DOI:**10.33099/2311-7249/2023-48-3-115-121**

УДК 004.8

Миколайчук Роман Антонович *(доктор технічних наук, доцент)* [1]
Миколайчук Віра Романівна [2]
Марченко Павло Андрійович [3]

[1] ***Національний університет оборони України, Київ, Україна***
[2] ***Державний університет інформаційно комунікаційних технологій, Київ, Україна***
[3] ***Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна***

ВИКОРИСТАННЯ МЕТОДІВ НАВЧАННЯ З ПІДКРІПЛЕННЯМ ДЛЯ РОЗРОБКИ МОДЕЛІ РОБОТИЗОВАНОГО ЗАСОБУ МОНІТОРИНГУ ІНТЕЛЕКТУАЛЬНИХ ДИНАМІЧНИХ ОБ'ЄКТІВ

Стаття присвячена вдосконаленню моделей роботизованих засобів для ефективного моніторингу інтелектуальних динамічних об'єктів. У контексті розвитку роботизованих систем моніторингу, особливо в умовах, де роботизовані засоби мають справу з інтелектуальними динамічними об'єктами моніторингу, що активно ухиляються від виявлення та ідентифікації, виникає новий виклик, що полягає у потребі розроблення складних моделей, здатних ефективно моделювати роботизовані засоби з метою протидії такій поведінці об'єктів моніторингу. Традиційні підходи до моделювання часто виявляються недостатньо гнучкими для вирішення таких завдань. Мета статті полягає у розробленні моделі роботизованих засобів для забезпечення ефективного моніторингу інтелектуальних динамічних об'єктів на основі використання методів навчання з підкріпленням. Під час проведеного дослідження були застосовані методи навчання з підкріпленням для адаптації поведінки роботизованих систем до динамічно змінюваних умов, комп'ютерного зору та згорткових нейронних мереж для визначення параметрів роботизованих засобів залежно від розташування в середовищі моделювання. Це дає змогу розробити моделі, здатні до самонавчання та самовдосконалення в реальному часі, що є ключовим для ефективного моніторингу та протидії з інтелектуальними динамічними об'єктами. Отримані результати включають розроблення моделі та відповідних алгоритмів, що демонструють здатність до швидкої адаптації, точності прогнозування поведінки об'єктів та ефективної протидії. На основі створеної моделі, був проведений експеримент, у межах якого роботизовані засоби та інтелектуальні об'єкти з випадково заданими параметрами протидіяли у контрольованому середовищі, що дало змогу зробити висновки про працездатність запропонованої моделі. Наукова новизна дослідження полягає у впровадженні сучасних методів навчання з підкріпленням для створення гнучких та адаптивних моделей роботизованих засобів для систем моніторингу, здатних до ефективної протидії з інтелектуальними об'єктами у різних умовах. Теоретична значущість полягає у розширенні розуміння можливостей машинного навчання у сфері робототехніки, а практична значущість – у потенціалі застосування розроблених моделей у військовій та технічній галузях для підвищення ефективності моніторингу та управління. В цілому, стаття розкриває важливі аспекти використання методів навчання з підкріпленням для оптимізації протидії роботизованих засобів з інтелектуальними динамічними об'єктами та надає практичні рекомендації для подальших досліджень у цій області.

Ключові слова: *моніторинг, інтелектуальний динамічний об'єкт, роботизований засіб, штучний інтелект, машинне навчання, навчання з підкріпленням (reinforcement learning), динамічне середовище, моделювання.*

Вступ

Швидкий розвиток інформаційних технологій висуває нові виклики та створює нові можливості у сфері автоматизації та роботизації, зокрема, під час моніторингу рухомих об'єктів. Це особливо актуально у сферах, де потрібно відстежувати рухомі об'єкти, такі як безпілотні літальні апарати (далі – БпЛА), автономні транспортні засоби, надводні (підводні) безпілотні човни тощо, у різних середовищах. Такі роботизовані системи моніторингу мають значний потенціал для підвищення ефективності, безпеки та автономності в різних областях.

Сучасний стан розвитку науки і техніки вимагає нових підходів до моніторингу рухомих об'єктів, не лише через зростаючу потребу в автоматизації процесів, але й через складність взаємодії з об'єктами моніторингу, що можуть демонструвати непередбачувану поведінку та «інтелектуальне» ухилення від зближення з роботизованими засобами зокрема. З огляду на швидкий розвиток технологій у сфері штучного інтелекту та

робототехніки, відкриваються нові можливості для створення більш ефективних систем моніторингу, що можуть знайти застосування за різними напрямами, – від військових операцій до спостереження за дикою природою.

Постановка проблеми. У контексті розвитку роботизованих систем моніторингу, особливо в умовах, де роботизовані засоби (далі – РЗ) мають справу з інтелектуальними динамічними об'єктами моніторингу (далі – ОМ), що активно ухиляються від виявлення та ідентифікації, виникає новий виклик, який полягає у потребі розроблення складних моделей, здатних ефективно моделювати РЗ з метою протидії такій поведінці ОМ. Традиційні підходи до моделювання часто виявляються недостатньо гнучкими для вирішення таких завдань, особливо в умовах, де поведінка ОМ може бути непередбачуваною та змінюватися динамічно.

Центральним аспектом цієї проблеми є розробленя моделі РЗ, що дасть змогу адекватно протидіяти активним маневрам ОМ. Така протидія вимагає від РЗ не лише здатності до швидкого реагування на зміни у поведінці ОМ, але й передбачення їх потенційних маневрів (рухів) для ефективного відстеження та ідентифікації. Одним з можливих підходів для вирішення такого роду проблем є методи навчання з підкріпленням (reinforcement learning (далі – RL)), оскільки вони дають змогу розробляти системи, здатні адаптуватися до динамічних змін середовища та оптимізувати поведінку РЗ на основі протидії з ОМ.

Таким чином, актуальність дослідження полягає у зростаючій потребі в автоматизації та підвищенні ефективності моніторингових систем. Сучасні виклики, такі як необхідність швидкого реагування на непередбачувані зміни у середовищі, вимагають від систем бути більш гнучкими та інтелектуальними. Роботизовані системи, оснащені алгоритмами RL, можуть забезпечити не лише автоматизацію, але й самонавчання та адаптацію до нових умов, що є важливим для ефективного моніторингу.

Аналіз останніх досліджень і публікацій. Питанням використання методів RL присвячена низка робіт. Так, у роботі [1] досліджуються питання використання методів глибокого навчання з підкріпленням для вибору та розміщення об'єктів у переповнених середовищах. У роботі [2] пропонується новий метод захоплення рухомих об'єктів за допомогою активної RGB-камери (камера для запису кольорових зображень), прикріпленої до робота, з використанням глибокого навчання з підкріпленням. Дослідження [3] зосереджено на пошуку цілей рухомих та невидимих об'єктів за допомогою кількох автономних підводних апаратів. Використовується алгоритм глибокого навчання з підкріпленням для покращення пошуку цілей.

У статті [4] розглянуто можливості використання алгоритмів ройового інтелекту при проектуванні систем управління автономних груп БпЛА, визначено, які недоліки сучасних систем можна подолати, наведено загальний алгоритм роботи ройового інтелекту, проведено огляд основних алгоритмів ройового інтелекту. Реалізацію системи нейронечіткого управління групою мобільних роботизованих платформ запропоновано виконувати на підставі проблемно-орієнтованого підходу, що передбачає поєднання програмного (універсального) і апаратного (спеціалізованого) забезпечення, який забезпечує високу ефективність використання обладнання. Вдосконалено метод часового розподілу ресурсів запам'ятовуючого середовища багатопортової пам'яті [5]. Але у роботах [4; 5] проблематика формалізації моделей РЗ не розглядалась.

У роботі [6] пропонується система управління роботизованим маніпулятором на основі глибокого навчання з підкріпленням для точного відстеження рухомих цілей.

На основі аналізу згаданих джерел, можна сформулювати такі висновки. Навчання з підкріпленням демонструє значний потенціал у розробленні роботизованих систем, що можуть адаптуватися до динамічних умов та автономно приймати рішення в реальному часі. Використання RL сприяє створенню гнучких алгоритмів, здатних оптимізувати стратегії поведінки робота для ефективного виявлення та відстеження об'єктів. Роботизовані системи часто працюють на обмежених ресурсах, особливо в контексті обчислювальних потужностей. Інтеграція алгоритмів RL у такі системи вимагає ретельного планування та оптимізації для забезпечення ефективної роботи без перевантаження ресурсів.

Разом із тим, існують невирішені питання, що потребують подальших досліджень, основними з яких є:

адаптивність і гнучкість (системи мають бути здатні адаптуватися до непередбачуваних змін у поведінці цілей, що вимагає високого рівня інтелектуальної автономії);

обчислювальна ефективність (обмежені обчислювальні ресурси вимагають ефективних алгоритмів для швидкого реагування на зміни в середовищі);

комунікація та координація (ефективна комунікація та координація між роботизованими засобами є ключовими для синхронізації дій і ефективного виявлення та ідентифікації цілей).

Тому, розроблення моделі роботизованого засобу моніторингу інтелектуальних динамічних об'єктів є актуальним науковим завданням, що може бути успішно виконане за рахунок застосування методів RL.

Мета статті полягає у розробленні моделі роботизованих засобів для забезпечення ефективного моніторингу інтелектуальних динамічних об'єктів на основі використання методів навчання з підкріпленням.

Виклад основного матеріалу дослідження

Навчання з підкріпленням (RL) – це підхід до машинного навчання, що полягає у навчанні агенту приймати рішення щодо послідовності виконаних дій, з урахуванням взаємодії з середовищем для досягнення максимальної винагороди. Перспективним напрямом RL є глибоке навчання з підкріпленням (далі – Deep RL), що передбачає інтеграцію глибоких нейронних мереж для покращення здатності RL-моделей розпізнавати складні шаблони та адаптуватися до складних середовищ.

У контексті роботизованих систем моніторингу, методи RL можуть бути використані для:

автономної навігації – навчання самостійного пересування РЗ в навколишньому середовищі;

взаємодії з ОМ – ідентифікація та відстеження цільових об'єктів, що ухиляються;

адаптації до динамічних умов – швидка адаптація до змін у середовищі та поведінці об'єктів.

Для побудови моделі РЗ потрібно визначити такі компоненти [7]:

Стан (S). Представляє поточне відображення середовища РЗ. Стан середовища може бути поданий як вектор $s \in S$, де S – простір усіх можливих станів.

Дія (A). Описує конкретні дії, які роботизований засіб може виконати. Дія подається як: $a \in A(s)$, де $A(s)$ – множина можливих дій у стані *s*

Політика (π). Є стратегією або правилом, за яким РЗ обирає дії, базуючись на поточному стані. Політика $\pi(a \mid s)$ визначає ймовірність вибору дії a у стані s.

Функція винагороди (R). Визначає винагороду або покарання за виконання певної дії в певному стані. $R(s,a)$ визначає винагороду за виконання дії a у стані s.

Функція переходу станів (P). Визначає ймовірність переходу з одного стану в інший після виконання певної дії. $P(s' \mid s,a)$ визначає ймовірність переходу в стан s' зі стану s після виконання дії a. Задача навчання з підкріпленням полягає у навчанні агенту (у нашому випадку, роботизованого засобу) вибору оптимальних дій в середовищі для максимізації загальної винагороди протягом часу. Агент взаємодіє із середовищем, яке реагує на його дії та змінює свій стан, надаючи агенту винагороду або покарання.

Функція цінності (V). Оцінює очікувану суму майбутніх винагород, починаючи з певного стану s, слідуючи політиці π. $V^{\pi}(s)$ визначає цінність стану s при політиці π. Зазвичай формується на основі рівняння Беллмана (1):

$$V^{\pi}(s) = \sum_{a \in A(s)} \pi(a \mid s) \times \sum_{s' \in S} P(s' \mid s,a)[R(s,a) + \gamma V^{\pi}(s')], \quad (1)$$

де γ – коефіцієнт дисконтування, що визначає важливість майбутніх винагород.

Ціль агента – знайти оптимальну політику π^*, яка максимізує очікувану суму винагород:

$$\pi^* = \arg\max_{\pi} \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R_t \mid \pi] \quad (2)$$

де R_t – винагорода в час t,

γ – коефіцієнт дисконтування (зазвичай між 0 та 1), що визначає важливість майбутніх винагород.

Для оцінювання та оптимізації дій агента в різних станах використовується Q-функція $Q^{\pi}(s,a)$, що визначає очікувану суму винагород за вибір дії a у стані s та слідування політиці π після цього:

$$Q^{\pi}(s,a) = \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid s_t = s, a_t = a, \pi] \quad (3)$$

Рівняння Беллмана для Q-функція матиме такий вигляд:

$$Q^{\pi}(s,a) = R(s,a) + \gamma \sum_{s' \in S} P(s' \mid s,a) \max_{a'} Q^{\pi}(s',a'), \quad (4)$$

де $P(s' \mid s,a)$ – ймовірність переходу в стан s' після виконання дії a у стані s.

Наведені відповідно до [7] математичні залежності, складають основу моделі, що дає змогу РЗ адаптуватися до змін у середовищі та поведінці об'єктів, оптимізуючи свої дії для досягнення поставлених цілей. Надалі надаються результати досліджень стосовно визначення складових залежностей (1–4).

Для опису та математичного формулювання середовища, в якому діють РЗ, використовується інтегрований підхід, заснований на початкових даних карти місцевості, отриманих з відкритих джерел, та на даних, що доповнюються в процесі діяльності РЗ.

Початкова конфігурація середовища створюється на основі карти місцевості M, за рахунок використання відкритих геоданих для створення базової карти місцевості.

Карта визначається як матриця $M = [m_{ij}]$, де кожен елемент m_{ij} репрезентує ділянку місцевості з певними атрибутами або параметрами P (наприклад, тип поверхні, перешкоди). Параметри місцевості подаються у формі $P = \{p_1, p_2, \dots, p_n\}$, де кожен p_i є характеристикою місцевості.

У процесі функціонування роботизованої системи моніторингу буде здійснюватися доповнення середовища на основі даних, отриманих під час функціонування РЗ – D_{RZ}, шляхом внесення відповідних змін до M та P: $A(M, D_{RZ}) \rightarrow M'$, де M' – оновлена карта місцевості.

Таким чином, у процесі навчання та адаптації РЗ використовується динамічне середовище, що неперервно оновлюється, для моделювання реальних умов та поліпшення стратегій взаємодії. Такий підхід дає змогу створити адаптивне та гнучке середовище, що відображає реальні умови місцевості та забезпечує ефективне навчання і функціонування РЗ. Важливим елементом є здатність середовища динамічно адаптуватися на основі зібраних РЗ даних, що дає змогу забезпечити актуальність та релевантність моделі середовища.

Для визначення параметрів РЗ відповідно до їх

місцезнаходження в середовищі, використовувались методи комп'ютерного зору (computer vision (далі – CV)) та згорткові нейронні мережі (convolutional neural networks (далі – CNN)):

Використовуються дані, отримані під час експлуатації РЗ, що можуть включати зображення, сенсорні дані та іншу інформацію про середовище. Для кожного типу ділянки (наприклад, міська місцевість, лісові масиви, водойми) заздалегідь визначаються типові параметри РЗ, такі як швидкість, енергоспоживання, ефективність сенсорів. Припустимо, що кожній ділянці d відповідає набір параметрів $P_d = \{p_1, p_2, \dots, p_n\}$, де p_i відображає конкретний параметр РЗ.

Тоді за допомогою CNN можна провести навчання нейронної мережі розпізнаванню різних типів ділянок місцевості та визначати на цій основі відповідні параметри РЗ за рахунок згортки параметрів отриманого зображення до типової ділянки місцевості $f_{CNN}(D_{RZ}) \to d$, де d – тип ділянки середовища. В подальшому, на основі визначеного типу ділянки d обчислюються відповідні параметри P_d. Такий підхід дає змогу здійснювати динамічне оновлення параметрів РЗ залежно від змін середовища та адаптувати поведінку РЗ залежно від оновлених параметрів, що забезпечує оптимальне використання ресурсів та ефективне виконання завдань у різних умовах середовища.

Такий підхід дає змогу РЗ адаптуватися до різних умов середовища, використовуючи передові технології комп'ютерного зору та машинного навчання для аналізу та відповідного налаштування їх параметрів. Використання CNN для аналізу зображень дає змогу точно визначати тип місцевості та адаптувати поведінку РЗ згідно з вимогами конкретної ділянки.

Під час моделювання розглядаються основні типи дій РЗ, що включають переміщення в середовищі, поворот, зупинку та зміну швидкості руху. За потреби множину дій може бути розширено.

Припустимо, що стан РЗ у середовищі описується як $S = (x, y, \theta)$, де x, y – координати РЗ, а θ – орієнтація. Тоді вплив дій РЗ на зміну стану S' можна формалізувати так:

Переміщення:

$$x' = x + d \cdot \cos(\theta)$$
$$y' = y + d \cdot \sin(\theta)$$
$$S' = (x', y', \theta), \quad (5)$$

де d – відстань переміщення;

Поворот:

$$\theta' = \theta + \phi$$
$$S' = (x, y, \theta'), \quad (6)$$

де ϕ – кут повороту.

Зупинка: $S' = S$ (без зміни стану). Зміна параметра швидкості РЗ, не впливає безпосередньо на S, але змінює динаміку переміщення.

Кожна дія A викликає зміну стану РЗ у середовищі відповідно до функції середовища $F_{\text{env}}(S, A) \to S'$.

Вибір дії A залежить від поточного стану S, цілей РЗ та умов середовища.

Цей підхід дає змогу формалізувати і регулювати поведінку РЗ у середовищі, забезпечуючи ефективне планування траєкторій та адаптивність до змін умов. Використання математичних моделей для представлення дій дає змогу точно моделювати поведінку РЗ і робить можливим використання складних алгоритмів навчання та оптимізації.

Політика агента є набором правил або стратегій, які визначають дії РЗ залежно від стану середовища та динаміки ОМ. Ця політика має на меті оптимізувати певні цілі, наприклад, ефективність моніторингу або мінімізацію часу для досягнення завдань. Політика π може бути визначена як функція, що відображає стан середовища S у дії A, які має виконати РЗ $\pi: S \to A$, де S – це вектор стану, що включає інформацію про розташування та орієнтацію РЗ, а також про стан та динаміку ОМ в даній місцевості.

Інтелектуальні динамічні ОМ постійно змінюють стан середовища, що вимагає від РЗ адаптації своєї стратегії, а саме:

реакція на зміни. Політика РЗ має включати механізми для адаптації до змін у поведінці ОМ;

прогнозування дій ОМ. Використання методів прогнозування потенційних змін у поведінці ОМ.

Використання навчання з підкріпленням на основі політики агента дає змогу РЗ ефективно адаптуватися до динамічних змін у поведінці інтелектуальних ОМ, забезпечуючи гнучкість та високу реактивність системи моніторингу.

Функція винагороди в контексті навчання з підкріпленням для РЗ враховує час і ймовірність наближення до ОМ та запобігання виходу ОМ за межі зони моніторингу, може бути описана й формалізована так. Функція винагороди $R(s, a)$ визначає винагороду або штраф, що отримує роботизований засіб за виконання дії a в стані s. Ця функція оцінює ефективність дій РЗ у контексті двох основних цілей:

наближення до ОМ: РЗ отримує позитивну винагороду за успішне наближення до ОМ, що збільшується з підвищенням ймовірності виявлення ОМ;

запобігання виходу ОМ із зони моніторингу: РЗ отримує штраф, якщо ОМ виходить за межі зони моніторингу.

Припустимо, що:

P_{approach} – ймовірність успішного наближення РЗ до ОМ.

T_{approach} – час, необхідний РЗ для наближення до ОМ.

P_{exit} – ймовірність виходу ОМ за межі зони моніторингу.

Тоді функція винагороди може бути визначена за виразом:

$$R(s, a) = w_1 f\left(P_{\text{approach}}, T_{\text{approach}}\right) - w_2 g\left(P_{\text{exit}}\right), \quad (7)$$

де $f(P_{\text{approach}}, T_{\text{approach}})$ – функція, що оцінює винагороду за наближення до ОМ та може включати час та ймовірність наближення;

$g(P_{\text{exit}})$ – функція, що оцінює штраф за вихід ОМ за межі моніторингу;

w_1, w_2 – вагові коефіцієнти, що визначають важливість кожної цілі.

Для практичного моделювання поведінки РЗ доцільно використовувати TF-Agents у межах фреймворку TensorFlow. Процес моделювання включає встановлення необхідних бібліотек TensorFlow та TF-Agents. Після цього створюється відповідне середовище, що відображає умови, в яких буде функціонувати РЗ, включно з типами дій і станом РЗ. Дані, отримані за допомогою CV, та інтеграція моделей CNN допомагає у визначенні стану РЗ на основі зображень.

Наступним кроком є налаштування агента TF-Agents, що найкраще відповідає його завданням та умовам середовища, у якому оперує РЗ. Агент навчається на основі взаємодій з середовищем, збираючи досвід через серію дій та відповідних винагород.

Процес навчання агента включає збір даних про взаємодію, оптимізацію політики на основі цих даних та регулярне оновлення політики агента для покращення його ефективності. Після навчання важливо оцінити ефективність агента, запустивши його в середовищі та проаналізувавши отримані результати. Це дає змогу визначити, наскільки добре агент адаптується до змін умов та виконує поставлені задачі. Для послідовного тренування ОМ та РЗ можна використати підхід навчання з підкріпленням, де обидва типи агентів – ОМ і РЗ – тренуються за допомогою послідовностей реалізації, що використовують протилежні функції винагород:

ініціалізація ОМ. Об'єкти моніторингу спочатку тренуються без протидії з боку РЗ. Наступне завдання – навчитися уникати ідентифікації, за таких умов функція винагороди має стимулювати уникати РЗ та досягнути межі зони моніторингу;

тренування ОМ. Об'єкти моніторингу тренуються за допомогою обраного алгоритму навчання з підкріпленням, намагаючись максимізувати свою винагороду.

Після досягнення певного рівня вмінь ОМ, розпочинається тренування РЗ, що проводиться подібно до тренування ОМ із протилежною функцією винагороди.

Процес тренування містить декілька ітерацій:

оцінка та адаптація. Після кожного циклу тренування проводиться оцінка ефективності та адаптація стратегій як РЗ, так і ОМ;

моніторинг параметрів. За результатами ітерації визначаються ймовірність і час ідентифікації ОМ;

критерії завершення. Тренування продовжується доти доки параметри не стабілізуються.

Такий підхід дає змогу створити балансовану взаємодію між ОМ та РЗ, де обидва типи агентів оптимізують свої стратегії у відповідь на дії один одного. Використання протилежних функцій винагород дає змогу симулювати реалістичну взаємодію, де ОМ намагаються уникати, а РЗ здійснити ідентифікацію.

Отже, під час досліджень було сформовано модель роботизованого засобу моніторингу інтелектуальних динамічних об'єктів, що формалізована виразами (2–7). Моделювання фокусується на протидії між РЗ та ОМ у динамічному середовищі. Суть моделі полягає у використанні методів навчання з підкріпленням, де РЗ та ОМ оптимізують свої стратегії відповідно до функцій винагороди, що враховують ймовірність ідентифікації та час. Модель призначена для створення ефективного механізму моделювання моніторингу та протидії між РЗ та ОМ, з можливістю адаптації до змінних умов середовища. Послідовність реалізації моделі є ітеративним процесом, що включає налаштування середовища, тренування агентів, оцінку їхньої взаємодії та адаптацію стратегій.

Розроблена модель є корисною для розроблення систем моніторингу, де важлива адаптивність до змін умов та ефективне взаємодіюче навчання РЗ та ОМ.

Для оцінювання працездатності розробленої моделі РЗ проведено експеримент, що полягав у моделюванні протидії між РЗ та ОМ, де параметри визначалися випадковим чином. Використовуючи фреймворк TensorFlow як основний інструмент для обчислень та аналізу даних, експеримент проводився протягом 20 ітерацій. Кожна ітерація включає оцінку взаємодії між РЗ та ОМ, аналізуючи їх поведінку, стратегії руху, та реакції на зміну умов. Це дало змогу сформувати та дослідити ефективність різних стратегій протидії в динамічному середовищі. Результати проведених експериментальних досліджень наведено на рис. 1, 2.

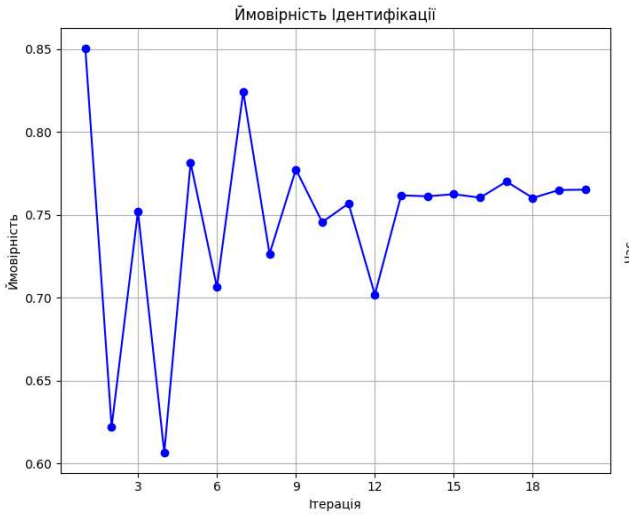

Рисунок 1 – Результати експериментального визначення ймовірності ідентифікації.

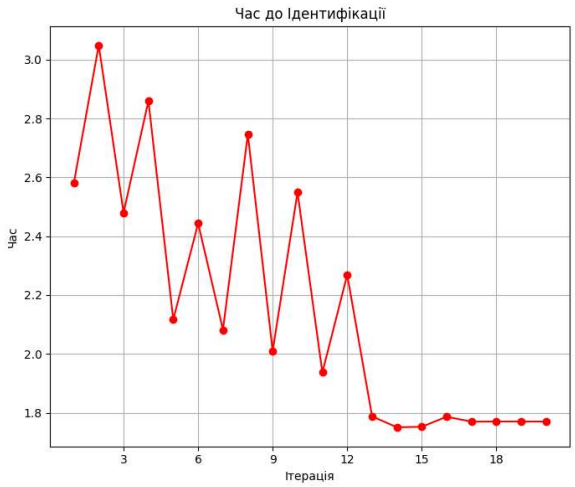

Рисунок 2 – Результати експериментального визначення часу ідентифікації

Отримані результати експериментів дають змогу підбити такі підсумки:

модель демонструє високу точність у відстеженні ОМ, особливо у складних умовах;

РЗ швидко адаптуються до змін у поведінці ОМ та умовах середовища;

модель показує хорошу стійкість до помилок та непередбачуваних ситуацій.

Ці результати свідчать про високу ефективність та адаптивність розробленої моделі, що робить її перспективною для застосування в реальних умовах застосування роботизованих засобів моніторингу інтелектуальних динамічних об'єктів.

Висновки й перспективи подальших досліджень

На основі проведених досліджень можна зробити висновок, що ціль статті, яка полягала у розробленні моделей роботизованих засобів для ефективного моніторингу інтелектуальних динамічних об'єктів на основі використання методів навчання з підкріпленням, була досягнута. Експериментальне моделювання взаємодії між роботизованим засобом і інтелектуальним динамічним об'єктом з випадково заданими параметрами протягом 20 ітерацій демонструє значний потенціал у використанні алгоритмів машинного навчання та глибокого навчання для адаптації та оптимізації поведінки роботизованих систем. Результати свідчать, що такі системи можуть ефективно адаптуватися до змінних умов і виконувати задачі моніторингу з високою точністю та ефективністю.

Наукова новизна дослідження полягає у впровадженні передових методів RL для створення гнучких і адаптивних моделей роботизованих систем, здатних до ефективної взаємодії з інтелектуальними об'єктами у змінних умовах. Теоретична значущість полягає у розширенні розуміння можливостей машинного навчання у сфері робототехніки, а практична значущість – у потенціалі застосування розроблених моделей у військовій і технічній галузях для підвищення ефективності моніторингу та управління.

Подальші дослідження можуть бути спрямовані на розроблення та впровадження більш складних алгоритмів адаптивного навчання, що дають змогу роботизованим засобам краще пристосовуватися до динамічних змін у поведінці інтелектуальних об'єктів. Також важливим напрямом подальших досліджень є вдосконалення моделей протидії між роботизованими засобами та інтелектуальними об'єктами, зокрема, підвищення точності прогнозування та реакції на непередбачувані зміни. Інтеграція розроблених моделей з іншими передовими технологіями, такими як Інтернет речей (IoT) та штучний інтелект, а також проведення експериментів у реальних умовах для перевірки ефективності моделей у різних сценаріях, включно з міським середовищем, промисловими об'єктами та іншими умовами, може значно розширити можливості моніторингу.

Список бібліографічних посилань

1. Imtiaz M. B., Qiao Y. and Lee B. Prehensile and non-prehensile robotic pick-and-place of objects in clutter using deep reinforcement learning. *Sensors*. 2023. Т. 23. № 3. С. 1513. DOI: https://doi.org/10.3390/s23031513. **2. Seongwon J., Hyemi J. and Hyunseok Y.** Vision-based reinforcement learning: moving object grasping with a single active-view camera. *2022 22nd international conference on control, automation and systems (ICCAS)*. 232–237. DOI: 10.23919/ICCAS55662.2022.10003899. **3. Wang G., Wei F., Jiang Y., Zhao M., Wang K. and Qi H.** A multi-auv maritime target search method for moving and invisible objects based on multi-agent deep reinforcement learning. *Sensors*. 2022. Т. 22. № 21. С. 8562. DOI: https://doi.org/10.3390/s22218562. **4. Каратанов О. В., Устименко О. В., Єна М. В., Бова Є. А., Калашнікова В. І.** Використання алгоритмів ройового інтелекту при проектуванні систем управління груп безпілотних літальних апаратів. Харків : Молодий вчений 2021. № 10 (98). С. 98–103. **5. Цмоц І. Г., Опотяк Ю. В., Штогрінець Б. В., Дзюба А. О., Олійник Ю. Ю.** Базова структура нейронечіткої системи керування групою мобільних роботизованих платформ. *Український журнал інформаційних технологій*. 2023. Том 5. № 1. С. 77–85. **6. Yingqi L., Xiaomei W. and Ka-Wai K.** Towards adaptive continuous control of soft robotic manipulator using reinforcement learning. *2022 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2022. P. 7074–7081. DOI: https://doi.org/10.1109/IROS47612.2022.9981335. **7. Kuang W.** Fundamentals of reinforcement learning. Texas : University of Texas Rio Grande Valley. 2021. 111 p. URL: https://faculty.utrgv.edu/weidong.kuang/book/RL.pdf (accessed: 15 September 2023).

USING REINFORCEMENT LEARNING METHODS TO DEVELOP A MODEL OF A ROBOTIC MEANS OF MONITORING INTELLIGENT DYNAMIC OBJECTS

Mykolaichuk Roman *(Doctor of Technical Sciences, Associate Professor)* [1]
Mykolaichuk Vira [2]
Marchenko Pavlo [3]

[1] ***The National Defence University of Ukraine, Kyiv, Ukraine***
[2] ***State University of Information and Communication Technologies, Kyiv, Ukraine***
[3] ***National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine***

The article is devoted to the improvement of models of robotic means for effective monitoring of intelligent dynamic objects. In the context of the development of robotic monitoring systems, especially in conditions where robotic means deal with intelligent dynamic monitoring objects that actively evade detection and identification, a new challenge arises, which is the need to develop complex models capable of effectively simulating robotic means with in order to counteract such behavior of monitoring objects. Traditional modeling approaches are often not flexible enough to solve such tasks.

Formulation of the problem in general. *The purpose of the article is to improve existing models of robotic means to ensure effective monitoring of intelligent dynamic objects.*

Analysis of recent researches and publications *In the course of the research, reinforcement learning methods were used to adapt the behavior of robotic systems to dynamically changing conditions, computer vision and convolutional neural networks to determine the parameters of robotic tools depending on the location in the simulation environment. This made it possible to develop models capable of self-learning and self-improvement in real time, which is key for effective monitoring and countermeasures with intelligent dynamic objects.*

Presenting the main material *The obtained results include the development of a model and corresponding algorithms that demonstrate the ability to quickly adapt, accurately predict the behavior of objects and effective countermeasures. Based on the created model, an experiment was conducted in which robotic means and intelligent objects with randomly set parameters interacted in a controlled environment, which allowed us to draw conclusions about the performance of the proposed model.*

Elements of scientific novelty. *The scientific novelty of the research consists in the introduction of modern RL methods to create flexible and adaptive models of robotic means for monitoring systems capable of effective countermeasures with intelligent objects in various conditions.*

Practical significance of the article *lies in expanding the understanding of the possibilities of machine learning in the field of robotics, and the practical significance lies in the potential of applying the developed models in the military and technical fields to increase the effectiveness of monitoring and management.*

Conclusion and the perspectives of future researches *Overall, the paper reveals important aspects of using RL methods to optimize the interaction of robotic vehicles with intelligent dynamic objects and provides practical recommendations for further research in this area.*

Keywords*: monitoring, intelligent dynamic object, robotic tool, artificial intelligence, machine learning, reinforcement learning, dynamic environment, simulation.*

References

1. Imtiaz, M. B., Qiao, Y., Lee, B., (2023)**.** Prehensile and non-prehensile robotic pick-and-place of objects in clutter using deep reinforcement learning. *Sensors*. 23 (3), 1513. DOI: 10.3390/s23031513. **2**. **Seongwon, J., Hyemi, J., and Hyunseok, Y.,** (2022)**.** Vision-based reinforcement learning: moving object grasping with a single active-view camera. *2022 22nd international conference on control, automation and systems (ICCAS)*. 232–237. DOI: 10.23919/ICCAS55662.2022.100038993.9981335.

3. Wang, G, Wei, F., Jiang, Y., Zhao, M., Wang, K. and Qi, H., (2022). A multi-auv maritime target search method for moving and invisible objects based on multi-agent deep reinforcement learning. *Sensors*. 22(21), 8562. DOI: 10.3390/s22218562. **4. Karatanov, O. V., Ustymenko, O. V., Yena, M. V., Bova, Ye. A., Kalashnikova, V. I.,** (2021). The use of swarm intelligence algorithms in the design of control systems for groups of unmanned aerial vehicles. Kharkiv : Molodyy vchenyy, 10 (98), 98–103. **5. Tsmots, I. G., Opotyak, Yu. V., Shtogrinets, B. V., Dzyuba, A. O., Oliynyk, Yu. Yu.,** (2023). The basic structure of a neurofuzzy control system for a group of mobile robotic platforms. *Ukrayins'kyy zhurnal informatsiynykh tekhnolohiy,* 5, 1, 77–85. **6. Yingqi, L., Xiaomei, W. and Ka-Wai, K.,** (2022). Towards adaptive continuous control of soft robotic manipulator using reinforcement learning. *2022 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 7074–7081. DOI: 10.1109/IROS47612.2022. **7. Kuang, W.,** (2021). Fundamentals of reinforcement learning. Texas: University of Texas Rio Grande Valley [online]. Available at: https://faculty.utrgv.edu/weidong.kuang/book/RL.pdf [Accessed : 25 September 2023].