



МЕТОДИ НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти *Третій (доктор філософії)*

Галузь знань	12 Інформаційні технології
Спеціальність	123 Комп'ютерна інженерія
Освітня програма	Комп'ютерна інженерія
Статус дисципліни	Вибіркова
Форма навчання	очна(денна)
Рік підготовки, семестр	2 курс, осінній семестр
Обсяг дисципліни	4 кредити
Семестровий контроль/ контрольні заходи	Залік
Розклад занять	Лекцій 18 (36 годин) Лабораторних 9 (18 годин)
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com Лабораторні: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com
Розміщення курсу	

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна “Методи навчання з підкріпленням” забезпечує набуття необхідних знань щодо основних методів, які дозволяють використовувати принципи навчання з підкріпленням у різних сферах наукової та практичної діяльності. В рамках вивчення даної дисципліни також розглядаються основні технологічні рішення у сфері науки, які забезпечують акумулятивний підхід до навчання з підкріпленням, який орієнтований на цілеспрямоване навчання через взаємодію з навколишнім середовищем.

Вивчення даної дисципліни майбутніми науковцями дозволить їм набути важливих компетенцій, оскільки, в зв'язку з накопиченням значних обсягів даних у різних сферах діяльності людини, виникає широкий спектр задач, для вирішення яких актуальною стає така поведінка об'єкта в навколишньому середовищі, яка характеризується оптимальністю.

Метою вивчення дисципліни «Методи навчання з підкріпленням» є підготовка фахівців, здатних розв'язувати комплексні проблеми в галузі дослідницько-інноваційної діяльності у сфері навчання з підкріпленням шляхом вивчення теоретичних та практичних положень побудови поведінкових моделей, які характеризуються закритим циклом навчання, відсутністю прямих інструкцій щодо конкретних дій у кожний момент часу та використанням сигналів винагороди, які

можуть формуватися протягом тривалих періодів часу, що передбачає глибокого усвідомлення наявних та створення нових підходів.

Предметом дисципліни є:

- методи навчання з підкріпленням для вирішення задачі з єдиним станом;
- методи навчання з підкріпленням для вирішення марковських задач прийняття рішення;
- методи на основі динамічного програмування та методи Монте-Карло;
- наближені методи навчання з підкріпленням з урахуванням політики;
- наближені методи навчання з підкріпленням поза політик

Основні результати навчання

Здобувачі наукового ступеня доктора філософії після засвоєння навчальної дисципліни мають продемонструвати такі **компетентності**.

1. *Загальні компетентності*: здатність до абстрактного мислення, аналізу і синтезу; здатність до пошуку, оброблення та аналізу інформації з різних джерел щодо точних та наближених методів навчання з підкріпленням; здатність набуття універсальних навичок усної та письмової презентації власного наукового дослідження за результатами практичного застосування сучасних інформаційних технологій у науковій діяльності.

2. *Спеціальні компетентності*: здатність ефективно застосовувати основні методи навчання з підкріпленням, а саме: методи навчання з підкріпленням для вирішення задачі з єдиним станом, методи навчання з підкріпленням для вирішення марковських задач прийняття рішення, методи на основі динамічного програмування та методи Монте-Карло, наближені методи навчання з підкріпленням з урахуванням політики, наближені методи навчання з підкріпленням поза політикою, здатність аргументувати вибір методу розв'язання наукової задачі, критично оцінювати отримані результати та захищати прийняті рішення щодо вибору методів та навчання з підкріпленням.

За результатами вивчення навчальної дисципліни «Методи навчання з підкріпленням» мають бути отримані такі **знання**.

1. Мати передові концептуальні та методологічні знання у сфері навчання з підкріпленням, а також дослідницькі навички, достатні для проведення наукових і прикладних досліджень в галузі машинного навчання на рівні останніх світових досягнень з комп'ютерної інженерії, ІТ-інфраструктур, інформаційних технологій.

2. Знати сучасні методи проведення досліджень у сфері навчання з підкріпленням для побудови ефективних поведінкових моделей шляхом вивчення теоретичних та практичних підходів до їх створення, використання результатів досліджень для уточнення наукових висновків та запровадження ефективних механізмів майбутньої поведінки об'єктів в навколишньому середовищі.

3. Знати і розуміти наукові і математичні положення, що лежать в основі методів навчання з підкріпленням для створення ефективних алгоритмів машинного навчання, методів побудови та дослідження математичних моделей та технологій адаптивних та інтелектуальних обчислень при добуванні даних.

Уміння, які мають бути отримані у рамках вивчення навчальної дисципліни «Методи навчання з підкріпленням».

1. Вміти ефективно здійснювати пошук та критичний аналіз інформації з різних джерел щодо методів та технологій навчання з підкріпленням.

2. Вміти розв'язувати задачі синтезу та аналізу об'єктів дослідження при вирішенні задач навчання з підкріпленням.

3. Вміти створювати та реалізовувати математичні моделі поведінки об'єкта в навколишньому середовищі в залежності від параметрів, які характеризують стан навколишнього середовища;
4. Вміти застосовувати технології адаптивних та інтелектуальних обчислень при реалізації алгоритмів машинного навчання.

Здобувачі наукового ступеня також мають бути **здатні**.

1. Застосовувати прикладні бібліотеки та програмні системи, які використовуються при машинному навчанні.
2. Володіти методами та технологіями програмування з використанням прикладних бібліотек та програмних систем, призначених для машинного навчання при реалізації методів навчання з підкріпленням

Таке поєднання загальних та спеціальних компетентностей, теоретичних та практичних знань, умінь та здатностей сприяє підвищенню науково-практичного рівня здобувачів наукового ступеня доктора філософії задля здійснення ними ефективних наукових досліджень.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Для успішного оволодіння дисципліною необхідні знання:

- основ математичного аналізу, теорії функцій та математичної статистики;
- основ функціонування операційних систем;
- основ програмування мовою Python.

Відповідно до освітньої програми необхідно попередньо оволодіти знаннями з дисциплін: "Програмування", "Об'єктно-орієнтоване програмування", "Системне програмування", "Структури даних та алгоритми", "Інженерія програмного забезпечення", "Алгоритми та методи обчислень", "Дискретна математика".

Компетентності, знання та вміння, отримані в рамках вивчення даної дисципліни, можуть бути застосовані для отримання обґрунтованих результатів досліджень та підвищення наукового рівня дисертаційних робіт.

3. Зміст навчальної дисципліни

Розділ 1. Задача навчання з підкріпленням

Тема 1.1. Основні поняття про навчання з підкріпленням.

Тема 1.2. Приклади навчання з підкріпленням.

Тема 1.3. Елементи навчання з підкріпленням.

Тема 1.4. Обмеження навчання з підкріпленням та сфера його застосування.

Тема 1.5. Розширений приклад навчання з підкріпленням: гра в хрестики-нулики

Розділ 2. Задачі про «багаторуких бандитів»

Тема 2.1. Задача про «n-рукого бандита».

Тема 2.2. Методи типу «дія-цінність».

Тема 2.3. Інкрементальна реалізація методу «дія-цінність».

Тема 2.4. Поняття про нестационарні задачі.

Тема 2.5. Оптимістичні початкові цінності.

Тема 2.6. Верхня межа довіри при виборі дій.

Тема 2.7. Градієнтні «бандити».

Тема 2.8. Асоціативний пошук («контекстний бандит»).

Розділ 3. Скінченні Марківські процеси прийняття рішень

- Тема 3.1. Інтерфейс агент-середовище.
- Тема 3.2. Цілі і винагороди.
- Тема 3.3. Поняття про розрахунок прибутку
- Тема 3.4. Уніфікована нотація для епізодичних та постійних задач.
- Тема 3.5. Властивість Маркова.
- Тема 3.6. Марківський процес прийняття рішення.
- Тема 3.7. Ціннісні функції.
- Тема 3.8. Оптимальні ціннісні функції.
- Тема 3.9. Оптимальність та апроксимація.

Розділ 4. Динамічне програмування

- Тема 4.1. Оцінка політики.
- Тема 4.2. Удосконалення політики.
- Тема 4.3. Ітерації політики.
- Тема 4.4. Ітерація цінності.
- Тема 4.5. Асинхронне динамічне програмування.
- Тема 4.6. Узагальнена ітерація політики.
- Тема 4.7. Ефективність динамічного програмування.

Розділ 5. Методи Монте Карло

- Тема 5.1. Прогноз Монте-Карло.
- Тема 5.2. Оцінка цінностей дій методом Монте-Карло .
- Тема 5.3. Керування з використанням методу Монте-Карло.
- Тема 5.4. Монте-Карло керування без стартових досліджень.
- Тема 5.5. Поза політичний прогноз через вибірку важливості.
- Тема 5.6. Додаткова (інкрементальна) реалізація методом Монте-Карло
- Тема 5.7. Монте-Карло керування поза політикою
- Тема 5.8. Вибірка важливості для усіченого прибутку

Розділ 6. Навчання за методом часових відмінностей

- Тема 6.1. TD передбачення.
- Тема 6.2. Переваги методів прогнозування TD
- Тема 6.3. Оптимальність методу TD(0)
- Тема 6.4. Sarsa: TD керування в рамках політики
- Тема 6.5. Q-навчання: метод керування TD поза політикою
- Тема 6.6. . Ігри, післястани та інші спеціальні випадки

Розділ 7. Сліди придатності

- Тема 7.1. n-кроковий TD метод прогнозування
- Тема 7.2. Погляд вперед TD(λ)i
- Тема 7.3. Перегляд назад для TD(λ)
- Тема 7.4. Еквівалентності переглядів вперед та назад
- Тема 7.5. Метод Sarsa(λ)
- Тема 7.6. Метод Воткінса Q(λ)
- Тема 7.7. Сліди придатності поза політикою з використанням вибірки важливості
- Тема 7.8. Питання впровадження
- Тема 7.9. Змінна λ

Розділ 8. Планування та навчання за допомогою табличних методів

- Тема 8.1. Моделі та планування

Тема 8.2. Інтеграційне планування, дії та навчання.
Тема 8.3. Планування у випадку помилкової моделі
Тема 8.4. Пріоритетна прогонка
Тема 8.5. Повне vs. вибіркоче резервне копіювання
Тема 8.6. Траєкторні вибірки
Тема 8.7. Динамічне програмування в реальному часі
Тема 8.8. Планування під час прийняття рішення.
Тема 8.9. Евристичний пошук
Тема 8.10. Алгоритми поширення
Тема 8.11. Монте-Карло для пошуку у дереві.

Наближені методи розв'язування

Розділ 9. Наближення за політикою цінностей дій

Тема 9.1. Прогнозування цінності наближеними функціями
Тема 9.2. Методи градієнтного спуску
Тема 9.3. Лінійні методи
Тема 9.4. Управління з наближеними функціями
Тема 9.5. Принципи застосування бутстрепа
Тема 9.6. Нелінійні функції наближення: штучні нейронні мережі.

Розділ 10. Керування за політикою з наближенням

Тема 10.1. Епізодне семі-градієнтне керування
Тема 10.2. Семі-градієнтний n-кроковий метод Sarsa
Тема 10.3. Диференціальний семі-градієнтний n-кроковий метод Sarsa

Розділ 11. Методи поза політикою з наближенням

Тема 11.1. Семі-градієнтні методи
Тема 11.2. Градієнтний спуск для помилки Беллмана
Тема 11.3. Градієнтні TD – методи.
Тема 11.4. Емпфатні TD – методи.

Розділ 12. Методи градієнта політики

Тема 12.1. Наближена політика та її переваги.
Тема 12.2. Теорема градієнта політики
Тема 12.3. Методи «актор-критик»
Тема 12.4. Градієнт політики для неперервних задач
Тема 12.5. Параметризація політики для неперервних задач.

4. Навчальні матеріали та ресурси

Базова:

1. Новотарський М.А., Нестеренко Б.Б. Штучні нейронні мережі: обчислення.– К.: Інститут математики НАН України, 2004.– 408 с.

2. Новотарський М.А., Нестеренко Б.Б., Формальні засоби моделювання паралельних процесів та систем // Праці Інституту математики НАН України.– Т.90.–Київ: Ін-т математики НАН України, 2012.–334с.
3. Novotarskyi M., Kuzmich V. USAK Method for the Reinforcement Learning // Information, Computing and Intelligent Systems. – 2020. –№1. – p.18-21, DOI:10.20535/2708-4930.1.2020.216042
4. Novotarskyi M.A., Stirenko S.G., Gordienko Y.G., Kuzmych V.A. Deep reinforcement learning with sparse distributed memory for “Water World” problem solving //Radio Electronics, Computer Science, Control.– 2021.– Vol.1 № 1.– P.136-143. DOI: <https://doi.org/10.15588/1607-3274-2021-1-14>
5. Sutton R.S., Barto A.G. Reinforcement Learning, second edition: An Introduction.– Bradford Books.– 2018.– 552p.
6. Winder P. Reinforcement Learning: Industrial Applications of Intelligent Agents.– O'Reilly Media– 2020.–408 p

Додаткова:

1. Graesser L., Keng W.L. Foundations of Deep Reinforcement Learning: Theory and Practice in Python.– Addison-Wesley Professional, 2019. – 416 p.
2. Fu X., Wang L. Data Mining with Computational Intelligence. – Springer, 2005. – 287 p.
3. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних. – К.: Знання, 2014. – 599 с.

Навчальний контент

7. Методика опанування навчальної дисципліни (освітнього компонента)

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Розділ 1. Задача навчання з підкріпленням Тема 1.1. Основні поняття про навчання з підкріпленням. Тема 1.2. Приклади навчання з підкріпленням. Тема 1.3. Елементи навчання з підкріпленням. Тема 1.4. Обмеження навчання з підкріпленням та сфера його застосування. Тема 1.5. Розширений приклад навчання з підкріпленням: гра в хрестики-нулики	10	4	2	4
Розділ 2. Задачі про «багаторуких бандитів» Тема 2.1. Задача про «n-рукого бандита». Тема 2.2. Методи типу «дія-цінність». Тема 2.3. Інкрементальна реалізація методу «дія-цінність».	10	4	2	4

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Тема 2.4. Поняття про нестационарні задачі. Тема 2.5. Оптимістичні початкові цінності. Тема 2.6. Верхня межа довіри при виборі дій. Тема 2.7. Градієнтні «бандити». Тема 2.8. Асоціативний пошук («контекстний бандит»).				
Розділ 3. Скінченні Марківські процеси прийняття рішень Тема 3.1. Інтерфейс агент-середовище. Тема 3.2. Цілі і винагороди. Тема 3.3. Поняття про розрахунок прибутку Тема 3.4. Уніфікована нотація для епізодичних та постійних задач. Тема 3.5. Властивість Маркова. Тема 3.6. Марківський процес прийняття рішення. Тема 3.7. Ціннісні функції. Тема 3.8. Оптимальні ціннісні функції. Тема 3.9. Оптимальність та апроксимація.	10	4	2	4
Розділ 4. Динамічне програмування Тема 4.1. Оцінка політики. Тема 4.2. Удосконалення політики. Тема 4.3. Ітерації політики. Тема 4.4. Ітерація цінності. Тема 4.5. Асинхронне динамічне програмування. Тема 4.6. Узагальнена ітерація політики. Тема 4.7. Ефективність динамічного програмування.	10	4	2	4
Розділ 5. Методи Монте Карло Тема 5.1. Прогноз Монте-Карло. Тема 5.2. Оцінка цінностей дій методом Монте-Карло. Тема 5.3. Керування з використанням методу Монте-Карло. Тема 5.4. Монте-Карло керування без стартових досліджень. Тема 5.5. Поза політичний прогноз через вибірку важливості. Тема 5.6. Додаткова (інкрементальна) реалізація методом Монте-Карло. Тема 5.7. Монте-Карло керування поза політикою. Тема 5.8. Вибірка важливості для усіченого прибутку.	10	4	2	4
Розділ 6. Навчання за методом часових відмінностей Тема 6.1. TD передбачення. Тема 6.2. Переваги методів прогнозування TD Тема 6.3. Оптимальність методу TD(0)	10	2	2	6

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Тема 6.4. Sarsa: TD керування в рамках політики Тема 6.5. Q-навчання: метод керування TD поза політикою Тема 6.6. . Ігри, післястани та інші спеціальні випадки				
Розділ 7. Сліди придатності Тема 7.1. n-кроковий TD метод прогнозування Тема 7.2. Погляд вперед TD(λ) Тема 7.3. Перегляд назад для TD(λ) Тема 7.4. Еквівалентності переглядів вперед та назад Тема 7.5. Метод Sarsa(λ) Тема 7.6. Метод Воткінса Q(λ) Тема 7.7. Сліди придатності поза політикою з використанням вибірки важливості Тема 7.8. Питання впровадження Тема 7.9. Змінна λ	8	2	0	6
Розділ 8. Планування та навчання за допомогою табличних методів Тема 8.1. Моделі та планування Тема 8.2. Інтеграційне планування, дії та навчання. Тема 8.3. Планування у випадку помилкової моделі Тема 8.4. Пріоритетна прогонка Тема 8.5. Повне vs. вибіркоче резервне копіювання Тема 8.6. Траєкторні вибірки Тема 8.7. Динамічне програмування в реальному часі Тема 8.8. Планування під час прийняття рішення. Тема 8.9. Евристичний пошук Тема 8.10. Алгоритми поширення Тема 8.11. Монте-Карло для пошуку у дереві.	10	2	0	8
Розділ 9. Наближення за політикою цінностей дій Тема 9.1. Прогнозування цінності наближеними функціями Тема 9.2. Методи градієнтного спуску Тема 9.3. Лінійні методи Тема 9.4. Управління з наближеними функціями Тема 9.5. Принципи застосування бутстрепа Тема 9.6. Нелінійні функції наближення: штучні нейронні мережі.	10	2	0	8

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Розділ 10. Керування за політикою з наближенням Тема 10.1. Епізодне семі-градієнтне керування Тема 10.2. Семі-градієнтний n-кроковий метод Sarsa Тема 10.3. Диференціальний семі-градієнтний n-кроковий метод Sarsa	10	2	2	6
Розділ 11. Методи поза політикою з наближенням Тема 11.1. Семі-градієнтні методи Тема 11.2. Градієнтний спуск для помилки Беллмана Тема 11.3. Градієнтні TD – методи. Тема 11.4. Емпфатні TD – методи.	10	2	2	6
Розділ 12. Методи градієнта політики Тема 12.1. Наближена політика та її переваги. Тема 12.2. Теорема градієнта політики Тема 12.3. Методи «актор-критик» Тема 12.4. Градієнт політики для неперервних задач	10	2	2	6
Залік	2	2	0	0
Всього в семестрі:	120	36	18	66

8. Самостійна робота аспіранта

Метою проведення циклу лабораторних робіт є набуття студентами необхідних практичних навичок використання методів та способів представлення та опрацювання даних для отримання прихованої інформації, методів дослідження математичних моделей опису шуканої інформації, технології добування даних для розроблення прогнозів щодо досліджуваних об'єктів та процесів.

№ з/п	Назва лабораторної роботи	Кількість ауд. годин
1	<i>Лабораторна робота № 1.</i> Створення зошита в середовищі Jupyter notebook.	2
2	<i>Лабораторна робота № 2.</i> Практичне застосування атрибутів та функцій модуля NumPy.	2
3	<i>Лабораторна робота № 3.</i> Вивчення та практичне застосування бібліотеки Matplotlib.	2
4	<i>Лабораторна робота № 4.</i> Вивчення та практичне застосування бібліотеки Pandas.	2

5	Лабораторна робота № 5. Первинна обробка набору даних "Yelp Dataset JSON" для тематичного дослідження.	2
6	Лабораторна робота № 6. Фільтрування великого набору даних фрагментами	2
7	Лабораторна робота № 7. Збереження структури DataFrame у форматі csv	2
8	Лабораторна робота № 8. Практичний аналіз текстових представлень великих обсягів даних	2
9	Лабораторна робота № 9. Візуалізація одержаних даних та підготовка презентації за результатами досліджень	2

Політика та контроль

9. Політика навчальної дисципліни (освітнього компонента)

Під час занять з навчальної дисципліни «Методи навчання з підкріпленням» аспіранти повинні дотримуватись певних дисциплінарних правил:

- забороняється запізнюватись на заняття;
- при вході викладача, на знак привітання, особи, які навчаються в КПІ ім. Ігоря Сікорського повинні встати;
- не допускаються сторонні розмови або інший шум, що заважає проведенню занять;
- виходити з аудиторії під час заняття допускається лише з дозволу викладача.
- не допускається користування мобільними телефонами та іншими технічними засобами без дозволу викладача.

Лабораторні роботи здаються особисто з попередньою перевіркою теоретичних знань, які необхідні для виконання лабораторної роботи. Перевірка практичних результатів включає перевірку коду та виконання тестових завдань.

В процесі навчання викладач має право нарахувати до 5 заохочувальних балів за дострокове виконання лабораторної роботи, за проявлений творчий підхід при виконанні індивідуального завдання або за активну участь у обговоренні питань, що пов'язані з тематикою лекції або практичного заняття.

За виконання та здачу лабораторної роботи після зазначеного дедлайну, за значну кількість пропущених занять, або за порушення правил поведінки на заняттях викладач може призначити до 5 штрафних балів.

При проведенні контрольних заходів та при виконанні лабораторних робіт аспіранти повинні дотримуватися правил академічної доброчесності. При виявленні значного відсотку списування або плагіату викладач може відмовити у прийнятті даної роботи та вимагати доброчесного виконання навчального плану.

10. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Види контролю з навчальної дисципліни «Методи навчання з підкріпленням» включають:
Поточний контроль: тестування закритими тестами.

Календарний контроль: провадиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Семестровий контроль: залік

Умови допуску до семестрового контролю: семестровий рейтинг більше 40 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

11. Додаткова інформація з дисципліни (освітнього компонента)

Викладання дисципліни «Методи навчання з підкріпленням» для спеціальності «Комп'ютерна інженерія» має свою специфіку, яка пов'язана з тим, що сфера застосування методів машинного навчання з підкріпленням постійно розширюється. Повсюдна інформатизація приводить до накопичення величезних обсягів даних у наукових дослідженнях, виробництві, на транспорті, в охороні здоров'я. Задачі прогнозування, управління та прийняття рішень часто потребують машинного аналізу даних та машинного навчання, оскільки раніше такі задачі або не ставились взагалі, або вирішувались точними методами, які вимагали великих обчислювальних потужностей через складність відповідних алгоритмів

Умова зарахування додаткових балів.

В рамках вивчення навчальної дисципліни «Методи навчання з підкріпленням» допускається зарахування балів, одержаних в результаті дистанційних курсів на платформі "Coursera", за умови попереднього погодження програми даного курсу з викладачем та за умови отримання офіційного сертифікату.

Робочу програму навчальної дисципліни (силабус):

Складено , д.т.н, професор, Новотарський Михайло Анатолійович

Ухвалено кафедрою обчислювальної техніки (протокол № 18 від 25.05.2021)

Погоджено Методичною комісією ФІОТ (протокол № 10 від 14.06.2021)