УДК 004.832.2, 004.023

Методи розв’язання задачі управління запасами на основі нейро-асоціативного навчання і навчання з підкріпленням

Євген Федоров
професор, д-р техн. наук
ORCID: 0000-0003-3841-7373
y.fedorov@chdtu.edu.ua

Черкаський державний технологічний університет

Тетяна Уткіна
доцент, канд. техн. наук
ORCID: 0000-0002-6614-4133
t.utkina@chdtu.edu.ua

Черкаський державний технологічний університет

Ольга Нечипоренко
доцент, канд. техн. наук
ORCID: 0000-0002-3954-3796
o.nechyporenko@chdtu.edu.ua

Черкаський державний технологічний університет

Ключові слова:
нейро-асоціативне навчання;
навчання з підкріпленням;
управління запасами;
обмежена машина Коші;
метод Q-навчання;
метод SARSA.

Сьогодні актуальним завданням є розробка інтелектуальних методів, спрямованих на розв’язання задач управління запасами. Багато сучасних компаній для вдосконалення та оптимізації своїх бізнес-процесів використовують технологію теорії обмежень, яка забезпечує динамічне управління буфером запасів і використовується для управління ланцюгами поставок. Метою роботи є підвищення ефективності управління запасами за допомогою нейро-асоціативного навчання на основі обмеженої машини Коші та навчання з підкріпленням на основі Q-навчання та SARSA. Для досягнення поставленої мети створено метод на основі обмеженої машини Коші для управління буфером запасів, метод на основі Q-навчання та метод на основі SARSA для загальних задач управління запасами. Запропонована нейромережева модель обмеженої машини Коші має гетероасоціативну пам’ять без обмежень на ємність і забезпечує високу точність управління буфером запасів. Модель використовує розподіл Коші, що покращує збіжність методу параметричної ідентифікації, порівняно з традиційною обмежено машиною Больцмана. На відміну від повної машини Коші, обмежена машина Коші дає змогу працювати з більшим обсягом пам'яті. Модифікація методів Q-навчання та SARSA за рахунок динамічних параметрів дає змогу підвищити швидкість навчання за заданого рівня середньоквадратичної помилки. Проведені обчислювальні експерименти показали, що керування важливістю нагороди, параметром швидкості навчання, параметром для ε-жадібного підходу для методів Q-навчання та SARSA дає змогу на початкових стадіях зробити пошук рішення більш глобальним, а на заключних стадіях зробити пошук рішення більш локальним. Запропоновані методи дають змогу розширити сферу застосування нейро-асоціативного навчання та навчання з підкріпленням, що

підтверджується їх адаптацією для задач управління запасами та сприяє підвищенню ефективності інтелектуальних комп'ютерних систем загального і спеціального призначення. Перспективами подальших досліджень є застосування запропонованих методів для інших задач прийняття рішень, зокрема й у області штучного інтелекту.

DOI: 10.31558/2786-9482.2024.1.5

Вступ

Сьогодні актуальним завданням є розробка методів, спрямованих на вирішення задач управління запасами, які використовуються в інтелектуальних комп'ютерних системах загального та спеціального призначення. Все більше компаній прагнуть вдосконалювати та оптимізувати свої бізнес-процеси на основі впровадження технології теорії обмежень, яка забезпечує динамічне управління буфером запасів і використовується для управління ланцюгами поставок [1–4]. Внаслідок цього істотно зростає актуальність розробки методів оптимізації для технології теорії обмежень. Методи оптимізації, що знаходять точний розв'язок, мають високу обчислювальну складність. Методи оптимізації, що знаходять наближений розв'язок за допомогою спрямованого пошуку, часто потрапляють у локальний екстремум. Методи випадкового пошуку не гарантують збіжності. Через це виникає проблема недостатньої ефективності методів оптимізації, яка потребує вирішення. Для прискореного знаходження розв'язку для задач управління запасами та зниження ймовірності попадання в локальний екстремум використовуються нейро-асоціативне навчання та навчання з підкріпленням.

Метою роботи є підвищення ефективності пошуку розв'язків задач управління запасами за допомогою нейро-асоціативного навчання та навчання з підкріпленням. Для досягнення поставленої мети необхідно вирішити такі завдання:

1) створити нейромережеву модель управління буфером запасів;
2) запропонувати критерій ефективності нейромережевої моделі управління буфером запасів;
3) створити метод параметричної ідентифікації нейромережевої моделі управління буфером запасів;
4) запропонувати критерій ефективності управління запасами;
5) створити метод на основі Q-навчання для задач управління запасами;
6) створити метод на основі SARSA для задач управління запасами;
7) провести чисельне дослідження запропонованих методів.

Постановка проблеми

Нехай для моделі управління буфером запасів задано навчальну вибірку $S = \{(\mathbf{x}_m^{in}, \mathbf{d}_m^{out})\}$, $m \in \overline{1, M}$, де $\mathbf{x}_m^{in}$ – m-й вектор ознак, $\mathbf{d}_m^{out}$ – m-й вектор виду дії, яка змінює розмір буферу запасів. Тоді проблему підвищення точності управління буфером запасів за моделлю обмеженої машини Коші (RCM-моделлю) запишемо так: $g(\mathbf{x}^{in}, \mathbf{w})$, де $\mathbf{x}^{in}$ – вектор ознак,

$\mathbf{w}$ – вектор параметрів. Потрібно підібрати такий вектор параметрів $\mathbf{w}^*$, який забезпечує мінімум критерію: $F = \frac{1}{M}\sum_{m=1}^{M}(g(\mathbf{x}_m^{in}, \mathbf{w}^*) - \mathbf{d}_m^{out})^2 \to \min$.

Проблема підвищення ефективності вирішення задач управління запасами на основі методів навчання з підкріпленням (Q-навчання та SARSA) представляється як проблема знаходження такого розв'язку x^*, щоб $F(x^*) \to \min$, де x – вектор кількості закуповуваного товару у постачальника, а $F(\cdot)$ – цільова функція, яка пов'язана з витратами на закупівлю та зберігання товарів.

З погляду економічного зиску підвищення ефективності вирішення задач управління запасами дає змогу зменшити витрати від дефіциту товару та витрати від зберігання товару.

Огляд літератури

Серед конекціоністських методів, що використовуються для управління різними об'єктами [5–8], важливу роль відіграють методи на основі асоціативних нейромереж. Методи нейро-асоціативного навчання мають один або більше з таких недоліків: 1) не мають гетероасоціативної пам'яті [9–11]; 2) не працюють із дійсними даними [12–14]; 3) не мають високої ємності асоціативної пам'яті [15–17]; 4) не мають високої точності [18–20]; 5) мають високу обчислювальну складність [21–23]. Через це постає завдання побудови ефективних методів нейро-асоціативного навчання.

Методи навчання з підкріпленням мають один або більше з таких недоліків: 1) є лише абстрактний опис методу чи опис методу орієнтовано на рішення лише певного завдання [24, 25]; 2) не гарантується збіжність методу [26, 27]; 3) не враховується вплив номера ітерації на процес пошуку рішення [28, 29]; 4) відсутня можливість вирішувати завдання умовної оптимізації [30]; 5) недостатня точність методу [31, 32]; 6) не автоматизовано процедуру визначення значень параметрів [33]. Через це постає завдання побудови ефективних методів навчання з підкріпленням.

Метод нейро-асоціативного навчання для задач управління буфером запасів

Сформуємо вхідні та вихідні змінні. Вхідними змінними обрано:

x_1 – поточний обсяг запасу;

x_2 – час знаходження у червоній зоні буфера запасів;

x_3 – час знаходження у зеленій зоні буфера запасів.

Вихідною змінною y обрано номер виду дії (збільшити, зменшити, не міняти), що змінює розмір буфера запасів.

Вхідні та вихідні змінні подаються у бінарному вигляді.

Структурна схема нейромережевої моделі управління буфером запасів у формі RCM-моделі зображена на рис. 1. Вона являє собою рекурентну нейронну мережу з одним видимим та одним прихованим шарами.

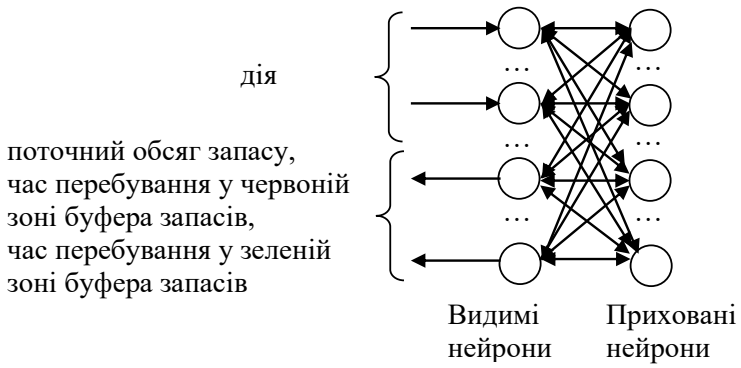

Рисунок 1. Структурна схема RCM-моделі

Компонентами RCM є приховані стохастичні нейрони, стан яких описується на основі розподілу Бернуллі у вигляді:

$$x_j = \begin{cases} 1, & \text{з ймовірністю } P_j \\ 0, & \text{з ймовірністю } 1 - P_j \end{cases}.$$

Ймовірність переходу j-го стохастичного нейрона у стан 1 розраховується так:

$$P_j = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\Delta E_j\right),$$

де ΔE_j – збільшення енергії нейронної мережі внаслідок зміни стану j-го стохастичного нейрона з 0 на 1.

Переваги RCM-моделі:

1) на відміну від більшості нейронних мереж, має гетероасоціативну пам’ять;
2) на відміну від машини Больцмана, не має обмежень на ємність пам’яті;
3) на відміну від машини Больцмана, має меншу обчислювальну складність.

Визначимо нейромережеву модель управління буфером запасів.

Позитивна фаза (кроки 1–2)

1. Ініціалізація стану видимих вхідних нейронів, які відповідають вхідним змінним x_1, x_2 та x_3 у бінарному вигляді $\mathbf{x1}^{in} = \mathbf{x}^{in}$.

2. Обчислення стану прихованих нейронів:

$$P_j = \frac{1}{2} + \frac{1}{\pi} \arctan\left(b_j^h + \sum_{i=1}^{N^{in}} w_{ij}^{in-h} x1_i^{in} \right), \; j \in \overline{1, N^h};$$

$$x1_j^h = \begin{cases} 1, & P_j \geq U(0,1) \\ 0, & P_j < U(0,1) \end{cases}, \; j \in \overline{1, N^h},$$

де $U(0,1)$ – функція, що повертає рівномірно розподілене випадкове число в діапазоні [0, 1].

Негативна фаза (крок 3)

3. Обчислення стану видимих вихідних нейронів, які відповідають виду дії у бінарному вигляді:

$$P_j = \frac{1}{2} + \frac{1}{\pi}\arctan\left(b_j^{out} + \sum_{i=1}^{N^h} w_{ij}^{out-h} x1_i^h\right),\ j \in \overline{1, N^{out}};$$

$$x2_j^{out} = \begin{cases} 1, & P_j \geq U(0,1) \\ 0, & P_j < U(0,1) \end{cases},\ j \in \overline{1, N^{out}},$$

де b_j^h – поріг для j-го нейрона прихованого шару;

b_j^{out} – поріг для j-го нейрона видимого вихідного шару;

w_{ij}^{in-h} – вага зв'язку від i-го нейрона у видимому вхідному шарі до j-го нейрона прихованого шару;

w_{ij}^{out-h} – вага зв'язку від i-го нейрона у видимому вихідному шарі до j-го нейрона прихованого шару;

N^h – кількість нейронів у прихованому шарі;

N^{in} – кількість нейронів у видимому вхідному шарі;

N^{out} – кількість нейронів у видимому вихідному шарі.

Виберемо критерій ефективності нейромережевої моделі управління буфером запасів. Відповідно до цільової функції навчання RCM-моделі у знаходженні таких значень вектора параметрів $\mathbf{w} = (w_{11}^{in-h}, ..., w_{N^{in}N^h}^{in-h}, w_{11}^{out-h}, ..., w_{N^{out}N^h}^{out-h})$, які мінімізують середньоквадратичну помилку на вибірці даних:

$$F = \frac{1}{M(N^{in} + N^{out})} \sum_{m=1}^{M} \left\| \mathbf{x}2_m^{out} - \mathbf{d}_m^{out} \right\|^2 \to \min_{\mathbf{w}},$$

де $\mathbf{x}2_m^{out}$ – m-й оціночний вектор виду дії за моделлю;

$\mathbf{d}_m^{out}$ – m-й вектор виду дії.

Запропонуємо метод параметричної ідентифікації нейромережевої моделі управління буфером запасів на основі алгоритму CD-1 (one-step contrastive divergence). Метод параметричної ідентифікації нейромережевої моделі управління буфером запасів з урахуванням алгоритму CD-1 (рис. 2) складається з восьми блоків.

1. Ініціалізація

Номер ітерації навчання $n = 1$: ініціалізація зсувів (порогів) $b_i^{out}(n)$, $i \in \overline{1, N^{out}}$, $b_j^h(n)$, $j \in \overline{1, N^h}$, і ваг $w_{ij}^{in-h}(n)$, $i \in \overline{1, N^{in}}$, $j \in \overline{1, N^h}$, $w_{ij}^{out-h}(n)$, $i \in \overline{1, N^{out}}$, $j \in \overline{1, N^h}$, $w_{ii}^{in-h}(n) = 0$, $w_{ii}^{out-h}(n) = 0$, $w_{ij}^{in-h}(n) = w_{ji}^{in-h}(n)$, $w_{ij}^{out-h}(n) = w_{ji}^{out-h}(n)$ за рівномірним розподілом на інтервалі (0, 1) або [–0.5, 0.5].

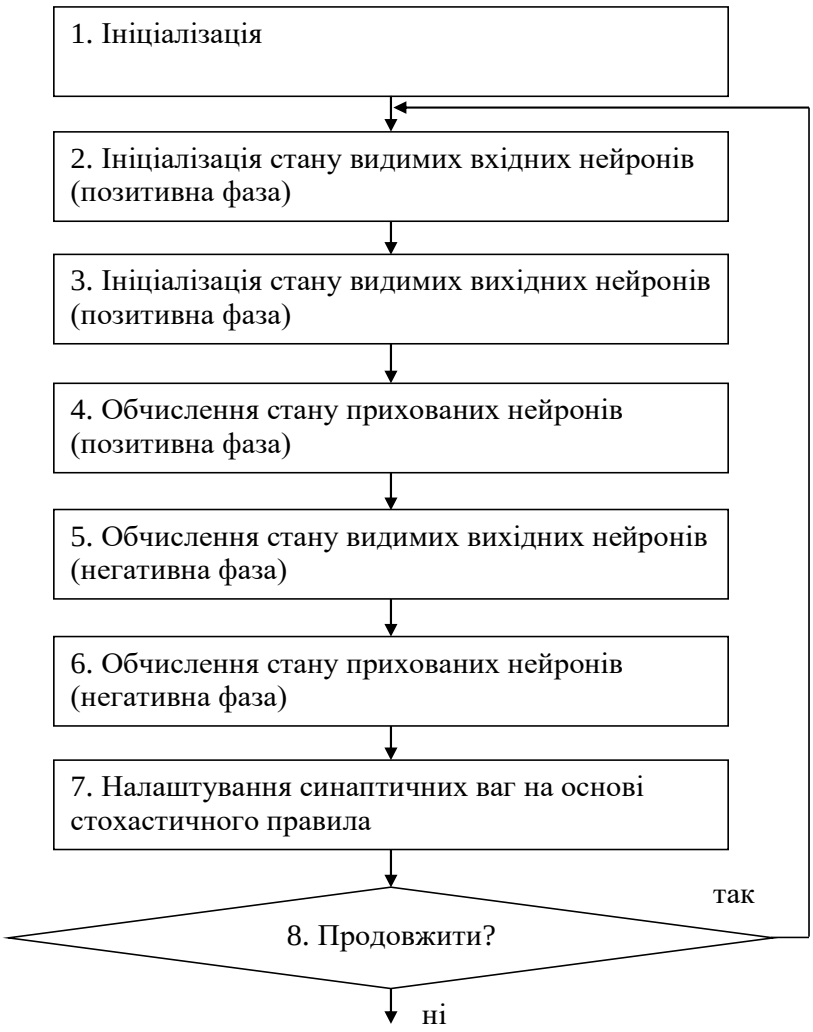

Рисунок 2. Послідовність процедур методу параметричної ідентифікації нейромережевої моделі управління буфером запасів на основі CD-1

Задається навчальна вибірка $\{(\mathbf{x}_m^{in}, \mathbf{x}_m^{out}) \mid \mathbf{x}_m^{in} \in (0,1)^{N^{in}}, \mathbf{x}_m^{out} \in (0,1)^{N^{out}}\}$, $m \in \overline{1,M}$, де $\mathbf{x}_m^{in}$ – m-й вхідний вектор зі значеннями у бінарному вигляді, $\mathbf{x}_m^{out}$ – вектор виду дії в бінарному вигляді, M – розмір навчальної вибірки.

Позитивна фаза (кроки 2–4)

2. Ініціалізація стану видимих вхідних нейронів, які відповідають вхідним змінним x_1, x_2 та x_3 у бінарному вигляді:

$$\mathbf{x}1_m^{in} = \mathbf{x}_m^{in},\ m \in \overline{1,M}.$$

3. Ініціалізація стану видимих вихідних нейронів, які відповідають виду дії у бінарному вигляді:

$$\mathbf{x}1_m^{out} = \mathbf{x}_m^{out},\ m \in \overline{1,M}.$$

4. Обчислення стану прихованих нейронів:

$$P_{mj} = \frac{1}{2} + \frac{1}{\pi}\arctan\left(b_j^h(n) + \sum_{i=1}^{N^{in}} w_{ij}^{in-h}(n) x1_{mi}^{in} + \sum_{i=1}^{N^{out}} w_{ij}^{out-h}(n) x1_{mi}^{out} \right),\ m \in \overline{1, M};$$

$$x1_{mj}^h = \begin{cases} 1, & P_{mj} \geq U(0,1) \\ 0, & P_{mj} < U(0,1) \end{cases},\ m \in \overline{1, M},\ j \in \overline{1, N^h}.$$

Негативна фаза (кроки 5–6)

5. Обчислення стану видимих вихідних нейронів, які відповідають дії у бінарному вигляді:

$$\mu_{mj}^{out} = b_j^{out}(n) + \sigma_j^{out} \sum_{i=1}^{N^h} w_{ij}^{out-h}(n) x1_{mi}^h,\ m \in \overline{1, M},\ j \in \overline{1, N^{out}};$$

$$x2_{mj}^{in} = \mu_{mj}^{in} + \sigma_j^{in} N(0,1),\ m \in \overline{1, M},\ j \in \overline{1, N^{out}}.$$

6. Обчислення стану прихованих нейронів:

$$P_{mj} = \frac{1}{2} + \frac{1}{\pi}\arctan\left(b_j^h(n) + \sum_{i=1}^{N^{in}} w_{ij}^{in-h}(n) x1_{mi}^{in} + \sum_{i=1}^{N^{out}} w_{ij}^{out-h}(n) x2_{mi}^{out} \right),\ m \in \overline{1, M};$$

$$x2_{mj}^h = \begin{cases} 1, & P_{mj} \geq U(0,1) \\ 0, & P_{mj} < U(0,1) \end{cases},\ m \in \overline{1, M},\ j \in \overline{1, N^h}.$$

7. Налаштування порогів та синаптичних ваг на основі стохастичного правила:

$$b_i^{out}(n+1) = b_i^{out}(n) + \eta\left(\frac{1}{M}\sum_{m=1}^{M} x1_{mi}^{out} - \frac{1}{M}\sum_{m=1}^{M} x2_{mi}^{out} \right),\ i \in \overline{1, N^{out}};$$

$$b_i^h(n+1) = b_i^h(n) + \eta\left(\frac{1}{M}\sum_{m=1}^{M} x1_{mi}^h - \frac{1}{M}\sum_{m=1}^{M} x2_{mi}^h \right),\ i \in \overline{1, N^h};$$

$$\rho_{ij}^+ = \frac{1}{M}\sum_{m=1}^{M} x1_{mi}^{in} x1_{mj}^h,\ i \in \overline{1, N^{in}},\ j \in \overline{1, N^h};$$

$$\rho_{ij}^- = \frac{1}{M}\sum_{m=1}^{M} x1_{mi}^{in} x2_{mj}^h,\ i \in \overline{1, N^{in}},\ j \in \overline{1, N^h};$$

$$w_{ij}^{in-h}(n+1) = w_{ij}^{in-h}(n) + \eta(\rho_{ij}^+ - \rho_{ij}^-),\ i \in \overline{1, N^{in}},\ j \in \overline{1, N^h};$$

$$\rho_{ij}^{+} = \frac{1}{M}\sum_{m=1}^{M} x1_{mi}^{out}\, x1_{mj}^{h}\,,\ i \in \overline{1, N^{out}}\,,\ j \in \overline{1, N^{h}}\,;$$

$$\rho_{ij}^{-} = \frac{1}{M}\sum_{m=1}^{M} x2_{mi}^{out}\, x2_{mj}^{h}\,,\ i \in \overline{1, N^{out}}\,,\ j \in \overline{1, N^{h}}\,;$$

$$w_{ij}^{out-h}(n+1) = w_{ij}^{out-h}(n) + \eta(\rho_{ij}^{+} - \rho_{ij}^{-})\,,\ i \in \overline{1, N^{out}}\,,\ j \in \overline{1, N^{h}}\,.$$

8. Перевірка умови завершення:

якщо $\dfrac{1}{M \cdot N^{out}}\sum_{m=1}^{M}\sum_{i=1}^{N^{out}} | x1_{mi}^{out} - x2_{mi}^{out} | > \varepsilon$, тоді $n = n+1$ і перехід до 2.

Методи навчання з підкріпленням для задач управління запасами

Визначимо цільову функцію для задач управління запасами як добуток двох функцій:

$$F(x,z) = F1(x,z) + F2(x,z) \to \min_{x}\,;$$

$$F1(x,z) = \sum_{m=1}^{M} w1 \cdot \max\left(0, z^{\min} - \left(x_m + z_{m-1} - D_m\right)\right);$$

$$F2(x,z) = \sum_{m=1}^{M} w2 \cdot \max\left(0, x_m + z_{m-1} - D_m - z^{\max}\right),$$

$$z_m = x_m + z_{m-1} - D_m\,,$$

де $F1(\cdot)$ – витрати від дефіциту товару,

$F2(\cdot)$ – витрати від зберігання товару,

$w1$ – прибуток від продажу однієї одиниці товару;

$w2$ – витрати на зберігання однієї одиниці товару;

x_m – кількість товару, що закуповується у постачальника протягом m-го етапу;

z_m – кількість запасів товару в кінці m-го етапу;

z_0 – вихідна кількість запасів товару;

$z^{\min}, z^{\max}$ – мінімально допустимі та максимально допустимі запаси товару в кінці кожного етапу;

D_m – кількість товару, що продається протягом m-го етапу;

M – кількість етапів.

Кількість закуповуваного товару в постачальника обмежено в такий спосіб:

$$x^{\min} \le x_m \le x^{\max},\ m \in \overline{1,M},$$

де $x^{\min}, x^{\max}$ – мінімальна і максимальна кількість товару, що закуповується у постачальника протягом кожного етапу.

Пропонований метод на основі Q-навчання з динамічними параметрами складається з 13 кроків.

1. Ініціалізація

1.1. Задаємо параметри системи управління запасами – N, M, $w1$, $w2$, $x^{\min}, x^{\max}$, $z^{\min}, z^{\max}$, z_0, D_m, $m \in \overline{1,M}$ а також параметри алгоритму: $\rho^{\min}, \rho^{\max}$ для управління швидкістю навчання, $0 < \rho^{\min} < \rho^{\max} < 1$; $\varepsilon^{\min}, \varepsilon^{\max}$ для ε-жадібного підходу, $0 < \varepsilon^{\min} < \varepsilon^{\max} < 1$; $\theta^{\min}, \theta^{\max}$ для визначення важливості майбутньої винагороди, $0 < \theta^{\min} < \theta^{\max} < 1$.

1.2. Ініціалізуємо таблицю винагороди $Q = [Q(i, j)]$, $Q(i, j) = 0$, $i, j \in \overline{1,M}$.

2. Встановлюємо номер ітерації $n = 1$.

3. Обчислюємо параметри:

$$\rho(n) = \rho^{\max} - (\rho^{\max} - \rho^{\min}) \frac{n-1}{N-1};$$

$$\varepsilon(n) = \varepsilon^{\max} - (\varepsilon^{\max} - \varepsilon^{\min}) \frac{n-1}{N-1};$$

$$\theta(n) = \theta^{\min} + (\theta^{\max} - \theta^{\min}) \frac{n-1}{N-1}.$$

4. Задаємо початковий стан $s = z_0$.

5. Задаємо початковий номер етапу $m = 1$.

6. Вибирається дія a (кількість товару, що закуповується у постачальника), за якою виходимо зі стану s за ε-жадібним підходом. Якщо $U(0,1) < \varepsilon(n)$, тоді обираємо дію a випадковим способом із множини дозволених дій $\{x^{\min}, x^{\max}\}$, інакше обираємо дію a, таку, щоб $a = \arg\max_b Q(s, b)$, $b \in \{x^{\min}, x^{\max}\}$. Обрана дія a стає новим компонентом вектора кількості товару, що закуповується у постачальника, тобто $x_m = a$.

7. Якщо відбувається останній етап, тобто $m = M$, тоді переходимо до кроку 13, інакше $m = m + 1$.

8. Обчислюємо елемент таблиці поточної винагороди $R(s, a)$:

$$R(s,a) = -\left(w1 \cdot \max\left(0, z^{\min} - \left(a + s - D_m\right)\right) + w2 \cdot \max\left(0, a + s - D_m - z^{\max}\right)\right).$$

9. Оновлюємо запаси товару: $e = a + s - D_m$.

10. Обчислюємо елемент таблиці винагороди $Q(s,a)$:

$$Q(s,a) = (1-\rho(n))Q(s,a) + \rho(n)\left(R(s,a) + \theta(n) \cdot \max_b Q(e,b) \right),\ b \in \{x^{\min}, \ldots, x^{\max}\}.$$

11. Встановлюємо новий поточний стан $s = e$ і переходимо до кроку 6.

12. Якщо найкраще значення цільової функції на поточній ітерації менше від кращого значення за попередніми ітераціям, тобто $F(x) < F(x^*)$, то оновити вектор кількості товару, що закуповується у постачальника: $x^* = x$.

13. Якщо $n < N$, то перейти до кроку 3, інакше завершити роботу.

Пропонований метод на основі SARSA з динамічними параметрами збігається з попереднім методом на основі Q-навчання з динамічними параметрами за всіма кроками, окрім дев'ятого. Це крок реалізується так:

9. Вибирається дія c (кількість товару, що закуповується у постачальника), за якою потрібно перемістити зі стану e, використовуючи ε-жадібний підхід (якщо $U(0,1) < \varepsilon(n)$, то вибрати дію c випадковим способом із множини дозволених дій $\{x^{\min}, \ldots, x^{\max}\}$, інакше вибрати дію c як $c = \arg\max_b Q(e,b)$, $b \in \{x^{\min}, \ldots, x^{\max}\}$). Вибрана дія c стає новим компонентом вектора кількості товару, що закуповується у постачальника, тобто $x_m = c$. Обчислюється елемент таблиці винагороди за формулою:

$$Q(s,a) = (1-\rho(n))\, Q(s,a) + \rho(n)\left(R(s,a) + \theta(n)\, Q(e,c) \right).$$

Експерименти

Чисельне дослідження запропонованих методів проведемо з використанням пакету Python. Для методів Q-навчання та SARSA встановимо такі параметри навчання $\rho^{\min} = 0.1, \rho^{\max} = 0.9$ $\theta^{\min} = 0.1, \theta^{\max} = 0.9$. Експерименти проведемо на основі даних логістичної компанії "Ekol Ukraine"– це інтегрована логістична компанія, яка надає інтелектуальні клієнтоцентричні рішення в управлінні ланцюгом постачання, спрямовані на масштабування та розвиток бізнесу. Обсяг датасету становить 1000 реалізацій комп'ютерної техніки протягом одного року.

Залежність параметра $\theta(n)$ задамо зростаючою: $\theta(n) = \theta^{\min} + (\theta^{\max} - \theta^{\min})\frac{n-1}{N-1}$. Залежності параметрів $\rho(n)$ та $\varepsilon(n)$ задамо спадними $\rho(n) = \rho^{\max} - (\rho^{\max} - \rho^{\min})\frac{n-1}{N-1}$ та $\varepsilon(n) = \varepsilon^{\max} - (\varepsilon^{\max} - \varepsilon^{\min})\frac{n-1}{N-1}$. Графіки залежностей цих параметрів представлено на рис. 3. Зміна параметра $\theta(n)$ початкових ітераціях зменшує важливість нагороди, що робить пошук рішення більш глобальним, а на заключних ітераціях підвищує важливість нагороди, що робить пошук рішення більш локальним. Зміна параметра $\rho(n)$ на початкових ітераціях підвищує швидкість навчання, що робить пошук рішення більш глобальним, а на заключних

ітераціях зменшує швидкість навчання, що робить пошук рішення більш локальним. Зміна параметра $\varepsilon(n)$ на початкових ітераціях підвищує ймовірність випадкового вибору стану, що робить пошук рішення більш глобальним, а на заключних ітераціях зменшує ймовірність випадкового вибору стану, що робить пошук рішення більш локальним.

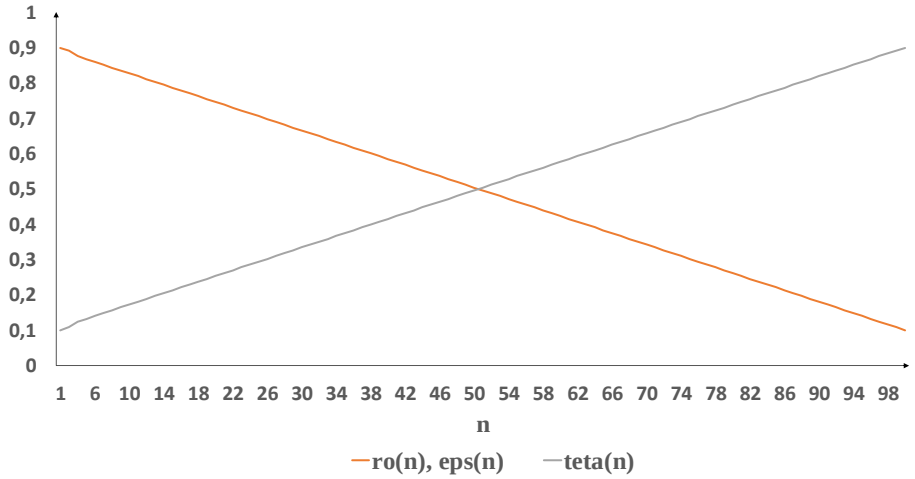

Рисунок 3. Залежність параметрів $\theta(n)$, $\rho(n)$ та $\varepsilon(n)$ від номера ітерації n

Результати навчання по тестовій вибірці запропонованої нейромережевої RCM-моделі з нейромережевою моделлю багатошарового персептрона (MLP) з двома шарами на основі критерію середньоквадратичної помилки представлено в табл. 1. З таблиці видно, що використання RCM зменшує середньоквадратичну помилку, чим підвищує точність управління буфером запасів. Порівняння по тестовій вибірці запропонованого методу Q-навчання з динамічними параметрами та з традиційним методом Q-навчання за середньоквадратичною помилкою і за кількістю ітерацій для вирішення задачі управління запасами здійснено в табл. 2. Аналогічні результати отримано для запропонованого методу SARSA з динамічними параметрами та з традиційним методом SARSA (табл. 3). З табл. 2–3 видно, що модифікація методів Q-навчання та SARSA за рахунок динамічних параметрів збільшує швидкість навчання за збереження середньоквадратичної помилки методу.

Таблиця 1. Порівняння нейромережевих моделей RCM та MLP

Середньоквадратична помилка нейромережевої моделі	
RCM	MLP
0.05	0.1

Таблиця 2. Порівняння запропонованого методу Q-навчання з традиційним

Середньоквадратична помилка методу		Кількість ітерацій	
запропонованого	традиційного	запропонованого	традиційного
0.05	0.05	300	2000

Таблиця 3. Порівняння запропонованого методу SARSA з традиційним

Середньоквадратична помилка методу		Кількість ітерацій	
запропонованого	традиційного	запропонованого	традиційного
0.05	0.05	300	2000

Висновки

Запропонована нейромережева модель обмеженої машини Коші має гетероасоціативну пам’ять та не має обмежень на ємність пам’яті. Вона використовує розподіл Коші, що підвищує збіжність параметричної ідентифікації та забезпечує високу точність управління буфером запасів.

Запропонована модифікація методів Q-навчання та SARSA за рахунок використання динамічних параметрів у правилі оновлення таблиці винагороди дає змогу підвищити швидкість навчання. Запропоновані модифікація завдяки дослідженню всього простору пошуку на початкових ітераціях та спрямованості пошуку на заключних ітераціях забезпечують високу точність розв’язання задачі управління запасами.

Запропоновані методи розширюють сферу застосування нейро-асоціативного навчання та навчання з підкріпленням, що підтверджується їх адаптацією для задач управління запасами. Це сприятиме підвищенню ефективності інтелектуальних комп’ютерних систем загального та спеціального призначення. Перспективами подальших досліджень є масштабування запропонованих методів на широкий клас задач штучного інтелекту.

Література

1. Mayo-Alvarez, L., Del-Aguila-Arcentales, S., Alvarez-Risco, A., Chandra Sekar, M., Davies, N. M., & Yáñez, J. A. (2024). Innovation by integration of Drum-Buffer-Rope (DBR) method with Scrum-Kanban and use of Monte Carlo simulation for maximizing throughput in agile project management. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1). https://doi.org/10.1016/j.joitmc.2024.100228
2. Melendez, J. R., Zoghbe Nuñez, Y. A., Malvacias Escalona, A. M., Almeida, G. A., & Layana Ruiz, J. (2018). Theory of constraints: A systematic review from the management context. *Espacios*, *39*(48).
3. Stopka, O., Zitrický, V., Ľupták, V., & Stopková, M. (2023). Application of specific tools of the theory of constraints – a case study. *Cognitive Sustainability*, *2*(1). https://doi.org/10.55343/cogsust.48
4. Bart, A., Delahaye, B., Fournier, P., Lime, D., Monfroy, É., & Truchet, C. (2018). Reachability in parametric interval Markov chains using constraints. *Theoretical Computer Science*, *747*, 48–74. https://doi.org/10.1016/j.tcs.2018.06.016
5. Haykin, S. (2008). Neural Networks and Learning Machines. *Pearson Prentice Hall New Jersey USA 936 pLinks* (vol. 3, p. 906). https://doi.org/978-0131471399
6. Shlomchak, G., Shvachych, G., Moroz, B., Fedorov, E., & Kozenkov, D. (2019). Automated control of temperature regimes of alloyed steel products based on multiprocessors computing systems. *Metalurgija*, *58*(3–4), 299–302.
7. Shvachych, G. G., Ivaschenko, O. V., Busygin, V. V., & Fedorov, Y. Y. (2018). Parallel computational algorithms in thermal processes in metallurgy and mining. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, (4), 129–137. https://doi.org/10.29202/nvngu/2018-4/19
8. Fedorov, E., Utkina, T., Nechyporenko, O., & Korpan, Y. (2020). Development of technique for face detection in image based on binarization, scaling and segmentation methods. *Eastern-*

European Journal of Enterprise Technologies, *1*(9–103), 23–31. https://doi.org/10.15587/1729-4061.2020.195369

9. Singh, U. P., Jain, S., Tiwari, A., & Singh, R. K. (2019). Gradient evolution-based counter propagation network for approximation of noncanonical system. *Soft Computing*, *23*(13), 4955-4967. https://doi.org/10.1007/s00500-018-3160-7
10. Sonika, Pratap, A., Chauhan, M., & Dixit, A. (2017). New technique for detecting fraudulent transactions using hybrid network consisting of full-counter propagation network and probabilistic network. In *Proceeding – IEEE International Conference on Computing, Communication and Automation, ICCCA 2016* (pp. 177–182). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CCAA.2016.7813713
11. Baggenstoss, P. M. (2019). Applications of projected belief networks (PBN). In *European Signal Processing Conference* (Vol. 2019 – September). European Signal Processing Conference, EUSIPCO. https://doi.org/10.23919/EUSIPCO.2019.8902708
12. Baggenstoss, P. M. (2019). On the duality between belief networks and feed-forward neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(1), 190–200. https://doi.org/10.1109/TNNLS.2018.2836662
13. Sountsov, P., & Miller, P. (2015). Spiking neuron network Helmholtz machine. *Frontiers in Computational Neuroscience*, *9*(APR). https://doi.org/10.3389/fncom.2015.00046
14. Kohonen, T. (2012). *Self-Organization and Associative Memory*, (3rd ed.). Berlin; New York: Springer-Verlag. 311 p. https://doi.org/10.1007/978-3-642-88163-3
15. Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018
16. Lobo, R. A., & Valle, M. E. (2020). Ensemble of binary classifiers combined using recurrent correlation associative memories. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 12320 LNAI, pp. 442–455). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-61380-8_30
17. Kobayashi, M. (2017). Quaternionic Hopfield neural networks with twin-multistate activation function. *Neurocomputing*, *267*, 304–310. https://doi.org/10.1016/j.neucom.2017.06.013
18. Du, K. L., & Swamy, M. N. S. (2014). *Neural Networks and Statistical Learning* (Vol. 9781447155713, pp. 1–824). Springer-Verlag London Ltd. https://doi.org/10.1007/978-1-4471-5571-3
19. Javidmanesh, E. (2017). Global stability and bifurcation in delayed bidirectional associative memory neural networks with an arbitrary number of neurons. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, *139*(8). https://doi.org/10.1115/1.4036229
20. Park, Y. (2010). Optimal and robust design of brain-state-in-a-box neural associative memories. *Neural Networks*, *23*(2), 210–218. https://doi.org/10.1016/j.neunet.2009.10.008
21. Khristodulo, O. I., Makhmutov, A. A., & Sazonova, T. V. (2017). Use algorithm based at hamming neural network method for natural objects classification. In *Procedia Computer Science* (Vol. 103, pp. 388–395). Elsevier B.V. https://doi.org/10.1016/j.procs.2017.01.126
22. Fischer, A., & Igel, C. (2014). Training restricted Boltzmann machines: An introduction. *Pattern Recognition*, *47*(1), 25–39. https://doi.org/10.1016/j.patcog.2013.05.025

23. Wang, Q., Gao, X., Wan, K., Li, F., & Hu, Z. (2020). A novel restricted Boltzmann machine training algorithm with fast Gibbs sampling policy. *Mathematical Problems in Engineering*, *2020*. https://doi.org/10.1155/2020/4206457
24. Bertsekas, D. P. (2019). *Reinforcement Learning and Optimal Control*. Belmont, MA: Athena Scientific.
25. François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. *Foundations and Trends in Machine Learning*, *11*(3–4), 219–354. https://doi.org/10.1561/2200000071
26. Goldberg, D. A., Katz-Rogozhnikov, D. A., Lu, Y., Sharma, M., & Squillante, M. S. (2016). Asymptotic optimality of constant-order policies for lost sales inventory models with large lead times. *Mathematics of Operations Research*, *41*(3), 898–913. https://doi.org/10.1287/moor.2015.0760
27. Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., & Silver, D. (2018). Rainbow: Combining improvements in deep reinforcement learning. In *32nd AAAI Conference on Artificial Intelligence, AAAI 2018* (pp. 3215–3222). AAAI Press. https://doi.org/10.1609/aaai.v32i1.11796
28. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236
29. Graesser, L., & Keng, W. L. (2019). *Foundations of Deep Reinforcement Learning: Theory and Practice in Python*. Boston: Addison-Wesley Professional.
30. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*, (2nd ed.). Adaptive Computation and Machine Learning. Cambridge: The MIT Press.
31. Matta, M., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Re, M., Silvestri, F., & Spanò, S. (2019). Q-RTS: A real-time swarm intelligence based on multi-agent Q-learning. *Electronics Letters*, *55*(10), 589–591. https://doi.org/10.1049/el.2019.0244
32. Kar, S., Moura, J. M. F., & Poor, H. V. (2013). 2D-learning: A collaborative distributed strategy for multi-agent reinforcement learning through consensus + innovations. *IEEE Transactions on Signal Processing*, *61*(7), 1848–1862. https://doi.org/10.1109/TSP.2013.2241057
33. Ottoni, A. L., Nepomuceno, E. G., Oliveira, M. S., & Oliveira, D. C. (2022). Reinforcement learning for the traveling salesman problem with refueling. *Complex and Intelligent Systems*, *8*(3), 2001–2015. https://doi.org/10.1007/s40747-021-00444-4

Рукопис отримано – 15/05/2024; прийнято до публікації – 20/06/2024.

Methods of solving the problem of stock management based on neuro-associative learning and reinforcement learning

Eugene Fedorov, Tetyana Utkina, Olga Nechyporenko

Abstract

Today, the development of intelligent methods aimed at solving inventory management problems is an urgent task. Many modern companies use the technology of constraint theory to improve and optimize their business processes, which provides dynamic inventory buffer management and is used for supply chain management. The aim of the work is to improve the efficiency of inventory management through neuro-associative learning based on the constrained Cauchy machine and reinforcement learning based on Q-learning and SARSA. To achieve this goal, a method based on a constrained Cauchy machine for inventory buffer management, a method based on Q-learning, and a method based on SARSA for general inventory management tasks are created. The proposed neural network model of the constrained Cauchy machine has a hetero-associative memory with no capacity limitations and provides high accuracy for inventory buffer management. The model uses the Cauchy distribution, which improves the convergence of the parametric identification method by comparison with the traditional restricted Boltzmann machine. Compared to the full Cauchy machine, the constrained Cauchy machine allows working with a larger memory size. Modification of the Q-learning and SARSA methods by means of dynamic parameters allows to increase the learning speed at a given level of the mean square error. Computational experiments have shown that controlling the importance of the reward, the learning rate parameter, and the parameter for ε-greedy approach for the Q-learning and SARSA methods allows making the solution search more global at the initial stages and more local at the final stages. The proposed methods allow expanding the scope of neuro-associative learning and reinforcement learning, which is confirmed by their adaptation for inventory management tasks, and contributes to the efficiency of intelligent computer systems for general and special purposes. Prospects for further research are the application of the proposed methods to other decision-making tasks, including the ones in the field of artificial intelligence.

Keywords: neuro-associative learning; learning with reinforcement; inventory management; bounded Cauchy machine; Q-learning method; the SARSA method.