

변이형 오토인코더를 이용한 탄도미사일 궤적 증강기법 개발

이동규¹⁾, 홍동욱^{1)*}

1) 한화시스템

Development of Augmentation Method of Ballistic Missile Trajectory using Variational Autoencoder

Dong Kyu Lee¹⁾, Dong Wg Hong^{1)*}

1) Hanwha Systems

Abstract : Trajectory of ballistic missile is defined by inherent flight dynamics, which decided range and maneuvering characteristics. It is crucial to predict range and maneuvering characteristics of ballistic missile in KAMD (Korea Air and Missile Defense) to minimize damage due to ballistic missile attacks, Nowadays, needs for applying AI(Artificial Intelligence) technologies are increasing due to rapid developments of DNN(Deep Neural Networks) technologies. To apply these DNN technologies amount of data are required for supervised learning, but trajectory data of ballistic missiles is limited because of security issues. Trajectory data could be considered as multivariate time series including many variables. And augmentation in time series data is a developing area of research. In this paper, we tried to augment trajectory data of ballistic missiles using recently developed methods. We used TimeVAE(Time Variational AutoEncoder) method and TimeGAN(Time Generative Adversarial Networks) to synthesize missile trajectory data. We also compare the results of two methods and analyse for future works.

Key Words : Ballistic Missile Trajectory, Auto Encoder, Deep Learning, AI, Data Augmentation

Received: November 20, 2023 / **Revised:** December 20, 2023/ **Accepted:** December 26, 2023

* Corresponding Author: Dong Wg Hong / Hanwha Systems / dongwg.hong@hanwha.com

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non- commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

1. 서론

탄도미사일의 궤적은 고유의 비행운동역학에 따라 사거리 및 운동 특성이 결정된다.[1] 탄도미사일 방어를 위한 미사일 방어체계에서 탄도미사일의 특성 및 사거리를 정확히 예측하는 것은 필수적이다. 현재 국내에서는 한국형 미사일 방어체계의 구축을 위하여 한국형 탄도탄작전통제소(KTMO-Cell)을 운용하고 있으며, 탄도탄 작전 통제소에서는 탐지 체계로부터 탐지된 탄도탄 정보 들을 수신하고 이를 종합하여 상황 전파 및 경보 발령, 요격 체계들을 이용한 적극 방어 등을 수행한다. 이러한 일련의 작전에 있어서 가장 중요한 것은 탐지된 탄도탄 정보에 대한 정확한 식별 및 사거리에 대한 예측이다. 이러한 미사일 방어 작전 능력의 향상을 위하여 최근에 발전하고 있는 인공지능 기술의 적용 요구가 증대되고 있다.[2]

인공지능 기술은 여러 가지로 분류될 수 있지만, 최근에 가장 주목받고 있는 기술은 심층신경망에 기반한 기술이다. 심층 신경망 모델은 학습 기반의 지도학습 모델과 비지도학습 모델, 강화학습 모델 등으로 구분된다. 학습 기반의 신경망 모델은 대용량의 학습데이터가 필수적으로 요구되는데, 이는 다양한 데이터를 학습에 이용하여 신경망의 과적합을 피할 수 있기 때문이다. 심층 신경망을 이용하여 가장 괄목할만한 성과를 거둔 분야는 이미지 처리 분야인데, 이러한 성과는 이미지 분야에서 학습 데이터로 사용할 수 있는 방대한 데이터가 있었기 때문이다. 한편 이미지 처리 분야에서 거둔 성과를 다른 분야에 활용하는 연구가 지속적으로 수행되고 있는데 대표적인 분야가 시계열 데이터이다.

시계열 데이터는 시간의 흐름에 따라 변하는 특성을 갖는 데이터들을 지칭하는데 대표적으로 주가, 바이오 신호, 장치의 모니터링을 위한 센서 신호, 자연어 등이 있다. 시계열 데이터는 일정한 크기의 이미지 데이터에 반하여 그 크기가 유동적이며, 시간의 흐름에 따라 변화하는 특성을 내재하고 있어 기존의 이미지 데이터에 적용되었던 신경망 모델을 적

용하기 어렵다. 따라서 시계열 데이터 처리를 위한 신경망 모델이 제안되었으며 RNN (Recurrent Neural Network), LSTM (Long Short Term Memory), GRU (Gated Recurrent Unit), Transformer 등이 대표적이다.

시계열 데이터에 지도학습 모델을 적용할 때, 이미지 데이터에 비하여 상대적으로 학습 데이터 양이 적거나 보유하고 있는 데이터가 비대칭성 등을 포함하는 경우, 학습 모델의 성능이 과적합 문제로 인하여 성능이 떨어지는 경우가 있다. 이러한 문제를 해결하기 위하여 다양한 시계열 데이터 증강 기법에 대하여 최근에 많은 연구가 이루어지고 있다.[3] - [5]

본 연구에서는 학습 기반의 심층 신경망 기술을 적용하여 미사일 방어 작전 능력을 향상시키기 위한 연구의 시작으로, 미사일 궤적 데이터를 증강하기 위하여 심층신경망 기술을 적용하였다. 본 연구에서는 미사일 궤적 데이터를 시계열 데이터로 간주하고 시계열 데이터를 증강하기 위한 다양한 방법 중에서 최근 연구 결과인 변이형 오토인코더를 이용한 방법과 생성적 적대신경망 (Generative Adversarial Networks, GAN) 을 적용하였다.[6],[7] 현재 보유하고 있는 여러 종류의 탄도 미사일 모의 궤적 데이터를 이용하여 변이형 오토인코더 및 생성적 적대신경망을 학습시키고 학습시킨 모델에 기반하여 유사한 훈련용 데이터를 합성하였다.

인공지능 모델의 개발을 위한 시스템 프로세스 정립을 위하여 데이터 수집->데이터 특성 분석->데이터 전처리->모델 학습 및 최적화->모델 배포 및 서비스 등으로 구성되는 프로세스를 정립하고 이를 자동화하기 위한 시스템 엔지니어링 기법으로 MLOps(Machine Learning Operations)방법론과 도구들이 사용되고 있다. 본 연구는 이러한 방법론을 적용하기 위하여 딥러닝 프로세스 중 가장 첫 번째 단계인 데이터 수집에서 기존의 수집된 데이터를 기반으로 학습 데이터 양을 늘이기 위한 증강 기법을 다루고 있다.

본 논문의 구성은 다음과 같다. 1장에서는 본 연

구에 대한 필요성 및 개요를 정리하였고, 2장에서는 본 연구와 관련된 연구들을 정리하였다. 3장에서는 본 연구에서 다루는 탄도 미사일의 궤적 특징에 대하여 기술하였고 4장과 5장에서는 본 연구에서 적용하는 변이형 오토인코더 및 생성적 적대신경망에 대하여 기술하였다. 6장에서는 본 연구에서 수행한 시험방법과 그 결과를 분석하였고 7장에서 본 연구의 결론과 향후 연구를 기술하였다.

2. 관련 연구

탄도 미사일의 비행 특성에 대한 해석이 권용수, 최봉식(2006)의 연구에서 제시되었다. 해당 연구에서는 탄도 미사일의 비행 단계를 3단계로 구분하고 각 단계별로 요구되는 비행 특성에 대한 수학적 모델을 제시하였다. 오주호, 강동수(2022)에서는 순환신경망(RNN) 모델의 한 종류인 LSTM을 이용하여 탄도미사일의 사거리를 예측하는 연구를 수행하였다. 해당 연구에서는 본 연구와 비슷하게 탄도 미사일의 비행궤적 데이터를 모의하여 생성하고, 모의 데이터를 이용하여 LSTM 모델을 학습하여 모의 사거리와 예측 사거리를 비교하는 연구를 수행하였다. 당 연구에서는 전체 모의 궤적 데이터 중에서 거리, 고도, X축 속도, Y축 속도 데이터를 이용하였고, 전체 구간이 아닌 특정 구간의 데이터만을 이용하여 학습을 수행하고 사거리를 예측하였다. 본 연구에서는 X축 속도, Y축 속도, Z축 속도 데이터를 이용하고 특정 구간이 아닌 전체 구간에 대하여 학습을 수행하고 유사 학습 데이터를 생성하였다.

박인희 외(2021)에서는 본 연구에서 비교 방법으로 적용한 생성적 적대 신경망을 이용하여 비행체 궤적의 합성 데이터를 생성하였다. 당 연구에서는 생성적 적대 신경망으로 생성된 비행체 궤적과 2차원 장단기 메모리를 이용하여 비행체의 궤적 예측 작업을 수행하였다. 생성된 비행체 궤적에 대해서는 판별 스코어를 이용하여 기존의 수행된 여러 시계열 데이터와 품질을 비교하였다.

심층 신경망을 이용한 시계열 데이터에 대한 연구는 다양한 분야에서 이루어졌다. Qingsong et al.(2020)에서는 시계열 데이터의 특성을 설명하고 심층 신경망을 이용한 시계열 데이터의 응용 부분을 크게 시계열 분류 및 예측, 그리고 시계열의 이상 탐지 등 3가지 분야로 정의하였다. 그리고 각 응용 부분에서 공통적으로 시계열 데이터의 증강으로 시계열 인공신경망 모델의 성능이 향상됨을 보여주고 있다.

박지연 외(2021)에서는 심층 신경망이 아닌 기계학습 기반의 시계열 데이터 분류 알고리즘인 K-NN(K-Nearest Neighbor), SAX-VSM(Symbolic Aggregate Approximation in Vector Space Model) 등을 이용하여 시계열 데이터의 분류 성능을 평가하였다. Hassan et al.(2019)에서는 심층 신경망을 이용한 시계열 데이터 분류기의 성능 비교 및 분석을 수행하였다. 다계층 신경망 및 합성공 신경망 등 다양한 형태의 신경망에 대한 기존 연구 결과를 리뷰하고 생성형 모델과 분류기 모델로 구분하여 구조를 설명하였다. 9개의 분류기 모델에 대하여 공개된 데이터에 대하여 시험을 수행하고 시계열 분류기의 성능을 높이기 위한 다양한 방법 등을 기술하였으며, 시계열 분류기에 대한 공개 소스 기반의 프레임워크를 제공하였다. Hong Yang, Travis Desell(2022)에서는 단순한 변환 방법 기반의 시계열 데이터 증강 기법을 적용하여 시계열 데이터를 증강시키고 다양한 시계열 분류 모델을 적용하여 성능을 비교 분석하였다. 각 분류 모델 및 시계열 데이터의 종류에 따라 가장 최적의 성능을 보이는 증강 기법을 정리하였다. 김수희, 이기용(2022)에서는 시계열 데이터의 길이를 증강하는 증강 기법을 적용하여 데이터를 증강하고 시계열 분류 모델의 성능을 향상시키기 위한 대조 학습 기반 사전 훈련을 수행하여 분류 모델의 정확도를 향상시킬 수 있음을 보였다. 해당 연구에서는 단순한 증강 방법을 적용하고 증강된 시계열 데이터를 훈련하는 방법을 고도화하여 분류기의 성능을 향상시킬 수 있음을 보여준다.

또 다른 연구 분야로 다양한 시계열 응용 분야에 대하여 증강 기법 및 분류기 등을 적용한 사례에 대한 연구가 수행되었다. 이창현 외(2021)에서는 시계열 데이터의 증강을 위하여 생성적 적대 신경망 모델을 설계하고 교통 데이터를 이용한 실험으로 시계열 데이터 증강 기법이 유효함을 보였다. Francisco J. Bald n, Jos M.Ben tez(2021)에서는 다변량 시계열 분류 작업을 위하여 시계열 데이터의 특성을 포함하는 새로운 특징 벡터들을 정의하고 기존의 분류기를 이용하여 시험을 수행하였고 Kasra et. al(2019)에서는 시계열 이상탐지를 위해 오토인코더를 이용한 시계열 데이터 증강을 수행하였다. 해당 연구에서는 공개된 시계열 데이터를 기반으로 시험한 결과 기존의 데이터 증강 기법보다 오토인코더 기반의 데이터 증강 기법이 더 안정적으로 이상탐지하는 것을 확인하였다. 이수형 외(2022)에서는 게임 사용자의 초기 게임 행동을 분석하여 사용자 분류를 위해 오토인코더를 이용하여 데이터를 증강하고 증강된 데이터를 이용하여 사용자 분류 성능이 높아짐을 보였다. 김승환, 전성혜(2022)에서는 변이형 오토인코더를 사용하여 특허 문서의 전처리를 위하여 데이터를 증강하였다.

MLOPS는 Machine Learning(ML) 모델 사용의 서비스를 제공하는 시스템 개발과 운영을 위한 DevOps(Development Operations) 기반 접근 방식으로 ML 모델을 사용하지 않는 시스템에 비해 복잡해진 시스템 개발 및 운영에 이점을 제공한다. MLOps는 DevOps와 달리 학습을 위한 데이터 처리 절차, 모델 개발과 테스트, 배포 절차, 모델의 모니터링 지원을 포함하며, 이를 통해 많은 ML 시스템이 서비스되고 있다. 시계열 데이터 및 연합학습 시스템 등에 대한 MLOps 구축 사례가 있다.[21]

3. 탄도미사일 궤적 특성

본 연구에서는 시계열 데이터 분석의 기초 자료로 전통적인 탄도미사일 궤적과 활공기동형 탄도미

사일 궤적을 시뮬레이션한 자료를 사용하였다. 일반적으로 탄도미사일의 비행단계는 로켓의 추진력을 이용하는 초기부터 추력이 종료된 후 포물선 운동에 의해 자유낙하하면서 목표에 도달하는 비행종료까지를 추진 및 상승단계, 중간단계 및 종말단계로 구분하는데 추진 및 상승단계는 미사일 발사지점부터 추력이 종료되는 시점까지이며 추력단계(Powered phase)라고도 한다. 중간단계는 추력종료 또는 추진체의 연료소진 시점에서부터 중력에 의해 자유낙하가 시작되는 시점까지로 탄도궤적(Ballistic Trajectory) 또는 궤도단계(Orbit Phase)라고도 하며, 이 구간에서는 탄도미사일 중간유도비행을 통해 궤도조정이 이루어진다. 종말단계(Terminal phase)는 중간단계의 연장선상에 있긴 하지만, 탄도미사일이 비행궤도의 정점에서 낙하하면서 대기권으로 진입시 대기효과(effect of atmosphere)가 탄도미사일의 낙하속도와 궤도에 크게 영향을 미치게 되는 단계로 재진입 단계(Reentry phase)라고도 한다.[18] 탄도미사일 시뮬레이션 궤적은 비행경로각 조절을 통한 궤적성형 방법을 사용하였다. 탄도미사일 궤적은 300km, 500km, 1000km를 대상으로 최소에너지궤적(Minimum Energy Trajectory)으로 구현하였으며, 활공기동형 탄도미사일은 낙하구간에서 경로각 조절을 통해 일정고도 활공을 유지하도록 궤적을 구현하였다. 물리적 입력 제원은 홍동욱(2019)의 데이터를 기반으로 하였으며, 계산의 정확도를 높이기 위해 궤적계산의 추력 및 중량제원을 조정하였다. 궤적 시뮬레이션은 탄도미사일의 운동방정식과 항력특성을 고려하였으며, 이를 토대로 궤적계산을 위한 x,y,z축 성분별 운동방정식은 식 1 ~ 식 7과 같이 나타낼 수 있다. [20]

$$\dot{r}_x = v_x, \dot{r}_y = v_y, \dot{r}_z = v_z \quad (1)$$

$$\dot{v}_x = a_T \frac{v_x}{\sqrt{v_x^2 + v_y^2 + v_z^2}} - a_D \frac{v_x}{\sqrt{v_x^2 + v_y^2 + v_z^2}} - a_G \frac{v_x}{\sqrt{v_x^2 + v_y^2 + v_z^2}} + 2wv_y + w^2r_x \quad (2)$$

$$\dot{v}_y = a_T \frac{v_y}{\sqrt{v_x^2 + v_y^2 + v_z^2}} - a_D \frac{v_y}{\sqrt{v_x^2 + v_y^2 + v_z^2}} - \quad (3)$$

$$a_G \frac{v_y}{\sqrt{v_x^2 + v_y^2 + v_z^2}} + 2wv_x + w^2r_y$$

$$\dot{v}_z = a_T \frac{v_z}{\sqrt{v_x^2 + v_y^2 + v_z^2}} - a_D \frac{v_z}{\sqrt{v_x^2 + v_y^2 + v_z^2}} - \quad (4)$$

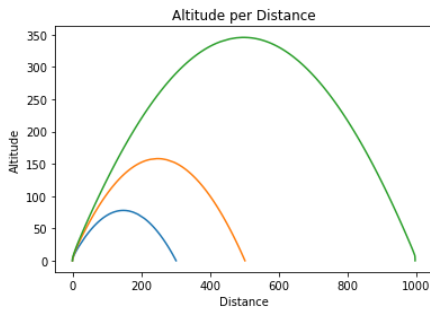
$$- a_G \frac{v_z}{\sqrt{v_x^2 + v_y^2 + v_z^2}}$$

$$a_T = \frac{g_0 I_{sp} \dot{m}}{m(t)} \quad (5)$$

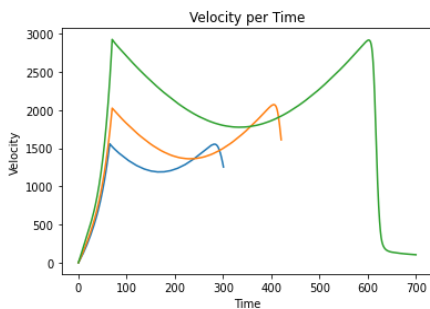
$$a_D = \frac{C_D S \rho(h) v^2}{2m(t)} \quad (6)$$

$$a_G = \frac{\mu}{r^2} \quad (7)$$

$g_0, I_{sp}, C_D, S, \rho$ 및 μ 는 각각 중력가속도, 비추력, 항력계수, 미사일 단면적, 대기밀도 및 중력 파라미터를 의미한다. 궤적의 시뮬레이션 결과는 그림 1과 같다.



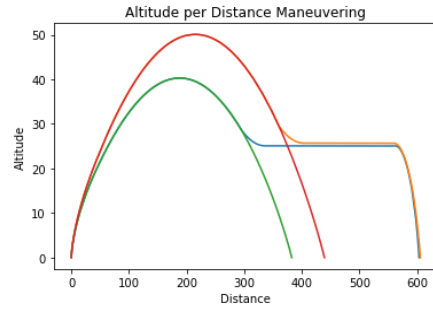
(a) Distance vs Altitude



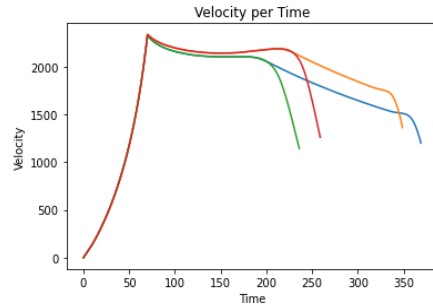
(b) Time vs Velocity

[Figure 1] General Ballistic Missile Trajectory Simulation Results

그림 1은 일반적인 탄도미사일 궤적 시뮬레이션 결과를 나타내었다. 시뮬레이션 결과 오주호, 강동수(2022)에서 수행한 최소에너지 방식의 사거리별 최대속도와 거의 일치함을 알 수 있었다.



(a) Distance vs Altitude



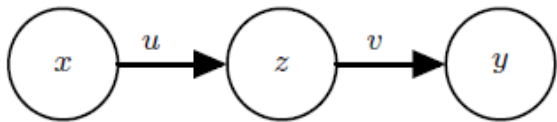
(b) Time vs Velocity

[Figure 2] Low Altitude Glide Maneuvering Trajectory Simulation Results

그림 2는 저고도 활강 기동형 탄도미사일 궤적 시뮬레이션 결과를 나타내었다. 시뮬레이션 결과 동일한 최고 고도를 가지는 일반적인 최소에너지 방식의 탄도미사일 궤적은 최대 사거리가 각각 380km와 430km를 나타낸 반면, 활강 기동형 탄도미사일의 경우는 최대 600km까지 사거리 증가가 가능하였고 최대사거리 증가에 따라 속도 감소가 일반 탄도미사일에 비해 완만하게 이루어짐을 알 수 있었다.

4. 변이형 오토인코더

오토인코더는 비지도 학습 방식의 다계층 신경망으로 구성되며 인코더와 디코더로 구성된다. 인코더는 다차원의 입력 벡터를 입력으로 받아 저차원의 특징 공간으로 변환하는 기능을 수행한다. 디코더는 저차원의 특징 공간을 입력으로 받아 입력 데이터로 다시 복원하는 역할을 수행한다. 디코더의 출력을 입력 데이터와 비교하여 오차를 분석하고 오차를 최소화하도록 과정을 반복하며 인코더와 디코더를 학습시킨다.[15]



[Figure 3] The Structure of a Basic Autoencoder[15]

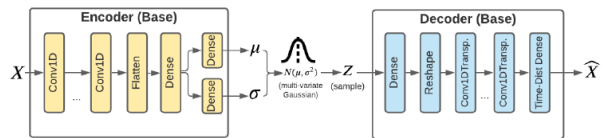
위의 그림에 나타난 바와 같이 1개의 잠재적 레이어(latent layer)로 구성된 오토인코더는 입력 x 를 받아 함수 u 를 이용하여 표현벡터 z 로 변환한다. 그리고 디코더는 표현벡터 z 를 입력으로 받아 함수 v 를 이용하여 y 데이터로 복원한다. 인코더와 디코더는 다층 신경망으로 구현되어 입력 데이터 x 와 디코더의 출력값 y 의 오차를 이용하여 오차가 최소화 되도록 신경망의 파라미터를 학습한다. 이 과정에서 신경망은 데이터들의 주요 특성을 학습하게 되며 학습이 완료된 모델을 이용하여 새로운 학습 데이터를 생성하거나 입력 데이터의 노이즈를 제거할 수 있다. 오토인코더는 인코더와 디코더를 구성하는 신경망 모델을 다양하게 조합하여 많은 모델이 연구되어지고 있으며 변이형 오토인코더도 이러한 연구의 결과로 제시되었다.

변이형 오토인코더는 오토인코더에서 인코더와 손실함수가 다른 모델이다. 변이형 오토인코더의 인코더는 표현벡터 z 로 변환하는 대신 잠재 공간의 분포를 출력한다. 변이형 오토인코더의 목적은 원본 데이터의 분포를 정확하게 나타낼 수 있는 샘플들을

생성하는 것이다. 변이형 오토인코더의 손실 함수는 오토인코더의 손실함수에 추가적으로 아래의 수식과 같은 쿨백-라이블러 발산 (Kullback-Leibler divergence)을 사용한다.

$$D_{KL}[N(\mu, \sigma) || N(0, 1)] = -\frac{1}{2} \sum (1 + \log(\sigma^2) - \mu^2 - \sigma^2) \quad (8)$$

변이형 오토인코더는 입력 데이터와 분포적으로 매우 유사한 새로운 데이터를 생성하기 때문에 머신러닝에서 사용되는 대표적인 생성모형이다.[6] 본 연구에서는 Abhyuday et. al(2021)에서 시계열 생성용으로 제시한 변이형 오토인코더 모델(TimeVAE) 중에서 기본 모델을 사용한다. 그림 4에 기본 TimeVAE에 대한 구조를 나타내었다.



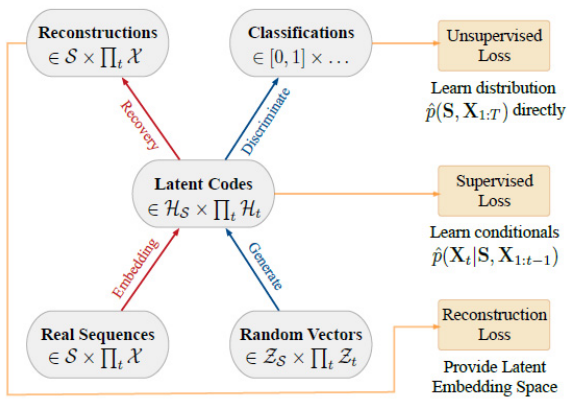
[Figure 4] Block diagram of components in Base TimeVAE[6]

5. 생성적 적대신경망(GAN)

생성적 적대신경망은 준지도 및 비지도 학습을 위한 신경망 중 하나로서, 합성 데이터 생성을 통한 지도 학습(판별) 방법을 이용한다. 생성적 적대 신경망은 두 개의 신경망으로 구성되는데, 생성 신경망은 랜덤 노이즈로 구성된 벡터를 입력받아 실제 데이터와 비슷한 특성을 갖는 합성 데이터를 생성한다. 판별 신경망은 실제 데이터와 합성 데이터를 분류하는 것을 목표로 한다. 생성적 적대신경망은 이미지 데이터 뿐 아니라 자연어 처리, 음성 및 오디오 합성, 음악 생성 등과 같은 시계열 데이터를 다루는 분야에서도 많은 연구가 수행되었다.

본 연구에서는 Jinsung et. al(2019)에서 제시한 시계열 생성 모델 (TimeGAN: Time-series

Generative Adversarial Networks)을 적용하였다. TimeGAN은 4개의 신경망 Embedding function, Recovery function, Sequence generator, 그리고 Sequence discriminator로 구성된다. 처음의 2개 신경망은 오토인코더를 구성하며 뒤에 있는 2개의 적대적 신경망과 함께 훈련된다. TimeGAN은 각각의 신경망을 훈련하기 위한 목적 함수를 정의하고, 랜덤하게 데이터를 생성함과 동시에 판별망을 이용하여 원본 데이터와 비교함으로써 동시에 4개의 신경망을 훈련시킨다. 아래 그림에 TimeGAN의 구성을 나타내었다.



[Figure 5] Block Diagram of Component Function and Objectives in TimeGAN[7]

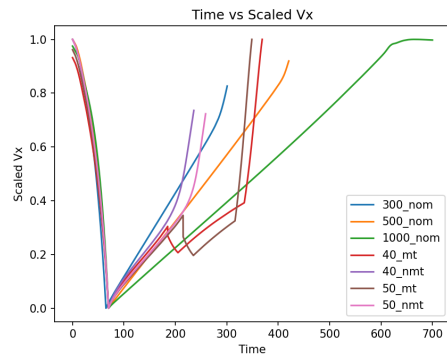
6. 시험 및 결과

본 연구에서는 탄도 미사일의 궤적을 사거리와 기동 특성 등을 이용하여 모두 7 종류의 모의 데이터를 훈련용 데이터로 사용하였다. 일반적인 탄도미사일의 탄도비행을 하면서 사거리가 각각 300km, 500km, 1000km인 3종류의 궤적 데이터와 최대 고도가 각각 40km, 50km일 때 정상 탄도비행을 하는 궤적과 회피 기동을 하는 궤적을 갖는 4종류의 궤적 데이터를 이용하였다. 각 궤적들은 시간대별로 X축, Y축, Z축 좌표값과 X축, Y축, Z축 속도 벡터 등 모두 13개의 특성 값을 갖는다. 표 1에 입력 데이터를 나타내었다.

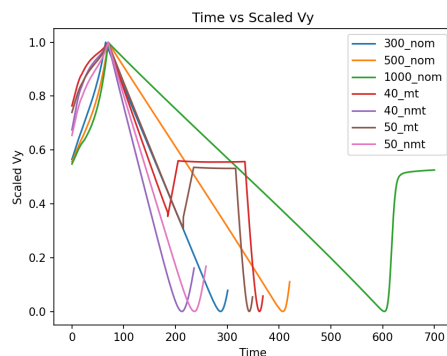
<Table 1> Specifications of Simulated Trajectory

| No. | Distance (km) | Altitude (km) | Maneuvering |
|-----|---------------|---------------|-------------|
| 1 | 300 | 80 | - |
| 2 | 500 | 160 | - |
| 3 | 1000 | 350 | - |
| 4 | 383 | 40 | - |
| 5 | 603 | 40 | 0 |
| 6 | 439 | 50 | - |
| 7 | 606 | 50 | 0 |

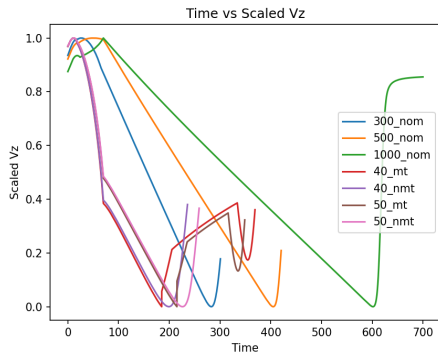
7종류의 모의 데이터를 대상으로 특성값 중에서 X축 속도 벡터, Y축 속도 벡터, Z축 속도 벡터를 증강시켰다. 속도 벡터를 증강 대상으로 선택한 것은 모의 데이터에서 가장 중요한 것이 속도 벡터이기 때문이다. X, Y, Z 위치값과 고도값은 탄도 미사일의 발사 시점을 선택한 후 속도 벡터를 이용하여 계산 가능한 값이다. 그림 6에 각 신호별 속도 벡터의 값과 고도 값을 정규화하여 나타내었다.



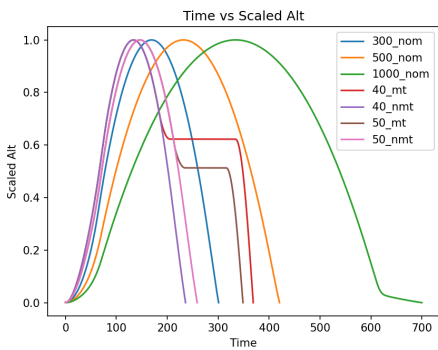
(a) Vx data



(b) Vy data



(c) Vz data



(d) Alt data

[Figure 6] Normalized Specifications of Simulated Signals

본 연구에서는 Abhyuday et. al(2021)에 기술된 변이형 오토인코더를 사용한 생성 모델(TimeVAE)과 Jinsung et. al(2019)에 제시된 적대적 생성망을 이용한 생성 모델(TimeGAN)을 사용하여 그 결과를 비교하였다. 기존의 연구 결과와의 비교를 위하여 기존 연구에서 사용되었던 판별 스코어(Discriminative Score)와 예측 스코어(Predictive Score)를 계산하여 비교를 수행하였다. 판별 스코어는 (accuracy-0.5)값으로 정의되며 0에 가까울수록 좋은 성능을 나타낸다. 예측 스코어는 사후 시퀀스 예측 모델을 이용하여 예측된 값과 원본 값과의 절대에러를 측정된 값으로 0에 가까울수록 좋은 성능을 나타낸다.[7] 또한 시각화 도구로써 PCA(Principal Component Analysis)와 t-SNE(t-distributed Stochastic Neighbor Embedding)를 적용한 2차원 그래프를 사용하였다. 신경망 모델에 적용되는 세대 수 등의 하이퍼 파라

미터는 비교를 위하여 동일한 값을 사용하였다.

전처리 과정에서 CSV 파일로부터 읽어들이는 2차원의 입력 데이터들을 3차원 벡터로 표시하기 위하여 사용자가 지정하는 시퀀스 길이를 이용하여 (N, T, D)의 3차원 벡터로 만드는데 (N: batch size, T: Number of Time steps (Sequence length), D: Number of feature dimension), TimeVAE에서는 모두 동일한 크기를 사용하였고, TimeGAN에서는 입력 데이터에 따라 크기를 변경하면서 시험을 수행하였다. TimeVAE의 경우 원본 데이터로 훈련시킨 각각의 모델을 저장한 후, 훈련에 사용한 입력 데이터를 다시 입력하여 생성된 시퀀스와 모델에 정규분포를 갖는 랜덤한 값을 입력으로 주어 생성된 시퀀스를 같이 분석하였다. 표 2와 그림 7에 본 연구에서 수행한 시험 결과를 나타내었다.

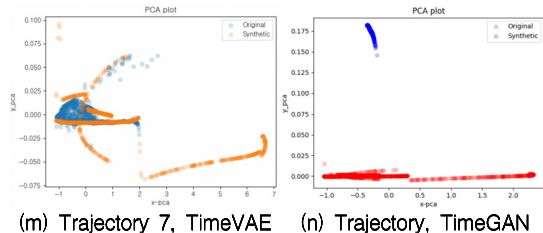
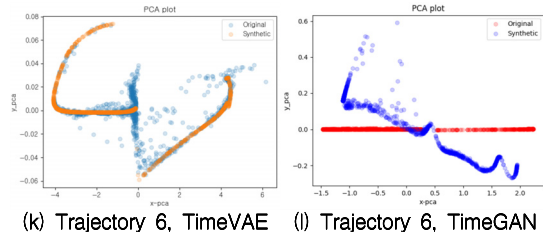
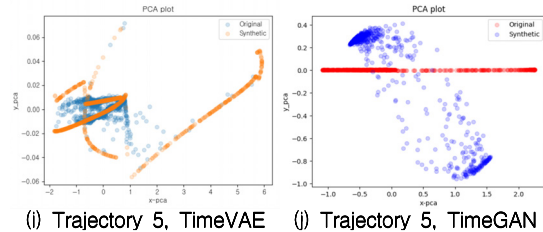
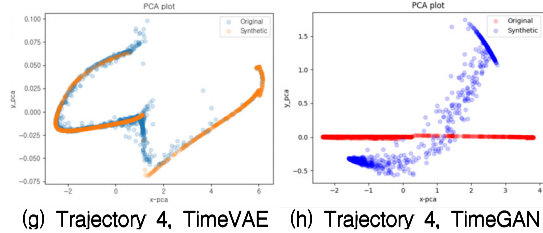
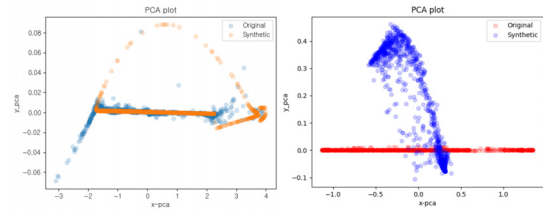
<Table 2> Discriminative(disc.) and Predictive(pred.) Scores of Synthesized Data

| No. | TimeVAE (Original Trajectory Input) | | TimeVAE (Normal Distribution Random Input) | | TimeGAN | |
|-----|--|-------|---|-------|---------|-------|
| | disc. | pred. | disc. | pred. | disc. | pred. |
| 1 | 0.021 | 0.025 | 0.28 | 0.051 | 0.283 | 0.134 |
| 2 | 0.018 | 0.033 | 0.204 | 0.085 | 0.272 | 0.134 |
| 3 | 0.024 | 0.029 | 0.099 | 0.06 | 0.307 | 0.162 |
| 4 | 0.021 | 0.031 | 0.056 | 0.027 | 0.27 | 0.077 |
| 5 | 0.021 | 0.016 | 0.202 | 0.028 | 0.251 | 0.175 |
| 6 | 0.021 | 0.031 | 0.115 | 0.025 | 0.299 | 0.098 |
| 7 | 0.014 | 0.042 | 0.224 | 0.113 | 0.287 | 0.219 |
| avg | 0.02 | 0.03 | 0.17 | 0.06 | 0.28 | 0.14 |

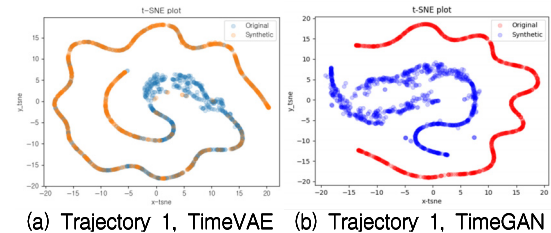
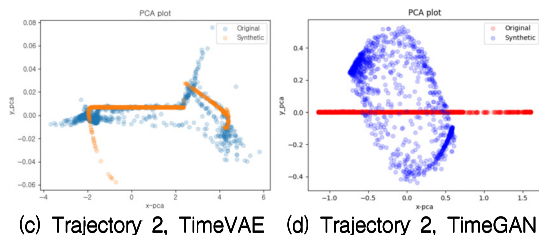
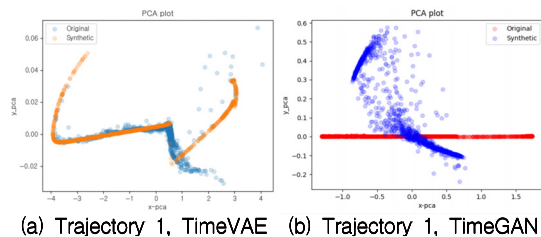
위의 결과를 보면 모든 7가지 종류의 모의 신호에 대하여 적대적 생성망을 이용한 생성 데이터에 비하여 변이형 오토인코더를 적용한 결과가 우수함을 알 수 있다. 원본 입력 데이터를 이용하여 변이형 오토인코더로 합성된 데이터가 가장 우수한 결과를 보이고 있고 정규 분포를 갖는 임의의 난수를 입력으로 하여 생성된 결과도 적대적 생성망을 이용한 경우와 비슷하거나 우수한 성능을 보이고 있다. 정규 분포를 갖는 임의의 난수를 입력으로 할 경우 학습된 변이형 오토인코더 모델을 이용하여 원하는 만

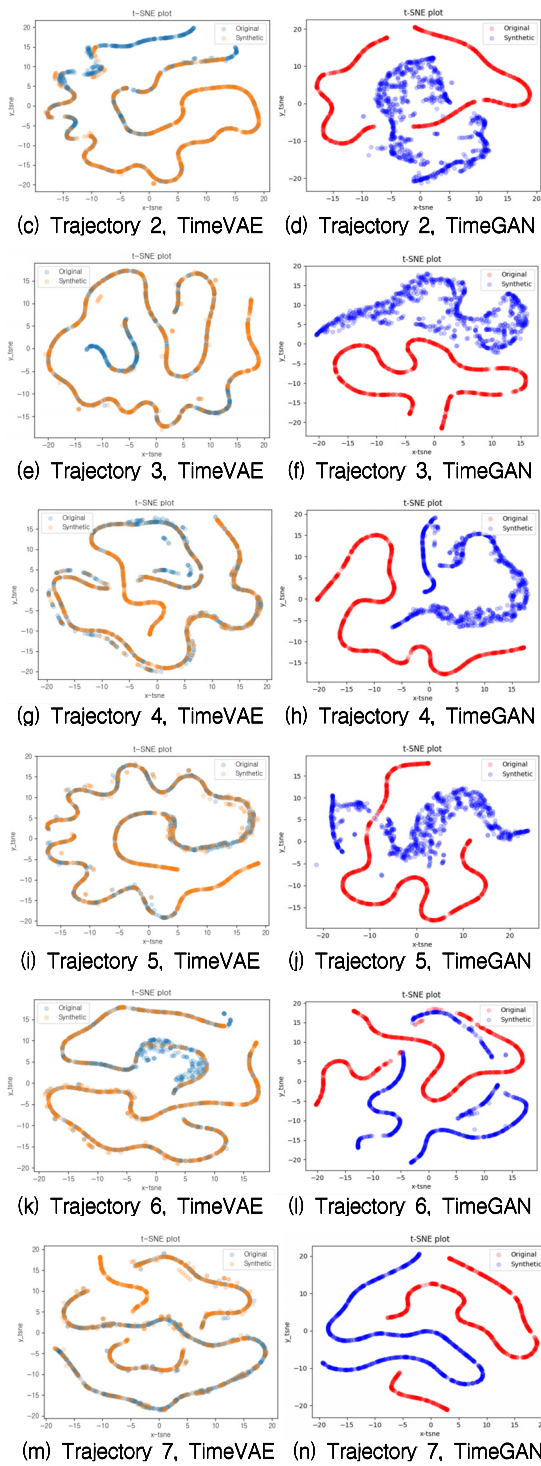
컴의 더욱 더 많은 합성 데이터를 생성할 수 있을 것이다.

시각화 도구로 각각의 PCA 결과와 t-SNE 결과를 비교해도 변이형 오토인코더를 적용한 방법이 더 우수한 성능을 갖고 있음을 알 수 있다. 그림 7에 각 모의신호별 TimeVAE와 TimeGAN의 PCA 결과를 나타내었고 그림 8에 각 모의신호별 TimeVAE와 TimeGAN의 t-SNE 결과를 나타내었다. PCA와 t-SNE는 고차원의 데이터를 저차원으로 축소하여 표시하는 방법으로 원본 데이터와 합성된 데이터가 잘 일치할수록 좋은 성능을 나타낸다. 그림 7과 8의 그래프에서 다른 색으로 표시된 그래프는 원본 항적과 본 연구의 결과로 합성된 항적의 X, Y, Z 3차원의 속도 벡터를 2차원으로 축소하여 나타낸 것이다. 원본 데이터의 특성에 따라 PCA는 선형 방식으로 차원을 축소하고 t-SNE는 t-분포를 기반으로 차원을 축소한다. 이렇게 차원을 축소하여 각 특성을 그래프로 비교할 수 있다. 원본 데이터와 합성된 데이터의 분포를 비교하여 그 일치되는 정도로 합성된 데이터의 품질을 비교할 수 있다. 그림 7과 8에 나타낸 결과는 TimeGAN의 결과와 정규 분포를 갖는 임의의 난수를 입력으로 하여 생성된 변이형 오토인코더의 결과를 나타내었다. 그림에 포함하지 않은 원본 입력에 대한 변이형 오토인코더의 그림은 원본과 합성 데이터가 가장 일치하는 결과를 나타내고 있다.



[Figure 7] PCA Plots of synthesized trajectory





[Figure 8] t-SNE Plots of synthesized trajectory

본 연구에서는 TimeVAE와 TimeGAN을 구현하기 위하여 Windows 10 기반의 노트북에서 파이썬을 사용하였다. 사용한 노트북의 사양은 Intel Core

i5 2.4GHz CPU 기반에 8 GBytes의 메모리를 탑재하였으며 별도의 GPU는 사용하지 않았다. 파이썬 개발도구로 Pycharm 2023.1.1. Community Edition을 사용하였으며, 아나콘다 가상환경으로 파이썬 환경을 구축하였다. TimeVAE는 3.8 버전의 python 환경에서 tensorflow 2.12.0 모듈을 사용하였고, TimeGAN은 3.6 버전의 python 환경에서 tensorflow 1.15.0 모듈을 사용하였다. 본 연구에서 사용한 하드웨어는 빠른 학습을 위한 별도의 그래픽 가속 장치를 사용하지 않아 실행시간이 최적화되었다고 하기는 어렵지만 2가지 방법의 상대 비교를 위해 실행시간을 측정하였다. 표 3은 각 생성 모델의 실행시간을 나타내었다. 실행 시간은 모델의 학습을 위한 다양한 파라미터에 의존적이기 때문에 절대적인 비교가 어렵지만 비슷한 파라미터 상에서도 TimeVAE의 실행속도가 상대적으로 빠르다는 것을 확인할 수 있다. 향후 그래픽 가속 장치를 이용한 하드웨어를 이용하면 실행시간을 더욱 단축할 수 있을 것이다.

<Table 3> Training Times for Both Models

| No. | Times (min) | |
|-----|-------------|---------|
| | TimeVAE | TimeGAN |
| 1 | 14.87 | 25.66 |
| 2 | 28.21 | 24.38 |
| 3 | 23.56 | 25.72 |
| 4 | 10.54 | 47.17 |
| 5 | 14.17 | 23.72 |
| 6 | 10.54 | 102.41 |
| 7 | 14.42 | 26.26 |
| Avg | 16.62 | 39.33 |

7. 결론

본 연구에서는 미사일 궤적 데이터를 시계열 데이터로 간주하고 시계열 데이터의 증강 기법 중에서 가장 우수한 성능을 갖고 있다고 보고된 최근의 연구 결과들을 이용하여 그 성능을 비교하였다. 최근

에 이미지 처리나 음성 처리 분야에서 많은 진전이 이루어진 생성형 모델을 이용한 증강 방법과 변이형 오토인코더를 이용한 증강 방법을 적용하여 각각의 성능을 비교하였다. 비교 결과 미사일 궤적 데이터의 증강데이터는 기존에 가장 우수한 성능을 보였던 적대적 생성망을 이용한 방법에 비해서 변이형 오토인코더를 적용한 방식이 입력 데이터에 따라 비슷하거나 우수한 성능을 보이는 것을 확인하였다.

향후의 연구로서 본 연구에서 개발한 증강 기법을 더욱 발전시키는 작업을 수행해야 한다. 우선 GPU 기반의 프로세서에서 병렬 처리가 가능하도록 개발하여 학습에 소요되는 시간을 줄일 수 있어야 할 것이다. 그래서 대량의 모의 데이터에 대해서도 빠른 속도로 실행이 가능하며 생성된 데이터들을 효율적으로 관리할 수 있는 시스템을 구축해야 한다. 이를 이용하여 본 연구에서 개발된 생성 모델에 대해서도 반복적인 시험을 통해 모델의 하이퍼파라미터를 조절하여 생성 모델의 최적화를 수행해야 한다. 이를 위한 시스템 엔지니어링 기법인 MLOps 기반의 자동화 파이프라인 구축할 것이다.

그리고 최종적으로는 본 연구에서 개발한 증강 기법으로 데이터를 증강하여 성능이 향상된 심층망 기반의 시계열 분류기를 개발해야 한다. 증강된 데이터를 이용하여 미사일의 사거리 및 기동 특성을 분류할 수 있는 심층망 기반의 시계열 분류기 성능을 향상시킬 수 있을 것이다. 이를 이용하면 본 연구의 최종 목표인 미사일 궤적 데이터에 대하여 조기에 미사일의 사거리 및 기동 특성을 정확하게 분류할 수 있는 진보된 미사일 방어 기술을 확보할 수 있을 것이다.

References

1. 권용수, 최봉식, 탄도미사일의 비행궤적 특성 해석, 한국국방경영분석학회지, Vol. 32, No. 1, p.176-187, 2006.
2. 오주호, 강동수, LSTM을 이용한 탄도미사일 사

- 거리 예측, 정보과학회 컴퓨팅의 실제 논문지, Vol. 28, No. 8, p.405-412, 2022.8.
3. Qingsong Wen, Liang Sun, Fan Yang, Xiaomin Song, Jingkun Wang, Huan Xu, Time Series Data Augmentation for Deep Learning: A Survey, Proceedings of the 30 International Joint Conference on Artificial Intelligence (IJCAI-21) Survey Track, 2020.
4. Brian Kenji Iwana, Seiichi Uchida, An empirical survey of data augmentation for time series classification with neural networks, PLOS ONE, July 15, 2021.
5. 김수희, 이기용, 기계학습 성능 향상을 위한 효과적인 시계열 데이터 증강 기법, 한국정보과학회 학술발표 논문집, p. 703-705, 2021.
6. Abhyuday Desai, Cynthia Freeman, Ian Beaver, Zuhui Wang, Timevae: A variational auto-encoder for multivariate time series generation, arXiv preprint arXiv:2111.08095, 2021.
7. Jinsung Yoon, Daniel Jarrett, Mihaela van der Schaar, Time-series Generative Adversarial Networks, 33rd Conference on Neural Information Processing Systems (NeurIPS), 2019.
8. 박인희, 이창진, 정찬호, 시계열 생성적 적대 신경망을 이용한 비행체 궤적 합성 데이터 생성 및 비행체 궤적 예측에서의 활용에 관한 연구, 전기전자학회 논문지, Vol. 25, No. 4, p. 766-769, December 2021.
9. 박지연, 서동호, 최병찬, 남해운, 시계열 데이터 분류 알고리즘 성능 비교 및 분석, 한국통신학회 학술대회 논문집, p. 111-112, 2021.
10. Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, Pierre-Alain Muller, Deep learning for time series classification: a review, Data mining and Knowledge Discovery, Vol. 33, No. 4,

- p. 917-963, 2019.
11. Hong Yang, Travis Desell, Robust Augmentation for Multivariate Time Series Classification, arXiv preprint arXiv:2201.11739, 2022.
 12. 김수희, 이기용, 시계열 분류 모델의 성능 향상을 위한 시계열 데이터 증강 및 대조 학습 기반 사전 훈련 기법, 정보과학회 컴퓨팅의 실제 논문지, Vol.28, No.11, p.550-556, 2022.
 13. 이창현, 신동훈, 전호석, 정진우, 권기원, 임태호, 생성적 적대 신경망(GAN)을 이용한 시계열 데이터 증강 기법 적용 및 성능 평가에 관한 연구, 한국통신학회 학술대회 논문집, p. 1076-1077, 2021.
 14. Francisco J. Baldo n, Jos M.Ben tez, Multivariate times series classification through an interpretable representation, Information Sciences 569, p.596-614, 2021.
 15. Kasra Babaei, ZhiYuan Chen, Tomas Maul, Data augmentation by AutoEncoders for Unsupervised Anomaly Detection, arXiv: 1912.13384v1, 21 Dec, 2019.
 16. 이수형, 김범성, 이희성, 오토인코더 기반의 데이터 증강과 초기 게임 행동을 이용한 구매 예측 시스템, 한국지능시스템학회 논문지, Vol. 32, No. 4, p. 286-291, 2022.
 17. 김승환, 전성해, 변이형 오토인코더를 이용한 영과잉 특허 데이터 분석, 한국지능시스템학회 논문지, Vol. 32, No. 1, p.44-50, 2022.
 18. 오현식, 김윤식, 이순주, 노성규, 궤적 성형을 통한 극초음속 활강 비행체 모델링 및 시뮬레이션, 한국항공우주학회 학술발표회 초록집, p.314-315, 2022.
 19. 홍동욱, 탄도미사일 궤적 시뮬레이션을 이용한 방어영역 산출 및 응용, 박사학위 논문, 한남 대학교, 2019.
 20. 김지원, 권용수, 항력을 고려한 탄도미사일 비행 궤적 특성 해석, 한국항공학회 논문지, 20(2), p.134-140, 2016.
 21. 홍승후, 이강윤, 연합학습시스템에서의 MLOps 구현방안 연구, 한국인터넷정보학회 논문지, 23(3), 2022.