

Research on the application of Machine Learning to threat assessment of combat systems

Seung-Joon Lee*

*Engineer, Naval R&D Center, Hanwha Systems, Gumi, Korea

[Abstract]

This paper presents a method for predicting the threat index of combat systems using Gradient Boosting Regressors and Support Vector Regressors among machine learning models. Currently, combat systems are software that emphasizes safety and reliability, so the application of AI technology that is not guaranteed to be reliable is restricted by policy, and as a result, the electrified domestic combat systems are not equipped with AI technology. However, in order to respond to the policy direction of the Ministry of National Defense, which aims to electrify AI, we conducted a study to secure the basic technology required for the application of machine learning in combat systems. After collecting the data required for threat index evaluation, the study determined the prediction accuracy of the trained model by processing and refining the data, selecting the machine learning model, and selecting the optimal hyper-parameters. As a result, the model score for the test data was over 99 points, confirming the applicability of machine learning models to combat systems.

▶ **Key words:** Artificial Intelligence, Machine Learning, Gradient Boosting Regressor, Support Vector Regressor, Combat Systems

[요 약]

본 논문에서는 전투체계 위협지수를 머신러닝 모델 중 Gradient Boosting Regressor, Support Vector Regressor를 통해 예측하는 방법을 제시한다. 현재 전투체계는 안전성과 신뢰성이 증시되는 소프트웨어이므로 신뢰성이 보장되지 않은 AI 기술의 적용을 정책상 제한하고 있으며, 이로 인하여 전력화된 국내 전투체계는 AI 기술을 탑재하고 있지 않다. 하지만 AI의 전력화를 목표로 하는 국방부의 정책 방향에 대응하기 위하여, 전투체계의 머신러닝 적용에 필요한 기반 기술을 확보하기 위한 연구를 실시하였다. 이 연구는 위협지수 평가에 필요한 데이터를 수집한 뒤 데이터 가공 및 정제, 머신러닝 모델 선정 및 최적의 하이퍼 파라미터를 선정하여 학습된 모델의 예측 정확도를 판단하였다. 그 결과 테스트 데이터에 대한 모델 점수가 99점 이상으로 도출되었으며 전투체계에 머신러닝 모델의 적용 가능성을 확인하였다.

▶ **주제어:** 인공지능, 머신러닝, Gradient Boosting Regressor, Support Vector Regressor, 전투체계

-
- First Author: Seung-Joon Lee, Corresponding Author: Seung-Joon Lee
 - *Seung-Joon Lee (sj0543.lee@hanwha.com), Naval R&D Center, Hanwha Systems
 - Received: 2023. 06. 22, Revised: 2023. 07. 19, Accepted: 2023. 07. 21.

I. Introduction

현재 국방부는 급격하게 변화하고 있는 미래 안보환경에 대비한 주변국의 군사혁신 추진 동향을 인식하고, AI·무인전투체계와 같은 첨단과학 기술을 신속히 군에 적용하여, 미래를 대비한 국방정책·전략 발전에 집중하기 위해 ‘미래국방혁신구상’을 추진하고 있다. 국방 연구개발진흥정책을 수립하며, 국방부 차원의 미래 신기술 육성 분야 선정 및 지원방안을 제시하고 AI 기반 무인체계의 신속한 획득 및 전력화를 추진하는 등 적극적인 움직임을 펼치고 있다[1]. ‘국방부 국방정책방향’ 세부 추진과제 이행방안 업무보고의 ‘국방혁신 4.0을 통한 첨단과학기술군 육성’ 측면에서, AI 기술 수준과 발전단계를 고려하여 Fig.1 과 같이 ‘국방AI 발전모델’을 정립하였으며, 이에 따라 우리 군에 대한 AI 기술 적용을 단계적으로 확대해 나갈 것이라고 밝혔다. 이에 따라 함정 전투관리체계(Combat Management Systems, 이하 전투체계)의 AI 기술 습득에 대한 방향성 또한 확고해지고 있는 상황이다[2].



Fig. 1. Defense AI three-stage evolutionary model[3]

전투체계의 교전과정은 Fig. 2와 같이 표적식별, 위협평가, 무장할당, 명중평가로 이루어진다. 과정별 AI 적용이 필요하나 이 중 위협평가에 대한 AI 적용이 데이터 수집, 모델의 선택, 운용자 직접 개입 등의 측면에서 AI 적용의 첫 단계로 적절하다 판단하였다. 현재 전투체계에서의 위협평가 기능은 현재 운용 중인 방법으로도 전투 상황에서의 유효성은 검증이 되었으나 국방정책의 흐름에 따라 무인전투체계의 신속한 전력화를 위한 방법에 대한 연구개발은 필요한 상황이다.



Fig. 2. Warfare Flow

본 논문에서는 전투체계 도메인에서의 위협평가 기능에 AI 적용에 대한 가능성을 확인하고, AI 중 머신러닝 예측 기법 중 Gradient Boosting Regressor(이하 GBR), Support Vector Regression(이하 SVR)으로 지도 학습을 해보며, 각 모델의 성능을 측정하였다.

본 논문의 구성은 다음과 같이 이루어진다. 2장에서는 전투체계, 머신러닝, GBR, SVR 모델에 대한 개념 및 scikit-learn API에 대해 설명한다. 3장에서는 연구 내용에 대한 데이터 타입 및 정규화 적용 내용, 하이퍼 파라미터 설정에 대해 설명한다. 4장에서는 GBR, SVR 학습 결과 및 평가에 대해 설명하고 마지막 5장에서는 전투체계에 머신러닝 적용 가능성 판단 및 추후 연구과제에 대한 결과를 설명한다.

II. Preliminaries

1. Related works

1.1 Machine Learning in Combat Management Systems Domain

함정 전투체계는 공중, 수상 및 수중 적의 위협을 조기에 탐지 및 추적하는 센서와 자함 및 구역 방어를 위한 무장 등을 통제 및 관리하는 함정 전투력의 핵심이 되는 체계이다[4]. 레이더, 전자광학체계 등 함정에 탑재된 센서장비를 통해 표적을 분석하고 함포 등의 함정에 탑재된 무장체계에 교전 명령을 내림으로써 대함, 대잠전에서 최적의 공격 및 방어 수단을 실시간으로 제공한다. 함정 전투체계는 Fig. 3과 같이 센서체계, 항해체계, 무장체계 그리고 통신체계로 구성되며, 실시간으로 처리된 전술정보를 관리하고 전술상황 감시, 센서통제, 무장할당 및 발사통제 등을 포함한 각종 전술대응을 용이하도록 지원한다[5,6]. 함정의 전투를 담당하는 교전 모듈의 위협평가 기능은 센서를 통해 수집된 표적에 대해 위협도를 평가하여 교전의 우선순위를 판별하여주는 기능을 수행한다.



Fig. 3. Naval Combat Management Systems[7]

함정 전투체계는 현재 대부분의 기능이 알고리즘에 의한 자동화가 되어 있으며, 운용자의 판단 및 수동 운용이 필요한 부분 또한 존재하나 현재 전투체계 도메인에서 실제로 AI 적용 사례는 확인된 바가 없으며, 실제로 적용된 사례가 없는 근거를 제시하는 연구 결과 또한 존재한다. 해당 연구에서 말하길 오늘날의 해군 작전은 기술이 기하급수적으로 발전하고 위협 환경이 더욱 복잡해짐에 따라 어려운 의사 결정에 직면하는 경우가 많으며, AI 시스템 개발 및 운영에 필요한 데이터의 요구 충족 제한, 이에 따른 새롭고 비결정적인 시스템을 엔지니어링 하는 것, 적들이 제기하는 도전에 직면하는 것 등의 해군 작전을 위한 AI에 대한 설계가 힘든 부분이 있어 AI 시스템에 대한 심층적인 이해와 투자 기술 발전기간에 대한 현실적인 기대치가 필요하다고 하며 공감하는 바이다[15].

전투체계 AI 적용에 앞서 전투체계의 교전 과정을 분류하여 과정별 AI의 모델의 적합성을 판단한 연구도 존재한다[16]. 해당 연구에서 말하길 본 연구에서 연구하는 부분인 표적의 위협도를 판단함에 있어서는 연관 모델을 제시하였으나 본 연구에서는 표적의 위협지수 평가라는 명확한 결과값이 존재하므로 지도학습 중 회귀를 사용하여 연구하였다. 또한 전투체계에 AI 적용에 대한 구조적인 관점의 연구도 존재한다[17,18]. 하지만 본 연구는 구조적인 관점의 연구보다 실제 데이터셋을 이용한 모델 학습에 더욱 중점을 두었다.

1.2 Combat Management Systems Threat Assessment

위협평가 기능은 현재 국내 개발 전투체계에서는 계산식이 있는 알고리즘을 통해 위협지수를 계산하게 되어 있다. 이 알고리즘은 표적의 거리, 속도와 가중치를 위한 상수를 활용하여 만들어진 계산식이다. 하지만 이 부분에서 센서를 통해 수집한 표적의 모든 정보를 판단하지 않고 있으며, 함정 전투체계를 운용할 때 주변 상황 및 표적 정보를 운용자가 판단하여 추가적인 가중치를 입력해주고 있다.

- 위협지수 = (속도, 거리, 가중치 상수에 의한 계산식) + (운용자 입력 가중치)

위협평가에 사용되는 데이터들은 전투체계의 센서들을 통해 수집할 수 있는 정형화된 데이터들이며, 추가로 운용자의 판단으로 정해지는 결과 데이터가 존재하여 머신러닝의 적용에 용이하며 적합성 또한 만족한다고 판단하였다. 머신러닝 중 지도학습의 경우 입력 데이터, 이에 따른 결과 데이터 모두를 학습 모델에 전달해줌으로써 입력된 데이터를 기반으로 결과값을 예측할 수 있는 모델이 있다. 앞서 말한 것과 같이 입력 데이터의 수집이 가능하고 위협

평가에서 위협지수라는 결과값을 얻기 위해 머신러닝의 지도학습 중 회귀 모델을 사용하여 AI 적용을 기대할 수 있다고 판단한다.

1.3 Machine Learning

머신러닝은 인공지능 분야에서 데이터를 분석하고, 패턴을 찾아내며, 그 패턴을 사용하여 예측이나 분류 등의 작업을 수행하는 알고리즘의 집합이다. 기존의 프로그래밍 방식과 달리, 머신러닝은 데이터를 기반으로 자동으로 학습하고, 경험을 통해 스스로 개선해나가는 능력을 가지고 있다. Fig. 4와 같이 크게 지도학습, 비지도학습, 강화학습으로 분류된다. 지도학습은 입력과 출력 데이터가 존재하여 이를 이용해 모델을 학습시키고 예측하는 방식이며, 비지도학습은 출력 데이터가 없이 입력 데이터만을 이용해 데이터의 패턴을 찾아내는 방식이다. 강화학습은 환경과 상호작용하며 행동을 통해 보상을 얻어가며 학습하는 방식이다.



Fig. 4. Machine Learning Classification

1.4 GBR(Gradient Boosting Regressor)

GBR은 부스팅 모델 중 하나로, 부스팅 모델은 결정 트리를 사용하는 약한 학습기를 결합하여 강한 학습기를 만드는 앙상블 기법 중 하나이다. GBR은 첫 번째 결정 트리를 학습시키고, 이를 기반으로 예측값을 계산한다. 이전 결정 트리의 오차와 현재 모델의 예측값의 차이를 계산한다. 계산한 오차를 예측 변수로 추가하여 새로운 결정 트리를 학습시킨다. 앞의 과정을 반복하여 모델을 계속 보완한다. GBR은 다른 부스팅 모델에 비해 속도가 느리고, 하이퍼 파라미터 튜닝이 어려울 수 있다. 또한, GBR은 과적합에 민감하기 때문에 적절한 하이퍼 파라미터 값 설정과 데이터 전처리가 필요하다. 그러나 GBR은 예측 성능이 높고 과적합을 해결하기 위한 다양한 방법이 있으며, 결정 트리 알고리즘을 기반으로 하기 때문에, 결정 트리 알고리즘의 장점인 해석력이 뛰어나다는 점도 있다. 이를 통해 모델이 어떤 결정을 내리는지 이해하고, 결과를 설명하는 데 도움이 된다.

1.5 SVR(Support Vector Regressor)

SVR은 Support Vector Machine(SVM)의 변형 알고리즘이며, 지도학습 알고리즘 중 하나로 분류 문제를 푸는 SVM과 달리 예측, 회귀 문제를 푸는데 사용된다. 데이터 포인트들을 고차원 공간으로 매핑하고, 데이터 포인트들과 가까운 마진 내에서 선형 회귀를 수행한다. 이때 마진의 폭은 하이퍼 파라미터로 설정되며, SVR은 이 마진 내에서 오차를 최소화하는 초평면을 찾는다. 이 초평면은 회귀식을 나타내며, 새로운 데이터 포인트가 들어올 때 이 식을 사용하여 예측한다. 즉, 커널 함수를 사용하여 고차원 공간으로 데이터를 매핑하며, 이러한 커널 함수는 데이터의 특징들을 유지하면서, 고차원 공간으로 매핑한다. 적은 수의 훈련 데이터에 대해서도 높은 예측 정확도를 보이며, 이상치에 대한 강인함을 가지고 있다. 하지만 데이터가 많은 경우 모델 구축에 시간이 많이 걸릴 수 있으며, 하이퍼 파라미터 튜닝이 필요하다.

1.6 Scikit Learn

scikit-learn은 머신러닝 및 데이터 분석 작업에 가장 널리 사용되는 라이브러리 중 하나이며, 다양한 머신 러닝 모델을 제공하며 데이터 분석, 데이터 마이닝, 자연어 처리 등 다양한 분야에서 사용된다. 그리고 개발자들이 사용하기 쉽게 API를 제공하며, 머신러닝 및 데이터 분석에 필요한 다른 라이브러리와 연동, 확장 및 사용이 편리하게 구성되어 있다. scikit-learn에서 제공하는 여러 모델을 활용하여 앞선 연구로 커피 프랜차이즈 실패 요인 분석[8], 해체 폐기물 발생량 예측[9], 농산물 가격 예측[10], 타이타닉 생존자 예측[11], 기상정보를 활용한 전력 수요 예측 [12], 이산화탄소 농도 데이터를 활용한 실내 재실 인원 추정[13] 등 여러 회귀 모델들을 사용함을 확인하였다. 본 연구에서 사용할 예정인 GBR, SVR 모델 또한 scikit-learn에서 제공하여 선택하게 되었다. 본 연구에서 사용한 scikit-learn 라이브러리 내 사용한 API들은 Estimator, Predictor, Model Selection, Metrics 이다. 모델의 학습에는 Estimator의 fit 함수를 사용하였다. Predictor는 학습 데이터를 기반으로 모델의 매개변수를 최적화하여 학습을 수행하는 데 도움이 되었다. 그리고 입력 데이터에 대한 결과를 예측하기 위해 Predictor의 predict 함수를 활용하였다. Predictor는 학습된 매개변수를 활용하여 입력 데이터에 대한 목표값을 예측하는 역할을 수행하였다. 또한, 모델 선택과 검증을 위해 Model Selection에서 제공하는 메소드를 활용하였으며, 이를 통해 다양한 모델을 평가하고 적절한 하이퍼 파라미터를 선택하는 데 도움이

되었다. 마지막으로, Metrics에서 제공하는 모델 점수(R2 Score)와 Root Mean Square Error(RMSE) 함수를 사용하여 학습된 모델의 성능을 수치적으로 측정하였고, 이러한 지표들은 최종 학습된 모델의 결과를 수치로 평가하는데 활용하였다.

III. The Proposed Scheme

본 연구는 Fig. 5와 같은 절차로 진행하였다.

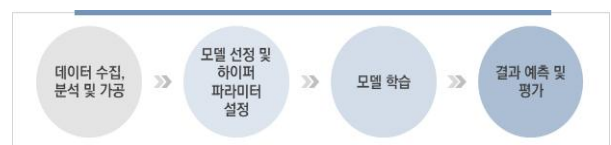


Fig. 5. Procedure for conducting a study

1. Data collection, analysis, and processing

실제 해군에서 사용하는 전투체계의 실제 데이터 수집은 보안상 제한되므로 위협도를 평가하기 위해 실제 전투체계 교전 상황을 재현하여 전투체계에서 발생하는 데이터들을 약 10만개 정도 수집하였다. 그리고 학습 결과의 이상치가 발생하는 오차를 줄이기 위해 속성별 편향적인 데이터를 추가로 약 26만개 정도 생성하여 총 36만개의 데이터로 학습 데이터셋과 시험 데이터셋을 구성하였다.

본 데이터셋은 6개의 속성으로 이루어져 있으며, 위협 지수를 예측한다. 데이터셋은 Table. 1에서 보는 바와 같이 7개의 속성으로 이루어져 있으며, 이 중 threat_point는 결과 데이터이다. Fig. 6은 테스트 데이터의 샘플이다.

	dis	vel	WF_ability	inner_WF_available_range	under	over	threat_point
42123	91179.22576568318	-1612.8844531695543	0.5		1	0	999
42124	129964.07022114073	808.7345522589347	0.5		0	1	0
42125	207926.56952865323	-2531.568720706523	0.5		0	0	1
42126	50016.399331992936	-526.8585781577768	0.5		1	0	1
42127	123727.57301615398	224.4634962564659	0.5		0	1	0
42128	170009.40154603467	-3109.937003065687	0.5		0	0	1
42129	133664.00067882976	2023.6531427700702	0.5		0	0	0
42130	127481.15526967793	705.4792043250081	0.5		0	1	0
42131	17047.630674966873	228.65722823609997	0.5		1	0	0

Fig. 6. Datasets Sample

Table 1. Features of datasets

feature	Description
dis	Distance of the target relative to this ship - unit : m
vel	Speed of the target relative to this ship - unit : kts
WF_ability	Target's armed capabilities - range : 0.00 ~ 1.00
inner_WF_available_range	Whether or not this ship is within target armament range - range : 0(=false), 1(=true)
under	Whether the threat point is the minimum or not - range : 0(=false), 1(=true)
over	Whether the threat point is the maximum or not - range : 0(=false), 1(=true)
threat_point	Threat Index of the target relative to this ship - range : 1~999

‘dis’, ‘vel’, ‘WF_ability’, ‘inner_WF_available_range’ 속성들은 전투체계에서 수집하여 가공 및 정제한 속성이며, 거리, 속도 등은 자함을 기준으로 한 표적의 상대적인 값이다. 실제로 전투체계를 통해 수집되는 정보는 ‘dis’, ‘vel’과 플랫폼의 종류(ex, Mig-21, LYNX, B747, F-16, P3-C) 뿐이나 학습을 위해 추가적인 2개의 속성을 생성하였다. 아래 Fig. 7을 참고한다.

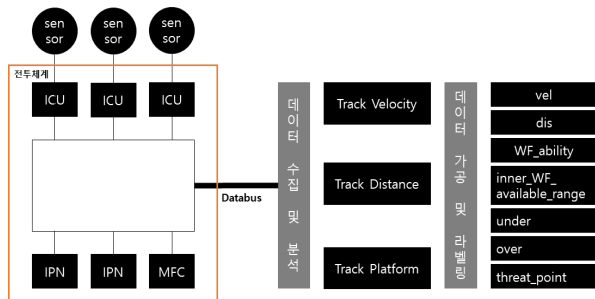


Fig. 7. Data collection, analysis and processing

‘dis’ 와 ‘vel’은 표적의 속도와 거리는 나타내는 속성이며, ‘under’ 과 ‘over’ 속성들은 전투체계에서 표시하는 위협지수의 값이 1~999이므로 실질적으로 판단하였을 때 위협지수가 1미만 혹은 999 초과일 경우에 1로 세팅된다. 즉 ‘under’ 속성의 경우 1미만 일 경우 1, 아닐 경우 0, ‘over’ 속성의 경우 999초과 일 경우 1, 아닐 경우 0 으로 세팅된다. 학습하는 모델에서 예측하는 위협지수의 값이 1~999로 예측될 수 있게 ‘under’, ‘over’ 속성을 추가하였으며, 해당 속성이 없을 경우의 예측 결과는 1~999 범위 보다 초과하는 값들이 많이 예측됨을 확인하였다. 특히 결

정 트리를 사용하는 GBR 모델에서 해당 속성이 중요하게 쓰일 것으로 판단하였다. ‘WF_ability’, ‘inner_WF_available_range’ 속성은 표적이 가진 무장 능력 및 사거리에 대한 필드이다. 해당 표적 플랫폼 정보는 전투체계를 통해 수집 가능하며, 수집된 플랫폼의 종류에 따라 ‘WF_ability’ 속성에 0~1 사이의 값으로 부여되며 백분율로 사용된다. 자함이 플랫폼의 무장 사거리 안에 위치 할 경우 ‘inner_WF_available_range’의 값은 1으로 사용되며, 아닐 경우 0으로 사용된다.

학습 데이터셋 속성별 아이템 개수는 Table. 2와 같다. 결과 속성과 ‘WF_ability’ 속성의 데이터셋이 균등하게 분포할 수 있도록 하였으며, 실제 데이터에서 무장 사거리 밖인 상황이 많아 무장 사거리 밖인 데이터를 다수 수집 및 생성하였다. 데이터를 속성별로 균등하게 생성한 이유는 모든 상황에 대해 모델이 학습하기 위함이며 균등하게 생성하지 않았을 경우 예측률이 떨어짐을 확인하였다. 추가로 본 연구에서는 R2 Score에 100 곱하여 R2 Score를 표현하였으며, R2 Score는 100이 최댓값이며, 높을수록 높은 예측률을 가진다.

- 균등 데이터 모델 점수(R2 Score) : 약 99.9
- 균등하지 않은 데이터 모델 점수(R2 Score) : 약 97.5

Table 2. The Number of training dataset

Feature	The Number
By inner_WF_available_range property	
inner_WF_available_range == 1	50008
inner_WF_available_range == 0	298593
By WF_ability property	
WF_ability == 1	69868
WF_ability == 0.9	69655
WF_ability == 0.5	69872
WF_ability == 0.2	69565
WF_ability == 0.0,1	69641
By threat_point property	
0 <= threat_point < 200	69782
200 <= threat_point < 400	70540
400 <= threat_point < 600	63194
600 <= threat_point < 800	64667
999 <= threat_point < 999	80418

2. Selecting a model, setting hyperparameters

머신 러닝의 경우 학습 데이터에 따라 최적의 모델을 다룰 수 있으며, 선택된 모델 중에서도 하이퍼 파라미터의 설정에 따라 그 성능을 달라질 수 있다. 이에 최적의 모델 및 하이퍼 파라미터를 확인하였다.

GridSearchCV API를 사용하여 GBR 모델의 하이퍼 파라미터 최적값을 확인하였다. 22개의 하이퍼 파라미터 중

4개의 하이퍼 파라미터를 사용하였으며, 그 값은 Table. 3 과 같다.

Table 3. GBR Optimal Hyper Parameters

Feature	Input List	Optimal Value
learning_rate	[0.01, 0.02, 0.03, 0.04]	0.01
max_depth	[4, 6, 8, 10]	10
n_estimators	[100, 500, 1000, 1500]	1500
subsample	[0.9, 0.5, 0.2, 0.1]	0.2

각 하이퍼 파라미터 속성별 내용은 아래와 같다.

- learning_rate: 각 트리의 기여도를 제어하는 매개변수이다. 이 값이 작을수록 모델이 안정적이지만, 더 많은 트리가 필요하다.

- max_depth: 트리의 최대 깊이이다. 이 값이 커질수록 모델이 더 복잡해지며, 과적합 가능성이 커질 수 있으나 적절한 값은 데이터에 따라 다를 수 있다.

- n_estimators: 부스팅 단계 수, 즉 트리의 개수이다. 이 값이 클수록 모델이 복잡해지며, 과적합 가능성이 커진다.

- subsample: 각 트리에서 사용할 샘플의 비율이다. 이 값이 작으면 더 안정적인 모델이 된다.

하이퍼 파라미터별 영향성 확인을 위한 결과는 Table. 4와 같다. learning_rate가 고정일 때 (max_depth * n_estimators)의 값이 클수록 즉, 트리가 많을수록 모델의 정확도가 높아짐을 확인하였으며, learning_rate가 높을수록 트리의 개수에 영향을 적게 받음을 확인하였고, subsample은 현재 데이터에서 크게 영향을 주는 하이퍼 파라미터가 아님을 확인하였다. 현재 위협평가 데이터와 같은 선형적이지 않은 데이터의 경우 트리의 개수에 영향을 많이 받음을 확인하였다.

Table 4. GBR fit results

hyperparameters				R2 Score
learning_rate	max_depth	n_estimators	subsample	
0.01	10	1500	0.1	99.9449
0.01	10	1000	0.1	99.9442
0.01	4	500	0.1	97.6394
0.01	10	100	0.1	86.0545
0.04	10	1500	0.1	99.9365
0.04	10	1000	0.1	99.9388
0.04	4	500	0.1	99.6356
0.04	10	100	0.1	99.8816
0.01	4	100	0.9	75.0275
0.01	4	100	0.5	75.0287
0.01	4	100	0.2	75.0018
0.01	4	100	0.1	75.0450

GridSearchCV API를 사용하여 SVR 모델의 하이퍼 파라미터 최적값을 확인하였다. 22개의 하이퍼 파라미터 중 3개의 하이퍼 파라미터를 사용하였으며, 그 값은 Table. 5 과 같다.

Table 5. SVR Optimal Hyper Parameters

Feature	Input List	Optimal Value
kernel	['rbf', 'poly']	rbf
C	[0.1, 1, 10]	10
gamma	[0.1, 1, 10]	0.1

각 하이퍼 파라미터 속성별 내용은 아래와 같다.

- kernel: SVR에 사용할 커널 함수이며, 'rbf', 'linear', 'poly', 'sigmoid' 중 하나를 선택할 수 있다. 기본값은 'rbf'이다. 'linear'는 현재 데이터에서의 특성을 생각해보았을 때 성능이 좋지 않을 것으로 판단하였고, 'sigmoid'는 적절하지 않다고 판단하여 GridSearchCV 테스트에서 제거하였다.

- C: 손실 함수의 규제 강도를 결정하는 매개변수이며, 작은 값은 많은 오차를 허용하고, 큰 값은 작은 오차를 허용한다. 모델의 복잡성과 성능 사이의 트레이드오프를 조절하는 중요한 요소이다.

- gamma: 'rbf', 'poly' 및 'sigmoid' 커널의 계수이며, gamma가 클수록 학습 데이터의 영향력이 커지고, 작을수록 작아진다.

하이퍼 파라미터별 학습 결과는 Table. 6과 같으며, 전체적인 점수가 낮음을 확인하였으며, 현재 데이터셋에선 하이퍼 파라미터별 영향성이 의미가 없다고 판단하였다.

Table 6. SVR fit results

kernel	hyperparameters		R2 Score
	C	gamma	
rbf	0.1	0.1	29.60
rbf	0.1	1	29.60
rbf	0.1	10	29.60
rbf	1	0.1	29.66
rbf	1	1	29.66
rbf	1	10	29.66
rbf	10	0.1	26.79
rbf	10	1	26.78
rbf	10	10	26.78
poly	0.1	0.1	17.50
poly	0.1	1	17.50
poly	0.1	10	17.50
poly	1	0.1	18.40
poly	1	1	18.40
poly	1	10	18.40
poly	10	0.1	15.30
poly	10	1	15.30
poly	10	10	15.30

3. Training models

학습에 scikit learn API를 사용하였으며, 2개 모델 (sklearn.ensemble.GradientBoostingRegressor, sklearn.svm.SVR)을 사용하여 성능을 비교하였다. 3.1에 작성된 것과 같이 데이터 수집, 생성 및 정제하여 학습시켰으며, 하이퍼 파라미터는 3.2에 표현된 것처럼 GridSearchCV를 통해 적절한 하이퍼 파라미터를 설정하였다. 모델 학습의 경우 본 연구를 위해 최적의 하이퍼 파라미터를 구하는 과정에서 동시에 진행되었다.

IV. Test Results and Evaluation

1. GBR

아래는 GBR 학습의 결과인 평균오차, 모델 점수이다.

- 평균 오차(RMSE) : 6.515831562802117
- 모델 점수(R2 Score) : 99.95946458178291

RMSE 방식으로 평균오차를 구했고, R2 Score로 모델의 성능을 확인하였다. 예측 값의 범위는 1~999이며 테스트 데이터셋 별 오차범위는 Table. 7과 같다. 70 이상의 오차가 발생하는 경우가 약 100건으로 확인되었으나 높은 성능 및 신뢰할 만한 결과가 도출되었다. 전투체계 위협지수를 예측하는 데 필요한 데이터셋은 GBR 모델을 활용하면 신뢰할 만한 사용성이 있음을 입증하였다.

Table 7. The Number of discriminatory data

Range	Count
Errors < 10	67290
10 <= Errors < 30	1687
30 <= Errors < 50	401
50 <= Errors < 70	238
70 <= Errors < 100	98
100 <= Errors < 200	6
200 <= Errors < 300	1
300 <= Errors	0

Fig. 8은 테스트 데이터와 예측 데이터의 Scatter Plot이다. 실제 데이터와 모델을 통해 학습된 데이터의 분포도 및 성능을 육안으로 확인할 수 있는 표이며, 파란색 세모 점과 주황색 동그라미 점이 많이 겹쳐 있을 때 예측 결과가 우수하다고 판단할 수 있다. 육안으로 볼 수 있듯이 GBR 모델의 학습 결과가 우수함을 확인하였으며, 본 연구에 사용한 데이터셋이 GBR 모델에 적절하다고 판단할 수 있다.

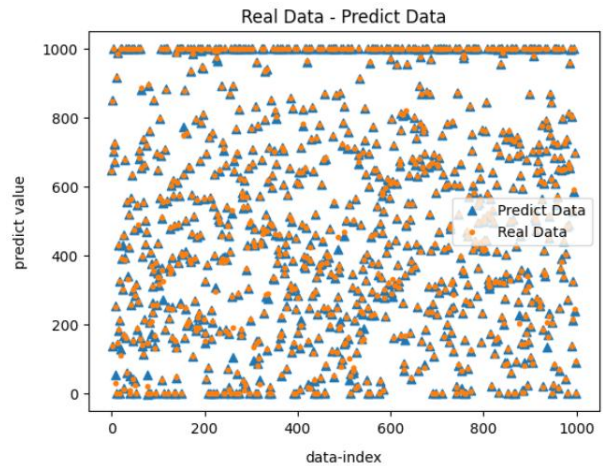


Fig. 8. GBR Scatter Plot

Fig. 9의 속성 중요도를 보면 'over' 속성의 값이 유독 중요도가 높은 것으로 확인된다. 테스트 데이터에 포함된 데이터의 예측 threat_point 값이 999가 넘어가는 데이터가 많아 'over' 속성이 중요한 것으로 예측되었다. 실제 전투체계에서도 자함 기준으로 위협도가 999 이하인 경우보다 999가 넘어가는 케이스가 많다. 그렇기에 위협도가 999가 넘어가는 데이터를 0, 1과 같은 boolean 타입의 속성을 추가하여 전처리하는 방식이 중요도가 높은 것을 확인하였다.

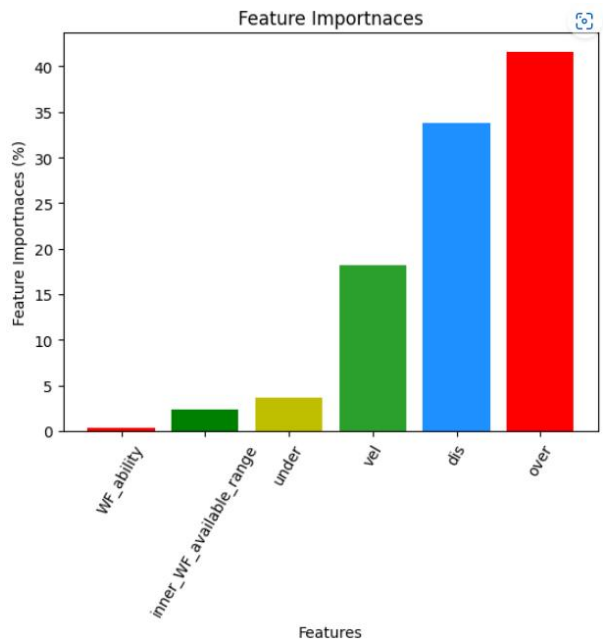


Fig. 9. GBR Feature Importances

2. SVR

아래는 SVR 학습의 결과인 하이퍼 파라미터 별 최고, 최저 모델 점수이다.

- 최저 모델 점수(R2 Score) : 15.30
- 최고 모델 점수(R2 Score) : 26.79

모델 점수는 100점이 최댓값이나 최고 모델 점수가 26.79인 SVR 모델 학습 결과는 현재 데이터에 SVR 모델이 적합하지 않음을 알 수 있었다. Fig. 10과 같이 GBR 모델과는 대조적으로 오차가 심각함을 알 수 있고, 해당 데이터 자체가 SVR 모델 및 설정된 하이퍼 파라미터에서는 학습이 제대로 이루어지지 않음을 확인하였다. 전투체계 표적 데이터셋의 경우 SVR 모델에서는 제대로된 학습이 이루어지지 않았으며, 해당 모델은 현재 데이터셋에 적합하지 않음을 확인하였다. 모델별로 적합한 데이터셋이 있음을 확인하였다.

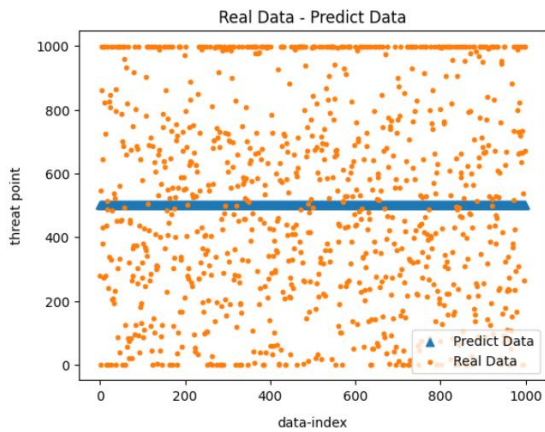


Fig. 10. SVR Scatter Plot

V. Conclusions

본 논문에서는 AI의 전력화를 목표로 하는 국방부의 정책 방향에 대응하기 위하여, 전투체계의 머신러닝 적용에 필요한 기반 기술을 확보하기 위한 연구를 실시하였다. 전투체계 도메인에서의 위협평가 기능에 머신러닝 예측 기법 중 GBR, SVR으로 지도 학습을 해보며 모델의 성능을 측정하였다.

GBR로 학습된 모델의 R2 Score가 약 99.9점 및 평균 오차가 1% 미만인 결과를 확인하였다. 전투체계를 통해 수집된 데이터셋을 GBR 모델이 높은 성능으로 예측하였으므로 충분히 전투체계 위협평가 기능에 적용 가능한 수준으로 확인하였다. 반면에 SVR로 학습된 모델의 R2 Score가 약 29.7점인 결과를 확인하였다. 좀 더 다양한

하이퍼 파라미터의 수정 및 데이터 속성의 다양화, 전처리가 필요할 것으로 보이며, 결과만으론 현재 데이터셋이 해당 모델에 적합하지 않을 수도 있다는 예측이 가능하다. 결과적으로 현재 전투체계를 통해 수집할 수 있는 데이터로 적용할 수 있는 모델은 GBR 모델인 것으로 확인하였으며, 전투체계 내 다른 기능에서의 AI 적용 가능성을 확인하였다.

추후 전투체계에서 센서를 통해 수집할 수 있는 표적들 (ex 전투기, 헬기, 미사일 등) 대공 표적에 대한 움직임(기동 패턴)을 수집, 학습하여 대공 표적의 위협 여부를 연구할 예정이다. NASA Engineering Safety Center (NESC)에서 오픈소스로 제공하는 simupy flight API[14]를 이용하여 Fig. 11과 같이 기동 패턴을 모의해보았다. 실제로 수집할 수 없는 비행 패턴 정보에 대해서는 오픈소스를 이용하여 진행할 예정이다.

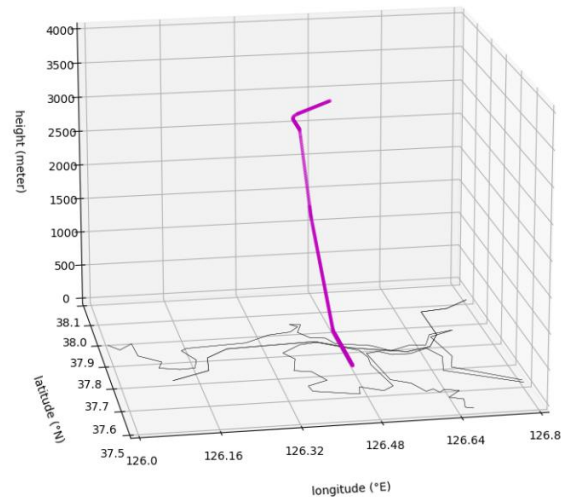


Fig. 11. Flight Movement Patterns

REFERENCES

- [1] Department of Defense Editorial Board, "Promoting 'Future Defense Innovation Initiative' to leapfrog advanced science and technology forces : Focus on rapid application of advanced technologies such as AI and unmanned systems, and development of future-proof defense policies and strategies". Defense & Technology, Vol. 511, pp. 13-15, Sep. 2021
- [2] Department of Defense Editorial Board, "Report on the implementation plan for the detailed implementation of the National Defense Policy Direction : Establishment of a three-axis system, strengthening of ROK-US joint exercises, and creation of a defense AI center". Defense & Technology, Vol. 522,

- pp.11-13, Aug. 2022.
- [3] Se-Yong Kim, Heung-Sun Park. "Recommendations for the development of the Defense AI Center to foster an AI-based science and technology powerhouse". *Defense & Technology*, Vol. 528, pp. 84-91. Feb. 2023
- [4] Jeong-Woo Son, "A Study on the Design Plan of Naval Combat System Software to Reduce Cost of Hardware Discontinuation Replacement", *Journal of the Korea Society of Computer and Information*, Vol. 28, No. 1, pp. 71-78, Jan. 2023. DOI: <https://doi.org/10.9708/jksci.2023.28.01.071>.
- [5] K. T. Kwon, K. P. Kim, and H. J. Choi, "Design of the Scalable Naval Combat System Software using Abstraction and Design Pattern" *Journal of The Korea Society of Computer and Information*, Vol. 24, No. 7, pp. 101-108, July 2019.
- [6] Cheol-Gyu Yi, Young-Gab Kim, "A Study on the Implementation of Software Security Test of Naval Ship Combat System" *Journal of The Korea Institute of Communications and Information Sciences*, Vol. 45, No. 03, pp. 628-637, March 2020. DOI: 10.7840/kics.2020.45.3.628
- [7] Naval Combat Management system, https://www.hanwhasystems.com/kr/business/defense/naval/combat_index.do
- [8] Yelyn Ahn, Hyunhee Lee, Sungmin Ryu, Seoyeon Kim, Minseo Park, "Analysis of Coffee Franchise Failure Factor using Machine Learning Algorithms", *Journal of the Korean Institute of Industrial Engineers*, Vol. 49, No. 1, pp. 37-45, Feb. 2023. DOI: <https://doi.org/10.7232/JKIIIE.2023.49.1.037>
- [9] Cha, Gi-Wook, Hong, Won-Hwa, "Development of Machine Learning Predictive Model for Forecasting Demolition Waste Generation", *Journal of the Architectural Institute of Korea*, Vol. 39, No. 2, pp. 229-235, Feb. 2023. DOI: <https://doi.org/10.9708/jksci.2022.27.05.029>
- [10] Jung-Ju Im, Tae-Wan Kim, Ji-Seoup Lim, Jun-Ho Kim, Tae-Yong Yoo, Won Joo Lee, "A Design and Implement of Efficient Agricultural Product Price Prediction Model", *Journal of The Korea Society of Computer and Information*, Vol. 27, No. 5, pp. 29-36, May. 2022. DOI: <https://doi.org/10.9708/jksci.2022.27.05.029>.
- [11] Tai-Sung Hur*, Suyoung Bang*, "A Comparative Analysis of the Pre-Processing in the Kaggle Titanic Competition", *A Comparative Analysis of the Pre-Processing in the Kaggle Titanic Competition*, Vol. 28 No. 3, pp. 17-24, March. 2023. DOI: <https://doi.org/10.9708/jksci.2023.28.03.017>.
- [12] Hyun-Jong Cha, Ah Reum Kang, "Machine Learning-Based Power Demand Forecasting Model Using Meteorological Information", *International JOURNAL OF CONTENTS*, Vol. 23, No. 2, pp. 117-124, Jan. 2023. DOI: <https://doi.org/10.5392/JKCA.2023.23.02.117>
- [13] Xiguan Liang, Owen Anderton, Sowoo Park, Doosam Song, "Indoor occupancy estimation from CO2 concentration data based on machine learning", *Korean Journal of Air-Conditioning and Refrigeration Engineering*. pp. 72-77, Nov. 2022.
- [14] NASA Engineering Safety Center (NESC), *nasa/simupy-flight*, *simupy-flight*, <https://github.com/nasa/simupy-flight>
- [15] Dr. Bonnie Johnson, "Challenges In Implementing Artificial Intelligence For Naval Warfare", *NAVAL ENGINEERS JOURNAL*, Vol. 135, No. 1, pp. 95-103, March. 2023. <https://bt.e-ditionsbyfry.com/publication/?m=63540&i=788511&p=96&ver=html5>
- [16] Dr. Bonnie Johnson, Dr. John M. Green, Gregory Burns, Todd Collier, Richard Cornish, Kyle Curley, Allan Freeman, Jared Spears, "Mapping Artificial Intelligence to the Naval Tactical Kill Chain", *NAVAL ENGINEERS JOURNAL*, Vol. 135, No. 1, pp. 155-167, March. 2023. <https://bt.e-ditionsbyfry.com/publication/?m=63540&i=788511&p=156&ver=html5>
- [17] Kwon Pan Gum, Jang Kyoung Sun, Kim Seung Woo, Kim Jun Young, Yun Won Hyuk, Rhee Kye Jin, "A Study on the Implementation Method of Artificial Intelligence Shipboard Combat System", *Journal of Information and Security*, Vol. 20, No. 2, pp. 123-135, June 2020. DOI: <https://doi.org/10.33778/kcsa.2020.20.2.123>
- [18] Youngmook Kim, "Development and application of artificial intelligence for military training modeling and simulation in Republic of Korea", *Journal of Advances in Military Studies*, Vol. 4, No. 2, pp. 21-36. Aug. 2021. DOI: <https://doi.org/10.37944/jams.v4i2.114>

Authors



Seung-Joon Lee received the B.S. degrees in Information and Communication Engineering from Yeungnam University, Korea, in 2014. He is currently working in Hanwha Systems Co. from 2015.

He is interested in Naval Combat Management Systems, Artificial Intelligence and so on.