

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

**ГРИШКО АНДРІЙ ОЛЕКСАНДРОВИЧ**

УДК 519.6:004.89

**ГІБРИДНІ МЕТОДИ МАШИННОГО НАВЧАННЯ В СИСТЕМАХ  
ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ДАНИХ**

05.13.23 – системи та засоби штучного інтелекту

Автореферат дисертації на здобуття наукового ступеня  
кандидата технічних наук

Харків – 2013

Дисертацією є рукопис.

Робота виконана в Харківському національному університеті радіоелектроніки Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор  
Удовенко Сергій Григорович,  
Харківський національний університет  
радіоелектроніки, м. Харків, професор кафедри  
електронних обчислювальних машин

Офіційні опоненти: доктор технічних наук, професор  
Каргін Анатолій Олексійович  
Донецький національний університет, м. Донецьк,  
завідувач кафедри комп'ютерних технологій;

доктор технічних наук, доцент  
Сіра Оксана Володимирівна  
Національний технічний університет «ХПІ»,  
м. Харків, професор кафедри  
комп'ютерного моніторингу та логістики

Захист відбудеться «26» червня 2013 р. о 13<sup>00</sup> годині на засіданні спеціалізованої вченої ради Д 64.052.01 Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

З дисертацією можна ознайомитись у бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

Автореферат розісланий «22» травня 2013 р.

Учений секретар спеціалізованої  
вченої ради Д 62.052.01

Є.І. Литвинова

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Розвиток інформаційних технологій на основі використання методів обчислювального інтелекту сприяв появі комп'ютерних систем інтелектуальної обробки інформації різного функціонального призначення (зокрема, систем обробки електронної біржової інформації та систем інтелектуального керування в умовах невизначеності). Сучасні дослідження свідчать про ефективність застосування в таких системах методів машинного навчання. Серед цих методів останнім часом найбільше розповсюдження отримали базовий та модифіковані методи навчання з підкріпленням (reinforcement learning (RL)), що базуються на взаємодії агента з зовнішнім середовищем, внаслідок якої агент самостійно визначає найбільш доцільну поведінку (стратегію) для досягнення оптимального в сенсі заданого критерію результату. При цьому в процесі навчання передбачається можливість короткочасових втрат з тим, щоб в подальшому отримати максимальній виграш на значному часовому інтервалі. Завдяки цьому навчання з підкріпленням є методом, концептуально придатним для ефективної роботи в інтелектуальних системах з високим рівнем внутрішніх та зовнішніх збурень (наприклад, в трейдингових системах, системах керування запасами, системами керування динамічними об'єктами в умовах невизначеності тощо)

Втім, не зважаючи на загальну методологічну основу, застосування методів машинного навчання в таких системах має свої особливості, що визначаються, насамперед, характером та цілями конкретних типів систем.

Слід відзначити, що використання машинного навчання з підкріпленням як для трейдингових систем, так і для систем інтелектуального керування передбачає необхідність вирішення в реальному часі оптимізаційних задач при визначенні стратегій поведінки агента з урахуванням мінімізації можливого ризику. Для цього можуть бути застосовані як традиційні методи оптимізації (наприклад, динамічне програмування), так і методи обчислювального інтелекту (еволюційна оптимізація). Реалізація процедури пошуку екстремуму Q-функції, пов'язаної з визначенням оптимальної стратегії, найбільш успішно може бути здійснена за допомогою нейромережових моделей. Останнім часом з'явилися численні наукові публікації, присвячені теоретичним дослідженням та практичному застосуванню методів машинного навчання з підкріпленням в інтелектуальних системах. Значний вплив на розвиток даного напрямку штучного інтелекту здійснили, зокрема, R. Sutton, A. Barto, T. Prescott, О.П. Мосалов, Д.В. Прохоров, В.Г. Редько та деякі інші вчені. В той же час все ще є відсутнім універсальний підхід до вирішення задач інтелектуальної обробки даних, пов'язаних з комбінованим застосуванням методів машинного навчання, нейромережових прогнозуючих моделей та еволюційних процедур оптимізації. Таким чином, актуальною та важливою науково-технічною задачею є розробка гібридних методів, що об'єднують можливості RL- методів, нейромережових моделей та методів оптимізації (зокрема, генетичної оптимізації та динамічного програмування) та дозволяють підвищити ефективність роботи інтелектуальних систем трейдингового прогнозування та керування стохастичними

процесами в умовах невизначеності. Задачі, що при цьому виникають, обумовили напрям досліджень даної дисертаційної роботи.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційну роботу виконано в рамках держбюджетної теми №ДР0110U000458 «Еволюційні гібридні методи і моделі інтелектуальної обробки інформації зі змінною структурою за умов невизначеності», що виконувалась у Харківському національному університеті радіоелектроніки, де автор брав участь як виконавець.

**Мета і задачі дослідження.**

Мета дослідження – розробка гібридних методів машинного навчання інтелектуальних систем різного функціонального призначення (на прикладі трейдингових систем та систем інтелектуального керування) для підвищення ефективності стратегій прийняття рішень.

Для досягнення зазначеної мети в роботі вирішуються такі задачі:

- дослідження задачі застосування методів машинного навчання в системах інтелектуальної обробки даних;
- розробка гібридних методів машинного навчання в трейдингових системах;
- розробка гібридних методів машинного навчання в системах інтелектуального керування стохастичними об'єктами;
- імітаційне моделювання роботи запропонованих гібридних методів та вирішення практичних задач.

*Об'єкт дослідження* – процеси інтелектуальної обробки даних з застосуванням методів машинного навчання.

*Предмет дослідження* – гібридні методи машинного навчання з підкріпленням в системах інтелектуальної обробки даних.

*Методи дослідження:* методи машинного навчання з підкріпленням – для визначення оптимальних стратегій прийняття рішень в інтелектуальних системах досліджуваних типів; методи генетичної оптимізації – для пошуку оптимальних схем обробки інформації в трейдингових системах, що використовують навчання з підкріпленням; методи стохастичного динамічного програмування – для пошуку оптимальних рішень в системах керування запасами, що використовують навчання з підкріпленням; методи нейромережевого моделювання – для перевірки теоретичних результатів та практичної реалізації запропонованих гібридних алгоритмів.

**Наукова новизна отриманих результатів.**

1. Вперше запропоновано метод адаптації структури технічних індикаторів до поточного стану біржового ринку з подальшим формуванням стратегій трейдингової інтелектуальної системи, що базується на використанні комбінованого RL-навчання та генетичних алгоритмів. Метод дозволяє агенту трейдингової системи аналізувати та оперативно використовувати для формування сигналів підкріплення найбільш важливу інформацію щодо стану зовнішнього середовища.

2. Вперше запропоновано гібридний метод керування запасами з використанням стохастичного динамічного програмування та техніки навчання з підкріпленням, що є сумісною з несепарабельним критерієм. Ідея методу полягає в апроксимації функції оцінювання за допомогою послідовної реєстрації усереднених

витрат. Стратегії, що при цьому застосовуються, гарантують збіжність обчислювального алгоритму.

3. Удосконалено метод нейромережевої апроксимації Q-функцій RL-алгоритму, який на відміну від аналогів дозволяє здійснювати корекцію конфігурації апроксимуючого багатозарового перцептронну.

4. Одержав подальший розвиток гібридний метод інтелектуального керування динамічними об'єктами, що є поширенням SARSA-методу та Q-методу та дозволяє брати до уваги як сусідні так і суттєво віддалені стани середовища. Нейромережева реалізація такого методу дозволяє зменшити середньоквадратичне відхилення оцінок Q-функції в околу точки, що аналізується, від номінального значення.

5. Удосконалено модель прогнозування часових рядів, процес навчання якої базується на застосуванні нейромережевого фільтра-предиктора, що на відміну від аналогів забезпечує високу швидкодію та якість прогнозів в умовах невизначеності. Запропонована модель може бути використана для прогнозування тренду сигналів підкріплення при інтелектуальному керуванні динамічними об'єктами.

**Практичне значення отриманих результатів.** Практичне значення полягає в тому, що отримані теоретичні результати реалізовано за допомогою нових процедур машинного навчання, що забезпечує високу якість обробки інформації в інтелектуальних системах. Зокрема, тестування запропонованої гібридної трейдингової системи на базі модуля «Genetic-Trade-M» підтверджує її працездатність та перспективи практичного використання для прийняття рішень щодо здійснення біржових транзакцій. Розроблена система за функціональними можливостями та якості отримуваних результатів перевищує можливості відомих некомерційних систем електронної біржової торгівлі, що використовують методи машинного навчання (акт ПП «Софтвеа Експерт» від 05.09.2012). Окремі положення, висновки та рекомендації дисертаційної роботи використано в навчальному процесі для підготовки та модифікації курсів «Методи та засоби обчислювального інтелекту» та «Системи цифрової обробки інформації» на кафедрі електронних обчислювальних машин Харківського національного університету радіоелектроніки, що підтверджено відповідним актом (акт від 12.09.2012 р.).

**Особистий внесок здобувача.** Всі основні результати, що виносяться на захист, отримані автором особисто. В опублікованих зі співавторами роботах здобувачу належать: [1] – розробка гібридного метода машинного навчання в трейдингових системах, що базується на навчанні з підкріпленням та генетичній оптимізації; [2, 9] – аналіз та модифікація існуючих методів машинного навчання в системах електронної біржової торгівлі; [3] – розробка модифікованого метода визначення стратегій з використанням машинного навчання з підкріпленням; [4, 17] – синтез інтелектуальної системи формування стратегій з використанням комбінованих індикаторів; [5] – аналіз та синтез гібридного метода RL-навчання в системах керування динамічними об'єктами; [6] – аналіз та синтез методів нейромережевої апроксимації Q-функцій; [8, 15] – розробка структури прогнозуючого нейромережевого фільтра для інтелектуальних систем прийняття рішень; [10, 19] – синтез та тестування гібридних методів машинного навчання для

прогнозу електроспоживання; [11, 13, 16] – моделювання та тестування інтелектуальних систем прийняття рішень з використанням RL-навчання.

**Апробація результатів дисертації.** Основні результати дисертаційної роботи доповідалися й обговорювалися на міжнародній конференції AUPEC – 2004 (Брісбен, Австралія), міжнародній конференції IDEAL - 2005 (Берлін, Німеччина), 8-й та 9-й міжнародних науково-технічних конференціях «Проблеми інформатики та моделювання» (Харків, 2008, 2009), 10-й та 11-й міжнародних науково-практичних конференціях «Сучасні інформаційні та електронні технології» (Одеса, 2009, 2010), 14-му міжнародному молодіжному форумі «Радіоелектроніка та молодь у XXI сторіччі» (Харків, 2010), 5-й міжнародній школі «Теорія прийняття рішень» (Ужгород, 2010), 1-й та 2-й міжнародних науково-технічних конференціях «Сучасні напрямки розвитку інформаційно-комунікаційних технологій та засобів управління» (Київ, 2010, 2011), 1-й та 2-й міжнародних науково-технічних конференціях «Інформаційні технології в навігації и керуванні: стан та перспективи розвитку» (Київ, 2010, 2011), 2-й міжнародній науково-практичній конференції «Информатика, математическое моделирование, экономика» (Росія, Смоленськ, 2012).

**Публікації.** За матеріалами дисертації опубліковано 20 печатних робіт, з них: 7 статей (1 одноосібно) у наукових фахових виданнях України, 2 статті за кордоном, 11 публікацій у збірниках праць міжнародних наукових конференцій та семінарів.

**Структура та обсяг дисертації.** Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та двох додатків. Повний обсяг дисертації становить 169 сторінок; обсяг основного тексту 153 сторінок; 37 ілюстрацій; 10 таблиць; список використаних джерел, що включає 115 найменувань та займає 12 сторінок; один додаток на 4 сторінках.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми, сформульовано мету та задачі дослідження, наведено відомості про зв'язки вибраного напрямку дослідження із планами організації, де виконана робота. Дано стислу анотацію отриманих в дисертації результатів, зазначено їх практичну цінність, обґрунтованість і достовірність. Наведено дані про використання проведених досліджень у практичних додатках.

**Перший розділ** присвячено аналізу стану застосування методів RL-навчання в системах інтелектуальної обробки даних. Відзначено, що інтелектуальні системи, де найбільш активно застосовуються такі методи, можна поділити на два класи: трейдингові системи (клас А) та системи керування стохастичними процесами за умов невизначеності (клас Б). Показано доцільність та актуальність розробки нових та модифікації існуючих гібридних методів RL-навчання в інтелектуальних системах різного функціонального призначення (на прикладі трейдингових систем та систем інтелектуального керування) з метою підвищення ефективності прийняття рішень. До трейдингових систем належать системи, що здійснюють поточний аналіз стану середовища (фінансового ринку, ринку цінних паперів, вартості енергетичних ресурсів тощо) та виробляють в реальному часі (на основі функцій оцінок, що

застосовують сигнали підкріплення) рекомендації щодо здійснення найменш ризикованих транзакцій. До систем інтелектуального керування стохастичними процесами, що можуть бути ефективно реалізовані з використанням методів RL-навчання, належать, наприклад, системи інтелектуального керування технологічними процесами з марковською моделлю поведінки, системи керування мобільними об'єктами тощо. Алгоритм Q-навчання ідентифікує дискретний набір станів довколишнього середовища  $S$  та виконує одну з можливих дій з множини  $A$ . У відповідь на дію  $a_t$  в час  $t$  при поточному стані  $s_t$ , агент системи отримує сигнал підкріплення  $r_t = r(s_t, a_t)$  від середовища, після чого воно переходить до нового стану  $s_{t+1} = \delta(s_t, a_t)$ . В алгоритмі використовуються функції переходу  $\delta(s_t, a_t)$ . Задача базового алгоритму Q-навчання – визначити та реалізувати стратегію  $\pi: S \rightarrow A$  згідно з поточним станом  $s_t$  (тобто  $\pi(s_t) = a_t$ ), що відповідає максимальному значенню суми сигналів підкріплення впродовж значного часу. Для формалізації цієї задачі використовують функцію  $V^\pi(s_t)$ , що є сумою всіх сигналів, отриманих алгоритмом, який стартує зі стану  $s_t$  з використанням стратегії  $\pi$ :

$$V^\pi \leftarrow \sum_{i=0}^{\infty} \gamma^i \cdot r_{t+i}, \quad (1)$$

де  $r_{t+i}$  – послідовність сигналів підкріплення;  $\gamma \in [0,1]$  – коефіцієнт дисконтування, що визначає поточну вагу майбутніх доходів. Оптимальна стратегія, що максимізує повний виграш, починаючи з будь-якого стану, може бути представлена у наступному вигляді:

$$\pi^* \leftarrow \arg \max_{\pi} V^\pi(s), \quad s \in S. \quad (2)$$

Агент в змозі реалізувати стратегію (2) на підставі оптимального значення функції  $V^\pi(s_t)$ , якщо є в наявності сигнал підкріплення та функція переходу. Зазвичай замість функції переходу використовують Q-функцію, значення якої зберігаються у двовимірній таблиці  $Q(s, a)$ . Через неможливість застосування таблиць значної вимірності при вирішенні практичних задач Q-функцію апроксимують за допомогою нейромереж. Існує чимало методів прискорення машинного навчання (алгоритм TD( $\lambda$ ), використання адаптивних критиків тощо). Втім аналіз показує, що актуальним залишається питання розробки гібридних методів обробки даних в інтелектуальних системах класів А та Б, пов'язаних з комбінованим застосуванням методів машинного навчання, нейромережових прогнозуючих моделей та еволюційних процедур оптимізації. За результатами такого аналізу сформульовано завдання наукового дослідження, здійснюваного в дисертації.

**Другий розділ** присвячено розробці гібридних методів машинного навчання з підкріпленням (RL-навчання) в системах електронної біржової торгівлі (трейдингових системах). Сутність RL-навчання, що використовується для систем цього класу, полягає в наступному: агент трейдингової системи має досліджувати поточні біржові ситуації та приймати рішення навіть при неповній інформації щодо цих ситуацій. Єдиний зворотний зв'язок, що отримується агентом від біржового ринку – скалярний сигнал підкріплення (позитивний, коли дії агенту є

прибутковими для трейдера та негативними в іншому випадку). Однією з функцій агенту може бути прогнозування фінансових рядів (наприклад, курсів валют) з використанням методів технічного аналізу, що базуються на застосуванні різних торгових моделей та спеціальних індикаторів стану ринку.

В роботі наведено приклад формування оптимальних стратегій з використанням алгоритма RL-навчання за результатами аналізу Q-значень, що дає змогу проаналізувати принцип роботи трейдингової системи з двома індикаторами фінансового ринку та двома типами транзакцій агенту (купівля (buy) або продаж (sell)). Слід зазначити, що в сучасному технічному аналізі використовують велику сукупність індикаторів, що допомагають визначити тенденції зміни показників біржового ринку. На цей час існує більш, ніж триста технічних індикаторів, що активно використовуються. Проаналізувати поведінку такої кількості індикаторів у реальному часі неможливо навіть із застосуванням комп'ютерної техніки.

В дисертаційній роботі вперше запропоновано метод адаптації структури технічних індикаторів до поточного стану біржового ринку з подальшим формуванням стратегій трейдингової інтелектуальної системи, що базується на використанні комбінованого RL-навчання та генетичних алгоритмів (ГА). Показано, що для оцінки ефективності роботи трейдингових систем, побудованих на основі RL-навчання, найбільш доцільним є використання коефіцієнту Стерлінга ( $Kc$ ), що визначається відношенням прогнозованої прибутковості до максимально можливих втрат, або коефіцієнту позитивних наслідків ( $Kпн$ ), що визначається відсотковим відношенням прибуткових та збиткових транзакцій трейдингової системи. На практиці часто використовують як сигнали для входу до ринку або виходу з нього перетин двох ковзних середніх з різними періодами. Якщо швидке (з меншим періодом) ковзне середнє перетинає повільне (з більшим періодом) ковзне середнє знизу догори – це сигнал до купівлі, що відповідає входу до довгої позиції або виходу з короткої позиції. Якщо ж швидке ковзне середнє перетинає повільне згори донизу, то це сигнал до продажу. Ця техніка має назву методу подвійних перетинів. Таким чином, трейдингова система має відображувати два типи торгових стратегій – відкриття (вхідна стратегія) та закриття (вихідна стратегія).

Розглянемо деякі технічні індикатори, комбінації яких можуть бути ефективно застосовані в трейдинговій системі, що працює на основі алгоритмів RL і ГА за ітераційними обчислювальними схемами Q-навчання. При використанні ковзного середнього (moving average (MA)) усереднюються ціни акції за деякий період. Так, 9-денні та 40-денні індикатори MA визначаються наступним чином:

$$MA_9(n) = \frac{1}{9} \sum_{i=0}^9 C(n-i), \quad MA_{40}(n) = \frac{1}{40} \sum_{i=0}^{40} C(n-i),$$

де  $C(n)$  – остання ціна закриття.

Рекомендації до купівлі або продажу відповідають наступним умовам:

$MA_{Buy}(n) = 1$  если  $(MA_9(n-1) < MA_{40}(n-1))$  и  $(MA_9(n) > MA_{40}(n))$ .

$MA_{Sell}(n) = 1$  если  $(MA_9(n-1) > MA_{40}(n-1))$  и  $(MA_9(n) < MA_{40}(n))$ .

Якщо в період  $n$   $MA_{Buy}(n) = 1$ , то формується сигнал купівлі; відповідно, якщо в період  $n$   $MA_{Sell}(n) = 1$ , то формується сигнал продажу.

Індикатор імпульсу (Momentum (n)) характеризує швидкість зміни цін.



Стохастичний індикатор (Stochastic (n)) відображує положення поточної ціни відносно діапазону цін за певний період в минулому.

При використанні ГА для визначення торгових стратегій кожному хромосому можна інтерпретувати як можливе рішення трейдингової системи у поточній ситуації (при цьому в якості таких рішень визначимо вибір позицій трейдера). Алгоритм прийняття рішень використовує значення набору індикаторів ринків, що об'єднуються до бінарних строк ГА. Наприклад, в разі згаданого вище методу двійних перетинів формується сигнал продажу, коли короткі середні рухи нижче ніж більш довгі. Індикатор «MASell», що відповідає цьому сигналу, приймає значення 1, якщо умову для продажу виконано, та значення 0 в протилежному випадку. Відзначимо, що команда з продажу в цьому правилі кодується як «0» в молодшому біті строки, а правило прийняття довгої позиції для купівлі формує «1» в цьому біті. Правила для прийняття короткої або довгої позицій називають правилами входів. Довжина правил та кодуєчих їх бінарних строк збільшується зі збільшенням кількості індикаторів, що використовуються у трейдинговій системі. Сукупність таких правил можна розглядати як стратегію. Вочевидь, що при великій кількості комбінацій значень індикаторів та операторів, що їх пов'язують, ГА не в змозі розглянути весь простір стратегій для знаходження оптимуму. В запропонованому методі використовуються 10 індикаторів як для купівлі, так і для продажу (тобто всього 20 індикаторів). Для базового варіанту гібридної трейдерної системи випадковим чином сформуємо по 150 можливих правил кожного типу (вход та вихід). З цієї сукупності правил (також випадковим чином) комбінуємо 150 пар стратегій, що ранжуються згідно зі своєю прибутковістю, а потім беруться до уваги при створенні початкового покоління правил. Згідно з ГА здійснюється покроковий процес генерування нових правил трейдингової системи. Обмін генетичного матеріалу здійснюється за допомогою оператора кросовера, що комбінує пару правил-батьків та формує пару нових правил-потомків. В алгоритмі використовується також оператор мутації для активізації оновлення бінарних строк. Процес продовжується до виконання заданої умови останову ГА, після чого визначається найкраща пара правил (найкраща стратегія). Застосування ГА в трейдинговій системі дає можливість оперативного визначення для поточної біржової ситуації комбінацій найбільш корисних індикаторів та правил, що можуть бути використані для представлення стану зовнішнього середовища. Приклад правила продажу, що використовує індикатори «MASell», «MomentumSell» та «StochasticBuy», наведено в таблиці 1.

Таблиця 1

Приклад бінарної строки правила продажу

MASell	Або	MomSell	Та	StochBuy	Позиція
1	0	1	1	0	0

В цьому прикладі рішенням є правило прийняття короткої позиції при продажу. Логічний запис цього правила: «ЯКЩО MASell = 1 АБО (MomentumSell = 1 ТА StochasticBuy = 0), ТО прийняти коротку позицію».

На рис. 1. наведено загальну схему прийняття рішень в запропонованому варіанті трейдингової системи. Згідно з цією схемою відібрані з використанням ГА

стратегії надходять до RL-модулю, основанийого на алгоритмі Q-навчання.

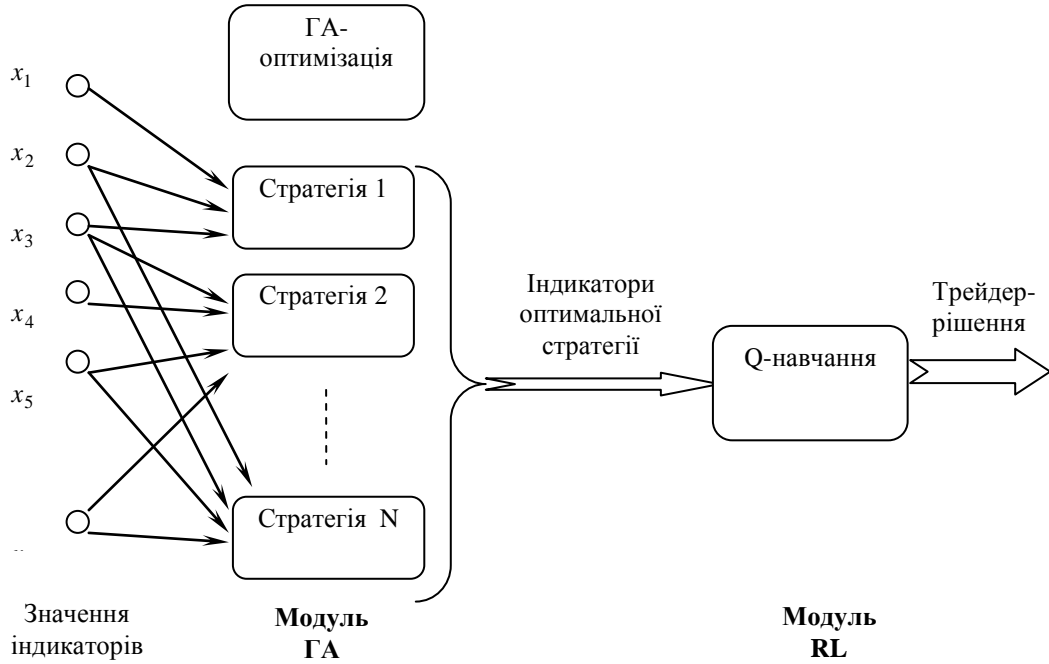


Рисунок 1 – Схема визначення оптимального набору індикаторів

Набір індикаторів та станів дозволяє визначити Q-значення, що використовуються на етапі прийняття трейдинговою системою відповідних рішень (SELL (продаж), BUY (купівля) або NEUTRAL (відсутність операцій)). В процесі реалізації системою прийнятої стратегії сигнал підкріплення визначає різницю між значеннями функції пристосованості у суміжні моменти часу. Аналіз знаку цієї різниці використовується для прийняття довгих та коротких позицій.

В другому розділі також модифіковано метод нейромережевої апроксимації Q-функцій RL-алгоритму, що дозволяє здійснювати корекцію конфігурації апроксимуючого багаточарового перцептронну (БШП), що використовується в подальшому для прийняття стратегій на основі гібридних алгоритмів машинного навчання. Для навчання БШП, що застосовується в трейдинговій системі, пропонується використання модифікованого алгоритму Левенберга-Марквардта. Для отримання раціональної конфігурації БШП пропонується наступний підхід. Маємо набір моделей з  $N$  БШП с різними архітектурами. Навчання мереж здійснюється на наборі  $\mathcal{T} = \{(x_i, y_i) : i = \overline{1, n}\} \subset \mathbf{R}^d \times \mathbf{R}$ . Вектор параметрів мережі позначимо як  $\underline{w}_j \in \mathbf{R}^{p_j}$ , а нейромережеву реалізацію Q-функції як  $\eta_j(x, \underline{w}_j)$ ,  $j = \overline{1, N}$ . Як критерій редукції моделей приймемо середньоквадратичну похибку

$$e_j(\underline{w}_j) = \frac{1}{n} \sum_{i=1}^n [y_i - \eta_j(x_i, \underline{w}_j)], \quad (3)$$

мінімум якої досягається відповідним настроюванням параметрів мережі.

Редукція моделей БШП, що застосовуються для апроксимації Q-функцій, здійснюється наступним чином: задаються  $N$  БШП різної архітектури (кількість циклів навчання  $k$ , період дискримінації  $m < k$ ); кожні  $m$  циклів перевіряється умова редукції та видаляються всі БШП, для яких ці умови виконуються.

Загальну структуру гібридної системи вибору торгових стратегій з

використанням запронованого підходу наведено на рис.2. Після пуску системи дані з серверу передаються до модулю аналізу поточної ситуації на біржевому ринку (АС), розраховуються значення індикаторів та визначається поточний стан. Далі визначається доцільність корекції набору індикаторів на поточний час (модуль ГА) та оцінюється необхідність корекції БШП, що застосовується для апроксимації Q-функцій (модуль БШП). З урахуванням результатів роботи модулів ГА та БШП дані надходять до модулю RL, який поновлює Q-значення з використанням БШП та виробляє рекомендації трейдеру з вибору оптимальних торгових стратегій. Трейдер намагається максимізувати виграш з урахуванням значення та знаку сигналу підкріплення  $r_t$ , що визначає різницю між значеннями прогнозованого виграшу в моменти часу  $t-1$  і  $t$ .

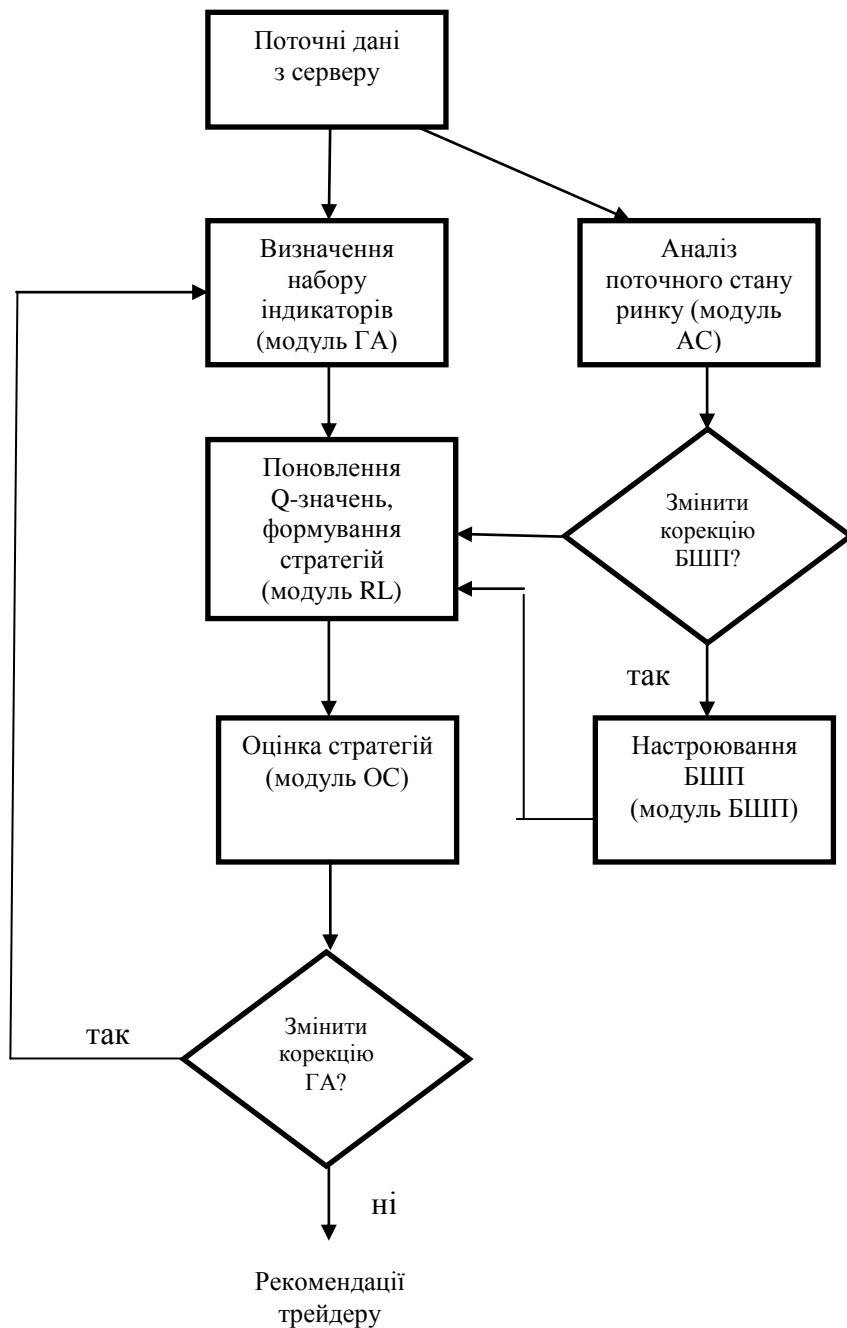


Рисунок 2 – Схема визначення стратегій в гібридній трейдинговій системі

В третьому розділі розглядається можливість використання гібридних методів RL-навчання з підкріпленням при вирішенні задач інтелектуального керування для важливих класів стохастичних керованих систем: управління запасами та цифрового керування динамічними об'єктами (тобто інтелектуальних систем класу Б).

Регулювання рівня запасів у складних технологічних комплексах у значній мірі впливає на динаміку їх поведінки (при цьому стан визначається рівнем запасів), а невизначеність даних (запитів, надходжень, вартості, наявності засобів виробництва) додає таким комплексам стохастичні властивості. Оптимальне управління такою системою вимагає розробки стратегії використання запасів, що дозволяє мінімізувати математичне очікування вартості на деякій сукупності стратегій. Класичне рішення для оптимізації вибору стратегії за сепарабельним критерієм може базуватися на динамічному програмуванні (зокрема, математичне очікування є сепарабельним). Нажаль, для більшості випадків оптимізація математичного очікування є занадто ризикованою. Доцільно вибрати критерій оптимізації, що забезпечував би робастні рішення, навіть якщо він не є сумісним з класичною технікою динамічного програмування. До статистичних характеристик випадкової змінної величини  $C$ , що складають інтерес для задачі, яка розглядається, крім математичного очікування  $E(C)$  та дисперсії  $var(C) = E(C - E(C))^2$  слід віднести поширені в системах управління ризиками характеристики Risk-At-Value (RAV) та Value-At-Risk (VAR). RAV означає (для деякого порогу витрат  $C'$ ) ймовірність того, що  $C \geq C'$ , а VAR для заданого порогу ризику відповідає мінімальному  $C'$ , для якого  $P(C > C') \leq r$ .

Розглянемо можливість використання в системах управління запасами техніки навчання з підкріпленням. В дисертаційній роботі для вирішення цієї задачі вперше запропоновано гібридний метод з використанням стохастичного динамічного програмування (СДП) та техніки навчання з підкріпленням, що є сумісною з несепарабельним критерієм. На прикладі RL-навчання визначимо стратегію, що є оптимальною за компромісним критерієм «математичне очікування/ризик» несепарабельного типу  $-\alpha E + \alpha VAR; 0 \leq \alpha \leq 1$

Для опису цієї проблеми використовуються: стохастичний процес  $p(\cdot)$  (в часі від 0 до  $T$ ); рівняння еволюції системи:  $x_{t+1} = f(t, u_t, x_t)$ ,  $f(\cdot)$  та стратегія  $u(\cdot)$ , що мінімізує математичне очікування сумарних витрат  $c_1(t_1, u_1) + c_2(t_2, u_2) + \dots + c_T(t_T, u_T)$  ( $u_t$  означає  $u(t, x_t)$ ). Основою принципу декомпозиції Беллмана є спосіб визначення функції  $V(t, x_t)$  для знаходження оптимальної стратегії  $u(\cdot)$ , що забезпечує виконання наступної рівності:

$$V(t, x_t) = E [c_t(t, u_t) + c_{t+1}(t+1, u_{t+1}) + \dots + c_T(t_T, u_T) | p(t) = \tilde{p}], \quad (4)$$

де  $x_t = \tilde{x}$  та  $x_{t+1} = f(t, u_t, x_t)$ .

Для вирішення цієї задачі достатньо в кожний момент часу визначити значення  $u_t \in \arg \min_u [c_t(t, u) + V(t+1, x_{t+1})]$ . Функція  $V(\cdot)$  називається функцією Беллмана або функцією оцінювання, або ж (з позицій RL-

навчання) функцією вартості. Будемо розглядати  $V(\cdot, \cdot)$  як компроміс вигляду  $-\alpha \underline{E} + \alpha VaR$ , що бере до уваги не лише математичне очікування витрат, але й ризик. Ідея, що пропонується, полягає в апроксимації  $V(\cdot, \cdot)$  за допомогою сумарної вартості. При цьому застосовується стратегія, що базується на повільних процедурах (послідовно здійснюється деяка кількість симуляцій перед початком використання  $V(\cdot, \cdot)$ ). Такий метод є робастним по відношенню до збіжності стратегій. Введення поняття ризику здійснюється шляхом застосування представлення типу «*VAR*». Модуль екстраполяції, що дозволяє екстраполювати «*VAR*» для неперервних рівнів запасів, передбачає необхідність введення обмежень на функції оцінювання, що приводить до задачі квадратичною оптимізації з лінійними обмеженнями. Комбіновану схему формування функції оцінювання наведено на рис.3. СДП реалізує першу функцію оцінювання без «*VAR*» та з дещо спрощеним стохастичним процесом (початковий пункт RL-навчання). Комбінування RL-навчання (за алгоритмом  $TD(0)$ ) та СДП дозволяє використовувати гарантовану стійкість СДП та можливість RL-методів трансформувати управління за математичним очікуванням витрат до стратегії, що використовує поняття ризику (після RL-навчання під час фази симуляції). Модуль ФО реалізує екстраполяцію опуклими функціями. СДП реалізується лише для попередніх розрахунків, що не беруть до уваги поняття ризику. Модуль «Критик» дозволяє більш повно враховувати поняття *RaV* та *VAR*.



Рисунок 3 – Комбінована схема формування функції оцінювання

Фіксованими є наступні параметри: частота прийняття рішень (кількість симуляцій між кожним значенням  $V(\cdot, \cdot)$  та  $\pi$ ; крок спуску; поріг довіри (близько

5%); параметр компромісу між математичним очікуванням та VAR. Модуль «Критик» має можливість побудови ломаної кривої функції оцінювання  $V(s, t, x)$  для заданої дії. Це дозволяє реалізувати оптимізацію вибору рішення, як квадратичну опуклу оптимізацію з лінійними обмеженнями. В роботі наведено послідовність реалізації базового варіанту запропонованого методу інтелектуального управління запасами та показано, що стратегії, які при цьому застосовуються, гарантують збіжність обчислювального алгоритму.

Для дослідження проблеми використання RL-методів в інтелектуальних цифрового керування динамічними об'єктами запропоновано гібридний метод, що є подальшим розвитком сумісного використання SARSA-алгоритму та Q-алгоритму та дозволяє брати до уваги як сусідні так і суттєво віддалені стани середовища.

Після  $n$  застосувань TD-методу можна отримати виграш, що визначається наступним рівнянням :

$$R_t = R_t^n = r_{t+1} + \gamma r_{t+2} + \dots + \gamma^n (r_{t+n+1} + \gamma V(s_{t+n+1})) . \quad (5)$$

Використання оцінки (7) дозволяє реалізувати процедуру вигляду:

$$V(s_t) \leftarrow V(s_t) + \alpha [R_t^n - V(s_t)] . \quad (6)$$

Введемо до (8) усереднені зважені оцінки  $R_t$  :

$$R_t = \sum_i a_i R_t^i ; \sum a_i = 1 . \quad (7)$$

Окремим випадком зважених оцінок тут є експоненційно зважені оцінки, що підвищують значущість майбутніх результатів з часом:

$$R_t = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^i R_t^i . \quad (8)$$

Це дозволяє спростити реалізацію RL-навчання шляхом використання попередніх значень сигналів замість оцінок майбутніх значень. Для цього введемо допоміжний параметр, що визначається за наступною рекурсією:

$$e_t(s) = \begin{cases} \gamma \lambda e_{t-1}(s), & s \neq s_t \\ \gamma \lambda e_{t-1}(s) + 1, & s = s_t \end{cases} . \quad (9)$$

Перерахунок узагальнених функцій виграшу буде тепер здійснюватися для кожного стану пропорційно значенню його коефіцієнту (11):

$$V(s_t) \leftarrow V(s_t) + \alpha e(s_t) [r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] . \quad (10)$$

Суть запропонованого підходу, що представлено в роботі послідовністю обчислювальних операцій з реалізації процедур (5-10), полягає в заміщенні станів, що використовують прогнозовані значення сигналів підкріплення попередніми зваженими станами, що використовують поточні значення сигналів підкріплення. Відповідний алгоритм є розширенням SARSA-алгоритму та  $Q(\lambda)$ -алгоритму.

В заключній частині третього розділу запропоновано удосконалену структуру моделі прогнозування, алгоритм навчання якої базується на застосуванні нейромережевого фільтра-предиктора, що на відміну від існуючих забезпечує високу швидкодію та якість прогнозів в умовах нестаціонарності та невизначеності.

Запропонована модель може бути використана для прогнозування тренду сигналів підкріплення при інтелектуальному керуванні динамічними об'єктами згідно з розглянутим вище алгоритмом. Фільтр складається з  $n$  паралельно з'єднаних адаптивних лінійних асоціаторів з  $nd + 1$  входами кожний, де  $d \geq 1$  – глибина передісторії прогнозування часового ряду. Кожен з асоціаторів використовує  $nd + 1$  настроюваних синаптичних ваг  $w_{ij}$ ,  $j = 1, 2, \dots, n$ ;  $i = 0, 1, \dots, nd$ .

У четвертому розділі розглядаються результати практичної реалізації інтелектуальних систем класів А та Б з машинним RL-навчанням, що використовують методи, запропоновані в теоретичній частині дисертаційної роботи.

Для оцінки ефективності гібридної трейдингової системи, запропонованої в розділі 2, були проведені тестові експерименти з використанням даних FX-ринку. В першій серії експериментів для тестування системи були обрані дані, зареєстровані з 5-хвилинним періодом дискретизації (ці же дані були використані для тестування відомої трейдингової системи Демпстера, що дає можливість проведення порівняльного аналізу). Дані (43 700 добових записів) було розбито на дві частини: in-sample та out-sample періоди. In-sample період застосовувався для того, щоб оптимізувати ефективність трейдингової системи з метою визначення найкращої стратегії, використовуючи гібридний метод, оснований на ГА та RL-навчання. Out-sample період використовувався для тестування цієї стратегії. Після початкового ГА-навчання індикатори подавалися до модулю RL. При тестуванні були задані наступні параметри ГА та RL: початкова популяція – 180 стратегій; ймовірність кросоверу – 0.5; ймовірність мутації – 0.02;  $\alpha = 0.5$ ,  $\gamma = 0.85$ ,  $\lambda = 0.25$ .

Рис. 4 підтверджує ефективність досліджуваної системи за out-sample період для двох цінних рядів (обмінні курси EUR/USD та AUD/USD).

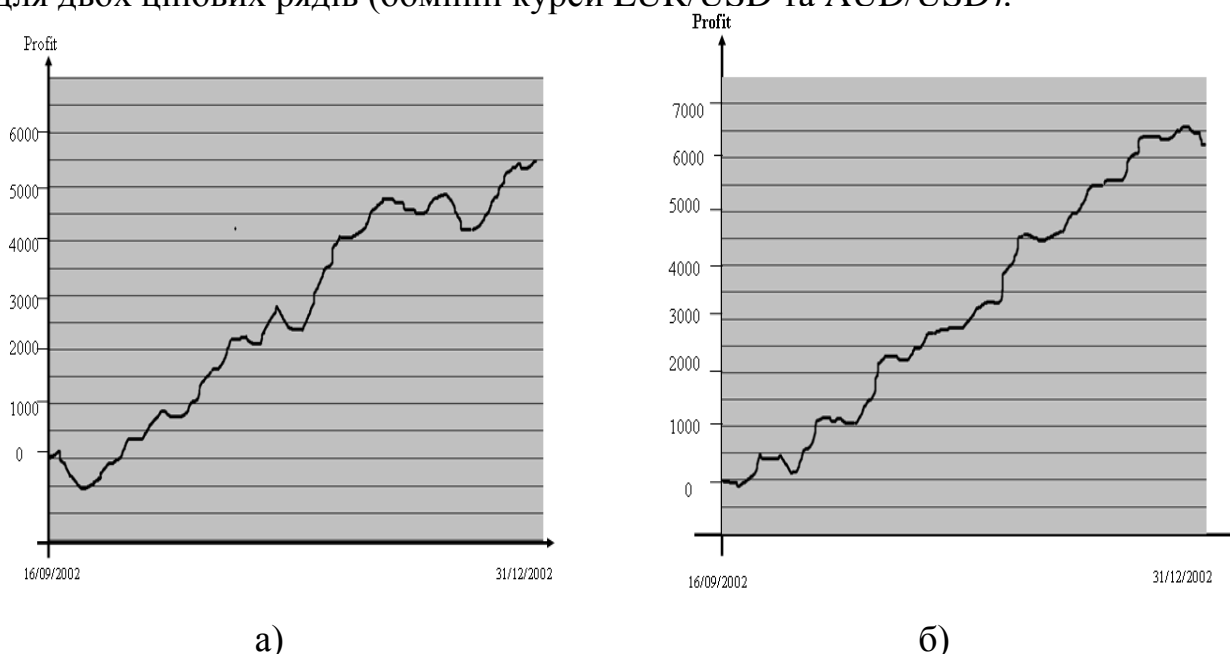


Рисунок 4 – Прибуток, отриманий гібридною системою на даних FX-ринку: а) EUR/USD; б) AUD/USD

У другій серії експериментів, крім статистичних даних FX-ринку, для тестування системи були використані дві стохастичні моделі: NM (нестационарний випадковий обхід) та AR (стаціонарний AR(2)-процес. Було продовжено порівняння

можливостей запропонованої гібридної трейдингової системи та системи Демпстера. Відзначимо, що розроблена система (на відміну від системи Демпстера) дозволяє приймати нейтральні позиції (НП). В таблиці 2 наведено деякі результати тестування трьох можливих варіантів трейдингової системи (ГА з НП; ГА+RL без НП; ГА+RL з НП). Тут відображені значення річного прибутку, а також значення коефіцієнтів Стерлінга та позитивних наслідків (КПН) для періодів in-sample (1) та out-sample (2). Результати тестування підтверджують ефективність розробленої системи. Використання зміни комбінацій технічних індикаторів та введення НП дозволяють підвищити прибутковість стратегій. Так прибутковість роботи RL-системи з НП (в режимі out-sample) перевищила прибутковість системи Демпстера на 3.6%, 11.7% та 10.6% для пар GBP/USD, EUR/USD и AUD/USD відповідно.

Таблиця 2

Результати тестування для курсу AUD/USD

AUD/USD	ГА з НП		ГА+RL без НП		ГА+RL з НП	
	1	2	1	2	1	2
Річний прибуток	5.3%	3.7%	13.0%	12.2%	18.9%	14.3%
КПН	53%	52%	56%	59%	63%	62%
Коефіцієнт Стерлінга	0.05	0.04	0.32	0.28	0.46	0.34

Систему було реалізовано на базі оригінального модуля «Genetic-Trade-M».

Схема комбінованого застосування RL-навчання та СДП, запропонована в розділі 3, була протестована для задачі управління гідравлічними запасами при використанні різних типів електростанцій в умовах стохастичності запитів на поставку електроенергії. Під час моделювання були отримані графіки для комбінованого критерію  $(1-\alpha)E + \alpha VaR$  при  $\alpha = 0; 0.1; 0.2; 0.4; 0.6; 0.8$  з використанням 1800 сценаріїв. Аналіз результатів показав суттєве зменшення рівня ризику отримання незадовільних рішень (в середньому на 55 %) при використанні комбінованого метода в порівнянні з результатами, отриманими з використанням стандартної процедури СДП з сепарабельним критерієм.

В четвертому розділі запропоновано також структуру модифікованої модульної системи моделювання процесів цифрового керування (СМ-RL) з використанням комбінованих методів машинного навчання (рис. 5).

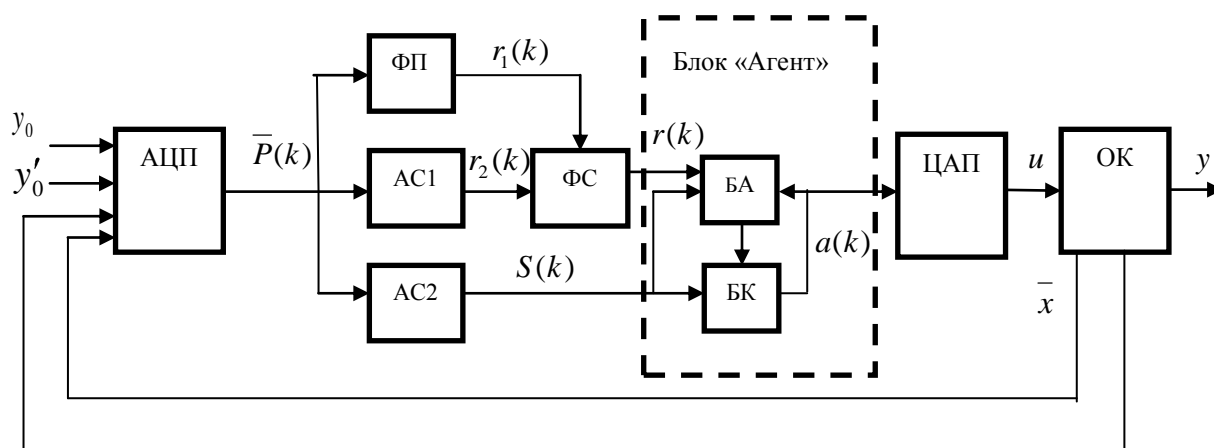


Рисунок 5 – Структура модифікованої системи моделювання



Вектор вхідних сигналів системи складається з завдання  $u_0$ , швидкості зміни завдання  $u'_0$ , виходу  $u$  та вектора змінних стану об'єкту керування (ОК). В результаті обробки вектора вхідних сигналів система формує вектор управління  $u$ . Схема містить модуль аналого-цифрового перетворювання (АЦП), на виході якого формується вектор дискретних сигналів, що надходить далі на аналізатори сигналів (АС1 и АС2) та на фільтр-предиктор ФП. Модуль АС1 визначає значення сигналу підкріплення  $r_2(k)$ , а модуль АС2 – сигналу стану середовища  $s(k)$ . Блок керування БК формує сигнал дії на ОК  $a(k)$ , а блок адаптації БА здійснює корекцію Q-функції згідно з модифікованим алгоритмом RL-навчання. Метою функціонування блоку «Агент» є максимізація сумарної оцінки виграшу. Розглянуту систему було використано для моделювання систем цифрового керування деякими технічними ОК, зокрема, процесом керування газоперекачувальним агрегатом.

У додатках наведено акти про впровадження та практичне застосування результатів, отриманих у дисертаційній роботі.

## ВИСНОВКИ

У дисертаційній роботі наведено результати, які відповідно до поставленої мети є вирішенням науково-технічної задачі розробки гібридних методів, що об'єднують можливості RL-методів, ШНМ-моделей та методів оптимізації (зокрема, генетичної оптимізації та динамічного програмування) та дозволяють підвищити ефективність роботи інтелектуальних систем трейдингового прогнозування та керування стохастичними процесами в умовах невизначеності. На підставі проведених теоретичних та експериментальних досліджень отримано такі найсуттєвіші результати:

1. Новий метод адаптації структури технічних індикаторів до поточного стану біржового ринку з подальшим формуванням стратегій трейдингової інтелектуальної системи, що базується на використанні машинного навчання з підкріпленням та генетичних алгоритмів. Запропонований метод дозволяє агенту трейдингової системи аналізувати та оперативно використовувати для формування сигналів підкріплення найбільш важливу поточну інформацію.

2. Новий гібридний метод керування запасами з використанням стохастичного динамічного програмування та машинного навчання з підкріпленням, що є сумісною з несепарабельним критерієм та бере до уваги ризик виникнення негативних наслідків. Ідея методу полягає в апроксимації функції оцінювання за допомогою послідовної реєстрації усереднених витрат. Стратегії, що при цьому застосовуються, гарантують збіжність відповідної обчислювальної процедури.

3. Удосконалений метод нейромережевої апроксимації Q-функцій RL-алгоритму, який на відміну від аналогічних методів дозволяє корегувати конфігурацію апроксимуючого багатозарового перцептронну.

4. Розвинений гібридний метод інтелектуального керування динамічними об'єктами, що є поширенням SARSA-методу та Q-методу та дозволяє брати до уваги як сусідні так і суттєво віддалені стани середовища. Реалізація такого методу за допомогою багатозарового перцептронну дозволяє зменшити середньоквадратичне

відхилення оцінок Q-функції в околу точки, що аналізується, від номінального значення.

5. Удосконалена модель прогнозування часових рядів (зокрема, тренду сигналів підкріплення), процес навчання якої базується на застосуванні нейромережевого фільтра-предиктора, що на відміну від аналогів забезпечує високу швидкодію та якість прогнозів в умовах невизначеності. Запропонована модель може бути використана при інтелектуальному керуванні динамічними об'єктами.

6. Розроблені методи було програмно реалізовано та використано для ряду практичних впроваджень. Зокрема, тестування запропонованої гібридної трейдингової RL-системи на базі модулю «Genetic-Trade-M» підтверджують її працездатність та перспективи практичного впровадження. Розроблена система за функціональними можливостями та якістю отримуваних результатів перевищує можливості відомих некомерційних систем електронної біржової торгівлі. Окремі положення, висновки та рекомендації дисертаційної роботи використано в навчальному процесі для підготовки та модифікації курсів «Методи та засоби обчислювального інтелекту» та «Системи цифрової обробки інформації» на кафедрі електронних обчислювальних машин Харківського національного університету радіоелектроніки.

## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Гришко А.А. Применение гибридных методов машинного обучения в компьютерных трейдинговых системах / А.А. Гришко, С.Г. Удовенко, Л.Э. Чалая // Системні технології. – №3 (68). – 2010. – С. 84 - 92.

2. Гришко А.А. Комбинированные методы машинного обучения в системах электронной биржевой торговли / А.А. Гришко, С.Г. Удовенко, Л.Э. Чалая // Вестник Херсонского национального технического университета. – №2 (38). – 2010. – С. 187 - 191.

3. Гришко А.А. Определение стратегий в трейдинговых системах на основе методов машинного обучения / А.А. Гришко, С.Г. Удовенко, Л.Э. Чалая // Бионика интеллекта. – 2010. – № 1 (72). – С. 18 - 23.

4. Гришко А.А. Применение методов машинного обучения в стохастических системах управления запасами / А.А. Гришко // Автоматизированные системы управления и приборы автоматики. – 2011. – Вып.154. – С. 91-96.

5. Гришко А.А. Интеллектуальная система формирования биржевых стратегий с использованием комбинированных индикаторов / А.А. Гришко, С.Г. Удовенко, Л.Э. Чалая // Біоніка інтелекту. –2011. –№1(75) – С.88-93.

6. . Гришко А.А. Нейросетевая аппроксимация Q-функций в трейдинговых системах / А.А. Гришко, С.Г. Удовенко // Системні технології. – №3 (74). – 2011. – С. 118 - 125.

7. Гришко А.А. Гибридные методы машинного обучения в системах управления динамическими объектами / А.А. Гришко, С.Г. Удовенко, Л.Э. Чалая // Біоніка інтелекту. – 2012. –№1(78). – С.78-84.

8. Гришко А.А. Адаптивное краткосрочное прогнозирование нестационарных временных рядов в интеллектуальных системах принятия решений / А.А. Гришко,

С.Г. Удовенко, Л.Э. Чалая // Сборник научных статей «Информатика, математическое моделирование, экономика» – Т.3. – Смоленск, Смоленский филиал РУК. – 2012. – С.43-49.

9. Hryshko A. A System for Electricity Trading using Genetic Algorithm and Reinforcement Learning / A. Hryshko, T. Downs // Proc. of Int. Conf. «AUPEC». – Brisbane (Australia).– 2004.– P.139-144.

10. Hryshko A. A Machine Learning Approach to Intraday Trading on Foreign Exchange Markets / A. Hryshko, T. Downs // Proc. of Int. Conf. «IDEAL-2005». – Berlin-Heidelberg. – 2005.– P. 588-595.

11. Hryshko A. Development of Machine Learning Software for High Frequency Trading in Financial Markets / A. Hryshko, T. Downs // In Book «Business Applications and Computational Intelligence» (Chapter XX): Idea Group Inc. Australia.– 2006. – P.406-430.

12. Гришко А.А. Адаптивные стратегии принятия решений в системах электронной торговли / А.А. Гришко // Материалы 8-й Международной научно-технической конференции «Проблемы информатики и моделирования». – Харьков: НТУ «ХПИ». – 2008. – С. 51.

13. Гришко А.А. Анализ сетевого трафика с использованием средств объектно-ориентированного программирования / А.А. Гришко, А.С. Пискун, С.Г. Удовенко // Материалы 10-й международной научно-практической конференции «Современные информационные и электронные технологии». – Одесса – 2009.– С. 136.

14. Гришко А.А. Система управления параллельным вычислительным процессом / А.А. Гришко // Материалы 9-й Международной научно-технической конференции «Проблемы информатики и моделирования». – Харьков: НТУ «ХПИ»– 2009 –С. 47.

15. Гришко А.А. Нейронечеткое краткосрочное прогнозирование в трейдинговых системах / А.А. Гришко, С.Г. Удовенко // Материалы 11-й международной научно-практической конференции «Современные информационные и электронные технологии» – Одесса. – 2010.– С. 70.

16. Гришко А.А. Многошаговый модифицированный алгоритм Q-обучения / А.А. Гришко, С.Г. Удовенко // Праці V-ї міжнародної школи-семінару «Теорія прийняття рішень». – Ужгород. – 2010. – С.73-74.

17. Гришко А.А. Комбинированные технические индикаторы в компьютерных системах трейдингового прогнозирования / А.А. Гришко, С.Г. Удовенко // Материалы научн.-техн.конф. «Сучасні напрямки розвитку інформаційно-комунікаційних технологій та засобів управління». – Київ. – 2010. – С. 72.

18. Гришко А.А. Методы аппроксимации таблицы Q-значений в алгоритмах обучения с подкреплением / А.А. Гришко // Материалы 1-й международной научно-технической конференции «Информационные технологии в навигации и управлении: состояние и перспективы развития». – Киев. – 2010. – С.58.

19. Гришко А.А. Модели принятия решений в интеллектуальных системах краткосрочного прогнозирования электропотребления / А.А. Гришко, С.Г. Удовенко // Материалы 2-ї Міжнародної науково-технічної конференції. «Сучасні напрямки

розвитку інформаційно-комунікаційних технологій та засобів управління». – Київ. – 2011. – С.48.

20. Гришко А.А. Анализ сходимости алгоритмов интерактивного машинного обучения в системах принятия оптимальных решений / А.А. Гришко // Материалы 2-й международной научно-технической конференции «Информационные технологии в навигации и управлении: состояние и перспективы развития». – Киев. – 2011. – С.39.

### АНОТАЦІЯ

**Гришко А.О. Гібридні методи машинного навчання в системах інтелектуальної обробки даних. – На правах рукопису.**

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. – Харківський національний університет радіоелектроніки, Міністерство освіти та науки України, Харків, 2013.

Дисертацію присвячено розробці гібридних методів машинного навчання в інтелектуальних системах різного функціонального призначення (на прикладі трейдингових систем та систем інтелектуального керування), що дозволяють підвищити ефективність стратегій прийняття рішень.

Запропоновано метод адаптації структури технічних індикаторів до поточного стану біржового ринку з подальшим формуванням стратегій трейдингової інтелектуальної системи, що базується на використанні комбінованого RL-навчання та генетичних алгоритмів. Запропоновано гібридний метод керування запасами з використанням стохастичного динамічного програмування та техніки навчання з підкріпленням, що є сумісною з несепарабельним критерієм. Модифіковано метод нейромережевої апроксимації Q-функцій RL-алгоритму, що дозволяє здійснювати корекцію конфігурації апроксимуючого багатозарового перцептрон.

Удосконалено структуру моделі прогнозування, алгоритм навчання якої базується на застосуванні нейромережевого фільтра-предиктора, що на відміну від існуючих забезпечує високу швидкодію та якість прогнозів в умовах нестаціонарності та невизначеності. Запропонована модель може бути використана для прогнозування тренду сигналів підкріплення при інтелектуальному керуванні динамічними об'єктами.

Розроблені методи було програмно реалізовано та використано для ряду практичних впроваджень.

**Ключові слова:** машинне навчання, сигнал підкріплення, гібридний метод, трейдингова система, нейромережева апроксимація, генетичний алгоритм, динамічне програмування, оптимальні стратегії.

### АННОТАЦИЯ

**Гришко А.А. Гибридные методы машинного обучения в системах интеллектуальной обработки данных. – На правах рукописи.**

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.23 – системы и средства искусственного интеллекта. –

Харьковский национальный университет радиоэлектроники, Министерство образования и науки Украины, Харьков, 2013.

Диссертация посвящена разработке гибридных методов, объединяющих возможности RL-методов, ИНС-моделей и методов оптимизации (в частности, генетической оптимизации и динамического программирования) и позволяющих повысить эффективность работы интеллектуальных систем трейдингового прогнозирования и управления стохастическими процессами в условиях неопределенности.

Объект исследования – процессы интеллектуальной обработки данных с применением методов машинного обучения.

Предмет исследования – гибридные методы машинного обучения с подкреплением в системах интеллектуальной обработки данных.

Показана целесообразность и разработки гибридных методов машинного обучения в интеллектуальных системах различного функционального назначения (на примере трейдинговых систем и систем интеллектуального управления), которые позволят повысить эффективность стратегий принятия решений. Определены основные задачи реализации процедур машинного обучения с подкреплением в трейдинговых системах, решение которых создает предпосылки для повышения эффективности электронной биржевой торговли. Предложен метод адаптации структуры технических индикаторов к текущему состоянию биржевого рынка при реализации трейдером комбинированного RL-обучения с применением генетических алгоритмов. Модифицирован метод нейросетевой аппроксимации Q-функций RL-алгоритма, позволяющий осуществлять коррекцию конфигурации аппроксимирующего многослойного персептрона.

Предложен метод решения задачи интеллектуального управления запасами с использованием стохастического динамического программирования и техники обучения с подкреплением, совместимой с несепарабельным критерием. В отличие от известных алгоритмов предлагаемый алгоритм обучения формирует реальное распределение прогнозируемых затрат, соответствующих рассматриваемым состояниям.

Предложен метод управления динамическим объектом, основанный на замене состояний, использующих прогнозируемые значения сигналов подкрепления предыдущими взвешенными состояниями, использующими текущие значения сигналов подкрепления. Метод является расширением SARSA-алгоритма и  $Q(\lambda)$ -алгоритма и позволяет учитывать оценки состояний, удаленных на большее расстояние друг от друга. Предложена нейросетевая прогнозирующая модель, обладающая фильтрующими свойствами и позволяющая обеспечить высокое качество прогноза тренда сигнала подкрепления в процессе работы системы управления динамическим объектом.

Разработанные методы были программно реализованы и использованы для ряда практических внедрений. В частности, результаты тестирования предложенной гибридной трейдинговой RL-системы на базе модуля «Genetic-Trade-M» подтверждают ее работоспособность и перспективы практического внедрения. Разработанная система по функциональным возможностям и качеству полученных результатов превышает возможности известных некоммерческих систем электронной биржевой торговли. Предложена структура модифицированной

модульной системы моделирования процессов управления техническими объектами с применением комбинированных методов машинного обучения. Модифицированная структура, в отличие от базового варианта, содержит оригинальные модули, расширяющие функциональные возможности системы моделирования.

Отдельные положения, выводы и рекомендации диссертационной работы использованы в учебном процессе для подготовки и модификации курсов «Методы и средства вычислительного интеллекта» и «Системы цифровой обработки информации» на кафедре электронных вычислительных машин Харьковского национального университета радиоэлектроники.

**Ключевые слова:** машинное обучение, сигнал подкрепления, гибридный метод, трейдинговая система, нейросетевая аппроксимация, генетический алгоритм, динамическое программирование, оптимальные стратегии.

### ABSTRACT

**Hryshko A.A. Hybrid machine learning methods in systems of intellectual processing of data. – Manuscript.**

Thesis for the candidate of technical degree of science on the specialty 05.13.23 – systems and means of artificial intelligence. – Kharkiv National University of Radioelectronics, Kharkiv, 2013.

The thesis is devoted to developing methods for the synthesis of hybrid machine learning methods in systems of intellectual processing of data (on the example of the trading systems and intellectual stochastic control systems).

The principles of trading on financial markets are introduced, followed by the development of hybrid trading strategies based on the machine learning methods of genetic algorithms and reinforcement learning. The new methods are implemented in software and allow a user to connect directly to a broker. The hybrid learning system that combine Q-learning and method of Bellman optimization dynamic is proposed. The principles of design of system are based on the adaptation recursive of estimation functions. In addition, the thesis considers the optimal strategies in control systems for dynamic objects using machine learning methods for reinforcement. The proposed approach allows to obtain high-quality approximation of the optimal strategies for evaluating functions by using multi-layer artificial neural networks. All the methods had been implemented as software and applied for several practical implementations.

**Key words:** machine learning, signal of reinforcement, hybrid method, trading system, neural network approximation, genetic algorithm, dynamic programming, optimal strategies.

Підп. до друк. \_\_\_\_\_ 2013      Формат 60x80      Папір. друк.  
Умов. друк. арк. 1,2    Облік вид. арк. 1,0      Зам. № \_\_\_\_\_  
Тираж 100 прим.

Надруковано у навчально-науковому видавничо-поліграфічному центрі ХНУРЕ  
61166, Харків, просп. Леніна, 14