УДК 004.8: 004.646

Олексюк О.С., магістрант,
Марчук Г. С., ст. викл.
Державний університет «Житомирська політехніка»

ЗАСТОСУВАННЯ НАВЧАННЯ З ПІДКРІПЛЕННЯМ ДЛЯ ПОБУДОВИ ІНТЕЛЕКТУАЛЬНОГО АВТОПІЛОТА В АВІАСИМУЛЯТОРІ

Розвиток технологій штучного інтелекту відкриває нові можливості для створення автономних систем керування складними динамічними об'єктами. Традиційні методи автоматичного керування, засновані на жорстко запрограмованих правилах або лінійних регуляторах, часто не здатні ефективно діяти в умовах високої невизначеності та нелінійної аеродинаміки. Метою роботи є розробка та теоретичне обґрунтування моделі інтелектуального автопілота для ведення повітряного бою в симуляційному середовищі з використанням методів навчання з підкріпленням.

Задача керування літаком формалізується як Марковський процес прийняття рішень (Markov Decision Process, MDP) і описується п'ятьма базовими компонентами: стани, дії, переходи, винагорода та політика.

Метою агента є максимізація очікуваної сумарної дисконтованої винагороди протягом епізоду взаємодії із середовищем. Процес оптимізації зводиться до пошуку таких вагових коефіцієнтів нейронної мережі, при яких літак надійно відстежує ціль та уникає пошкоджень [1].

Вектор стану агента формується з чотирьох логічних блоків. Блок 1 – параметри власного стану літака. Блок 2 - просторова орієнтація. Блок 3 - відносне положення цілі. Блок 4 - кінематика зближення. Ці параметри консолідуються у єдиний тензор, який обробляється MLP або рекурентною мережею GRU [1].

Керування реалізується через безперервний простір дій у діапазоні [-1, 1], який включає чотири канали: тягу двигуна, крен, тангаж та рискання. Така модель забезпечує плавне і фізично коректне маневрування літака. Правильне проєктування функції винагороди є одним із найважливіших етапів навчання агента.

Сумарна винагорода формується як лінійна комбінація зазначених складових, що стимулює агента оптимізувати траєкторії одночасно за декількома критеріями.

Для подолання проблеми «рідкісних винагород» (sparse rewards) в архітектуру агента інтегровано модуль внутрішньої мотивації (Curiosity-driven exploration), який генерує додаткову внутрішню

винагороду за непередбачуваність стану середовища. Це дозволяє агенту освоїти базову фізику польоту ще до моменту, коли він починає регулярно отримувати зовнішню винагороду за перехоплення цілі.

Практичну верифікацію підходу здійснено у власному симуляційному середовищі на базі Unity / ML-Agents. Автопілот на алгоритмі PPO з модулем Curiosity навчався просторовому перехопленню статичних та рухомих цілей. Конвергенція моделі підтверджується чотирма метриками навчання (рис.1).

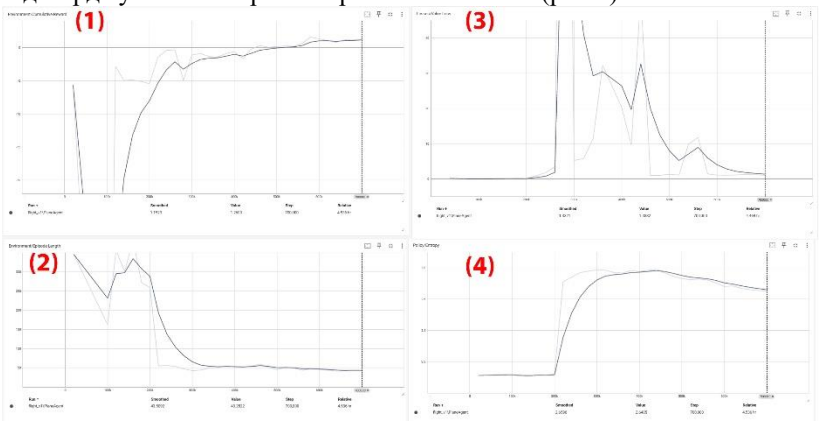

Рис. 1. Динаміка навчання RL-агента (Cumulative Reward, Episode Length, Value Loss та Policy Entropy) у середовищі Unity

Результати дослідження підтверджують ефективність застосування DRL для побудови інтелектуального автопілота. Завдяки ітеративному навчанню у симуляторі алгоритм здійснює керування на мікрорівні з точністю та частотою, недосяжними для людського сприйняття. Запропонований підхід формує потужний фундамент для розробки автопілотів-напарників, здатних брати на себе критичне просторове маневрування та знижувати когнітивне навантаження на пілота. Досягнення цього напрямку знайшли практичне підтвердження у випробуваннях AlphaDogfight (DARPA), де автономні агенти перевершили досвідчених пілотів-людей [2].

Список використаних джерел

1. Hu, D., Yang, R., Zuo, J., Zhang, Z., Wu, J., & Wang, Y. (2021). Application of deep reinforcement learning in maneuver planning of beyond-visual-range air combat. IEEE Access, 9, 32282–32297.

2. Bae, J. H., Jung, H., Kim, S., Kim, S., & Kim, Y. D. (2023). Deep reinforcement learning-based air-to-air combat maneuver generation in a realistic environment. IEEE Access, 11, 26427–26440.