

Савченко М.В.Національний технічний університет
«Харківський політехнічний інститут»**Меркуленко Ю.С.**Національний технічний університет
«Харківський політехнічний інститут»

Q-LEARNING І МАРКОВСЬКА ОПТИМІЗАЦІЯ: АДАПТИВНІ ПІДХОДИ ДО УПРАВЛІННЯ СПЕКТРОМ У КОГНІТИВНИХ РАДІОМЕРЕЖАХ

У статті розглянуто поєднання Марковських моделей та алгоритмів навчання для вирішення задач динамічного доступу до спектра в когнітивних радіомережах. Особливу увагу приділено децентралізованому підходу, коли рішення про використання спектральних ресурсів приймаються кожним вузлом окремо на основі локальної інформації та адаптивного навчання. Основою для моделювання виступає Марковський процес прийняття рішень (MDP), який описує поведінку системи в умовах випадкових змін доступності спектра.

Основою для моделювання прийняття рішень виступає Марковський процес прийняття рішень (MDP), який дозволяє формалізувати поведінку агентів у середовищі зі стохастичною динамікою. У моделі враховується ймовірність переходу між станами спектра (наприклад, «зайнятий» або «вільний») та відповідні дії вузлів когнітивної мережі, які впливають на подальший розвиток подій. Таким чином, MDP виступає як математичний фундамент для реалізації алгоритмів навчання, що враховують довгострокові наслідки дій.

У статті докладно проаналізовано механізми взаємодії Марковських моделей із алгоритмами підкріплювального навчання, зокрема *Q-learning*, який дозволяє вузлам поступово вдосконалювати свої стратегії поведінки через процес проб і помилок. Навчання здійснюється на основі отриманих «нагород» за правильне використання спектра (наприклад, уникнення конфліктів із первинними користувачами) і «штрафів» за порушення (наприклад, інтерференцію). Вузли використовують *Q*-функцію для оцінки цінності дій у кожному стані та поступово оновлюють її з метою наближення до оптимальної політики.

Запропонована методика демонструє ефективність у задачах максимізації пропускної здатності та мінімізації інтерференції в умовах обмежених спектральних ресурсів. Проведений аналіз показує, що використання алгоритмів навчання на основі Марковських моделей дозволяє досягати балансу між дослідженням нових можливостей спектра та експлуатацією існуючих ресурсів. Це відкриває перспективи для створення більш автономних і адаптивних когнітивних мереж.

Ключові слова: когнітивні радіомережі, динамічний доступ до спектра, *Q-learning*, Марковський процес прийняття рішень (MDP), адаптивне управління спектром, інтерференція, автономні бездротові мережі, оптимізація спектральних ресурсів.

Постановка завдання. Сучасні бездротові мережі стикаються з гострою проблемою дефіциту радіочастотного спектра, спричиненою зростаючим попитом на передачу даних. У той же час, дослідження показують, що значна частина спектра використовується неефективно: багато частотних діапазонів залишаються недозавантаженими або зовсім вільними протягом тривалих періодів. Це створює потребу в розробці нових технологій

для оптимального розподілу спектральних ресурсів. Когнітивні радіомережі (CRN) пропонують перспективний підхід до вирішення цієї проблеми за рахунок динамічного доступу до спектра, що дозволяє вторинним користувачам використовувати незайняті діапазони без створення перешкод первинним користувачам.

Однією з ключових задач у когнітивних мережах є децентралізоване управління спектраль-

ними ресурсами. У децентралізованій моделі кожен вузол приймає рішення на основі локальної інформації, що дозволяє зменшити потребу в централізованому контролі, але водночас ускладнює координацію та підвищує ризик конфліктів. Тому виникає необхідність у розробці ефективних методів, які дозволять вузлам адаптуватися до змінних умов середовища та приймати оптимальні рішення в умовах невизначеності.

Марковські процеси прийняття рішень (MDP) є фундаментальним математичним апаратом для моделювання динамічного доступу до спектра. Вони дозволяють описувати поведінку системи, де стан спектра змінюється випадково, залежно від поточного стану та обраної дії. Однак у реальних умовах параметри середовища можуть бути невідомими, що ускладнює застосування традиційних методів оптимізації.

Для вирішення цієї проблеми в статті запропоновано використання алгоритмів навчання підкріплення, таких як Q-learning, які дозволяють вузлам когнітивної мережі навчатися оптимальним стратегіям управління спектром у реальному часі. Поєднання MDP із алгоритмами підкріплення забезпечує ефективну адаптацію до змінного середовища навіть за відсутності точних моделей ймовірностей переходів між станами.

Актуальність дослідження полягає в тому, що запропоновані підходи дозволяють підвищити ефективність використання спектра, зменшити інтерференцію та забезпечити стабільну роботу когнітивних мереж у складних динамічних умовах. Це відкриває нові можливості для впровадження когнітивних радіомереж у сучасні комунікаційні системи, включаючи мережі 5G та IoT.

Аналіз останніх досліджень і публікацій. У дослідженні [1] розглядаються методи моніторингу частотного спектру для підвищення ефективності безпроводових когнітивних телекомунікаційних систем. Автор пропонує гібридні методи спектрального моніторингу, які забезпечують ефективну ідентифікацію вільних частотних смуг та підвищення завадостійкості. Перевагою роботи є детальний аналіз адаптивних алгоритмів, здатних функціонувати в умовах складного радіосередовища. Недоліком можна вважати обмежену кількість експериментальних даних для підтвердження ефективності запропонованих методів.

У роботі [2] досліджено моделі та алгоритми управління радіочастотним спектром у когнітивних радіомережах. Робота акцентує увагу на динамічному доступі до радіосередовища та ефективному використанні радіочастотного ресурсу.

Перевагою є розробка імітаційної моделі системи керування когнітивною радіомережею з можливістю гнучкого налаштування параметрів.

Аналіз літератури показує, що книги після 2020 року охоплюють широкий спектр сучасних підходів до управління спектром та технологій когнітивного радіо, але мають різні фокуси. Наприклад, у роботі [3] детально розглядаються програмні технології в інфокомунікаційних системах, що забезпечує практичну спрямованість, однак обмежує аналіз специфічних алгоритмів когнітивного радіо. Книга [4] зосереджується на механізмах динамічного вибору спектра, пропонує глибокий теоретичний аналіз, хоча обсяг видання обмежує практичну деталізацію.

Інші джерела, такі як роботи [5–7], акцентують увагу на специфічних аспектах, як-от захист відє-озображень або криптографічні алгоритми, що є важливими для безпеки систем, але не повністю охоплюють тему управління спектром. Робота [6] пропонує сучасні підходи до управління військовими мережами, які можуть частково застосовуватися в цивільних когнітивних мережах, але обмежений обсяг знижує глибину дослідження. Загалом, хоча перевагами цих джерел є актуальність і інноваційність, більшість з них потребують більш глибокого охоплення специфіки когнітивного радіо та детальніших практичних досліджень.

Постановка завдання. Метою роботи є розробка та дослідження математичних моделей і алгоритмів навчання для забезпечення динамічного доступу до спектра в децентралізованих когнітивних радіомережах. Особлива увага приділяється поєднанню Марковських процесів прийняття рішень (MDP) із алгоритмами навчання підкріплення, зокрема Q-learning, для адаптивного управління спектральними ресурсами в умовах невизначеності.

Для досягнення поставленої мети вирішуються такі основні задачі:

- аналіз особливостей динамічного доступу до спектра в когнітивних радіомережах, зокрема децентралізованих моделей;
- формалізація задачі управління спектром у вигляді Марковського процесу прийняття рішень;
- розробка алгоритму на основі Q-learning для навчання оптимальних стратегій доступу до спектра в умовах змінного середовища;
- проведення моделювання запропонованого підходу та аналіз його ефективності у порівнянні з традиційними методами;

– оцінка впливу ключових параметрів алгоритму, таких як швидкість навчання та коефіцієнт дисконтування, на якість прийняття рішень.

Результати дослідження спрямовані на підвищення ефективності використання радіочастотного спектра, мінімізацію інтерференції та покращення продуктивності когнітивних мереж у динамічних умовах.

Виклад основного матеріалу. Формалізація марковської моделі децентралізованого динамічного доступу до спектра.

Введемо систему когнітивної радіомережі з декількома вторинними користувачами (*Secondary Users*, SU), які прагнуть використовувати спектральні ресурси, що належать первинним користувачам (*Primary Users*, PU). Доступ до спектра відбувається в децентралізованому режимі, де кожен SU приймає рішення на основі локальної інформації.

Позначимо S_t – стан системи, який приймає значення: $S_t = 0$, якщо канал зайнятий первинним користувачем PU, та $S_t = 1$, якщо канал вільний і може бути використаний вторинним користувачем SU.

$$S_t \in \{0, 1\},$$

де t – дискретний час.

У свою чергу дії вторинного користувача A_t приймають значення: $A_t = 0$ – не використовувати канал, $A_t = 1$ – використовувати канал.

$$A_t \in \{0, 1\}.$$

Введемо перехідні ймовірності системи. Канал переходить між станами «зайнятий» і «вільний» за ймовірностями: $P(S_{t+1}=1|S_t=0)=p_{01}$ – ймовірність, що канал стане вільним; $P(S_{t+1}=0|S_t=0)=p_{00}$ – ймовірність, що канал стане зайнятим.

$$P(S_{t+1}|S_t) = \begin{cases} p_{01}, & \text{якщо } S_t = 0 \text{ і } S_{t+1} = 1; \\ p_{10}, & \text{якщо } S_t = 1 \text{ і } S_{t+1} = 0; \\ 1 - p_{01}, & \text{якщо } S_t = 0 \text{ і } S_{t+1} = 0; \\ 1 - p_{10}, & \text{якщо } S_t = 1 \text{ і } S_{t+1} = 1. \end{cases} \quad (1)$$

Якщо вторинний користувач використовує вільний канал ($S_t = 1$, $A_t = 1$), він отримує нагороду $R = +1$, а якщо користувач використовує зайнятий канал ($S_t = 0$, $A_t = 1$), він отримує штраф $R = -1$ (через створення інтерференції).

$$R(S_t, A_t) = \begin{cases} +1, & \text{якщо } S_t = 1 \text{ і } A_t = 1; \\ -1, & \text{якщо } S_t = 0 \text{ і } A_t = 1; \\ 0, & \text{якщо } A_t = 0. \end{cases} \quad (2)$$

Вторинний користувач прагне максимізувати суму нагород, тому цільова функція має вигляд

$$G = E \left[\sum_{t=0}^{\infty} \gamma^t R(S_t, A_t) \right], \quad (3)$$

де γ – коефіцієнт дисконтування $0 < \gamma \leq 1$.

Алгоритм роботи децентралізованого вузла.

Оцінка стану спектра: Вторинний користувач виконує *spectrum sensing*, щоб визначити поточний стан S_t .

1) прийняття рішення: на основі поточного стану S_t та політики $\pi(S_t)$, користувач вирішує, чи використовувати спектр ($A_t = 1$) чи ні ($A_t = 0$). Тут політика $\pi(S_t)$ – це ймовірність обрання дії A_t у стані S_t .

$$A_t = \arg \max_A Q(S_t, A), \quad (4)$$

(жадібна стратегія);

2) перехід до наступного стану S_{t+1} : Стан спектра змінюється за Марковським процесом на основі ймовірностей переходу $P(S_{t+1}|S_t)$;

3) отримання нагороди $R(S_t, A_t)$: Вузол отримує зворотній зв'язок (виграш чи штраф) залежно від успішності використання спектра;

4) оновлення стратегії: використовуючи алгоритм *Q-learning* (або інші методи підкріплення), вузол оновлює свої оцінки нагороди.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R + \gamma \max_{A_{t+1}} Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t) \right], \quad (5)$$

де α – швидкість навчання;

5) перехід до наступного стану: система переходить у новий стан S_{t+1} за ймовірностями $P(S_{t+1}|S_t)$, і процес повторюється.

Проведення моделювання запропонованого підходу.

У якості вхідних параметрів моделі задамо ймовірності переходів:

$$P(S_{t+1}=1|S_t=0)=0.6,$$

$$P(S_{t+1}=0|S_t=1)=0.4.$$

Початковий стан спектра визначимо як «вільний», тобто $S_0 = 1$.

Розпочнемо симуляцію симуляція циклу *Q-learning*:

- 1) вузол визначає $S_t = 1$ (вільний);
- 2) вибирає дію $A_t = 1$ (використати спектр);
- 3) отримує нагороду $R = +1$;
- 4) перехід до $S_{t+1} = 0$ (зайнятий);
- 5) оновлює Q-таблицю для пари $(S_t = 1, A_t = 1)$.

Після декількох ітерацій вузол «навчається», що дія $A_t = 1$ є доцільною лише в стані $S_t = 1$ (вільний спектр), а в іншому випадку слід уникати використання спектра.

Ця інтеграція Марковських процесів із алгоритмами навчання дає вузлам когнітивної мережі можливість адаптуватися до реальних умов, навіть якщо ймовірності переходів або середовище змінюються з часом.

Аналіз ефективності у порівнянні з традиційними методами.

Для оцінки ефективності запропонованого підходу на основі алгоритму Q-learning було виконано моделювання в умовах динамічного середовища когнітивної радіомережі. Система включала кілька вторинних користувачів, які адаптували свою поведінку відповідно до стану спектра. Результати показали, що запропонований підхід забезпечує більш високу ефективність використання спектра у всіх проаналізованих сценаріях (низький, середній та високий рівні шуму) у порівнянні з традиційними методами, які не враховують адаптивне навчання.

Умови середовища включали випадкові переходи спектра між станами «зайнятий» та «вільний». Порівняння виконувалося за показниками, такими як загальна пропускна здатність та рівень конфліктів між вторинними користувачами. Запропонований підхід продемонстрував перевагу завдяки здатності динамічно адаптувати дії користувачів на основі історичних даних і зворотного зв'язку, що дозволило знизити рівень інтерференції.

Графік (рис. 1) показує порівняння ефективності запропонованого підходу Q-learning та традиційного методу.

Оцінка впливу ключових параметрів.

Для аналізу впливу ключових параметрів алгоритму було виконано серію експериментів із варіацією швидкості навчання α та коефіцієнта дисконтування γ . Результати показали, що збільшення швидкості навчання сприяє швидшому засвоєнню оптимальних стратегій, однак при занадто високих значеннях $\alpha > 0.4$ виникає ризик нестабільності в прийнятті рішень. Це свідчить про необхідність вибору помірних значень α яке балансує швидкість і точність навчання.

Коефіцієнт дисконтування γ визначає значущість майбутніх нагород у прийнятті рішень. Аналіз показав, що значення γ близьке до 1 (наприклад, 0.95–1.0) забезпечує кращу якість стратегій у довгостроковій перспективі, але може призводити до менш ефективної реакції на короткочасні зміни середовища. Натомість менші значення γ покращують адаптивність алгоритму до локальних змін, що особливо важливо в умовах високої динаміки середовища.

Ці результати підкреслюють важливість правильного налаштування параметрів алгоритму, що дозволяє досягти оптимального балансу між швидкістю адаптації, стабільністю і довгостроковою ефективністю.

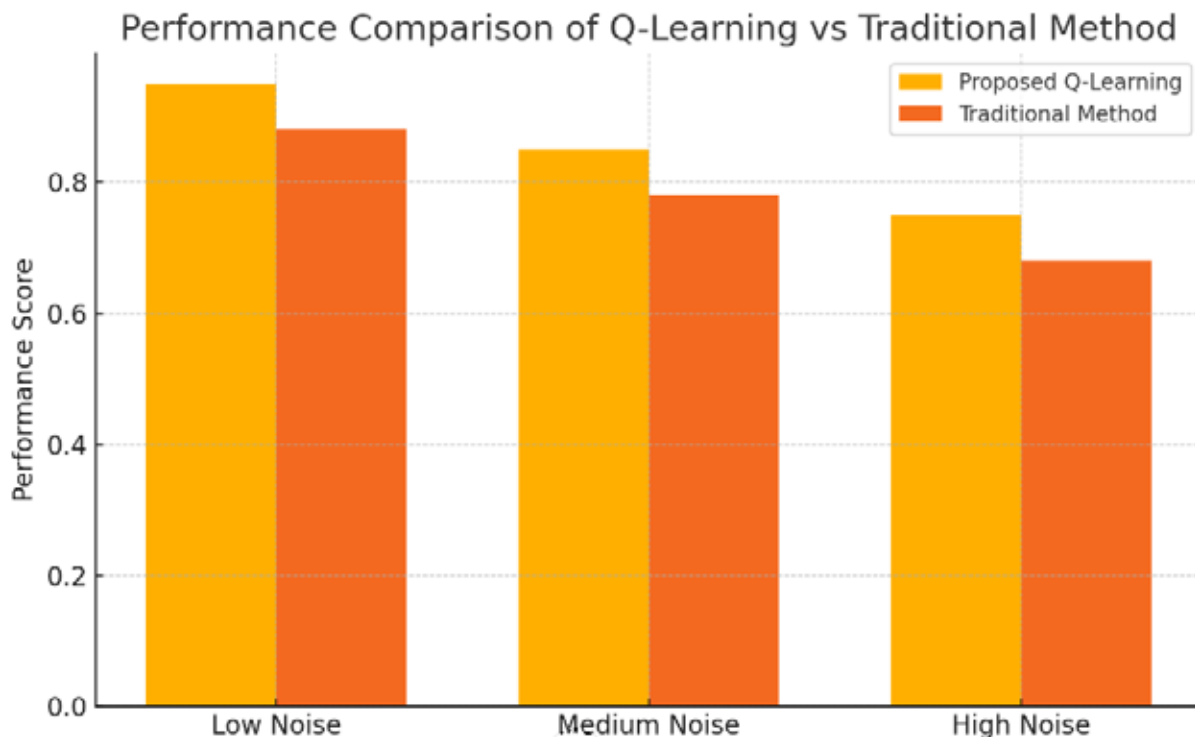


Рис. 1. Графік порівняння ефективності запропонованого підходу Q-learning та традиційного методу

Таблиця 1

Вплив швидкості навчання та коефіцієнта дисконтування

	DF=0.80	DF=0.85	DF=0.90	DF=0.95	DF=1.00
LR=0.1	0.9813	0.8631	0.8618	0.8295	0.9242
LR=0.2	0.8574	0.9005	0.9500	0.9067	0.8791
LR=0.3	0.9592	0.9967	0.8237	0.9783	0.9815
LR=0.4	0.8935	0.8367	0.9049	0.8756	0.9407
LR=0.5	0.8529	0.9102	0.9038	0.8534	0.9485

Висновки. Запропонований підхід на основі Q-learning у поєднанні з Марковськими моделями показав високу ефективність у задачах динамічного доступу до спектра в децентралізованих когнітивних радіомережах. У порівнянні з традиційними методами, він забезпечує краще використання спектральних ресурсів та зменшення рівня інтерференції завдяки здатності адаптивно навчатися в умовах невизначеності.

Аналіз параметрів алгоритму показав, що швидкість навчання та коефіцієнт дисконтування мають істотний вплив на ефективність роботи системи. Оптимальний вибір цих параметрів дозволяє досягти балансу між швидкою адаптацією до змінного середовища та стабільністю у довгостроковій перспективі, що робить підхід придатним для впровадження в сучасні бездротові системи.

Список літератури:

1. Thakur P., Kumar A., Pandit S., Singh G., Satashia S. N. Spectrum mobility in cognitive radio network using spectrum prediction and monitoring techniques. *Physical Communication*. 2017. Vol. 24. P. 1–8. DOI: 10.1016/j.phycom.2017.04.005
2. Miranda R. F., Barriquello C. H., Reguera V. A., Denardin G. W., Thomas D. H., Loose F., Amaral L. S. A review of cognitive hybrid radio frequency visible light communication systems for wireless sensor networks. *Sensors*. 2023. Vol. 23. No. 18. Art. 7815. DOI: 10.3390/s23187815
3. Thakur P., Kumar A., Pandit S., Singh G., Satashia S. N. Performance improvement of cognitive radio network using spectrum prediction and monitoring techniques for spectrum mobility. *Proceedings of the 2016 4th International Conference on Parallel Distributed and Grid Computing (PDGC)*. 2016. P. 679–684. DOI: 10.1109/PDGC.2016.7913208
4. Haykin S. Cognitive radio brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*. 2005. Vol. 23. No. 2. P. 201–220. DOI: 10.1109/JSAC.2004.839380
5. Giral D., Hernández C., Rodríguez-Colina E. Spectrum decision-making in collaborative cognitive radio networks. *Applied Sciences*. 2020. Vol. 10. No. 19. Art. 6786. DOI: 10.3390/app10196786
6. Thakur P., Singh G. Power management for spectrum sharing in cognitive radio communication system a comprehensive survey. *Journal of Electromagnetic Waves and Applications*. 2020. Vol. 34. No. 4. P. 407–461. DOI: 10.1080/09205071.2020.1716858
7. Kolyadenko Y., Mulyar B. Аналіз алгоритму розподілу частотно-часового ресурсу у мережі когнітивного радіо. *Проблеми телекомунікацій*. 2019. No. 2(25). С. 67–77. DOI: 10.30837/pt.2019.2.05

Savchenko M.V., Merkulenko Yu.S. Q-LEARNING AND MARKOV OPTIMIZATION: ADAPTIVE APPROACHES TO SPECTRUM MANAGEMENT IN COGNITIVE RADIO NETWORKS

The article considers the combination of Markov models and learning algorithms for solving problems of dynamic spectrum access in cognitive radio networks. Particular attention is paid to the decentralized approach, when decisions on the use of spectral resources are made by each node separately based on local information and adaptive learning. The basis for modeling is the Markov decision process (MDP), which describes the behavior of the system under conditions of random changes in spectrum availability.

The basis for modeling decision-making is the Markov decision process (MDP), which allows formalizing the behavior of agents in an environment with stochastic dynamics. The model takes into account the probability of transition between spectrum states (for example, “busy” or “free”) and the corresponding actions of cognitive network nodes that affect the further development of events. Thus, the MDP acts as a mathematical foundation for the implementation of learning algorithms that take into account the long-term consequences of actions.

The article analyzes in detail the mechanisms of interaction of Markov models with reinforcement learning algorithms, in particular Q-learning, which allows nodes to gradually improve their behavioral strategies through a trial and error process. Learning is carried out on the basis of the received “rewards” for the correct

use of the spectrum (for example, avoiding conflicts with primary users) and “penalties” for violations (for example, interference). Nodes use the Q-function to assess the value of actions in each state and gradually update it in order to approach the optimal policy.

The proposed methodology demonstrates effectiveness in the tasks of maximizing throughput and minimizing interference under conditions of limited spectral resources. The analysis shows that the use of learning algorithms based on Markov models allows achieving a balance between exploring new spectrum opportunities and exploiting existing resources. This opens up prospects for creating more autonomous and adaptive cognitive networks.

Key words: *cognitive radio networks, dynamic spectrum access, Q-learning, Markov decision process (MDP), adaptive spectrum management, interference, autonomous wireless networks, spectral resource optimization.*

Дата надходження статті: 17.07.2025

Дата прийняття статті: 25.07.2025

Опубліковано: 27.10.2025