

심층 강화학습을 이용한 선도 추격자 대항 전투기 회피 기동 연구

박준우^{1*}, 채혁주¹, 박수정¹, 최한림¹, 방효충¹, 안경수², 김영곤²한국과학기술원¹, 한화시스템(주)²

Study on Evasive Maneuver of a Fighter against Lead Pursuer using Deep Reinforcement Learning

Junwoo Park^{1*}, Hyeok-Joo Chae¹, Su-Jung Park¹, Han-Lim Choi¹, Hyochoon Bang¹, Kyeong-Soo An², Young-Gon Kim²

Key Words : Within Visual Range Combat(근접 공중전), Lead Pursuit(선도 추격), Evasive Maneuver(회피 기동), Reinforcement Learning(강화 학습), Basic Fighter Maneuver(기본 전투 기동)

서론

인공지능(AI)을 전투기 공중전 등의 의사결정에 사용하여 교전 우위를 점하려는 시도는 최근 국방 분야의 주요 관심사 중 하나⁽¹⁾이다. G-LOC (G-induced Loss of Consciousness) 상황에서의 TF (Terrain Following) 혹은 GCAS (Ground Collision Avoidance System) 등과 같이 비행체 전반에서 조종사를 보조하거나 아예 조종사의 역할을 대신하는 등의 시도도 활발히 이루어지고 있다. 이런 흐름에 맞추어 볼 때, 인공지능 기반의 의사결정이 가능하여 조종사가 필요 없는 가상군(Computer Generated Force, CGF)은 그 활용도가 특히 높다고 할 수 있다.

T-50에는 내장형(Embedded) 조종사 훈련시스템이 탑재되어 있어, 별도의 시뮬레이터 없이 기체에 탑재한 채로 가상의 임무 수행이 가능한 것으로 알려져 있으나 기본 전투 기동(Basic Fighter Maneuver, BFM) 등의 규칙 기반의 기동을 하는 CGF들로 훈련 시나리오를 구성한다면 훈련생이 도출하거나 학습할 수 있는 전술 또한 제한적일 것이다. 최근에는 강화 학습을 활용하여 격자 형태로 간소화된 상태 공간(State space) 내의 교전 상황에서 준최적의 기동 전술을 획득하거나⁽²⁾ 전문가의 교전 궤적으로부터 교전 우위를 점하는 기동을 학습하는 모방 학습⁽³⁾ 등 학습 기반의 인공지능형 CGF를 구성하는 시도가 다수 이루어지고 있는데, 본 연구는 이와 같은 인공지능형 개체를 조종사 훈련시스템에 활용할 가능성을 확인해보고자 한다. 개종에서 공격적인 전투 기동을 훈련하는 시나리오의 경우에는 이를 효과적으로 회피하는 상대가 필요하고, 본 연구는 특히 비례 항법 유도 법칙과 선도 추격(Lead Pursuit) 유도 법칙을 혼용하여⁽⁴⁾ 기종 교전의 우위를 점하려는 추격자(red, 첨자 r)가 근접할 때 이를 효과적으로 회피하는 회피자(blue, 첨자 b)의 기동을 강화 학습을 통해 도출하고자 한다. 이는 전투기 조종사 훈련 모의 시뮬레이터 내에서 상대편 CGF에 적용하여 특정 시나리오로 삽입될 수 있을 것이다. 학습된 기동의 모의실험 결과 궤적을 이용해 기본 전투 기동 간의 비교 분석을 수행한다.

본론

추격자와 회피자는 실제 전투기의 거동 혹은 이에 준하는 Full-6DoF 동역학 모델과의 충실성(Fidelity)은 확보하되, 강화 학습 시 탐색이 필요한 상태 공간을 최대한 줄이기 위해 식 (1)의 Pseudo-6DoF 운동 모델을 활용한다. 식 (1)에서 \mathbf{X} , V 는 각각 관성 좌표계 기준 기체의 위치와 속력, T, L, D 는 각각 추력, 양력, 그리고 항력의 크기를 나타내며 비행 고도 및 속도, 받음각 α 의 함수이다. χ, γ 는 각각 방향각 및 비행 경로각이며, ϕ 는 기체의 뱅크각을 의미한다. 제어 입력 \mathbf{u} 는 식 (3)의 받음각과 뱅크각의 명령값을 사용하며, 각각의 거동은 식 (2)와 같이 1차 시스템으로 가정한다. 강화학습에서 사용할 Action a 는 회피자의 제어 입력 \mathbf{u} 로 설계한다. ($a = \mathbf{u}_b$) 단, 기체는 고도에 따라 낼 수 있는 최대 추력 T_{\max} 가 정해지고 실수 $\eta \in [0, 1]$ 을 이용해 현재의 추력 T 를 ηT_{\max} 로 모델링 하면 Action a 를 증강하여 $a' = [a^T, \eta_b]^T$ 와 같이 사용할 수 있지만 본 연구에서는 최대 추력을 이용한 회피를 상정하여 η 는 1로 고정하여 사용한다. 아울러, 강화 학습 Agent의 관측값은 식 (4)와 같이 설계한다.

$$\dot{\mathbf{X}} = \begin{bmatrix} \cos(\gamma)\cos(\chi) \\ \cos(\gamma)\sin(\chi) \\ -\sin(\gamma) \end{bmatrix} V \quad (1)$$

$$\begin{aligned} \dot{V} &= \frac{T\cos(\alpha) - D}{m} - g\sin(\gamma) \\ \dot{\chi} &= \frac{(L + T\sin(\alpha))\sin(\phi)}{mV\cos(\gamma)} \\ \dot{\gamma} &= \frac{(L + T\sin(\alpha))\cos(\phi) - mg\cos(\gamma)}{mV} \end{aligned}$$

$$\dot{\alpha} = \frac{\alpha_{\text{cmd}} - \alpha}{\tau_{\alpha}}, \dot{\phi} = \frac{\phi_{\text{cmd}} - \phi}{\tau_{\phi}} \quad (2)$$

$$\mathbf{u} = [\alpha_{\text{cmd}}, \phi_{\text{cmd}}]^T \quad (3)$$

$$x_{RL} = \begin{bmatrix} \overline{V}_b, c_{\chi_b}, s_{\chi_b}, c_{\gamma_b}, s_{\gamma_b}, c_{\alpha_b}, s_{\alpha_b}, c_{\phi_b}, s_{\phi_b}, \dots \\ \overline{V}_r, c_{\chi_r}, s_{\chi_r}, c_{\gamma_r}, s_{\gamma_r}, c_{\alpha_r}, s_{\alpha_r}, c_{\phi_r}, s_{\phi_r}, \dots \\ \overline{X}_b, \overline{X}_r - \overline{X}_b, \overline{X}_r \end{bmatrix}^T \quad (4)$$

여기서 $\overline{}$ 표기는 각각의 변수를 사전에 지정한 최댓값으로 나누어서 정규화해 $[0, 1]$ 범위로 대응시킨 것을 의미하며, c, s 는 각각 삼각함수 $\cos(\cdot)$ 과 $\sin(\cdot)$ 의 줄임말로 마찬가지로 $[-\pi, \pi]$ 혹은 $[0, 2\pi]$ 의 범위를 가지는 각도 변수들을 정규화해 $[-1, 1]$ 로 대응시킨 것이다. 입력 정규화는 일반적으로 상대 공간 내에서 단위, 규모 그리고 운용 범위에 따른 비균질성(heterogeneity)을 해소하여 학습 속도 및 안정성을 높이는 데 쓰인다.

Agent가 획득하는 보상은 식 (5)와 같이 설계한다. 지표면보다 아래로 기동하거나, 무한히 상승하는 형태의 회피를 지양하기 위해서 사전에 설정된 고도 경계를 넘어서면 에피소드를 종료하고 매우 큰 음의 보상을 줌으로써 부정 강화(Negative Reinforcement)를 유도한다. 아울러, 기총 타격 가능한 범위를 Fig. 1과 같은 원뿔형으로 가정하여 식 (6)의 조건을 이용해 피격 여부를 결정하는데 회피자가 해당 범위 밖에서 기동할 경우 지속적인 양의 보상을 받도록 한다. 식 (6)에서 C_n^r 은 관성 좌표계에서 추격자 동체 좌표계의 좌표 변환 행렬로 식 (7)과 같이 계산할 수 있으며, 모두 추격자의 변수를 사용한다. 식 (1)의 Pseudo-6DoF 모델은 점질량의 운동을 기술한 것이기 때문에 옆 미끄러짐각 β 가 0이라는 가정으로부터 회전 변환 행렬을 도출한다. w 는 특히 바람좌표계를 의미한다.

$$r = \begin{cases} 1 & \text{evading} \\ -5 \times 10^2 & \text{being shot} \\ -10^4 & \text{altitude limit} \end{cases} \quad (5)$$

$$\begin{bmatrix} x_{rb}^r \\ y_{rb}^r \\ z_{rb}^r \end{bmatrix} = C_n^r (X_b - X_r) \quad (6)$$

$$0 < x_{rb}^r < x_{gun}, \frac{\sqrt{y_{rb}^r{}^2 + z_{rb}^r{}^2}}{x_{rb}^r} < \tan(\theta_{gun})$$

$$C_n^r = C_w^r C_n^w = \begin{bmatrix} C_\alpha & 0 & -S_\alpha \\ S_\phi S_\alpha & C_\phi & S_\phi C_\alpha \\ C_\phi S_\alpha - S_\phi C_\phi C_\alpha \end{bmatrix} \begin{bmatrix} C_\gamma C_\chi C_\gamma S_\chi - S_\gamma \\ -S_\chi C_\chi & 0 \\ S_\gamma C_\chi S_\gamma S_\chi & C_\gamma \end{bmatrix} \quad (7)$$

학습은 결정론적 정책 기울기, DDPG⁽⁵⁾ 기법을 이용해 회피기의 기동 정책을 학습하였는데, 연속적인 Action 공간에 적용할 수 있다는 점과 Off-policy로 학습하여 잘못된 특정 지역(Local) 정책에 수렴할 확률이 낮다는 특성 때문에 본 연구에 특히 적합하다.

Fig. 2에는 추격자가 회피자보다 높은 속력을 가지고 회피자의 후방 위쪽에서 추격을 시작한 경우에 본 연구를 통해 학습한 회피자가 이를 회피한 결과를 나타내었다. 상대적으로 높은 고도에 있는 추격자에게 우세한 위치를 내어주지 않기 위해 교전 시작 직후 뱅

크각을 크게 틀어($\|\phi_{cmd}\| \gg 1$) 배면 비행을 시작하며 이후 기수각을 들어 종축 선회하는 모습을 보인다. 높은 속력을 가진 추격자는 달성하기 어려운 작은 선회 반경으로 선회하며, 추격자가 다시 근접해올 경우 연속적인 종축 극기동($\alpha_{cmd} = \alpha_{max}$)을 통해 추미를 내어주지 않는다. 전체적으로 추격자의 에너지 낭비를 유도해 초기 조건보다 월등한 교전 기하를 확보한 것을 확인할 수 있다. 단, 선도 추격자의 경우 선도율과 가시선(Line of sight) 변화를 조절하는 파라미터가 존재하는데 이를 변경하면서 학습을 수행한다면 다양한 성능을 가진 추격자에 대비할 수 있을 것으로 보인다.

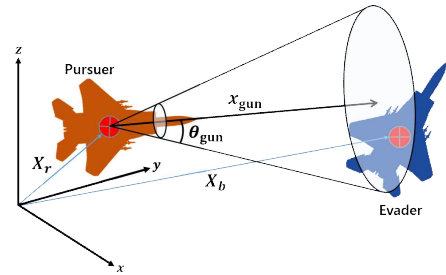


Fig. 1. 3D depiction of 3d cone shaped gunfire range

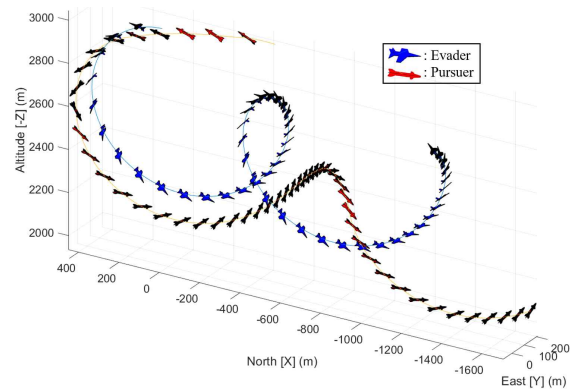


Fig. 2. Sample evasion of a trained agent against lead pursuer

본 연구에서는 선도 추격자에게 교전 우위를 내어주지 않는 회피자의 기동을 적합한 게임 설계와 DDPG 학습 기법을 통해 획득하였으며 교전 궤적 결과를 통해 그 방법이 효과적임을 보였다.

후 기

이 논문은 2021년도 한화시스템(주)의 재원을 지원 받아 수행된 연구임

참고문헌

- 1) Moon, I., Jung, M., Kim, D., "Modeling and Simulation on One-vs-One Air Combat with Deep Reinforcement Learning," *Journal of the Korea Society for Simulation*, Vol. 29, No. 1, 2020, pp. 39-46

2) Lee, D., Bang, H., "Within Visual Range Air Combat of UCAV using Reinforcement Learning," *Proceeding of the 2009 KSAS Conference*, 2009, pp. 249-252.

3) Park, S. J., Park, S. S., Choi, H., An, K., Kim, Y., "An Expert Data-Driven Air Combat Maneuver Model Learning Approach," *AIAA Scitech 2021 Forum*, 2021, pp. 526-537.

4) You, D. I., and Shim, D. H., "Design of an aerial combat guidance law using virtual pursuit point concept," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, Vol. 229, No. 5, 2015, pp. 792-813.

5) Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971, 2015.