

머신러닝 기반의 해군 정비인시 예측 모델

유정민^o

^o국방대학교, 해군 군참부

e-mail: code_nol@naver.com^o

Man Hour Prediction Model for ROK Navy's Maintenance Based on Machine Learning

Jun-Min Yoo^o

^oKorea National Defense University, Dept. of Logistic Branch, ROK Navy

● 요약 ●

해군에서 운용하는 각 함정은 여러 가지 무기체계를 동시에 탑재하고 있는, 고가의 복합무기체계이다. 주어진 기간동안 효과적인 정비를 수행함으로써, 가동률을 극대화하는 것이말로 복합무기체계인 함정을 효과적으로 운용하는 방법이며, 경제적인 국방운용이라 부를 수 있을 것이다. 정비인시는 여러 무기체계가 동시에 탑재된 복합무기체계 정비의 핵심이다. 정비인시를 정확하게 알고 있어야만, 제한된 정비기간을 각 무기체계에 정확히 할당할 수 있을 것이며, 최적의 요소에 대한 정비가 수행될 수 있을 것이다. 본 연구에서는 해군에서 운용중인 특정장비에 대한 약 10년간의 정비자료를 이용하여 정비인시를 예측하는 모델을 제안하였다. 모델의 성능은 R2 Score를 통해 0.69의 준수한 수치를 보였다. 이 모델을 통해 조금더 세밀하고, 정확한 정비인시 예측과 정비계획 수립이 가능하리라 판단하였다.

키워드: 머신러닝(Machine Learning), 회귀(Regression), 인시(Manhour)

I. Introduction

무기체계의 가동률이란 해당 무기체계를 사용 가능한 시간과 그렇지 않은 기간의 비율을 의미한다. 당연히 높은 가동률을 갖춘 무기체계가 전투에서 효과적으로 사용될 수 있다. [1] 해군에서 운용하는 전투함, 지원함 등의 각종 선박은 통상 함정이라 부르며, 이러한 함정은 대개 고가의, 여러 무기체계가 동시에 탑재된 복합무기체계이다. 고가의 함정에 대한 원활한 운용유지를 위해서는 체계적이고, 효과적인 정비지원이 필요하다. 결론적으로 효과적인 정비지원을 통해 함정 가동률을 높이는 것이, 경제적인 국방운용의 한 방법이라 말할 수 있는 것이다. 본 연구에서 주목한 것은 정비지원에 필요한 정비기간이다.

가동률을 높이는 방법에는 완성도 높은 정비수행을 통해 고장률을 낮추는 방법과, 정비수행 자체에 필요한 정비기간을 낮추는 방법 등이 있다. 본 연구에서 다루는 비와 같이, 정확한 정비인시를 예측할 수 있다면, 정비와 정비사이의 휴지기간이 최소화되도록 정비계획을 수립할 수 있을 것이고, 이는 결국 함정 가동률 향상으로 이어질 수 있을 것이다.

실제로 해군의 정비지원부대에서는 특정 항목을 수행하는데 필요한 기간과 대략적인 자원들을 미리 정해두고 있는 실정이다. 물론, 대부분 정해진 기간안에 수행이 가능하도록 여유시간을 두고 있음을 말할

필요가 없을 것이다. 본 연구에서 제안하는 모델을 활용한다면, 이렇게 정비를 수행할 때 필요한 기간을 보다 정확히 예상할 수 있을 것이고, 기술변화에 따라 정비기간을 유연하게 적용할 수 있을 것이다.

II. Preliminaries

1. Man Hour of ROK Navy's Maintenance

해군에서의 정비는 통상 공시라는 단어로 불리며, 정비에 필요한 인시는 공시라고 불린다. 1인이 1시간당 수행할 수 있는 정비를 1공시라고 부르며, 하루에 수행하는 8공시를 1공수라고 부른다. 이러한 공시는 여러 가지 요인에 의해 변화될 수 있다. 예를 들면, 기술력의 변화, 정비를 위한 지원장비의 변화 등에 의해 변화될 수 있다. 또한 동일한 장비에 대한 점검 수준의 정비가 수행되더라도, 정비지원 간 수리부속 교체여부 및 교체수량에 따라 공시가 변화될 수 있을 것이다. 이렇게 공시라는 것은 1인이 1시간당 수행한다라는 개념만이 정해진 것이며, 실제 수행되는 양은 상황과 대상에 따라 변화하고 또 기술발전에도 따라 변화하므로 효과적인 정비지원을 위해서는 해당 상황에 맞는 최선의 사전 예측이 절실한 상황이다.

2. Related Works

2.1 Regression using Machine Learning

회귀는 입력변수와 출력변수 사이의 관계를 모델링하는 통계적인 방법이다. [2] 특정함수를 통해 모델링된 모델을 이용하여 새로운 값에 대한 출력을 예측하는 것이다. 머신러닝을 통한 회귀모델은 예측의 정확도 향상, 다양한 변수 처리, 과적합 방지, 대규모 데이터 처리 등의 이점이 있기에 다양한 분야에서 사용되는 방법이다. 회귀 분석의 성능과 유연성을 향상시키며, 모델의 정확도를 높일 수 있는 것이다. 본 연구에서는 정비인시에 대한 수치적 예측이 필요로하기에 회귀 모델을 적용하였으며, 대규모 데이터 처리가 요구되므로 머신러닝을 이용하였다.

2.2 Algorithms of Regression

회귀에 사용되는 대표적인 알고리즘에는 여러 가지가 있다. 본 연구에서는 선형, 의사결정트리, 랜덤포레스트, 그래디언트 부스팅 회귀 등 4가지 알고리즘을 적용하였다. 선형 회귀는 입력 데이터와 출력 데이터간 선형 관계를 모델링하는 것이며, 의사결정트리 회귀는 입력데이터를 분할, 각 구간에서의 출력값을 예측하는 것이다. 랜덤포레스트는 여러 의사결정트리를 결합하여 앙상블 학습을 수행하는 방식이며, 그래디언트 부스팅은 일련의 결정트리를 학습하는 방식으로 진행된다. 이 외에도 다양한 알고리즘이 있을 수 있으나, 본 연구에서는 대표적인 위 4가지 알고리즘을 적용하고, 결과를 비교하였다.

2.3 R2 Score

R2 스코어는 회귀 모델의 예측 성능을 평가하기 위한 지표 중 하나이다. [3] 결정계수라고도 불리우며, -∞과 1사이의 값을 갖는다. 이 값은 실제값과 예측값의 차이를 분산의 비율로 나타낸 것이며, 수식으로 나타내면 아래와 같다.

$$R^2 = 1 - (SSE/SST)$$

여기서 SSE는 잔차의 제곱합, SST는 총 제곱합을 의미한다. Mean Squared Error 등 회귀모델에서 사용할 수 있는 평가 지표들은 다양하게 존재하지만, 본 연구에서는 모델의 설명력을 표현할 수 있는 평가 지표인 R2 스코어를 이용하였다. 일반적으로 R2 스코어는 0.5 이상인 경우 보통, 0.7 이상인 경우 우수한 성능으로 평가한다.

III. The Proposed Scheme

1. Tools and Dataset

본 연구에서는 데이터 수집을 위해 SQL을 사용하였고, 전처리 및 머신러닝 모델 구축을 위해 Python을 사용하였다. Python에는 numpy, pandas, sklearn, matplotlib 등 라이브러리를 같이 사용하였다.

연구에 사용한 데이터는 데이터의 축적 정도와 품질을 고려하여 해군에서 운용중인 특정 추진 디젤엔진 1종으로 하였다. 해군 함정 9척에서 약 15년 이상 운용중인 장비이며, 모두 23개의 개체가 존재한

다. 정비는 총 4개 부대에서 지역별로 수행되었고, 기능이 유사한 부서를 통합하여 집계한 결과 세부적으로는 99개 부서에서 부분별 정비를 수행하였다.

장비를 구성하는 그 부분품을 통상 구성품이라고 부르는데, 이러한 구성품은 입력항목에 별도로 설정된 것은 아니므로, 원 데이터에서 자연어처리를 통해 추출된 단어들로, 이전 연구에서와 같이 62개를 추출하여 적용하였다. [4] <표 1>은 이와 같은 구성품 목록을 정리한 내용이다.

Table 1. Assembly list of Equipment

num	Assembly	num	Parts
1	유체커플링	10	인터쿨러
2	과급기	11	해수펌프
3	조속기	12	노즐
4	발칸커플링	13	LOP
5	청수펌프	14	밸브시트
6	청수쿨러	15	청수에열기
7	차폐장치
8	윤활유쿨러
9	스터디Brg'	62	W-5

2. Preprocessing

앞서 설명한 바와 같이, 데이터들을 모두 수집한 결과 17,355열 × 12행 수준으로 축적되었다. 수집된 최초 변수(행)은 <표 2>에서 나타내었다.

Table 2. List of Columns

num	Assembly	num	Parts
1	함정 차선	7	작업 일자
2	적용 함명	8	정비지원부대
3	운용 부서	9	정비 부서
4	적용 엔진	10	사용 수리부속 명칭
5	적용 구성품	11	사용 수리부속 수량
6	표준정비 여부	12	정비인시

최초 적용된 변수 중 사용 수리부속의 경우 여러 중복데이터가 존재하거나, 서로 다른 내용이 포함된 내용들이 많았다. 단순히 제거하기에는 해당 변수가 모델 설명력에 미치는 영향이 클 것으로 생각되어 등장하는 수리부속의 명칭을 새로운 컬럼으로 만드는 방법을 수행하였다. 결론적으로 기존 수리부속 명칭 및 정비인시 컬럼을 삭제하고, 식별된 수리부속 178개의 자료를 추가하였다.

최종적으로 중복과 이상치, 결측치 제거를 실시하고, 최소-최대 정규화 등의 전처리를 수행한 결과 9,455열 × 188행의 데이터를 확보하였다.

IV. Results

1. Exploratory Data Analysis

먼저 데이터 특성에 대한 이해를 위해 탐색적 데이터 분석을 진행하였다. 탐색적 데이터 분석은 통계분석, 상관관계 분석 순으로 진행하였다.

1.1 Statistics Analysis

먼저, 9척의 함정과 23대의 엔진을 기준으로 데이터의 통계적 특성을 확인하였다. 데이터의 기준연도와 각 함정의 취역연도의 차이로 인해, 함정별 누적 공사의 건수는 최소 600여건에서 최대 1400여건까지 분포하였으며, 비교적 조기에 획득된 1차선의 함정에서 정비데이터가 많이 누적되어 있었다. 반대로 1공사당 평균 정비인시는 최소 18시간에서 최대 23시간까지 근사한 차이를 보였다. 엔진의 경우 18대의 사용엔진 대비 5대의 휴지기간 중인 엔진이 있었기에 정비데이터의 누적수는 편차가 큰 편이었으나, 1공사당 평균 정비인시는 최소 17시간에서 최대 25시간까지 유사한 수치를 보였으며, 각 함정 및 엔진당 평균 정비인시는 20시간으로 동일한 수치를 보였다. <그림 1>은 이러한 결과를 그래프로 표현한 결과이다.

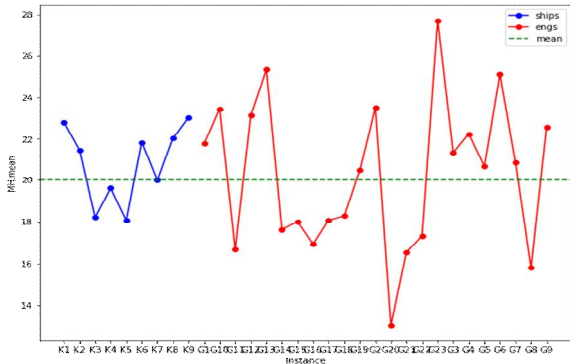


Fig. 1. MH,mean of Instance

1.2 Correlation Analysis

소모된 자재를 새로운 변수로 추가하기 전, 수집된 데이터를 통해 상관관계를 확인하였다. 상관계수를 확인한 결과 최대값은 함명과 차선사이에서 나타났으나, 이 결과는 크게 의미가 없는 것이었으며, 정비인시와 관계있는 변수 중 수리부속 교체 수량이 0.5이상의 상관계수를 보였다.

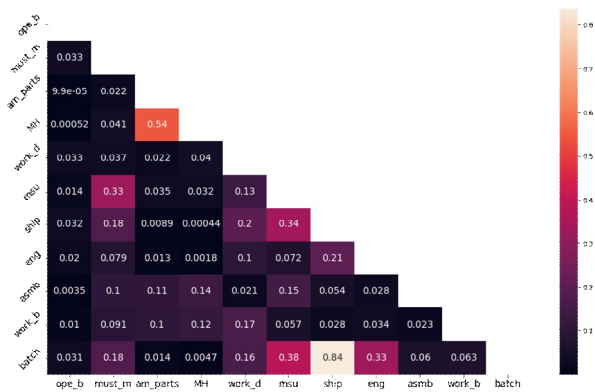


Fig. 2. Heatmap for Correlation

2. Prediction and Evaluation

다음으로는 선형회귀와 의사결정트리, 랜덤포레스트 및 그래디언트 부스팅 알고리즘을 사용하여 모델을 학습하고, 예측 및 R2 스코어를 확인하였다. 먼저 선형회귀를 사용한 결과, R2 스코어는 음수의 값을 기록하였으며, -636,640의 값으로 설명력이 전혀 없는 모델로 판단하였다.

의사결정나무 알고리즘을 사용하여 분석한 결과 R2 스코어는 0.407의 값으로 선형회귀보다는 향상되었지만 여전히 설명력이 부족한 모델로 평가하였다. <그림 3>은 의사결정나무 분석결과를 그래프로 나타낸 결과이다.

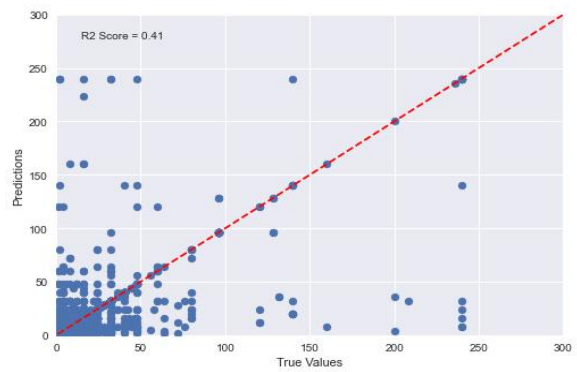


Fig. 3. Result of Decision Tree Classification

랜덤포레스트 회귀 모델의 경우, 의사결정나무 알고리즘보다 향상된 0.697의 결과를 보였다. 설명력을 갖춘, 사용할 수 있는 모델로 판단하였다. 또한 그래디언트 부스팅을 이용한 결과 0.692의 결과를 보여 유사한 수준을 보였다. <그림 4>와 <그림 5>는 이 결과를 나타낸 그래프이다.

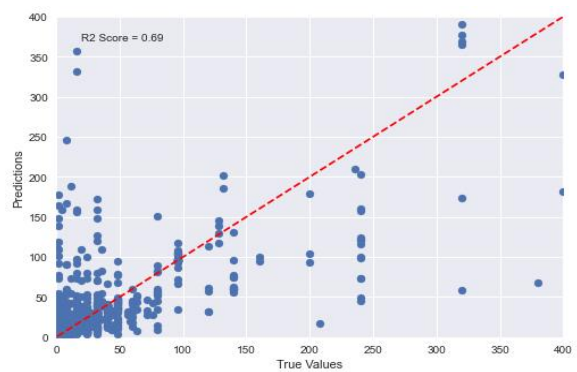


Fig. 4. Result of RandomForest Classification

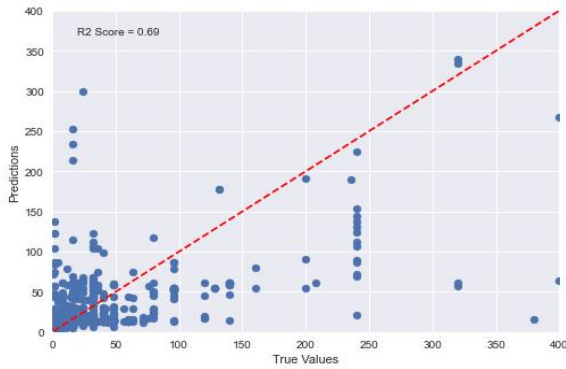


Fig. 5. Result of GradientBoosting Classification

V. Conclusions

본 연구에서는 해군의 특정 장비에 대한 정비인사를 예측하기 위한 머신러닝 모델을 제안하였다. 또한 연구 결과를 R2 스코어를 통해 제시하였으며, 확인된 R2 스코어는 정비데이터를 이용하여 학습된 이러한 머신러닝 모델이 충분히 설명력이 있음을 나타내고 있다. 과정 중 사용된 알고리즘 중에서는 랜덤포레스트와 그래디언트 부스팅 알고리즘을 적용한 회귀모델이 가장 우수한 성능을 보였다.

향후에는 머신러닝 모델의 설명력을 보다 높이기 위해 다양한 정비데이터가 추가된다면 더 높은 정확도를 보일 수 있으리라 예상된다. 예를들면, 해당시점의 정비시설 및 공구현황, 정비인력의 근무기간 등 숙련도와 같은 변수가 추가될 수 있을 것이다.

마지막으로, 이와 같은 모델의 확대적용을 통해 보다 효과적이고 효율적인 정비가 수행될 수 있기를 기대한다.

REFERENCES

- [1] M. P. Enright, et.al, "A New Method for Assessing Weapon System Operational Availability," IEEE Transactions on Reliability, 2004.
- [2] Joseph Hilbe, "Regression Analysis of Count Data," Cambridge Series in Statistical and Probabilistic Mathematics, 2014.
- [3] J. Neter, et.al, "The Analysis of Variance," Journal of the American Statistical Association, 1976.
- [4] Jung-Min Yoo, ea.al, "SNA-based Trend Analysis of Naval Ship Maintenance," Journal of The Korea Society of Computer and Information, Vol. 24, No. 6, pp. 165-174, June, 2019.