



Received: 2023/11/19
Revised: 2023/11/30
Accepted: 2023/12/25
Published: 2023/12/31

***Corresponding Author:**

Chan-Ha Lee

E-mail: lch1577@naver.com

함정선체 정비인시 예측 모델 연구

A Study on the Prediction Model of Warship's Hull Maintenance Manhour

안원철¹, 유동훈², 이찬하^{3*}, 신승민⁴, 오경원⁵, 신일식⁶

¹(주)베가스 데이터 책임분석가

²(주)베가스 데이터 수석분석가

³(주)베가스 국방데이터 전문가, 공학박사

⁴부산대학교 선박해양플랜트기술연구원 초빙교수

⁵호원대학교 항공정비공학과 교수

⁶중소조선연구원 민군협력센터 센터장

Wonchul Ahn¹, Donghoon Yoo², Chan-Ha Lee^{3*}, Sengmin Shin⁴, Kyungwon Oh⁵,
Shin Ilsik⁶

¹Data analyst, Begas Co., Ltd.

²Data analyst, Begas Co., Ltd.

³Ph.D/National Defense Data Consultant, Begas Co., Ltd.

⁴Visiting professor, Research Institute of Ships & Ocean Engineering,
Pusan University

⁵Professor, Dept. of Aircraft MRO Engineering, Howon University

⁶Center director, Civil-military Support Center, Research Institute of
Medium & Small Shipbuilding

Abstract

해군정비창은 선제적인 정비지원을 위하여 함정선체 정비인시 예측 모델을 개발하고자 한다. 우선 업무 프로세스 이해를 통해 필요한 데이터를 식별하고, 전처리 과정을 거쳐 데이터셋을 준비하였다. 이후 데이터 규모 확대를 위하여 정비항목 군집화를 수행하였다. 모델은 전통기법과 인공지능 기법을 활용하였으며, MAE를 최소로 하는 관점에서 성능을 평가하였다. 그 결과 인공지능 기법이 우수한 성능을 제공하지 못하였다. 이는 데이터 부족과 특성이 제대로 고려되지 못한 것으로 해석되며 데이터에 대한 인식이 개선되어야 할 것으로 사료된다.

The Naval Shipyard intends to develop a predict model of Warship's hull maintenance manhour for preemptive maintenance support. First, understand the business rule to identify the data you need, a data set was prepared through a preprocessing process. After that, maintenance item clustering was performed to expand the data scale. The model used traditional techniques and AI techniques, Performance was evaluated from the viewpoint of minimizing MAE. As a result, AI techniques did not provide excellent performance. This is interpreted as a lack of data and characteristics not properly considered, It is believed that the perception of data should be improved.

Keywords

함정선체(Warship's Hull),
정비인시(Maintenance Manhour),
정비항목 군집화(Maintenance Item Clustering),
인공지능 기술(AI Techniques),
데이터 부족 및 특성(Lack of Data and Characteristics)

Acknowledgement

본 논문은 2021 민군협력개발사업 "IoT 기반 함정 정비 통합관제 플랫폼 개발"에 의해 수행되었음 (21-CM-TN-12).

1. 서론

인류는 20세기 후반 컴퓨터, 인터넷 기반의 정보 기술(IT: information technology) 정보혁신을 거쳐 현재는 데이터 기술(DT: data technology) 혁명 시대로 접어들고 있다. 21세기의 화두는 현실세계(physical system)의 다양한 현상을 ICBM(IoT, cloud, big-data, mobile) 기술과 인공지능(AI: artificial intelligence) 기술을 융합하여 디지털 사이버세계(cyber system)와 결합한 CPS(cyber-physical system)시대를 여는 것이다. 이는 그동안 정보통신기술을 통해 축적해 온 방대한 데이터를 분석하여 개별 요구에 부응하는 새로운 가치를 창출하고 미래를 예측하는 시대를 만드는 것으로, 데이터의 활용 가치는 더욱 높아지고 세상은 데이터를 기반으로 급격하게 발전하게 될 것이다.

이러한 시대적 변화에 발맞춰 국방부는 데이터 기반의 디지털 국방 구축을 위해 신개념·신기술과 연계한 관련 법제도 마련, 민군 협

력체계 강화, 국방업무의 표준화 및 디지털 전환, 디지털 전문인력 양성, 데이터 분석·활용 기반의 업무 일상화 등 다양한 정책과 연구개발을 시행하고 있다[1-5].

데이터 분석·활용과 관련된 연구로, 해군 군수사에서 군수업무의 디지털화를 위해 빅데이터 기술이 적용 가능한 분석주제 발굴 연구 및 시범 구축 사업[6,8]을 수행하였으며, 해군함정기술연구소는 함정장비의 센서 데이터를 활용한 고장예측 시스템에 관한 연구[7,9]를 진행하였다. 국방부에서는 국방 빅데이터 인공지능 공통 분석 서비스 사업으로 피복, 급식 등에 대한 수요예측 분석모델 개발을 위한 연구[10]를 진행하였다. Hong et al.[11]은 심층신경망(DNN)을 활용한 선박 전력부하 예측방법을 제안하였으며, Choi et al.[12]은 디지털 트윈을 활용한 수리부속 예측 방법을 제안하였다.

이러한 다양한 연구개발 노력에도 불구하고 공감대 형성의 어려움, 전문인력의 부족, 이해당사자 간 충돌, 일하는 문화의 변화 지연 등은 국방업무 디지털화를 위해 여전히 해결해야 할 숙제로 남아 있다.

따라서 본 연구에서는 국방업무 디지털 전환 업무의 일환으로 해군정비창의 정비데이터를 활용하여 입찰 예정 함정에 대한 선제적인 정비지원을 준비하고자 한다.

이 목표를 달성하기 위한 분석주제는 하향식 접근법(top-down approach)을 적용하여 발굴하고, 새로운 문제 탐색에 있어 필요시 상향식 접근법(bottom-up approach)도 병행한다.

하향식 접근법에서 모델 개발을 위해서 본 연구는 우선 데이터 분석을 통해 연고자 하는 목표를 설정한다. 다음으로 관련된 업무 프로세스를 명확히 이해하고 분석주제를 정의한다. 그리고 필요한 데이터를 선정하여 수집하고 업무 프로세스 확인을 통해 적합성을 점검한다. 이후 분석용 데이터셋(data set)을 준비하고, 데이터 분석을 통해 최적의 모델을 식별한다. 다음으로 수집한 데이터셋을 대상으로 시뮬레이션을 시

행하여 모델의 성능을 검증한다. 마지막으로 본 연구에서 최종적으로 얻은 결론을 요약하고, 정리한다.

2. 비즈니스 이해

2.1. 분석 목표 설정

해군정비창의 최종 목표는 데이터 분석을 통해 선제적인 함정 정비지원을 준비하는 것이다. 이를 위해 Table 1과 같이 자재 소요를 수반한 정비 수요예측을 분석 수행 방향으로 제기하였다.

Table 1. 정비 수요예측 분석 수행 방향

기반	정비대상	분석 수행 방향
정비실적 기반 수요예측	동일장비	자체정비단계 초과 PMS 목록 식별
		표준정비항목 목록 식별
		비표준공사 목록 식별
장비상태 기반 수요예측	동일장비	고장예측 S/W 기반 정비항목 식별
		상동
		상동

2.2 업무 프로세스 이해

해군정비창에서 발생하는 정비 수요예측을 위해서는 정비대상, 정비개념 등이 먼저 정의되어야 한다.

해군은 정비(maintenance)의 개념을 사용할 수 있는 장비 및 보급품을 항상 사용할 수 있는 상태로 유지하는 행위와 사용 불가능한 장비 및 보급품을 원상태로 복구하는 일체의 행위(손질, 검사, 수리, 재생, 보존, 시험, 개조, 수정)로 정의하고 있다[13].

이러한 정비의 개념은 Fig. 1과 같이 고장정비, 점진정비, 주기정비 형태로 변화하였으며, 대한민국 해군에서는 1968년 거제함(PCE-1003)을 미 해군으로부터



Fig. 1. 정비개념의 변천사

터 인수해 오면서 같이 도입된 계획정비시스템(PMS: planned maintenance system)을 기반으로 하여 발전하였다.

이러한 정비개념의 변천사를 미루어 보면, PMS는 운전시간을 기반으로 고장 간격을 도출하고 이를 기초로 정비를 수행하는 시간기반정비(TBM: time-based maintenance) 개념의 시스템이라 할 수 있다. 정비교범 및 야전 운용실적 기반의 해군 핵심 정비시스템인 PMS는 신형함정의 증가, 작전 소요의 변동성과 시급성으로 인한 고기술 정비 물량 증가, 고장예측의 불확실성 등 정비환경의 변화와 데이터 기술의 발전과 더불어 장비 상태를 진단하여 예방·예견·선행 정비가 수행되는 상태기반정비(CBM: condition-based maintenance) 개념이 가미되었고, 근래에는 수명주기 중심의 CBM+ 기반 예측정비(predictive maintenance) 개념과 융합하여 점차 고도화되고 있다.

정비대상이 되는 장비는 일련번호별로 개별 관리하는 개체장비와 종류별로 다수의 장비를 일괄적으로 관리하는 모듈장비로 분류할 수 있다. 또한 해군의 정비시스템은 수행 주체에 따라 함정과 정비창으로 분류할 수 있으며, 각 장비의 정비단계로는 부대정비, 야전정비, 창정비로 구분한다.

정비항목별로는 함정 자체 수행 PMS 항목과 정비창 이관(지원) PMS 항목, 정비창 지원 표준공사 및 비표준 공사 항목으로 분류할 수 있다. 함정의 경우 정비주기가 다른 다수의 장비가 장착 및 탑재되어 있고, 함정 운용 특성상 일정 주기별로 수리기간을 설정하여 정비지원부대에 입창, 일괄적인 정비를 시행한다. 따라서 함정의 정비지원부대(정비창, 수리창 등)는 각 장비의 정비단계별로 야전정비와 창정비를 병행하여 지원한다. 그리고 정비항목별로는 정비창 이관(지원) PMS 항목, 표준공사, 비표준 공사 항목을 지원한다.

Table 2는 해군 정비시스템 분류 결과를 정리하여 나타낸 것이다.

Table 2. 해군 정비시스템 분류 결과

장비분류	수행주체	정비단계	정비항목	
			PMS 정비항목	공사항목
개체정비, 모듈정비	함정, 정비창	부대정비, 야전정비, 창정비	함정 자체 PMS, 정비창 지원 PMS	표준공사, 비표준공사

이러한 정비시스템이 지속 고도화되고 조기에 안정화되기 위해서는 정비 수행 간 생성되는 운용 데이터가 더 세밀하게 수집·관리되어야 하며, 분석을 통해 도출한 인사이트를 업무 프로세스에 더욱 빠르게 환류시켜야 한다.

2.3. 분석주제 정의

본 연구에서는 적용 대상이 해군정비창인 점과 업무의 우선순위를 고려하여 정비대상을 장비로 제한하였고, 정비항목은 정비창에서 지원하는 표준공사 중심의 정비 수요예측 부문에 한정하여 분석을 진행하였다.

개체장비의 정비 소요는 발행된 공사명령서(정비지시서)에 TBM 개념을 적용하여 정비항목별 정비주기를 추정하고, 정비주기 도래 시점을 고려하여 식별할 수 있다. 그러나 함정 선체의 경우, 정비범위가 광범위하고 정비항목이 다양하지만 정비이력은 동일 장비코드로 일괄 관리되고, 모듈장비의 경우도 다수의 장비에 대한 정비이력이 일괄 관리됨에 따라 동일 장비별, 동일 정비항목별 정비간격 계산이 어렵다. 따라서 함정선체와 모듈장비의 정비 소요는 과거 소비인시 데이터를 활용한 정비인시 예측을 통해 식별할 수 있다.

본 연구에서는 함정선체에 대한 정비 소요를 식별하기 위한 정비인시 예측 모델을 분석주제로 정의하고 연구를 진행한다.

3. 데이터 준비

3.1 필요 데이터 식별

함정선체 정비인시 예측 모델 개발에 필요한 데이터는 해군정비창에서 생성된 데이터 중 업무 프로세스, 국방군수통합정보체계(DELIS) Table 정의서, 관련 유관기관 등의 업무 관계자 인터뷰 결과를 토대로 식별하였다.

개발 모델에 활용되는 데이터의 규모는 현재 해군 정비창 데이터의 품질이 수집기간을 과거로 확대할수록 저하되는 현상을 고려하여 최상의 신뢰도를 확보할 수 있는 8개년(2016년 1월~2023년 7월)으로 선정하였다.

Table 3는 업무 프로세스 이해를 통해 모델 개발에 필요한 주요 데이터를 정리하여 나타낸 것이다.

Table 3. 주요 필요 데이터

Table		Column
한글명	영문명	
품목제원	TACO_ITEM_PDQI	국가품목식별번호, 품목제원번호, 한글품명, 품류코드 등
국방장비목록편람	TACO_DELD	국가품목식별번호, 국방장비코드 등
표준공정	TLNO_STDD_PROCS	정비부대코드, 품목제원번호, 예방기능코드, 표준예방문서번호, 표준공사명 등
해상정비지시서	TLNO_SEA_MTORD	정비부대코드, 정비지시서번호, 품목제원번호, 예방기능코드, 표준예방문서번호, 공사명 등
해상정비지시서 상세	TLNO_SEA_MTORD_DTL	정비부대코드, 정비지시서번호, 예방인시 등
공사일보	TLNO_CSTWRK_DAYRPT_PTCLS	일보일자, 정비지시서번호, 소비인시 등

3.2 데이터 전처리 및 정합성 점검

수집된 필요 데이터의 정합성 점검을 위한 전처리 과정은 Table 3의 각 column을 대상으로 운용 특성을 고려하여 이상치 및 결측치를 식별하고 정제작업 (data cleansing)을 수행하는 것이다.

이 과정에서는 이상치(outliers) 제거, 결측치 (missing data) 처리(삭제, 대체, 예측 등), 중복 데이터 제거, 데이터 형식 통일화(숫자, 문자, 날짜 등의 형식을 표준화), 데이터 정규화(데이터를 비교할 수 있는 단위로 변환), 데이터 유효성 검사(데이터의 논리적인 오류나 모순을 찾아 수정) 등이 수행된다.

수집된 필요 데이터는 업무 프로세스 확인을 통해 Fig. 2와 같이 결합하고, 정제작업을 통해 모델 개발을 위한 분석용 데이터셋으로 변환된다.

Table 4는 최종적으로 구축한 분석용 데이터셋의 규모를 나타낸 것이다.



Fig. 2. 데이터 전처리 프로세스

Table 4. 분석용 데이터셋 구축 결과

구분	데이터 규모 (건)	데이터 기간	대상
정비대상마트	25,962	2016년 1월 ~	해군정비창
정비이력마트	500,438	2023년 7월	

4. 데이터 분석

탐색적 자료 분석(EDA: exploratory data analysis)은 데이터를 탐색하고 이해하기 위한 과정이다. 분석 대상 데이터의 특성, 구조, 패턴, 이상치, 관계 등을 파악하기 위한 방법론 및 절차를 의미하며, 데이터 이해 단계에서 변수의 분포와 특성을 파악하고, 변수 생성 단계에서 분석 목적에 맞는 주요한 요약 및 파생변수를 생성하며, 변수 선택 단계에서 목적변수에 의미 있는 후보 변수를 선택하기 위해 활용된다.

본 장에서는 함정선체 정비인시 예측 모델 개발을 위하여 데이터의 규모와 주요 변수의 특성 검토 등에 대한 탐색적 자료 분석을 통해 분석 수행 방향을 구체화하고, 최적의 모델을 식별한다.

4.1 입력 데이터 정의

모델 개발에 필요한 입력 데이터는 품목제원번호, 예량기능코드 등 기준정보와 분석 결과 데이터 도출을 위한 변수로 구분할 수 있다.

Table 5는 모델에 필요한 주요 입력 데이터를 정리하여 나타낸 것이다.

Table 5. 주요 입력 데이터

Input data		Source data	
Type	Variable	Data mart	Column
기준	정비지시서번호	정비이력	MTORDNO
	국방장비코드		FAMC
	품목제원번호		ITMDNO
	국가품목식별번호		NIIN
	한글품명		HANGL_ITMNM
	품류코드		KICD
	예량기능코드		NAVY_ESTMQTY_FCN_CD
	표준예량문서번호		STDD_EQDOCNO
변수	정비일자	STD_CSTWRK_NM	
	총 소비인시	MNTNC_DT	
		TOT_CNSPTN_MH	

4.2 입력 데이터 수준 평가

입력 데이터 수준 평가는 분석용 데이터셋의 규모, 품질 및 구조를 파악하는 과정으로, 데이터 품질 문제 및 분석모델 개발에 필요한 조치 등을 파악할 수 있다. 여기서는 결측치 등을 제외하고 분석모델에 실질적으로 활용할 수 있는 데이터 규모를 중점적으로 살펴보고자 한다.

Table 6는 정비인시 예측을 위한 입력 데이터의 수준 평가 결과를 정리하여 나타낸 것이다.

Table 6. 정비인시 예측용 입력 데이터 수준평가 결과

입력데이터 규모	분석항목별 데이터 규모		
	정비이력 빈도	분석항목 수	비중 (%)
71,234	10건 이상	1,655	12.9
	5건 이상~10건 미만	1,911	14.9
	1건 이상~5건 미만	9,294	72.3
	계	12,860	100.0

입력 데이터(정비지시서) 규모만 고려하면 약 7만 건 수준으로 분석을 위한 일정 수준의 데이터를 확보한 것으로 판단할 수 있다. 그러나 정비인시를 예측하기 위해서는 함정선체의 표준공사 단위(이하 “분석항목”)의 모델 개발이 필요하다. 따라서 모델 개발을 위한 가용 데이터는 분석항목별로 검토해야 한다. 분석항목 단위로 데이터 규모를 살펴보면, 전체 중 5건 이상의 데이터가 확보된 표준공사는 약 27% 수준에 불과하다.

분석모델 개발을 위한 최소한의 데이터 규모는 분석기법 또는 데이터 특성에 따라 차이가 있을 수 있으나, 일반적으로 최소 수십 건 이상의 데이터가 확보되어야 분석모델의 성능을 보장할 수 있다. 따라서 분석항목별 원천 데이터 확보 및 분석항목 통합 등의 조치를 통한 충분한 데이터 확보가 선행되어야 한다.

4.3 동형/유사장비별 군집화

본 절에서는 분석항목별로 부족한 데이터를 확보하기 위하여 동일장비별 표준공사(표준정비항목) 대상으로 발행된 공사명령서(정비지시서)를 동형/유사장비별로 추가 군집화한다.

이때 동형/유사장비는 동일 장비와 구성품 일부가 상이한 장비로 구성품이 동일한 부분에 대한 운용 데이터에 한해 통합 분석할 필요가 있으며, 이를 위해 장비코드별·정비항목별 단위로 상세화하여 데이터를 통합, 분석한다.

군집화는 3단계로 나누어 진행한다. 우선 자연어 처리 기술을 활용하여 공사명령서(정비지시서)의 공사명 정보를 통합하고, 다음으로 현업 전문가 인터뷰를 통해 동형/유사장비별로 그룹화한다. 그리고 그 결과를 현장 실무자가 검증하는 방식으로 수행한다.

Table 7은 동형/유사장비별 정비항목 군집화 결과 일부를 정리하여 나타낸 것이며, 본 연구의 정비인시 예측 모델은 동형/유사장비 군집화한 장비코드와 군집화 코드 단위로 개발한다.

Table 8은 동형/유사장비 군집화 후 입력 데이터 수준평가 결과를 나타낸 것이며, 기존과 대비하여 입력 데이터 규모가 증가하였다. 이는 군집화 과정에서 개체장비로 잘못 발행된 정비항목이 함정 선체로 정상화되었기 때문이다. 또한 분석항목을 통합함에 따

Table 7. 동형/유사장비별 정비항목 군집화

동일장비별 정비항목					동형/유사장비별 정비항목		
장비코드	장비명	예량기능	문서번호	정비항목명	군집화 장비코드	군집화코드	정비항목명
05111104	레이더 세트	T	1	추적레이더 공기건조기 양륙/탑재(FFG)			
40100102	헬기구축함-I(2차선)	T	589	30mm 골키퍼 양륙/탑재(DDH-I)	40100201	G000003	정비품 양륙/탑재 지원
40100201	헬기구축함-II	T	632	청수탱크 양륙/탑재			
40100101	헬기구축함(DDH)	A	409	캡 제작(소형: 30 mm 이하) 5개			
40210201	잠수함구조함	A	409	캡 제작(소형: 30 mm 이하) 5개	40100101	G000007	캡 제작
40350102	예인정 300TON	A	409	캡 제작(소형: 30 mm 이하) 5개			

Table 8. 정비인시 예측용 입력 데이터 군집화 전후 비교

정비이력 빈도	분석항목별 데이터 규모 (군집화 전)			분석항목별 데이터 규모 (군집화 후)		
	입력데이터 규모	분석항목 수	비중 (%)	입력데이터 규모	분석항목 수	비중 (%)
10건 이상	71,234	1,655	12.9	90,807	502	34.6
5건 이상~10건 미만		1,911	14.9		269	18.5
1건 이상~5건 미만		9,294	72.3		681	46.9
계		12,860	100.0		1,452	100.0

라 데이터가 5건 이상 확보된 분석항목의 약 53 %로 증가하였다.

4.4 데이터 변환

데이터 변환은 입력된 원천 데이터를 분석하기 유용한 형태로 가공하거나, 분석 목적에 맞게 조정하는 작업이다. 분석모델 개발에 필요한 데이터 중 분석용 데이터셋에서 직접 추출이 불가능한 데이터에 일련의 변환규칙을 활용하여 새로운 변수를 유도한다.

함정선체 정비인시 예측을 위한 입력 변수로는 총 소비인시를 활용하였다. 다만, 총 소비인시는 현재 공사명령서(정비지시서) 단위로 수집되기 때문에, 분석항목별로 정보의 개수(또는 길이)가 지나치게 다양하다(1개~6,000개). 따라서 본 연구에서는 공사명령서(정비지시서) 단위의 총 소비인시 정보를 월 단위로 집계하여 각 분석항목의 변수 길이를 규격화하며, 이를 모델 개발을 위한 변수로 활용한다.

따라서 각 분석항목에서 활용할 수 있는 월 단위별

총 소비인시 정보는 데이터 수집기간과 같이 최대 91개가 된다.

Table 9은 현재 구축한 데이터셋을 기준으로 각 분석항목에서 활용할 수 있는 월 단위별 총 소비인시의 분포를 나타낸 것이다. 월 단위별 총 소비인시 정보가 20개 이상 존재하는 분석항목은 전체 중 약 17 %이며, 과반수의 분석항목은 활용할 수 없는 월 단위별 총 소비인시 정보가 5개 미만이다.

Table 9. 월 단위별 총 소비인시 수준평가 결과

분석항목별 데이터 규모		
정비이력 빈도	분석항목 수	비중 (%)
20개 이상	243	16.7
10개 이상~20개 미만	160	11.0
6개 이상~10개 미만	207	14.3
5개 이하	842	58.0
계	1,452	100.0

Figs. 3-4는 분석항목 중 일부에 대한 월 단위별 총 소비인시의 시계열적 추세를 나타낸 것이다. Fig. 3와 같이 데이터가 충분히 확보되고, 규칙성이 보이는 분석항목은 예측 모델 개발이 적절해 보이나, Fig. 4와 같이 부족한 경우는 모델보다는 산술평균 등의 기술 통계를 활용하는 방안도 고려해 볼 필요가 있다.

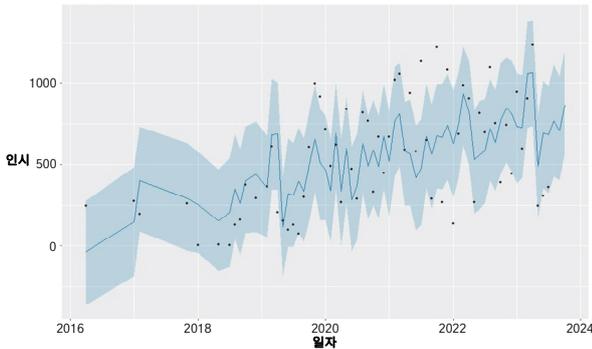


Fig. 3. 데이터가 충분한 경우(64개)

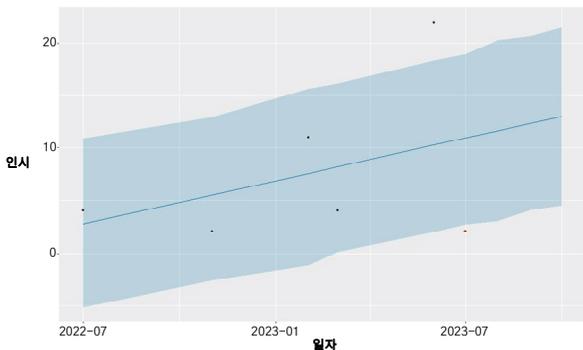


Fig. 4. 데이터가 부족한 경우(5개)

5. 시뮬레이션 및 고찰

본 연구에서는 분석항목(군집화한 함정선체 장비 코드, 군집화코드) 단위로 모델을 개발하여 미래의 정비인시를 예측하고자 하였다. 모델의 입력 변수는 앞서 언급한 바와 같이 총 소비인시를 월 단위로 집계한 월 단위별 총 소비인시 정보를 활용한다.

개발 모델에는 현재 해군에서 시계열 데이터 분석에 사용 중인 이동평균 및 최소자승법 등의 전통적인 기술과 LSTM(long short-term memory) 등의 인공지능(AI) 기술을 병행 활용하고자 한다.

5.1 분석 기술

분석 기술은 Table 10과 같이 크게 인공지능(AI)

기술과 통계적 추론(statistical inference), 기술통계(descriptive statistics) 등의 전통적인 기술로 구분할 수 있다.

Table 10. 분석 기술 분류

분류	분석 알고리즘/기법
인공지능(AI)	Prophet, LSTM
통계적 추론	최소자승법, 지수평활법, 추세보정지수평활법
기술통계	산술평균, 이동평균, 가중이동평균

인공지능(AI) 기술은 인간이 가진 인지·추론·판단 등의 지능을 컴퓨터로 구현하고자 시작된 기술로 기계학습(ML: machine learning)과 딥러닝(DL: deep learning)이 대표적이다.

이중 본 연구에 적용하는 Prophet은 Facebook에서 개발한 ML 기반의 시계열 예측 알고리즘으로 직관성이 높고, 결측값이 포함된 데이터에도 적용할 수 있는 장점을 갖고 있다[14]. 주요 구성요소는 trend, seasonality, holiday이며 다음과 같은 수식으로 정의할 수 있다.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_i \tag{1}$$

여기서, $g(t)$ 는 주기적이지 않은 변화인 트렌드를 나타내며, 부분적으로 선형 또는 로지스틱 곡선을 표현한다. $s(t)$ 는 주 단위, 월 단위, 연 단위 등 주기적으로 나타나는 패턴을 나타낸다. $h(t)$ 는 휴일과 같이 불규칙한 이벤트를 나타내며, ϵ_i 는 정규분포라고 가정된 오차를 나타낸다.

LSTM은 순환신경망(RNN: recurrent neural network)의 한 종류로 시계열 데이터의 예측, 자연어 처리, 음성 인식 등에 활용하는 기법이다. LSTM을 이해하기 위해서는 전통적인 신경망과 순환신경망에 대한 이해가 필요하다.

전통적인 신경망은 입력과 출력 간의 관계를 학습하는 데에는 효과적이지만, 순차적이고 시간에 따라 변하는 데이터를 처리하는 데에는 어려움이 있다. 이러한 한계로 인해 순환신경망은 순차적인 데이터에 잘 적응하는 루프 구조로 설계되어 각각의 단계에서 이전 단계의 출력을 현재 단계의 입력으로 사용함으로써 이전 정보를 보존하고 활용할 수 있도록 하였다.

그러나 일반적인 순환신경망도 장기의존성에 대한 처리가 어렵다는 문제가 제기되었고, 해법으로 등장한 LSTM은 메모리 셀(memory cell)과 여러 게이트를 사용하여 장기의존성을 학습하고 유지할 수 있게 하였다[15,16]. 이로써 순환신경망이 시계열 데이터나 문장 등 긴 시퀀스(sequence)에 대한 학습에서 뛰어난 성과를 거둘 수 있게 되었다.

LSTM은 정보를 저장하고 전달하는 역할을 담당하는 메모리 셀이 순차적으로 연결되어 결과를 출력하는 모델이다. 메모리 셀은 입력, 출력, 망각 게이트로 구성된다. 입력 게이트는 새로운 정보를 메모리 셀에 추가하여 어떤 정보를 받아들일지를 결정하는 역할을 담당하며, 망각 게이트는 불필요한 정보를 삭제하는 역할을 담당한다. 마지막으로 출력 게이트는 메모리 셀의 내용을 기반으로 최종 출력을 생성한다.

통계적 추론은 관측된 표본에 기초하여 관측 불가능한 매개변수를 추정하는 방법으로 본 연구에서는 최소자승법, 지수평활법, 추세보정지수평활법을 적용한다.

기술통계는 표본 자체의 속성을 파악하는 데 주안점을 두는 데이터 분석 방법으로 산술평균, 이동평균, 가중이동평균 등이 이에 속한다.

5.2 평가 지표

본 연구에서는 예측값의 정확도를 나타내기 위해 평균절대오차(MAE: mean absolute error)를 사용한다. 이 지표는 회귀 평가에 주로 사용되며, "0"에 가까울수록 좋은 모델이라고 할 수 있다. MAE의 수식은 식 (2)와 같다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (2)$$

여기서, n 은 전체 데이터의 개수, x 는 i 번째까지의 실제값, \hat{x}_i 는 i 번째에서의 예측값을 나타낸다.

MAE는 이상치에 민감하게 반응할 수 있으므로 이상치 항목이 많거나 중요한 경우에는 다른 평가 지표를 함께 고려할 필요가 있다. 또한 큰 값을 가진 일부 케이스가 MAE를 상당히 높게 만드는 한계가 있다. 이러한 특성을 고려하여 본 연구에서는 MAE의 절대적인 수치보다는 각 분석기법을 비교하기 위한 수단으로 활용하고자 한다.

5.3 성능평가

본 연구에서는 각각의 분석항목 단위로 Table 10에 언급한 8개 알고리즘(AI 기술 2개, 통계적 추론 3개, 기술통계 3개)을 적용하여 예측 모델을 개발하고, MAE를 사용하여 각 알고리즘의 성능을 평가한다. 다만, 앞서 군집화 과정을 거쳐 도출한 1,452개 분석항목 중에서 월 단위별 총 소비인시 정보가 6개 이상 존재하는 610개에 대해서만 모델 개발을 진행한다.

Table 11은 적용 알고리즘별 성능평가 결과를 나타낸 것이다. Prophet의 경우는 610개 분석항목 중 211개 항목에서만 예측이 수행되었으며, 나머지 항목에서는 데이터의 부족 또는 불규칙성 등의 문제로 모델이 개발되지 않았다. 나머지 알고리즘의 경우에는 모든 항목에서 예측이 진행되었다. MAE 결과를 살펴보면, 이동평균법의 성능이 가장 좋은 것으로 파악되며, 전반적으로 기술통계에 속한 알고리즘의 성능이 상대적으로 우수해 보인다.

Table 12는 월 단위별 총 소비인시 정보가 20개 이상 존재하는 분석항목 243개에 대한 성능평가 결과이다. MAE 결과를 살펴보면, 지수평활법의 성능이 가장 좋은 것으로 파악되며, Table 11과 같이 인공지능 기술에 속한 알고리즘의 성능이 가장 낮은 것으로 파악된다.

Table 13은 분석항목별(610개)로 정비인시 예측 결과가 가장 우수한 기법을 나타내고 있다. 분석항목 중 약 27%에서 인공지능 기술이 채택되었으며, 통계적 추론의 26%, 기술통계의 47%를 차지했다.

Tables 11-13에 나타난 바와 같이 상위기술로 분류되는 인공지능 기술이 기술통계, 통계적 추론 등 전통적 기술과 비교하여 성능이 우수하게 나오지 못한 이유는 크게 2가지로 해석된다.

첫 번째는 데이터의 규모이다. 대체로 LSTM과 같은 인공지능 기술은 많은 데이터가 있어야 모델이 데이터로부터 패턴을 학습하여 일반화할 수 있다고 한다. 이런 관점에서 본 연구에서 개발한 인공지능 모델은 학습이 제대로 수행되지 못했을 가능성이 있다.

두 번째는 데이터 특성이다. 데이터 특성은 2가지 관점으로 생각할 수 있다. 하나는 데이터가 단순한 추세나 주기성이 뚜렷한 경우이고, 다른 하나는 데이터의 추세나 주기성이 없는 경우이다. 데이터가 단순한 추세나 주기성이 뚜렷한 경우는 이동평균법 등과 같

Table 11. 알고리즘 성능평가 결과 1

분류	분석알고리즘/기법	적용항목 수	MAE
인공지능	Prophet	211	158.2
	LSTM	610	54.5
통계적 추론	최소자승법	610	49.4
	지수평활법	610	41.9
	추세보정지수평활법	610	48.9
기술통계	산술평균법	610	45.8
	이동평균법	610	39.5
	가중이동평균법	610	40.4

Table 12. 알고리즘 성능평가 결과 2

분류	분석알고리즘/기법	적용항목 수	MAE
인공지능	Prophet	134	141.4
	LSTM	243	94.5
통계적 추론	최소자승법	243	83.9
	지수평활법	243	65.4
	추세보정지수평활법	243	66.2
기술통계	산술평균법	243	81.8
	이동평균법	243	68.9
	가중이동평균법	243	82.7

Table 13. 알고리즘 성능평가 결과 3

분류	분석알고리즘/기법	채택항목 수	비중 (%)
인공지능	Prophet	21	3.4
	LSTM	142	23.3
통계적 추론	최소자승법	59	9.7
	지수평활법	56	9.2
	추세보정지수평활법	45	7.4
기술통계	산술평균법	136	22.3
	이동평균법	114	18.7
	가중이동평균법	37	6.1
계		610	100.0

은 기법이 잘 동작할 수 있으며, 이상치나 결측치 등의 영향으로 데이터의 추세나 주기성이 식별되지 않는 경우에도 산술평균과 같은 단순한 기법이 더 효과적일 수 있다.

효과적인 시계열 모델 개발을 위해서는 데이터가 충분한 규모로 확보되어야 하고, 과거 데이터로부터 추세 또는 주기성 파악이 가능해야 하며, 입력 데이터의 연속성이 보장되어야 한다.

본 연구에서 분석항목별로 부족한 데이터를 확보하기 위하여 동형/유사장비별로 분석항목을 추가 군집화했음에도 불구하고, 월 단위별 총 소비인시 정보가 충분하지 못한 것으로 해석된다. 그리고 사용한 총 소비인시 정보를 월 단위로 변환했음에도 다수의 부분에서 입력 데이터의 연속성이 보장되지 못했다.

따라서 현재의 함정선체 정비지원 업무 프로세스가 정상이라면, 정비인시 예측 방법으로는 시계열 기법보다는 인과관계에 기반한 회귀분석 방법 적용을 고려할 필요가 있다.

반면 업무 프로세스와 달리 데이터 입력관리에 왜곡이 포함되어 있다면, 분석모델 개발에 앞서 데이터 수집 및 품질관리체계 개선이 필요하다.

6. 결론

해군정비창은 정비데이터를 활용하여 선제적인 함정 정비지원을 준비하고자 한다. 본 연구에서는 이 목표를 달성하기 위하여 관련 업무 프로세스 이해를 통해 분석주제를 정의하고, 모델 개발에 필요한 데이터를 식별하였다. 이후 데이터 품질 확보를 위한 데이터 전처리 과정을 통해 분석용 데이터셋을 준비하였다. 탐색적 자료 분석을 통해 입력 데이터를 정의하고 입력 데이터의 수준 평가를 활용 가능한 데이터의 규모를 중점적으로 확인하였으며, 부족한 데이터를 확보하기 위하여 동일장비별 표준공사 대상으로 발행된 공사명령서(정비지시서)를 동형/유사장비별로 추가 군집화하였다. 그리고 원천 데이터를 분석에 유용한 형태로 변환규칙을 활용하여 데이터 변환을 시행하였다.

함정선체 정비인시 예측 모델은 통계적 추론, 기술통계 등의 전통적인 기술과 시계열의 인공지능 기술을 병행 활용하였다. 각 알고리즘의 성능은 MAE를 최소로 하는 관점에서 평가하였다.

각 알고리즘의 성능평가 결과, 군집화된 표준공사 중 인공지능 기술이 약 27%, 통계적 추론이 26%, 기술통계가 47%를 차지했다.

일반적으로 상위기술로 분류되는 인공지능 기술이 우수한 성능을 제공한다. 그러나 본 연구에서는 전통적 기술들과 비교하여 우수한 성능을 제공하지 못하였다. 그 사유는 데이터 부족과 데이터의 추세 또는 주기성 파악이 불가, 입력 데이터의 연속성 미흡 등으로 해석된다. 이를 통해 분석모델이 우수한 성능을 제공하기 위해서는 충분한 데이터 확보와 데이터 특성을 고려한 최적의 분석기술을 적용해야 한다는 결론을 도출하였다.

그리고 해군정비창이 목표를 달성하기 위해서는 단기적으로는 현재의 정비데이터 수준에 부합하는 최적의 분석 기술을 활용하여 정비 소요를 예측하고, 부족한 신뢰성은 실무자의 정성적 판단으로 보완하여 업무에 활용하도록 해야 한다. 중·장기적으로는 업무 프로세스 진단을 통해 업무 수행 중 생성되는 데이터의 왜곡 발생 여부를 확인한다. 그리고 정상이라면 함정선체 정비인시 예측 방법으로 인과관계에 기반한 회귀분석 방법 적용을 고려하고, 비정상일 경우는 분석 모델 개발에 앞서 데이터 수집 및 품질관리체계 개선을 수행할 필요가 있다.

그리고 이와 같은 부분이 보완되기 위해서는 관리자의 데이터에 대한 인식이 개선되어야 할 것으로 사료된다. 그리고 예측 모델에서 비표준공사를 적용하지 못한 점은 본 연구에서 한계로 남아 있다.

참고문헌

- [1] 국방부, 「군수빅데이터 기반 구축 추진계획」, 2020.
- [2] 국방부, 「4차 산업혁명 스마트 국방혁신을 위한 스마트팩토리 구축추진 계획」, 2020.
- [3] 국방부, 「국방 모바일 환경 및 서비스 구축·운영에 관한 지시」, 2020.
- [4] 해군본부, 「SMART Navy 종합발전계획」, 2020.
- [5] 국방부, 산업부, 해경청, 「국가연구개발사업 협약서 (21-CM-TN-12) IoT기반 함정 정비 통합관제플랫폼 개발」, 2021.
- [6] 해군군수사, 「군수업무 빅데이터 기술 적용 연구보고서」, 2015. 12.
- [7] 해군함정기술연구소, 「장비 고장예측 시스템 개발방안 기초연구 보고서」. 2016.6.
- [8] 해군군수사, 「군수업무 빅데이터 분석활용체계 구축 시범사업 연구보고서」, 2017. 10.
- [9] 해군함정기술연구소, 「통합기관제어감시장치 고장예측 시스템 구축 연구보고서」. 2018.12.
- [10] 국방부, 국방 빅데이터 인공지능 공통 분석서비스 사업 결과 보고서, 2021.11
- [11] C. W. Hong, K. Hur, "Electric Power Load Forecasting of DP ships by Using CNN-RNN based DNN," Journal of the KNST, Vol.4, No.2, pp.121-126. 2021.
- [12] J. W. Choi, J. M. Jang, S. H. Park, "Spare Parts Prediction Using Digital Twins of Naval Systems," Journal of the KNST, Vol. 6, No. 1, pp. 84-91. 2023.
- [13] 해군 정비관리규정
- [14] Sean J. Taylor & Benjamin Letham, "Forecasting at Scale," The American Statistician, Vol. 72(1), pp. 37-45. 2018.
- [15] Sepp Hochreiter, Jürgen Schmidhuber, "Long Short-term Memory," Neural Computation, Vol. 9(8), pp. 1735-1780. 1997.
- [16] Christopher Olah, "Understanding LSTM Networks," 2015.