

차륜 및 차축베어링 고장진단을 위한 빅데이터 기반 머신러닝 기법 연구

정훈*, 박문성

한국전자통신연구원 초연결통신연구소 우정기술연구센터

A Study of Big data-based Machine Learning Techniques for Wheel and Bearing Fault Diagnosis

Hoon Jung*, Moonsung Park

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

요약 본 철도 유지보수 산업의 효율화를 위해서는 핵심부품의 적시 관리를 통한 부품 가동률 향상 및 철도 운행의 안정성 향상이 필요하다. 또한 유지보수 시스템 고숙화에 따른 신뢰성 향상과 핵심부품의 유지보수 비용 절감의 두 가지 측면을 모두 만족시키기 위해, 부품 이력관리와 대규모 빅데이터의 자동화된 분석 기술을 활용한 부품 상태 진단 기술 수요가 증가하고 있다. 이 논문에서는 철도차량의 차상 및 지상 장치로부터 발생하는 실시간 빅데이터 수집, 처리, 분석을 위해서 빅데이터 플랫폼 기반의 철도차량 부품의 상태 데이터 관리시스템을 개발하였으며, 이 시스템의 활용으로 철도차량의 부품 상태정보 및 시스템 리소스에 대한 실시간 모니터링이 가능하다. 또한 빅데이터 플랫폼으로부터 수집된 상태 데이터를 기반으로 분산/병렬처리 및 자동화된 부품 고장진단이 가능한 머신러닝 기법을 제안하였다. 실험결과, 분산/병렬처리 기술이 적용된 알고리즘의 실행시간 단축을 아마존 웹서비스의 가상 인스턴스 생성 시스템을 통해 증명하였으며, random forest 머신러닝 기법을 활용한 고장 진단 모델의 베어링 및 차륜 부품에 대한 상태 예측 정확도가 83%임을 확인하였다.

Abstract Increasing the operation rate of components and stabilizing the operation through timely management of the core parts are crucial for improving the efficiency of the railroad maintenance industry. The demand for diagnosis technology to assess the condition of rolling stock components, which employs history management and automated big data analysis, has increased to satisfy both aspects of increasing reliability and reducing the maintenance cost of the core components to cope with the trend of rapid maintenance. This study developed a big data platform-based system to manage the rolling stock component condition to acquire, process, and analyze the big data generated at onboard and wayside devices of railroad cars in real time. The system can monitor the conditions of the railroad car component and system resources in real time. The study also proposed a machine learning technique that enabled the distributed and parallel processing of the acquired big data and automatic component fault diagnosis. The test, which used the virtual instance generation system of the Amazon Web Service, proved that the algorithm applying the distributed and parallel technology decreased the runtime and confirmed the fault diagnosis model utilizing the random forest machine learning for predicting the condition of the bearing and wheel parts with 83% accuracy.

Keywords : Big data, Bearing/Wheel, Fault Diagnosis, Machine Learning, Rolling Stock

본 연구는 국토교통부 철도기술연구사업의 연구비지원(17RTRP-B103990-04, 철도차량 상태 데이터 분석기반 유지보수 지원시스템 개발)에 의해 수행되었습니다.

*Corresponding Author : Hoon Jung(ETRI)

Tel: +82-42-860-6379 email: hoonjung@etri.re.kr

Received December 6, 2017

Revised December 21, 2017

Accepted January 5, 2018

Published January 31, 2018

1. 서론

철도차량의 고속화 및 노선확대로 이용객수가 증가함에 따라 철도 차량에 이상이 발생할 경우 국민의 안전에 위협을 초래할 수 있을 뿐만 아니라 운행지연에 따른 고객 불편, 수송량 감소 문제가 발생된다. 이러한 철도운행의 사용중단 상태를 야기하는 부품의 손상과 불필요하게 발생하는 유지보수 비용을 감소하기 위해 철도차량의 유지보수 기술은 유럽 선진국의 주도로 검사주기 기반 예방정비로부터 예지정비 기술로 발전하고 있으며, 시스템 고속화에 따른 신뢰성 및 안전성 향상과 유지보수 비용 절감이라는 두 가지 측면을 모두 충족시키기 위한 기술 수요가 증가하고 있다. 철도 유지보수 분야는 2004년 고속철도의 도입에 따른 급격한 철도기술의 발전에 대한 역효과인 원천기술 부재로 어려움을 겪어왔고, 미 이전 기술획득을 위한 컨설팅 비용 등 철도운영사의 경영에도 부정적인 영향을 미치고 있다.

국내 철도 차량의 유지보수 시스템은 정보화 시스템 구축을 통해 통계에 기반한 RCM(Reliability Centered Maintenance) 기술을 활용할 수 있는 수준에 도달하였으나 우발 고장에 대한 대응이 어렵기 때문에 승객 안전과 운행지연의 직접적인 원인이 되는 모터블럭, 차축베어링, 차륜 등과 같은 핵심 장치에 대해서는 운행 중 고장 예방을 위한 지속적인 기술 개발이 필요하다[1, 2]. 이렇듯 핵심부품의 적시관리를 통한 부품 가동률 향상과 철도 운행의 안정성 향상을 위해서는 부품 이력관리와 대용량 데이터의 자동화된 분석 기술을 활용한 부품 상태진단 기술 개발이 요구되며, 실시간 빅데이터 발생, 처리 및 머신러닝 기술 발전으로 향후 철도산업 분야 4차 산업혁명 출현 전망에 따라 전문가적 지능을 갖는 머신러닝 및 인공지능 엔진 원천기술 연구가 필요한 실정이다. 독일 지멘스, 프랑스 알스톰 등 해외 선진 기업은 빅데이터 및 인공지능을 활용한 철도 유지관리와 교통수단 연계 기술 등 4차 산업혁명에 따른 혁신 기술을 개발 중에 있다. 국내 철도 유지보수 효율화를 위해서는 철도 차량 내 핵심부품에 대한 이해와 상태 감지, 분석 및 판단 기술에 대한 연구가 필요하며, 수집된 대용량의 데이터를 저장, 추출, 분류 및 관리하는 빅데이터 기술이 접목되어야 한다. 또한 각 기술의 융합을 위해서는 인메모리 기반의 분산/병렬처리와 같은 고속 데이터 처리 기법이 필수적이다.

이러한 핵심기술들을 효과적으로 융합하기 위해서 이 논문에서는 다음과 같은 세 가지의 핵심 연구를 수행한다.

- 빅데이터 분석 플랫폼 기반의 철도차량(고속열차 및 물류화물차량) 핵심부품 상태정보 관리시스템을 개발한다.
- 머신러닝 기반의 철도차량 핵심 부품의 모니터링 및 유지보수 시스템 개발은 철도차량 핵심부품의 이상 유무를 사전에 감지하여, 장치의 고장을 미연에 방지할 수 있는 자동화된 유지보수 지원을 위해 머신러닝의 감독학습 기반 고장진단 시스템을 제안한다.
- 대용량 빅데이터를 활용하는 머신러닝 기법의 효율적인 고장진단 모델 생성을 위해 분산/병렬처리가 가능한 아파치 Spark 및 맵 패턴을 적용한 기법을 제안한다.

이 논문에서 제안하는 머신러닝 기반 고장진단 시스템의 워크플로우는 Fig. 1과 같다.

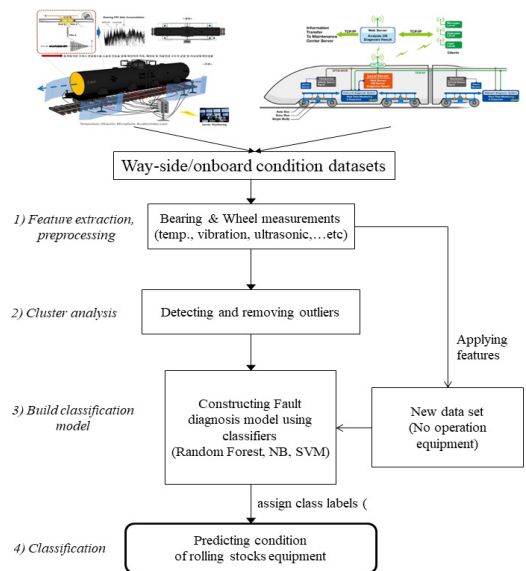


Fig. 1. Machine Learning work flow for detecting bearing & wheel failure.

- 빅데이터 플랫폼에 분산 저장된 물류차량 지상(way-side) 장치 및 고속열차 차상(onboard) 장치로부터 수집된 데이터에서 고장진단을 위한 인자를 추출한다.

- 고장진단 인자들 중 일부 이상치(outlier) 데이터의 발견 및 제거를 위해 군집화(clustering) 기법을 적용하며, 형성된 군집에 포함되지 않은 데이터를 제거한다.
- 머신러닝의 분류(classification) 기법을 이용하여 고장진단 모델을 생성하고, 아직 부품상태가 대기 중인 새로운 데이터에 대해 분류(클래스 라벨 할당을 통한 고장진단)를 수행한다.

논문의 효율적인 이해를 위한 2장 본론의 내용 구성은 다음과 같다.

2.1절에서는 빅데이터 기반 상태데이터 관리시스템의 설계 및 구현 내용을 기술하고, 2.2절에서는 철도차량 주요부품인 베어링(bearing) 및 차륜(wheel)의 고장진단을 위한 mapReduce 기반의 머신러닝 기법을 기술한다. 2.3절부터 2.5절까지는 본 논문에서 제안하는 분산/병렬 처리 프로그래밍 및 고장진단 모델의 성능평가 결과를 제시하고, 마지막으로 3장에서 결론을 맺는다.

2. 본론

2.1 빅데이터 기반 상태데이터 관리시스템

본 절에서는 대규모 철도차량 부품 상태 데이터 수집부터 관리 및 시각화 기능 제공을 위해서 빅데이터 분석 플랫폼을 활용한 상태 데이터 관리 시스템을 설계한다. 실시간 상태정보 모니터링을 위한 관리 시스템은 철

도차량의 부품 상태정보를 실시간으로 수집하여 분석할 수 있는 기능을 제공하고 웹 로그 및 서버 자원 로그를 대상으로 실시간으로 모니터링을 할 수 있는 시스템을 개발한다. 전체 시스템을 분석할 수 있는 대시보드와 개별 서비스 단위의 대시보드, 부품의 고장진단 및 시스템의 장애 사전 감지를 위한 대시보드 등으로 구성된다.

시스템의 설계는 3계층 구조로 설계한다. 저장계층에서는 물류차량/고속열차에서 측정된 상태정보를 분산구조 형태로 저장하며, 백업을 위한 기존 DB 시스템을 이용한다. 두 번째 계층에서는 데이터 처리 및 분석을 위한 계층으로 상태정보의 고장진단 인자 추출, 통계분석 및 이벤트 알림/경고 기능을 제공하며, 머신러닝 S/W를 탑재하고 있다. 마지막 세 번째 계층에서는 사용자에게 상태데이터 관리 기능을 제공하기 위한 시각화 기능 및 리소스 관리 기능을 포함한다. 철도차량 핵심부품 상태 데이터 관리시스템의 구조도는 Fig. 2와 같다. 시스템의 기능은 대시보드를 포함 리소스 관리로 구분되며, 현장 사용자들을 위한 UI가 구성된다.

전체 시스템 실시간 모니터링 분석을 위한 대시보드는 다음과 같다.

- 1) 물류차량 대상의 지상 장치 기반 베어링 및 차륜(휠) 부품의 상태 측정 정보 및 통계 지표를 표출한다(Fig. 3).
- 2) 시스템 운영서버의 리소스 관제를 위한 각 요소(CPU, RAM 호스트수, CPU/RAM/DISK 사용률 및 트렌드)의 사용률 분석 기능을 제공한다(Fig. 4).

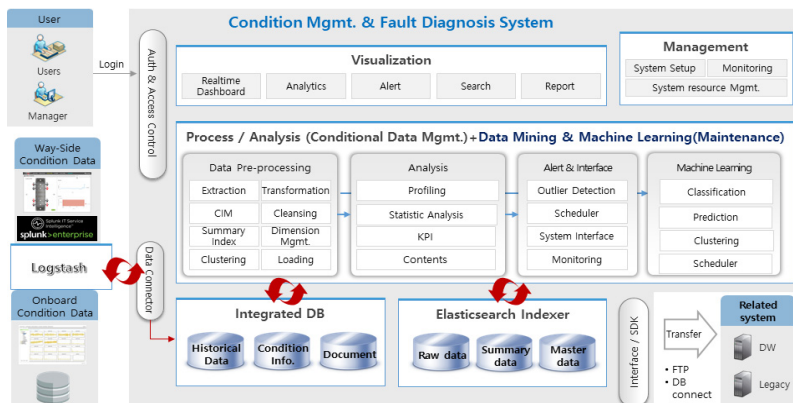


Fig. 2. Condition mgmt. & fault diagnosis system architecture.

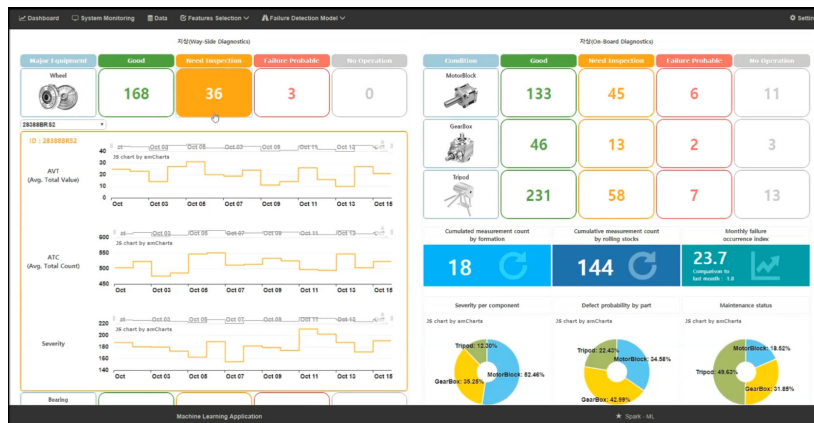


Fig. 3. Condition mgmt. & fault diagnosis system (condition information & statistics index).



Fig. 4. Condition mgmt. & fault diagnosis System (utilization).

2.2 MapReduce 기반 머신러닝 기법

MapReduce [3, 4]의 기본 자료구조는 key-value의 쌍으로 구성되며, key와 value는 정수, 실수, 문자열, 바이트열 또는 임의의 복잡한 자료구조로서 정의될 수 있다. MapReduce의 입력 데이터는 MapReduce 작업을 위해 특정 크기의 작은 블록으로 나누는 작업이 선행되어야 한다. 다음으로 특정 크기로 나누어진 데이터 블록에 대한 배포(distribute)를 수행한다. 이후 map 함수와 reduce 함수 수행을 위한 mapper 작업과 reducer 작업이 독립적으로 시작한다. Mapper는 주어진 데이터로 읽고 처리를 위한 기본 자료구조인 key-value의 쌍을 생성한다. Reducer는 동일한 중간 key와 관련된 모든 값을

key-value의 쌍으로 출력을 생성한다. 이 과정에서는 중간값에 대한 group by 연산이 수행된다. Mapper 작업과 reducer 작업은 데이터를 병렬적으로 처리한다. Fig 5는 mapReduce 처리 과정을 나타낸다. mapper와 reducer사이의 partitioner는 map 작업 후 생성된 <key, value> 쌍의 출력을 특정 reducer 작업으로 보내져야 하는데 레코드 키의 키값을 해싱해서 그 해싱값을 reduce 작업의 수로 나누어 그 레코드가 어느 reducer 작업으로 보내질지를 결정하는 역할을 한다. 즉 같은 키를 갖는 레코드들은 같은 reducer 작업으로 보내지게 된다.

철도차량 핵심부품의 상태정보를 활용하여 자동화된 고장진단 모델을 생성하기 위해, 머신러닝의 비감독학습

및 감독학습 기법이 활용된다. 먼저, 차상 및 지상 장치로부터 수집된 데이터를 전처리하여 부품의 고장진단에 활용될 수 있는 인자들(진동수, 온도, 속도 등)을 추출하며, 이를 상태 데이터라고 정의한다. 이 논문에서 머신러닝 기법의 입력으로 사용될 상태 데이터의 모든 특징(고장인자) 및 설명은 Table 1에 기술하였다.

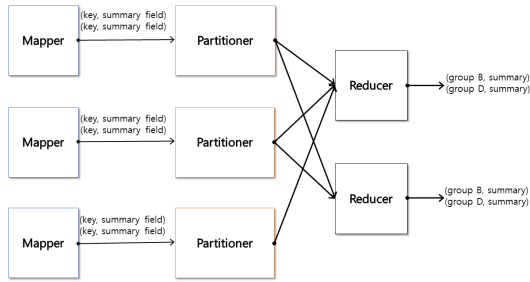


Fig. 5. The structure of the mapReduce pattern.

Table 1. The extracted features (condition info.).

| Feature | Description |
|-------------|---|
| type | Bearing or wheel |
| location | Bearing or wheel location number |
| velocity | Speed measurement |
| vib | Vibration measurement |
| load | Load measurement |
| temp | Bearing or wheel temperature |
| sound | Noise measurement value |
| RUL | Remaining useful life |
| Class Label | {Good, Need Inspection, Failure Probable} |

2.3 분산처리 가능한 밀도기반 군집화 기법

상태데이터는 주로 센서들을 활용하여 측정된 값을 전처리하여 획득한 것으로 일부 이상치 값을 포함할 수 있으며, 이 데이터로 인해 감독학습 알고리즘을 이용한 고장진단 모델의 분류 정확도에 악영향을 줄 수 있다. 따라서, 고장진단 모델 생성 전에 이는 반드시 제거되어야 하며, 일반적으로 통계적 및 비감독학습 방법이 사용된다. 이 논문에서는 비감독학습 기법으로 밀도 기반 DBSCAN 군집화를 수행하여 이상치를 탐지한다. DBSCAN 알고리즘의 분산/병렬처리가 가능하도록 mapReduce 패턴이 적용된다[5].

첫 번째 단계에서는 전체 데이터를 각각의 서브셋으로 나누는 작업을 수행하며, 각각의 서브셋에서의 로컬 클러스터링은 독립적으로 수행한다. 다음으로 서브셋에

서의 로컬 클러스터링은 map 함수를 통해 수행되며, 클러스터 구성은 기존의 DBSCAN 방식과 동일하다. 마지막으로 전체 클러스터 구성을 위해서 통합 작업이 수행된다. 먼저, 교차되는 모든 로컬 클러스터를 찾은 후에, 전체 클러스터를 구성한다. 이후 로컬 클러스터가 서로 다른 클러스터로 중복되어 멤버로 구성될 경우, relabel data를 통하여 조정하게 된다. MR-based DBSCAN 알고리즘의 프로세스는 Fig. 6과 같다. 본 논문에서는 철도 차량 부품의 상태 데이터에 대한 전처리 단계로서 [5]에서 제안한 군집화 기법을 적용하여, 부품의 특정 상태에도 포함되지 않거나 이상치 값을 포함한 데이터를 제거하는데 활용하고자 한다. 또한, 분산처리 기법을 적용함으로써 대규모 실시간 수집되는 철도차량 부품 정보인 빅데이터의 처리 능력을 기존방식과 비교하여 평가한다.

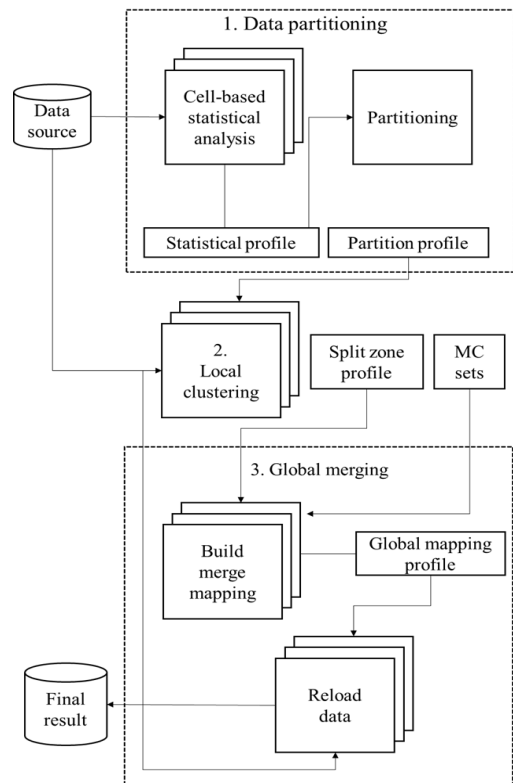


Fig. 6. An overview of MR-based DBSCAN method.

2.4 분산처리 가능한 Random Forest를 적용한 고장진단 모델

고장진단 모델 생성을 위한 감독학습 기법은 논문 [6]에서 검토를 하였으며, 그 결과를 바탕으로 본 절에서는

random forest 알고리즘을 사용한다.

RF(Random Forest)는 앙상블 학습 기법을 사용하는 대표적인 트리 구조의 감독학습 모델이다. 일반적으로 RF는 배깅(bagging) 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.

배깅은 Bootstrap Aggregation의 약자인데 주어진 훈련 데이터 셋에서 랜덤하게 부분집합(subset)을 N번 샘플링해서(더 정확하게는 observations과 features들을 random하게 sampling) N개의 예측모형을 만들어 개별 예측모형이 voting하는 방식으로 예측결과를 결정하여 Low Bias는 유지하고 High Variance는 줄이는 방법이다. RF는 이런 배깅 계열의 가장 대표적이고 예측력이 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[6]. 전처리된 상태정보로 구성된 대규모 훈련데이터에 대해 기존의 RF 방식은 모델 생성 시간이 오래 걸리는 문제가 발생한다. 따라서, 비감독학습과 동일하게 분산/병렬처리가 가능한 MR(Map-Reduce) 패턴 적용이 필요하다. 논문 [7]에서는 클래스 분포가 불균형을 이루는 빅데이터의 클래스 분류를 위한 MR-based RF 알고리즘을 제안하고 있으며, 기존 알고리즘에 분산처리를 위한 분할정복 기법인 map, reduce 함수를 이용한다.

학습 단계에서의 중요한 map 프로시저는 Fig. 6와 같다. 먼저 준비단계로 샘플링을 통해 추출된 공유 훈련셋으로부터 마스터 머신에서 공유 포리스트를 생성한다. 공유 훈련셋과 공유 포리스트는 분산처리에 의해 생성된 학습기에 의한 과잉적합(overfit) 문제를 최소화하기 위함이다[8]. 다음으로 학습을 위한 전체 훈련데이터는 마스터 머신에 저장되며, 이 학습데이터는 하위 머신들로 할당되며, 이때 동시에 미리 생성된 공유 훈련셋과 공유 포리스트를 각각의 하위 머신들에 배포한다. 각각의 하위 머신들에서는 공변량 손실(covariate loss) 계산을 통해 트리를 학습시키며, 각 머신의 트리가 목적 포리스트

를 위한 후보 포리스트들로 등록된다. 목적 포리스트는 공변량 손실을 통해 갱신되며, 후보 공유 포리스트가 가장 많이 생성될 때까지 반복된다. 이후 최종적으로 후보 중 절반이 목적 포리스트로 선택되어진다.

2.5 실험 및 성능 평가

철도차량 주요부품 중 베어링 및 차륜의 상태 기반 고장진단 판별을 위해서 비감독학습을 통한 전처리를 수행한 후, 감독학습 기법인 RF를 적용하여 부품의 상태를 분류한다.

먼저, 빅데이터 분석을 위한 분산/병렬처리를 위한 mapReduce 기반 알고리즘의 실행속도 평가를 위해서 아마존 웹서비스(AWS)의 EMR 분산/병렬처리 실험환경을 이용한다. EMR에 개발 알고리즘과 베어링 및 차륜 데이터셋(총 25개 필드를 갖는 16,000건 데이터 파일)을 업로드한 후, 가상 인스턴스(EC2)를 1에서부터 최대 16가지 생성하여 단독머신에서 실행할 때의 시간과 2, 4, 8, 16대의 머신에서 동시에 분산/병렬처리 할 때의 실행시간 비교를 수행하였다. EMR 구성도는 Fig. 7과 같다.

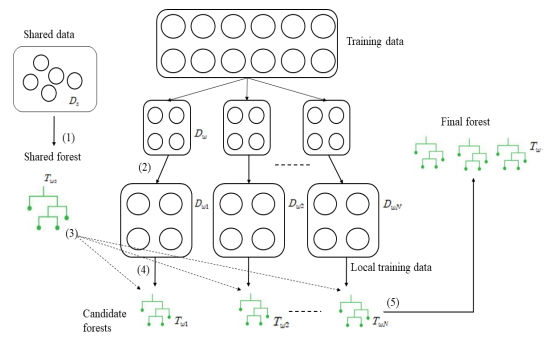


Fig. 7. An overview of Map operation for RF method.

실험에 사용된 데이터는 철도차량의 기계적 결함 부품인 베어링과 차륜의 상태정보를 활용한다. 총 데이터 건수는 246이며, 8개의 특징벡터와 1개의 클래스 속성 필드로 구성된다.

AWS의 가상인스턴스를 생성에 따라 분산/병렬처리 과정은 Fig. 8과 같다. Fig. 9에서는 노드 4개와 최대 생성 가능한 노드 수인 16개를 이용하여 각 프로세스별 분산/병렬처리 상황을 확인 할 있는 화면을 제공한다.

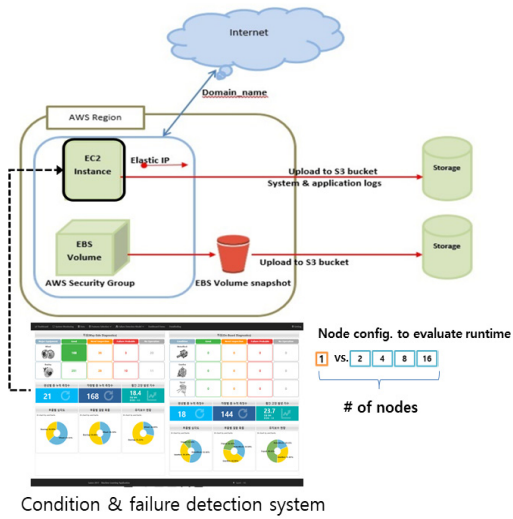


Fig. 8. AWS virtual instants generation.

Fig. 10은 단일노드와 다수의 노드 생성에 따른 분산/병렬처리 기법의 실행시간 비교 결과이다. 최소 1개의 노드인 단일 실행에서는 60초 만에 알고리즘이 종료되어 결과를 출력한다. 그러나 노드 2개 실행 시에는 약 1/2 정도로 실행시간이 줄어들며, 노드 16개 구성 시에 분산/병렬처리 되어 약 9.1초에 알고리즘이 수행됨을 확인할 수 있다.

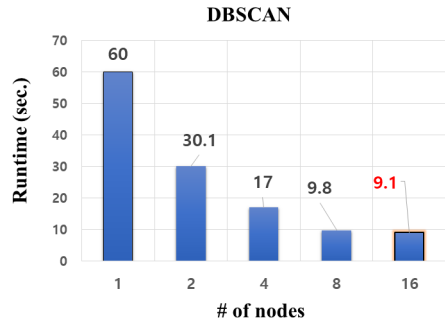


Fig. 10. A comparison of runtime for DBSCAN algorithm.

Fig. 11은 DBSCAN 군집화를 이용하여 데이터에 포함된 이상치 데이터를 추출한 결과이다. cluster - 1로 표시된 객체들은 모두 이상치 값이다.

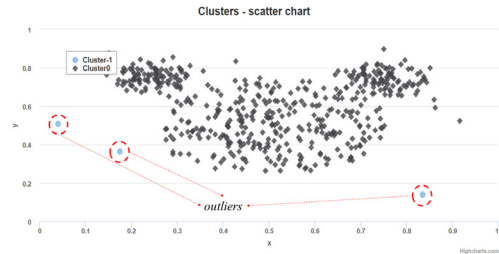


Fig. 11. Outlier detection using DBSCAN algorithm.

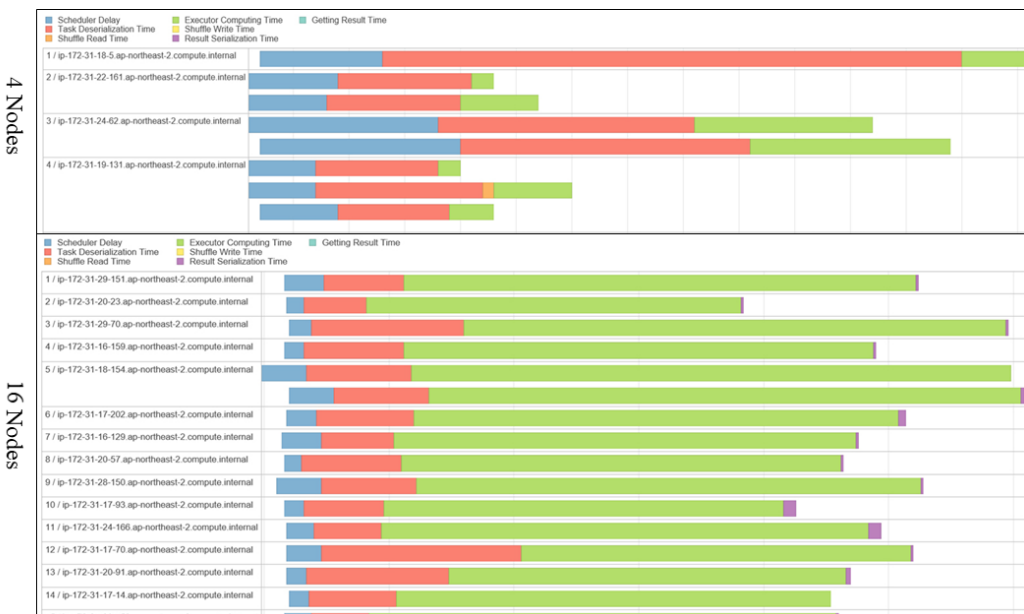


Fig. 9. A screenshot of distributed processing using multiple nodes.

DBSCAN과 동일하게 분류/예측 기법인 RF에 대해서도 분산/병렬처리 프로그래밍의 성능평가를 수행하였다. 단, 분류모델의 경우, 훈련 및 테스트 셋으로 원 데이터를 나누어 훈련셋을 이용하여 학습을 시키고 테스트 셋으로 모델의 정확도를 평가한다. 본 실험에서는 학습단계의 실행시간만을 평가하며, 원 데이터의 70%인 189건의 훈련데이터를 사용하였다. 실험결과 단독실행일 경우, 6초에 학습이 종료되며, 노드 16개를 구성하여 분산/병렬처리 할 경우 2.3초만에 학습이 완료됨을 확인하였다(Fig. 12).

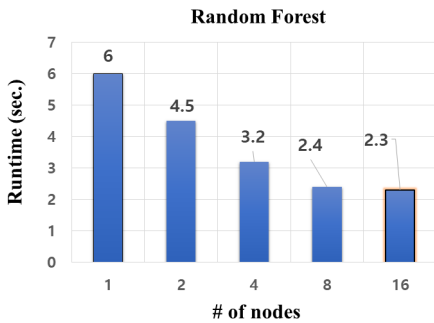


Fig. 12. A comparison of runtime for Random Forest algorithm.

마지막 제안한 기법의 성능평가로서 RF 분류기를 활용한 고장진단 모델의 정확도 평가를 수행한다. 3가지 클래스 상태(Good, Need Inspection, Failure Probability)에 대한 RF 기반 고장진단 학습모델의 예측 정확성을 평가하기 위해서 총 3가지의 평가 지표인 precision, recall, F1-value를 각 클래스 예측결과로서 평가한다. 또한 전체적인 진단 모델의 성능 평가는 정확도 지표를 통해 RF 알고리즘의 최적 성능을 제시한다. RF 기반의 베어링 및 차륜 부품 대상 고장모델의 모든 평가 척도들의 식은 다음과 같다.

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

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

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

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

성능평가 결과 정보는 Fig 13과 같이 3개 그룹으로 개발된 프로그램의 UI로 호출된다. 먼저, 1) 평가지표 결과에서는 정확도 83%이며, 정밀도, 민감도와의 조합평가 지표인 F1-value는 81%를 보였다. 또한 3가지 클래스에 대한 테스트셋을 활용한 평가에서 2) 혼잡 매트릭스를 호출하여 특정 클래스의 오분류와 정분류된 정보를 확인할 수 있다. 3)번 영역에서는 예측된 클래스 대비 실제 클래스 매칭 정보를 보여준다(초록색으로 표시된 것이 정확히 고장상태를 예측한 것이며, 붉은색 표시된 셀은 잘못 예측된 결과이다).

3. 결론

국내 철도차량 운영분야에서의 지능형 유지보수 시스템 효율화의 핵심 사항은 운행 중인 철도차량의 주요부품의 결함에 대한 실시간 고장진단을 위한 모니터링 기술 및 정확한 고장진단의 결과를 활용하는 것으로서, 기존의 예방적 정비 체계에서 적시정비 또는 예지정비 체계의 구축으로 불필요하게 발생하는 유지보수 비용의 감소가 필수적이다.

이를 위해서, 본 논문에서는 차상 및 지상에 부착된 센서 장치로부터 다양한 고장진단 파라미터를 추출하여 유지보수 현장 작업자의 검수 전에 그 상태를 실시간 모니터링과 결함 상태를 예측할 수 있는 빅데이터 기반 상태데이터 관리시스템과 머신러닝 기반의 유지보수지원 시스템을 개발하였다. 또한 향후 축적될 대용량 상태정보의 효율적인 유지보수지원 시스템의 결함진단을 위해서 분산/병렬처리가 가능한 mapReduce 기반의 알고리즘을 활용하였다.

고장진단 대상은 기계적 부품인 베어링 고장, 차륜 파손의 상태에 대한 머신러닝 기반 분류이다. 총 256건의 상태 데이터에서 기계적 결함의 여부 판단 기준인 3가지 클래스에 대한 머신러닝 기법의 평가를 수행하였다. 적용된 머신러닝 기법은 데이터 전처리를 위해 DBSCAN 군집화가 사용되었으며, 고장진단을 위한 감독학습 모델로서 random forest 기법을 적용하였다.

본 연구 결과를 바탕으로 한 향후 연구로는 고속열차 및 물류차량 대상 결함 진단 부품을 확대하여 기계적 및 전기적 부품의 상태정보에 대한 최적의 고장진단 알고리즘을 개발하는 것이다.



Fig. 13. A screenshot of performance evaluation results.

References

- [1] S. Ha, S. Chang, and W. Yoo, "Development of the Preventive Maintenance System for an Urban Transit," *Journal of the Society of Korea Indus. and Sys. Eng.*, vol. 30, no. 1, pp. 1-7, 2007.
- [2] Y. Cho, J. Kim, and J. Kim, "Performance Improvement of Bearing Fault Diagnosis Using a Real-time Training Method," *Journal of Multimedia Services Convergent*, vol. 7, no. 4, pp. 551-559, 2017.
DOI: <http://dx.doi.org/10.14257/AJMAHS.2017.04.33>
- [3] F. Li, B. C. Ooi, M. T. Özsü, and S. W. Zhejiang, "Distributed data management using MapReduce," *ACM Computing Surveys*, vol. 6, no. 3, 2014.
DOI: <https://doi.org/10.1145/2503009>
- [4] H. G. Lee, Y. H. Choi, J. Hoon, and Y. H. Shin, "Spatial Subspace Projected Clustering Method & MapReduce based Temporal Mining for Direct Marketing Service in Korea Post," *ETRI Journal*, vol. 37, no. 2, 2015.
DOI: <https://doi.org/10.4218/etrij.15.2314.0068>
- [5] Y. HE, H. TAN, W. LUO, S. FENG, and J. FAN, "MR-DBSCAN: a scalable MapReduce -based DBSCAN algorithm for heavily skewed data," *Front. Comput. Sci.*, vol. 8, no. 1, pp. 83-99, 2014.
DOI: <https://doi.org/10.1007/s11704-013-3158-3>
- [6] H. Jung, and J. Kim, "A Machine Learning Approach for Mechanical Motor Fault Diagnosis," *Society of Korea Indus. and Sys. Eng.*, vol. 40, no. 1, pp. 57-64, 2017.
DOI: <https://doi.org/10.11627/jkise.2017.40.1.057>
- [7] S. Rio, V. Lopez, J. M. Benitez, and F. Herrera, "On the use of MapReduce for imbalanced big data using Random Forest," *Journal of Information Sciences*, vol. 285, pp. 112-137, 2014.
DOI: <https://doi.org/10.1016/j.ins.2014.03.043>
- [8] S. Tsuchiya, R. Yumiba, Y. Yamauchi, T. Yamashita, and H. Fujiyoshi, "Transfer forest based on covariate shift," *Proc. of IAPR Asian Conf. on Pattern Recognition*, 2015.
DOI: <https://doi.org/10.1109/ACPR.2015.7486605>

정 훈(Hoon Jung)

[정회원]



- 1997년 5월 : Iowa State University
산업공학과 (석사)
- 2001년 8월 : University of Missouri
산업공학과 (박사)
- 2002년 4월 ~ 2011년 2월 : ETRI
선임연구원
- 2011년 3월 ~ 2016년 12월 :
ETRI 책임연구원

• 2017년 1월 ~ 현재 : ETRI 센터장

<관심분야>

물류 정보화/자동화 시스템, 고장진단, 드론시스템

박 문 성(MoonSung Park)

[정회원]



- 1993년 2월 : 숭실대학교 전기전자
공학과 (석사)
- 2003년 2월 : 배재대학교 컴퓨터공
학과 (박사)
- 1983년 6월 ~ 현재 : ETRI 책임연
구원

<관심분야>

물류 정보화/자동화 시스템, 고장진단, 드론시스템