공학석사 학위논문

실선 데이터를 이용한 기계학습 기반의 연료소비 예측모델의 개발

Development of a Fuel Consumption Prediction Model Based on Machine Learning Using Ship's Data

지도교수 박 준 범

2019 년 8 월

한국해양대학교 대학원

항해학과
김 영 롱

본 논문을 김영롱의 공학석사 학위논문으로 인준함.

위원장 : 공학박사 정 연 철 (인)

위 원 : 공학박사 문 성 배 (인)

위 원 : 공학박사 박 준 범 (인)

2019 년 6 월 20 일

한국해양대학교 대학원

목 차

List of Tables

List of Figures

Development of a Fuel Consumption Prediction Model Based on Machine Learning Using Ship's Data

Kim, Young-Rong

Division of Navigation Science
Graduate School of Korea Maritime and Ocean University

Abstract

As the environmental regulations of the international organizations are being strengthened and interest in the economic operation of the ship is being increased, many studies have been done to reduce the amount of fuel consumed by ships. In the ship operational measures, management of navigation performance, hull and propeller condition and ship system have been performed to improve the energy efficiency of ships. In recent years, due to the development of collection and storage of big data and communication technologies, there have been great demands for real-time monitoring techniques predicting ship operational performances through ship data.

In this study, it is intended to develop a prediction model for fuel consumption rate based on the real-time ship operational data. In previous studies, all the operational data collected from the ship were used without being aware, or the main data to estimate fuel consumption of a ship were selected relying on only the expert's experience. These cases could cause multicollinearity and overfitting

problems, and the complexity of the calculations has been also led to inefficiencies in the model generation process.

In order to resolve the weakness of existing prediction models, this study performed overall data processing to recognize the characteristics of ship data and then selected independent variables for implementing the fuel consumption prediction model logically through dimensional reduction methods such as correlation analysis, variance inflation factor, principal component analysis, and Lasso regularization. Finally, the regression model and ANN(Artificial neural network) algorithm were applied to complete prediction models, and the performance of those models was analyzed.

It is sure that the fuel consumption rate prediction model could support the operator's decision-making during the route planning, detect hull and equipment anomalies, identify performance degradation due to long-term operations and help users understand operational data. The prediction model developed in this study would be a basic stepwise study of the energy efficiency optimization system to support the operator's decision making. In future studies, it will be possible to establish a sophisticated prediction system by improving the accuracy and reliability of the model based on data on various types of ships and operating conditions.

KEY WORDS: Ship energy efficiency; Fuel consumption prediction; Machine learning; Dimension reduction method; Neural network.

실선 데이터를 이용한 기계학습 기반의 연료소비 예측모델의 개발

김 영 롱

한국해양대학교 대학원
항해학과

초 록

국제기구의 환경 규제 강화와 더불어 선박의 경제적 운항에 대한 관심이 높아지면서 선박에서 소모되는 연료량을 저감하기 위한 연구가 활발히 진행되고 있다. 선박의 에너지효율을 향상시킬 수 있는 운항적 조치에는 항해성능관리, 선체 및 프로펠러 상태관리, 선박 시스템관리와 같은 방법들이 있으며 최근에는 빅데이터의 수집 및 저장 그리고 통신기술의 발달로 인하여 선박 데이터를 활용한 실시간 운항성능 모니터링 및 예측 기술에 관한 수요가 많아지고 있다.

본 연구에서는 실선 운항데이터를 기반으로 연료소모율을 예측하는 모델을 개발하고자 하였다. 지금까지는 선박의 연료소모량을 예측하기 위하여 선박으로부터 수집가능한 모든 운항데이터를 사용하거나 일부 연구에서는 전문가의 경험에 의존하여 변수를 선정하기도 하였다. 하지만 이러한 경우 모델의 다중공선성 및 과적합 문제가 발생할 수 있으며 계산의 복잡성으로 인하여 비효율적인 측면이 있었다.

이러한 기존 방법들의 문제점을 해결하기 위하여 선박데이터의 특성을 고려한 전반적인 데이터 처리 과정을 제시하였으며 상관 분석 및 분산팽창지수, 주성분 분석, 라소 정규화와 같은 차원축소방법을 도입하여 예측모델을 구현하기 위한 독립변수를 선정하였다. 마지막으로 실선 데이터에 회귀모형 및 인공 신경망 알고리즘을 적용하여 예측모델을 구현하고 그 성능을 분석하였다.

연료소모율 예측모델은 항로계획 시 운항자의 의사 결정을 지원하고 선체 및 기기의 이상상태를 탐지할 수 있으며 장기 운항에 따른 성능저하 파악 및 이용자의 운항데이터에 대한 이해를 도울 수 있을 것으로 사료된다. 본 연구에서 개발한 예측모델은 운항자의 의사결정을 지원하기 위한 에너지효율 최적화 시스템의 기초 단계적인 연구라고 할 수 있다. 향후 연구에서는 다양한 선종과 운항 조건에 대한 데이터를 바탕으로 모델의 정확도와 신뢰도를 개선하여 고도화된 예측 시스템을 수립할 수 있을 것이라고 판단된다.

KEY WORDS: 선박 에너지효율; 연료소비 예측; 기계학습; 차원 감소법; 인공 신경망.

제 1 장 서 론

1.1 연구의 배경

선박의 연료소모량 관리는 해운 사회의 주요한 관심사 중 하나이다. 선박의 운항 경비는 선박의 종류, 크기, 운항 해역 등에 따라 차이가 있을 수 있지만 연료비가 전체 운항 경비의 약 50~60%를 차지하는 것으로 밝혀졌다 (Stopford, 2009; Eide et al., 2011). 따라서 해운회사들은 에너지 소비량을 줄이기 위한 효율적인 선박 운영 절차를 개발하여 관리 비용을 절감하고 시장에서 경쟁력을 유지하기 위해 노력하고 있다.

한편, 해상 운송에 의한 배기가스 배출량이 지속적으로 상승하고 있으며 이에 따른 국제환경규제의 강화로 선박의 운항 에너지효율에 대한 필요성이 대두되고 있는 상황이다 (IMO, 2009a). 국제해사기구(International Maritime Organization;IMO)의 해양환경보호위원회(Marine Environment Protection Committee;MEPC)에서는 신조선의 탄소가스 배출을 규제하기 위하여 에너지효율 설계 지수(Energy Efficiency Design Index;EEDI)를 도입하였다. 에너지효율 설계 지수는 선박이 1톤의 화물을 1마일 운송하는데 발생하는 CO2의 양을 의미하며 신조선의 설계 및 건조 시에 적용하여 국제해사기구의 요구 조건을 충족하지 못하는 선박은 운항이 전면 금지된다. 또한 현존선에 대해서는 에너지효율 운항 지수(Energy Efficiency Operational Index;EEOI) 및 선박 에너지효율 관리 계획(Ship Energy Efficiency Management Plan;SEEMP)을 사용하여 선박 운항자가 효율적으로 CO2 배출량을 줄일 수 있도록 규제하고 있다 (IMO, 2009b; 2012a; 2012b). 이와 같은 경제적, 환경적인 요인으로 선박의 연료 저감은 에너지효율에 있어 가장 중요한 요소가 되었다.

ABS (2013)에 따르면 선박의 에너지효율을 관리하기 위한 운항 조치는 크게 항해성능관리, 선체 및 프로펠러 상태관리, 선박 시스템관리로 분류할 수 있다.

선박 항해성능관리에는 주어진 시간 이내에 목적지까지 도착할 수 있는 최적

속력의 설정, 기상 및 해상을 고려한 항로 계획을 통한 선박의 부가 저항 감소, 자동조타장치의 적합한 모드 설정에 따른 타기 사용의 최소화, 평형수 적재량 및 트림의 최적화를 통한 선체 저항 최소화 등이 포함된다. 연료효율을 높이기 위해서는 이러한 여러 운항 요인들을 동시에 고려하여 관리하는 것이 굉장히 중요하다. 가장 기본적이면서 효과적인 방법은 선박의 운항 속력을 줄이는 것이다. 선속의 3승은 주기관에 의해서 소비되는 전력에 비례하며 이는 선박 연료 사용량에 가장 큰 영향을 미치는 것으로 알려져 있다 (Ronen, 1982; Fagerholt et al., 2010; Norstad et al., 2011). 이는 주어진 항차에서 기존의 선속을 10% 정도만 감속하여 운항하게 되면 약 20% 정도의 연료를 절약할 수 있음을 의미한다. 이처럼 선박의 다양한 운항조건에 대하여 최적의 속도로 운항하는 것은 에너지효율적인 측면에서 굉장히 효과적인 방법이다. 적절한 양의 평형수를 채워 선박의 흘수 및 트림을 조정하는 것도 쉬우면서도 저렴한 방법이며, 선박에서 소모되는 에너지는 트림의 조건에 따라 상당히 다르다는 것이 입증되었다 (Journé et al., 1987). 또한 다양한 기상, 해상 조건하에서 목적지까지 최단 시간에 최소한의 연료를 사용하여 도달할 수 있도록 지원해주는 최적항로지원 서비스는 연료소모량을 최대 3%까지 절감할 수 있다고 보고되었다 (Armstrong, 2013).

선체 및 프로펠러 상태 관리는 수면하부의 선체와 프로펠러의 표면을 깨끗한 상태로 유지하는 것을 말한다. 선박의 운항 중 자연적으로 선저부착물 및 해양유기체 등으로부터 선체의 중량과 마찰저항이 증가되거나 또는 외부의 충격에 의한 손상 등으로 추진효율이 저하될 수가 있다. 적절한 주기로 폴리싱을 수행하고 도료를 도장함으로써 선박의 추진 효율을 향상시킬 수 있으며 이는 오염된 선체에 비해 선체 성능을 최대 10% 증가시키는 효과를 가질 수 있다 (ABS, 2013; Demirel et al., 2016).

선박에 탑재된 각 기기들은 선내 전력을 소모하고 이러한 전력을 생산하기 위한 발전기의 운용은 선박의 연료소모량과 연관된다. 제조사의 지침에 따라 최적의 성능을 발휘하도록 기기들은 조정될 수 있기 때문에 시기적절한 유지보수를 통해 선박 시스템을 관리하는 것이 중요하다.

이와 같이 선박의 연료효율을 증진시키기 위하여 현재까지 다양한 방법들이 연구되고 실제로 적용되어져 왔다. 최근에는 빅데이터의 수집 및 저장 그리고 통신 기술의 발달 등으로 인하여 선박 데이터를 활용한 운항성능의 모니터링과 예측에 관한 관심이 대두되고 있다. 특히 선박의 실시간 운항조건과 해역의 기상정보를 활용한 운항성능 예측모델의 개발이 시도되어 왔고 기존에 많은 분야에서 활용되었던 회귀 모형 뿐만 아니라 기계학습 알고리즘을 접목한 연구도 늘어나고 있는 추세이다.

1.2 선행 연구 고찰

선박의 운항데이터로부터 에너지효율과 관련된 변수를 예측하기 위한 다양한 연구들이 수행되어 왔다. 고전적인 연구 방법으로는 변수들의 물리적인 관계를 나타내는 실험식을 바탕으로 변수간의 관계를 파악할 수 있는 간단한 회귀 분석을 활용한 경우가 많았다. 일부 연구에서는 변수간의 상관에 의한 모델의 다중공선성을 방지하고 효율성을 높이기 위하여 라소(Least Absolute Shrinkage and Selection Operator;LASSO), 부분최소제곱법(Partial Least Square;PLS)와 같은 차원감소 방법을 적용한 회귀 모형을 사용하기도 하였다. 최근에는 선박 빅데이터의 수집과 저장 및 실시간 원격 모니터링에 대한 관심이 높아지면서 예측력이 우수한 인공 신경망(Artificial Neural Network;ANN), 서포트 벡터 머신(Support Vector Machine;SVM), 가우시안 프로세서(Gaussian Process;GP)와 같은 기계학습 알고리즘을 적용하는 연구가 많아지고 있다. Table 2.1에서는 선박의 에너지효율 예측과 관련한 선행연구를 나타낸다.

Petersen et al. (2012)는 연안페리선의 2달간 운항데이터를 활용하여 연료소모량 예측모델을 개발하였다. 데이터 전처리 과정에서 주성분 분석(Principal Component Analysis;PCA)을 수행하여 변수간의 관계를 파악하였으며 인공 신경망과 가우시안 프로세스를 활용하여 예측모델을 만들었다.

Yoo et al. (2016)는 기존에 알려진 여러 선박 관련 물리량의 역학적 관계를 바탕으로 회귀 분석을 수행하였다. 실선 운항데이터를 사용하여 선박의 엔진출력과 속력의 예측모델을 만들었으며 그 결과를 ISO 15016과 비교하여 모델의

유효성을 검증하였다.

Kim et al. (2017)은 컨테이너선박에 설치된 센서로부터 축적된 운항데이터를 이용하여 선박의 연료소비에 영향을 미치는 변수를 파악하고 예측모델을 개발하고자 하였다. 외력의 영향을 보퍼트 풍력계급을 기준으로 구분하고 부분최소제곱회귀 분석을 통하여 컨테이너선의 연료소비패턴을 파악하는 방법을 제안하였다.

Yan et al. (2018)은 항해구간에 따른 최적의 선속을 결정하고자 k-means 군집 알고리즘을 적용하여 선박의 항로를 분할하였다. 또한 기존의 실험식을 바탕으로 연안페리선의 전저항 및 환경 요소의 영향을 계산하고 다양한 구간과 환경 요소에서의 에너지효율 최적화 모델을 정립하였다.

Wang et al. (2018)은 다양한 선박 운항 변수 중에서 적합한 변수를 선정하기 위하여 라소 정규화를 적용하였으며 회귀모형에 의한 예측 결과를 인공 신경망, 서포트 벡터 머신, 가우시안 프로세스에 의한 모델과 비교하여 정확성을 평가하였다.

Table 1.1 Previous studies on the prediction of ship energy efficiency

Author	Type of ship	Independent variable	Dependent variable	Method
Pedersen & Larsen. (2009)	Oil tanker	STW, Wind speed, Wind direction, Wave height, Wave direction, Water depth, Water temperature, Air temperature	Shaft power	ANN
Petersen et al. (2012)	Coastal ferry	STW, Propeller pitch, Mean draft, Trim, Distant to the water, Wind speed, Wind direction, Rudder angle	Fuel consumption	ANN, GP
Lu et al. (2013)	Oil tanker	Shaft power, Ship resistance	Specific fuel consumption	Empirical formula

Beşikçi et al. (2016)	Oil tanker	RPM, SOG, Mean draft, Trim, Cargo quantity, Wind speed, Sea state	Fuel consumption	ANN, Regression
Wang et al. (2016)	Cruise ship	RPM, SOG, Shaft power, Wind speed, Water depth, Fuel consumption	Wind speed, Water depth	WNN
Yoo et al. (2016)	Container ship	RPM, SOG, M/E Torque, Ship resistance	Shaft power, Ship Speed	Empirical formula
Kim et al. (2017)	Container ship	RPM, SOG, STW, Shaft power, Mean draft, Trim, Displacement, Wetted surface area, Propeller immersion, BFS, Wind resistance	Fuel consumption	PLS regression
Yu et al. (2018)	Merchant ship	RPM, SOG, STW, Shaft power, Mean draft, Trim, Displacement, Wetted surface area, Propeller immersion, BFS, Wind resistance	Fuel consumption	ANN, Regression
Shengzheng et al. (2018)	Container ship	RPM, SOG, STW, Shaft power, Mean draft, Trim, Cargo quantity, GM, Displacement, BFS, Wind direction, Current, Wind wave height, Sig.Wave height, Swell height, Swell direction, CO2 efficiency, Heading, LOA, Beam	Fuel consumption	LASSO regression, SVM regression, ANN, GP
Yan et al. (2018)	Coastal ferry	SOG, STW, Wind speed, Wind direction, Water depth, Water speed, Ship position	Fuel consumption	Empirical formula
Parkes et al. (2018)	Merchant ship	SOG, STW, Mean draft, Trim, Wind speed, Wind direction, Wave height, Heading	Shaft power	ANN

1.3 연구의 목적 및 기대효과

지금까지 선행된 연료소모량의 예측과 관련된 연구에서는 연료소모량 예측모델을 구현하기 위하여 서로 다른 운항변수들이 사용되었으며, 실험식, 회귀 모형, 인공 신경망, 가우시안 프로세스 등 다양한 학습 기법들이 적용되었다. 예측모델에 활용되는 변수는 선박에서 수신가능한 모든 변수를 사용하거나 또는 전문가의 경험에 의존하여 선정하는 경우가 많았다. 무분별한 운항데이터의 사용은 예측모델의 과적합 문제(overfitting)와 변수 간의 높은 상관관계로 인한 다중공선성 문제(multicollinearity) 등을 야기할 수 있었으며 계산 시간과 비용 차원에서도 비효율적인 측면이 있었다. 따라서 선박 운항데이터의 상관관계와 특성을 고려한 합리적인 변수 선정에 대한 연구가 필요할 것으로 사료된다 (IMO, 2009b).

선박의 운항데이터는 진동이나 충격 또는 외부 환경적인 요인 등의 영향으로 종종 정상범주를 벗어나는 이상값(outlier)을 포함하며, 여러 센서들로부터 데이터가 취합되거나 정비 상의 목적 등으로 결측되는 구간(missing value) 등이 있어 불안정한 특성을 가진다 (Perera et al., 2017; Abbasian et al., 2018). 이러한 데이터의 특성으로 인하여 선박으로부터 수집되는 원시 데이터를 운항성능 판단 및 예측에 그대로 활용하기에는 무리가 있었다. 따라서 본 연구에서는 선박 데이터에 대한 충분한 이해를 바탕으로 적절한 정제 방법을 수행하고자 하며 정제된 데이터에 상관 분석 및 분산팽창지수, 주성분 분석 및 라소 정규화와 같은 차원 감소기법을 적용하여 연료소모량에 영향을 미치는 주된 운항변수를 찾아내고 다양한 변수들의 특징을 파악하고자 한다. 또한 운항데이터를 회귀모형과 인공 신경망에 적용하여 연료소모율을 예측하는 모형을 구축하고 분석하고자 한다.

선박의 에너지효율 예측모델로부터 다음과 같은 효과를 기대할 수 있다.

1) 항로계획시 운항자의 의사 결정지원 : 항해 시작 전 차항구의 입출항 일정, 선박의 하중상태, 해역의 기상상태 등에 따른 최소한의 연료를 소모하면서 주어진 시간 안에 목적지까지 도착할 수 있는 항로의 선정에 대한 운항자의 의

사결정을 지원해준다. 또한 항해 중 새롭게 업데이트되는 정보를 반영하여 변화되는 조건에 따른 실시간 예측을 수행하여 항해 최적화를 가능하게 해준다.

2) 선체 및 기기의 이상상태 탐지 : 선체의 결함 또는 주기관의 상태 이상이 발생하게 되면 연료효율이 감소하게 될 것이며 이는 운항 중 실시간으로 선박에서 관측되는 연료소모량과 예측모델에 의해 예측되는 값 사이에 차이가 발생하게 될 것이다. 또한 예측모델에 입력되는 기기의 센서에 이상이 발생하게 되면 예측 결과가 상이하기 때문에 이상상태를 탐지할 수 있게 될 것이다.

3) 장기 운항에 따른 성능저하 파악 : 장시간 선박 운항에 따른 선저부착물 및 해양유기체 등으로 인한 선체 저항 증가나 주기관 및 기타 장비의 성능 저하로 인한 추진효율 감소 등을 예측모델로부터 파악 가능하다.

4) 운항데이터에 대한 운항자의 이해도 향상 : 여러 변수들에 의한 연료효율을 미리 예측함으로써 선박 운항자가 전반적인 운항 요소와 연료효율 간의 상관에 대한 이해도를 향상시켜줄 수 있다.

1.4 논문의 구성

본 논문은 총 5장으로 구성되어 있다. 제 2장에서는 연구에 사용될 선박 운항데이터에 대한 소개와 이를 활용하여 연료소모율 예측모델을 생성하기까지의 전반적인 연구 과정을 서술한다. 제 3장에서는 예측모델을 생성하기 전 선박 원시 데이터의 전처리 과정에 관하여 다룬다. 제 4장에서는 제 3장에서 전처리한 데이터에 다양한 학습 알고리즘을 적용하여 예측모델을 구현하고 각 방법에 대한 성능과 한계를 분석한다. 마지막으로 제 5장에서는 본 연구에 대한 결론과 향후 연구방향에 대하여 기술한다.

제 2 장 연구 재료 및 방법

2.1 데이터 소개

본 연구에서는 13k TEU(Twenty-foot Equivalent Unit) 급 컨테이너선을 대상으로 연료소모율 예측모델에 관한 연구를 수행하였다. Table 2.1은 대상선박의 주요 제원이며, Fig. 2.1은 선박의 일반배치도(general arrangement)를 나타낸 것이다. 2014년 1월 4일부터 2014년 6월 29일까지 약 6개월간 선박이 운항하면서 발생한 데이터를 선박 감시 제어 시스템(Alarm Monitoring and control System; AMS)로부터 수집하였으며 한 항차 기준 약 83일 정도로 Fig. 2.2 및 Table 2.2와 같이 아시아-유럽 노선을 항해하였다. 수집된 변수의 개수는 총 392개이며 1분 간격으로 데이터가 측정 되었다. 선박의 운항변수를 활용한 예측모델 구현이라는 본 연구의 목적에 부합하기 위하여 수집된 데이터 중 선박에서 조정 가능한 변수 및 기상상태 등을 포함한 Table 2.3과 같은 운항변수를 취득하였다.

Table 2.1 Principal particulars of the target ship

Particular	Target Ship
LOA [m]	360.0
LBP [m]	345.0
Breadth [m]	46.0
Moulded depth [m]	28.0
Summer draft [m]	15.5
Deadweight [ton]	142,000.0
Displacement [ton]	185,000.0

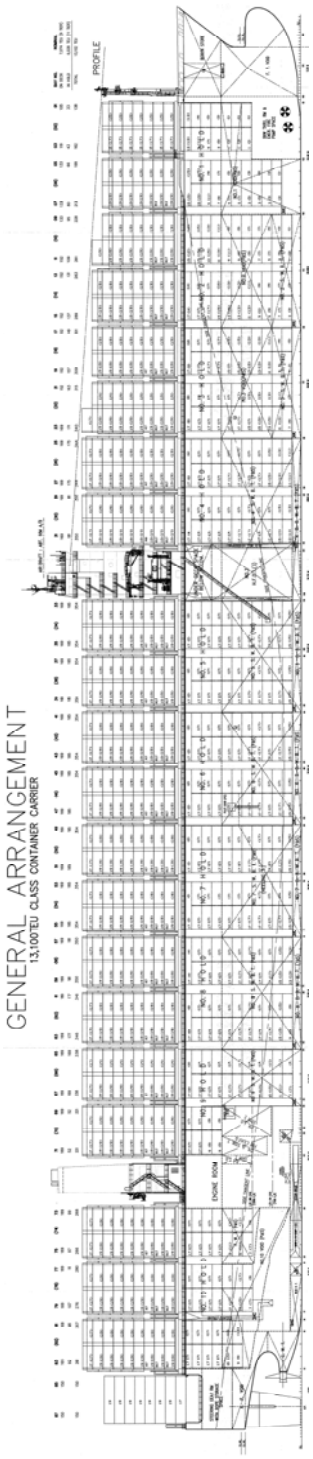

Fig. 2.1 General arrangement of the target ship

Fig. 2.2 Operational route of the target ship

Table 2.2 Port rotations of the target ship

	Port Rotation
West Bound	Xingang(China)-Kwangyang(Korea)-Pusan(Korea)-Shanghai(China)-Xiamen(China)-Yantian(China)-Singapore(Singapore)-Suez(Egypt)-Algeciras(Spain)-Hamburg(Germany)
East Bound	Hamburg(Germany)-Rotterdam(Netherland)-Le havre(France)-Suez(Egypt)-Singapore(Singapore)-Yantian(China)-Hongkong(China)-Xingang(China)

Table 2.3 Data list collected from the target ship

No.	Variable	Unit	Remark
1	M/E RPM		
2	Speed of the ground	knot	
3	Speed through water	knot	
4	Course of the ground	degree	
5	Heading	degree	000/090/180/270 degree refers to north/east/south/west direction
6	True wind speed	m/s	
7	True wind direction	degree	000/090/180/270 degree refers to northerly/easterly/southerly /westerly wind
8	Rudder angle	degree	
9	Mean draft	meter	
10	Trim	meter	(+) : Trim by the stern (-) : Trim by the head
11	Displacement	ton	
12	Wetted surface area	m^2	
13	Shaft power	kW	
14	M/E fuel consumption	ton/h	

2.2 연구 방법

Fig. 2.3은 연구의 흐름도를 나타낸 것이다. 선박으로부터 수집한 원시 데이터를 데이터 통합, 정제, 변환 및 축소의 전처리 과정을 통해서 분석에 적합한 양질의 데이터로 처리하였다. 분석단계에서는 전처리가 완료된 데이터를 다중선형 회귀 및 인공 신경망 모형에 적용하여 선박의 연료소모율 예측모델을 개발하였다. 마지막으로 평가단계에서는 개발한 예측모델의 예측 결과를 새로운 평가 데이터와 비교분석하였다.

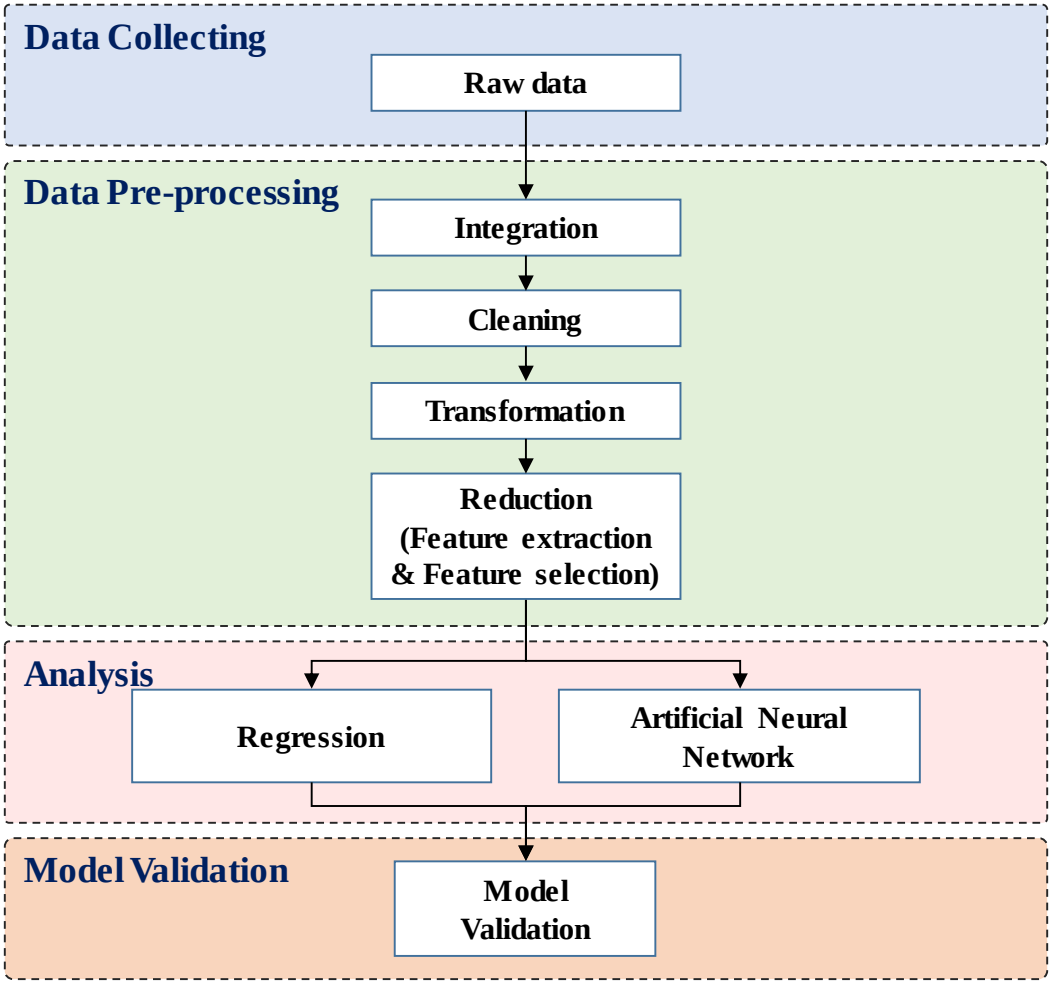

Fig. 2.3 Flow chart of the study

제 3 장 데이터 수집 및 전처리

3.1 데이터 수집

본 연구에서는 2014년 1월 4일부터 2014년 6월 29일까지 약 6개월간 13k TEU급 컨테이너선박이 운항하면서 발생한 데이터를 선박 감시 제어 시스템으로부터 수집하였다. 1분 간격으로 데이터가 저장되었으며 수집된 변수의 개수는 총 392개이다. 본 연구에서는 선박의 운항중 각 변수들의 관계를 파악하고 연료소모율 예측모델을 구현하는 것이 목적이기 때문에 전체 데이터 수집 기간 중 실제로 선박이 기관을 사용하여 항해를 한 기간의 데이터만을 취득하였다. 또한 선박의 입출항에 해당하는 항내 조선 구간의 경우 잦은 엔진 사용 및 여러 변수들의 작용으로 불안정한 데이터를 다수 포함하고 있기 때문에 분석 대상에서 제외하였다. Table 3.1은 선박데이터의 전체 수집 기간 중 항해, 정박/입출항 및 결측 구간에 대한 구분을 나타낸 것이다. 대상선박의 항내전속전진(maneuvering full ahead)시 주기관의 분당 회전수(M/E RPM) 50 미만은 정박/입출항 구간으로 취급하였다.

Table 3.1 Configuration of ship data

Ship operational data								
At sea			At Berth & Arrival & Departure			Missing section		
Month	Day	Hour	Month	Day	Hour	Month	Day	Hour
3	17	0	0	28	17	1	0	0

3.2 데이터 통합

데이터 통합 단계에서는 원시 데이터를 취합하여 분류하고 다양한 로그 파일의 형식을 통일시켜 일관성 있는 형태로 변환한다. Fig. 3.1은 처리 및 분석의 용이성을 위하여 수집된 데이터를 CSV(Comma Separated Values)파일 형식으로 취합한 예시이다.

NO.	Full date	GLL:Lat. - (	GLL:Lat. - (	GLL:Lat.De	Latitude	GLL:Long.	GLL:Long.	GLL:Long.[	Longitude	GLL: UTC (	GLL: UTC (	GLL:Status	1=A 2=D 3	VTG:Cours	VTG:Cours	VTG:Spee	VTG:Spee
1	2014-01-01 0:0	543	9146	1	5°42.8914(	9004	4531	1	90°4.1453	14	58.27	A	D	91	92.7	13	24.1
2	2014-01-01 0:1	543	8853	1	5°42.7885	9004	859	1	90°4.2085	14	59.12	A	D	91	92.6	13	24.1
3	2014-01-01 0:2	543	8394	1	5°42.7839	9004	3906	1	90°4.4390	15	0.16	A	D	91	92.8	13	24.1
4	2014-01-01 0:3	543	7871	1	5°42.7787	9005	8594	1	90°4.7859	15	1.15	A	D	91.4	93.1	13	24.1
5	2014-01-01 0:4	543	7417	1	5°42.7741	9005	391	1	90°4.9039	15	2.16	A	D	90.8	92.5	13	24.1
6	2014-01-01 0:5	543	7095	1	5°42.7709	9005	2266	1	90°5.2226	15	3.24	A	D	90.9	92.7	13.1	24.3
7	2014-01-01 0:6	543	6768	1	5°42.7676	9005	156	1	90°5.4015	15	4.03	A	D	91.2	92.9	13.1	24.3
8	2014-01-01 0:7	543	6392	1	5°42.7639	9006	6094	1	90°5.6609	15	5.21	A	A	91.1	92.8	13.1	24.3
9	2014-01-01 0:8	543	6001	1	5°42.7600	9006	5156	1	90°5.8515	15	6.16	A	A	90.9	92.6	13.1	24.3
10	2014-01-01 0:9	543	5610	1	5°42.7561	9006	9375	1	90°5.9937	15	7.13	A	D	90.4	92.1	13.1	24.3
11	2014-01-01 0:10	543	5342	1	5°42.7534	9006	156	1	90°6.2015	15	8.14	A	D	90.5	92.2	13.1	24.3
12	2014-01-01 0:11	543	5059	1	5°42.7505	9006	1406	1	90°6.4140	15	9.17	A	D	90.9	92.6	13.1	24.4
13	2014-01-01 0:12	543	4800	1	5°42.7480	9007	4922	1	90°6.7492	15	10.21	A	D	90.7	92.5	13.1	24.3
14	2014-01-01 0:13	543	4341	1	5°42.7434	9007	8359	1	90°6.9835	15	11.25	A	A	90.7	92.4	13.1	24.4
15	2014-01-01 0:14	543	4063	1	5°42.7406	9007	1094	1	90°7.2109	15	12.27	A	D	91.2	92.9	13.2	24.4
16	2014-01-01 0:15	543	3545	1	5°42.7354	9007	3125	1	90°7.4312	15	13.27	A	D	91.2	92.9	13.2	24.3
17	2014-01-01 0:16	543	3301	1	5°42.7330	9008	3984	1	90°7.5398	15	14.24	A	D	90.5	92.2	13.2	24.3
18	2014-01-01 0:17	543	3218	1	5°42.7321	9008	3516	1	90°7.7351	15	15.12	A	D	90.3	92	13.2	24.3
19	2014-01-01 0:18	543	3203	1	5°42.7320	9008	6875	1	90°7.9687	15	16.14	A	D	90.5	92.2	13.1	24.3
20	2014-01-01 0:19	543	2852	1	5°42.7285	9008	547	1	90°8.1054	15	17.25	A	D	91.1	92.7	13.1	24.3
21	2014-01-01 0:20	543	2344	1	5°42.7234	9008	938	1	90°8.4093	15	18.11	A	D	90.8	92.5	13.1	24.3
22	2014-01-01 0:21	543	2148	1	5°42.7214	9009	4297	1	90°8.7429	15	19.24	A	A	90.7	92.4	13.1	24.3
23	2014-01-01 0:22	543	1914	1	5°42.7191	9009	3750	1	90°8.9375	15	20.08	A	A	90.9	92.6	13.1	24.4
24	2014-01-01 0:23	543	1484	1	5°42.7148	9009	7656	1	90°9.1765	15	21.14	A	D	90.6	92.3	13.1	24.4
25	2014-01-01 0:24	543	1318	1	5°42.7131	9009	9375	1	90°9.2937	15	22.16	A	D	90.4	92.1	13.2	24.4
26	2014-01-01 0:25	543	1265	1	5°42.7126	9010	469	1	90°9.6046	15	23.14	A	A	90.3	92	13.2	24.4
27	2014-01-01 0:26	543	1050	1	5°42.7105	9010	2891	1	90°9.7289	15	24.2	A	D	90.7	92.3	13.2	24.4
28	2014-01-01 0:27	543	762	1	5°42.7076	9010	6719	1	90°9.9671	15	25.2	A	A	90.5	92.2	13.2	24.3
29	2014-01-01 0:28	543	596	1	5°42.7059	9010	7813	1	90°10.278	15	26.17	A	A	90.5	92.2	13.2	24.4
30	2014-01-01 0:29	543	435	1	5°42.7043	9010	7344	1	90°10.373	15	27.12	A	D	90.9	92.6	13.2	24.4
31	2014-01-01 0:30	543	29	1	5°42.7002	9011	1953	1	90°10.719	15	28.15	A	D	90.6	92.3	13.2	24.3
32	2014-01-01 0:31	543	9678	1	5°42.6967	9011	5391	1	90°10.853	15	29.26	A	D	90.8	92.4	13.2	24.3

Fig. 3.1 An example of operational data collected from the ship

3.3 데이터 정제

데이터 세트에는 일반적으로 불필요 데이터, 이상치 및 결측값 등이 존재한다. 이러한 데이터는 분석 시 계산복잡성을 증가시키며 예측모델의 정확성을 감소시키는 역할을 한다. 따라서 데이터가 부족한 부분을 보완 또는 대체하거나 센서의 오작동이나 외력에 의해 잘못 관측된 이상치를 제거할 필요가 있다.

3.3.1 결측값

선박에 장착된 기기들은 센서의 노후, 불량 또는 기타 외부환경 등의 영향으로 고장이 발생하거나 일정기간동안 안전상 또는 정비 등의 이유로 전원이 차

단되는 경우가 있다. 이로 인하여 일정 기간에 대한 센서 데이터에 불연속한 결측값이 발생하게 된다. 결측값은 결측 구간의 데이터세트를 제거하거나 평균, 중앙값 또는 회귀 분석을 이용한 예측값으로 대체하는 방법 등이 일반적으로 활용된다. 결측치를 다른 값으로 대체하는 경우 실제와 다른 부적절한 값이 입력될 수도 있기 때문에 해당하는 데이터 특성에 대한 충분한 이해와 배경지식을 가지고 적절하게 처리할 필요가 있다. 현재 연구에서는 운항 변수들 간의 관계를 파악하고 정확도 높은 예측모델을 파악하기 위하여 결측이 있는 데이터 세트를 임의로 대체하지 않고 제거하였다.

Fig. 3.2는 수집한 선박 운항데이터의 결측 구간 예시를 보여준다. 2014년 2월 8일부터 2월 14일간 수집한 선박의 대지속력 데이터이며 점선으로 표시한 2월 9일 00시부터 2월 14일 00시까지의 데이터가 결측된 것을 알 수 있다. 해당하는 데이터 수집기간 전후로 일정한 선속이 관측된 것으로부터 센서의 오류가 발생하였거나 또는 기타 목적으로 전원이 차단된 것으로 판단된다.

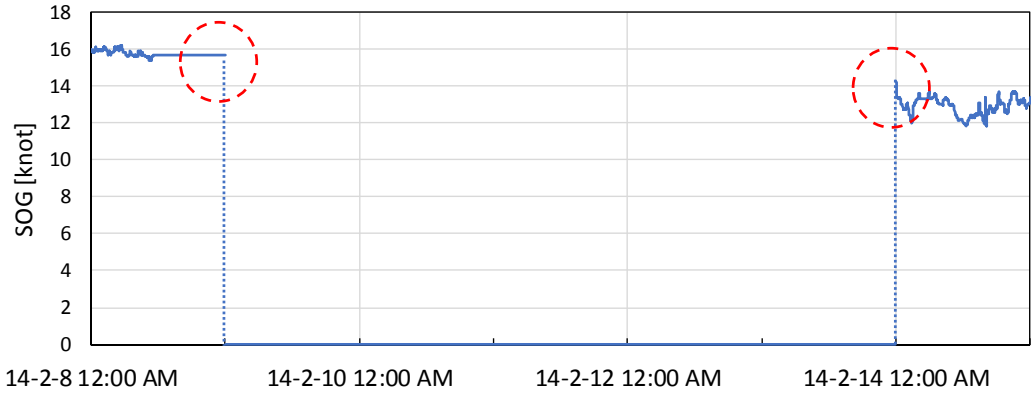

Fig. 3.2 Missing section part in collected data

3.3.2 이상값

이상값은 측정된 값에서 임의의 오류나 변화가 발생하는 것으로서 데이터의 처리를 방해하는 신호를 의미한다. 선박의 에너지효율 계산에 활용되는 Noon report 데이터의 경우 일반적으로 선박 운항자들이 직접 시스템 상에 입력하기

때문에 실수로 오정보가 입력될 가능성을 배제할 수 없으며 이는 일관성 없는 데이터 샘플을 야기한다. 또한 선박에서 수집되는 데이터는 선박의 운항 특성상 진동이나 충격 또는 외부 환경적인 요인 등에 영향을 받을 수 있기 때문에 종종 불안정하거나 정상범주를 벗어나는 잡음값이 발생되기도 한다.

본 연구에서는 다음과 같은 방법을 적용하여 선박 운항데이터의 이상값을 처리하였다.

1) Fig. 3.3은 대지속력의 시계열 데이터의 예시를 나타낸 것이다. 2월 6일 05시경부터 19시경까지 데이터 값이 변동 없이 일정한 것을 확인할 수 있다. 이처럼 일정시간동안 특정한 이유 없이 값이 일정한 경우 기기나 센서의 오작동으로 간주하여 해당하는 데이터 세트를 제거하였다.

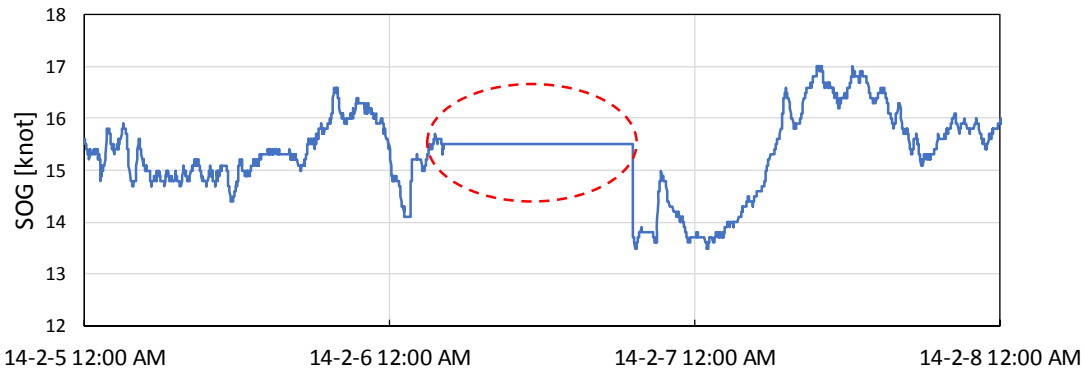

Fig. 3.3 Sensor fault part in collected data

2) 기기 및 센서의 매뉴얼 및 전문가의 경험을 토대로 정상적인 작동 범위를 크게 벗어나는 경우 이상값으로 판단하여 제거하였다. 예를 들어, 풍향 또는 선수방위가 0-360도를 벗어나는 경우 연료소모량과 같이 데이터의 특성상 음의 값을 가질 수 없음에도 불구하고 음수인 경우 등이 있다.

3) Fig. 3.4는 정규분포의 3시그마 법칙을 나타낸 것이다. 정규 분포의 데이터에서 전체의 약 68%는 평균으로부터 1 표준 편차, 약 95%는 2 표준 편차, 약 99.7%는 3 표준 편차 안에 있는 것으로 알려져 있으며 데이터의 이상값을 식별하는데 많이 활용되고 있다. 관측값이 평균으로부터 일정 수의 표준 편차만큼

떨어져 있는 경우 해당 데이터 지점을 이상값으로 식별하며 통상적으로 3 표준편차가 사용된다 (Pukelsheim, 1994).

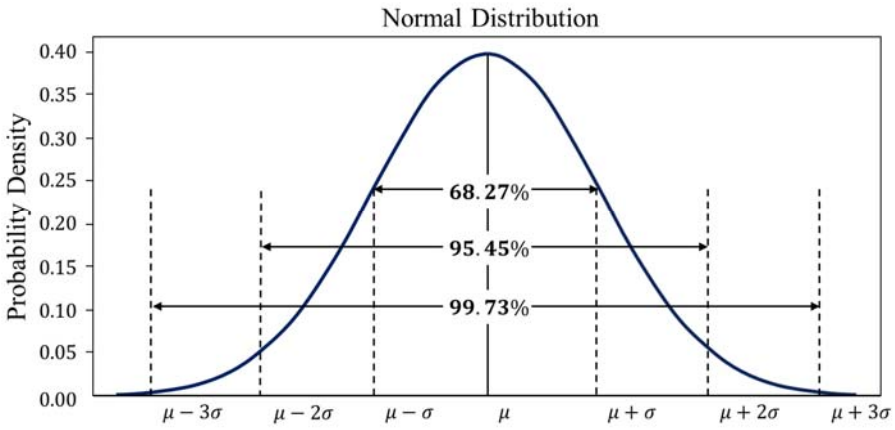

Fig. 3.4 Confidence intervals by the 3-sigma rule of the normal distribution

본 연구에서는 관측치와 예측치의 차(residual)가 정규분포를 이룬다는 가정하에 평균으로부터 3 표준편차를 벗어난 범위의 데이터를 이상치로 판단하였으며 Fig. 3.5와 같이 운항변수 중 선형관계의 데이터 분포를 나타내는 기관출력과 연료소모량에 대한 회귀 모형을 사용하여 이상값을 탐지하였다. 기관출력과 연료소모량의 원시 데이터에 대한 산점도를 별표로 나타내었으며 실선이 관측치들의 평균값으로 나타낸 회귀선(regression line)을 의미한다. 회귀선을 기준으로 ± 3 표준편차(standard deviation) 이내에 해당하는 값을 원으로 표기하였다.

Fig. 3.6은 Fig. 3.5에서 ± 3 표준편차를 벗어나는 관측값 중에 파선 원으로 표기한 점에서의 전후 시계열 데이터를 예시로 나타낸 것이다. 기관 출력과 연료소모량에 대한 변화 추세를 볼 때 해당하는 점에서만 과도한 값이 관측된 것을 알 수 있다. 이와 같이 회귀 모형의 잔차에 대한 표준편차로부터 탐지한 이상값 데이터는 원시 데이터로부터 제거하였다.

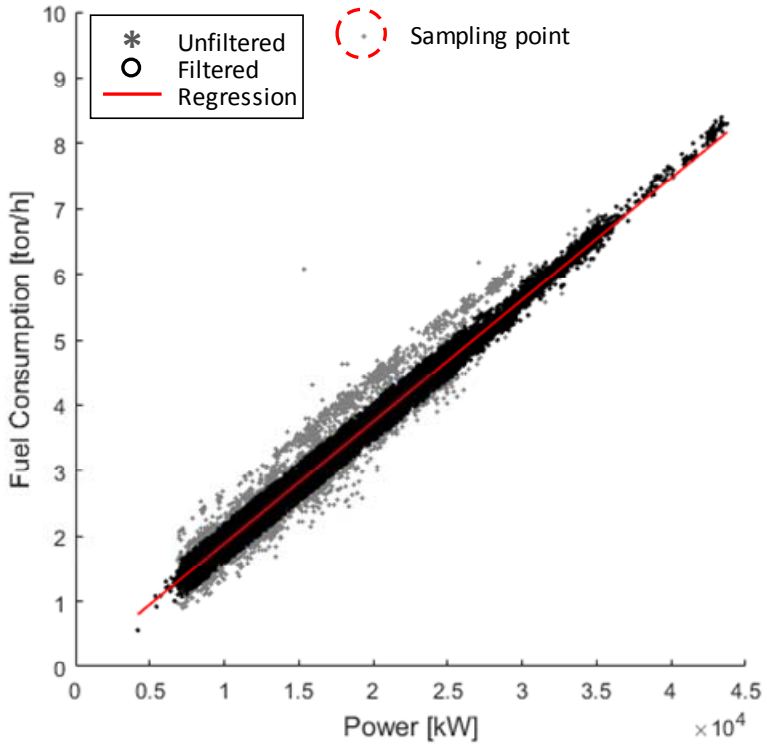

Fig. 3.5 Outlier detection by the relationship between engine power and fuel consumption

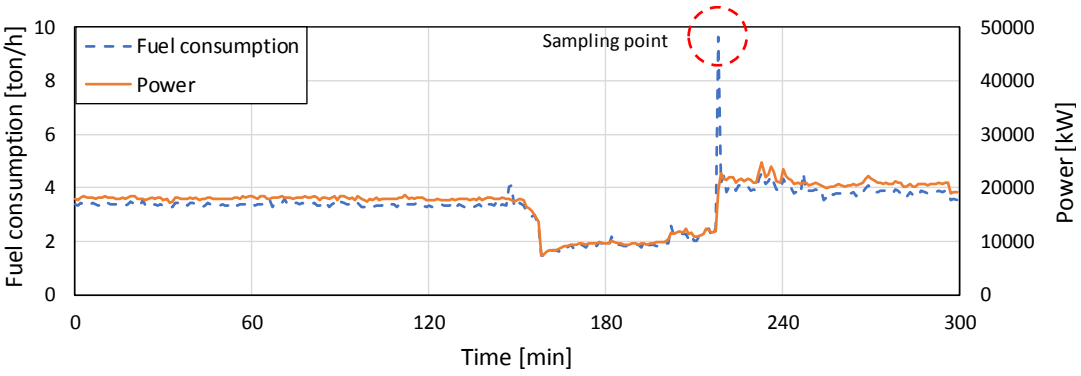

Fig. 3.6 Analysis of data identified as an outlier by the 3-sigma rule

4) 데이터의 잡음을 처리하기 위한 필터링 방법에는 이동평균 필터(moving average filter), 중앙값 필터(median filter), 가우시안 필터(gaussian filter) 등 다양한 기법들이 존재한다. 이동평균 필터의 경우 특정 시간 간격의 윈도우를 이동하면서 해당하는 시간동안 관측값들을 평균하여 현재 값을 대체함으로써 경향선을 구하는 방법이다. 데이터 세트 내의 불규칙한 변동을 완화해 주는 장점이 있지만 항상 잡음 데이터를 포함하여 평균을 계산하기 때문에 불규칙적으로 큰 잡음이 섞인 데이터의 경우 필터링 효율이 떨어지는 경향이 있다. 반면에 중앙값 필터는 관측값의 주변값들을 오름 또는 내림차순으로 정렬하여 해당하는 관측값을 정렬의 중앙값으로 대체하는 방법이다. 이는 관측한 데이터가 전부 계산에 반영되는 것이 아니기 때문에 과도한 신호잡음을 제거하는데 탁월하며 전반적인 추세를 표현할 수 있다 (Pratt, 2007).

Fig. 3.7(a)-3.7(c)는 선박의 시계열 데이터에 중앙값 필터를 적용한 결과를 원시 데이터와 비교한 것이다. 파선이 필터를 적용하기 전의 원시 데이터이며 실선으로 나타낸 부분이 중앙값 필터를 적용한 결과이다. Pedersen and Larsen (2009) 및 Perera and Mo (2017)는 선박으로부터 수집한 운항데이터를 10-15분 간격으로 평균하여 분석에 활용한 바 있다. 본 연구에서는 원시데이터에 중앙값 필터를 적용하였으며 10분 간격의 윈도우를 사용함으로써 필터링을 보다 정밀하게 수행하고자 하였다. 윈도우 간격에 따라서 데이터의 필터링 결과가 달라질 수 있기 때문에 이와 관련해서는 추후 연구를 수행할 필요가 있을 것으로 판단된다. 평균흘수나 트림 데이터의 경우 실제로 선박의 동요주기 또는 외력 등에 의하여 수시로 값이 변화하나 1분 간격의 데이터를 수집하였기 때문에 전체적으로 데이터가 불안정하며 일정한 추세변화를 파악하기가 어렵다. 또한 연료소모량의 경우 일부 구간에서 값이 튀는 현상이 있으며 이는 주기관으로 입출력되는 연료의 유량계로부터 계산되는 과정에서 발생한 잡음으로 파악된다. 따라서 본 연구에서는 중앙값 필터 방법을 적용하여 데이터의 잡음 및 불안정한 경향성에 대한 문제점을 보완하고자 하였다. 중앙값 필터에서는 원시데이터의 특성을 유지하면서 잡음을 효과적으로 제거할 수 있는 적절한 윈도우 크기를 설정할 필요가 있다.

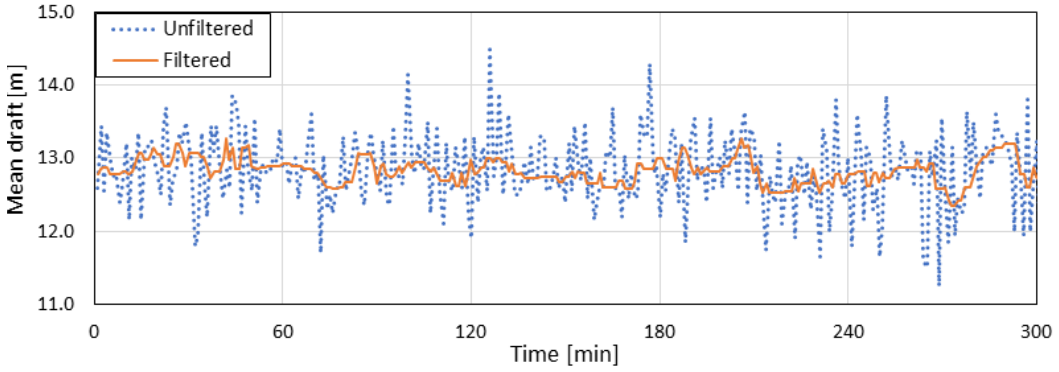

Fig. 3.7(a) Time series data of mean draft filtered by median filter

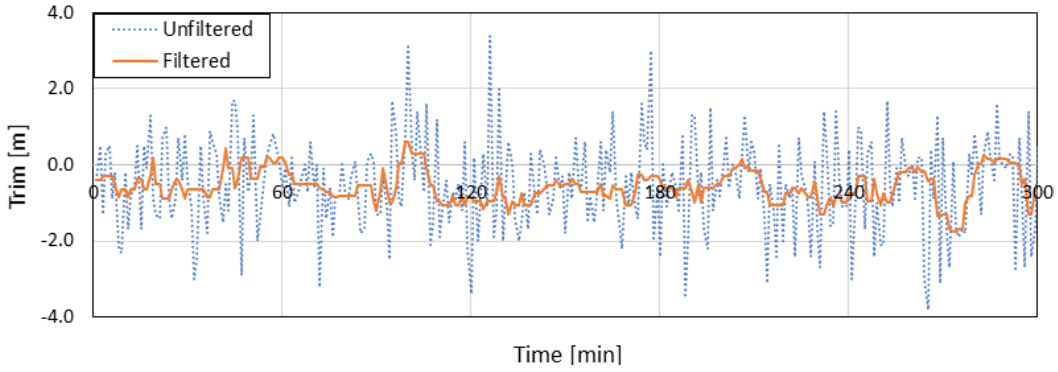

Fig. 3.7(b) Time series data of trim filtered by median filter

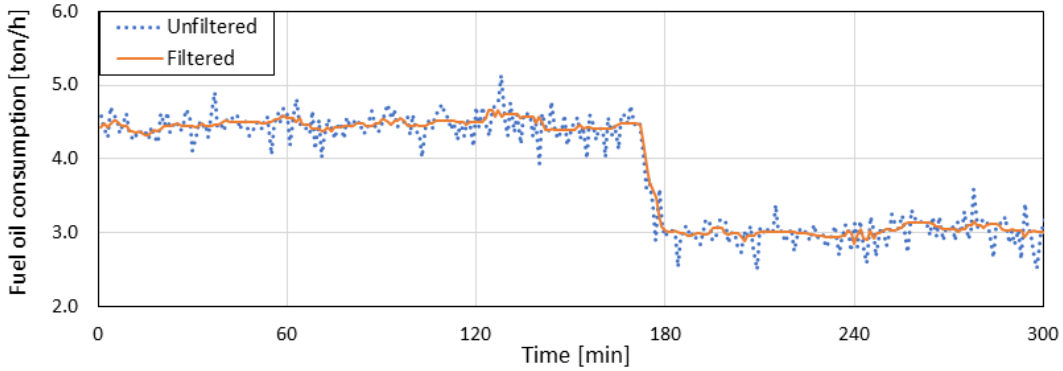

Fig. 3.7(c) Time series data of fuel consumption filtered by median filter

3.4 데이터 변환

3.4.1 변수 변환

데이터 변환은 기존의 변수를 조합하여 새로운 변수로 변환하거나 새로운 속성을 가진 데이터로 변형해주는 것을 의미한다. 선행 연구나 실험식 등으로부터 변수들의 관계가 명확히 밝혀져 있고 기존의 변수를 조합한 새로운 변수가 종속변수에 더 직접적인 연관이 있다면 데이터 변환을 통해서 모델의 예측 성능 향상을 도와줄 필요가 있다. 이러한 데이터 변환 과정은 모델 생성에 불필요한 변수의 개수를 줄이고 계산 복잡성을 해소시켜주는 역할을 한다.

1) 선박의 진침로, 대지속력, 진풍향, 진풍속과 같은 각 변수들은 선박이 바람 성분으로부터 받는 외력과 관련이 있지만 이러한 효과를 직접적으로 나타내지는 못하는 단점이 있다. 선박의 속력 및 침로를 고려하여 상대풍향 및 상대풍속을 계산함으로써 선박의 연료소모량이 바람성분으로부터 받는 영향을 고려하고자 하였다. 겉보기 바람 A는 운항하는 배 위에서 측정한 풍속으로 식 (3.1)과 같이 실제 바람의 속력과 선박 이동속도의 벡터 합으로 나타낼 수 있다. 겉보기 바람의 방향 β는 식 (3.2)와 같이 계산할 수 있다.

$$A = \sqrt{W^2 + V^2 + 2WV\cos\alpha} \tag{3.1}$$

$$\beta = \arccos\left(\frac{W\cos\alpha + V}{A}\right) \tag{3.2}$$

여기에서 V는 선박의 속력, W는 실제 바람의 속력, α는 선박의 선수방위를 기준으로 계산한 실제 바람의 풍향, A는 겉보기 바람의 속력, β는 겉보기 바람의 방향이다.

기존의 진풍향과 진풍속 데이터는 상대풍향과 상대풍속으로 대체하였으며 대지침로 및 선수방위 데이터는 상대풍향 및 상대풍속의 계산에 반영되었고 각각의 변수는 연료소모량에 관계가 없기 때문에 분석 대상에서 제외하였다. 또한 선박에서 조우하는 바람은 좌현 또는 우현의 각도가 동일한 경우 영향이 같기 때문에 선수방위를 기준으로 0-180도로 변환하였다. 따라서 상대풍향은 0도가 본선을 기준으로 선수에서 불어오는 바람이며 180도가 선미에서 불어오는 바람을 의미한다.

2) 프로펠러축의 각속도는 식 (3.3)과 같이 기관 분당 회전수로 계산할 수 있으며 기관 출력은 식 (3.4)와 같이 프로펠러축의 토크와 각속도의 곱으로 나타낼 수 있다. 따라서 중복되는 변수의 사용을 방지하기 위하여 분석 대상 변수에서 기관 출력과 기관 분당 회전수 데이터 중 하나를 제외함이 바람직하다. 기관의 출력은 선박 운항자가 선박의 운항중 즉각적으로 확인하기에는 용이하지 못하며 통상적으로 주기관의 분당 회전수를 이용하여 속력을 조정하기 때문에 기관 출력을 제외하였다.

$$\omega = \frac{M/E\,RPM}{60} * 2\pi \tag{3.3}$$

$$\begin{aligned} Shaft\ power &= T \times \omega \\ &= \frac{T \times M/E\,RPM \times 2\pi}{60} \end{aligned} \tag{3.4}$$

3) 기존 연구에서는 단위 시간 당 연료소모량을 종속변수로 하여 예측모델을 생성하는 경우가 많았다. 하지만 항해사가 선박의 운항 전 항해 계획을 세우거나 운항 중에 선박의 상태를 모니터링 하는 경우 단순히 관측되는 연료소모량만으로 선박의 효율적인 운항 상태를 판단하기는 어려웠다. 단편적인 예로, 선박에서는 주기관의 출력을 높일수록 외력의 영향을 많이 받을수록 연료소모량

이 증가하지만 연료소모량이 크더라도 단위 연료소모량 당 항해한 거리가 큰 경우는 오히려 연료효율이 좋을 수도 있기 때문이다. 이처럼 실질적인 선박의 운항효율을 판단하기 위해서는 연료소모대비 효율을 파악할 필요가 있다. 따라서 본 논문에서는 식 (3.5)와 같이 단위 항해거리 당 연료소모량을 계산하여 연료효율을 파악하였다. 이는 단위 시간 당 연료소모량을 단위 시간 당 항해거리로 나누어 계산하였으며 기존의 단위 시간 당 연료소모량 변수를 대체하였다.

$$Fuel\ consumption rate\ [ton/mile] = \frac{Fuel\ oil\ consumption\ per\ hour}{Navigation distance\ per\ hour} \tag{3.5}$$

Table 3.2는 데이터 정제과정 및 변수 변환을 수행한 후의 각 운항변수에 대한 기술통계량을 나타낸 것이다. 주기관의 분당 회전수, 대지속력, 대수속력, 상대풍속, 상대풍향, 타각, 평균흘수, 트림, 배수량, 침수표면적이 선박 연료소모량을 예측하기 위한 모델의 독립변수가 되며 단위 항해거리 당 연료소모량이 종속변수이다.

Table 3.2 Descriptive statistics of operational variables

No	Variable Name	Min	Max	Average	Standard deviation
1	M/E RPM	50.00	90.00	63.45	7.47
2	Speed of the ground [knot]	8.00	22.40	14.71	2.01
3	Speed through water [knot]	7.40	21.80	14.54	1.88
4	Relative wind speed [knot]	0.00	72.89	19.61	10.15
5	Relative wind direction [degree]	0.00	180.00	43.07	43.45
6	Rudder angle [degree]	0.00	36.10	1.64	2.43
7	Mean draft [meter]	11.25	15.70	14.26	0.94
8	Trim [meter]	-2.30	2.15	-0.05	0.58
9	Displacement [ton]	123466.40	184009.60	163225.08	13240.03
10	Wetted surface Area [m^2]	13108.02	14514.28	14047.80	292.93
11	Fuel consumption rate [ton/mile]	0.08	0.48	0.21	0.05

Fig. 3.8은 데이터의 필터링 전과 후를 히스토그램으로 비교한 것이다. 데이터의 통합, 정제, 변환 과정을 거치면서 분석에 불필요한 데이터, 이상값 및 결측값 등이 제거되었다.

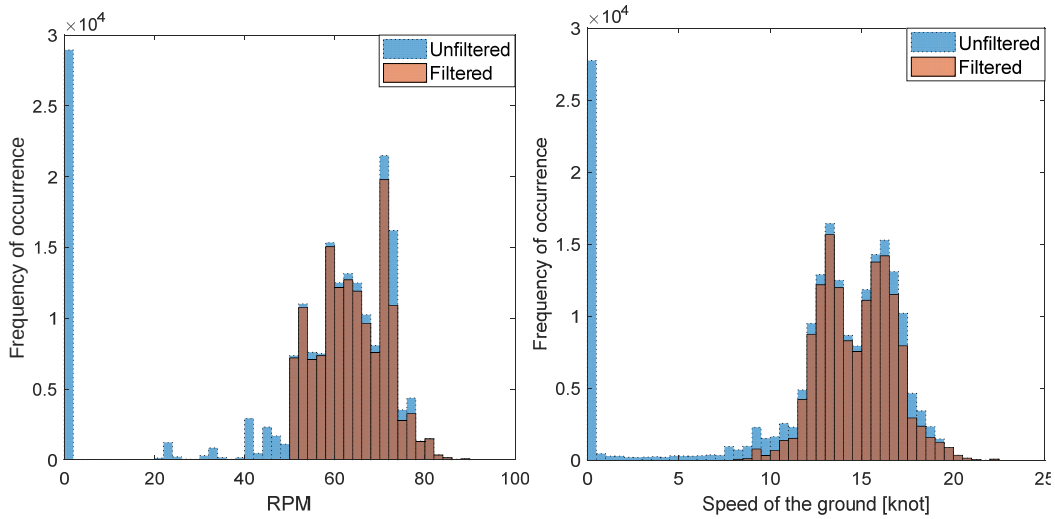

Fig. 3.8(a)(b) Comparison of filtered and unfiltered histograms of (a)M/E RPM (b)speed of the ground

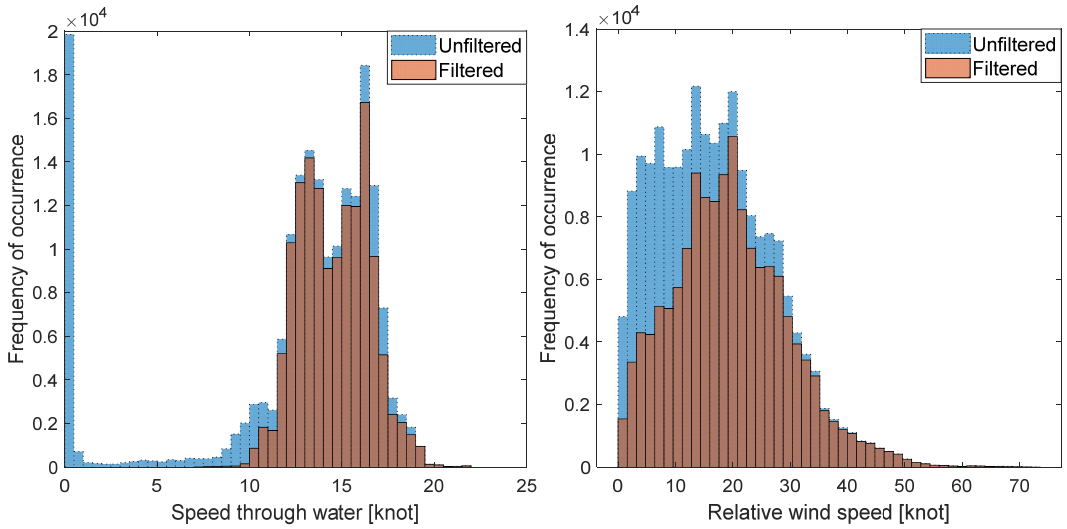

Fig. 3.8(c)(d) Comparison of filtered and unfiltered histograms of (c)speed through water (d)relative wind speed

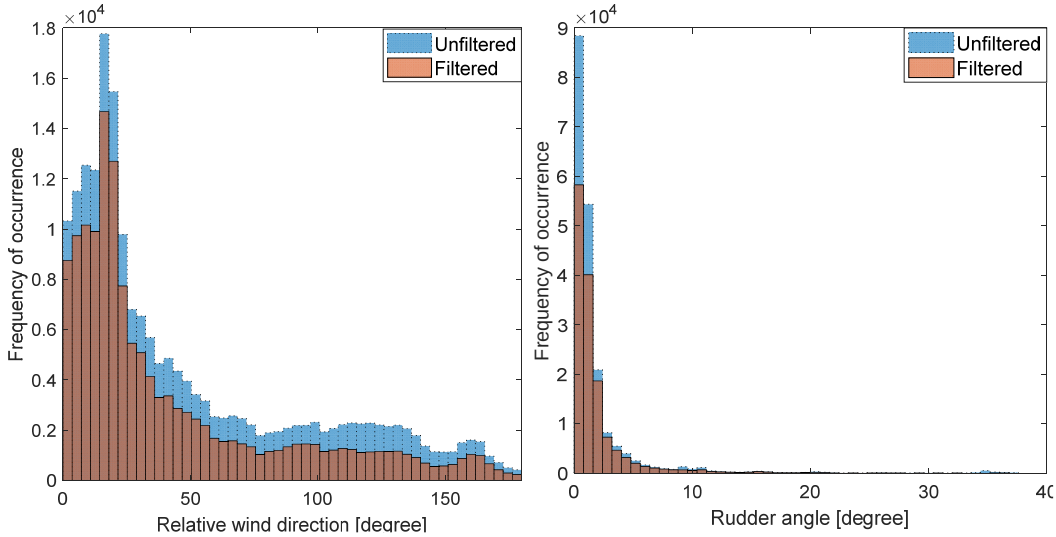

Fig. 3.8(e)(f) Comparison of filtered and unfiltered histograms of (e)relative wind direction (f)rudder angle

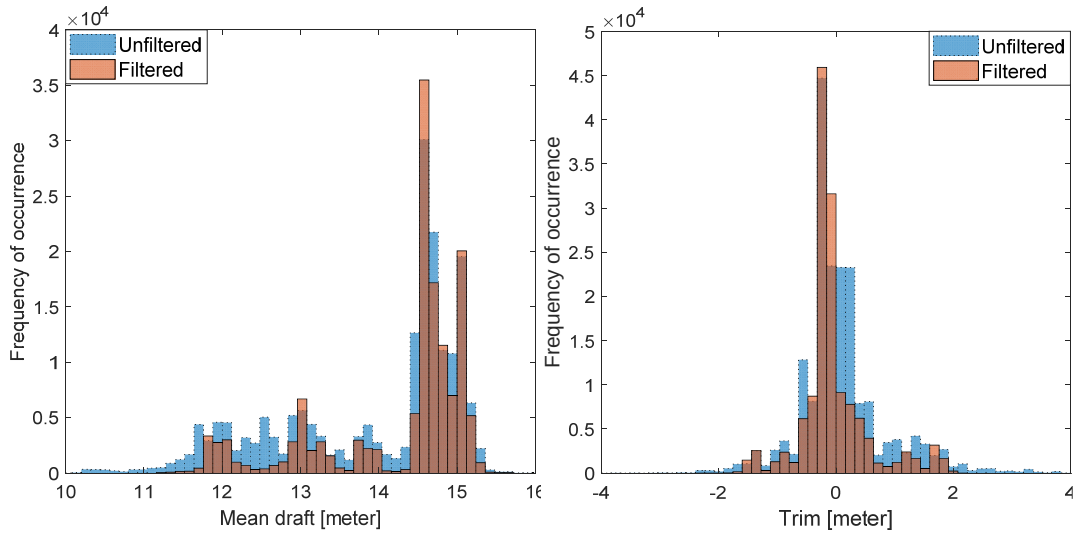

Fig. 3.8(g)(h) Comparison of filtered and unfiltered histograms of (e)mean draft (f)trim

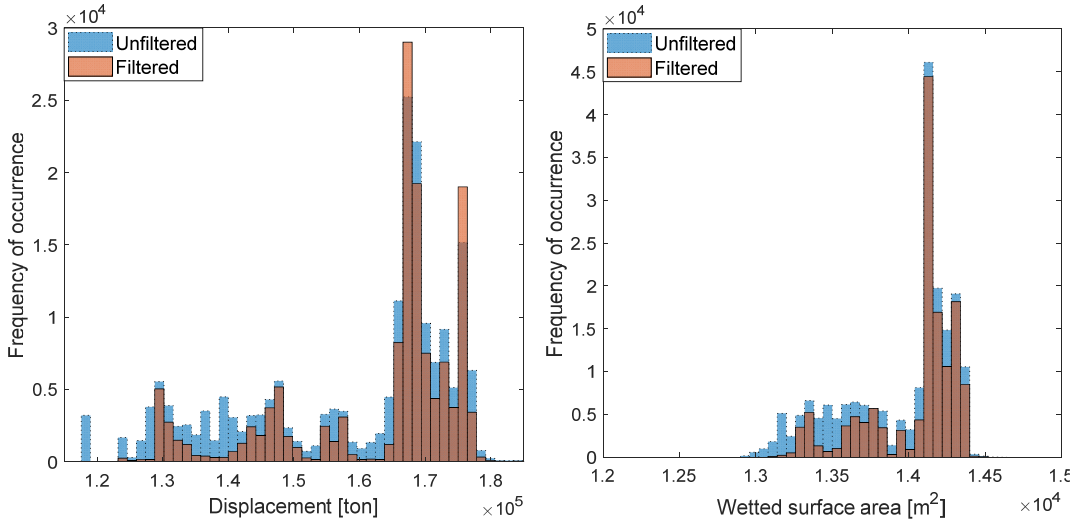

Fig. 3.8(i)(j) Comparison of filtered and unfiltered histograms of (i)displacement (j)wetted surface area

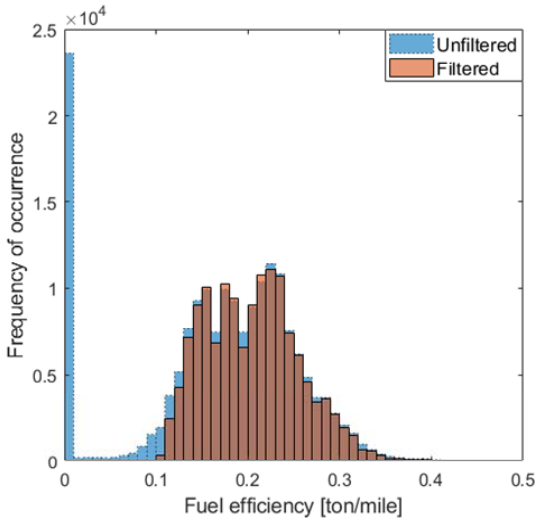

Fig. 3.8(k) Comparison of filtered and unfiltered histograms of fuel consumption rate

3.4.2 표준화

일반적으로 수집된 데이터의 변수들은 각각 다른 단위체계를 사용하고 있기 때문에 추후 회귀계수 추정시 각 변수들이 종속변수에 미치는 영향을 회귀계수

로부터 파악하기 어렵게 만든다. 따라서 본 연구에서는 표준화 기법 중 데이터 분석에 가장 일반적으로 사용되는 Z값 변환(Z-score)을 수행하여 데이터를 변환하고자 하였다. Z값 변환은 주어진 데이터의 산술 평균과 표준 편차를 사용하며 식 (3.6)과 같이 각 관측치와 평균값의 차를 표준 편차로 나누어서 계산한다. 정규화된 변수 Z는 평균이 0이고 표준편차가 1인 특징을 갖는다.

$$z = \frac{x - \mu}{\sigma} \tag{3.6}$$

여기에서 x는 원시 데이터의 관측값, σ는 원시 데이터의 표준편차, μ는 평균이며, z는 표준화된 값이다.

Fig. 3.9는 모델의 독립변수에 해당하는 운항 변수들의 Z값 변환을 수행한 것이다. 각 변수의 평균이 0으로 표준편차가 1로 변환된 것을 확인 할 수 있다.

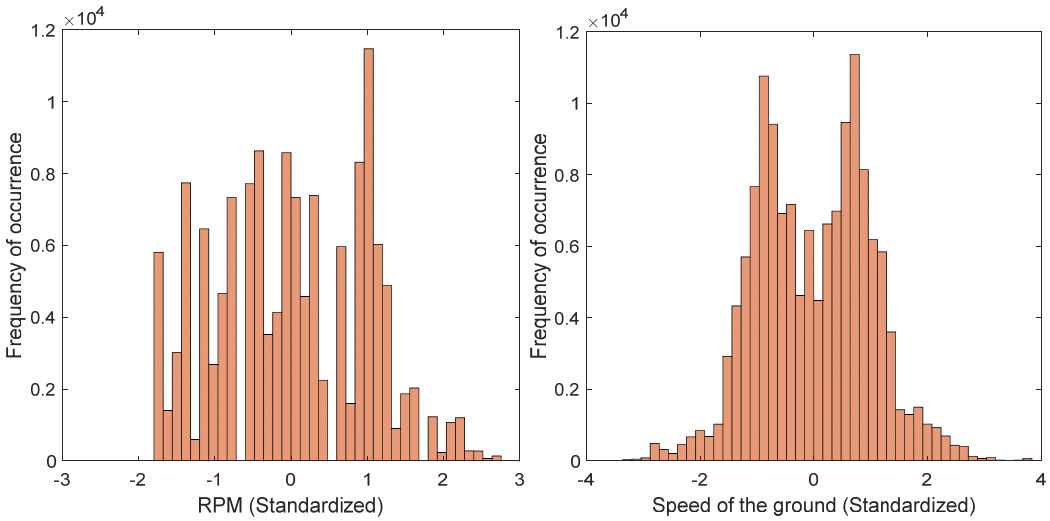

Fig. 3.9(a)(b) Histogram of standardized (a)M/E RPM (b)speed of the ground

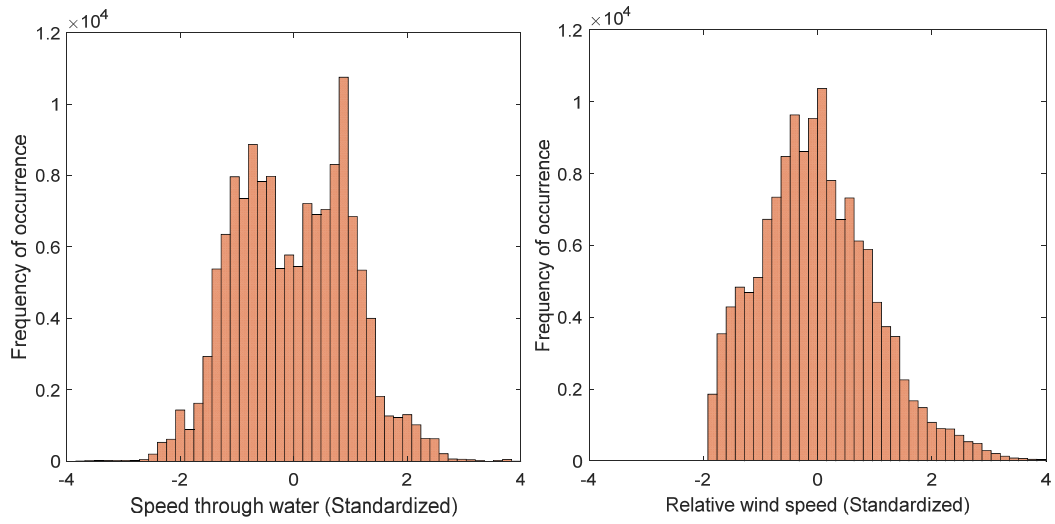

Fig. 3.9(c)(d) Histogram of standardized (c)speed through water (d)relative wind speed

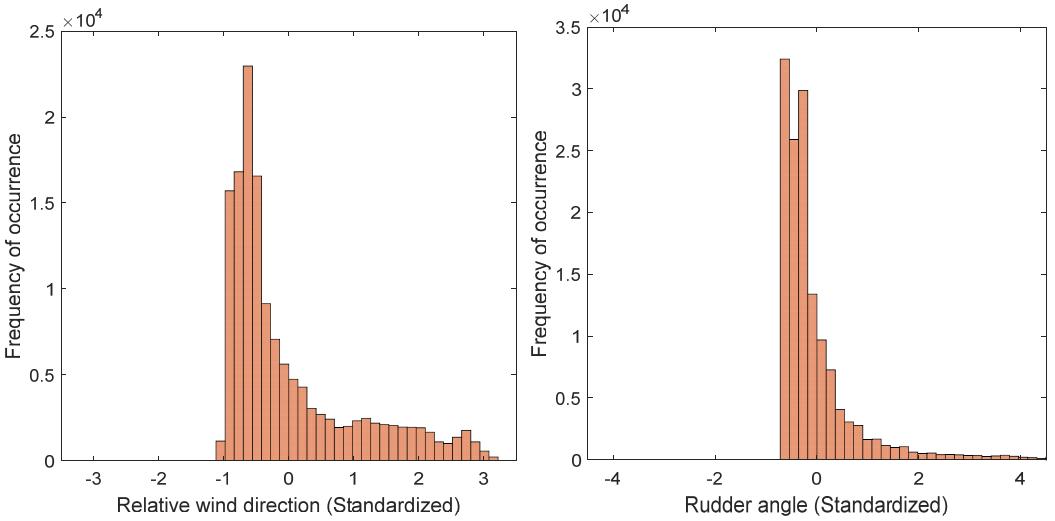

Fig. 3.9(e)(f) Histogram of standardized (e)relative wind direction (f)rudder angle

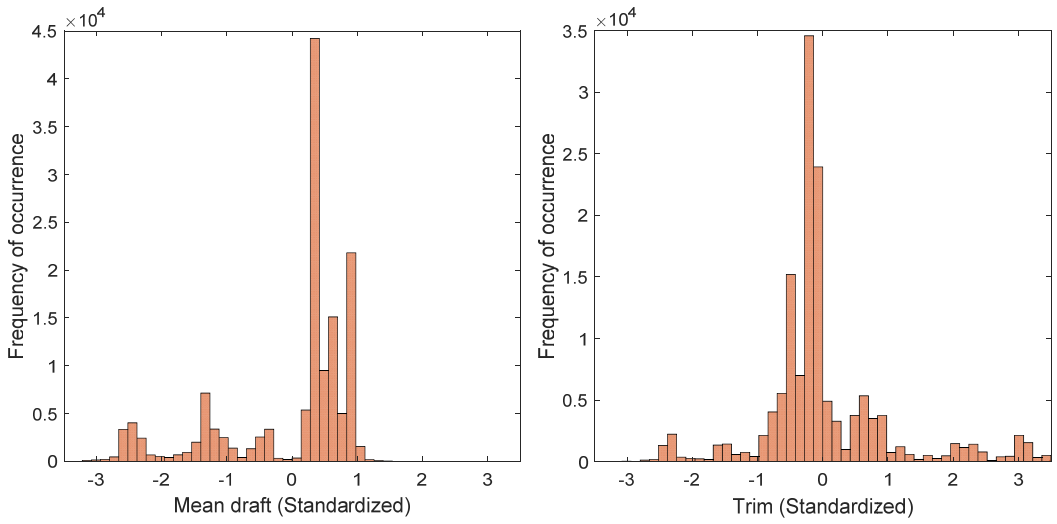

Fig. 3.9(g)(h) Histogram of standardized (g)mean draft (h)trim

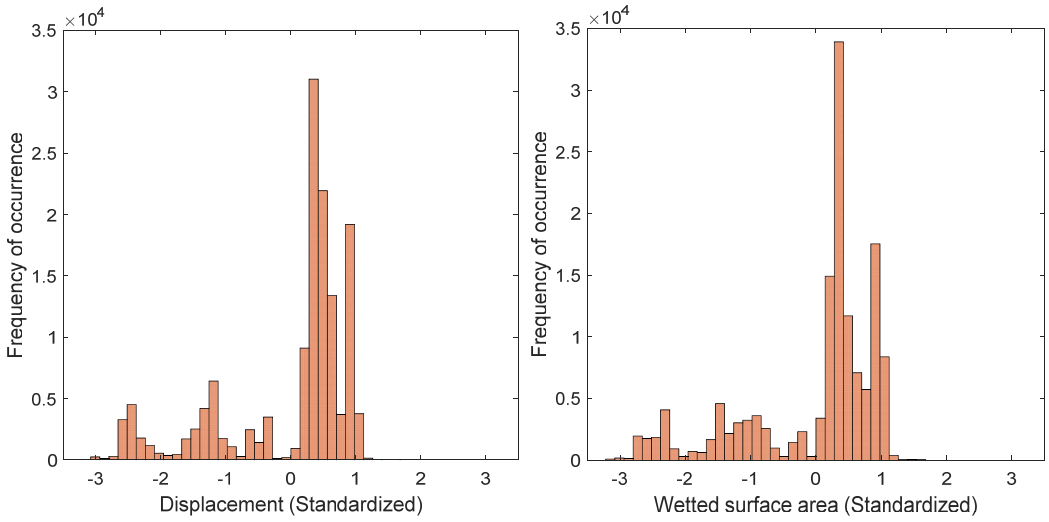

Fig. 3.9(i)(j) Histogram of standardized (i)displacement (j)wetted surface area

3.5 데이터 축소

데이터 축소는 일반적으로 데이터 세트의 부피를 줄이거나 고차원의 데이터를 저차원의 데이터로 변환하는 것을 말한다. 차원 축소의 방법은 크게 특징 선택기법(feature selection)과 특징 추출 기법(feature extraction)으로 구분할 수

있다. 특징 선택기법은 원시 데이터에서 모델 구축에 관련된 특징들의 부분 집합을 직접적으로 선택하는 방법으로 라소 정규화, 리지 정규화, Elastic Net과 같은 벌점 함수를 활용한 회귀기법들이 있다. 특징 추출기법은 원시 데이터에서 특징들의 선형 결합으로 이루어진 새로운 특징을 만들어 내는 것을 말하며 그 예로는 주성분 분석, 부분최소제곱법 등이 있다.

차원축소방법은 다음과 같은 문제점을 해결하는데 이점을 가진다.

1) 고차원 데이터에 대한 모델 생성시 불필요한 데이터를 포함할 수 있기 때문에 국부적으로 과적합될 수가 있으며 이는 새로운 데이터에 대한 예측을 어렵게 한다.

2) 종속변수를 설명하는 독립변수가 많아질수록 예측력은 높아지나 독립변수들 간의 상관관계가 크게 존재하는 경우 중복되는 정보 사용 등으로 인하여 다중공선성 문제가 발생할 수 있으며 이는 모형을 불안정하고 예측 성능을 저하시킨다.

3) 데이터 세트의 부피가 커질수록 이를 처리하기 위한 상당한 시간과 비용이 소요된다.

3.5.1 상관 분석 및 분산팽창지수에 의한 변수 선택

데이터 변환과정을 거쳐서 전처리된 데이터로 모델을 구현하기에 앞서, Fig. 3.10과 같이 전체 운항변수에 대한 상관 분석을 수행하였다. 상관 분석의 이론적 배경은 부록 A.1에 제시하였다. 상관계수의 절대값이 클수록 변수간의 강한 상관관계가 있음을 나타내며 이는 각 배열 칸 내부의 색으로도 확인이 가능하다.

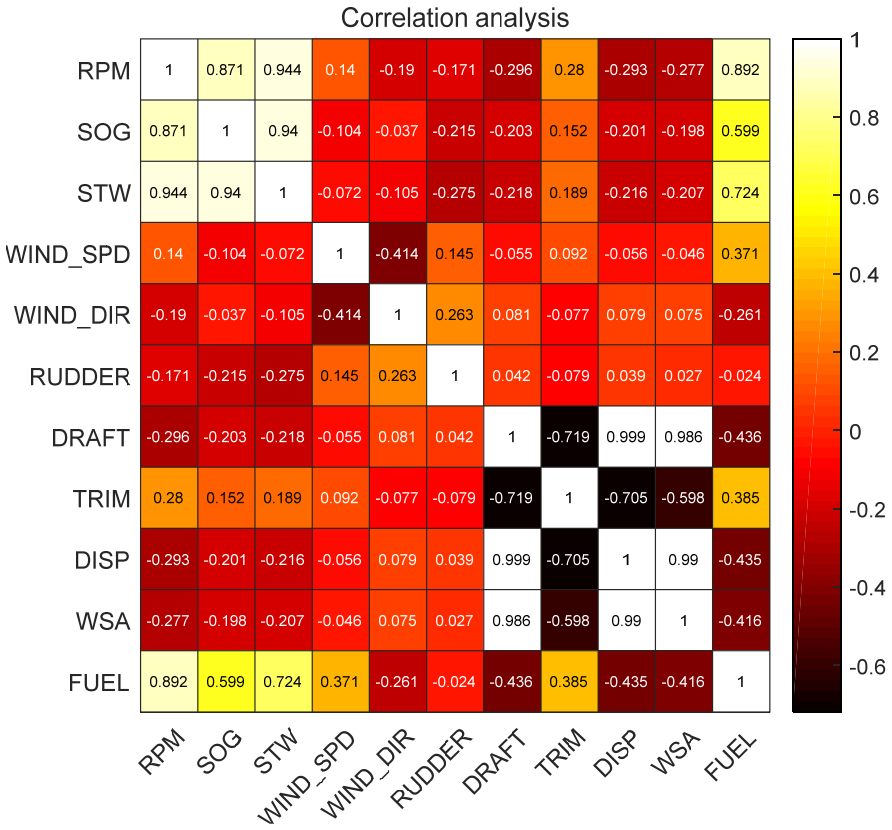

Fig. 3.10 Correlation analysis of operational variables

Fig. 3.10으로부터 주기관 분당 회전수와 대지속력, 주기관 분당 회전수와 대수속력, 대지속력과 대수속력, 평균 흘수와 배수량, 평균 흘수와 침수표면적, 배수량과 침수표면적 등 많은 변수들이 상호간에 높은 상관관계를 가지는 것을 알 수 있다. 이처럼 변수 간에 높은 상관계수를 가질 경우 한 독립변수를 다른 독립변수의 선형 함수로 표현할 수 있으며 이는 추정한 회귀계수의 신뢰성을 떨어뜨리기 때문에 높은 상관계수를 갖는 변수를 적절히 처리할 필요가 있다. 하지만 이러한 피어슨 상관계수는 두 변수간의 상관성만 파악할 수 있으며 한 변수가 2개 이상의 다른 독립변수의 선형 결합으로 표현되어 발생하는 다중공선성은 확인할 수가 없다. 이러한 다중공선성은 독립변수의 회귀계수 및 유의확률로부터는 일반적으로 확인할 수 없으며 다중공선성 진단을 수행하지 않고 모든 변수를 활용하여 예측모델을 구축하는 경우 모델의 회귀계수가 불안정해

져 각각의 독립변수가 종속변수에 미치는 영향력에 대한 해석을 어렵게 한다. 따라서 각 예측변수들의 분산팽창지수(Variance Inflation Factor;VIF) 값을 도출하여 다중공선성 문제를 진단하고자 하였으며 통계 소프트웨어인 SPSS를 이용하여 분석하였다. 각 변수들의 상관계수 및 분산팽창지수를 고려하여 가장 영향력이 작은 변수부터 차례대로 제거하는 방법을 활용하였다.

식 (3.7)은 k번째 회귀계수의 추정량 $\hat{\beta}_k$에 대한 분산팽창지수를 나타낸 것이다. 통상적으로 통계학에서는 분산팽창지수 값이 10 이상인 경우 다중공선성의 문제가 있는 것으로 간주한다 (Neter, 1996).

$$VIF_k = \frac{1}{(1 - R_k^2)} \tag{3.7}$$

여기에서 R_k^2은 x_k를 반응변수로 하고 나머지를 독립변수로 하는 회귀 모형의 결정계수를 나타낸다.

Table 3.3은 회귀 분석을 수행하여 10개의 독립변수들에 대한 회귀계수, 표준오차, 유의확률 및 분산팽창지수 값을 나타낸 것이다. 회귀 분석과 관련된 내용은 부록 A.2에 제시하였다. 각 변수들의 유의확률을 보면 유의수준 0.01에서 평균흘수를 제외한 모든 변수가 유의함을 알 수 있다. 또한 주기관 분당 회전수, 대지속력, 대수속력, 평균흘수, 트림, 배수량, 침수표면적의 분산팽창지수 값이 커서 예측변수들 사이에 다중공선성의 가능성이 있을 것으로 판단된다.

Table 3.3 VIF values of independent variables

Variable Name	Regression coefficients	Standard error	p-value	Correlation coefficients	VIF
Constant	0.000	0.000	0.993		
M/E RPM	1.575	0.002	0.000	0.892	18.9
Speed of the ground	-0.504	0.001	0.000	0.599	9.2
Speed through water	-0.298	0.003	0.000	0.724	29.6
Relative wind speed	0.072	0.001	0.000	0.371	1.9
Relative wind direction	0.016	0.001	0.000	-0.261	1.4
Rudder angle	0.044	0.001	0.000	-0.024	1.4
Mean draft	0.020	0.012	0.091	-0.436	598.2
Trim	0.058	0.004	0.000	0.385	72.9
Displacement	0.376	0.025	0.000	-0.435	2499.0
Wetted Surface Area	-0.498	0.020	0.000	-0.416	1718.8

Table 3.4는 Table 3.3에서 분산팽창지수가 10이상인 변수 중에서 연료소모율과 가장 상관계수가 낮은 변수를 차례로 제거하는 과정을 나타낸 것이다. 트림, 침수표면적, 배수량, 대지속력, 대수속력 순서로 변수를 제거하여 회귀 모형을 생성하였다. 세 번째 단계에서 대지속력의 분산팽창지수가 9.2로 10미만이지만 다중공선성의 가능성이 높을 것으로 판단하여 임의로 제거하였다. 마지막 단계에서는 주기관의 분당 회전수, 상대풍속, 상대풍향, 타각, 평균흘수의 총 5개 변수가 남았으며 각 변수들의 분산팽창지수를 확인하면 모두 10이하로 다중공선성이 발생하지 않을 것으로 사료된다. Fig. 3.11은 최종적으로 남은 변수들의 상관 분석을 수행한 결과이다. 모든 독립변수 간의 상관계수가 0.5 이하로써 최초의 상관 분석 결과와는 상이한 것을 알 수 있다.

Table 3.4 The process of variable selection by multicollinearity test

Variable Name	Regression coefficients	Standard error	p-value	Correlation coefficients	VIF
Constant	0.000	0.000	0.992		
M/E RPM	1.576	0.002	0.000	0.892	18.9
Speed of the ground	-0.504	0.001	0.000	0.599	9.2
Speed through water	-0.298	0.003	0.000	0.724	29.6
Relative wind speed	0.072	0.001	0.000	0.371	1.9
Relative wind direction	0.016	0.001	0.000	-0.261	1.4
Rudder angle	0.044	0.001	0.000	-0.024	1.4
Mean draft	-0.013	0.012	0.280	-0.436	574.2
Displacement	0.096	0.014	0.000	-0.435	785.2
Wetted Surface Area	-0.223	0.004	0.000	-0.416	61.0

Variable Name	Regression coefficients	Standard error	p-value	Correlation coefficients	VIF
Constant	0.000	0.000	0.991		
M/E RPM	1.555	0.002	0.000	0.892	18.4
Speed of the ground	-0.499	0.002	0.000	0.599	9.1
Speed through water	-0.284	0.003	0.000	0.724	29.4
Relative wind speed	0.073	0.001	0.000	0.371	1.9
Relative wind direction	0.014	0.001	0.000	-0.261	1.4
Rudder angle	0.047	0.001	0.000	-0.024	1.4
Mean draft	0.265	0.011	0.000	-0.436	479.2
Displacement	-0.404	0.011	0.000	-0.435	477.9

Variable Name	Regression coefficients	Standard error	p-value	Correlation coefficients	VIF
Constant	0.000	0.001	0.991		
M/E RPM	1.551	0.002	0.000	0.892	18.4
Speed of the ground	-0.498	0.002	0.000	0.599	9.1
Speed through water	-0.282	0.003	0.000	0.724	29.4
Relative wind speed	0.074	0.001	0.000	0.371	1.9
Relative wind direction	0.014	0.001	0.000	-0.261	1.4
Rudder angle	0.048	0.001	0.000	-0.024	1.4
Mean draft	-0.139	0.001	0.000	-0.436	1.2

Variable Name	Regression coefficients	Standard error	p-value	Correlation coefficients	VIF
Constant	0.000	0.001	0.993		
M/E RPM	1.644	0.003	0.000	0.892	18.0
Speed through water	-0.845	0.003	0.000	0.724	17.7
Relative wind speed	0.068	0.001	0.000	0.371	1.9
Relative wind direction	-0.005	0.001	0.000	-0.261	1.4
Rudder angle	0.022	0.001	0.000	-0.024	1.3
Mean draft	-0.131	0.001	0.000	-0.436	1.2

Variable Name	Regression coefficients	Standard error	p-value	Correlation coefficients	VIF
Constant	0.000	0.001	0.998		
M/E RPM	0.818	0.001	0.000	0.892	1.2
Relative wind speed	0.223	0.001	0.000	0.371	1.3
Relative wind direction	-0.024	0.001	0.000	-0.261	1.4
Rudder angle	0.098	0.001	0.000	-0.024	1.2
Mean draft	-0.184	0.001	0.000	-0.436	1.1

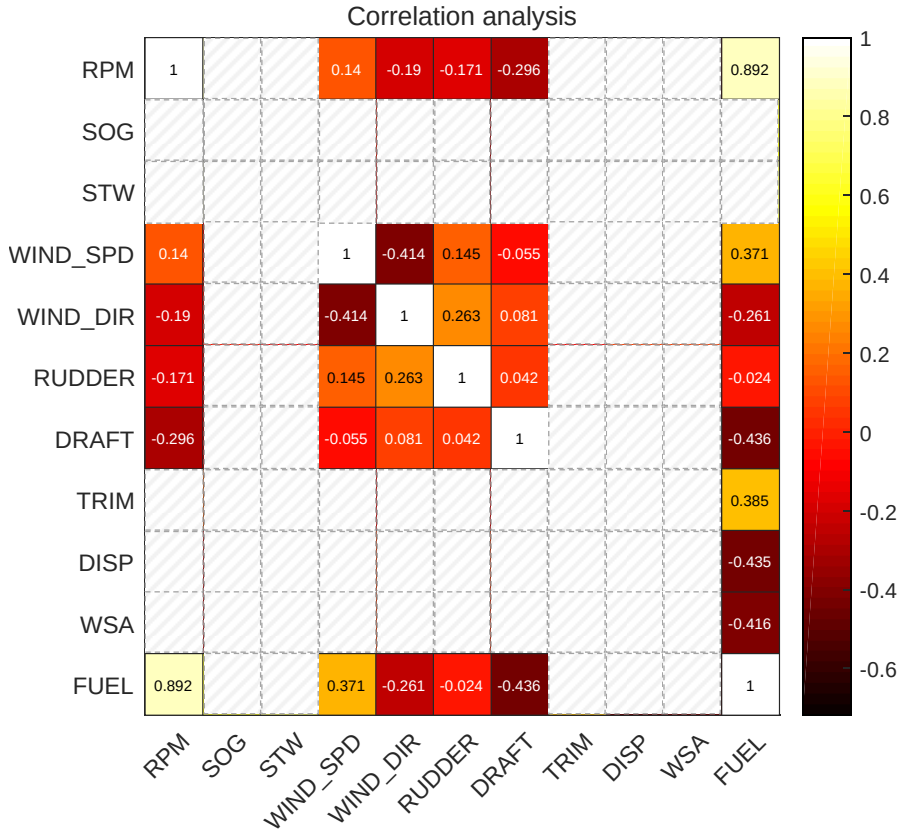

Fig. 3.11 Correlation analysis of variables selected by multicollinearity test

3.5.2 주성분 분석에 의한 특징 추출

앞 절의 상관계수 및 분산팽창지수로부터 운항 변수들 간의 다중공선성의 가능성을 확인하였다. 본 절에서는 주성분 분석을 수행하여 데이터를 적절히 설명할 수 있으면서도 다중공선성을 피할 수 있는 주성분을 도출하고자 하였다. 주성분 분석을 통한 주성분의 추출은 데이터의 차원을 축소할 수 있어 모델의 복잡성을 피하고 계산의 효율성을 향상시킬 수 있는 이점이 있었다. 주성분 분석의 이론적 배경과 관련한 내용은 부록 A.3에 제시하였다.

주성분 분석을 수행하여 예측모델의 입력데이터에 대한 주성분을 추출하기에 앞서, 각 변수들 간의 물리적인 관계를 파악하고 데이터 처리에 이해를 돕기 위하여 종속변수를 포함한 모든 데이터세트에 대한 주성분 분석을 수행하였다.

주성분 분석을 통해서 얻은 각 변수들의 주성분 점수의 크기와 부호로부터 변수들의 상호 연관성을 파악할 수 있다. 또한 주성분 분석으로부터 파악한 변수들의 관계는 센서의 오류 구간 탐지에도 충분히 활용이 가능하다. 각 변수들의 시계열 데이터를 분석하여 주성분 분석으로부터 파악한 변수들 간의 유의미한 경향을 벗어나는 구간은 수집한 데이터의 이상을 의심해 볼 필요가 있다 (Perera & Mo, 2016).

종속변수인 연료소모율을 포함한 전체 운항변수를 대상으로 주성분 분석을 실시하였으며 그 결과는 Table 3.5와 같다. Fig. 3.12는 주성분의 개수에 따른 고유값(eigen value)과 누적 분산값(cumulative variance)을 나타낸 것이다.

Table 3.5 Eigen values and cumulative variances of principal components (all variables)

	Eigen value	Proportion	Cumulative
PC1	4.691	0.426	0.426
PC2	2.499	0.227	0.654
PC3	1.503	0.137	0.790
PC4	1.135	0.103	0.893
PC5	0.504	0.046	0.939
PC6	0.412	0.037	0.977
PC7	0.218	0.020	0.996
PC8	0.030	0.003	0.999
PC9	0.008	0.001	1.000
PC10	0.001	0.000	1.000
PC11	0.000	0.000	1.000

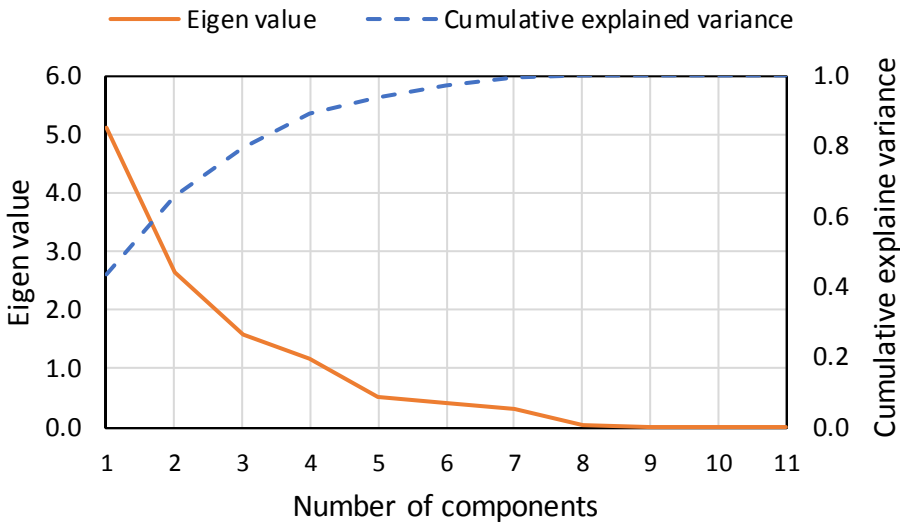

Fig. 3.12 Eigen values and cumulative variances corresponding to the number of components (all variables)

Table 3.5와 Fig. 3.12를 보면 추출한 고유치는 각각 4.691, 2.499, 1.503, 1.135로 4개의 주성분까지는 1이상의 값을 보여주고 있으며 5번째 주성분부터 그 값이 확연히 떨어지는 것을 알 수 있다. 또한 누적 분산 비율도 전체에 대하여 89.3%를 설명할 수 있으므로 제 4성분까지의 변수들에 대한 해석을 수행하였다.

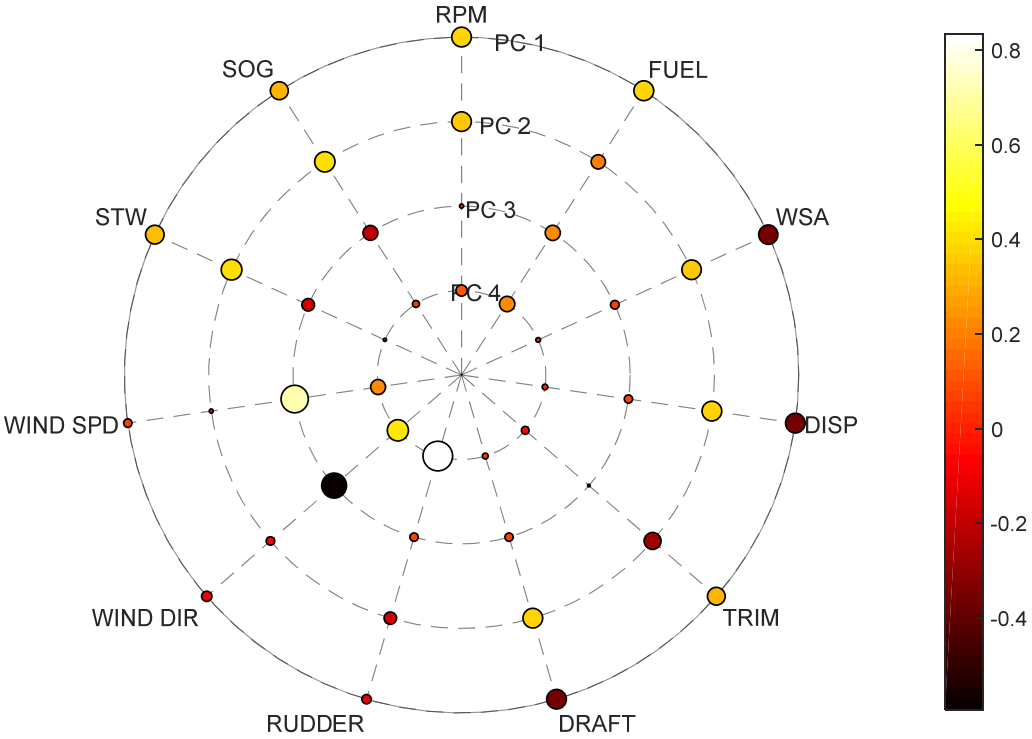

Fig. 3.13 Score plot of each variable according to the principal component

Fig. 3.13은 각 변수들에 대한 주성분 점수를 주성분 별로 나타낸 것이다. 가장 바깥쪽 원부터 제1 주성분의 각 변수에 대한 주성분점수를 나타내었으며 가장 안쪽에 있는 원이 제 4주성분이다. 해당하는 운항변수의 원이 클수록 주성분점수의 절대값이 큰 것을 의미하며 주성분점수가 (+)인 경우 주성분 점수를 나타내는 원의 명암이 밝아지며 (-)인 경우 어두워진다.

주성분 분석으로부터 파악한 변수 간의 관계는 다음과 같다.

1) 제 1성분 : 선박 주기관의 분당 회전수가 증가하면 선박의 대지속력과 대수속력이 증가하며 단위 항해거리 당 연료소모량이 증가한다. 즉, 선속이 높을수록 연료효율이 감소하는 경향이 있다.

2) 제 2성분 : 선박의 평균흘수가 증가하면 배수량 및 선체 수선면 하부의 표면적이 증가하고 연료효율은 감소한다. 이는 흘수 증가로 인하여 선박의 저항이 증가하기 때문에 단위 항해거리 당 연료소모량을 증가시키는 역할을 하는

것으로 판단된다. 또한 평균흘수가 증가하면 트림이 감소(선수트림)하는 경향이 있으며 이는 흘수 증가분을 조정하기 위한 트림의 조정으로 보인다.

3) 제 3성분 : 상대풍향이 작아지면 상대풍속이 증가하고 연료효율도 감소한다. 상대풍향과 상대풍속은 바람의 진풍속과 선박의 속도 벡터의 합성으로 계산되기 때문에 상대풍향이 감소하는 경우는 선박이 강한 선수풍을 조우하거나 선속이 빠른 경우에 해당한다.

4) 제 4성분 : 타각이 커지면 연료효율이 감소한다. 선박이 운항시 타각을 크게 사용하게 되면 타판에 작용하는 항력에 의해서 선속이 떨어지는 효과가 있으며 이는 단위 거리 당 연료사용량을 증가시킨다.

이러한 변수간의 관계는 각 변수들을 시계열로 나타낸 그래프와 상호 비교함으로써 확인이 가능하며 파악한 변수 간의 관계로부터 크게 벗어나는 데이터의 경우 센서의 오류로 식별할 수 있다.

다음으로 주성분 분석을 수행하여 연료소모율 예측모델을 생성하기 위한 주성분을 추출하고자 하였다. 앞서 실시한 주성분 분석과 다르게 종속변수인 연료소모율 변수는 제외하였으며 나머지 예측모델의 입력변수가 되는 운항변수들에 대해서만 분석을 수행하였다. 주성분 분석의 결과는 Table 3.6과 같으며 주성분의 개수에 따른 고유값과 누적 분산값을 나타낸 것이 Fig. 3.14이다.

Table 3.6 Eigen values and cumulative variances of principal components (independent variables)

	Eigen value	Proportion	Cumulative
PC1	3.315	0.395	0.395
PC2	2.617	0.312	0.706
PC3	1.038	0.124	0.830
PC4	0.621	0.074	0.904
PC5	0.403	0.048	0.952
PC6	0.286	0.034	0.986
PC7	0.095	0.011	0.997
PC8	0.020	0.002	1.000
PC9	0.001	0.000	1.000
PC10	0.000	0.000	1.000

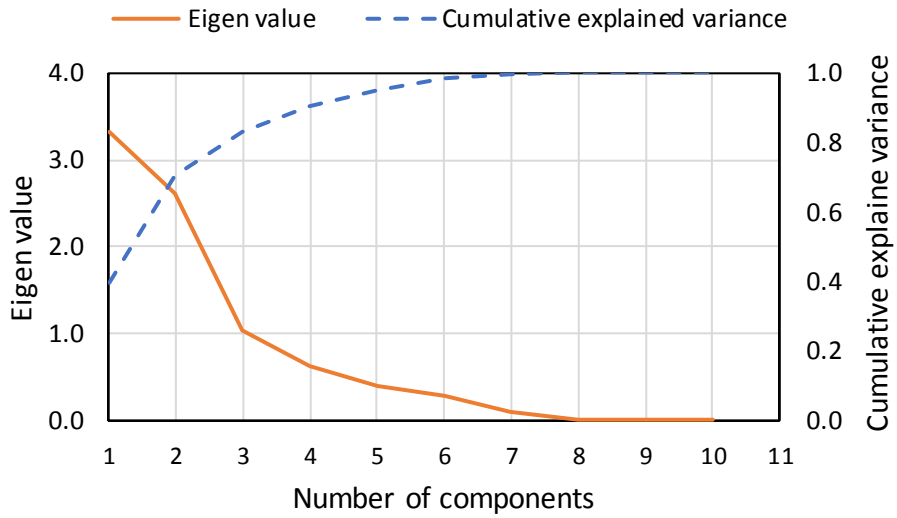

Fig. 3.14 Eigen values and cumulative explained variances corresponding to the number of components (independent variables)

고유값이 작은 주성분은 앞에서 언급한 바와 같이, 각각의 변수에 대한 유용한 정보를 나타내지 못할 수 있으며 선박의 운항데이터에 대한 이상 정보를 포함할 수 있다. Table 3.6과 Fig. 3.14를 보면 3개의 주성분까지는 각각 3.315, 2.617, 1.038로 1이상의 값을 가지고 있으며, 제 4주성분부터는 0.621로 감소하는 폭이 다소 줄어드는 것을 알 수 있다. 또한 누적 분산 비율을 고려하면 제 3주성분까지는 전체의 약 83.0%, 제 4주성분까지 포함하면 약 90.4%를 설명할 수 있다. 제 3주성분의 고유값이 1 이상이긴 하나 누적분산비율이 0.830으로 설명력이 다소 부족할 것으로 판단하였다. 따라서 본 연구에서는 Fig. 3.14의 주성분 개수에 따른 고유값과 누적 분산 비율을 고려하여 제 4주성분까지를 예측모델의 입력변수로 추출함으로써 운항데이터에 대한 설명력을 충분히 반영하고자 하였다. Table 3.7은 각 주성분을 이루는 운항변수들의 주성분점수를 나타낸 것이다.

Table 3.7 Principal component scores of each variable

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
RPM	0.453	0.323	0.126	0.141	0.091	-0.051	-0.566	-0.568	-0.005	-0.003
SOG	0.436	0.379	-0.111	0.033	-0.073	-0.014	0.770	-0.232	0.001	-0.001
STW	0.440	0.386	-0.041	0.027	0.005	-0.017	-0.211	0.781	0.005	0.003
WIND SPD	-0.007	-0.019	0.836	0.406	-0.026	0.324	0.147	0.090	-0.003	0.000
WIND DIR	0.022	-0.074	-0.486	0.454	0.185	0.718	-0.040	-0.014	-0.002	0.001
RUDDER	-0.048	-0.127	-0.173	0.771	-0.186	-0.566	0.018	0.047	-0.002	0.000
MEAN DRAFT	-0.352	0.417	-0.012	0.058	0.056	-0.004	-0.011	-0.019	0.832	-0.064
TRIM	0.144	-0.215	0.068	0.035	0.909	-0.228	0.130	0.051	0.117	0.124
DISP	-0.354	0.420	-0.012	0.056	0.087	-0.014	-0.005	-0.014	-0.312	0.768
WSA	-0.372	0.431	0.000	0.070	0.282	-0.061	0.022	0.002	-0.444	-0.625

주성분 점수로부터 각 주성분을 구성하는 주요 요인들을 분석해보면 제 1주성분은 주기관 분당 회전수, 대지속력, 대수속력이 0.453, 0.436, 0.440로 강한 양의 상관관계를 가지고 있어 선박의 속도에 영향을 미치는 추진성분을 의미하는 것으로 볼 수 있다. 제 2주성분은 평균흘수, 트림, 배수량 및 침수표면적이 0.417, -0.215, 0.420, 0.431의 상관관계를 보여 선박의 하중과 관련된 인자로 판단하였으며 제 3주성분은 상대풍속과 상대풍향이 0.836, -0.486으로 선박에 작용하는 외력 성분을 의미함을 알 수 있다. 마지막으로 제4주성분은 타각성분을 의미한다.

Fig. 3.15는 Table 3.7에서 추출한 각 주성분을 구성하는 주요 인자들을 도식화하여 나타낸 것이다. Table 3.8은 연료소모율 예측모델에 대한 각 주성분의 회귀계수를 나타낸 것이다.

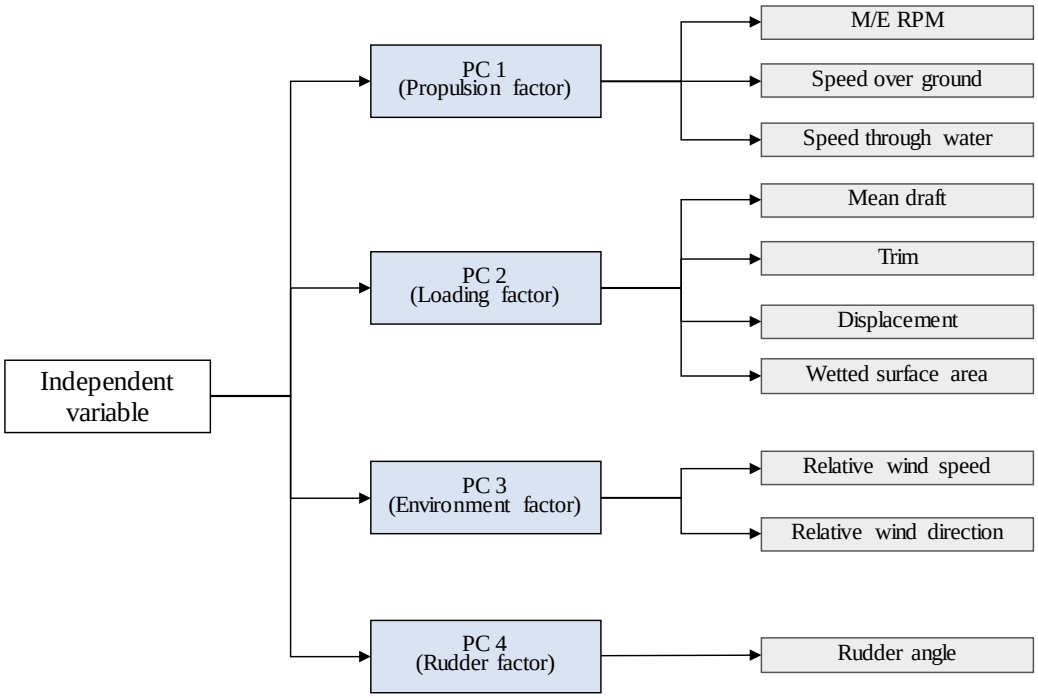

Fig. 3.15 Meaning of each principal component extracted by principal component analysis

Table 3.8 Regression coefficients of independent variables determined by PCA

Variable name	Coefficients
Constant	0.2058
PC 1	0.0160
PC 2	0.0107
PC 3	0.0119
PC 4	0.0187

3.5.3 라소 정규화에 의한 변수 선택

본 절에서는 라소 정규화를 수행하여 다중공선성에 의한 회귀계수의 과도한 추정을 방지하고 국부적인 과적합 문제를 완화시키고자 하였다. 이와 관련한 라소 정규화의 이론은 부록 A.4에 추가로 서술하였다. Fig. 3.16은 조절 모수(tuning parameter) 값에 따른 각 독립변수들의 회귀계수를 나타낸 것이며 각각의 선들은 독립변수를 의미한다. 조절모수 값이 0에 가까워질수록 회귀계수가 최소제곱법 결과에 가까워지고 값이 증가할수록 종속변수에 미치는 영향력이 작은 변수부터 회귀계수가 0으로 축소됨을 알 수 있다.

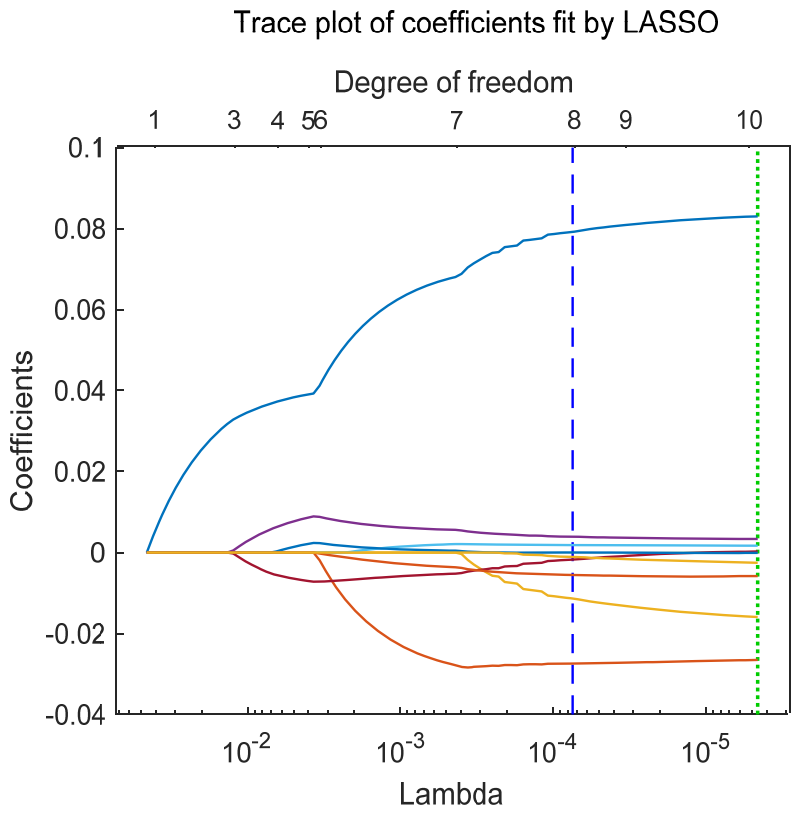

Fig. 3.16 Regression coefficients of independent variables according to the tuning parameter

조절모수 값을 정하기 위하여 Fig. 3.17과 같이 10-fold 교차 검증(cross validation)를 수행하고 평균제곱오차(Mean Squared Error;MSE)를 최소로 해주는 값을 찾았다. Fig. 3.17의 x축은 조절모수의 값이며 y축은 평균제곱오차 값을 나타낸다. 점선은 평균제곱오차가 최소가 되는 지점에서의 조절 모수를 나타낸 것이며 해당하는 지점으로부터 1 표준오차만큼 떨어진 조절모수를 파선으로 표시하였다. Breiman et al. (1984)의 1 표준오차 법칙에 의하면 평균제곱오차가 가장 작은 모델보다 1 표준오차 이내의 범위에서 가장 정규화 된 모델을 선택하는 방법으로써 회귀의 교차 검증에 통상적으로 많이 활용되어진다. 본 연구에서는 1 표준오차 법칙을 적용하여 λ를 0.000074로 선정하였다. 그림 상단에 표시된 숫자는 독립변수의 개수이며 이를 참조하면 평균제곱오차가 최소가 되는 지점에서는 분석에 사용된 모든 독립변수들의 회귀계수가 0보다 커서 그대

로 남아있으며 조절 모수가 증가하여 최소평균제곱오차로부터 1 표준오차 지점에서는 8개의 변수가 선택된 것을 확인할 수 있다.

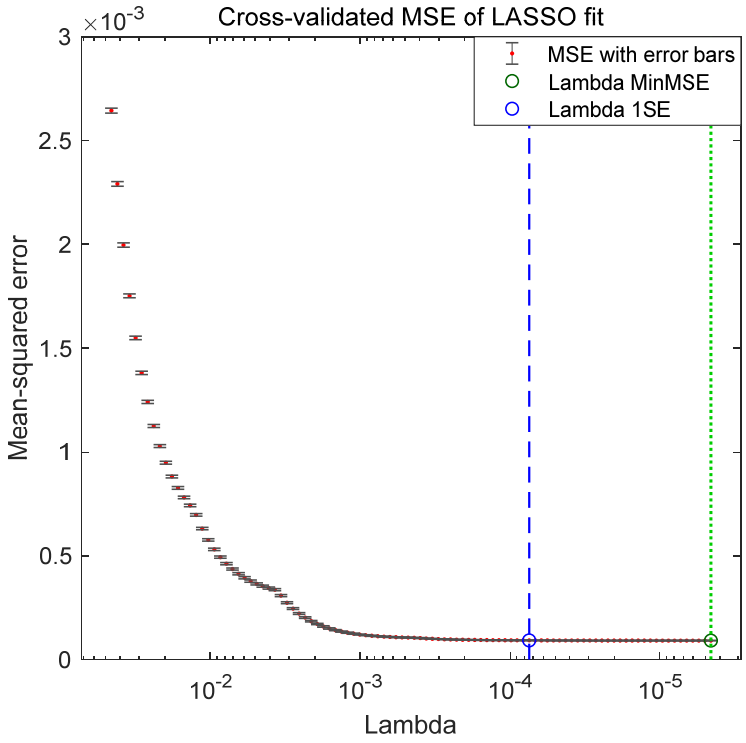

Fig. 3.17 Mean squared errorcross according to the tuning parameter (10-fold validation)

Table 3.9는 최소평균제곱오차로부터 1 표준오차 위치에서 각 독립변수들의 회귀계수를 나타낸 것이다. 각 변수들의 회귀계수를 비교해보면 주기관의 분당 회전수가 0.0792로 연료소모율에 가장 큰 영향을 미치며 그 뒤를 이어서 대지속력, 대수속력, 배수량, 상대풍속, 흘수, 타각, 침수표면적 순서로 회귀계수를 가졌다. 상대풍향과 트림의 경우 회귀계수가 0으로 축소되어 변수 선택에서 제외되었다. 이는 연료소모율에 미치는 설명력이 부족하여 제외된 것으로 판단되며 변수 간에 서로 상관성을 가지는 변수들은 연료소모율에 미치는 효과가 중복되기 때문에 상대적으로 계수가 감소한 것을 알 수 있다.

Table 3.9 Regression coefficients of independent variables determined by LASSO

Variable name	Coefficients
Constant	0.2055
M/E RPM	0.0792
Speed of the ground	-0.0274
Speed through water	-0.0114
Relative wind speed	0.0039
Relative wind direction	0.0000
Rudder angle	0.0018
Mean draft	-0.0018
Trim	0.0000
Displacement	-0.0056
Wetted Surface Area	-0.0011

3.5.4 경험적인 판단에 의한 변수 선택

선박의 항로 계획은 통상적으로 입출항 일정, 운항 해역의 선박통항량, 해상 및 기상 상태 등을 고려하여 수립되며 이는 주로 해운선사의 운항담당자 또는 선박 운항자의 지식과 경험에 근거하여 수행되어진다.

실제 선박에서 연료소모율 예측모델을 활용하여 항로계획을 수행하는 경우 선박 운항자에 의해서 조정가능한 운항변수 또는 운항변수를 조정함으로써 영향을 받을 수 있는 환경변수가 예측모델의 대표적인 입력변수로 선정될 필요가 있으며 이러한 입력값들은 사용자가 계획하는 값을 직접 입력하거나 선내 시스템으로부터 자동으로 입력 가능하여야 할 것이다. 따라서 본 절에서는 항해 계획시 에너지효율 예측 시스템을 활용한다고 가정하여 Table 3.2와 같은 운항변수를 선정하였다. 실제 항해계획시에는 운항자가 대수속력이나 주기관의 분당

회전수와 같은 변수를 미리 예측하여 입력하기는 어렵기 때문에 계획하고자 하는 선속인 대지속력을 사용하였으며, 출항 기준의 선박 흘수와 트림, 운항시 예상되는 해상상태(시스템 자동 입력)를 임의의 입력변수로 사용하였습니다. 또한 본 연구에서 수집한 선박 데이터에는 운항 당시의 기상 및 해상상태에 관한 정보가 선박의 풍속계로부터 측정한 풍향 및 풍속 데이터가 유일하였다. 해류 및 조류와 같은 파도에 의한 외력 성분을 고려하기 위하여 대지속력과 대수속력의 차를 외력성분으로 가정하여 반영하였다. 향후 대상 선박으로부터 운항해역의 해상 및 기상정보 등 다양한 변수의 데이터가 수신된다면 본 절에서 선정한 예측모델의 입력변수에 추가되어 더욱 정확한 해석이 가능할 것이라 판단된다.

본 절에서 선정한 운항변수에 대하여 상관 분석과 다중공선성 진단을 수행하였다. Fig. 3.18은 해당하는 변수들의 상관 계수를 나타낸 것이며 Table 3.10은 선형 회귀 모형에서 각 변수들의 회귀계수, 표준오차, 유의확률 및 분산팽창지수 값을 나타낸다. 독립변수 간의 상관계수가 높지 않으며 모든 변수의 분산팽창지수 값이 10이하이기 때문에 다중공선성 문제는 발생하지 않을 것으로 판단하여 해당하는 변수를 예측모델의 학습에 활용하였다.

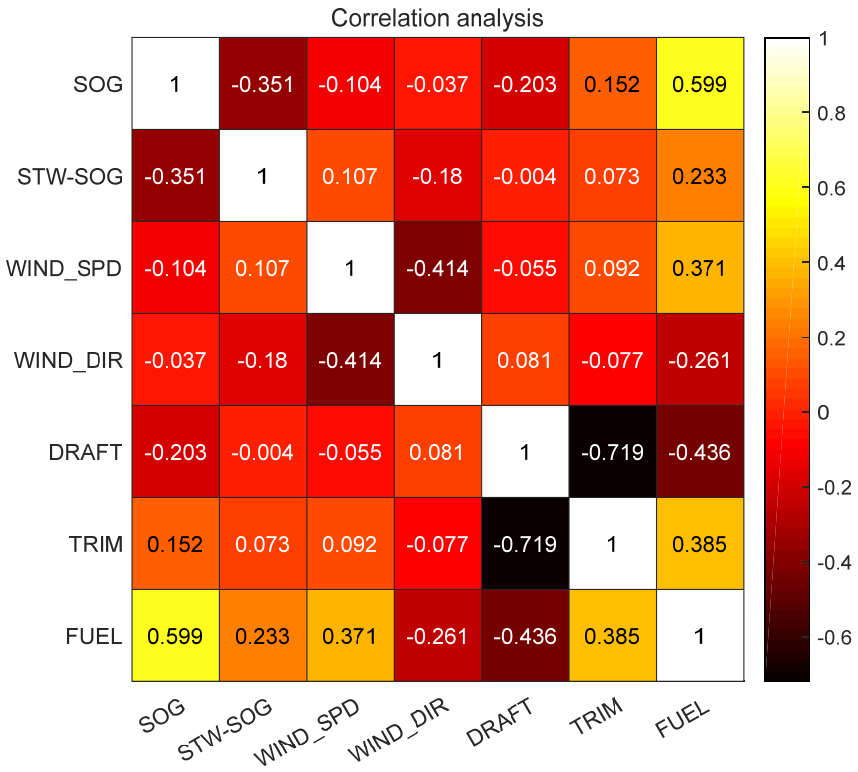

Fig. 3.18 Correlation analysis of independent variables selected by empirical method

Table 3.10 VIF values of independent variables selected by empirical method

Variable Name	Regression coefficients	Standard error	p-value	Correlation coefficients	VIF
Constant	0.205	0.000	0.000		
Speed of the ground	0.038	0.000	0.000	0.599	1.2
STW-SOG	0.023	0.000	0.000	0.233	1.2
Relative wind speed	0.021	0.000	0.000	0.371	1.2
Relative wind direction	0.002	0.000	0.000	-0.261	1.3
Mean draft	-0.012	0.000	0.000	-0.436	2.1
Trim	0.002	0.000	0.000	0.385	2.1

제 4 장 예측모델 개발 및 평가

4.1 예측모델 개발

4.1.1 데이터 구분

본 연구에서는 연료소모율 예측모델의 학습을 위하여 전체 운항데이터세트를 학습용 데이터(training data set)와 평가용 데이터(test data set)로 구분하였다. 전체 데이터 중 70%를 학습용 데이터로 분할하여 예측모델을 생성하는데 사용하였으며 학습에 사용되지 않은 나머지 30%를 학습된 모델의 성능을 평가하기 위한 평가용 데이터로 활용하였다. 선박의 연료소모율 예측모델의 성능을 효과적으로 파악하기 위해서는 전체 데이터를 항차 단위 또는 만재, 공선항해 등으로 구분하여 사용하는 것이 바람직할 것으로 사료된다. 하지만 본 연구의 대상 선박인 컨테이너선박은 화물의 운송 특성상 만재, 공선항해의 구분에 따른 데이터의 분류가 쉽지 않다. 또한 데이터의 전처리 과정을 수행하면서 분석에 불필요한 일부 데이터세트 구간이 제거되어 학습용 데이터와 평가용 데이터의 비율이 정확하게 유지되지 않기 때문에 본 연구에서는 전처리 된 데이터를 7:3의 비율로 나누어 사용하였다. 예측모델의 입력값이 되는 독립변수는 3.4절의 데이터 축소방법을 통해서 선정하였으며 Table 4.1과 같다. 선정된 변수로 구성된 학습용 데이터에 다중선형 회귀와 인공 신경망 기법을 적용하여 예측모델을 학습하였으며 그 구성은 Table 4.2와 같다.

Table 4.1 Definition of variables for prediction models

	Method of selecting independent variables	Independent variables	Dependent variable
1	Correlation analysis & VIF	RPM, Wind speed, Wind direction, Rudder angle, Mean draft	Fuel consumption rate
2	Principal component analysis(PCA)	PC1, PC2, PC3, PC4	
3	Least absolute shrinkage and selection operator(LASSO)	SOG, STW, RPM, Wind speed, Rudder angle, Mean draft, Trim, Wetted surface area	
4	Empirical method	SOG, STW-SOG, Wind speed, Wind direction, Mean draft, Trim	

Table 4.2 Cases for analyzing the performance of prediction models

	Method of selecting independent variables	Learning method
Case 1	Correlation analysis & VIF	Multiple linear regression
Case 2		Artificial neural network
Case 3	PCA	Principal component regression
Case 4		Artificial neural network
Case 5	LASSO	LASSO regression
Case 6		Artificial neural network
Case 7	Empirical method	Multiple linear regression
Case 8		Artificial neural network

4.1.2 평가 기준

예측모델의 성능을 평가하기 위해서 다음과 같은 기준을 적용하였다.

1) 평균 제곱근 오차(Root Mean Square Error;RMSE)

모델의 예측값과 실제 관측값의 차이를 계산할 때 사용되는 측도로서 정밀도를 표현하는데 적합하다. 평균 제곱근 오차는 항상 양의 값을 가지며 값이 0이면 데이터에 완벽하게 적합함을 나타낸다. 잔차가 클수록 평균 제곱근 오차에 불균형적으로 큰 영향을 미치기 때문에 평균 제곱근 오차는 이상값에 민감한 특성을 가진다.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{4.6}$$

여기에서 y_i는 i번째 종속변수의 관측값, $\hat{y}_i$는 i번째 종속변수의 예측값, n은 관측값의 개수이다.

2) 조정 결정계수(adjusted coefficient of determination;adj-R^2)

결정계수는 총제곱합(Total Sum of Squares;SST) 중에서 회귀제곱합(Sum of Squares due to Regression;SSR)이 차지하는 비율로서 회귀 모형의 설명력을 나타낸다. 결정계수의 값은 0에서 1사이에 있으며 1에 가까울수록 회귀 모형의 설명력이 높다고 볼 수 있으며 0에 가까울수록 설명력이 낮다. 이러한 결정계수의 값은 일반적으로 독립변수의 개수에 비례하여 증가하는 경향을 가지기 때문에 식 (4.7)과 같이 자유도와 독립변수의 개수를 반영하여 문제점을 보완한 조정 결정계수를 사용한다.

$$Adjusted\,R^2 = 1 - \frac{SSE \times (n-1)}{SST \times (n-d-1)} \tag{4.7}$$

$$SSE = \sum (y_i - \hat{y_i})^2 \tag{4.8}$$

$$SST = \sum (y_i - \overline{y})^2 \tag{4.9}$$

여기에서 y_i는 i번째 종속변수의 관측값, $\hat{y}_i$는 i번째 종속변수의 예측값, $\overline{y}$는 관측값의 평균, SST는 총제곱합, SSE은 오차제곱합, n은 샘플의 개수이다.

4.1.3 다중선형 회귀 기반의 예측모델 개발

3.4절의 각 데이터 축소방법에 의하여 선정된 변수를 활용하여 다중선형 회귀 기반의 예측모델을 개발하였다.

1) 상관 분석 및 분산팽창지수에 의한 변수 선택

상관 분석 및 분산팽창지수를 고려하여 최종적으로 선택한 변수를 이용하여 다중선형 회귀를 구현하였으며 식 (4.1)과 같다. 선박 주기관의 분당 회전수가 가장 큰 회귀계수를 가지며 외력의 영향인 풍속과 하중성분인 선박의 평균흘수가 다음으로 큰 값을 가진다.

$$\begin{aligned} FUEL\ EFFICIENCY_{corr} = {} & 0.818RPM + 0.223\,WIND\,SPD - 0.024\,WIND\,DIR \\ & + 0.098RUDDER - 0.184DRAFT \end{aligned} \tag{4.1}$$

2) 주성분 분석에 의한 특징 추출

주성분 분석을 통해서 식 (4.2a)-(4.2d)와 같이 운항데이터의 특징을 4개의 주성분으로 추출하였으며 각 주성분을 독립변수로 하는 다중선형 회귀 모형은 식 (4.3)과 같다. 각각의 주성분은 선박의 추진성분, 하중성분, 외력성분, 타각성분

을 나타낸다.

$$PC1 = 0.453RPM + 0.436SOG + 0.440STW - 0.007\,WIND\,SPD + 0.022\,WIND\,DIR - 0.048RUDDER - 0.352DRAFT + 0.144\,TRIM - 0.354DISP - 0.372\,WSA \quad (4.2a)$$

$$PC2 = 0.323RPM + 0.379SOG + 0.386STW - 0.019\,WIND\,SPD - 0.074\,WIND\,DIR - 0.127RUDDER + 0.417DRAFT - 0.215\,TRIM + 0.420DISP + 0.431\,WSA \quad (4.2b)$$

$$PC3 = 0.126RPM - 0.111SOG - 0.041STW + 0.836\,WIND\,SPD - 0.486\,WIND\,DIR - 0.173RUDDER - 0.012DRAFT + 0.068\,TRIM - 0.012DISP \quad (4.2c)$$

$$PC4 = 0.141RPM + 0.033SOG + 0.027STW + 0.406\,WIND\,SPD + 0.454\,WIND\,DIR + 0.771RUDDER + 0.058DRAFT + 0.035\,TRIM + 0.056DISP + 0.070\,WSA \quad (4.2d)$$

$$FUEL\ EFFICIENCY_{pca} = 0.0160PC1 + 0.0107PC2 + 0.0119PC3 + 0.0187PC4 + 0.2058 \quad (4.3)$$

3) 라소 정규화에 의한 변수 선택

라소 정규화로부터 상대풍향과 트림이 제외되었으며 나머지 독립변수들은 종속변수에 미치는 영향력에 따라 회귀계수가 설정되었다. 라소 정규화로부터 추정한 회귀계수를 적용한 연료소모율의 회귀 모형은 식 (4.4)와 같다.

$$FUEL\ EFFICIENCY_{lasso} = 0.0792RPM - 0.0274SOG - 0.0114STW + 0.0039\,WIND\,SPD + 0.0018RUDDER - 0.0018DRAFT - 0.0056DISP - 0.0011\,WSA + 0.2055 \quad (4.4)$$

4) 경험적인 판단에 의한 변수 선택

실제 연료소모율 예측모델의 적용 가능성을 고려하여 선박 운항자에 의해서 조정가능한 운항변수 또는 운항변수를 조정함으로써 영향을 받을 수 있는 환경변수를 예측모델의 입력변수로 선정하였으며 회귀 모형은 식 (4.5)와 같이 나타낼 수 있다.

$$FUEL\ EFFICIENCY_{empirical} = 0.038SOG + 0.023(STW - SOG) + 0.021\,WIND\,SPD + 0.002\,WIND\,DIR - 0.012DRAFT - 0.002\,TRIM + 0.205 \quad (4.5)$$

4.1.4 인공 신경망 기반의 예측모델 개발

3.4절의 각 데이터 축소방법에 의하여 선정된 변수를 인공 신경망에 적용하여 예측모델을 개발하였다. 인공 신경망 이론에 대해서는 부록 B에 별도로 기술하였다.

Table 4.3과 같이 비선형 최소자승문제에서 안정적으로 해를 찾을 수 있는 Levenberg-Marquardt를 학습 알고리즘으로 사용하였으며 활성화 함수는 탄젠트-시그모이드 함수를 적용하였다. 앞서 변수의 선택 및 추출방법으로부터 인공 신경망의 입력층에 사용될 노드를 정의하였다. Table 4.3에 나타낸 바와 같이, 상관 분석은 5개, 주성분 분석은 4개, 라소 정규화는 8개, 경험기반의 변수 선택 방법은 6개의 노드를 입력층에서 사용하였으며 10개의 은닉 층을 적용하여 연료소모율을 예측하고자 하였다. Fig. 4.1은 인공 신경망을 이용한 연료소모율 예측모델의 구조를 도식화하여 나타낸 것이다.

Table 4.3 Main parameters of ANN models

Parameters	Method
Input layer node	Correlation & VIF : 5 PCA : 4 LASSO : 8 Empirical method : 6
Hidden neuron	10
Output layer node	1
Learning algorithm	Levenberg-Marquardt
Activation function	Tan-sigmoid
Performance function	MSE

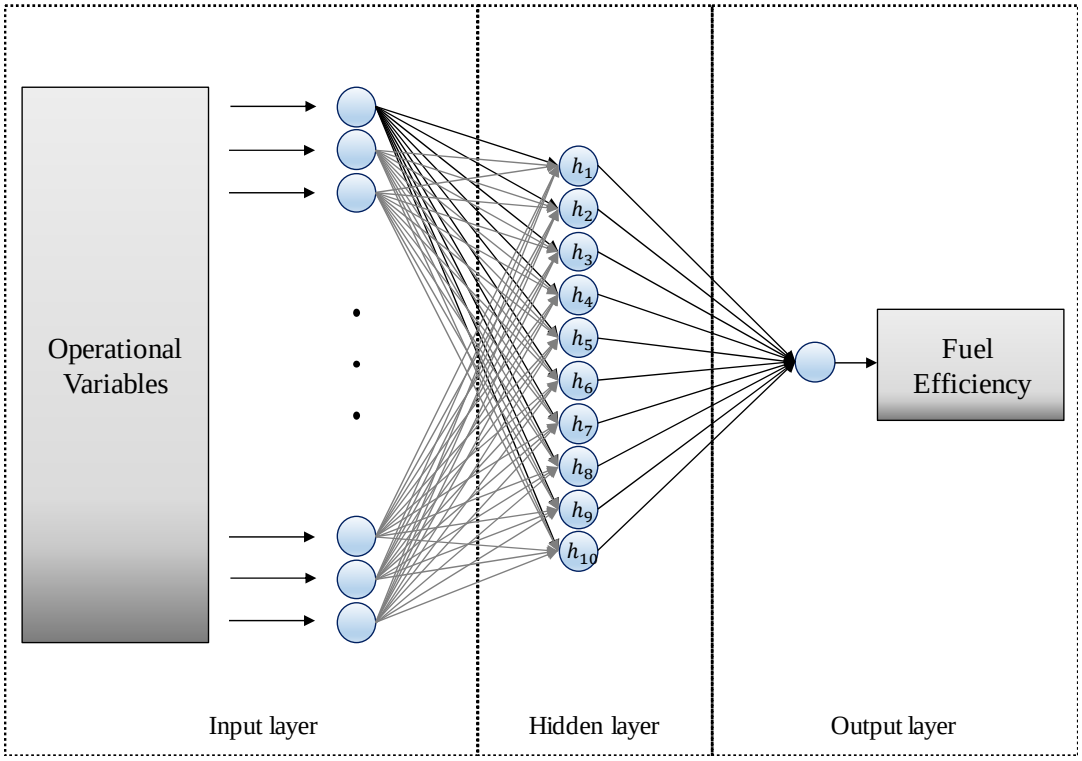

Fig. 4.1 Schematic diagram of ANN structure for prediction model

4.1.3절과 4.1.4절의 다중선형 회귀와 인공 신경망을 적용하여 Table 4.4와 같은 8가지 경우의 예측모델을 개발하였다. 표에 나타낸 평균 제곱근 오차와 조정 결정계수 값은 학습용 데이터에 대한 예측모델의 정확도를 나타내며 평가용 데이터에 대한 예측모델의 성능은 다음의 4.2절에서 다루도록 하겠다.

Table 4.4 Fuel consumption rate prediction models developed by the study

	Method of selecting independent variables	Learning method	Training data	
			RMSE	$adj-R^2$
Case 1	Correlation analysis & VIF	Multiple linear regression	0.0173	0.8867
Case 2		Artificial neural network	0.0154	0.9105
Case 3	PCA	Principal component regression	0.0237	0.7884
Case 4		Artificial neural network	0.0204	0.8426
Case 5	LASSO	LASSO regression	0.0096	0.9651
Case 6		Artificial neural network	0.0067	0.9831
Case 7	Empirical method	Multiple linear regression	0.0217	0.8225
Case 8		Artificial neural network	0.0157	0.9074

4.2 예측모델 평가

4.2.1 평가 결과

Fig. 4.2(a) ~ 4.2(h)는 4.1절에서 개발한 예측모델에 대하여 10일 동안의 연료 소모율을 예측한 결과를 평가용 데이터와 비교하여 나타낸 예시이다.

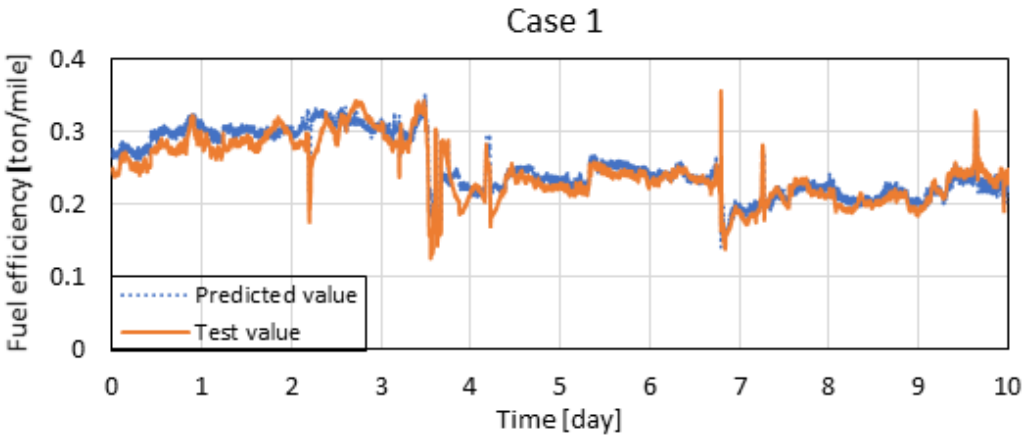

Fig. 4.2(a) Prediction accuracy of case 1 for 10 days of test data

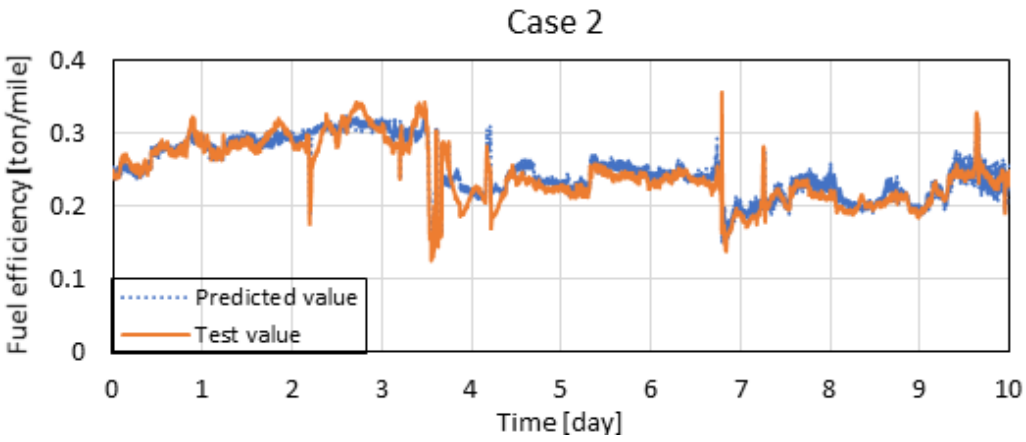

Fig. 4.2(b) Prediction accuracy of case 2 for 10 days of test data

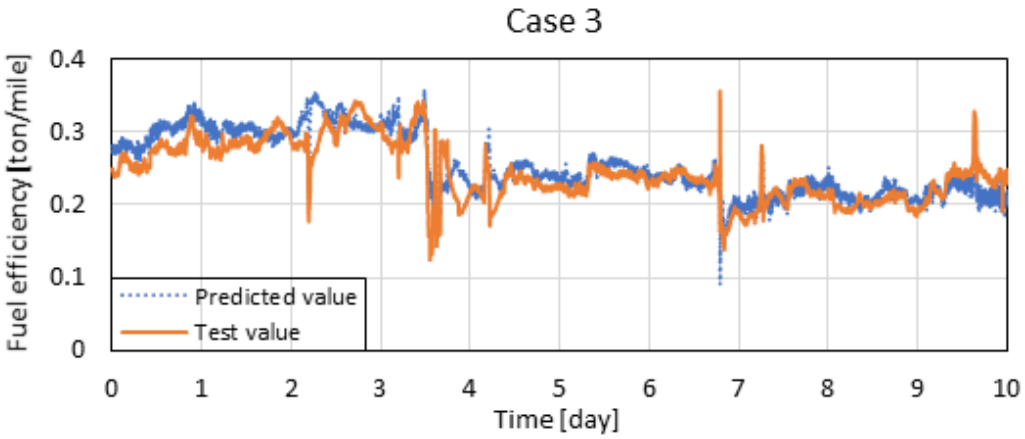

Fig. 4.2(c) Prediction accuracy of case 3 for 10 days of test data

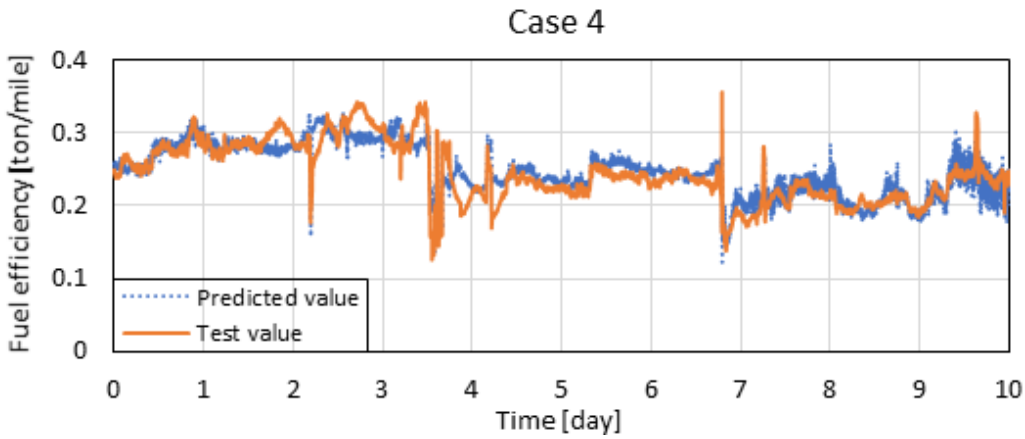

Fig. 4.2(d) Prediction accuracy of case 4 for 10 days of test data

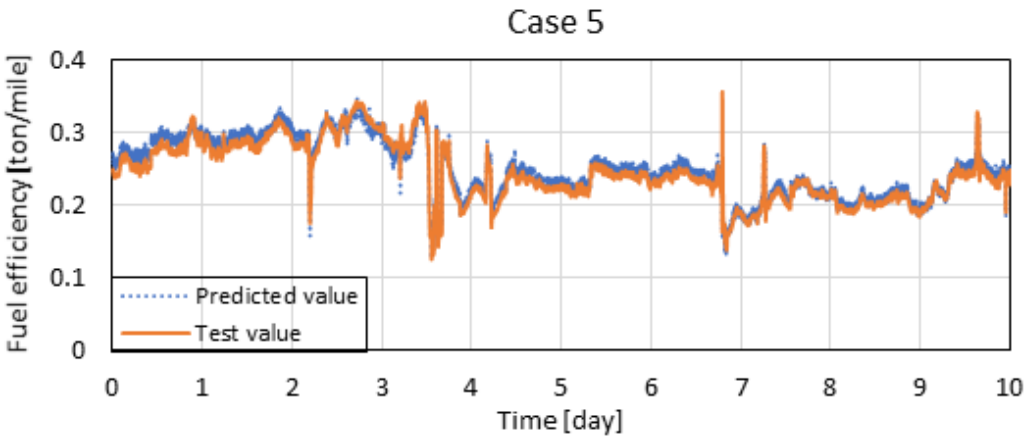

Fig. 4.2(e) Prediction accuracy of case 5 for 10 days of test data

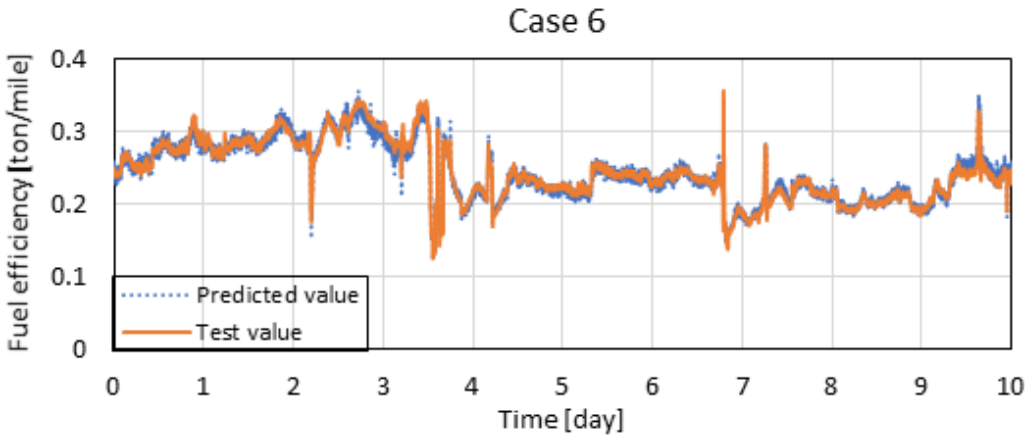

Fig. 4.2(f) Prediction accuracy of case 6 for 10 days of test data

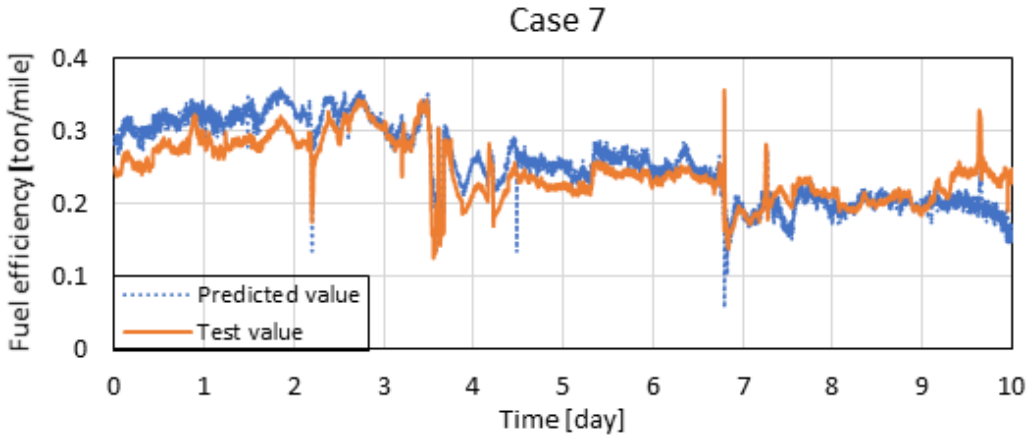

Fig. 4.2(g) Prediction accuracy of case 7 for 10 days of test data

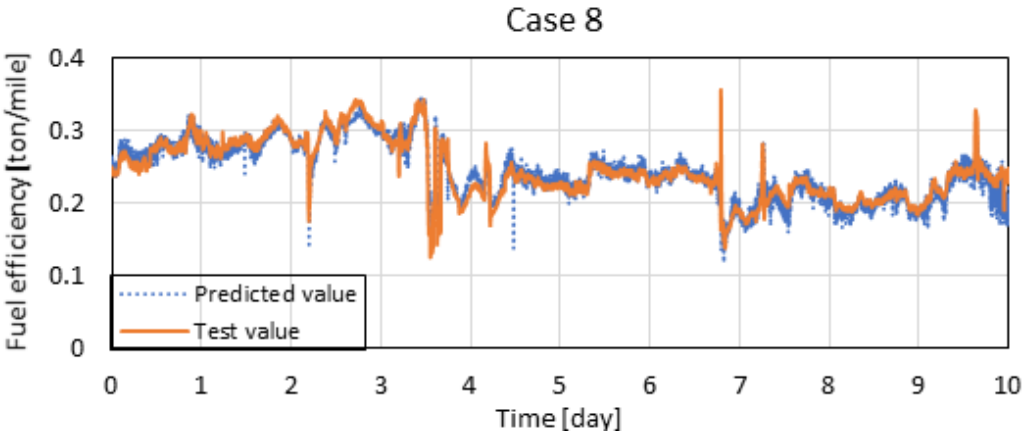

Fig. 4.2(h) Prediction accuracy of case 8 for 10 days of test data

Table 4.5 Performance results of each prediction model for test data

	Method of selecting independent variables	Learning method	Test data	
			RMSE	$adj-R^2$
Case 1	Correlation analysis & VIF	Multiple linear regression	0.0152	0.9119
Case 2		Artificial neural network	0.0119	0.9456
Case 3	PCA	Principal component regression	0.0211	0.8299
Case 4		Artificial neural network	0.0169	0.8901
Case 5	LASSO	LASSO regression	0.0106	0.9574
Case 6		Artificial neural network	0.0074	0.9793
Case 7	Empirical method	Multiple linear regression	0.0273	0.7140
Case 8		Artificial neural network	0.0135	0.9298

Table 4.5는 평가용 데이터에 대한 각 모델의 예측 성능을 나타낸 결과이다.

첫째. Table 4.5의 변수 선택 방법에 따른 모델의 평균 제곱근 오차 및 조정결정계수를 비교해보면 라소 정규화, 상관 분석, 주성분 분석, 경험 기반의 방법 순서로 성능이 우수하였다. 라소 정규화 기반의 예측모델은 불필요한 변수의 회귀계수를 축소시켜 제거하고 변수들의 영향력에 따라 회귀계수를 지정하였기 때문에 평가용 데이터에 대한 예측력이 0.9574, 0.9793로 상당히 높았다. 상관 분석 및 분산팽창지수에 의한 변수 선정방법도 예측력이 0.9119, 0.9456으로 높은 예측력을 보였다. 주성분 분석에 의한 변수 선택 방법이 다른 방법들에 비해 예측력이 다소 떨어지는 이유는 주성분을 추출하는 원리에 의한 것으로 판단된다. 주성분 분석은 종속변수와의 관계를 고려하는 것이 아니라 주어

진 입력변수의 데이터세트를 공통된 특징으로 묶어서 주요한 성분으로 추출해 주는 비지도 학습기반의 방법이다. 따라서 데이터 축소를 수행하는 과정에서 종속변수와 유의미한 데이터가 축소되어 종속변수를 대변하지 못하는 경우가 발생할 수도 있다.

둘째. 데이터의 학습 방법을 비교해보면 인공 신경망을 적용한 예측모델이 다중선형 회귀 모형에 비해 예측 성능이 우수하였으며 이는 인공 신경망 알고리즘이 변수들 간의 비선형적인 관계를 효율적으로 다루기 때문인 것으로 판단된다. 다중선형 회귀 모형의 경우 인공 신경망에 비해서 예측력이 다소 떨어지기는 하나 Fig. 4.2의 시계열 데이터에 대한 예측값을 보면 전반적인 추세를 예측하기에는 무리가 없었다. 또한 회귀계수로부터 각 변수들이 연료소모율에 미치는 영향력을 파악할 수 있어 해당하는 모델을 활용시 사용자의 이해를 도울 수 있는 장점이 있었다. 인공 신경망 기반의 예측모델은 회귀계수와 같이 입력변수들이 종속변수에 미치는 영향을 직관적으로 파악할 수는 없었으나 입력변수들을 고정하고 분석하고자 하는 변수만을 변경하면서 모델의 예측 결과를 비교하는 것과 같은 방법 등으로 생성된 경향성을 간접적으로 파악할 수 있었다. 인공 신경망 모델의 경향 분석과 관련된 내용은 부록 C에 제시하였다.

셋째. 경험에 기반하여 변수를 선정한 Case 7, Case 8의 예측모델을 상관 분석 기반으로 변수를 선정한 Case 1, Case 2와 비교해보면 상관 분석 기반의 예측모델보다 더 많은 변수의 데이터를 사용하였음에도 불구하고, 그 예측력은 0.7140, 0.9298로 상관 분석 기반의 모델을 하회하는 것을 알 수 있다. 이는 경험 기반으로 선택된 변수에 연료효율의 예측력을 높여줄 수 있는 주기관의 분당 회전수가 제외되었기 때문으로 보인다. 효율적이고 높은 정확도를 가지는 예측모델을 고려한다면 적절한 차원 축소 방법을 활용하는 것이 좋을 것으로 사료되나 예측모델의 사용 목적에 따라서는 이러한 임의의 변수 선정방법도 활용될 수 있을 것으로 판단된다.

Table 4.5의 예측모델 중 가장 우수한 예측력을 보여주는 라소 정규화에 의한 변수 선정과 다중선형회귀 및 인공신경망 학습으로 구현한 Case 5와 Case 6의 모델을 비교분석하고자 하였다. Fig. 4.2와 Fig. 4.3은 전체 평가용 데이터에

서 예측 오차가 가장 크게 발생하는 구간의 전후 시계열 데이터를 나타낸 것이다. 전반적으로 모델에 의한 연료소모율 예측 값이 실제 관측값에 잘 부합하는 것을 알 수 있다. 다중선형회귀 모델의 경우 전체 구간에 대한 설명력은 0.9574로 상당히 높으나 최대 오차가 발생하는 구간에서는 오차가 다소 나는 것을 알 수 있다. 반면에 인공신경망 모델은 최대 오차가 발생하는 구간을 포함한 모든 구간에서 오차가 거의 발생하지 않는 것을 확인할 수 있다.

연료소모율 예측모델의 활용 목적에 따라 일정 수준이상의 예측력을 확보하면서 운항변수들의 영향력을 파악하고자 한다면 라소 정규화 기반의 회귀 모형을 활용하는 것이 적합할 것이며 모델의 높은 정확도를 우선으로 한다면 변수간의 비선형 관계를 적절히 고려할 수 있는 인공 신경망을 비롯한 기계학습방법의 예측모델을 적용하는 것이 좋을 것으로 판단된다.

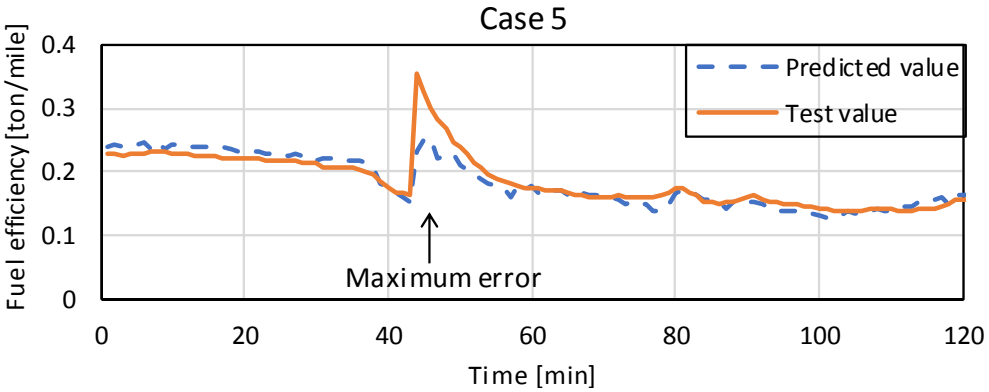

Fig. 4.3(a) Prediction accuracy of case 5 in the maximum error section

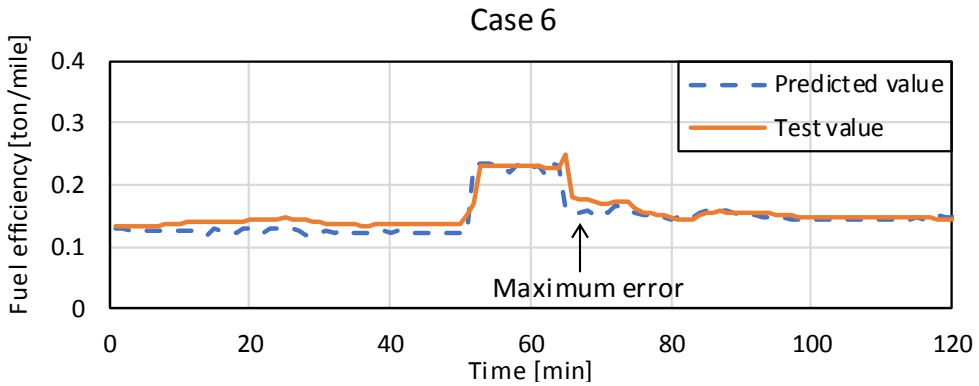

Fig. 4.3(b) Prediction accuracy of case 6 in the maximum error section

제 5 장 결론 및 제언

5.1 결론

본 논문에서는 13k TEU급 컨테이너선박으로부터 수집한 운항데이터를 활용하여 선박의 연료소모율 예측모델을 개발하고자 하였다. 이러한 예측모델은 향후 스마트선박의 운항시스템에서 항로계획시 최적항로선정, 선체 및 기기의 이상상태 탐지, 장기 운항에 따른 성능저하 파악 등에 활용될 수 있을 것으로 판단된다. 또한 운항데이터의 처리와 분석을 위해 본 연구에서 활용된 여러 접근방법들은 예지정비 및 안전 시스템 개발과 같은 선박의 다른 분야에도 확장 가능할 것으로 보인다.

본 논문의 연구 결과를 요약하면 다음과 같다.

1) 선박의 운항 특성상 수집되는 데이터 중 일부는 정상적인 범주를 벗어나는 이상값이나 결측 구간 등이 존재하였고 표준편차를 이용한 이상값 처리 및 중앙값 필터 등 다양한 데이터 전처리 방법을 적용하여 분석하기 적합한 상태로 데이터를 처리하였다.

2) 선행 연구에서는 연료소모량 예측 모델의 입력변수를 연구자의 경험 또는 대상 선박으로부터 수집가능한 모든 변수를 활용하는 경우가 많았다. 하지만 이러한 경우 변수들 간에 존재하는 상관관계로 인하여 다중공선성, 과적합과 같은 문제가 발생할 수 있었다. 따라서 본 논문에서는 상관 분석 및 분산팽창지수, 주성분 분석, 라소 정규화와 같은 차원감소방법을 활용하여 연료효율에 유의미한 변수를 선정하고자 하였으며, 그 결과 라소 정규화, 상관 분석 및 분산팽창지수, 주성분 분석, 경험적인 방법 순서로 변수 선택의 효과가 탁월하였다.

3) 연료소모율 예측모델을 생성하기 위하여 회귀 분석 및 인공 신경망 알고리즘을 적용하였으며 전반적으로 인공 신경망 모델이 다중선형회귀 모델에 비해 예측력이 우수하였다. 인공 신경망의 경우 변수들 간의 비선형적인 관계를 효율적으로 다루기 때문에 예측 결과의 정확도가 높은 반면 선형 회귀

모델은 다소 예측력이 떨어졌지만 추정한 회귀계수로부터 운항변수들의 영향력을 직관적으로 파악할 수 있다는 장점이 있었다.

4) 실제 선박에서 항로계획시 연료소모율 예측모델을 활용한다면 선박 운항자에 의해서 조정가능한 운항변수 또는 운항변수를 조정함으로써 영향을 받을 수 있는 환경변수가 예측모델의 대표적인 입력변수로 선정될 필요가 있으며 이러한 입력값들은 사용자가 계획하는 값을 직접 입력하거나 선내 시스템으로부터 자동으로 입력 가능하여야 한다. 따라서 경험에 의한 변수 선택 방법을 적용한 예측모델을 추가적으로 분석하였으며, 설명력이 0.7140, 0.9298로 다른 방법들에 비해서는 다소 떨어졌지만 예측모델의 사용 목적에 따라서는 이러한 임의의 변수 선정방법도 충분히 활용될 수 있을 것으로 판단되었다.

5) 본 연구에서 개발한 연료소모율 예측모델 중 라소 정규화에 의한 변수 선택과 인공 신경망 기반의 학습방법을 적용한 모델의 예측력이 0.9793으로 가장 높았으며 대상 선박으로부터 다양한 센서들의 정보와 충분한 양의 데이터를 수집할 수 있다면 해당하는 방법을 통한 예측모델 구현이 가장 효과적일 것으로 사료되었다.

5.2 제언

본 연구는 운항자의 의사결정을 지원하기 위한 에너지효율 최적화 시스템의 기초 단계인 연료소모율 예측모델을 개발하는데 의의가 있다. 현재 단계에서 개발된 연료소모율 예측모델은 선박의 운항 조건 및 외부 환경 조건이 주어지는 경우 단위 항해거리 당 연료소모량을 예측할 수 있기 때문에 운항하는 선박의 에너지 효율 모니터링에 충분히 활용될 수 있으리라 판단된다.

1) 실제로 선박 운항자가 연료소모율 예측모델을 활용하여 항로를 계획을 한다면 입력변수의 한 값을 조정하면 종속변수인 연료효율 뿐만 아니라 연관성을 가지는 다른 독립변수들도 영향을 받을 수 있기 때문에 정확한 예측 결과를 얻기 위해서는 각 변수들 간의 관계를 충분히 고려하여 반영할 필요가 있다. 또한 연료소모율을 최소로 하는 각 변수들의 최적화된 값을 찾는 방법에 관한 추

가적인 연구가 필요하다.

2) 데이터의 전처리 및 분석 과정에서도 향후 추가적으로 검토되어야 할 사항이 있었다. 본 연구에서는 컨테이너선박으로부터 수집한 데이터와 운항변수의 특성을 고려하여 중앙값 필터를 적용한 바 있다. 데이터 정제시 사용되는 필터의 종류, 필터의 간격에 따라 데이터의 필터링 결과가 달라질 수 있기 때문에 향후 연구에서는 다양한 필터 방법과 필터 간격에 대하여 고려할 예정이다. 또한 예측모델의 학습에 활용된 학습용 데이터와 평가용 데이터의 구분은 데이터의 전처리 완료 후 70%, 30%의 비율로 나누어서 활용하였지만 향후 연구에서는 항차 단위로 구분하여 활용할 수 있는 방안에 대하여 고찰해 볼 필요가 있었다.

3) 다양한 선종과 운항 조건에 대한 데이터가 축적된다면 데이터세트의 개수에 따른 모델의 예측성능을 비교분석하여 효율적인 모델 구현 조건에 관하여 파악할 수 있을 것이라고 생각되며 모델의 정확도와 신뢰도를 개선하여 고도화된 예측 시스템을 수립할 수 있을 것이라고 판단된다.

참고문헌

[1] 김경준, 이수동, 전치혁, 박개명, 변상수, 2017. *연료 소비 패턴 발견을 위한 컨테이너 선 운항데이터 분석의 통계적 절차*. 응용통계연구, 제 30권 제5호, pp.633-645.

[2] 유병현, 김진환, 김정중, 안영태, 2016. *실선운항기록을 이용한 환경외란중에서의 선박추진성능해석*. 2016년도 대한조선학회 춘계학술대회.

[3] 유은섭, 박개명, 문두환, 2018. *오픈소스 기반 빅데이터 플랫폼을 활용한 선박 운항 효율 예측 연구*. 한국 CDE 학회 논문집, 제 23권 제3호, pp.275-284.

[4] Abbasian, N. S., Salajegheh, A., Gaspar, H., & Brett, P. O., 2018. *Improving early OSV design robustness by applying 'Multivariate Big Data Analytics' on a ship's life cycle*. Journal of Industrial Information Integration, 10, pp.29-38.

[5] ABS, 2013. *Ship energy efficiency measures advisory*. Technical report.

[6] Armstrong, V. N., 2013. *Vessel optimisation for low carbon shipping*. Ocean Engineering, 73, pp.195-207.

[7] Beşikçi, E. B., Arslan, O., Turan, O., & Ölçer, A. I., 2016. *An artificial neural network based decision support system for energy efficient ship operations*. Computers & Operations Research, 66, pp.393-401.

[8] Breiman, L., Friedman, J., Olshen, R., & Stone, C., 1984. *Classification and regression trees*. Wadsworth Int. Group, 37(15), pp.237-251.

[9] Cattell, R. B., 1966. *The scree test for the number of factors*. Multivariate behavioral research, 1(2), pp.245-276.

[10] Eide, M. S., Longva, T., Hoffmann, P., Endresen, Ø., & Dalsøren, S. B., 2011. *Future cost scenarios for reduction of ship CO2 emissions*. Maritime Policy & Management, 38(1), pp.11-37.

[11] Fagerholt, K., Laporte, G., & Norstad, I., 2010. *Reducing fuel emissions by optimizing speed on shipping routes*. Journal of the Operational Research Society, 61(3), pp.523-529.

[12] Hotelling, H., 1936. *Simplified calculation of principal components*. Psychometrika, 1(1), pp.27-35.

[13] IMO, 2009a. *Second IMO GHG study 2009*, London.

[14] IMO, MEPC.1/Circ.683, 2009b. *Guidance for the development of a ship energy efficiency management plan (SEEMP).*

[15] IMO, MEPC.1/Circ. 684, 2009c. *Guidelines for voluntary use of the ship energy efficiency operational indicator (EEOI).*

[16] IMO, Resolution MEPC.203(62), MEPC 62/24/Add.1, Annex 19, 2011. *Amendments to the Annex of the Protocol of 1997 to Amend the International Convention for the Prevention of Pollution from Ships*, 1973, as Modified by the Protocol of 1978 Relating Thereto.

[17] IMO, Resolution MEPC.212(63), MEPC 63/23, 2012a. *Guidelines on the Method of Calculation of the Attained Energy Efficiency Design Index (EEDI) for New Ships.*

[18] IMO, MEPC.59/24/Add.1, Annex 19, 2012b. *Guidance for the development of a ship energy efficiency management plan (SEEMP).*

[19] Jolliffe, I. T., 1972. *Discarding variables in a principal component analysis. I: Artificial data.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 21(2), pp.160-173.

[20] Jolliffe, I. T., 1986. *Principal component analysis*. Springer, Newyork.

[21] Journée, J. M. J., Rijke, R. J., & Verleg, G. J. H., 1987. *Marine performance surveillance with a personal computer*. Technische Universiteit, pp.1-15.

[22] Kaiser, H. F., 1960. *The application of electronic computers to factor analysis.* Educational and psychological measurement, 20(1), pp.141-151.

[23] Lu, R., Turan, O., & Boulougouris, E., 2013. *Voyage optimisation: prediction of ship specific fuel consumption for energy efficient shipping*. In Low Carbon Shipping Conference, London, pp.1-11.

[24] Minsky, M., & Papert, S., 1969. *An introduction to computational geometry.* CambRIDGE tiass., HIT.

[25] Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W., 1996. *Applied linear statistical models*, 4, pp.318. Irwin:Chicago.

[26] Norstad, I., Fagerholt, K., & Laporte, G., 2011. *Tramp ship routing and scheduling with speed optimization.* Transportation Research Part C: Emerging Technologies, 19(5), pp.853-865.

[27] Parkes, A. I., Sobey, A. J., & Hudson, D. A., 2018. *Physics-based shaft power prediction for large merchant ships using neural networks.* Ocean Engineering, 166, pp.92-104.

[28] Pearson, K., 1901. *Principal components analysis.* Philosophical Magazine and Journal of Science, 6(2), pp.559.

[29] Pedersen, B. P., & Larsen, J., 2009. *Prediction of full-scale propulsion power using artificial neural networks.* In Proceedings of the 8th international conference on computer and IT applications in the maritime industries (COMPIT'09), Budapest, Hungary, pp.10-12.

[30] Perera, L. P., & Mo, B., 2016. *Marine engine operating regions under principal component analysis to evaluate ship performance and navigation behavior.* IFAC-PapersOnLine, 49(23), pp.512-517.

[31] Perera, L. P., & Mo, B., 2017. *Machine intelligence based data handling framework for ship energy efficiency.* IEEE Transactions on Vehicular Technology, 66(10), pp.8659-8666.

[32] Petersen, J. P., Jacobsen, D. J., & Winther, O., 2012. *Statistical modelling for ship propulsion efficiency.* Journal of marine science and technology, 17(1), pp.30-39.

[33] Pratt, W. K., 2007. *Digital image processing,* PIKS Scientific inside 4.

[34] Pukelsheim, F., 1994. *The three sigma rule.* The American Statistician, 48(2), pp.88-91.

[35] Ronen, D., 1982. *The effect of oil price on the optimal speed of ships.* Journal of the Operational Research Society, 33(11), pp.1035-1040.

[36] Rosenblatt, F., 1958. *The perceptron: a probabilistic model for information storage and organization in the brain.* Psychological review, 65(6), pp.386.

[37] Stopford, M., 2009. *Maritime economics.* 3rd Ed. Oxford:Routledge.

[38] Tibshirani, R., 1996. *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp.267-288.

[39] Turan, O., Demirel, Y. K., Day, S., & Tezdogan, T., 2016. *Experimental determination of added hydrodynamic resistance caused by marine biofouling on ships*. Transportation Research Procedia, 14, pp.1649-1658.

[40] Wang, K., Yan, X., Yuan, Y., & Li, F., 2016. *Real-time optimization of ship energy efficiency based on the prediction technology of working condition*. Transportation Research Part D: Transport and Environment, 46, pp.81-93.

[41] Wang, S., Ji, B., Zhao, J., Liu, W., & Xu, T., 2018. *Predicting ship fuel consumption based on LASSO regression*. Transportation Research Part D: Transport and Environment, 65, pp.817-824.

[42] Yan, X., Wang, K., Yuan, Y., Jiang, X., & Negenborn, R. R., 2018. *Energy-efficient shipping: An application of big data analysis for optimizing engine speed of inland ships considering multiple environmental factors*. Ocean Engineering, 169, pp.457-468.

부록 A 이론적 배경

A.1 상관 분석

상관 분석은 두 변수 사이의 연관성을 분석하는 통계기법으로써 상관계수가 클수록 밀접한 관계에 있다고 볼 수 있다. 상관계수는 변수의 분포특성(정규분포, t-분포 등)에 따라 모수적 상관 계수인 피어슨(pearson) 방법과 비모수적 상관 계수인 스피어만(spearman), 켄달(kendal) 등의 방법으로 구분할 수 있다.

피어슨 상관 분석은 두 변수가 갖는 선형관계를 분석하는 통계기법으로 식 (A.1)과 같이 나타낼 수 있다.

$$Pearson\ correlation\ coefficient = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2}\sqrt{\sum_i (y_i - \overline{y})^2}} \quad \text{(A.1)}$$

여기에서 $cov(X,Y)$는 두 변수의 공분산이며 σ_X, σ_Y는 두 변수의 표준편차이다. x_i는 변수 X에서 i번째 샘플의 값, y_i는 변수 Y에서 i번째 샘플의 값을 나타내며 $\overline{x}$, $\overline{y}$는 평균을 의미한다.

스피어만 상관 분석은 변수들의 상관관계를 순위를 매겨서 계산하는 방법으로서 연속형 변수가 아닌 순서형 변수인 경우에도 적용이 가능하다. 스피어만 상관계수를 구하는 식은 식 (A.1)의 상관계수 공식과 동일하나 피어슨 상관계수와는 다르게 x_i, 와 y_i의 값이 해당하는 변수의 i번째 순위값을 사용한다.

일반적으로 통계학에서는 상관계수의 크기에 따라 다음과 같이 관계를 정의할 수 있다.

1) 상관계수가 ± 1.0과 ± 0.7 사이이면 강한 양/음의 상관 관계

2) 상관계수가 ± 0.7과 ± 0.3 사이이면 뚜렷한 양/음의 상관 관계

3) 상관계수가 ± 0.3과 ± 0.1 사이이면 약한 양/음의 상관 관계

4) 상관계수가 ± 0.1과 0 사이이면 거의 무시할 정도의 양/음의 상관 관계

A.2 회귀 분석

회귀 분석은 독립변수와 종속변수간의 관계를 설명하거나 새로운 입력값에 대한 출력값을 예측하는데 사용된다. 독립변수의 개수에 따라 독립변수가 한 개인 경우는 단순회귀 분석, 둘 이상의 경우는 다중회귀 분석으로 구분하며 독립변수와 종속변수 사이의 함수관계가 선형을 이루고 있다고 가정하면 선형 회귀라 한다.

독립변수가 k개인 다중선형 회귀 분석의 기본 모형은 다음 식 (A.2)과 같이 나타낼 수 있다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon \tag{A.2}$$

여기에서 $X_1, X_2, \cdots, X_k$는 독립변수(independent variable), $\beta_0, \beta_1, \cdots \beta_k$는 회귀계수(regression coefficient), ϵ는 잔차(residual), Y는 종속변수(dependent variable)를 의미한다.

회귀 분석은 각 독립변수에 대응하는 회귀계수 β_0, β_k를 추정하는 것이다. 회귀계수를 구하기 위해서는 잔차 제곱합이 최소가 되게 하는 회귀식을 찾는 최소제곱법(least squares method)이 사용되며 식 (A.3)과 같다.

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} \beta_k x_{ik} \right)^2 \tag{A.3}$$

산출된 회귀식의 통계적 유의성은 회귀 모형의 F-검정, 종속변수에 대한 개별 독립변수는 t-검정을 수행하여 분석한다.

A.3 주성분 분석

주성분 분석(Principal Component Analysis;PCA)은 변수들 사이의 분산-공분산 관계를 이용하여 변수들의 선형결합으로 나타낼 수 있는 주성분을 찾고 데이터의 분산을 잘 나타내는 일부의 주성분을 활용하는 분석방법이다. 주성분 분석은 기존 변수들의 분산을 최대화하는 선형변환을 수행하여 새로운 주성분을 찾아냄으로써 차원을 감소시켜 해석을 용이하게 해준다. 또한 주성분은 서로 상관이 없거나 독립적인 새로운 변수들로 구성되며 정보의 손실을 최소화해준다(Jolliffe, 1986; Pearson, 1901; Hotelling, 1936). 이러한 주성분 분석은 다변량 자료에 존재하는 비정규성이나 이상치를 찾는데 사용되며 도출된 주성분은 서로 상관이 없는 독립적인 관계를 가지기 때문에 변수들 간의 상관관계로 인해 다중공선성이 의심되는 회귀 모형에도 응용할 수 있다.

주성분 분석을 통한 데이터의 차원 축소와 특징 추출 방법은 다음과 같은 순서를 통해서 진행된다.

입력 벡터 x가 p차원이고 데이터 샘플의 수가 k일 때 다음 식 (A.4)과 같이 입력 벡터의 평균을 계산한다.

$$\mu_i = \frac{1}{k}\sum_{i=1}^{k} x_i \tag{A.4}$$

입력 벡터와 평균 벡터의 차를 통해 공분산 행렬(covariance matrix)을 구한다.

$$C = \frac{1}{k-1}\sum_{i=1}^{k}(x_i - \mu_i)(x_i - \mu_i)^T \quad \text{(A.5)}$$

$\det(C - \lambda E) = 0$를 활용하여 공분산 행렬의 고유값(eigen value, λ_i)과 고유 벡터(eigen vector, V_i)를 구한다. 고유벡터는 분산이 가장 큰 방향을 나타내고 각각의 고유값은 해당하는 방향으로 축을 변환했을 때 분산의 크기를 나타낸다. p개의 변수로 이루어진 좌표계 $x(x_1, x_2, ...x_p)$를 정규직교행렬 $A_{(p*p)}$에서 회전하는 새로운 좌표계 $z(z_1, z_2, ...z_p)$로 변환할 때 z는 다음과 같이 나타낼 수 있다.

$$Z = AX \quad \text{(A.6)}$$

위의 크기 순서대로 정렬한 $\lambda_1 \geq \lambda_2 ... \geq \lambda_p \geq 0$를 행렬 A의 고유값이라 할 때 대응하는 고유벡터를 구하면 새로운 변수는 식 (A.7)과 같이 나타낼 수 있다.

$$z_k = a_{kp}x_p = a_{1k}x_1 + a_{2k}x_2 + ... + a_{pk}x_p \quad \text{(A.7)}$$

즉, 주성분 분석은 Fig. A.1과 같이 최적의 축으로 투영하여 p차원 공간을 그보다 더 낮은 차원의 공간으로 축소하는 것을 의미한다.

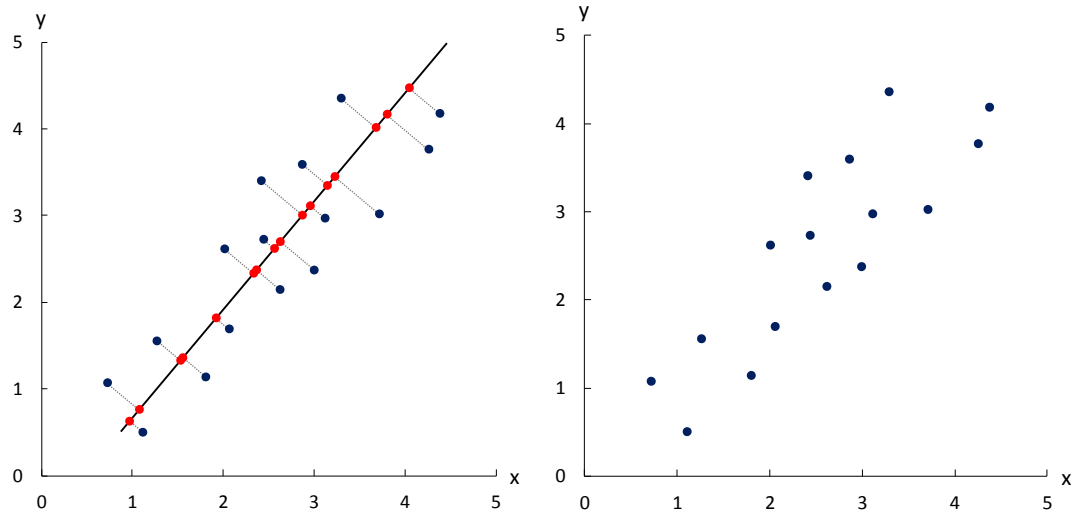

Fig. A.1 Linear transformation using principal component analysis

위 식 (A.7)에서 크기순으로 정렬한 고유값은 데이터의 특징을 나타내며 작은 고유값을 가지는 주성분은 입력 데이터와의 상관관계가 상대적으로 낮음을 의미한다. 차원축소를 통한 자료요약을 위해서는 설명력이 높은 주성분만을 선택할 필요가 있으며 이는 전체 분산에 대한 공헌도, 고유값의 크기, 스크리 산점도 등의 방법으로 판단할 수 있다.

1) 전체 분산에 대한 공헌도

여기에서 z_n를 제 n 번째 주성분이라 하면 전체 분산에 대한 해당 주성분의 기여율은 다음과 같이 나타낼 수 있다.

$$c_n = \frac{\lambda_n}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \tag{A.8}$$

최초 m개의 주성분에 의해 설명되는 누적기여율은 다음과 같다.

$$\sum_{k=1}^{m} c_k = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_m}{\lambda_1 + \lambda_2 + \ldots + \lambda_p}, m < p \tag{A.9}$$

통상적으로 최초 m개의 주성분(z_p) 중에서 누적기여율이 약 80~90%를 넘는 경우 나머지 $p-m$개의 주성분은 분석 대상에서 제외한다.

2) 고유값의 크기

공분산행렬 대신에 상관행렬을 사용하는 경우 통상적으로 각 주성분의 고유값(λ_i)이 1 이상인 주성분의 개수만큼 선택한다. 이는 상관행렬을 사용하는 경우 평균 고유값이 1이기 때문이다 (Kaiser, 1960). 그러나 고유값이 1보다 큰 주성분만 구성하기에는 그 개수가 너무 적어서 0.7보다 큰 주성분을 제안하는 연구도 있었다 (Jolliffe, 1972).

3) 스크리 도표

스크리 도표는 Fig. A.2 와 같이 고유값을 크기순으로 나열하고 x축에는 차원의 수, y축에는 차원의 고유값을 나타낸다. 스크리 산점도에서 고유값이 감소하는 추세가 급격히 완만해지거나 0에 가까워지는 것의 이전까지 주성분을 선택한다 (Cattell, 1966).

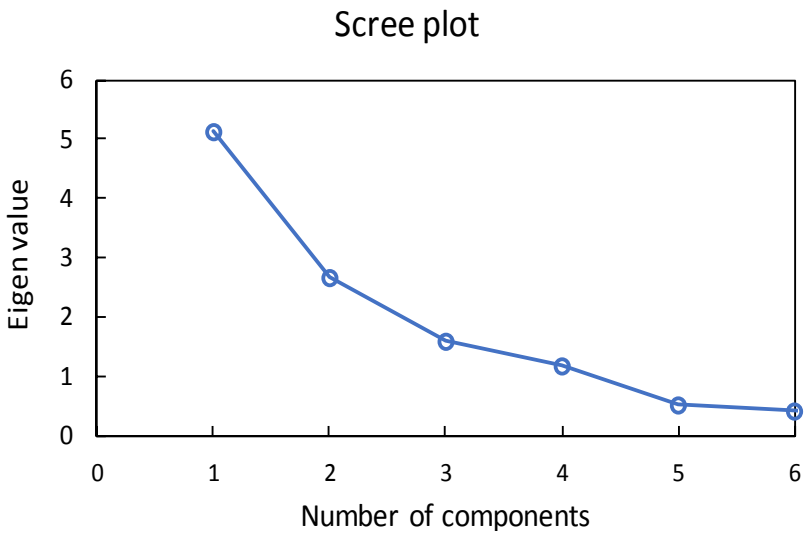

Fig. A.2 An example of scree plot

A.4 라소 정규화

일반적으로 회귀 분석에서는 독립변수의 개수가 많아지면 독립변수들 사이의 강한 상관관계로 인한 다중공선성(multicollinearity)이 발생할 수 있으며 이로 인하여 회귀계수 추정량의 분산이 커져 회귀 모형의 예측정확도가 떨어지는 문제가 발생할 수 있다. 또한 종속변수에 대한 해석력이 떨어지며 각 변수에 대한 역할을 판단하기 어렵게 한다. 이러한 문제점을 해결하기 위하여 라소 정규화(Least Absolute Square and Selection Operator;LASSO)는 회귀계수의 크기에 조절 모수(tuning parameter)를 부여함으로써 상대적으로 영향력이 적은 독립변수의 회귀계수 값을 축소해주는 방법이다. 영향력이 없는 변수의 회귀계수를 0으로 만들어 주기 때문에 변수 선택을 가능하게 하고 이에 따라 해석력이 뛰어난 모형을 만들어준다 (Robert, 1996).

라소 정규화는 식 (A.10)과 같이 기존의 잔차 제곱합에 추가적으로 회귀계수의 절대값의 합을 최소화하는 것을 제약조건으로 한다.

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2, \sum_{j=1}^{p} |\beta_j| \le t \tag{A.10}$$

부등식 제한조건이 있는 기존의 식을 라그랑주 승수법을 사용하여 변환하면 다음과 같이 나타낼 수 있다.

$$\hat{\beta}^{lasso} = arg\min_{\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{A.11}$$

여기에서 p는 독립변수의 개수, λ는 기존의 잔차 제곱합과 추가적인 제약 조건의 비중을 조절하기 위한 조절 모수이다.

λ가 크면 정규화 정도가 커지고 회귀계수들의 값이 작아진다. 반면에 λ가 작아지면 정규화 정도가 작아지며 0이 되면 일반적인 선형 회귀 모형이 된다. 잔차 제곱합 항과 제약조건 항의 합이 최소가 되게 하는 회귀계수 β와 λ를 찾는 것이 목적이다.

Fig. A.3는 라소 회귀 모형을 그림으로 나타낸 것이다. 좌표축은 각 회귀계수의 추정치를 의미하며 도형의 빗금 쳐진 부분은 λ가 포함된 제약조건을 의미한다. 마름모는 라소의 제약조건인 $|\beta_1|+|\beta_2| \leq t$을 나타낸다. 등고선의 중심에 있는 $\hat{\beta}$는 최소제곱추정치인 $\hat{\beta_1}, \hat{\beta_2}$을 의미하며 타원형 등고선은 잔차 제곱합이다. 등고선이 제약구간과 닿은 지점의 좌표가 추정된 회귀계수이다. 라소의 경우 사각형과 제약구간이 만나는 지점에서 추정된 회귀계수가 0으로 수렴하게 하여 변수 선택의 효과를 가질 수 있다. 변수를 제거함으로써 차원이 축소되며 모형의 해석력을 향상시킬 수 있다.

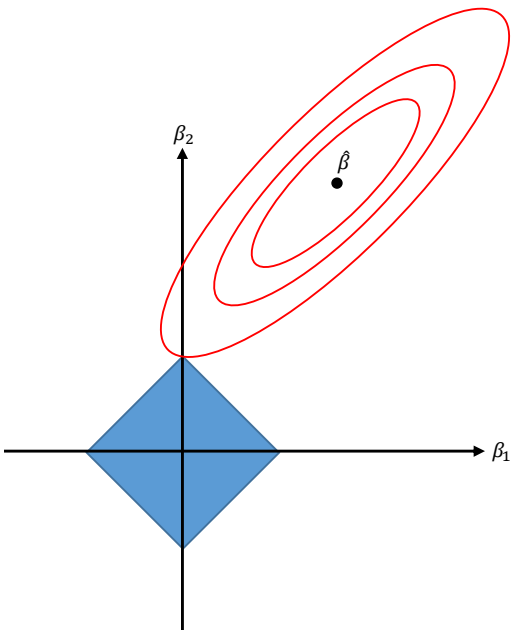

Fig. A.3 Geometric interpretation of LASSO regression

부록 B 인공 신경망 이론

사람의 뇌는 뉴런이라 불리는 신경세포의 복잡한 연결로 이루어져 있다. 하나의 뉴런은 다른 여러 뉴런과 복잡한 망으로 연결되어 있으며 망을 구성하는 기본물질은 시냅스로 이루어져 있다. 수상돌기는 시냅스로부터 신호를 입력받으며 세포체는 수상돌기로부터 신호를 받아들여 임계값 이상의 값을 가질 때 축삭으로 전달해준다. 축삭은 신호를 다음 시냅스로 전달하는 역할을 한다. 인공 신경망(Artificial Neural Network;ANN)은 이러한 인간의 생체 신경망 구조를 단순화하여 모델링한 것이다. Fig. B.1는 인공 신경망의 단일 뉴런을 나타낸 예시이다. 시냅스에서 들어오는 화학적 신호(x)인 입력값과 시냅스의 연결강도인 가중치(w)가 결합하여 합산된 후 활성화 함수(f)를 거쳐 최종적으로 출력값이 도출된다 (Frank, 1958).

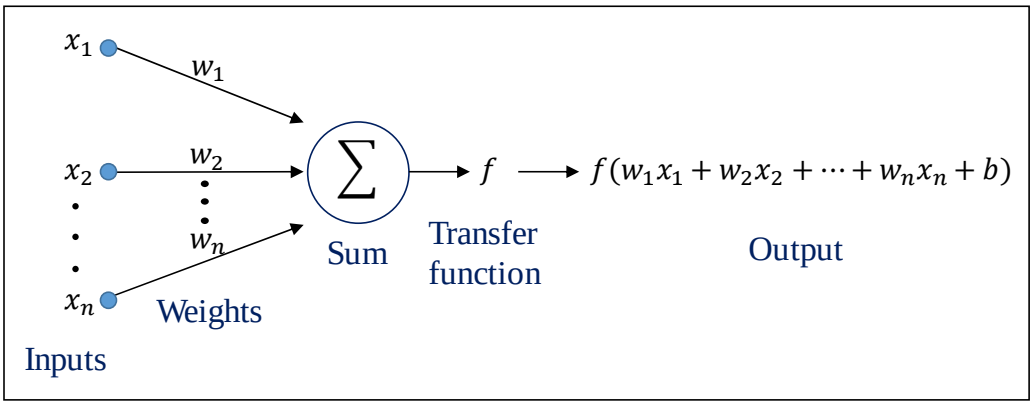

Fig. B.1 The basic concept of a single artificial neuron

이를 식으로 나타내면 다음과 같다.

$$v = b + \sum_{i=1}^{n} w_i x_i \tag{B.1}$$

$$y = f(v) \tag{B.2}$$

여기에서 x는 입력, w는 가중치, b는 편향, v는 합산된 출력, f는 활성화 함수이며, y는 출력을 나타낸다.

Minsky and Papert (1969)는 한 개의 층으로 이루어진 단층 퍼셉트론의 한계점을 극복하기 위하여 3개 이상의 층으로 구성된 다층 퍼셉트론을 제안하였다. 다층 퍼셉트론은 단층 퍼셉트론과 유사한 구조를 가지지만 입력층과 출력층 사이에 하나 이상의 은닉층이라 하는 중간층을 두어 비선형성을 가지는 데이터에 대해서도 학습이 가능하도록 하였다. 입력과 출력결과의 비선형 관계를 나타내기 위해서 계단, 임계논리, 시그모이드와 같은 활성화 함수를 은닉층에 사용한다. 역전파 알고리즘은 출력층에서 발생하는 오차 값을 역으로 은닉층으로 전파하여 출력값과 예측값의 오차를 작아지도록 가중치와 편향 값을 최적화하는 과정을 말한다.

미분이 간단하여 역전파 알고리즘에 적합한 시그모이드 함수 $f(x)$는 다음 식과 같이 나타낼 수 있다.

$$f(x) = \frac{1}{1 + e^{-\beta x}} \tag{B.3}$$

여기에서 β는 함수의 기울기 경사를 결정하는 상수이다.

Fig. B.2는 β의 값에 대한 시그모이드 함수의 형태를 나타낸 것이다.

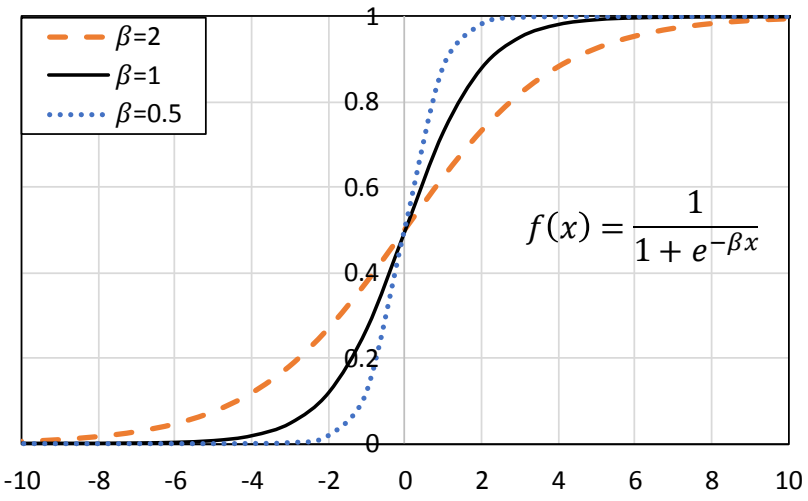

Fig. B.2 Sigmoid function with different beta values

부록 C 인공 신경망 모델의 경향 분석

회귀모델에서는 회귀계수로부터 각 독립변수가 한 단위 변화함에 따라 종속변수에 미치는 영향력을 파악할 수 있다. 반면 인공 신경망 모델의 경우 사용된 데이터의 패턴을 출력값에서 제시하지 못하기 때문에 블랙박스(black box)라고 불리기도 하였다. 따라서 본 논문에서는 사용된 모든 독립변수를 분석범위의 중간 수준으로 고정하고 분석하고자 하는 변수를 최소에서 최대값으로 변경하면서 모델의 예측 결과를 확인하여 경향성을 파악하고자 하였다.

Fig. C.1 ~ C.6은 Table 4.4의 Case 8에 해당하는 인공 신경망 모델에서 각 독립변수의 연료소모율에 대한 경향성을 분석한 결과이다. SOG=15.2knot, STW-SOG=-0.1knot, Relative wind speed=36.45knot, Relative wind direction=90deg, Mean draft=13.48m, Trim=-0.08m으로 고정하였으며 분석 대상이 되는 변수만 차례대로 최소값에서 최대값으로 변화시키면서 연료소모율 결과를 분석하였다.

1) Fig. C.1은 선박의 대지속력에 따른 연료소모율을 나타낸 그래프이다. 선속이 10-14노트 구간에서는 연료효율이 거의 일정하며 전체 선속 운용 구간 중에서 최고 높은 효율을 보인다. 14노트부터는 선속이 증가할수록 단위 항해거리당 연료소모량이 증가하여 연료효율이 좋지 않다. 10노트 미만의 구간에서도 10-14노트 구간에 비해서 연료효율이 좋지 않은 것을 알 수 있다. 이로부터 대상선박의 운항 일정을 충족시킬 수 있는 범위 내에서 저속 RPM으로 선박을 운용하는 것이 가장 에너지효율적이라고 할 수 있으며, 10노트 미만의 항내전진 속력에서는 잦은 주기관의 분당 회전수 변경으로 인하여 연료효율이 다소 떨어지는 경향이 있다. 다만, 이는 연료소모량의 절감만을 고려한 선속의 예시이며 실제 선박에서는 여러 변수들이 연료효율에 미치는 영향을 종합적으로 고려할 필요가 있다.

2) Fig. C.2는 선박의 대수속력과 대지속력의 차에 따른 연료소모율을 나타낸 그래프이다. 대지속력이 대수속력보다 클수록 즉, 대수속력과 대지속력의 차 값이 (-)일수록 연료효율이 증가하는 것을 알 수 있다. 이는 선박의 대지속력에

영향을 미칠 수 있는 조류 및 해류와 같은 기상환경요인이 선박의 연료효율과 크게 연관되어 있음을 나타낸다. 본 연구에서는 대수속력과 대지속력의 차로 외력성분을 대신하였으나 실제 운항해역의 해상 및 기상정보를 습득하여 모델에 반영한다면 더욱 정확한 해석이 가능할 것이라 판단된다.

3) Fig. C.3는 선박의 상대풍속에 따른 연료소모율을 나타낸 그래프이다. 상대풍속이 0-30노트 구간에서는 바람에 의한 연료소모율의 차이가 거의 없으나 30노트 이상부터는 연료소모율이 급격히 감소함을 알 수 있다. 따라서 대상선박에서는 가능하다면 상대풍속이 30노트 이내를 유지할 수 있도록 선박의 침로 및 속도를 조정하여 항해하는 것이 연료 저감에 효과적일 것으로 판단된다.

4) Fig. C.4는 선박의 상대풍향에 따른 연료소모율을 나타낸 그래프이다. 상대적으로 선수방향에서 바람을 조우할 때 단위 항해거리 당 연료소모량이 가장 많으며 조우각도가 커질수록 연료효율이 좋아진다. 특히, 0-45도까지의 선수 및 선수측면에서 불어오는 바람의 경우 연료소모율이 가장 떨어지며 해당하는 구간을 벗어나면서부터 급격히 연료소모율이 좋아지는 것을 알 수 있다. 따라서 대상선박에서는 가급적이면 상대풍향이 0-45도 구간을 피하도록 선박을 운용하는 것이 연료소모량 저감을 위한 좋은 방안이라고 할 수 있다.

5) Fig. C.5는 선박의 평균흘수에 따른 연료소모율을 나타낸 그래프이다. 평균흘수가 12.5-13.5m 구간에서 단위 항해거리 당 연료소모량이 가장 많으며 14.5m를 전후로 하여 연료효율이 가장 좋은 것을 알 수 있다. 본 연구의 대상선박은 계획만재흘수(Design draft)가 14.5m이며, 이는 일반적으로 가장 경제적이면서 최적의 운항 조건을 갖도록 설계된 흘수를 말한다. 이러한 인공 신경망의 예측결과로부터 실제 선박의 운항성능을 잘 나타내는 것으로 사료된다.

6) Fig. C.6는 선박의 트림에 따른 연료소모율을 나타낸 그래프이다. 최소 트림구간에서부터 최대 트림구간까지의 변화에 따른 연료소모율의 변동을 고려할 때 트림이 미치는 영향은 다른 운항변수들에 비하여 다소 미약한 것으로 판단된다. 그래프를 참조하면 선박의 운항 조건(SOG=15.2knot, STW-SOG=-0.1knot, Relative wind speed=36.45knot, Relative wind direction=90deg, Mean

draft=13.48m)에서는 선미트림 0-0.8m 구간에서 연료효율이 가장 좋은 것으로 나타난다.

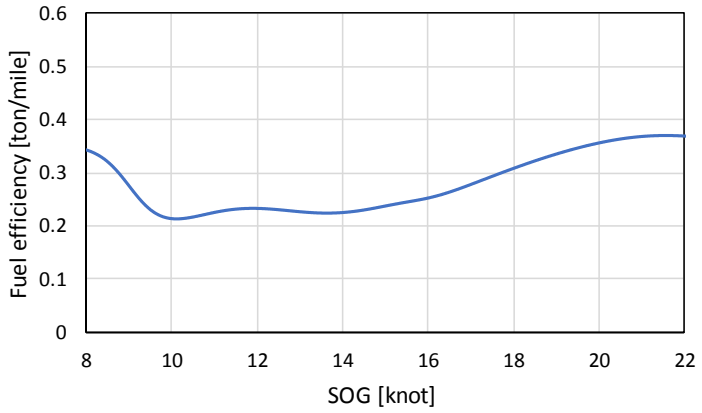

Fig. C.1 Trend analysis of fuel consumption rate by speed of the ground

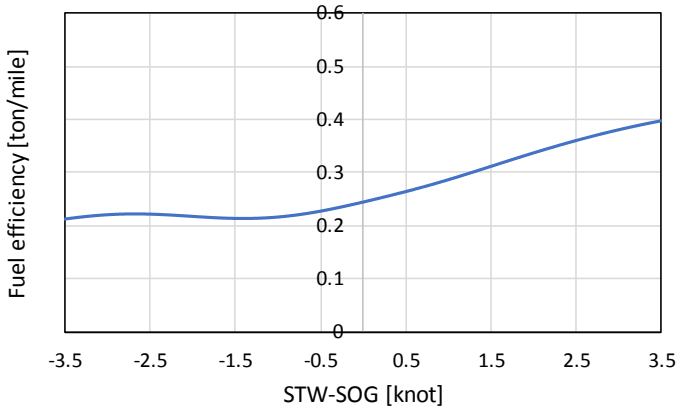

Fig. C.2 Trend analysis of fuel consumption rate by STW-SOG

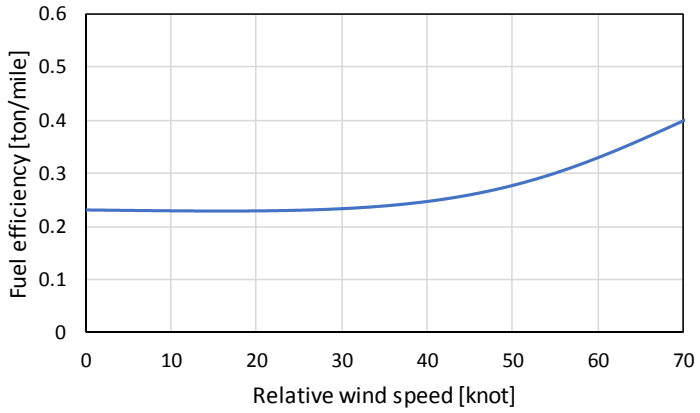

Fig. C.3 Trend analysis of fuel consumption rate by relative wind speed

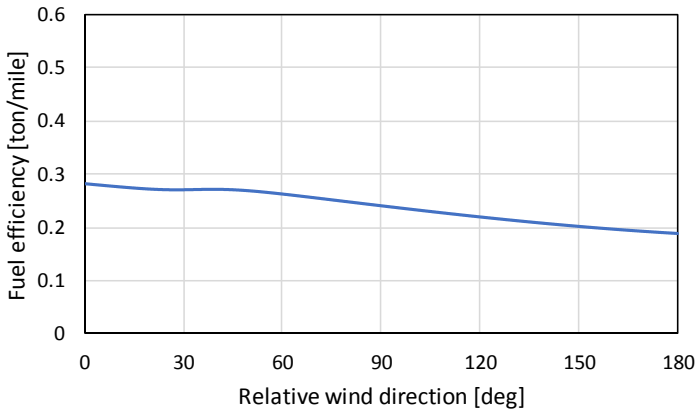

Fig. C.4 Trend analysis of fuel consumption rate by relative wind direction

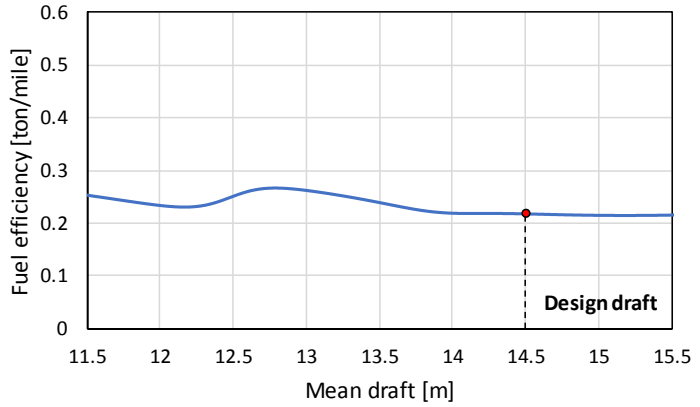

Fig. C.5 Trend analysis of fuel consumption rate by mean draft

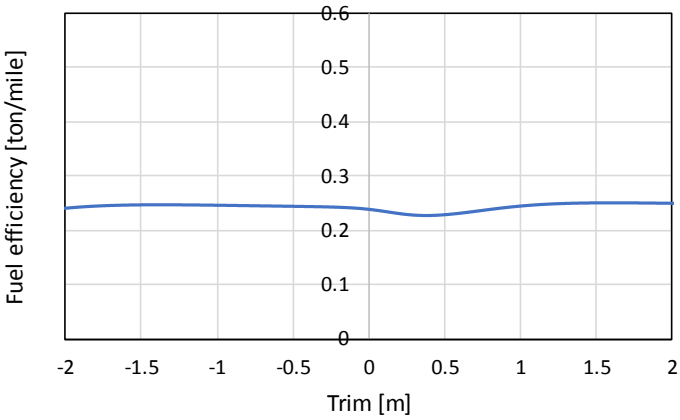

Fig. C.6 Trend analysis of fuel consumption rate by trim

감사의 글

어느새 2년의 석사과정을 마치고 학위 논문을 제출하게 되었습니다. 지난 시간을 돌이켜보면 아쉬운 일도 많았지만 스스로 성장할 수 있는 귀중한 시간이었습니다. 많은 분들의 관심과 도움이 있었기에 부족한 제가 대학원 생활을 무사히 해낼 수 있었습니다. 짧은 지면으로나마 그분들께 감사의 마음을 전합니다.

먼저 아낌없는 조언과 지도를 해주신 박준범 교수님께 진심으로 존경과 감사의 마음을 전합니다. 연구의 자세와 덕목에 대한 조언뿐만 아니라 인생의 선배로서 많은 것들을 배울 수 있었습니다. 교수님의 가르침을 마음속에 늘 간직하며 자랑스러운 제자가 되기 위해 노력하겠습니다.

귀한 시간을 내어 본 논문에 대하여 세심한 심사와 충고를 해주신 정연철 교수님과 문성배 교수님 그리고 항해학과 교수님들께도 감사드립니다.

또한 연구 과제를 진행하면서 많은 도움을 주신 한국형 e-Navigation 연구팀, 부산대학교 박종천 교수님 연구팀께 감사드리며 실험실에서 많은 조언과 격려를 해주신 전석희 교수님을 비롯한 선후배님들께도 고마움을 전합니다.

마지막으로 제가 선택한 길을 변함없이 믿고 응원해준 가족들에게 정말 감사드립니다.

미처 언급하지 못했지만, 저를 아끼고 응원해 주셨던 모든 분들께 진심으로 감사드립니다. 앞으로 더욱 정진하고 성장하여 세상에 많은 것을 베풀 수 있는 존재가 되도록 하겠습니다.

2019년 6월

김영롱 올림