

Рогушина Юлія, Гришанова Ірина. Розділ XIV. Використання семантичних технологій та reinforcement learning у побудові персональних шляхів навчання. Штучний інтелект у науці : монографія / [авт. колектив]; за ред. Яцишина Андрія та Яцишин Анни. – Київ: ФОП Ямчинський О.В., 2025. – С. 182-197. ISBN 978-617-8830-09-0

РОЗДІЛ XIV. ВИКОРИСТАННЯ СЕМАНТИЧНИХ ТЕХНОЛОГІЙ ТА REINFORCEMENT LEARNING У ПОБУДОВІ ПЕРСОНАЛЬНИХ ШЛЯХІВ НАВЧАННЯ

DOI: 10.33407/lib.NAES.id/748281

Рогушина Юлія^{1,2[0000-0001-7958-2557]}, **Гришанова Ірина**^{1[0000-0003-4999-6294]}

¹Інститут програмних систем Національної академії наук України, Київ, Україна

²Інститут цифровізації освіти Національної академії педагогічних наук України, Київ, Україна

i26031966@gmail.com, ladamandraka2010@gmail.com

Анотація. Досліджуються проблеми персоналізації навчального процесу в умовах цифрової трансформації та застосування сучасних семантичних технологій для її підтримки. Розглядаються існуючі стандарти та засоби збереження, пошуку та повторного застосування навчальних об'єктів на основі їх семантичного аналізу, що спрямований на розробку персональних шляхів навчання відповідно до індивідуальних потреб студента, його знань та освітніх цілей. Запропоновано застосування методів машинного навчання з підкріпленням для побудови персональних навчальних шляхів на основі концепції композиції сервісів: навчальні об'єкти інтерпретуються як сервіси з параметрами якості QoS. Розроблено модифікований метод машинного навчання Rule-based QoS-based Q-Learning, який забезпечує адаптивне вдосконалення стратегії вибору у динамічних середовищах, запобігання зациклювань та неефективних дій. Досліджено програмну реалізацію метода, проаналізовано перспективи подальших досліджень на основі глибокого Q-навчання та розширення механізмів роботи з динамічними середовищами.

Ключові слова: навчальний об'єкт, персональний навчальний шлях, Q-learning, навчання з підкріпленням.

USE OF SEMANTIC TECHNOLOGIES AND REINFORCEMENT LEARNING IN CONSTRUCTION OF PERSONAL LEARNING PATHWAYS

Abstract. This study explores the challenges of personalizing the learning process in the context of digital transformation and the use of modern semantic technologies that support this process. We consider existing semantic-based standards and tools for storing, retrieval and reuse learning objects. Their semantic analysis focuses on construction personal learning pathways (PLPs) according to individual needs, knowledge and learning goals of students. We propose the application of reinforcement learning methods for PLP constructing based on the concept of service composition: learning objects are interpreted as services with quality of service (QoS) parameters. A modified reinforcement learning method, Rule-based QoS-based Q-Learning, is developed to enable adaptive improvement of strategy selection in dynamic environments, prevent looping and eliminate inefficient actions. The study also represents a software implementation of the proposed method and analyzes prospects for further research, including deep Q-learning and enhanced mechanisms for handling dynamic environments.

Keywords: learning object, personal learning pathway, Q-learning, reinforcement learning.

Вступ. Цифрове освітнє середовище забезпечує доступ до великої кількості навчальних ресурсів, які можуть бути використані для персоналізації навчання відповідно до індивідуальних особливостей сприйняття інформації, уподобань та різноманітних додаткових знань та навичок. Але саме кількість, обсяг, динамічність та гетерогенність таких ресурсів потребують застосування сучасних семантичних технологій для автоматизації пошуку, аналізу та інтеграції цих *інформаційних об'єктів* (ІО), що супроводжується стандартизованими метаданими.

Процес цифрової трансформації освіти вимагає чіткої формалізації елементів екосистеми навчання та її основних задач для вибору відповідних інформаційних технологій обробки знань та створення відповідних програмних рішень.

Аналіз літератури та постановка проблеми. Персоналізація навчання (Groff, 2017) є однією з важливих тенденцій розвитку освіти впродовж десятиріч років, яка викликає велику зацікавленість наукової спільноти (Shemshack & Spector, 2020). Але цифрова трансформація сфери освіти значно змінює виклики та тенденції досліджень у цій сфері, потребуючи більш широкого запровадження семантичних технологій, Data Mining, обробки Big Data та елементів штучного інтелекту.

Можна виокремити кілька рівнів такої персоналізації:

- вибір майбутньої професії або сфери діяльності $s \in S$, що визначає набір потрібних компетенцій $\{q_i\} = s$ та вибір установи $ins \in Ins$, що надає релевантні освітні послуги $\forall q_i \exists LC(ins)$;

- визначення набору *навчальних курсів* (Learning Course, LC), які можуть бути використані здобувачем для отримання компетенцій $\{q_i\} = s$ для $s \in S$ (залежно від знань та навичок здобувача освіти) $\{LC_j\}$ таких, що $\forall LC_j, j = \overline{1, n}, \exists q_i \in s$ – цей етап традиційно пов'язують із побудовою *індивідуальної освітньої траєкторії* (IOT) здобувача;

- визначення припустимих послідовностей вивчення LC та вибір оптимальної з них для конкретного здобувача (відповідно до його індивідуальних цілей, здібностей та можливостей);

- вибір викладача $a = Staff(fns)$, співпраця з яким має забезпечити найбільш ефективний навчальний процес;

- визначення набору *навчальних об'єктів* (Learning Object, LO), що містять інформацію, що необхідна конкретному здобувачу для вивчення певного LC;

- визначення індивідуальної послідовності вивчення LO (PLP, (*Personal Learning Pathway*)) (Chen, 2008) з урахуванням наявних знань та навичок здобувача для успішного вивчення певного LO.

Усі ці етапи потребують застосування знань як щодо предметної області (ПрО) навчання та пертинентних LO, так і щодо самого здобувача. На вищих рівнях персоналізації здобувач має більше можливостей для особистого прийняття рішень відповідно до власних уподобань та уявлень про майбутню

професійну діяльність. Але нижчі рівні такої персоналізації потребують структурованого уявлення щодо Про та специфіку її вивчення, припустимі послідовності вивчення окремих дисциплін, розуміння взаємозв'язків між окремими компетенціями та навичками тощо, які повинен мати викладач. Тому на цих етапах процес персоналізації навчання має здійснюватися у співпраці між здобувачем та викладачем.

Слід відмітити, що з точки зору алгоритмів, які використовуються для визначення припустимих послідовностей вивчення LC та LO, немає принципової відмінності між вибором LO та LC. Але описи LC та LO не тільки різняться рівнем деталізації (для LC це LO та їх параметри, а для LO – компетенції відповідного курсу), а й кількістю елементів, що використовуються для співставлення: у побудові послідовності вибору LC у ІОТ, як правило, враховують не більше кількох десятків доступних LC, а зазвичай – значно менше, а їх вхідні та вихідні параметри теж обмежуються досить узагальненими компетенціями цих LC, тоді як вибір припустимих послідовностей вивчення LO для певного курсу може потребувати аналізу сотень об'єктів, для кожного з яких вхідні та вихідні параметри можуть обиратися з великої кількості компетенцій нижчого рівня деталізації – наприклад, знання окремих визначень, вміння доводити твердження та теореми.

LO – це базові елементи процесу навчання. Кожен LO – це автономна сукупність відомостей, що об'єднані на основі навчальної цілі. До LO відносяться лекційні матеріали (як текстові, так і мультимедійні), практичні завдання, засоби контролю та оцінювання отриманих знань та навичок. Основна відмінністю LO від інших типів ІО – наявність формалізованих метаописів, які визначають роль LO у навчальному процесі та характеризують їх контент. Певна автономність дозволяє повторно використовувати LO у складі різних навчальних програм (Korpi, Bogle & Bogle, 2005).

Для того, щоб надати можливість пошуку та повторного використання LO, вони супроводжуються метаданими, які відображають різні аспекти їх створення, тематики, доступу, форм подання та особливостей застосування в

освітньому процесі. Схеми метаданих, що використовуються в різних репозиторіях, можуть базуватися на різних загальноновживаних стандартах метаданих залежно від специфіки та цілей створення репозиторію. При цьому можуть використовуватися як універсальні стандарти опису метаданих (наприклад, Dublin Core), так і стандарти, які створені спеціально саме для опису навчальних матеріалів (наприклад, IEEE LOM та SCORM).

Такі метадані дозволяють формалізувати відомості про атрибути LO (формат LO, розробник LO, власник, ключові слова, релевантні LC, тривалість вивчення, дата створення тощо) та про LC, які використовують ці LO (розробник, результуючі та вхідні компетенції).

Наступною важливою проблемою у персоналізації навчання є складність LC та тих LO, що можуть використовуватися для його вивчення. В цьому дослідженні ми не аналізуємо складність LO для сприйняття або відносну складність певних галузей знань. Ми розглядаємо лише два структурні аспекти складності таких IO, а саме: 1. кількість об'єктів (компетенцій, LO, інших LC), що є необхідною умовою вивчення LO або LC; 2. кількість компетенцій (навичок, знань), які здобувач може отримати в результаті вивчення відповідного LO або LC. Саме ці параметри визначають обчислювальну складність та час побудови можливих маршрутів навчання. Складність LC певним чином корелюється з набором компетенцій здобувачів освіти та з їх віком – дорослі здобувачі частіше вдосконалюють свій професійний рівень, а не вивчають певну спеціальність “з нуля”. Крім того, дорослі здобувачі зазвичай мають значно більший професійний досвід та володіють більшим набором компетенцій, знань та навичок, що семантично пов'язані з тими LC, що вони планують вивчити (або хоча б мають значно глибше уявлення про потенційні результати навчання, ніж здобувачі освіти молодшого віку).

Задача побудови ІОТ, яку вирішує андрагог (викладач, що навчає дорослих), має більшу обчислювальну складність, тому що потребує співставлення складніших IO – метаописів LC та профілів здобувачів. Потрібно генерувати та порівнювати значно більший набір потенційно припустимих маршрутів

навчання, які можуть бути представлені через частково впорядковані множини LO. Крім того, потрібно брати до уваги, що LO та LC, що містяться у репозиторіях, а також умови доступу до них, теж можуть змінюватися.

В даному дослідженні ми використовуємо компетенції як основний механізм співставлення IO різних типів. Термін “Компетенція” використовується у розумінні (Rogushina & Priyma, 2017). В якості зовнішніх джерел знань щодо компетенцій та їх зв’язків можуть використовуватися ESCO (European Skills, Competences, Qualifications and Occupations, <https://ec.europa.eu/esco/portal/home>) – багатомовний класифікатор навичок, компетентностей, кваліфікації та професій, та інші стандарти, наприклад, RCD (Reasonable Competency Definition) (de-Marcos et al., 2008) і SRCM (Simple Reasable Competency Mapping) (Lundqvist, Baker & Williams, 2008).

Результати дослідження. Щоб знайти оптимальну послідовність вивчення LO конкретним студентом, викладачу потрібно виконати наступні кроки семантичного аналізу :

- побудувати множину компетенцій для LC, що поділяються на дві групи – вхідні (ті, що необхідні для вивчення LC) та вихідні (ті, які є результатом вивчення цього LC): саме елементи цієї множини використовуються для опису LO, що пертинентні курсу: для цього можуть бути використані зовнішні класифікатори компетенцій, такі як ESCO, але викладачу потрібно відібрати тільки ті компетенції, що потрібні для конкретного LC;

- на основі аналізу метаданих LO, їх ключових слів, структурованих анотацій тощо побудувати вхідні та вихідні компетенції для кожного релевантного LO з цього набору (можуть бути використані RDF-описи, елементи стандартів, семантична розмітка тощо);

- визначити набір компетенцій студента, для якого генерується план навчання;

- побудувати припустимі послідовності вивчення LO, де для вивчення кожного наступного LO студент повинен мати всі його вхідні компетенції,

отримавши їх в результаті вивчення попередніх LO або з власного досвіду до початку навчання.

З точки зору компетентнісного підходу будь-яке навчання може розглядатися як процес розширення набору компетенцій певної особи або колективу. Важливо звернути увагу, що при цьому розглядається не весь набір компетенцій, а тільки та його підмножина, що пертинентна певному LC. Відповідно до специфіки курсу та гетерогенності потенційних студентів можуть застосовуватися різні набори компетенцій, які базуються як на національних та міжнародних стандартах опису компетенцій, професій та кваліфікацій (більш характерно для формального навчання), так і на різноманітних внутрішніх правилах спільноти (більш характерно для неформального та інформального навчання). Крім того, викладач може самостійно поповнювати або зменшувати цей набір компетенцій.

Приміром, якщо виявляється, що не всі студенти мають достатній рівень розуміння всіх тих природних мов, якими викладено контент LO, то доцільно додати необхідний рівень розуміння певної мови для кожного LO. Це дозволить відбирати для кожного студента LO тими мовами, які він знає. Протилежний варіант – підвищення кваліфікації групи осіб певної спеціальності (тобто вони мають певний набір компетенцій), і тоді видалення цих компетенцій з вхідних умов LO дозволить зменшити час пошуку персональних шляхів навчання.

$St(l_0 \{C_{in}\}) \models \{C_{out}\}$, де l_0 – LO, $\{C_{in}\}$ – множина компетенцій, які обов'язково потребує здобувач до початку вивчення l_0 ; $\{C_{out}\}$ – множина компетенцій здобувача після початку вивчення l_0 , $\{C_{in}\} \subset \{C_{out}\}$. У персоналізованому навчанні додається додатковий параметр $St_{in}(l \in l \subseteq \{C_{in}\} \cup \{C_{out}\})$, що характеризує наявні компетенції здобувача $l \in LE$. Якщо $\exists c \in \{C_{in}\} \not\subseteq St_{in}(l \in l \subseteq \{C_{in}\} \cup \{C_{out}\})$, то здобувач має недостатній набір компетенцій для того, щоб вивчати певний курс. Якщо $\exists c \in \{C_{out}\} \not\subseteq St_{in}(l \in l \subseteq \{C_{in}\} \cup \{C_{out}\})$ (здобувач вже має певні результуючі компетенції курсу), то його PLP може бути скорочений.

Викладач може здобувати ці відомості з метаданих LO у репозиторіях (наприклад, у форматі RDF-трійок “LO-вхідні дані-компетенція” та “LO-вихідні дані-компетенція”) або з інших джерел, але має інтегрувати їх з набором компетенцій курсу, для якого потрібно створювати PLP (Rogushyna, & Gladun, 2024b).

Ми пропонуємо для автоматизації цього процесу використовувати індивідуальні вікірепозиторії, що будуються на основі семантичного розширення вікітехнології. В такому випадку для побудови наборів LO з відповідними властивостями потрібно спочатку внести до репозиторію потрібні об’єкти, створивши відповідні вікісторінки, та побудувати їх метаописи (в якості тегів семантичної розмітки використовуються компетенції LO), а потім зі сторінки LC виконувати вбудовані запити (Rogushina & Gladun, 2024b).

Такий запит може мати наступний вигляд:

```
{{#ask:
[[Категорія:LO]]
[[LC::{{{PAGENAME}}}
[[Language::Ukrainian]]
[[Рік розробки::>2020]] [[Доступ::>Free]]
|?Input competence
|?Output competence
|?Рік розробки
|?Definition
|format=broadtable |link=all |headers=show |searchlabel=...|class=sortable wikitable smwtable }}
```

В результаті виконання запиту генерується таблиця з полями, що визначені у запиті, яка містить актуальний опис LO у репозиторії. Важливо, що для створення переліку LO для нових курсів достатньо просто додавати цей код до сторінки LC без змін (рис.1). Важливо, що середовище Semantic MediaWiki дозволяє конвертувати результати семантичних запитів у RDF, і це забезпечує додаткові можливості їх використання в інших застосунках. Аналогічно виконується пошук компетенцій певного LC.

Прикладом такого репозиторію є ActiveBook, що дозволяє зберігати відомості щодо матеріалів з відкритих структурованих джерел (бібліотек

університетів та дослідницьких інститутів, репозиторіїв LO, корпоративних порталів). Семантична розмітка ресурсів базується на онтології освітніх ресурсів, що інтегрує стандарти схем метаданих для освітніх ресурсів та забезпечує інтероперабельність контенту. Система містить розширюваний набір шаблонів, які описують структуру різних типів LO, підтримує пошук та співставлення на семантичному рівні таких складних IO, як «Підручник», «Спеціальність», «Компетенція» тощо. Система реалізована на основі Semantic MediaWiki.

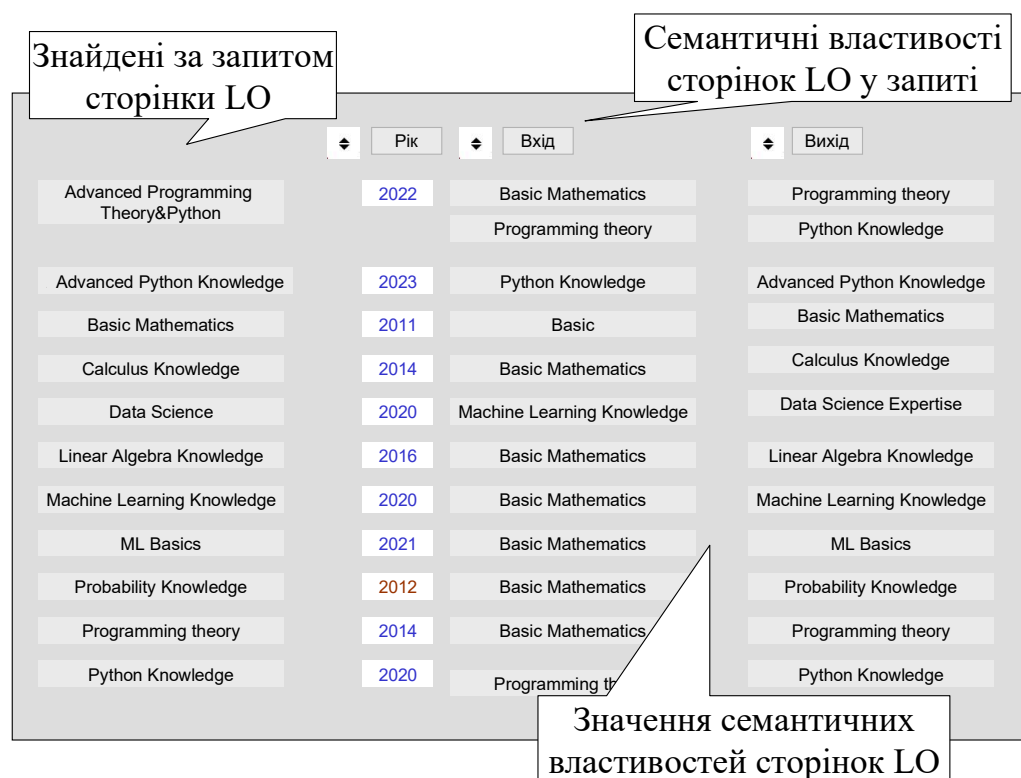


Рис.1. Пошук властивостей LO у репозиторії ActiveBook

Важливою перевагою такого технологічного рішення є те, що той самий LO може бути семантично розмічений різними викладачами відповідно до особливостей різних LC, і ці елементи розмітки не будуть конфліктувати між собою, а, навпаки, можуть використовуватися спільно. Крім того, використання відкритого програмного забезпечення та відкритих форматів забезпечує повторне використання знань щодо структури та призначення LO та відповідає принципам FAIR (Jacobsen et al., 2020).

Використання машинного навчання для побудови PLP

Після того, як визначено умови вивчення LC, результуючі компетенції курсу, наявні компетенції здобувача та характеристики наявних LO, що можуть бути використані в процесі навчання, потрібно побудувати PLP – послідовність того, як студент буде здобувати знання та навички з цих LO. Пропонується використовувати для цього інтерпретацію LO як сервісів (Uc-Cetina et al., 2015), а побудову PLP – як генерацію композитного сервісу за визначеними вимогами (Гришанова & Рогушина, 2024b). Це дозволяє абстрагуватися від окремих схем та стандартів опису LO та LC в різних репозиторіях, а також більш формально аналізувати критерії та обмеження на побудову навчального маршруту, включаючи послідовність навчальних об'єктів, залежності між ними, вимоги до попередніх знань, параметри якості навчання, такі як час освоєння, складність, вартість та адаптація до потреб студента.

Кожен сервіс визначається: входами – множиною вхідних параметрів (які дані необхідні для виклику сервісу), виходами – множиною вихідних результатів (які результати він повертає після виконання); функціоналом з перетворення середовища та множиною параметрів *якості обслуговування* (QoS). *Композитний сервіс* створюється шляхом поєднання інших сервісів в певну послідовність для досягнення певної мети. *Композиція сервісів* – це пошук послідовності викликів сервісів (*маршруту*), де вхід кожного наступного сервісу відповідає виходам попередніх сервісів або початковому стану, яка забезпечує досягнення цільового стану. LO можна розглядати як сервіс, де входи та виходи – це компетенції відповідного LC, а QoS – параметри якості навчання LO. Тоді PLP – це композитний сервіс, що враховує рівень підготовки студента і визначає послідовність його дій з вивчення LO для здобуття компетенцій LC.

Машинне навчання (*Machine learning*, ML) з підкріпленням (*Reinforcement Learning*, RL) (Zhao et al, 2016) — це підхід до навчання агентів, які взаємодіють із середовищем для знаходження оптимальної стратегії дій через отримання досвіду. *Агент* (автономна сутність, яка приймає рішення) адаптує свою поведінку, базуючись на винагородах за виконання певних дій у відповідь на зміну стану

середовища. *Середовище* – простір, в якому функціонує агент, що складається зі станів і дій, які агент може виконувати для досягнення заданого цільового стану, та визначає правила, стан системи, можливі дії та винагороди за ці дії. *Взаємодія* – процес, за допомогою якого агент вивчає середовище: обираючи дію, він отримує зворотний зв'язок (нагороду) та переходить до нового стану.

Стани – це комбінації доступних вхідних і вихідних параметрів, які змінюються в залежності від виконаних сервісів. Множина всіх можливих станів утворює простір станів, у якому виділяють *початковий* та *цільовий* стани. *Поточний стан агента* характеризується наявними вхідними та вихідними даними, доступними агенту після виконання певного сервісу. У разі досягнення цільового стану епізод завершується, і агент отримує винагороду. *Дії* – це сервіси, які може викликати агент, змінюючи поточний стан середовища. Агент шукає множину маршрутів, які забезпечують перехід з початкового стану в цільовий, та обирає з них *оптимальний*, використовуючи значення параметрів QoS сервісів у складі маршруту.

Особливості та переваги Q-learning

Q-learning (Jang et al., 2019) – один із класичних алгоритмів RL для пошуку оптимальної політики дій агента в середовищі без необхідності попереднього знання середовища. Метод працює на основі Q-таблиці, яка оцінює корисність $Q(s,a)$ кожної можливої дії a у кожному стані s , враховуючи винагороди і оцінку найкращих можливих дій у наступному стані, щоб максимізувати довгострокову винагороду за Белманом (Barto & Sutton, 2015). Для балансу між дослідженням та виконанням знайдених дій використовується ϵ -жадібна стратегія.

Q-learning розраховує миттєву винагороду r , яка визначається на основі виконання конкретної дії a у стані s (1 за досягнення цілі, -1 за неправильну дію). Ми пропонуємо для розрахунку винагороди використовувати параметри QoS, враховуючи необхідність їх максимізації або мінімізації, де винагорода залежить від параметрів QoS (якості сервісу), цільової функції оптимізації окремого параметра QoS (мінімізації або максимізації), та глобальних коефіцієнтів важливості, що дозволяє враховувати ефективність обраного шляху більш

коректно і наближено до реальних умов. При цьому важливість кожного параметра визначається глобальними ваговими коефіцієнтами QoS. У задачі побудови PLP ми використовуємо наступні QoS: вартість, тривалості навчання, рейтинг LO (набір параметрів може варіюватися залежно від вимог задачі). Агент знаходить маршрут з оптимальними QoS цільового стану (максимальним рейтингом і мінімальними вартістю і тривалістю), використовуючи формулу розрахунку винагороди:

$$R(s, a) = \sum_{i=1}^n w_i * r_i(s, a) - 1, \quad (1),$$

де стан $s \in S$, дія $a \in A$, w_i – вага i -го параметру QoS значення винагороди за стан s та дію a , $r_i(s, a)$ – нормалізована оцінка винагороди, яка визначається на основі значень параметрів QoS.

Ми пропонуємо наступну модифікацію Q-learning:

Динамічний вибір і адаптація дій: агент визначає послідовність дій, враховуючи специфічні параметри QoS залежно від поточних станів агента та враховуючи комбінації входів-виходів кожного сервісу, що дозволяє змінювати вибір дій в реальному часі.

Оцінка якості сервісів через QoS: параметри QoS гнучко застосовуються (з урахуванням коефіцієнтів їх важливості) при розрахунку миттєвої винагороди – кількість параметрів не є фіксованою, їх значення можуть бути нормалізовані, а кожен параметр має свою цільову функцію оптимізації.

Механізми запобігання неефективних дій: запобігання зациклювання через обмеження кількості кроків без прогресу для уникнення петлевих станів; обробка тупикових і неперспективних станів – якщо жодна дія не веде до подальшого розвитку, стан вважається тупиковим, а агент отримує штраф; автоматизація параметрів навчання – враховуючи розмір вхідних даних та складність середовища, автоматично визначаються такі параметри, як кількість епізодів (залежно від розміру таблиці станів і складності завдання), обмеження на кількість кроків (залежно від середньої довжини маршруту), кількість кроків без прогресу, гіперпараметри Q-learning та швидкість навчання.

Модифікація винагороди: застосовується функція винагороди (1), яка використовує значення QoS.

Запропоновані модифікації методу Q-learning (Гришанова & Рогушина, 2024a) не лише дозволяють агенту знаходити ефективні маршрути, але й забезпечують адаптивне вдосконалення стратегії вибору на основі зворотного зв'язку у динамічних середовищах, де характеристики сервісів можуть змінюватися. Цей метод, який доповнює Q-learning правилами запобігання неефективних дій та урахуванням якісних характеристик сервісів QoS, ми називаємо R-QoS-Q-Learning (Rule-based QoS-based Q-Learning).

Експериментальне дослідження R-QoS-Q-Learning

Щоб дослідити властивості та швидкість роботи алгоритму R-QoS-Q-Learning, створена його програмна реалізація мовою Python та згенерований набір з 1029 сервісів LO. Експеримент показав, що час виконання RQoS-Q-Learning залежить від якості вхідних даних – кількості визначених станів і існування можливих маршрутів, та кількості вибраних для навчання епізодів (час формування таблиць: від 1.4 с для 560 станів до 197.4923 с для 25913 станів; час виконання: від 2.5 до 114 с для 40000 епізодів) (Рис.2.А), та продемонстрував високий рівень якості побудованого PLP. Тестування виконувалось на Macbook Pro (Apple M1 Pro з 16 Гб пам'яті). Графік помилки різниці в часі (Рис.2.Б) Error (Temporal Difference Error, TD) при роботі на тестовому наборі даних з 1029 сервісів LO показує, що алгоритм поступово вчиться адаптуватися до запропонованого складного середовища.

Крива навчання при роботі R-QoS-Q-Learning (Рис.2.В) ілюструє процес навчання агента за допомогою Q-learning. На початку навчання винагорода є нерівномірною і низькою, але без значних коливань, що свідчить про фазу дослідження, коли агент вивчає середовище. Значення винагороди спочатку збільшується, а потім стабілізується, що підтверджує ефективність навчального процесу агента.

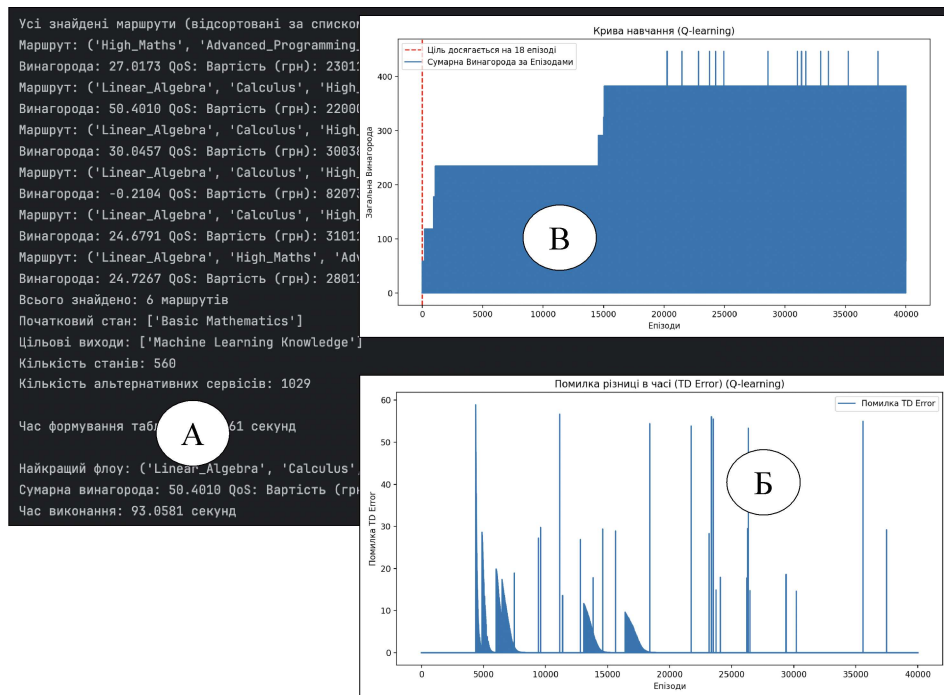


Рис.2. Результати роботи R-QoS-Q-Learning на тестовому наборі даних з 1029 сервісів LO

Висновки. Проведені експерименти виявили, що запропонована модифікація вдосконалює Q-learning, але залишає деякі проблеми: складність налаштування гіперпараметрів для стабілізації процесу навчання, масштабування для роботи з великими даними та у динамічних середовищах, потребу у пришвидшенні обчислень. Тому пропонується дослідити метод Deep Q-Network (DQN), який поєднає Q-learning із можливостями глибокого навчання для покращення продуктивності та масштабованості, та розробити інструменти автоматизованого багатокритеріального налаштування гіперпараметрів на основі методів Grid Search, Random Search, Bayesian Optimization.

Методи RL, зокрема Q-learning та DQN, є перспективними підходами для автоматизованої побудови PLP, забезпечуючи їх адаптивність, гнучкість та узагальнення знань та дають змогу гнучко адаптуватися до середовища і враховувати різні параметри QoS. Але їх успішне застосування потребує розв'язання проблем довгих епізодів і ефективності навчання, оптимізації балансу між дослідженням та експлуатацією та вдосконалення роботи з багатокритеріальною оптимізацією. Для цього планується дослідити можливі методи оптимізації і стабілізації процесу

навчання шляхом додавання правил, засоби автоматизації підбору гіперпараметрів, передбачити способи масштабування, наблизитись до роботи у динамічному середовищі без втрати знань та забезпечити модифікацію політики в реальному часі при появі нових LO.

Список джерел

Barto A., & Sutton R. (2015). *Reinforcement Learning: An Introduction*, Second edition, in progress. The MIT Press: Cambridge, Massachusetts. <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>

Chen, C. M. (2008). *Intelligent web-based learning system with personalized learning path guidance*. In: Computers & Education, 51(2), 787-814.

de-Marcos, L., Barchino, R., Martínez, J. J., Gutiérrez, J. A., & Hilera, J. R. (2008). *Competency-based intelligent curriculum sequencing: comparing two evolutionary approaches*. In: 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Vol. 3, pp. 339-342).

Groff, S. (2017) *Personalized Learning: The State of the Field & Future Directions*. https://curriculumredesign.org/wp-content/uploads/PersonalizedLearning_CCR_May2017.pdf.

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., & Schultes, E. (2020). *FAIR principles: interpretations and implementation considerations*. In: Data intelligence, 2(1-2), 10-29.

Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019) *Q-learning algorithms: A comprehensive classification and applications*. In: IEEE access, 6 7, 133653-133667.

Koppi, T., Bogle, L., & Bogle, M. (2005). *Learning objects, repositories, sharing and reusability*. In: Open Learning: The Journal of Open, Distance and e-Learning, 20(1), 83-91.

Li, J. , Zheng, X. L. , Chen, S. T. , Song, W. W. , & Chen, D. R. (2014): *An efficient and reliable approach for quality-of-service-aware service composition*. In: Information Sciences, 269, 238-254.

Lundqvist, K. O. Baker, K. D., & Williams, S. A. (2008) *An ontological approach to competency management*. <http://www.eife-l.org/publications/proceedings/ilf07/Contribution110.doc.pdf>.

Rogushina, J., & Gladun, A. (2024) *Repository of Complex Information Objects as a Source for Development of Semantic Applications*. In: XXIII International Scientific and Practical Conference Information Technologies and Security ITS 2023, CEUR Workshop Proceedings (2024), CEUR Vol-3887, 2024, P.119-133. <https://ceur-ws.org/Vol-3887/paper11.pdf>.

Rogushyna, J., & Gladun, A. (2024). *Transformation of university libraries into sources of open knowledge based on the semantic expansion of wiki technology*. In: Proc. of scientific works of the International Scientific Conference "Artificial Intelligence in Science and Education" AISE-24.

Rogushina, J., & Priyma, S. (2017). Use of competence ontological model for matching of qualifications // Chemistry: Bulgarian Journal of Science Education, V.26, N.2, P.216-228.

Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. Smart Learning Environments, 7(1), 33. link.springer.com/content/pdf/10.1186/s40561-020-00140-9.pdf.

Uc-Cetina, V., Moo-Mena, F., & Hernandez-UcanR. (2015) *Composition of web services using Markov decision processes and dynamic programming*. In: The Scientific World Journal, , (1), 545308.

Zhao, D., Wang, H., Shao, K., & Zhu, Y. (2016). Deep reinforcement learning with experience replay based on SARSA. In: 2016 IEEE symposium series on computational intelligence (SSCI), pp. 1-6.

Гришанова І.Ю., & Рогушина Ю.В., (2024) *Планування послідовностей сервісів навчальних об'єктів на основі машинного навчання з підкріпленням*. ІТБ-2024. С.74-77.

Гришанова І.Ю., & Рогушина Ю.В. (2024) *Технологія використання машинного навчання для побудови композиційного вебсервіса*. Проблеми програмування, № 4, 2024. С.3-13.