

# 하둡 및 스파크 기반 빅데이터 플랫폼을 이용한 선박 운항 효율 이상 상태 분석

이태현\*, 유은섭\*, 박개명\*\*, 유성상\*\*, 박진표\*\*\*, 문두환\*.#

\*경북대학교 정밀기계공학과, \*\*한국선급e-Navigation TFT, \*\*\* (주)소프트힐스

## Detection of Abnormal Ship Operation using a Big Data Platform based on Hadoop and Spark

Taehyeon Lee\*, Eun-seop Yu\*, Kaemyoung Park\*\*, Seongsang Yu\*\*, Jinpyo Park\*\*\*, Duhwan Mun\*.#

\*Dept. of Precision Mechanical Engineering, Kyungpook Nat' l UNIV., \*\*e-Navigation TFT, Korean Register of Shipping, \*\*\* Softhills

(Received 26 March 2019; received in revised form 15 April 2019; accepted 13 May 2019)

### ABSTRACT

To reduce emissions of marine pollutants, regulations are being tightened around the world. In the shipbuilding and shipping industries, various countermeasures are being put forward. As there are limits to applying countermeasures to ships already in operation, however, it is necessary for these vessels to use energy efficiently. The sensors installed on ships typically gather a very large amount of data, and thus a big data platform is needed to manage and analyze the data. In this paper, we build a big data analysis platform based on Hadoop and Spark, and we present a method to detect abnormal ship operation using the platform. We also utilize real ship operation data to discuss the data analysis experiment.

**Key Words :** Abnormal Status Detection(비정상 상태 감지), Big Data Platform(빅 데이터 플랫폼), Energy Efficiency(에너지 효율), Hadoop, Spark(하둡, 스파크), Ship Operation(선박 운영)

### 1. 서 론

전 세계적으로 해양 오염물질의 배출을 줄이기 위한 노력이 지속되고 있다. 국제해사기구는 2020년부터 세계 모든 해역을 지나는 선박 연료유의 함량유량에 대한 규제를 강화하고 있으며 배출규제해역을 설정하여 관리를 강화하고 있다. 이에 조선 및 해운 업계에서는 LNG(liquefied natural gas) 연료 사용 선박, 저유황 연료 사용, 배기가스 세정장치

장착 등의 대응방안을 선택하고 있다<sup>1-2)</sup>.

기존의 선박은 디젤엔진을 사용하기 때문에 LNG 연료를 사용하도록 전환하는 것이 힘들다. 저유황 연료는 고유황 연료보다 가격이 비싸고, 황 함량 정도에 따른 사용법의 변화 및 고착현상 등의 문제가 발생된다. 또, 배기가스 세정장치 장착에 있어서는 운용 에너지가 증가하고 이용가능공간이 축소되며 추가 투자비용이 발생하는 문제점이 있다. 이를 고려하면 기존 선박의 경우 저항 성능 향상<sup>12)</sup>, 유지보수 최적화 등으로 운항 시 에너지를 효율적으로 사용하는 것이 중요하다.

# Corresponding Author : dhmun@knu.ac.kr

Tel: +82-54-530-1271, Fax: +82-54-530-1278

선박 운항 과정에서 효율적으로 에너지를 사용하기 위해서는 선박에 설치된 센서로부터 얻어지는 각종 운항 데이터를 분석하여 적절한 대응을 하는 것이 필요하다. 하지만 장시간 운항하는 선박에서 실시간으로 얻어지는 운항 데이터의 양은 매우 방대하여 일반적인 방법으로는 분석에 어려움이 있으며 데이터의 관리 또한 어렵다. 따라서 이러한 빅데이터를 다룰 수 있는 분석 플랫폼이 필요하다.

이 논문에서는 빅데이터 분석 플랫폼을 구축하여 선박 운항 효율의 이상 상태를 분석하는 방법을 제시한다. 이를 위해 하둡과 스파크를 활용하여 데이터 저장, 관리 및 분석 도구를 포함하는 빅데이터 분석 플랫폼을 구성한다. 그리고 이를 기반으로 기존 선박 운항 데이터를 입력 받아 연료 소모율을 예측하는 회귀분석 모델을 생성한다. 그리고 회귀분석 모델을 신규 운항 데이터와 비교함으로써 운항 효율의 변동을 탐지하는 모듈을 개발한다. 마지막으로 구축된 빅데이터 분석 플랫폼을 이용하여 샘플 선박 운항 데이터에 대해서 이상 상태를 분석하는 실험을 수행한다.

이 논문은 다음과 같이 구성된다. 2절에서는 운항 데이터 분석에 관한 기존 연구들을 분석한다. 3절에서는 빅데이터 분석 플랫폼의 구성과 기능을 설명한다. 4절에서는 선박 운항 데이터의 구조와 데이터 분석 방법을 설명한다. 5절에서는 구축된 분석 플랫폼을 이용하여 선박 운항 효율 이상 상태를 분석한 결과를 논의한다. 마지막으로 6절에서 결론을 맺는다.

## 2. 관련연구

선박의 에너지 효율을 분석하기 위한 다양한 연구가 보고되었다. Yu et al.<sup>[3]</sup>은 하둡과 R을 연동하여 선박 운항 효율을 예측하는 회귀 모델을 개발하였다. Perera et al.<sup>[4]</sup>은 EM(expectation-maximization)과 GMM(gaussian mixture model)을 이용하여 운항 데이터를 분류한 후 오토인코더(autoencoder) 신경망을 이용하여 선박 성능을 모니터링하는 방법을 제안했다. Beşikçi et al.<sup>[5]</sup>은 선박 속도, RPM(revolutions per minute), 트림 등의 선박 운항 데이

터를 이용하여 인공 신경망을 적용하여 연료 사용률을 예측하는 의사결정 지원 시스템을 제안했다. 김용대 et al.<sup>[6]</sup>은 선박에서 획득되는 데이터를 활용하는 선박 에너지 효율 모니터링 시스템을 제안했다. Kwon et al.<sup>[7]</sup>은 배수량, 방형계수, 프루드수(Froude number)와 보퍼트 풍력계급(BFS; beaufort force scale)에 따른 선속 손실량 약산식을 제안했다. 제안된 선속 손실량 약산식은 모든 선박에 대해 범용적으로 간단하게 사용할 수 있지만 특정 조건의 선박에 대해서는 정확도가 떨어진다. Kim et al.<sup>[8]</sup>은 연료소모량, 속력, RPM, 엔진 부하의 관계를 분석했으나 외력에 의한 변화는 고려하지 않았다. Kim et al.<sup>[9]</sup>은 선박으로부터 획득한 운항 데이터를 이용하여 선박의 연료 소모량에 대한 회귀분석을 통해 연료 소모 패턴을 알아내는 방법을 제안했지만 방대한 양의 선박 운항 데이터를 관리하는 방법에 대해서는 다루지 않았다. 선박의 에너지 효율 분석에 관한 기존 연구들은 분석 기법에 초점을 두고 있다. 이 연구는 선행 연구<sup>[3]</sup>의 결과를 바탕으로 최신 오픈소스 기반 빅데이터 환경인 하둡과 스파크를 적용하여 선박 운항 효율의 이상 상태를 탐지하는 방법의 개발 및 플랫폼의 구축을 목적으로 한다.

## 3. 하둡과 스파크 기반 빅데이터 분석 플랫폼

빅데이터 분석 플랫폼은 Fig. 1과 같이 하둡(hadoop), 스파크(spark), 양(yarn), 제플린(zeppelin), 암바리(ambari), 주키퍼(zookeeper)로 구성되며 다중 클러스터 환경 구축을 통해 단일 서버에서 처리하지 못했던 빅데이터를 효율적이고 빠르게 처리할 수 있다.

하둡은 대용량 데이터를 분산 처리할 수 있는 자바 기반의 오픈소스 프레임워크로써 여러 대의 서버에 데이터를 저장하고 처리할 수 있다. 하둡의 구성요소는 크게 HDFS(hadoop distributed file system)과 맵리듀스(mapreduce)로 나뉜다. HDFS는 하나의 데이터를 여러 대의 서버에 나누어 저장하고 같은 데이터를 중복하여 저장하여 하나의 서버가 다운되거나 데이터가 손상되더라도 복구가 가

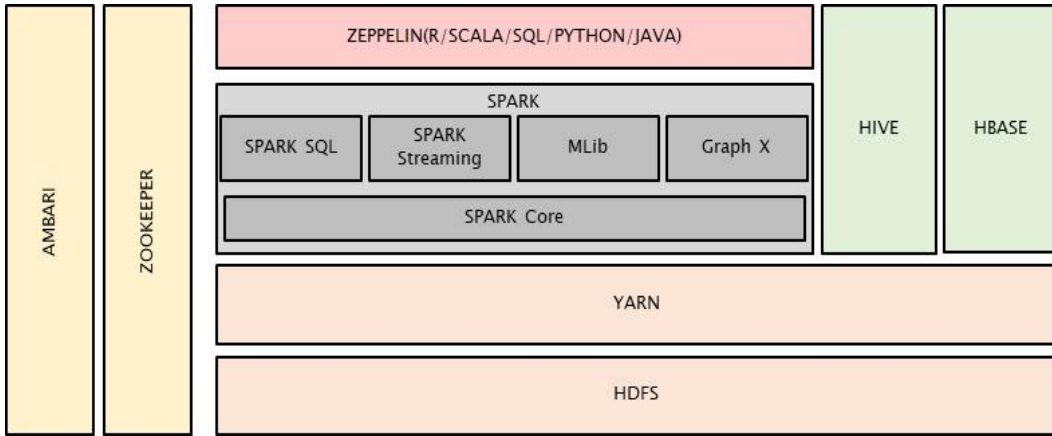


Fig. 1 Open source-based big data platform

능하다. 맵리듀스는 여러 대의 서버에 작업을 분배하여 분산된 데이터를 쉽고 빠르게 처리할 수 있는 기능이다. 이 연구에서는 하둡의 구성요소 중 HDFS를 적용한다.

스파크는 데이터를 분산 쿼리 및 처리할 수 있는 오픈소스 엔진이다. 기존의 맵리듀스는 연산을 위한 저장소로 디스크를 사용하지만 스파크는 연산을 위한 저장소로 메모리를 우선적으로 사용하고 메모리가 부족할 경우 디스크를 사용하는 인메모리 방식으로 데이터를 처리하기 때문에 기존의 맵리듀스보다 훨씬 빠른 데이터 처리속도를 가진다. 맵리듀스에서는 배치 (batch) 데이터를 처리하는 기능만을 제공하여 실시간 데이터를 처리하는 부족하지만 스파크는 배치와 실시간 데이터 처리를 위한 스트리밍 기능을 제공한다. 프로그래밍 언어에서도 맵리듀스는 Java, SQL만 지원하지만 스파크는 R, Scala, SQL, Python, Java를 지원한다.

얌은 클러스터 리소스 관리 시스템이다. 각 시스템에 필요한 리소스를 할당하고 모니터링 함으로써 분산 병렬 처리를 가능하게 해준다.

제플린은 데이터 분석 및 가시화 시스템으로 기존 CLS (command Line interface) 기반의 데이터 분석 환경을 개선하기 위해 개발된 시스템이다. 제플린을 이용하면 개발자가 웹에서 코드를 작성하고 실행 결과를 직접 가시화할 수 있다.

암바리는 하둡 전체 시스템의 관리 및 모니터링

을 지원하는 시스템이다. 그리고 주키퍼는 클러스터 코디네이션 시스템으로 분산 시스템 내에서 상태 정보나 설정 정보 등을 유지하고 관리한다.

#### 4. 선박 운항 효율 이상 상태 분석 방법

##### 4.1 선박 운항 데이터 구조

이 연구에 사용된 선박 운항 데이터는 한 척의 국제 상선에서 획득한 데이터로 그 수는 약 80,000개이다. 데이터는 Fig. 2와 같이 선박이 운항하는 방향 (동쪽과 서쪽)에 따라 나뉘어져 있다. 방향에 따라 나뉘어진 데이터 세트는 다시 보퍼트 풍력계급(BFS 0~12)별로 구분되며 이는 다시 입력(input) 데이터와 (output) 데이터로 나뉜다. 선박 운항 방향으로 나뉜 데이터의 수는 동쪽과 서쪽 각각 약 40,000개로 균등하게 구성된다. 그리고 동쪽과 서쪽으로 구분된 데이터의 보퍼트 풍력계급별 데이터 분포는 BFS 2와 9 사이에 대부분의 데이터가 집중된 정규분포를 이룬다.

데이터는 MATLAB 파일로 되어있으며 각 데이터는 Table 1과 같이 12개의 변수로 구성된다. 데이터를 구성하고 있는 변수는 보퍼트 풍력계급 (BFS), 배수량 (displacement), 선박이 물에 잠긴 깊이 (draft), 시간당 연료소모량 (FOC), 대지속력 (LGS),

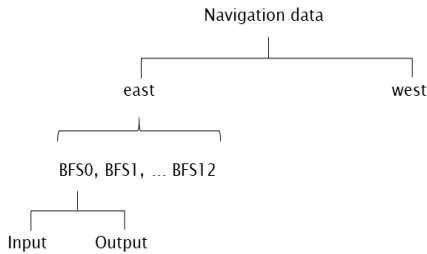


Fig. 2 Structure of ship navigation data

Table 1 Variables of ship navigation data

Variable	Detail	Unit
BFS	Beaufort wind Force Scale	
Displacement	Displacement of the ship	t
Draft	Draft of the ship	m
FOC	Fuel oil consumption	l/h
LGS	Longitudinal speed over ground	knots
LWS	Longitudinal speed through water	knots
Power	Power of engine	kW
Propeller immersion	Proportion of the submerged propeller	%
RPM	Revolutions per minute	rpm
Trim	Trim of the ship	m
Wetted surface	Wetted surface of the ship	m <sup>2</sup>
Wind resistance	Wind resistance acting on the ship	kN

대수속력 (LWS), 엔진출력 (power), 프로펠러가 물에 잠긴 정도 (propeller immersion), 분당 엔진회전 수 (RPM), 선박의 기울기 (trim), 선박의 침수면 (wetted surface) 및 바람저항 (wind resistance)이다.

## 4.2 선박 운항 데이터 구조

빅데이터 분석 플랫폼을 이용한 선박 운항 효율 이상 상태 분석 절차는 Fig. 3과 같다. 분석 절차는 크게 회귀 모델 생성 단계 및 이상 상태 탐지 단계로 구분된다.

회귀 모델 생성을 위해서 먼저 MATLAB 파일 형식의 기준 운항 데이터 (base navigation data)를 처리가 용이한 CSV (comma-separated values) 파일

형식으로 변환하고 BFS별로 구분된 데이터들을 하나로 병합한다 (Fig. 3의 *Converting to CSV*). 병합된 선박 운항 데이터는 로컬 영역에 저장이 되는데 셸 (shell) 상에서 HDFS 제공 명령어를 실행하여 HDFS에 저장한 후 로컬영역에 있는 데이터를 제거한다 (Fig. 3의 *Saving CSV in HDFS*). 분석을 수행할 때에는 스파크에서 직접 HDFS에 저장된 선박 운항 데이터를 가져온다 (Fig. 3의 *Retrieving CSV from HDFS*). 그리고 가져온 선박 운항 데이터에 대해서 노이즈 데이터를 수행한 후 곡선 피팅 및 인공신경망 학습을 통해 회귀 모델을 생성한다 (Fig. 3의 *Data filtering과 Regression analysis*). 생성된 회귀 모델은 우선 로컬영역에 저장되는데 셸 상에서 HDFS 제공 명령어를 실행하여 HDFS에 저장한다 (Fig. 3의 *Saving the model in HDFS*).

이상 상태 탐지 단계는 신규 운항 데이터 (new navigation data)가 획득된 이후에 시작되는데 먼저 HDFS에 저장된 회귀 모델을 로컬영역을 거쳐 스파크로 불러온다 (Fig. 3의 *Retrieving the model from HDFS와 Reading the regression model*). 그리고 신규 운항 데이터에 대해서 형식 변환 병합 작업을 수행한다 (Fig. 3의 *Converting to CSV*). 다음으로 기준 회귀 모델을 사용하여 신규 운항 데이터에 대한 오차를 계산하여 이를 가시화한다. 사용자는 이 오차를 기준 운항 데이터에 대해 계산된 오차와 비교하여 이 두 값의 차이가 클 경우 선박 운항 효율에 변동 (이상)이 발생한 것으로 판정한다 (Fig. 3의 *Verifying whether new data follows the regression model*).

선박 운항 효율의 변동 판단 척도가 되는 오차 계산은 RMSE(root mean square deviation) 계산법을 사용한다. 기준 회귀모델과 신규 운항 데이터간의 잔차를 구한 후에 RMSE 계산법을 적용한다. 오차는 절대오차(absolute error)와 상대오차(relative error) 두가지로 나뉜다. 절대오차는 위에서 구한 RMSE 계산 값을 그대로 사용한다. 상대오차는 기준 회귀 모델과 신규 운항 데이터간의 잔차와 신규 운항 데이터와의 비율을 계산한 후 RMSE 계산법을 적용한다. 운항 효율이 변동한 것으로 판단하는 기준은 신규 운항 데이터에서 계산된 오차가 기준 운항 데이터에서 계산된 오차보다 클 경우로 한다.

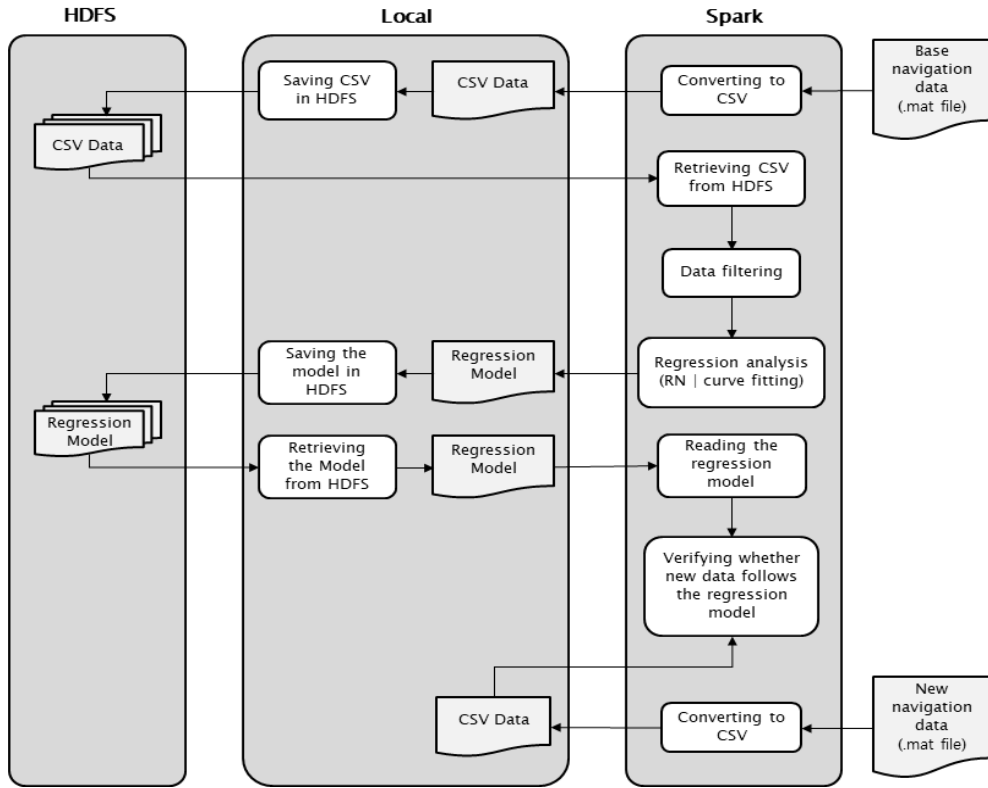


Fig. 3 Procedure for the abnormal status analysis of ship navigation efficiency

## 5. 빅데이터 분석 플랫폼을 활용한 운항 효율 이상 상태 분석 실험

### 5.1 빅데이터 분석 플랫폼 구축 및 실행

구축한 빅데이터 분석 플랫폼 환경은 Fig. 4와 같다. 분산 환경의 클러스터 시스템을 구현하기 위해 마스터 서버 1개와 슬레이브 서버 2개로 구성했다. 사용자는 마스터 서버에서 제플린을 이용하여 분석 코드를 작성하고 실행한다. 제플린에서 실행된 코드는 마스터 서버의 스파크로 전달되고 스파크에서는 슬레이브 서버로 작업 명령을 내린다. 작업이 완료되면 슬레이브 서버에서 마스터 서버로 분석 결과가 전달된다. CSV파일 형식의 선박 운항 데이터와 분석 결과물인 인공신경망 회귀모델 파일은 HDFS로 저장되고 사용할 때 데이터를 불러온다.

### 5.2 운항 데이터 전처리

운항 효율 이상 상태 분석 실험을 위해 전체 선박 운항 데이터를 기준 운항 데이터와 신규 운항 데이터로 나누었다. 전체 약 80,000개의 데이터를 시간 순서대로 정렬한 뒤 앞단의 70%를 기준 운항 데이터로 사용하고 후단의 30%를 신규 운항 데이터로 사용했다.

이 연구에서 사용된 선박 운항 데이터를 보면 BFS가 2에서 9사이의 구간에 이르는 데이터가 대부분이고 이 데이터들은 정규 분포를 갖는다. 그래서 유의미한 데이터 분석을 할 수 있다고 판단되어 이 구간의 데이터만을 분석에 사용하였다.

데이터 분석의 정확도는 데이터의 양과 질에 많은 영향을 받는다. 따라서 전체 데이터에서 센서의 오작동이나 잘못된 관측으로 인한 이상치가 존재할 경우 이를 제거해주어야 한다. 이 연구에서는 이상

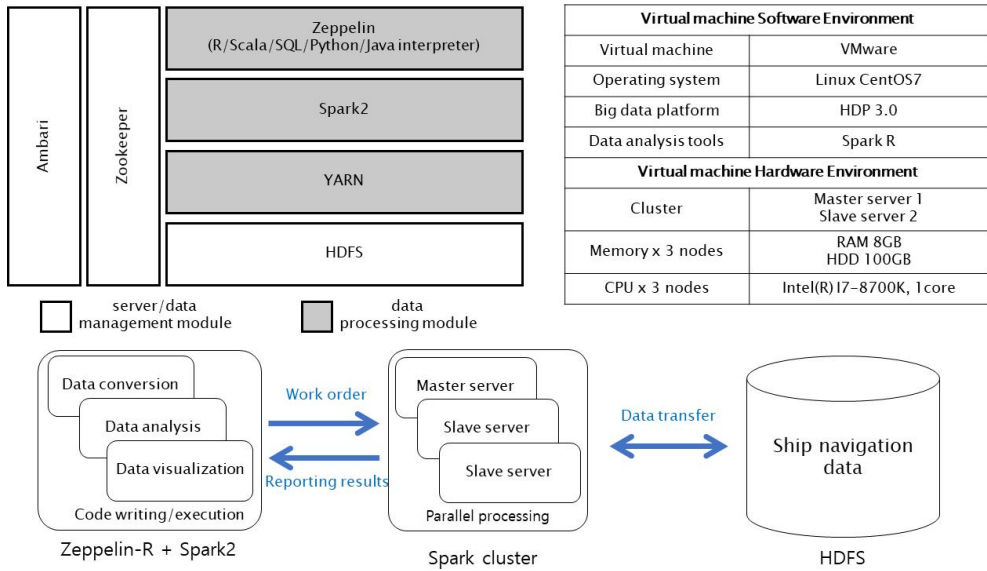


Fig. 4 Big data analysis platform for ship operation abnormal status detection

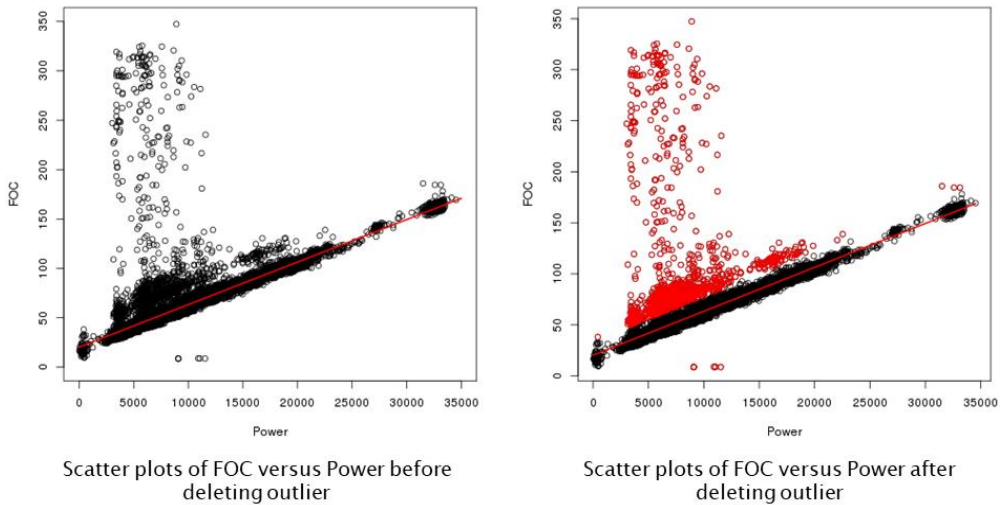


Fig. 5 Scatter plots of FOC versus Power

치를 제거하기 위한 기준으로 연료 소모량을 나타내는 FOC와 엔진 출력을 나타내는 Power의 관계를 이용했다. FOC와 Power의 관계는 이론적으로 선형에 가까우며, FOC와 속도 관련 변수인 LWS, LGS, RPM과의 관계와는 달리 BFS 등의 외부 요인의 영향이 적다. 그리고 해상 상태에 따라 흔들리는 선박의 상황을 고려할 때, 선박에 고정되어 있는 엔

진과 관련 있는 변수에 대한 측정값에 비해 선박의 흔들림에 따라 변동이 발생하는 액체 연료 소모량에 대한 측정값에 이상치가 많을 것으로 판단했다. 선박 운항 데이터에서 이상치를 제거하기 위하여, 먼저 FOC-Power 선형 회귀모델을 생성했다. 다음으로 생성된 선형 회귀모델과 데이터 사이의 잔차를 계산하여 정상구간을 벗어나는 값들을 이상치로 판

단하고 이를 제거하였다.

정상구간은 잔차의 분포가 정규분포를 이룬다고 가정한 상태에서 통계학의 표준 점수 계산 공식(1)을 사용하여 결정하였다. 표준 점수는 통계학적으로 정분포를 만들고 각각의 경우가 표준편차 상에서 어디에 위치하는지를 보여주는 차원 없는 값이다. Z는 표준 점수를 의미하고  $\mu$ 는 잔차의 평균,  $\sigma$ 는 잔차의 표준편차, x는 잔차의 값이다. 기준이 되는 잔차 값을 계산하기 위해서 공식(1)을 변형한 공식(2)를 사용했으며 Z값은 양측 검정 신뢰도 70%에 해당되는 1.04를 사용했다. 잔차가  $X_1$ 보다 클 경우 그리고  $X_2$ 보다 작을 경우 정상 구간에서 벗어나는 것으로 판단하여 이를 제거하였다.

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

$$X_1 = \mu + Z \times \sigma, X_2 = \mu - Z \times \sigma \quad (2)$$

### 5.3 인공신경망을 이용한 회귀분석

인공신경망<sup>[11]</sup>은 생물학의 신경망을 기계학습과 접합시킨 것으로 시냅스와 같은 결합방식의 네트워크를 형성한 수학적 모델이다. 일반적으로 사용하는 기본적인 인공신경망 모델인 다층인공신경망은 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 구성된다. 그리고 각 층들은 노드들로 구성된다. 입력층에서는 시스템 외부로부터 데이터를 입력 받아서 시스템으로 전송하고 은닉층에서는 입력된 값을 받아 처리한 다음 결과를 산출한다. 마지막으로 출력층에서는 입력된 값과 현재 시스템 상태에 기준하여 시스템 출력 값을 산출한다.

선행 연구에서 Yu et al.<sup>[3]</sup>은 선박 운항 효율을 계산하기 위하여 선박의 연료 소모량과 속도 관련 변수에 대한 커브 피팅 및 회귀분석을 수행하였으며, 선박 운항 효율을 가장 잘 표현하는 방법은 인공신경망으로 생성한 종속변수가 FOC, 독립변수가 RPM, BFS인 회귀모델이라고 보고했다. 이 연구에서는 Yu et al.<sup>[3]</sup>의 연구 결과를 참고하여 기준 선박 운항 데이터의 변수 중 FOC, RPM, BFS에 대해서 인공신경망을 이용한 회귀분석을 수행하였다. 인공신경망의 경우 충분한 데이터가 갖춰지면 다른 수학적 해석이 필요없이 다중 변수 및 비선형성을 포함한 데이터에 대한 학습이 가능하다. 그리고 생

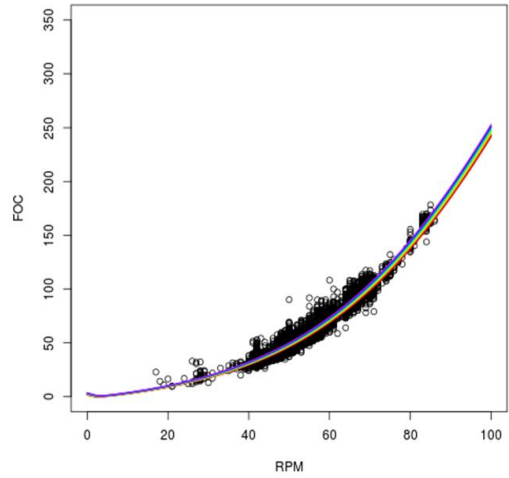


Fig. 6 Regression model between RPM and FOC based on neural network

성되는 회귀모델의 정확도 또한 준수하다. 또한 선박 운항 데이터 구조를 살펴볼 때 BFS와 같은 서열척도를 포함하여 학습이 가능하며 다중공선성 문제가 있는 데이터도 처리가 가능하다.

R에서 인공신경망 회귀분석을 사용하기 위해서 nnet패키지<sup>[10]</sup>를 사용했다. 전처리를 거친 기준 선박 운항 데이터에 대해 인공신경망 회귀분석을 통하여 회귀모델을 생성했다. 생성된 회귀모델을 데이터 분포와 함께 가시화하면 BFS별로 그래프가 나타난다. 회귀모델 생성결과는 Fig. 6 과 같다. 인공신경망 회귀분석으로 생성된 회귀모델은 BFS의 값에 따라 총 8개의 다른 색으로 분리되어 나타났다. 그래프의 형태는 RPM 값이 증가함에 따라 FOC 값이 점점 증가하는 곡선형태로 나타났으며 BFS 값이 높을수록 곡선의 기울기는 더 크게 나타났다.

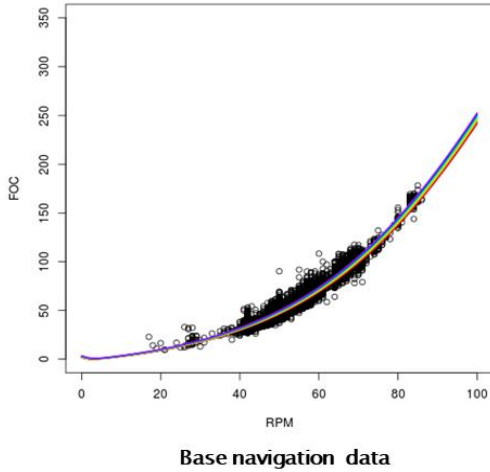
### 5.4 선박 운항 효율 이상 상태 탐지

회귀분석은 전체 데이터에 대해 가장 적합한 하나의 모델을 만드는 것이기 때문에 신규 선박 운항 데이터뿐만 아니라 기준 선박 운항 데이터에 대해서도 회귀모델 간에 오차가 발생하게 된다. 그래서 새로운 선박 운항 데이터와 기준 회귀모델 간에 계산된 오차의 값 en이 기준 선박 운항 데이터와 기준 회귀모델 간의 오차의 값 ep보다 크게 나타날 경우 운항 효율 변동(이상)이 발생한 것으로 했다.



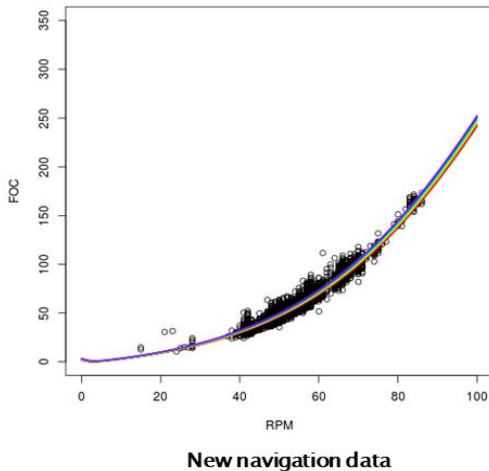
“ Absolute error : 3.527

“ Relative error : 6.166 %



“ Absolute error : 3.499

“ Relative error : 6.110 %



**Fig. 7 Detection of ship operation abnormal status by the comparison of RMSEs for base and new navigation data**

새로운 선박 운항 데이터를 기초로 운항 효율 이상 여부를 파악하기 위해서 먼저 입력 받은 선박 운항 데이터의 전처리를 수행하였다. 그리고 mnet패키지를 사용하여 새로운 선박 운항 데이터가 기준 회귀 모델을 따랐을 때의 값을 계산하여 실제 신규

선박 운항 데이터 값도 분포도와 함께 그래프로 가시화하였다. 그리고 이 그래프에는 기준 회귀 모델과 신규 선박 운항 데이터 사이에 RMSE값을 계산하여 절대오차와 상대오차를 같이 표시했다. 선박 운항 효율 이상 상태 탐지 결과는 Fig. 7과 같다.

기준 회귀모델에 대해서 기준 운항 데이터와의 절대오차와 상대오차는 3.527과 6.166%로 나타났고, 신규 운항 데이터와의 절대 오차와 상대오차는 3.499과 6.110%로 나타났다. 두 오차 간의 차이는 각각 0.028과 0.056%로, 변동이 아주 작았기 때문에 신규 운항 데이터에 대해서 선박 운항 효율의 변동이 없는 것으로 판단하였다. 기준 및 신규 운항 데이터가 동일 데이터 세트를 일정 비율로 나눠서 만든 것이므로 선박 운항 효율에 변동이 없는 것이 타당하다.

이 연구에서 사용한 운항 데이터의 양이 상대적으로 많지 않아 빅데이터 분석 플랫폼의 시간적인 측면이나 관리 효율성 측면의 논의는 다소 어려운 점이 있다. 그러나 일반적으로 클러스터 환경을, 단일 서버와 비교하면, 네트워크 상에 서로 연결된 노드(서버) 사이의 인터페이스, 작업 할당 및 제어 등의 분산 환경 관리로 인해서 시간 및 자원이 추가로 소요되지만 계산 업무를 각 노드에 분산 할당하여 병렬로 처리함으로써 계산 시간을 줄일 수 있게 된다. 이에 따라 많은 양의 데이터에 대해서 복잡한 다수의 연산을 해야 할 경우 빅데이터 분석 플랫폼이 효과적이다.

## 6. 결론

하둡과 스파크 기반의 오픈소스 기반 빅데이터 플랫폼을 활용하여 선박 운항 효율 이상 상태를 분석하는 방법을 제안하였다. 제안하는 선박 운항 효율 이상 상태 분석 절차 크게 회귀 모델 생성 단계 및 이상 상태 탐지 단계로 구분된다. 회귀 모델 생성 단계에서는 노이즈 데이터를 제거한 다음 인공 신경망 학습을 이용하여 회귀 모델을 생성하였다. 이상 상태 탐지 단계에서는 기준 회귀 모델을 이용하여 신규 선박 운항 데이터의 오차를 분석하여 그 값이 기준 선박 운항 데이터에 대해서 계산된 오차보다 클 경우 이상 상태로 판정하였다.



구축한 빅데이터 플랫폼을 이용한 운항 효율 이상 상태 탐지 실험을 수행하였다. 이를 위해서 먼저 실제 선박 운항 데이터를 7:3의 비율로 기준 데이터와 신규 데이터로 나누었다. 그리고 기준 데이터를 기반으로 생성된 회귀모델에 대해서 기준 데이터와의 오차  $ep$ 와 신규 데이터와의 오차  $en$ 를 비교했다. 사전에 예상한 것과 같이, 비교 결과 두 오차 간의 차이가 매우 작았기 때문에 선박 운항 효율의 변동이 없는 것으로 판단하였다.

이 연구에서 사용된 운항 데이터는 선박 한 척에 대해서 특정 기간에 수집한 데이터로써 빅데이터라고 하기에는 양이 많지 않다. 하지만 해운사들은 운영 중인 다수의 선박에 대해서 전 운용 기간 동안의 데이터를 수집하고 분석하여 운항의 효율성을 극대화하는 노력을 하고 있다. 이 연구는 이를 지원하기 위한 프로토타입을 개발한 것에 의미가 있으며 향후에는 다수의 선박에 대해서 장기간 수집된 데이터를 활용하여 운항 효율의 이상 상태 분석을 수행할 계획이다.

## 후 기

이 논문은 해양수산부 한국형 e-Navigation 사업의 “IMO 차세대 해양안전 종합관리 체계 기술개발” 과제와 산업통상자원부 산업핵심기술개발사업(과제번호: 20000725 및 10080662)의 지원을 받아 수행된 연구 결과 중 일부임을 밝힙니다.

## REFERENCES

1. Jung, R. T., “Recent International Development on the Technical and Operational Measures of IMOs CO 2 Emission Control From Ships”, Journal of the Korean Society for Marine Environment & Energy, Vol. 14, No. 1, pp. 65-71, 2011.
2. Kim, M., “Eco-friendly Technology Trends in Ocean Transportation”, Shipper’s Journal, pp. 27-38, 2013.
3. Yu, E., Park, K., and Mun, D., “Study on Prediction of Ship Navigation Efficiency Using Open Source-based Big Data Platform”, Korean Journal of Computational Design and Engineering, Vol. 23, No. 3, pp. 275-284, 2018.
4. Perera, L. P., and Mo, B., “Machine intelligence for energy efficient ships: A big data solution. Maritime Engineering and Technology III”, Guedes Soares & Santos (Eds.), Vol. 1, pp. 143-150, 2016.
5. Beşikçi, E. B., Arslan, O., Turan, O., and Ölçer, A. I., “An artificial neural network based decision support system for energy efficient ship operations”, Computers & Operations Research, Vol. 66, pp. 393-401, 2016.
6. Kim, Y., Yoon, H., and Kang, N., “Design and Implementation of Ship Energy Efficiency Monitoring System”, The Journal of Korea Navigation Institute, Vol. 20, No. 5, pp. 408-416, 2016.
7. Kwon, Y. J., and Kim, D. Y., “A Research on the Approximate Formulae for the Speed Loss at Sea”, Journal of Ocean Engineering and Technology, Vol. 19, No. 2, pp. 90-93, 2005.
8. Kim, S. K., Lee, Y. S., Kong, G. Y., Kim, J. P., and Jung, C. H., “A Study on the Ships Speed for Reducing the Fuel Oil Consumption in Actual Ships”, Journal of the Korean Society of Marine Environment & Safety, Vol. 18, No. 1, pp. 41-47, 2012.
9. Byeon, S., Jun, C., Kim, K., Lee, S., and Park, K., “A statistical procedure of analyzing container ship operation data for finding fuel consumption patterns”, The Korean Journal of Applied Statistics, Vol. 30, No. 5, pp. 633-645, 2017.
10. “R project” (2016), <https://cran.r-project.org/web/packages/nnet/index.html> (accessed 01 August, 2018).
11. Lee, T. H., Park, J. H., “Modelling for the Full Ranges of the Steam Table using Neural Networks”, J. Korean Soc. Mech. Technol, Vol. 20, No. 1, pp. 82~88, 2018.
12. Choi, H. J., Moon, I. K., "A Study on the Panel Sensitivity to Predict Resistance Performance of High Speed Ship", J. Korean Soc. Mech. Technol, Vol. 20, No. 2, pp.160~165, 2018.