

ПЕТРОВ ДМИТРО

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0001-9108-7903>e-mail: dmytro.petrov.mpfhv.2023@lpnu.ua**ШАХОВСЬКА НАТАЛІЯ**

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-6875-8534>e-mail: Nataliya.b.shakhovska@lpnu.ua**СКОПІВСЬКИЙ СТЕПАН**

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0008-4668-0140>e-mail: stepan.y.skopivskyi@lpnu.ua

ДОСЛІДЖЕННЯ ВПЛИВУ РІВНЯ ДРЕЙФУ РОЗПОДІЛУ ДАНИХ НА АГЕНТІВ НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Предметом досліджень є методи виявлення та оцінки дрейфу даних. Досягнення мети відбувається шляхом використання агента навчання з підкріпленням на даних з різним рівнем дрейфу даних та оцінці якості дій цієї нейронної мережі. Подальший опрацювання результатів відбувається із порівнянням найпопулярніших методів оцінки дрейфу та якістю агента на наборах даних з проведенням кореляційного аналізу та моделей прогнозування. Апробацію роботи впливу дрейфу даних на агентів навчання з підкріпленням здійснено з використанням різних алгоритмів виявлення дрейфу даних на основі реального агента покращення якості стільникового зв'язку та реального набору даних із обладнання стільникових сот.

Ключові слова: навчання з підкріпленням, дрейф даних, часові ряди, машинне навчання.

PETROV DMYTRO, SHAKHOVSKA NATALIYA, SKOPIVSKYY STEPAN

Lviv Polytechnic National University

INVESTIGATING THE EFFECT OF DATA DISTRIBUTION DRIFT LEVEL ON REINFORCEMENT LEARNING AGENTS

For the regression and classification tasks, as discussed in the first chapter, new approaches to circumvent the problem of data drift are actively being researched. However, for reinforcement learning, the question of how much data drift affects the quality of actions is still open. SAC (Soft Actor-Critic) - an RL agent that is trained to solve the problem of voice quality parameters optimization was chosen for the topic research. This is a state-of-the-art reinforcement learning algorithm designed to solve problems with a continuous action space. SAC is an extension of the original Actor-Critic architecture and includes several key improvements to enhance learning stability and sampling efficiency. The SAC agent was trained on GSM network statistical data received from base stations (BS). In total, data that were used for training was collected from 1,044 base stations for a total of 2.5 years. The statistics are averaged daily values of the network characteristics at the base station. The agent makes decisions based on the current values of the following parameters: HR Usage Rate, HR Usage Rate, TCH Blocking Rate, TCH Traffic, Number of Available TCH, and two thresholds for setting the use of FR (Full Rate). The agent target is to set those thresholds so that blocking does not occur and use BS resources as efficiently as possible. A system was built to analyze reinforcement learning agents. The activity diagram is given in the second section for a system collecting agent quality data for diverse data sets with different drift levels. The system has two implementation options, depending on whether the data known to the agent is used or not. For the case when the quality of the agent is evaluated on the training data, there are two implementation options: drift estimation and collection of SHAP analysis of the agent. In the variant of SHAP collection of values for the dataset, the agent is loaded and approximated to the regression model to find the contribution of each feature to the agent decision. In the course of performing this evaluation, data about the quality of the agent in the areas known to the agent is collected. After calculating all SHAP values on all episodes, these values are aggregated into a list of data features and their total absolute impact on the outcome of the agent's actions. This data will later be used to calculate the weighted data drift. After collecting data about the quality of the agent on the data it knows, this data is aggregated by episodes and stored for further analysis. A second use of the system for agent training data is to calculate agent quality with data drift accounting. The procedure will be identical to the case with new data from the moment the dataset is split into episodes, that it will be described in more detail in the variant for the data with drift in the second section.

Keywords: reinforcement learning, data drift, time series, machine learning.

Вступ

У сучасному світі, де зростає кількість галузей, що використовують машинне навчання, виникає проблема стійкості моделей до зміни у вхідних даних. Одним з факторів, що впливають на цю стійкість, є дрейф даних (Data drift), який виникає, коли структура вхідних даних змінюється з часом. Ця проблема є особливо актуальною у випадку агентів навчання з підкріпленням (Reinforcement Learning agents), які здатні взаємодіяти з навколишнім середовищем та змінювати свої стратегії на основі отриманої винагороди. Відповідно, невірно оброблений дрейф даних може призвести до зменшення ефективності таких агентів та навіть до повного руйнування їхньої стратегії.

Робота присвячена дослідженню впливу дрейфу даних на агентів навчання з підкріпленням та дослідженні стійкості таких агентів до змін вхідних даних. Для досягнення цієї мети будуть використані різні методи, такі як аналіз змін у розподілі вхідних даних, аналіз дій агента та розробка методів переоснащення моделей та застосування методів адаптації моделей до зміни вхідних даних.

Аналіз літературних джерел

У [1] автори пропонують рішення для виявлення дрейфу даних для двовимірного набору. Вони

використовують тест Колмогорова-Смірнова для двох вибірок, щоб виявляти дрейф даних. Сам метод є стандартним для пошуку дрейфу, однак вони пропонують рішення для двовимірного варіанту та в деталях розбирають спосіб його застосування. Однак, стаття розглядає приклад на моделі класифікації та не бере до уваги вплив дрейфу на результативність моделі.

Дослідження [2] присвячено підходу до боротьби з дрейфом даних за допомогою архітектурного рішення на основі CURL агенту, що динамічно оновлює основну модель автоматично підбираючи набори даних для перетренування підпорядкованої моделі, а також визначає пріоритет на перетренування певних класів, що позбавляє від необхідності проходити повний цикл перетренування з усім набором даних. Запропонований підхід автоматичного перерозгортання моделі МН при погіршенні метрик підпорядкованої моделі показав хороші результати при використанні із класифікаційною моделлю, а також суттєву перевагу запропонованої архітектури при дорозмічуванні даних. Ця архітектура запропонована лише для задачі класифікації, хоча сам підхід з оцінкою якості моделі та автоматичному перетренованню має бути дієвим і для інших задач, однак потребує глобального перепрацювання для них. Завдання підбору даних для перетренування моделей МН потребує свого рішення в кожному окремому випадку.

Стаття [3] вводить поняття Здоров'я МН, що є оцінкою якості передбачень моделі. Для цієї оцінки вони вводять поняття Similarity Score, що є власною метрикою оцінки чи точка належить розподілу. Згідно всіх результатів таблиці їх оцінка порівнювалася з KL-різницею та Відстанню Васерштейна. Варто відмітити, що дані метрики застосовували для оцінки також на наборах даних менше 1000, де показники цих метрик (як і більшості статистичних тестів) можуть показувати викривлену інформацію в силу особливостей розподілу даних. Як оцінку порівняння відповідності їх метрики до якості моделі, було вказано кореляцію, що варіювалася від 0.4 до 0.99, що було краще за інші методи від 1% до 51%. Не вистачає порівняння з іншими статистичними методами, та результатів для навчання з підкріпленням. Дана оцінка буде включена до аналізу впливу на рівні з іншими способами виявлення критичного дрейфу даних.

У роботі [4] показано вплив на якість використання моделей МН у оцінці точності та кількості ресурсів на використання моделі (при перетренуванні). Підхід передбачає тренування моделей на різних пакетах даних та оцінку доречності натренованих моделей на основі метрики, що обраховується за Бордом від коваріантного зсуву та оцінки дрейфу. Стаття недостатньо детально описує запропоновану метрику, приклад наведений лише для двох моделей МН, які також недостатньо описані і також немає аналізу даних на яких проводилося дослідження. Підхід пропонує підбирати моделі лише на основі оцінки того, наскільки той чи інший пакет даних підходить під поточні дані, однак не проводить оцінку впливу цієї невідповідності.

У [5] автори пропонують підхід до побудови моделі, яка "само-еволюціонує", для прийняття рішень на основі RL агенту. В дослідженні використовується Bayes-Adaptive Contextual Multi-Armed Bandit (BA-C-MAB) алгоритм. Результати запропонованого адаптивного алгоритму прийняття рішень на основі побудови оцінки відносно "ідеальної" політики показали для зміни початкових параметрів моделі прийняття рішень у середньому був на рівні із іншими розглянутими методами. Автори розглянули варіанти онлайн тренування, однак досі зберігалися значні падіння в точності прийнятих рішень. Також не було розглянуто варіанта моделі прийняття рішень на основі RL.

Дослідження [6] присвячено адаптивній системі в реальному часі на основі RL, що дає можливість покращити інтервали передбачення підбираючи параметри розподілу для квантильного передбачення. Також застосовується пріоритезоване повторення досвіду (PER) для онлайн частини архітектури. Серед мінусів відсутність порівняння впливу рівня дрейфу на ефективність моделей. Запропонована архітектура замкнутого кола також добре підходить для задач, що розв'язуються RL агентами.

Автори з [7] пропонують підхід базований на адаптивній пам'яті, що підлаштовується до частого дрейфу концепту у сфері онлайн моніторингу мережевого трафіку. Цей підхід оснований на комбінуванні ADAM (ADaptive Memory strategy) та RAL (Reinforced stream-based Active Learning). Підхід застосовувався до класифікацій за допомогою Random Forest та SVM, які є чутливими до змін у межах значень та їх розподілі. Підхід показав покращення стабільності. Однак сам підхід не застосовувався до RL агентів, не було також дослідження рівня дрейфу, хоча на результатах видно, що початкова точність починає просідати навіть із запропонованим алгоритмом для онлайн дотренування.

Методи та моделі

Проблема зміни розподілу даних є загальновідомою та критичною у галузі машинного навчання. Будь-які дані мають свій розподіл і він може різнитися для даних зібраних для тренування та реальним світом. Для оцінки різниці розподілів існує багато статистичних тестів, які застосовуються одразу на дані, або через їх перетворення з допомогою стиснення через такі методи як автокодувальники та SVM. Для дослідження теми обрано натренованого для розв'язування задачі оптимізації параметрів якості голосу агента навчання з підкріпленням - SAC (Soft Actor-Critic). Це найсучасніший алгоритм навчання з підкріпленням, розроблений для вирішення задач з неперервним простором дій. SAC є розширенням оригінальної архітектури Actor-Critic і містить кілька ключових покращень для підвищення стабільності навчання та ефективності вибірки.

SAC добре виконує широкий спектр завдань безперервного контролю та особливо ефективний у областях, де дослідження є складним. Однак важливо зазначити, що продуктивність SAC все ще може залежати від налаштувань гіперпараметрів, і для певних задач може знадобитися тонке налаштування.

Серед статистичних тестів оцінки різниці розподілу даних було використано:

1. Критерій узгодженості Колмогорова–Смирнова.
2. Відстань Вассерштейна (нормована).
3. Розбіжність Кульбака-Лейблера.
4. Індекс стабільності населення (PSI).
5. Відстань Дженсена-Шеннона.
6. Тест Крамера-фон-Мізеса.
7. Відстань Геллінгера (нормована).
8. Енергетична відстань.
9. Т-тест Стьюдента.

У статистиці тест Колмогорова-Смірнова (тест К-С або тест КС) є непараметричним тестом рівності неперервних чи дискретних одновимірних імовірнісних розподілів, який може бути використаний для порівняння значення з базовою вибіркою – тест К-С для однієї вибірки, або для порівняння двох вибірок – тест К-С для двох вибірок. Тест відповідає на питання "Як ймовірно, що ми побачимо такий набір вибірок, як цей, якщо вони були взяті з цього імовірнісного розподілу?" або, у другому випадку, "Як ймовірно, що ми бачимо два набори вибірок, як ці, якщо вони були взяті з одного (але невідомого) імовірнісного розподілу?".

Нас цікавить саме варіант тесту КС для двох вибірок, що має наступний вигляд:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|, \tag{1}$$

де $F_{1,n}(x)$ та $F_{2,m}(x)$ емпіричні функції розподілу першої та другої вибірок відповідно, n та m розміри першої та другої вибірок відповідно.

Спростування нульової гіпотези відбувається за умови:

$$D_{n,m} > \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1 + \frac{m}{n}}{2m}} \tag{2}$$

Відстань Вассерштейна також відома як відстань Землекопа, оскільки її можна розглядати як мінімальну кількість "роботи", необхідної для перетворення розподілу u в v , де "робота" вимірюється кількістю ваги розподілу, яку необхідно перемістити, помноженою на відстань, на яку її необхідно перемістити. Перша відстань Вассерштайна між розподілами u та v обраховується як:

$$l_1(u, v) = \inf_{\pi \in \Gamma(u, v)} \int_{\mathbb{R} \times \mathbb{R}} |x - y| d\pi(x, y), \tag{3}$$

де $\Gamma(u, v)$ це набір (імовірностей) розподілів у просторі $\mathbb{R} \times \mathbb{R}$ чий маргінали u та v для першого та другого факторів відповідно.

Якщо U та V відповіді кумулятивні функції розподілу для u та v , то ця відстань також дорівнює:

$$l_1(u, v) = \int_{-\infty}^{+\infty} |U - V| \tag{4}$$

Вхідні розподіли можуть бути емпіричними, тобто отриманими з вибірок, значення яких фактично є входами функції, або ж вони можуть розглядатися як узагальнені функції, і тоді вони є зваженими сумами дельта-функцій Дірака, розташованих на заданих значеннях.

Розбіжність Кульбака-Лейблера (також звана відносною ентропією та І-дивергенцією), позначається KL , це тип статистичної відстані: міра того, наскільки один розподіл ймовірностей P відрізняється від іншого, еталонного розподілу ймовірностей Q . Проста інтерпретація KL розбіжності P від Q – це очікувана надлишкова несподіванка від використання Q як моделі, коли фактичний розподіл є P . Хоча це відстань, однак вона не є симетричною для двох розподілів (на відміну від варіації інформації).

Загальна оцінка дрейфу даних зазвичай виконується наступним чином:

$$D_{score} = \frac{\sum_i^n d_i(x_i)}{|X|} \tag{5}$$

де $|X|$ – це розмір набору ознак, а x_i вибірка i -ої ознаки, а d_i це функція дрейфу вибірки, яка наведена нижче:

$$d_i(x) = \begin{cases} 1, & d(x) \geq t \\ 0, & d(x) < t \end{cases} \tag{6}$$

де $d(x)$ це функція відстані, а t порогове значення відстані для нульової гіпотези про ідентичність розподілів. Функція $d_i(x)$ повертає 1 при наявності дрейфу даних, нульова гіпотеза функції $d(x)$ спростована, чи 0 при ідентичності розподілів, відстань між взірцевим розподілом та розподілом x менша за

t.

З метою оцінки якості агента для покращення якості мобільного зв'язку було використано оцінку на основі використання HR і відсотка блокування, яка надає велике значення другому. Формула якісної оцінки навантаженості базової станції:

$$Quality\ Rate = \frac{(1 - \alpha) \cdot U + \alpha \cdot \ln(B + 1)}{(1 - \alpha) + \alpha \cdot \ln(2)} \quad (7)$$

у якій U та B нормалізовані значення HR Usage Rate та TCH Blocking Rate відповідно, α – константа, яка рівна 0.8. Це значення було знайдено імперичним методом і має на меті надати більшу вагу значенням блокування, ніж використанню HR. Оскільки набагато гірше, коли користувач не може додзвонитись, ніж розмова нижчої якості.

Результати експерименту

Для проведення аналізу було взято набір даних, на якому агент був натренований та ще 20% йому невідомих даних. Для кожної БС було розбито дані таким чином, щоб послідовно поділити часовий проміжок відомих та невідомих даних як 80% на 20%. Таким чином у нас немає протікання даних, і агент не бачив майбутні зміни за межами тренування.

Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
DATA	cat	Detected	Jensen-Shannon distance	0.819201

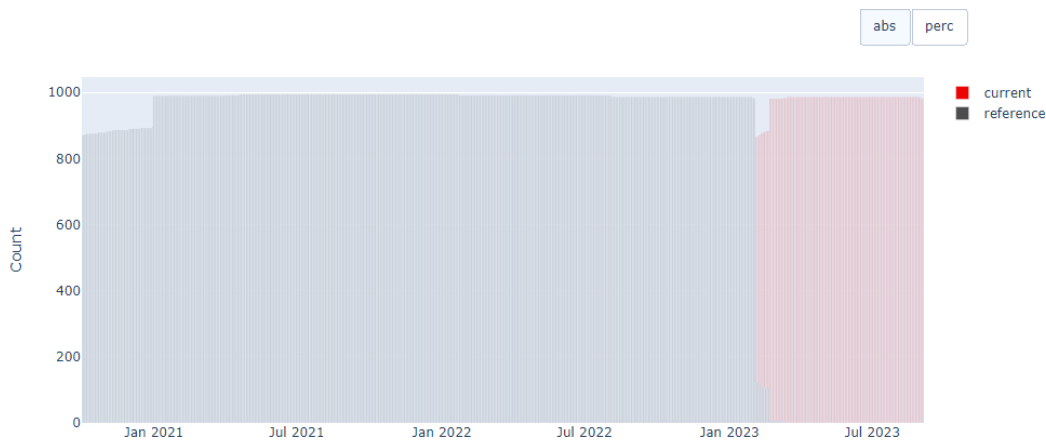


Рис. 1. Попередній аналіз дрейфу між відомими та невідомими даними для дати

Як зображено на рис. 1 дані розділені послідовно на невідому та відомі частини і вони мають лише декілька днів спільних у зв'язку з тим, що у деяких БС інша кількість даних була від початку, через що, їх поділ захоплює дані за інший часовий період. Однак це послідовний поділ у межах однієї БС, яка розглядається як одне середовище – агент може працювати лише в межах однієї базової станції за раз. При виконанні експериментів дані були поділені за БС і це вважалось однією вибіркою, оскільки агент здатен лише покращувати одну мережу за раз. Відповідно якість його дій та дрейф даних розраховувався для однієї БС за раз. Таким чином після конкатенації результатів виходять набори з 994 базових станцій, тобто 994 записи про якість та дрейф даних. На основі отриманих таким чином даних і проводився аналіз.

Для цього експерименту було використано підвибірку вибрану випадковим чином так, щоб взірцева вибірка була рівна невідомим даним для обраної БС. В результаті загальний розмір взірцевої вибірки у тестах мав зберігати оригінальний розподіл в межах розміру підвибірки, однак із розміром у середньому в 838 раз менший, що може погіршити якість деяких статистичних методів, але скоротить час виконання. Також було використано вікно вибірки – нові дані додавалися до взірцевого розподілу у розмірі підвибірки у 1000 рядків. Вікно вибірки мало би допомогти з фальшивим спрацюванням на дрейф даних через низьку кількість поточних(невідомих) даних для однієї станції. Оскільки бралось вікно з оригінального розподілу, то різниця значень мала би накладатися поверх оригінального розподілу і частотні значення мали би мати краще співвідношення.

Після чого дані були розбиті на набори по одній БС, на які агент мав покращувати якість зв'язку, що в результаті дало 994 записи про якість агента та дрейф даних між відомими та невідомими даними.

Для аналізу, які підходи до визначення дрейфу даних є найбільш ефективними, було проведено експерименти із 10-ма різними статистичними тестами для 4 різних підходів до збору взірцевої та поточної вибірок та із 5-ма способами збору результатів статистичних тестів про дрейф ознак.

В ході проведення експерименту із використанням підвибірки та застосуванням вікна для поточних значень із взірцевою вибіркою у 1000 значень було з'ясовано, що найкращим підходом на основі аналізу кореляції Пірсона є традиційний метод з часткою ознак, що перетнули порогові значення статистичного тесту Енергетичної Відстані із значенням кореляції у -0.64 до середнього значення якості.

Для наступного експерименту було прибрано використання підвибірки взірцевих даних розподілу при обрахунку дрейфу та було використано вікно вибірки у 1000 значень, що мало б додавати точок у поточний розподіл без істотної зміни його статистичних характеристик при обрахунку дрейфу з малим поточним розподілом. В результаті частка ознак із дрейфом (відсоток ознак, що мають інший розподіл) для цього варіанту збору дрейфу був найкращим методом для отримання інформації про зміну середньої якості агента та дрейфом з кращим показником кореляції у -0.65 за допомогою EB, що краще за попередній результат на $1,5\%$.

Експеримент із використанням підвибірки, без вікна вибірки підтвердив припущення, що значне зменшення вибірки взірцевого розподілу значно погіршує точність методів виявлення дрейфу. Це підтвердив експеримент з підвибіркою, але без вікна вибірки. Як результат, використання підвибірки саме із застосуванням стат. тестів за замовчуванням значно погіршує зв'язність тестів та оцінок якості агента, однак при використанні регресійних значень та інших способів агрегації, що втрачають менше даних про рівень дрейфу дає можливість отримати кращі результати із тестом Хеллінджера до $-0,58$ та навіть для тесту Вассерштейна $-0,59$, що чутливий, до змін взірцевої вибірки, особливо значного зменшення її, було отримано кореляцію на $10,5\%$ меншу ніж із тестом без підвибірки із використанням усереднення порогоування значень Вассерштейна, тобто кількості перевищень його порогу нульової гіпотези.

В експерименті без обмежень вибірок, було прибрано обмеження із зразкового розподілу повністю та не використовувалося вікно для поточного розподілу, що значно сповільнило процес виявлення дрейфу. Згідно результатів тест Вассерштейна має середню кореляцію у $-0,63$ з середнім значенням якості дій агента і пряму кореляцію у 0.64 із стандартним відхиленням, що свідчить про збільшення нестабільності якості при збільшенні дрейфу. Індекс Стабільності Населення має найбільші значення кореляції з поміж інших тестів із середнім квадратичним відхиленням якості зв'язку BC у 0.68 , середньої та максимальної кумулятивної винагороди у -0.65 та -0.62 відповідно, середньої моментальної винагороди та середнього квадратичного відхилення моментальної винагороди у -0.69 та 0.68 відповідно. Тест Хеллінджера із найбільшими кореляціями із мінімумом моментальної винагороди та СКВ у -0.60 та 0.67 відповідно.

Повнорозмірний взірцевий набір даних дозволяє методам виявлення дрейфу на основі тестів підтвердження входження чи подібності розподілів бути більш зв'язними із метриками якості агента. Приріст із середніми значеннями метрик якостей складає від $1,5\%$ до 15% відносно експериментів із застосуванням вибору підвибірок. Використання вікна вибірки може погіршити результати на 6% в обмін на швидший обрахунок метрик, однак при використанні із підвибіркою покращує результати зв'язності метрик дрейфу та якості на $10,5\%$ в порівнянні із використанням лише підвибірки, тому використанням підвибірки рекомендується із використанням вікна вибірки для мінімізування втрат.

Для виявлення, чи є практична цінність у поєднанні різних статистичних тестів було проведено ряд тестів із використанням лінійної регресії на регресійних значеннях статистичних тестів про дрейф даних. Із порівняння якості передбачення моделей на одному статистичному тесті та декількох впливало, що об'єднання тестів дає більше інформації про залежність у даних для формування передбачень. Набори із декількома значеннями мали точність до 15% кращу ніж з одним тестом. Після аналізу ваг моделей на багатьох тестів було визначено, що об'єднання різних тестів справді привносить додаткову інформацію про дані.

Для визначення зміни якості агента було зібрано дані про якість дій агента та його кумулятивну винагороду протягом застосування до тренувальних даних та невідомих даних із різним рівнем дрейфу.

Дослідження впливу дрейфу даних із пороговими значеннями дрейфу показали, що якість агента згідно функції якості при повному дрейфі погіршилася на 2% , що відповідно розподілу значень на відомих даних складає погіршення на $22,25\%$ від звичних значень, однак при частковому дрейфі помітних змін не відбулося. За таких же умов дослідження функції винагороди для агента у середньому мала на 76% відсотків нижчі значення винагороди при частковому дрейфі та на 78% нижчі значення винагороди при повному дрейфу простору ознак.

Висновок

У статті розглянуто особливості агента навчання з підкріпленням SAC та наведено обґрунтовані переваги у застосуванні до даних описаної області. Новизною дослідження є метод оцінювання впливу дрейфу даних на агентів навчання з підкріпленням. Практична цінність полягає у тому, що моделі навчання з підкріпленням матимуть чітку умову для перетренування при надмірній зміні даних та прогнозовану зміну поведінки. Досягнення мети відбувається шляхом використання агента навчання з підкріпленням на даних з різним рівнем дрейфу даних та оцінці якості дій цієї нейронної мережі. Подальший опрацювання результатів відбувається із порівнянням найпопулярніших методів оцінки дрейфу та якістю агента на наборах даних з проведенням кореляційного аналізу та моделей прогнозування. Апробацію роботи впливу дрейфу даних на агентів навчання з підкріпленням здійснено з використанням різних алгоритмів виявлення дрейфу даних на основі реального агента покращення якості стільникового зв'язку та реального набору даних із обладнання стільникових сот.

У результаті створено досліджено вплив дрейфу даних на агентів навчання з підкріпленням, оцінено

рівень зміни якості дій агента, виявлено оптимальний підхід до виявлення дрейфу, розроблено програмну реалізацію оцінки різного рівня дрейфу; проведено порівняльний аналіз найпопулярніших методів оцінки дрейфу та якості агента.

References

1. Porwik P., Dadzie B. M. Detection of data drift in a two-dimensional stream using the Kolmogorov-Smirnov test. *Procedia Computer Science*. 2022. 207. 168–175
2. Sack M. A data-centric reinforcement learning approach for self-updating machine learning models. *SPIE*, 2022.
3. Ghanta S., Subramanian S., Khernosh L., Sundararaman S., Shah H., Goldberg Y., Roselli D., Talagala N. *ML health monitor: Taking the pulse of machine learning algorithms in production*. 2019. ISBN 978-1-5106-2971-4.
4. Mallick A., Hsieh K., Arzani B., Joshi G. *Matchmaker: Data Drift Mitigation in Machine Learning for Large-Scale Systems*.
5. Costa R. D., Hirata C. M., Pugliese V. U. A Comparative Study of Situation Awareness-Based Decision-Making Model Reinforcement Learning Adaptive Automation in Evolving Conditions. *IEEE Access*. 2023. 11. 16166–16182.
6. Zhang Y., Wen H., Wu Q., Ai Q. Optimal Adaptive Prediction Intervals for Electricity Load Forecasting in Distribution Systems via Reinforcement Learning. *IEEE Transactions on Smart Grid*. 2022. 1–1.
7. Wassermann S., Cuvelier T., Mulinka P., Casas P. Adaptive and Reinforcement Learning Approaches for Online Network Monitoring and Analysis. *IEEE Transactions on Network and Service Management*. 2021. 18 (2). 1832–1849.