방식으로 예측결과를 결정하여 Low Bias는 유지하고 High Variance는 줄이는 방법이다. RF는 이런 배깅 계열의 가장 대표적이고 예측력 좋은 알고리즘이다. 예측결과와 정확성(Low Bias)은 개별 예측모형에 쓰이는 알고리즘(decision tree)의 평균값으로 유지되는 반면 낮은 안정성(High Variance)은 Central Limit Theorem에 의해 낮아진다.

의사결정 트리(decision tree)는 flowchart처럼 노드와 가지를 따라 데이터를 분류해가는 알고리즘이다. 대표적인 결정트리 기법은 CART(Classification and Regression Trees) [2, 4]이다. CART는 지니 지수(범주형 목표변수인 경우 적용) 또는 분산의 감소량(연속형 목표변수인 경우 적용)을 이용하여 이진분리(binary split)를 수행하는 알고리즘이다.

분리를 위한 척도인 지니 지수(Gini Index)는 불순도(impurity)를 측정하는 하나의 지수이다. 임의의 한 개체가 목표변수의 i 번째 범주로부터 추출되었고, 그 개체를 목표변수의 j 번째 범주에 속한다고 오분류(misclassification)할 확률은 $P(i)P(j)$ 가 된다. 여기에서 $P(i)$ 는 각 마디에서 한 개체가 목표변수의 i 번째 범주에 속할 확률이다. 이러한 오분류 확률은 모두 더하여 아래와 같이 구할 수 있다.

$$G = \sum_{j=1}^c \sum_{i=1, i \neq j}^c P(i)P(j)$$

이는 위와 같은 분류규칙 하에서 오분류 확률의 추정치가 된다. 여기서 c 는 목표변수인 클래스 라벨의 범주의 수를 말한다. 일반적으로 CART는 범주형 목표변수에 대해서는 지니 지수를 분리 기준으로 사용한다. 지니 지수는 각 마디에서의 불순도를 측정하는 척도 중의 하나로서,

$$G = \sum_{j=1}^c P(j)(1 - P(j)) = 1 - \sum_{j=1}^c P(j)^2 = 1 - \sum_{j=1}^c (n_j/n)^2$$

와 같이 표현될 수 있다. 여기에서 n 은 그 마디에 포함되어 있는 관찰치수를 말하고, n_i 는 목표변수의 i 번째 범주에 속하는 관찰치 수를 말한다. 지니 지수는 n 개의 원소 중에서 임의로 2개를 추출하였을 때, 추출된 2개가 서로 다른 그룹에 속해있을 확률을 의미한다. 목표변수의 범주가 2개인 경우에는 지니 지수는 다음과 같이 표현될 수 있으며,

$$G = 2P(1)P(2) = 2\left(\frac{n_1}{n}\right)\left(\frac{n_2}{n}\right)$$

이는 카이제곱 통계량을 사용하는 것과 같은 결과를 갖는다. CART 알고리즘은 지니 지수를 가장 감소시켜주는 예측변수와 그 변수의 최적분리를 자식마디로 선택하는데, 지니 계수의 감소량은 다음과 같이 계산 된다.

$$\Delta G = G - \frac{n_L}{n} G_L - \frac{n_R}{n} G_R$$

여기서 n 은 부모노드의 관측치 수를 말하고, n_R 과 n_L 은 각각 자식노드의 관측치 수를 의미한다. 즉, 자식노드로 분리되었을 때의 불순도가 가장 작도록 자식노드를 형성하는 것이다. 이는 다음과 같은 자식노드에서의 불순도가 중합을 최소화하는 것과 동일하다.

$$P(L) G_L + P(R) G_R = \frac{n_L}{n} G_L + \frac{n_R}{n} G_R$$

5. 실험 평가

철도주요 부품 중 모터의 상태기반 고장진단 판별을 위해서 3장에서 기술된 계층적 구조의 두 가지 데이터 셋(3클래스 및 6개 클래스)에 대하여 함수 및 트리기반 머신러닝 기법을 평가한다.

기본 클래스 및 조합 클래스에 대한 고장진단을 위한 함수 및 트리 기반 학습모델의 예측 정확성을 평가하기 위해서 precision, recall, F1-measure를 각 클래스 예측결과로서 평가한다. 또한 전체적인 진단 모델의 성능 평가는 정확도(accuracy) 및 모델의 오차율을 나타내는 MAE(Mean Absolute Error)와 RMSE(Root Mean Squared Error)를 적용한다. 고장모델의 모든 평가 척도들의 식은 다음과 같다.

$$\text{Precision} = \frac{TP}{TP+FP},$$

$$\text{Recall} = \frac{TP}{TP+FN},$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

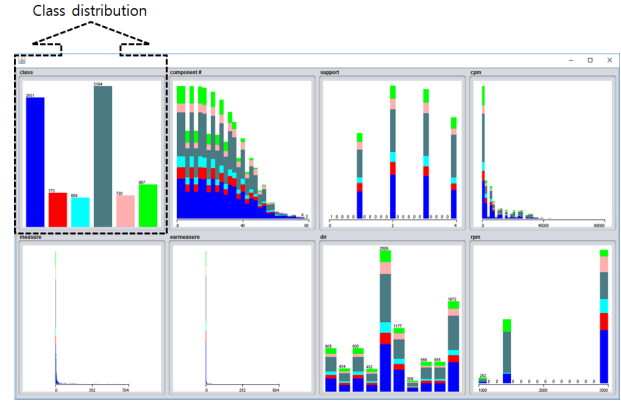
(※TP : True Positive, TN : True Negative, FP : False Positive, FN : False Negative)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i|$$

(※ f_i : prediction, y_i : true value)

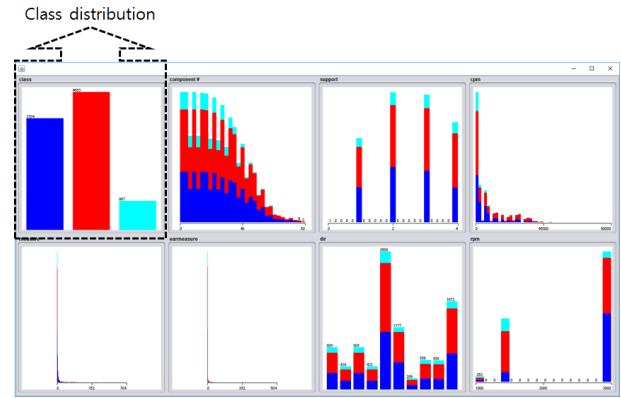
$$\text{RMSE} = \sqrt{E(\hat{\sigma}_t - \sigma_t)^2}$$

먼저 기본 클래스 6개를 포함한 데이터 셋에 대한 클래스 분포는 <Figure 5>와 같다. 분포에서 각 6개의 클래스는 불균형을 이루고 있음을 알 수 있다.



<Figure 5> Class Distribution for Basic Class Dataset

<Figure 6>은 조합 클래스에 대한 클래스 분포도이다.



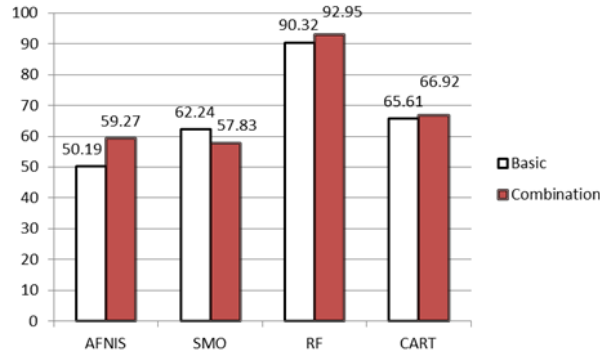
<Figure 6> Class Distribution for Combination Class Dataset

기본 클래스 데이터에 대한 각각의 진단모델의 precision, recall, F1-measure에 대한 혼잡매트릭스는 <Table 4>과 같다. 또한 <Table 5>는 조합 클래스인 2개의 고장 모드와 정상 클래스에 대한 결과이다.

<Table 4>의 경우 RF에서 클래스별 가장 최적의 분류한 결과를 볼 수 있다. 각 알고리즘별 세부적으로 C2, C3, C6 즉, faulty bearing, mechanical loosening, normal condition 클래스에 대한 각 척도값이 상대적으로 다른 클래스에 비해 낮게 측정된 것을 알 수 있다. 높은 계층구조 클래스 레벨에 대한 평가에도 역시 RF가 가장 적절한 진단 모델임을 보여주고 있으며, <Figure 3>의 구조도에 따라 고장모드를 조합하여 분류한 결과 <Table 5>에서는 정상상태 클래스 C3에 대한 척도가 낮음을 확인할 수 있다.

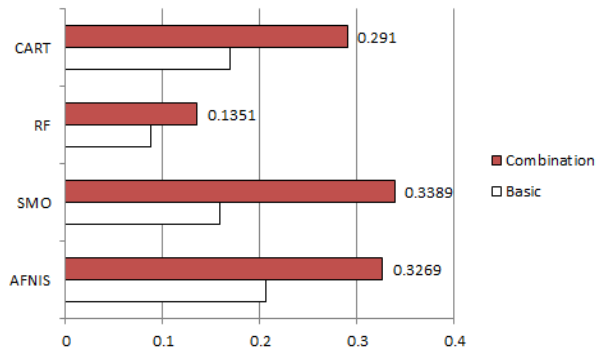
<Table 4> Confusion Metric for 6 Classes Dataset

Algorithm	Precision	Recall	F-measure	class
ANFIS	0.431	0.771	0.553	C1
	1.000	0.001	0.003	C2
	0.364	0.012	0.023	C3
	0.594	0.672	0.631	C4
	0.667	0.053	0.098	C5
	0.596	0.200	0.299	C6
SMO	0.550	0.801	0.652	C1
	0.630	0.295	0.402	C2
	0.598	0.265	0.367	C3
	0.692	0.759	0.724	C4
	0.661	0.376	0.480	C5
	0.760	0.322	0.452	C6
RF	0.902	0.935	0.918	C1
	0.875	0.845	0.860	C2
	0.875	0.849	0.862	C3
	0.922	0.928	0.925	C4
	0.911	0.857	0.883	C5
	0.880	0.846	0.862	C6
CART	0.581	0.818	0.680	C1
	0.711	0.331	0.452	C2
	0.684	0.200	0.310	C3
	0.727	0.796	0.760	C4
	0.688	0.358	0.471	C5
	0.692	0.501	0.581	C6



<Figure 7> Accuracy Comparison of Four Classification Methods

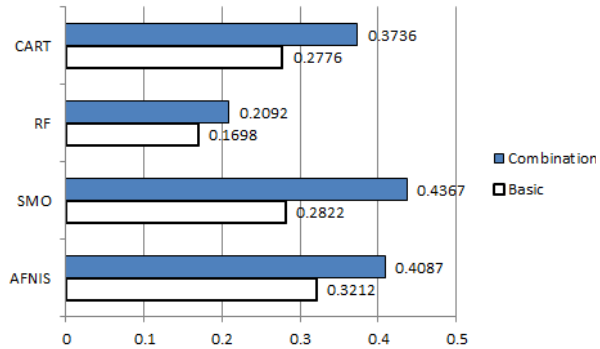
마지막으로 두 데이터 셋에 대한 각 고장진단 모델의 오류율을 측정하는 MAE와 RMSE는 <Figure 8>, <Figure 9>와 같다.



<Figure 8> MAE Comparison of Four Classification Methods

<Table 5> Confusion Metric for 3 Classes Dataset

Algorithm	Precision	Recall	F-measure	class
ANFIS	0.564	0.728	0.635	C1
	0.625	0.580	0.602	C2
	0.598	0.135	0.221	C3
SMO	0.546	0.742	0.629	C1
	0.622	0.533	0.574	C2
	0.541	0.165	0.253	C3
RF	0.928	0.942	0.935	C1
	0.937	0.939	0.938	C2
	0.895	0.836	0.864	C3
CART	0.625	0.760	0.686	C1
	0.718	0.649	0.681	C2
	0.666	0.418	0.513	C3



<Figure 9> RMSE Comparison of Four Classification Methods

6. 결론 및 향후 연구

전체적인 고장진단 모델의 성능을 직관적으로 평가할 수 있는 정확도 비교는 <Figure 7>에 나타내었다. 두 가지 데이터셋에 대해 트리 기반의 RF 기법에 가장 좋은 성능 (약 90% 이상)을 보임을 알 수 있다.

국내 철도차량 분야에서의 유지보수 시스템 효율화의 핵심은 현재 운행 중인 차량의 부품 손상에 대한 실시간 상태 진단을 위한 모니터링 기술 및 정확한 고장진단의 판별이며 불필요하게 발생하는 유지보수 비용의 감소가

필수적이다. 이를 위해서, 본 논문에서는 차량 내에 부착된 센서로부터 다양한 고장진단 파라미터를 추출하여 유지보수 현장 작업자의 검수 전에 그 상태를 예측할 수 있는 머신러닝 기반의 고장진단 모델의 설계 및 알고리즘을 평가하였다. 고장진단 대상 부품은 모터의 펌프에 대한 베어링 고장, 불균형, 조인트 문제, 기계적 풀림 현상 등의 부품 상태에 대한 정상 조건의 머신러닝 기반 분류이다. 총 9,254건의 센서 데이터에서 기계적 결합의 계층적 구조를 고려하여 기본 및 조합 데이터셋에 대한 머신러닝 기법의 평가를 수행하였다. 적용된 머신러닝 기법은 함수 및 트리 기반이며, 각 카테고리별 대표 알고리즘을 평가하였다. 실험결과, 두 데이터셋에 대하여 RF 기법에 가장 적합한 분류 모델임을 증명하였다.

본 연구를 토대로 KTX 차량 대상 부품을 확대하여 차륜 및 차축베어링, 트리포드 등 다양한 핵심 부품에 대한 센서 데이터 확보와 최적의 머신러닝 기법 적용을 통한 정확한 고장진단 모델 선정 및 고장진단 결과들을 활용한 부품의 잔존수명 예지 알고리즘 개발이 향후 연구이다.

Acknowledgement

This work has been supported by the railway technology Research project of MOLIT(Project No.15RTRP-B09 1404-02, Project name: Development of a early failure detection and maintenance technology for core parts of the rolling stock at onboard and wayside).

References

- [1] A book published by Korail, A study on a method of calculating the lifespan of freight train axle bearings, 2014.
- [2] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., Classification and regression trees, Belmont : Wadsworth, 1984.
- [3] Byeon, S.-K., Kang, C.-W., and Sim, S.-B., Defect Type Prediction Method in Manufacturing Process Using Data Mining Technique, *Journal of the Society of Korea Industrial and Systems Engineering*, 2004, Vol. 27, No. 2, pp. 10-16.
- [4] Choi, J. and Seo, D., Decision Trees and Its Applications, *Statistical Analysis Research*, 1999, Vol. 4, No. 1, pp. 61-83.
- [5] Choi, S.J., Kim M.H., and Kim, Y.H., Case study on the maintenance costs structure analysis for KTX high speed rolling stock system, *Proceedings of the Railway Society*, 2015 Korea, KSR2015A263.
- [6] Di, X., Han, T., and Yang, B.S., Random Forest Classifier for Machine Fault Diagnosis, *Proceedings of the Korean Society of Mechanical Engineers*, 2006, No.11, pp. 11-16.
- [7] Hayati, M., Seifi, M., and Rezaeivol, A., Double Gate MOSFET Modeling Based on Adaptive Neuro-Fuzzy Inference System for Nanoscale Circuit Simulation, *ETRI Journal*, 2010, Vol. 32, No.4, pp. 530-539.
- [8] Kim et al., The Life cycle cost estimation using the maintenance information of a propulsion control system in the high speed train(KTX-1), *The Transactions of The Korean Institute of Electrical Engineers*, 2011, Vol. 60, No. 11, pp. 2176-2181.
- [9] Kim, H.G. and Cho, H.S., A Production and Preventive Maintenance Policy with Two Types of Failures, *Journal of the Korean Society for Quality Management*, 2002, Vol. 30, No. 3, pp. 53-65.
- [10] Korail, Development of a early failure detection and maintenance technology for core parts of the rolling stock at onboard and wayside, Land and Transport Technology Research and Development Plan, 2015.
- [11] Shevade et al., Improvements to the SMO algorithm for SVM regression, *IEEE Transactions on Neural Networks*, 2000, Vol. 11, No. 5, pp. 1188-1193.
- [12] Thiago et al., MLP and ANFIS Applied to the Prediction of Hole Diameters in the Drilling Process : Artificial Neural Networks-Architectures and Applications, DOI : 10.5772/51629, 2013.

ORCID

Hoon Jung | <http://orcid.org/0000-0003-3962-9562>
 Ju-Won Kim | <http://orcid.org/0000-0003-2010-4491>