УДК 62-932:004.89 DOI: 10.30977/BUL.2219-5548.2025.108.0.78

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ В ІНТЕЛЕКТУАЛЬНІЙ СИСТЕМІ КЕРУВАННЯ КОВШЕМ НАВАНТАЖУВАЧА

Кириченко І.Г., Гурко В.О.
Харківський національний автомобільно-дорожній університет

Анотація. *Проаналізовано сучасні підходи щодо автоматичного керування фронтальними навантажувачами. Розглянуто перспективи застосування методів машинного навчання для створення інтелектуальної системи керування ковшем фронтального навантажувача. Особливу увагу приділено методам навчання з підкріпленням як перспективному інструменту для адаптивного керування навантажувачем в умовах невизначеності.*

Ключові слова*: фронтальний навантажувач, наповнення ковша, інтелектуальна система керування, машинне навчання, навчання з підкріпленням.*

Вступ

Підйомно-транспортні машини відіграють важливу роль у багатьох галузях промисловості. Вони використовуються на будівельних майданчиках, у гірничодобувній промисловості, сільському господарстві та на складах, забезпечуючи високу продуктивність робіт і зменшуючи фізичне навантаження на персонал. Серед цих машин особливо популярними є фронтальні навантажувачі (ФН), які завдяки своїй маневреності та багатофункціональності широко застосовуються для виконання завдань з навантаження, розвантаження та транспортування різноманітних матеріалів [1].

Сучасні тенденції розвитку ФН спрямовані на підвищення рівня їх автоматизації, що дає змогу не лише зменшити навантаження на оператора й забезпечити його безпеку, а й покращити продуктивність і енергоефективність машин за різних умов експлуатації [2–4].

Активний розвиток систем автоматизації ФН вимагає узагальнення підходів щодо їх побудови. Ця робота присвячена аналізу основних тенденцій у створенні та використанні систем автоматичного керування ФН.

Аналіз публікацій

Автоматизація ФН та інших підйомно-транспортних і будівельних машин є актуальним напрямом досліджень упродовж останніх трьох десятиліть. Загалом у процесі автоматизації ФН можна визначити п’ять етапів [4, 5]:

1) традиційний ФН з повністю ручним керуванням, коли оператор самостійно оцінює обстановку, приймає рішення та виконує всі дії щодо керування навантажувачем;

2) наявність асистентів, що за допомогою людино-машинного інтерфейсу рекомендують виконати ті або інші дії для підвищення продуктивності або економії палива;

3) ФН з напівавтоматичними системами, такими як круїз-контроль та система автоматичного заповнення ковша, коли частина робочого циклу повністю виконується самим навантажувачем з оптимальною продуктивністю або економічністю. Інші операції робочого циклу оператор здійснює самостійно за допомогою систем допомоги оператору, як на другому етапі;

4) напівавтономний ФН, де оператор вводить до системи керування ФН робоче завдання, а потім машина виконує завдання оптимальним способом, поки оператор не додасть наступне завдання;

5) повністю автономний навантажувач, що працює без участі оператора після того, як робоче завдання визначено «системним оператором», який керує кількома машинами на відстані з офісу.

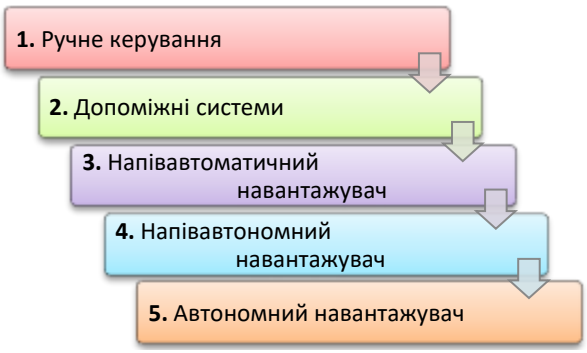

Рис. 1. Етапи автоматизації підйомно-транспортних і будівельних машин

Нині вчені активно досліджують 3 та 4-й етапи. Оскільки повна автоматизація ФН є

складним завданням, науковці зазвичай розглядають окремі завдання автоматизації.

Низка студій з автоматизації ФН присвячена розробленню систем підтримки оператора, які допомагають йому виконувати робочий процес з максимальною енергоефективністю [5, 6].

Під час руху від забою до місця розвантаження матеріалу ФН зазвичай рухається V-подібною траєкторіею [7]. Метою планування траєкторії є створення оптимальної кривої з огляду на початкове положення машини, розташування самоскида та наявності перешкод. Критеріями оптимальності можуть бути мінімізація відстані або витрати пального [8, 9].

Після цього постає завдання реалізації автономного руху навантажувача за заданою траєкторією. Це завдання розв'язується різними способами. Наприклад, у праці [10] застосовується класичний метод стеження за стіною, що може бути корисним у роботі навантажувачів у шахтах. У студіях [4, 11] для керування рухом ФН впроваджуються методи сучасної теорії керування.

Планування шляху й керування рухом ФН вимагають наявності оперативної інформації про стан робочого середовища, локалізацію машини в цьому середовищі, етапи виконання робочого процесу, стан основних параметрів машини тощо. З цією метою застосовуються різноманітні вимірювальні пристрої та системи: супутникові та інерціальні навігаційні системи, відеокамери, лазери, лідари, датчики переміщення штоків гідроциліндрів тощо. Кількість вимірювальних систем на борту та зовні ФН постійно збільшується, тому вивчаються способи оптимізації цих систем [12, 13].

Однією з основних операцій робочого процесу ФН є завантаження ковша матеріалом. Деякі автори намагалися реалізувати систему автоматичного наповнення ковша, використовуючи керування на основі фізичної моделі його взаємодії з палею [14–16]. Проте через складний характер цієї взаємодії та значну кількість невизначених збурень використання лише модельно-орієнтованих підходів може бути недостатньо ефективним [17].

У зв'язку з цим все більш поширеними є дослідження, у яких застосовуються методи, що не ґрунтуються на фізичних моделях. Зокрема завдяки розвитку штучного інтелекту в цих підходах почали активно використовувати нейронні мережі [17–20].

Мета й постановка завдань

Аналіз літератури показав, що методи штучного інтелекту, зокрема машинного навчання, є перспективним напрямом у створенні систем керування ФН, зокрема для наповнення ковша.

Метою роботи є аналіз технологій машинного навчання щодо їх придатності для створення інтелектуальної системи керування (ІСК) ковшем ФН.

Машинне навчання та його види

Машинне навчання (МН) є потужним інструментом для створення ІСК. Такі системи здатні адаптувати свої алгоритми до мінливих умов експлуатації: властивостей матеріалу, робочого середовища або параметрів руху машини. На відміну від традиційних методів керування, основаних на жорстко заданих правилах, МН не вимагає явного програмування всіх можливих сценаріїв, а аналізує накопичену інформацію для виявлення закономірностей, що робить його особливо ефективним для роботи в умовах невизначеності.

МН є частиною галузі штучного інтелекту [21], але зосереджене на розробленні методів автоматичного вилучення знань із даних і на механізмах самостійного навчання системи, тоді як штучний інтелект охоплює значно ширший спектр застосувань, таких як комп'ютерний зір, оброблення природної мови, роботизована автоматизація процесів тощо (рис. 2).

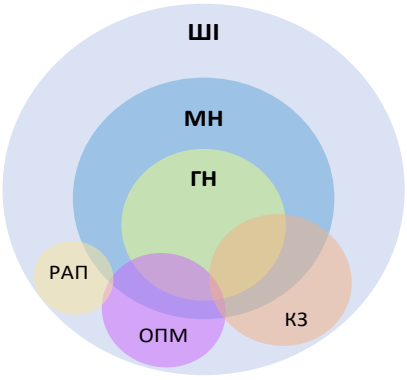

Рис. 2. Основні технології штучного інтелекту [21]: ГН – глибоке навчання; ОПМ – оброблення природної мови; КЗ – комп'ютерний зір; РАП – роботизована автоматизація процесів

Виокремлюють три основних типи МН, що розрізняються підходами до опрацювання інформації [22]. До них належать:

- навчання з учителем, де модель навчається на розмічених даних;

- навчання без учителя, де модель виявляє приховані структури в даних без явних міток;

- навчання з підкріпленням, де система навчається на основі взаємодії з навколишнім середовищем, отримуючи зворотний зв'язок у вигляді нагород або штрафів.

Навчання з учителем може застосовуватися для ІСК завантаженням ковша, але його ефективність обмежена доступністю розмічених даних. Інформація для навчання має містити дані про вхідні параметри (на основі яких приймаються рішення), так і про вихідні керуючі впливи. Вхідною інформацією є положення ковша, швидкість руху навантажувача, навантаження на ківш, орієнтація машини, характеристики палі (тип матеріалу, висоту) та інші параметри. Вихідні дані – це керуючі сигнали, що подаються на гідроциліндри ковша (рис. 3).

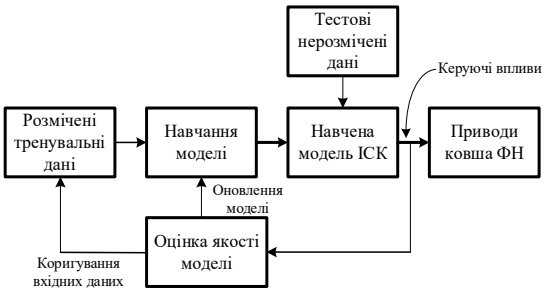

Рис. 3. Процес навчання з учителем

Джерелом даних може бути інформація, отримана з датчиків під час роботи навантажувача під керуванням досвідченого оператора, результати комп'ютерного моделювання або експертні оцінки. Однак ефективність навчання з учителем безпосередньо залежить від кількості та якості розмічених даних, що мають охоплювати роботу навантажувача за різних умов.

Якщо якість запропонованих ІСК рішень є незадовільною, необхідно або покращити тренувальні дані (збільшити їх обсяг, поліпшити їх якість), або вдосконалити алгоритми навчання моделі ІСК (рис. 3).

Алгоритми керування, навчені з використанням цього методу, можуть виявитися недостатньо адаптивними. Це пов'язано не лише з недостатністю інформації, а й з тим, що розмітка керуючих впливів може бути суб'єктивною і залежати від стилю роботи оператора (якщо дані отримано з реальної машини або від експерта). У комп'ютерному моделюванні точність моделі залежить від коректності закладених закономірностей, а, як уже було зазначено, створення достатньо точних моделей роботи ФН під час заповнення ковша є складним завданням.

Навчання без учителя дає змогу знаходити закономірності у вхідних даних без потреби в розміченій вибірці. У цьому підході модель аналізує особливості робочого процесу ФН і виявляє характерні шаблони поведінки без явного визначення вихідних керуючих впливів (рис. 4).

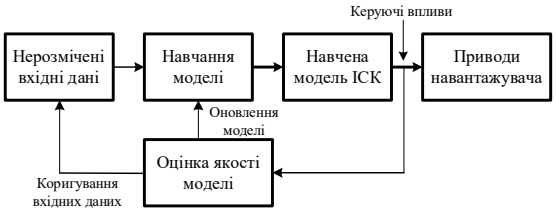

Рис. 4. Процес навчання без учителя

Вхідні дані, як і за умови навчання з учителем, можуть містити положення ковша, навантаження на ківш, швидкість руху та орієнтацію навантажувача, характеристики палі тощо. Однак, на відміну від навчання з учителем, вихідні сигнали не задаються – алгоритм має самостійно згрупувати інформацію або виявити приховані залежності між параметрами.

Застосування навчання без учителя дає змогу аналізувати широкий спектр сценаріїв і адаптувати модель до нових умов. Водночас складність інтерпретації знайдених закономірностей може ускладнювати практичне використання отриманих рішень.

Алгоритми керування, отримані за цим методом, часто є гнучкішими порівняно з навчанням з учителем, оскільки не залежать від суб'єктивної розмітки даних. Проте їх ефективність визначається точністю виявлених закономірностей і відповідністю побудованих моделей реальним умовам роботи ФН.

Навчання з підкріпленням дає змогу ІСК завантаження ковша самостійно знаходити оптимальні керуючі дії на основі принципу винагороди та штрафу. У такому підході агент (тобто ІСК) взаємодіє з навколишнім середовищем, отримуючи інформацію про стан машини та параметри її робочого процесу, випробовує різні керуючі стратегії та коригує їх на основі зворотного зв'язку (рис. 5).

На відміну від навчання з учителем, де оптимальні дії задаються заздалегідь, або навчання без учителя, яке лише знаходить приховані закономірності, метод підкріплення навчається шляхом проб і помилок.

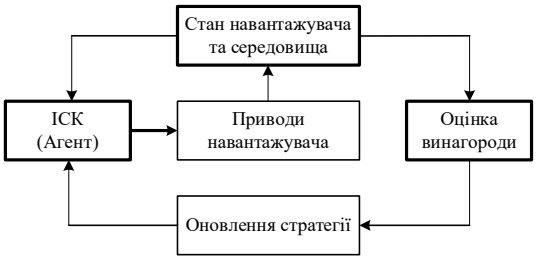

Рис. 5. Процес навчання з підкріпленням

Агент отримує винагороду або штраф залежно від ефективності своїх дій: наприклад, максимальне заповнення ковша за мінімальних витрат енергії може бути оцінене позитивно, тоді як надмірне буксування ФН – негативно.

Навчання з підкріпленням дає змогу отримати більш гнучку систему керування, здатну пристосовуватися до змінних умов роботи ФН [19, 22–26].

Однак для визначення оптимальної стратегії керування може знадобитися значна кількість ітерацій, що є обмеженням під час швидкої зміни властивостей матеріалу та інших умов роботи ФН.

Методи машинного навчання з підкріпленням

Як вже було зазначено, навчання з підкріпленням ґрунтується на принципі взаємодії агента із середовищем, де кожна дія оцінюється за допомогою системи винагород і штрафів (рис. 6).

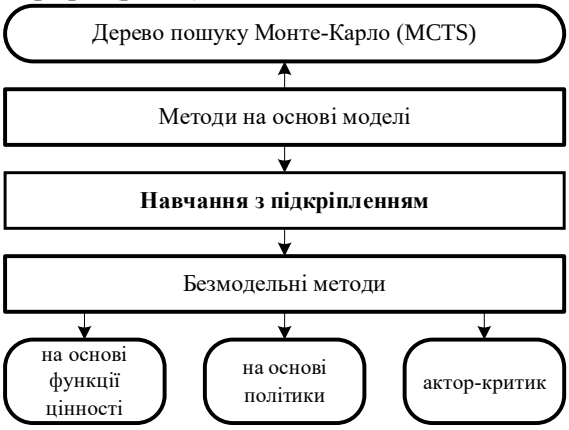

Рис. 6. Методи навчання з підкріпленням

Це дає змогу системі самостійно знаходити оптимальні стратегії керування. Проте в межах навчання з підкріпленням існують різні підходи до навчання та подання стратегії керування.

Більшість цих підходів не потребує моделі навколишнього середовища, хоча існують і алгоритми, основані на моделях. Деякі підходи орієнтовані на оцінку цінності станів або дій, інші безпосередньо шукають оптимальну політику, а деякі поєднують обидва підходи для досягнення кращих результатів.

Методи на основі функції цінності

Підходи цього типу спрямовані на визначення корисності перебування агента в певному стані або виконання конкретної дії. Одним із найпоширеніших алгоритмів є Q-навчання, що ґрунтується на оновленні значень функції цінності Q (Q від англ. *quality* – якість), яка визначає очікувану суму майбутніх винагород для кожного стану та дії. Потім будується Q-таблиця значень, що демонструє, наскільки вигідно виконати певну дію в кожному можливому стані. Алгоритм ітеративно оновлює значення функції Q за формулою

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \\ +\alpha\left(r_t + \gamma \max_{a'}(s_{t+1}, a') - Q(s_t, a_t)\right), \quad (1)$$

де s_t – поточний стан, у якому агент приймає рішення; a_t – дія, яку виконує агент у стані s_t; s_{t+1} – новий стан, до якого переходить агент після виконання дії a_t; a' – одна з можливих дій, яку агент може виконати в новому стані s_{t+1}; $Q(s_t, a_t)$ – поточне передбачене значення функції цінності для стану s_t і дії a_t; α – коефіцієнт швидкості навчання; r_t – винагорода за дію a_t;γ – коефіцієнт дисконтування, що визначає важливість майбутніх винагород щодо поточної винагороди; $\max_{a'}(s_{t+1}, a')$ – найбільше значення Q для наступного стану s_{t+1}.

Що вище значення α, то швидше оновлюється знання про середовище, але занадто велике α може призвести до нестабільності навчання. Значення γ, близьке до 1, змушує агента зважати на далекі майбутні винагороди, а низьке значення – орієнтуватися лише на негайні результати.

На рис. 7 наведений псевдокод, що реалізує алгоритм Q-навчання. У ньому для балансу між вибором найкращої відомої дії (експлуатацією) та вибором випадкової дії для пошуку нових можливостей (дослідженням) використовується ε-жадібна стратегія.

Алгоритм Q-навчання є основою більшості алгоритмів, що використовуються в навчанні з підкріпленням.

```
Ініціалізація Q-функції випадковими значеннями (Q(s, a) для всіх s, a)

while не досягнуто критерій завершення:
    s_t = поточний стан
    if випадкове число < ε:
        a = випадкова дія  // Дослідження
    else:
        a = argmax_a Q(s_t, a)  // Експлуатація

    виконати дію a
    отримати винагороду r і новий стан s_{t+1} за формулою (1)

    s_t = s_{t+1}
end
```

Рис. 7. Псевдокод для алгоритму Q-навчання

Однак у процесі керування ФН, коли ІСК має обробляти значну кількість вхідних параметрів, таблиця Q-значень стає надто великою, що робить класичний алгоритм Q-навчання неефективним. У цьому разі більш доцільним є використання покращеної версії Q-навчання – алгоритму глибокої нейронної мережі *Deep Q-Network* (DQN).

На відміну від класичного Q-навчання, де для кожного стану зберігається таблиця значень для можливих дій, в алгоритмі DQN функція цінності Q апроксимується за допомогою нейронної мережі. Це дає змогу уникнути необхідності зберігати в пам'яті таблицю для всіх можливих станів та дій системи, що значно зменшує обсяг пам'яті та обчислювальних ресурсів, необхідних для прийняття керуючих рішень.

Нейронна мережа, яка використовується замість таблиці Q-значень, приймає на вхід вектор стану системи (поточні висота і кут нахилу ковша, тиск у гідросистемі, швидкість і прискорення ФН, опір матеріалу палі) і повертає передбачувані значення Q для можливих дій (підйом або опускання ковша, зміна його кута нахилу, регулювання швидкості руху навантажувача).

Це дає змогу ІСК передбачити, які дії призведуть до максимізації очікуваної винагороди.

Навчання нейронної мережі в DQN здійснюється методом градієнтного спуску, під час якого мінімізується помилка середньоквадратичного відхилення між поточними та цільовими Q-значеннями. У цьому разі мережа, що навчається, і цільова мережа працюють незалежно: перша оновлюється на кожному кроці градієнтного спуску, а друга – лише після кількох ітерацій. Це сприяє зменшенню коливань і робить процес навчання більш стійким до помилок або «шуму» від частих оновлень.

Для підвищення ефективності навчання використовується так званий буфер відтворення досвіду (*Experience Replay*), в якому зберігаються останні переходи агента між станами. Це дає змогу повторно використовувати раніше зібрану інформацію про різні сценарії завантаження ковша. Під час оновлення моделі вибірка даних здійснюється випадковим чином, що знижує кореляцію між послідовними спостереженнями та запобігає перенавчанню. Це особливо важливо в умовах змінної щільності матеріалу палі та зміни кута нахилу ковша під його завантаження.

У разі використання пріоритетного відтворення досвіду (*Prioritized Experience Replay*) замість випадкового добору переходів з буфера досвіду надається пріоритет тим, які спричинили найбільшу помилку в передбаченні Q-значень. Такий підхід прискорює навчання, оскільки модель швидше коригує найбільш значущі помилки.

Ще однією модифікацією DQN є подвійне Q-навчання (*Double DQN*), що усуває проблему завищених оцінок Q-функції. У класичному DQN обчислення майбутньої винагороди ґрунтується на максимальному значенні Q-функції, що може призводити до переоцінки дій. У *Double DQN* одна мережа обирає дію з максимальним Q-значенням, а інша оцінює її цінність для оновлення, що знижує ймовірність перенавчання.

Оскільки процес завантаження ковша передбачає послідовність взаємопов'язаних рішень, перспективним є застосування в ІСК ФН глибокої рекурентної Q-мережі (*Deep Recurrent Q-Network* – DRQN). Це дає змогу брати до уваги часові залежності та додавати рекурентні шари до архітектури мережі.

Методи на основі політики

На відміну від методів, основаних на функції цінності, де агент спочатку оцінює мож-

ливі стани, а потім на підставі цих оцінок обирає дії, методи на основі політики шукають оптимальну стратегію керування: модель безпосередньо навчається передбачати оптимальні дії, не обчислюючи функції цінності для всіх станів.

Методи МН на основі політики можна поділити на дві групи:

1) методи градієнта політики;

2) методи з обмеженнями на оновлення політики.

Методи градієнта політики спрямовані на пряму оптимізацію стратегії (політики) агента. Для пошуку найкращих параметрів використовується градієнтний спуск. Найпростішим із цих методів є REINFORCE. У цьому разі політика $\pi_\theta(s_t)$ буде ймовірнісною функцією, яка обирає одну з дій на основі поточного стану ФН.

Наприклад, вона може бути нейронною мережею, яка приймає вектор станів (висота й кут нахилу ковша, швидкість ковша й навантажувача, навантаження на ківш), і на основі цього обирає ймовірність кожної можливої дії (зміна кута нахилу ковша, швидкості його підйому, швидкості машини).

Під час навчання агент обиратиме дії залежно від поточного стану й отримуватиме нагороду R_t.

Функція нагороди R_t може бути складною, що містить коефіцієнт заповнення ковша і витрати енергії з відповідними ваговими коефіцієнтами. Параметри політики оновлюватимуться на основі градієнтного спуску за формулою

$$\theta_{t+1} = \theta_t + \alpha \cdot \nabla_\theta \log \pi_\theta(a_t \mid s_t) R_t, \qquad (2)$$

де θ_t – вектор параметрів (ваг) нейронної мережі, що описує поточну політику агента на кроці t;

θ_{t+1} – вектор оновлених параметрів політики на кроці навчання $t + 1$;

α – коефіцієнт, що визначає швидкість оновлення параметрів мережі на кожному кроці навчання;

$\nabla_\theta \log \pi_\theta(a_t|s_t)$ – градієнт натурального логарифма ймовірності вибору дії a_t в стані s_t за поточних параметрів політики θ_t; він показує, як зміна параметрів політики θ_t впливає на ймовірність вибору дії a_t в стані s_t.

Процес оновлення триває доти, доки агент не навчиться обирати оптимальні дії, які максимізують загальну винагороду за виконання завдання (рис. 8).

Алгоритм REINFORCE простий у застосуванні, проте його основним недоліком є висока дисперсія оцінок градієнтів, що може призвести до нестабільності оновлень параметрів політики та сповільнення процесу навчання, особливо в задачах із великою кількістю станів і дій, як у разі керування наповненням ковша ФН.

Щоб усунути вплив високої дисперсії оцінок градієнтів на швидкість і якість навчання, можна брати до уваги геометрію простору параметрів. Такий підхід використовується в методі *Natural Policy Gradient* (NPG), який зважає на те, наскільки зміни параметрів політики впливають на розподіл імовірностей дій, що зі свого боку пов'язано з очікуваною нагородою. Це дає змогу коригувати кроки в процесі навчання й завдяки цьому точніше оновлювати параметри політики та підвищити ефективність навчання.

```
Ініціалізація параметрів політики випадковими значеннями

while не досягнуто критерій завершення:
    вибрати поточний стан s_t
    вибрати дію a_t за політикою

    виконати дію a_t
    отримати винагороду r_t і новий стан s_{t+1}

    зберегти винагороди для епізоду

    для кожного часу t:
        обчислити R_t як суму винагород від поточного моменту

        оновити параметри політики за формулою (2)

    оновити політику за новими параметрами θ

end
```

Рис. 8. Псевдокод для алгоритму REINFORCE

```
Ініціалізація параметрів політики θ і значень V(s) випадковими значеннями
while не досягнуто критерій завершення:
    вибрати поточний стан s_t
   // Вибір дії згідно з поточною політикою
    вибрати дію a_t за політикою π(a_t | s_t)
    виконати дію a_t
    отримати винагороду r_t і новий стан s_{t+1}
    Обчислення TD-помилки за формулою (5)
    // Оновлення критика: коригуємо параметри для мінімізації TD-помилки
    V(s_t) = V(s_t) + α_critic * δ_t
    Обчислення переваги A_t за формулою (4)
    Оновлення актора: коригуємо параметри політики θ за формулою (3)
    // Перехід до нового стану
    s_t = s_{t+1}
end
```

Рис. 9. Псевдокод для класичного алгоритму «Актор – критик»

Методи з обмеженнями на оновлення політики спрямовані на стабілізацію навчання, запобігаючи змінам політики на кожному кроці. Зокрема *Trust Region Policy Optimization* (TRPO) обмежує оновлення політики на основі того, наскільки нова політика відрізняється від старої, мінімізуючи ризик надмірного оновлення. Цей метод ефективний для складних завдань із високими вимогами до точності.

Одним із найпопулярніших методів цієї групи є *Proximal Policy Optimization* (PPO), який обмежує оновлення політики за допомогою «кліпінгу» – механізму, що контролює, як сильно може змінитися стратегія на кожному кроці. Це дає змогу досягти результатів, подібних до результатів TRPO, проте з меншими обчислювальними витратами.

Методи класу «Актор – критик».

Такі методи поєднують переваги підходів на основі функції цінності та методів політики, що дає змогу агентові одночасно оцінювати ефективність своїх дій і безпосередньо оновлювати стратегію керування. Як випливає з назви, модель класичного методу «Актор – критик» має два компоненти: «актор», який визначає дію на основі поточного стану системи, та «критик», що оцінює якість цієї дії, прогнозуючи сумарну майбутню винагороду.

Актор навчається оновлювати параметри стратегії θ, використовуючи функцію переваги A_t:

$$\theta_{t+1} = \theta_t + \alpha \cdot \nabla_\theta \log \pi_\theta(a_t \mid s_t) A_t . \qquad (3)$$

Функція переваги A_t показує, наскільки дія a_t краща або гірша за середню дію в стані s_t і обчислюється як

$$A_t = Q(s_t, a_t) - V(s_t), \qquad (4)$$

де $Q(s_t, a_t)$ – цінність дії a_t в стані s_t;

$V(s_t)$ – цінність стану s_t.

Відповідний псевдокод наведено на рис. 9.

Критик коригує свої параметри, щоб мінімізувати TD-помилку (*Temporal Difference Error*) δ_t, що є різницею між передбаченою цінністю стану та цільовим значенням:

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t), \qquad (5)$$

де r_t – фактична винагорода, отримана агентом; γ – коефіцієнт дисконтування, що визначає важливість майбутніх винагород щодо поточної винагороди; $V(s_t)$ – передбачена цінність поточного стану s_t; $V(s_t+1)$ – цінність наступного стану.

Метод «Актор – критик» є більш гнучким порівняно з Q-навчанням, оскільки він дає змогу агентові не лише запам'ятовувати значення Q-функції для кожного стану, а й формувати безперервну політику керування. Удосконаленням цього методу є алгоритми *Advantage Actor-Critic* (A2C) та *Asynchronous Advantage Actor-Critic* (A3C), які використовують паралельні обчислення для прискорення навчання.

Порівняння ефективності методів навчання з підкріпленням у процесі керування ковшем ФН

Отже, різні методи навчання з підкріпленням мають свої особливості та різну придатність до використання в ІСК ковшем ФН. На ефективність застосування цих методів у ІСК ФН впливають такі характеристики, як гнучкість, складність їх реалізації, адаптивність до змінних умов роботи, а також залежність від кількості та якості даних (табл. 1).

Таблиця 1 – Порівняння методів навчання з підкріпленням*

Метод	Гнучкість	Складність реалізації	Адаптивність	Залежність від даних
Q-навчання	Н	Н	С	В
DQN	С	С	С	С
REINFORCE	В	В	С	С
PPO	С	С	С	Н
Актор-критик	В	В	В	Н

Примітка: * Н – низька, С – середня, В – Висока

Гнучкість визначає, наскільки метод здатний працювати в різних умовах без необхідності перенавчання. Складність реалізації оцінює обчислювальні витрати, вимоги до апаратних ресурсів ІСК і складність налаштування. Адаптивність показує, наскільки швидко метод може коригувати свою поведінку в разі зміни умов роботи, а залежність від даних відтворює, наскільки метод вимагає значної кількості навчальних прикладів або заздалегідь розмічених даних.

Методи на основі функції цінності, до яких належить класичне Q-навчання, дають змогу ІСК оцінювати корисність різних станів і дій, спираючись на накопичений досвід взаємодії із матеріалом палі. Однак їх застосування в ІСК ФН ковшем навантажувача обмежене значним розміром Q-таблиці, необхідної для зберігання оцінок усіх можливих дій у кожному стані.

Це призводить до зростання обчислювальних витрат і ускладнює використання алгоритму в умовах змінних властивостей матеріалу.

Застосування глибокої нейронної мережі в алгоритмі DQN частково розв'язує цю проблему, даючи змогу апроксимувати функцію цінності без необхідності зберігання таблиці значень. Проте цей підхід вимагає значного обсягу даних для навчання, а його адаптивність залишається обмеженою фіксованою архітектурою мережі та стратегією оновлення ваг. Це може ускладнювати пристосування системи до різких змін властивостей матеріалу.

Методи на основі політики, зокрема REINFORCE, PPO і TRPO, здійснюють пряме оновлення керуючих дій без явного обчислення функції цінності. Це робить їх гнучкими, оскільки стратегія керування ІСК формується безпосередньо в процесі взаємодії ковша з матеріалом палі та коригується на основі отримуваних винагород.

Однак висока дисперсія градієнтних оцінок може спричинити нестабільність навчання, що ускладнює швидку адаптацію до змін навантаження на ківш і його гідропривод. Введення обмежень на оновлення політики в PPO і TRPO підвищує стійкість алгоритмів, запобігаючи різким змінам стратегії та знижуючи ризик нестійкої поведінки ІСК.

Методи класу «Актор – критик» дають змогу не лише формувати оптимальні стратегії керування ковшем, а й брати до уваги довгострокові наслідки прийнятих керуючих рішень. Завдяки використанню нейронних мереж для подання керівних політик вони забезпечують кращу адаптацію ІСК до змінних умов порівняно з методами, основаними на функції цінності. На відміну від підходів градієнта політики, алгоритми «Актор – критик» зменшують дисперсію градієнтів, що підвищує стабільність процесу навчання. Однак їх реалізація потребує значних обчислювальних ресурсів і ретельного налаштування параметрів, що може ускладнювати застосування в ІСК ФН.

Підвищити ефективність ІСК ФН та забезпечити її гнучкість у складних динамічних умовах експлуатації може поєднання методів «Актор – критик» з іншими підходами, зокрема еволюційними алгоритмами.

Висновки

Розроблення та впровадження системи автоматизації робочого процесу ФН, зокрема керування ковшем під час його навантаження матеріалом, є складним завданням внаслідок впливу значної кількості невизначених динамічних факторів. З огляду на це перспективним є використання в системах керування ФН методів МН.

У цьому разі найбільш доцільним є впровадження безмодельних методів у межах навчання з підкріпленням, зокрема методів класу «Актор – критик». Поєднання зазначених методів з іншими підходами МН сприяють підвищенню ефективності системи керування ФН та забезпечують їх гнучкість у динамічних умовах експлуатації.

Література

1. Разарьонов Л.В., Гурко В.О. Особливості роботи короткобазового малогабаритного навантажувача з бортовим поворотом при виконанні развороту. *Вісник НТУ «ХПІ». Тематичний випуск «Машинознавство та САПР»*. 2024. № 2. С.75-78.
2. Jassim H.S.H., Lu W., Olofsson T. Determining the environmental impact of material hauling with wheel loaders during earthmoving operations. *Journal of the Air & Waste Management Association*. 2019. Vol. 69, no. 10. P. 1195–1214. URL: https://doi.org/10.1080/10962247.2019.1640805
3. Research on the Trajectory and Operational Performance of Wheel Loader Automatic Shoveling / Y. Chen et al. *Applied Sciences*. 2022. Vol. 12, no. 24. P. 12919.
4. Гурко О.Г., Гурко В.О., Кучеренко А.Ю. Керування рухом фронтального навантажувача за заданою траєкторією. *Вісник ХНАДУ*. 2023. Вип. 101, т. 1. С. 26–34.
5. Frank B., Skogh L., Filla R. On increasing fuel efficiency by operator assistant systems in a wheel loader. *International conference on advanced vehicle technologies and integration*: Proceedings, Changchun. 2012. P. 155–161.
6. Liu X., Sun D., Qin D., Liu J. Achievement of fuel savings in wheel loader by applying hydrodynamic mechanical power split transmissions. *Energies*. 2017. Vol. 10, no. 9. P. 1267.
7. Filla R. An event-driven operator model for dynamic simulation of construction machinery. *The ninth Scandinavian international conference on fluid power*: Proceedings, Linköping. 2005. P. 3–20.
8. Nezhadali V., Frank B., Eriksson L. Wheel loader operation – Optimal control compared to real drive experience. *Control engineering practice*. 2016. Vol. 48. P. 1–9.
9. Planning the trajectory of an autonomous wheel loader and tracking its trajectory via adaptive model predictive control / J. Shi et al. *Robotics and autonomous systems*. 2020. Vol. 131. P. 103570.
10. Roberts J. M., Duff E. S., Corke P. I. Reactive navigation and opportunistic localization for autonomous underground mining vehicles. *Information sciences*. 2002. Vol. 145, no. 1-2. P. 127–146.
11. Gurko O., Kyrychenko I., Verbitsky V. Linear controller for wheel loader trajectory tracking. *Вісник ХНАДУ*. 2024. №. 107. С. 14–20.
12. Azar E.R., Kamat V.R. Earthmoving equipment automation: A review of technical advances and future outlook. *Journal of Information Technology in Construction*. 2017. Vol. 22, no. 13. P. 247–265.
13. Feature-based sensor configuration and working-stage recognition of wheel loader / L. Hou et al. *Automation in construction*. 2022. Vol. 151. P. 104401.
14. Meng Yu et al. Bucket trajectory optimization under the automatic scooping of LHD. *Energies*. 2019. Vol. 12, no. 20. P. 3919.
15. Frank B., Kleinert J., Filla R. Optimal control of wheel loader actuators in gravel applications. *Automation in construction*. 2018. Vol. 91. P. 1–14.
16. Filla R., Obermayr M., Frank B. A study to compare trajectory generation algorithms for automatic bucket filling in wheel loaders. *3rd Commercial Vehicle Technology Symposium*: Proceedings, Kaiserslautern, 2014. P. 357–374.
17. Field test of neural-network based automatic bucket-filling algorithm for wheel-loaders / S. Dadhich et al. *Automation in construction*. 2019. Vol. 97. P. 1–12.
18. Dadhich S., Bodin U., Sandin F., Andersson U. Machine learning approach to automatic bucket loading. *24th Mediterranean Conference on Control and Automation*: Proceedings, Athens, Greece, 2016. P. 1260–1265.
19. Data-Driven reinforcement-learning-based automatic bucket-filling for wheel loaders / J. Huang et al. *Applied sciences*. 2021. Vol. 11, no. 19. P. 9191.
20. Bucket loading trajectory optimization for the automated wheel loader/J. Yao et al. *IEEE transactions on vehicular technology*. 2023. P.1–11.
21. Gurko O., Miroshnyk V. Artificial intelligence in road construction: prospects and challenges. *Scientific papers of silesian university of technology. organization and management series*. 2024. Vol. 2024, no. 210. P. 191–203.
22. Мирошник В.А., Гурко В.О., Гурко О.Г. Використання навчання з підкріпленням для планування шляху будівельного робота. *Біоніка інтелекту*. 2024. Вип. 101, № 2. С. 34–38.
23. Continuous control of an underground loader using deep reinforcement learning / Backman S. et al. *Machines.* 2021. Vol. 9, no. 10. P. 216.
26. https://doi.org/10.3390/machines9100216. Azulay O., Shapiro A. Wheel Loader Scooping Controller Using Deep Reinforcement Learning. IEEE Access. 2021. Vol. 9. P. 24145–24154.
25. Eriksson D., Ghabcheloo R., Geimer M. Automatic Loading of Unknown Material with a Wheel Loader Using Reinforcement Learning. IEEE International Conference on Robotics and Automation, Yokohama, Japan, 2024. P. 3646–3652.
26. Jo T. Machine learning foundations. Supervised, Unsupervised, and Advanced Learning. Cham: Springer International Publishing. 2021. 410 p.

References

1. L. Razarenov, V. Hurko, "Peculiarities of operation of a short-base compact loader with lateral rotation when performing a U-turn ", *Bul. Nat. Technical Univ. «KhPI». Series:*

Engineering and CAD, no. 2, pp. 75-78, 2024 (in Ukrainian).
2. H.S.H. Jassim, W. Lu and T. Olofsson, "Determining the environmental impact of material hauling with wheel loaders during earthmoving operations", *J. Air & Waste Manage. Assoc.*, vol. 69, no 10, pp. 1195–1214, 2019.
3. Y. Chen, H. Jiang, G. Shi та T. Zheng, "Research on the trajectory and operational performance of wheel loader automatic shoveling", *Appl. Sci.*, vol. 12, no. 24, pp. 12919, 2022.
4. O. Gurko, V. Hurko and A. Kucherenko, "Controlling wheel loader motion along a desired trajectory", *Bull. Kharkov Nat. Automobile Highway Univ.*, issue 1, no 101, pp. 26, 2023 (in Ukrainian).
5. B. Frank, L. Skogh and R. Filla, "On increasing fuel efficiency by operator assistant systems in a wheel loader", in *Int. Conf. Adv. Vehicle Technol. Integration*, Changchun, China. 2012, pp. 155–161.
6. Xiaojun Liu, Dongye Sun, Datong Qin and Junlong Liu "Achievement of fuel savings in wheel loader by applying hydrodynamic mechanical power split transmissions", *Energies*, vol. 10, no. 9, pp. 1267, 2017.
7. R. Filla "An event-driven operator model for dynamic simulation of construction machinery", in *9th Scandinavian Int. Conf. on Fluid Power*, Linköping. 2005, pp. 3–20.
8. V. Nezhadali, B. Frank and L. Eriksson. "Wheel loader operation – Optimal control compared to real drive experience", *Control engineering practice*, vol. 48, pp. 1–9, 2016. https://doi.org/10.1016/ j.conengprac.2015.12.015
9. J. Shi et al. "Planning the trajectory of an autonomous wheel loader and tracking its trajectory via adaptive model predictive control", *Robotics and autonomous systems*, vol. 131, pp. 103570, 2020.
10. J.M. Roberts, E.S. Duff and P.I. Corke. "Reactive navigation and opportunistic localization for autonomous underground mining vehicles", *Information sciences*, vol. 145, no. 1–2, pp. 127–146, 2002.
11. O. Gurko, I. Kyrychenko, V. Verbitsky. "Linear controller for wheel loader trajectory tracking", *Bull. Kharkov Nat. Automobile Highway Univ.*, issue 107. pp. 14–20, 2024.
12. E.R. Azar and V.R. Kamat. "Earthmoving equipment automation: A review of technical advances and future outlook", *Journal of Information Technology in Construction*, vol. 22, no. 13, pp. 247–265, 2017.
13. L. Hou et al. "Feature-based sensor configuration and working-stage recognition of wheel loader", *Automation in construction*, vol. 151, pp. 104401, 2022.
14. Yu Meng et al. "Bucket trajectory optimization under the automatic scooping of LHD", *Energies*, vol. 12, no. 20, pp. 3919, 2019.
15. B. Frank, J. Kleinert and R. Filla. "Optimal control of wheel loader actuators in gravel applications", *Automation in construction*, vol. 91, pp. 1–14, 2018.
16. Filla R., Obermayr M., Frank B. "A study to compare trajectory generation algorithms for automatic bucket filling in wheel loaders", in *3rd Commercial Vehicle Technology Symposium*, Kaiserslautern, 2014, pp. 357–374.
17. S. Dadhich et al. "Field test of neural-network based automatic bucket-filling algorithm for wheel-loaders", *Automation in construction*, vol. 97, pp. 1–12, 2019.
18. S. Dadhich, U. Bodin, F. Sandin and U. Andersson, "Machine learning approach to automatic bucket loading", in *24th Mediterranean Conference on Control and Automation*, Athens, Greece, 2016, pp. 1260–1265
19. J. Huang et al. "Data-Driven reinforcement-learning-based automatic bucket-filling for wheel loaders", *Applied sciences*, vol. 11, no. 19, pp. 9191, 2021.
20. J. Yao et al. "Bucket loading trajectory optimization for the automated wheel loader" *IEEE transactions on vehicular technology*, pp. 1–11, 2023.
21. O. Gurko, V. Miroshnyk, "Artificial intelligence in road construction: prospects and challenges", *Scientific papers of silesian university of technology. Organization and management series*, vol. 2024, no. 210, pp. 191–203, 2024.
22. V.A. Miroshnyk, V.O. Hurko, O.G. Gurko. "The use of reinforcement learning to plan the path of a construction robot", *Bionics of intelligence, issue* 101, no 2, pp. 34-38, 2024 (in Ukrainian).
23. S. Backman et al. "Continuous control of an underground loader using deep reinforcement learning", *Machines*, vol. 9, no. 10, pp. 216, 2021.
24. J. Yao et al. "Bucket Loading Trajectory Optimization for the Automated Wheel Loader," *in IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 6948–6958, June 2023.
25. D. Eriksson, R. Ghabcheloo and M. Geimer. "Automatic Loading of Unknown Material with a Wheel Loader Using Reinforcement Learning", in *IEEE International Conference on Robotics and Automation*, Yokohama, Japan, 2024, pp. 3646–3652.
26. T. Jo. Machine learning foundations. Supervised, Unsupervised, and Advanced Learning. Cham: Springer International Publishing. 2021. 410 p.

Кириченко Ігор Георгійович, д.т.н., професор, кафедра будівельних і дорожніх машин, igk160450@gmail.com,
ORCID: https://orcid.org/0000-0002-2128-3500,
Гурко Володимир Олександрович, аспірант, volgurko@khadi.kharkov.ua,
ORCID: https://orcid.org/0009-0006-9216-6682.
Харківський національний автомобільно-дорожній університет, 61002, Україна, м. Харків, вул. Ярослава Мудрого, 25.

Machine learning applications in intelligent wheel loader bucket control systems

Abstract. Problem. *Wheel loaders are versatile machines widely used for loading, unloading and transporting various materials. Modern trends in the improvement of these machines focus on increasing the level of their automation. Automation reduces the operator's workload, enhances safety, and improves the productivity and energy efficiency of machines under various operating conditions.*

One of the tasks of wheel loader automation is controlling bucket loading. Due to the complex interaction of the bucket with soil or other loaded material and the influence of numerous uncertain dynamic factors, the use of model-free machine learning methods in bucket control systems is reasonable. ***Goal.*** *The goal of this paper is to analyze machine learning methods in terms of their suitability for developing an intelligent bucket control system for wheel loaders.* ***Methodology.*** *We examined three main types of machine learning – supervised, unsupervised, and reinforcement learning with a detailed evaluation of reinforcement learning methods such as Q-learning, Deep Q-Network, REINFORCE, and Proximal Policy Optimization. Then, we compared these methods based on flexibility, implementation complexity, adaptability, and data dependency to assess their suitability for an intelligent bucket control system.* ***Results.*** *The analysis identifies the Actor-Critic approach as the most suitable, with its potential for combination with other methods highlighted to enhance system performance in dynamic operating conditions.* ***Originality.*** *Through a systematic comparison of several reinforcement learning algorithms, we identified the Actor-Critic approach as the most effective method for controlling the loader bucket.* ***Practical value.*** *The results of the study will help to improve the efficiency of intelligent loader bucket control system.*

Key words: *wheel loader, bucket filling, intelligent control system, machine learning, reinforcement learning.*

Kyrychenko Igor, professor, Doct. of Science,
Department of Construction and Road machines,
igk160450@gmail.com,
ORCID: https://orcid.org/0000-0002-2128-3500,
Hurko Volodymyr, PhD student,
volgurko@khadi.kharkov.ua,
ORCID: https://orcid.org/0009-0006-9216-6682.
Kharkiv National Automobile and Highway University, 25, Yaroslava Mudrogo str.,
Kharkiv, 61002, Ukraine.