МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Навчально-наукового інституту атомної і теплової енергетики
Кафедра інженерії програмного забезпечення в енергетиці

МЕТОДИЧНІ ВКАЗІВКИ ДО ВИКОНАННЯ ЛАБОРАТОРНОЇ РОБОТИ №4

з дисципліни

«МЕТОДОЛОГІЇ РОЗРОБКИ ІНТЕЛЕКТУАЛЬНИХ КОМП'ЮТЕРНИХ ПРОГРАМ»

Тема: «Інтелектуальні агенти. Алгоритм Q-навчання»

КИЇВ 2023

Тема: Інтелектуальні агенти. Алгоритм Q-навчання

Мета: Ознайомитися з поняттям інтелектуального агента та одним з методів його навчання – Q-learning

Теоретичні відомості

Інтелектуальний агент (у штучному інтелекті) – сутність, яка одержує інформацію через систему сенсорів про стан зовнішнього середовища (керованих нею процесів) та здійснює вплив на нього, при цьому її реакції раціональні в тому розумінні, що її дії сприяють досягненню певної мети. Найбільш близьким аналогом у живій природі є примітивне інстинктивне поводження комах. Термін «інтелектуальний» не означає наявності інтелекту, але підкреслює більш високий рівень технології керування в порівнянні із примітивними тригерними системами автоматичного керування. Такий агент може бути як програмною системою, так і складною автоматизованоюсистемою, наприклад, комплексом керування технологічними, логістичними, фінансовими або будь-якими іншими процесами. Про "інтелектуальність" агента можна говорити, якщо його взаємодія з навколишнім середовищем є адекватною тій або іншій системі вимог. Ніякого відношення навіть до інтелекту вищих тварин і вже тим більше людини подібна функціональність не має.

У штучному інтелекті існує декілька типів агентів. Наприклад:

1) фізичний агент – агент, що сприймає навколишній світ через деякі сенсори й діє за допомогою маніпуляторів.

2) часовий агент – агент, що використовує інформацію, яка змінюється з ходом часу, і пропонує деякі дії або надає дані комп'ютерній програмі чи людині.

Бувають різні типи інтелектуальних агентів. Найпростіші з них діють по схемі:

IF (умова) THEN дія;

їх дії жорстко прописані.

Більш складні агенти можуть мати інформацію про зовнішнє середовище та принципи його функціонування, знати ступінь корисності тої чи іншої своєї дії, навчатися на основі інформації, одержуваної із зовнішнього середовища.

Агенти, що навчаються

У деякій літературі агенти, що навчаються (АН) *називаються автономними інтелектуальними* агентами (англ. *autonomous intelligent agents*), що підкреслює їхню незалежність і здатність до навчання й пристосовування до мінливих обставин.

Алгоритм Q-навчання інтелектуальних агентів

Q-навчання (Q-learning) - метод навчання інтелектуального агента, заснований на системі підкріплень (винагород від зовнішнього середовища при виборі вірних дій).

Спочатку агент випадковим чином вибирає свою поведінку, після чого отримує реакцію від середовища, на основі якої формує функцію корисності Q своїх дій. Функція Q дає йому можливість надалі уже не випадково вибирати стратегію поведінки, а враховувати досвід попередньої взаємодії із середовищем.

Одна з переваг Q-навчання – воно дає можливість порівнювати очікувану корисність доступних дій, не формуючи моделі навколишнього середовища.

Даний алгоритм оперує поняттями **дія** й **стан** інтелектуального агента.

У розпорядженні агента є деяка множина **дій** A(a_1, a_2 ... a_n). Дії агента впливають на середовище, і агент може визначати, у якому стані він перебуває в поточний момент, і одержувати винагороду від середовища за правильні дії.

Зовнішнє середовище представлене множиною можливих **станів** S(s_1, s_2 ... s_n), в яких може перебувати агент.

Цільовий стан агента - це такий стан, досягнення якого є кінцевою метою діяльності агента.

Завданням агента є знайти найкращу стратегію для досягнення **цільового стану**. В даному алгоритмі вона описується Q-значеннями, які визначають корисність виконуваної дії у відповідному стані.

Суть алгоритму Q-learning

Матриця R – містить дані про зовнішнє середовище. Матриця R представляє собою матрицю суміжності графа, в якому вершини – це можливі стани агента, а ребра – дії агента, які переводять його з одного стану в інший.

Матриця Q – пам’ять агента, що містить інформацію про корисність його дій. Це двомірна матриця, в якій кожний стан агента співставлений зпевною величиною винагороди, яку одержить агент за перехід у даний стан.

При ініціалізації програми матриця Q заповнюються нулями, якщо агент не має жодних знань про навколишнє середовище, або даними з матриці R, якщо початкові знання у нього є. Не маючи інформації про винагороди від того чи іншого стану агент обирає першу дію випадковим чином.

Формула оновлення матриці Q після кожної дії агента:

Q[s, a] = R[s, a] + Gamma * MAX(Q[s', a']),

де Q[s, a] – комірка матриці Q, що відповідає поточному стану агента;

R[s, a] – комірка матриці R, що відповідає поточному стану агента;

Gamma – швидкість навчання, рекомендоване значення = 0.8;

Q[s', a'] – комірка матриці Q, що відповідає наступному стану агента;

MAX(Q[s', a']) – вибір з множини можливих дій агента в поточному стані дії з максимальною винагородою.

Приклад 1. Дано схему будинку (рисунок 1). Інтелектуальний агент - робот, що повинен навчитися виходити на вулицю з будь-якої кімнати і бажано найкоротшою відстанню. Отримати знання про розташування кімнат та дверей він може лише на практиці, блукаючи по будинку.

Припустимо, що у нас є будинок з 5 кімнатами, які з'єднані дверима, як показано на рисунку 1. Пронумеруємо кожну кімнату від 0 до 4, а вулиці привласнимо номер 5.

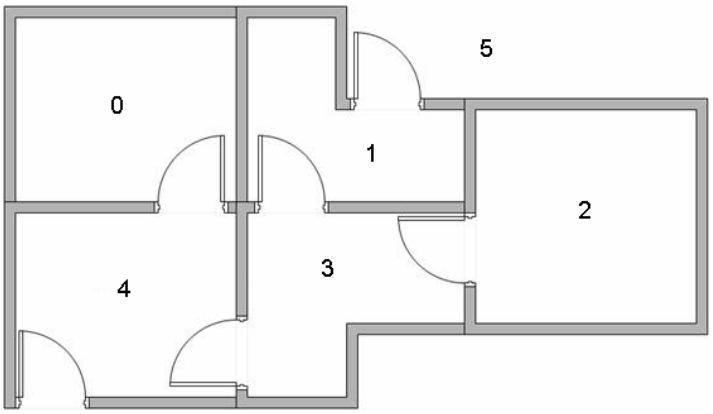


Рисунок 1

В даному прикладі станом агента буде перебування його в тій чи іншій кімнаті, або вулиці, а дією - вихід/вхід у ті чи інші двері.

Тобто в агента є 6 можливих станів:

0 стан - перебування у кімнаті №0;

1 стан - перебування у кімнаті №1;

2 стан - перебування у кімнаті №2;

3 стан - перебування у кімнаті №3;

4 стан - перебування у кімнаті №4;

5 стан - перебування на вулиці.

Цільовий стан - 5 стан (перебування на вулиці).

Представимо зовнішнє середовище агента у вигляді графа, в якому стани - вузли, а дії - ребра (рисунок 2).

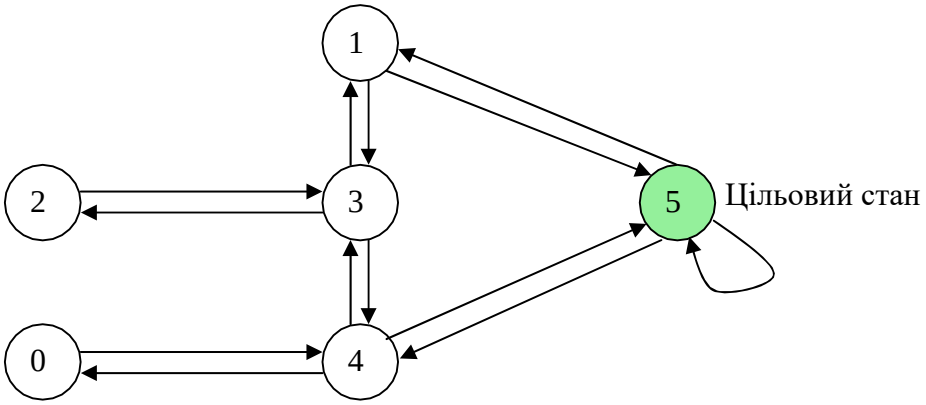


Рисунок 2

В якій би кімнаті не опинився агент, він повинен знайти вихід за межі будівлі. А якщо агент вже знаходиться на вулиці, то повинен залишатися там - тому 5-та вершина має петлю (рисунок 2).

Привласнимо кожному ребру вагу. Ребра, що ведуть до цільового стану, мають одержати вагу 100, всі інші - вагу 0 (рисунок 3).

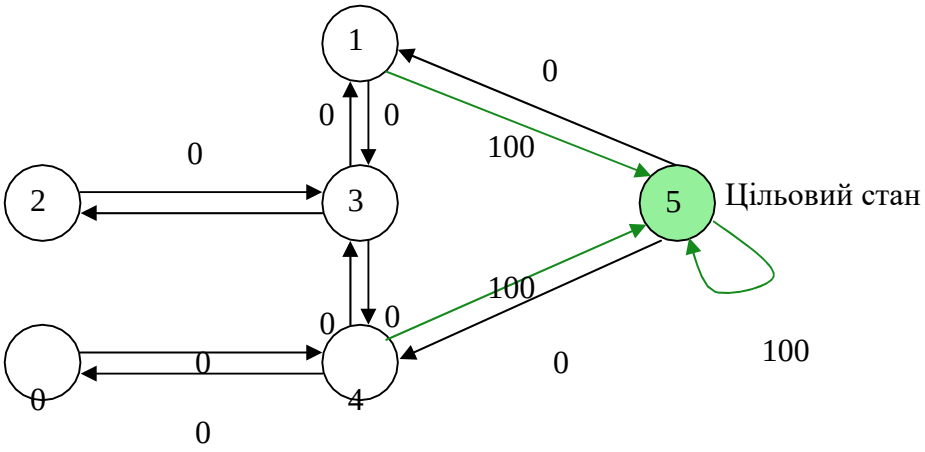


Рисунок 3

Тепер побудуємо матрицю суміжності даного графа (рисунок 4). Якщо вершини не зв’язані ребром, ставимо біля них у матриці (-1). Одержана матриця і є R матрицею.

		Стан					
		0	1	2	3	4	**5**
Дія	0	-1	-1	-1	-1	0	-1
	1	-1	-1	-1	0	-1	100
	2	-1	-1	-1	0	-1	-1
	3	-1	0	0	-1	0	-1
	4	0	-1	-1	0	-1	100
	5	-1	0	-1	-1	0	100

Рисунок 4

Індекси рядків R матриці вказують на номер стану агента, а індекси стовпчиків - на номер дії.

Розглянемо покроково одну з можливих послідовностей дій агента.

Ініціалізація.

1) Ініціалізація матриці R:

R=

	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

Рисунок 5

2) Ініціалізація матриці Q (для наочності припустимо, що пам’ять агента вже містить записи про розташування цільового стану):

Q =

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	100
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	100
5	0	0	0	0	0	100

Рисунок 6

3) Ініціалізація параметра швидкості навчання Gamma = 0,8.

1 крок. Агент випадковим чином опинився у кімнаті №3.

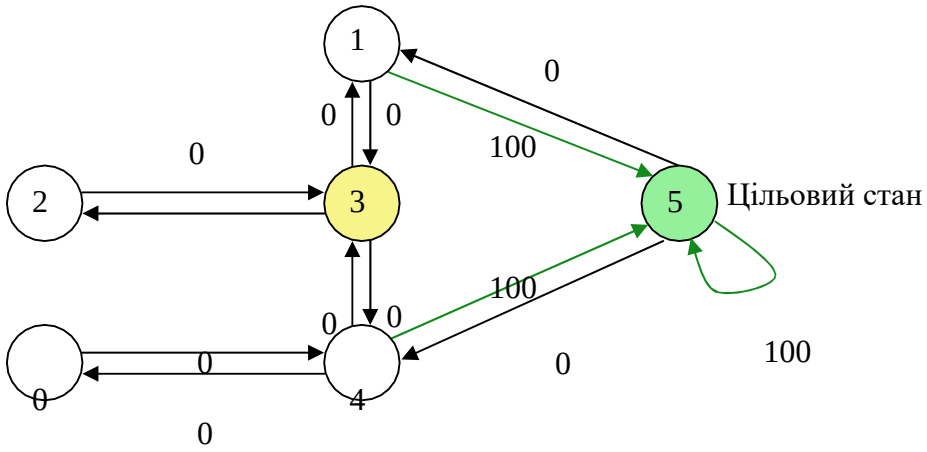


Рисунок 7

З матриці R агент дізнається про свої можливі дії. Оскільки агент перебуває в 2 стані, то його можливі дії містяться в 3-му рядку матриці. Зі стану 3 агент може перейти в стани: 1, 2 або 4, тобто здійснити дії (3, 1), (3, 2) або (3, 4). Оскільки агент ще не знає, який стан наблизить його до цільового стану, він обирає дію випадково. Припустимо, що агент випадковим чином обрав дію (3, 1).

R=

	0	1	2	3	4	**5**
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	**0**	**0**	-1	**0**	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

Рисунок 8

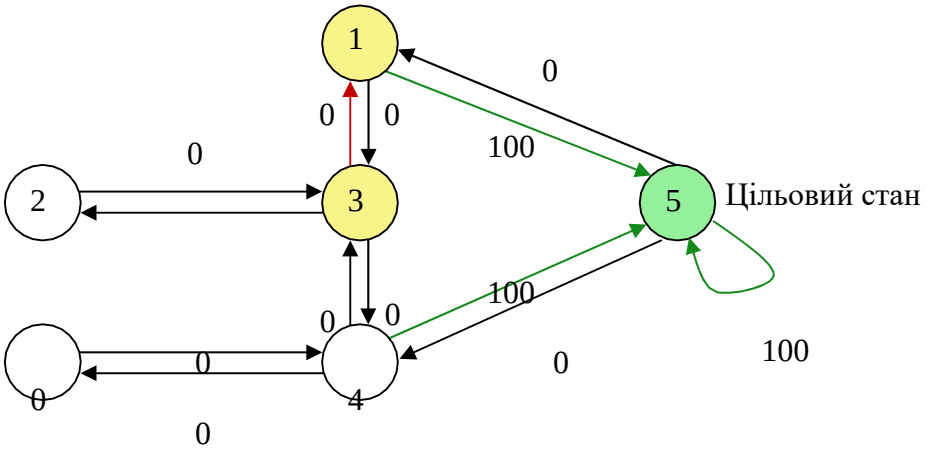


Рисунок 9

2 крок. На другому кроці агент обирає наступну дію та одержує винагороду за попередню дію.

Зараз агент перебуває в 1 стані. З даного стану він може перейти в стан 3 та стан 5. Оскільки стан п’ять відмічений вагою 100, то агент розуміє, що даний стан є цільовим. Отже агент видирає перехід у цільовий стан, тобто дію (1, 5).

R=

	0	1	2	3	4	**5**
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	**0**	-1	**100**
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

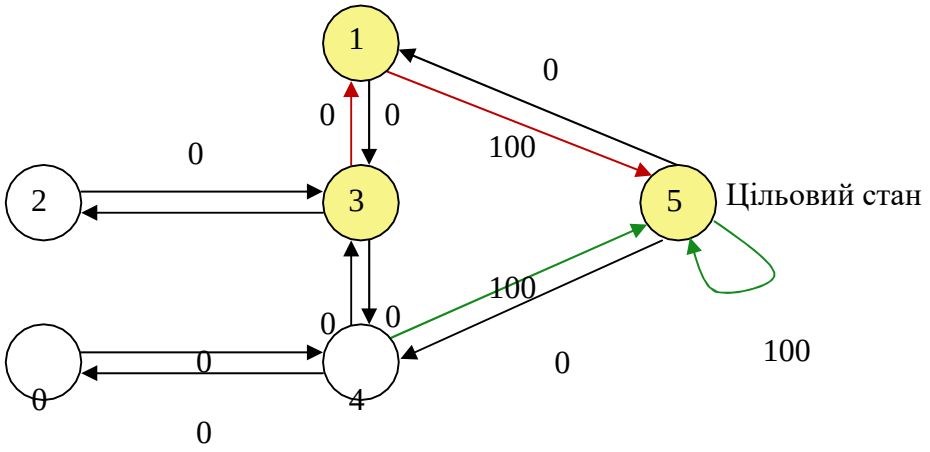


Рисунок 11

Винагорода за попередню дію (здійснену на кроці 1) обчислюється по наступній формулі:

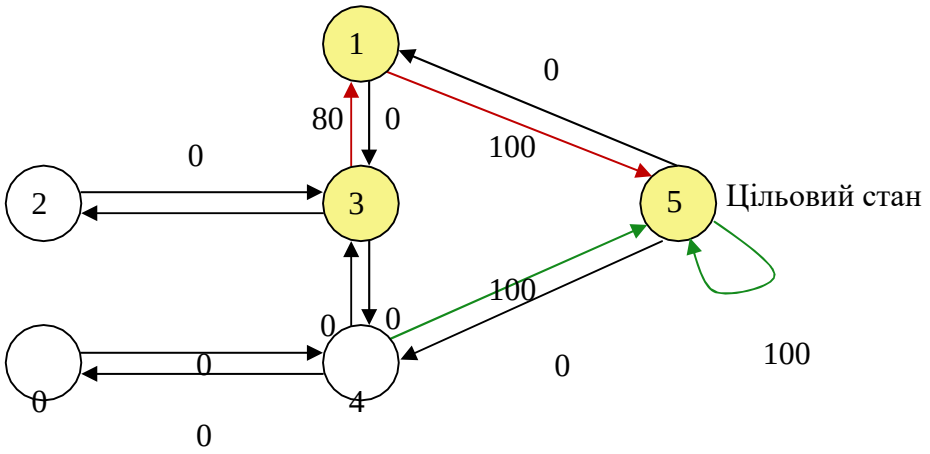
Q[3, 1] = R[3, 1] + 0,8 * MAX(Q[1, 3], Q [1, 5]) = 0 + 0,8 * 100 = 80

В матрицю Q, тобто пам’ять агента записується інформація про корисність дії (3, 1).

Q =

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	**80**	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

Рисунок 12



3 крок. На другому кроці агент знову обирає наступну дію та одержує винагороду за попередню дію.

Агент перебуває в 5 стані. Можливі дії - (5, 5), (5, 1) , (5, 4). Агент обирає дію (5, 5) як цільову дію.

Винагорода за попередню дію (здійснену на кроці 2) обчислюється по наступній формулі:

Q[1, 5] = R[1, 5] + 0,8 * MAX(Q[5, 5], Q [5, 1], Q [5, 4]) = 100 + 0,8 * 100 = 180

Q =

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	**180**
2	0	0	0	0	0	0
3	0	80	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

Рисунок 14

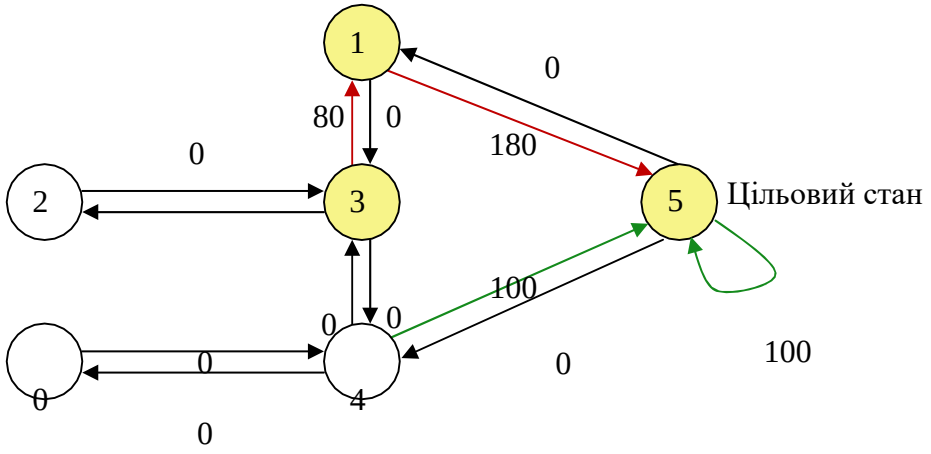


Рисунок 15

4 крок. Обчислимо винагороду за перебування в 5 стані та завершимо виконання алгоритму.

Q[5, 5] = R[5, 5] + 0,8 * MAX(Q[5, 5], Q [5, 1], Q [5, 4]) = 100 + 0,8 * 100 = 180

Q =

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	180
2	0	0	0	0	0	0
3	0	80	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	**180**

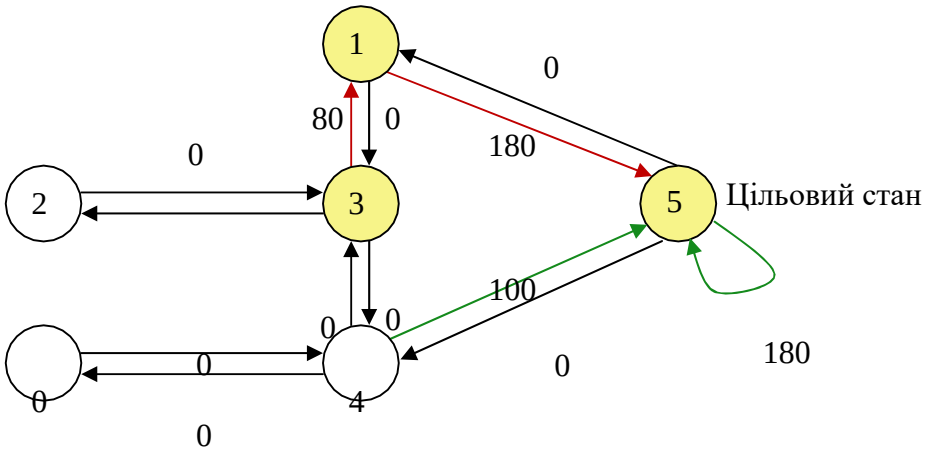


Рисунок 17

Тепер у пам’яті агента міститься інформація про корисність дій (3, 1), (1, 5) та (5, 5).

Чим більше разів агент буде шукати з різних кімнат вихід на вулицю, тим більше інформації запишеться в його пам’ять.

Наприклад, якщо агент здійснить 6 успішних спроб вийти з будинку з різних кімнат:

1 спроба -1 стан, 5 стан;

2 спроба - 3 стан, 4 стан, 5 стан;

3 спроба - 5 стан, 5 стан;

4 спроба - 2 стан, 0 стан, 4 стан, 5 стан;

5 спроба - 4 стан, 5 стан;

6 спроба - 0 стан, 4 стан, 5 стан.

То його пам’ять виглядатиме наступним чином:

Q =

	0	1	2	3	4	**5**
0	0	0	0	0	144	0
1	0	0	0	247	0	386
2	0	0	0	0	0	0
3	0	221	0	0	309	0
4	115	0	0	115	0	386
5	0	309	0	0	309	386

Рисунок 18

Завдання 1

Відповідно до свого варіанта **(див. Додаток А)** побудувати математичну модель інтелектуального агента та його зовнішнього середовища:

– записати множину станів інтелектуального агента та цільовий стан;

– зобразити зовнішнє середовище агента у вигляді графа, в якому стани – вузли, а дії – ребра;

– записати матрицю суміжності побудованого графа.

Проілюструвати роботу алгоритму Q-навчання розрахунками без програмної реалізації. Для цього показати 2 спроби агента досягти цільової мети. Перший раз, обираючи маршрут цілком довільним чином (так як пам'ять агента порожня), другий раз – враховуючи здобуту пам'ять агента. Розрахунки повинні містити обчислення винагород за дії агента, значення матриці Q та зображення графа зовнішнього середовища з позначенням шляху агента і змінених значень ваг ребер.

Завдання 2

Відповідно до свого варіанта **(див. Додаток А)** та побудованої в 1 завданні математичної моделі реалізувати програмне забезпечення для ілюстрації роботи алгоритму Q-навчання. Примітка: на початку програми ініціалізувати матрицю Q нулями.

Перед захистом звіт з лабораторної роботи надсилається на почту pis2020@ukr.net Тема листа «Група_Прізвище_ЛР №3» наприклад ТІ-01_Петренко І.І._ЛР_№4. Назва файлу «Група_Прізвище_ЛР №4» наприклад ТІ-01_Петренко І.І._ЛР_№4

Крайній термін захисту ЛР для ТІ-01, ТІ-02 15.05.2023 (для групи ТВ-01 17.05.2023). Теоретичні питання стосуватимуться лекцій №6-9. Якщо захист лабораторної роботи відбудеться пізніше оговореної дати, то оцінка знижується на 50 %.

Додаток А. Варіанти завдання

Варіант 1

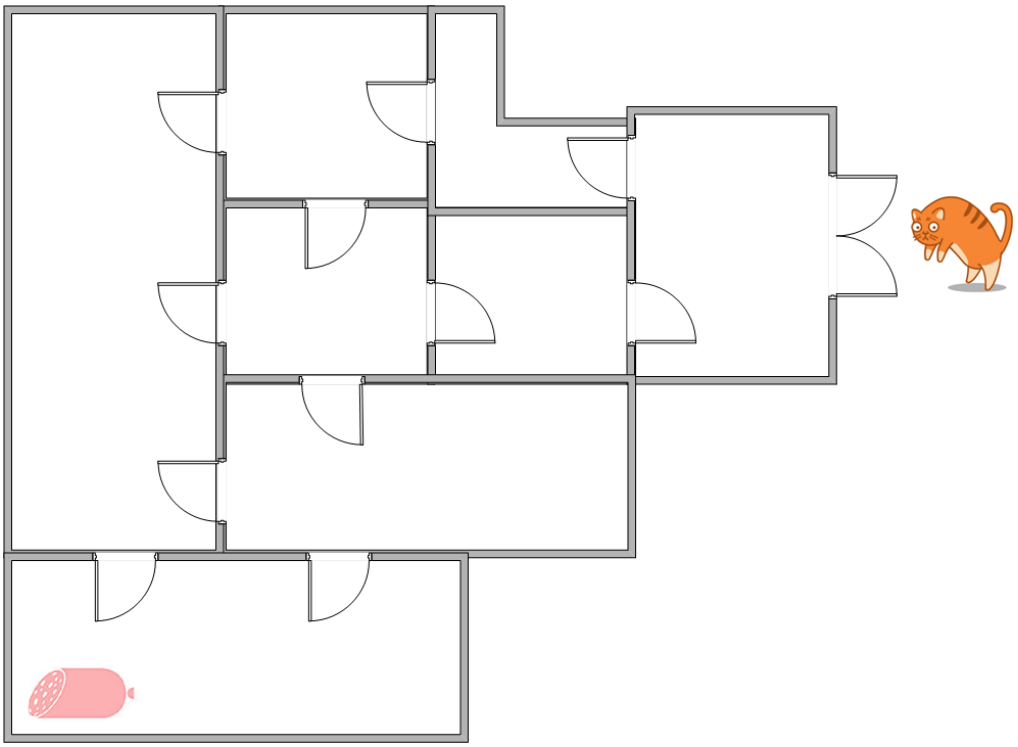


Рисунок А1

Варіант 2

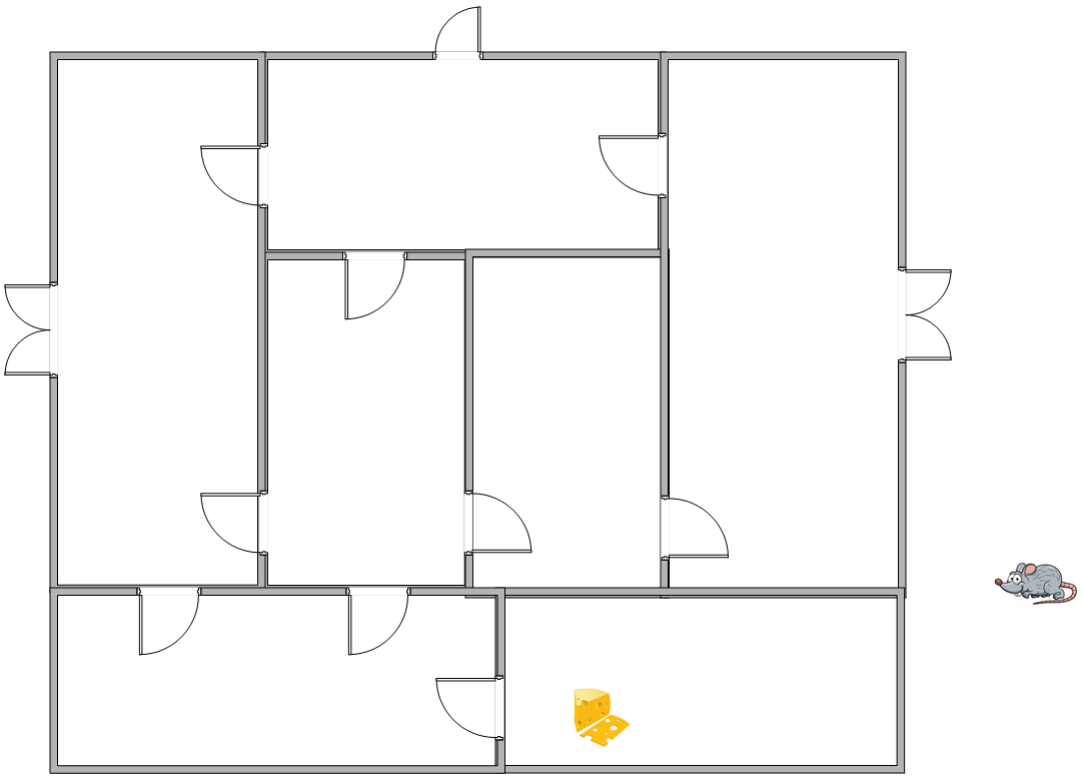


Рисунок А2

Варіант 3

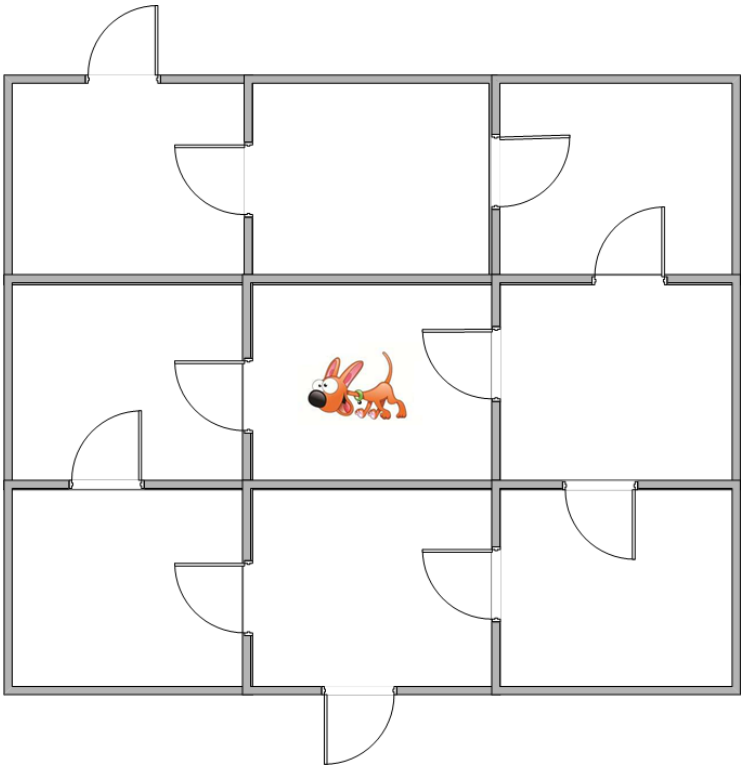


Рисунок А3

Ціль інтелектуального агента - вийти на вулицю.

Варіант 4

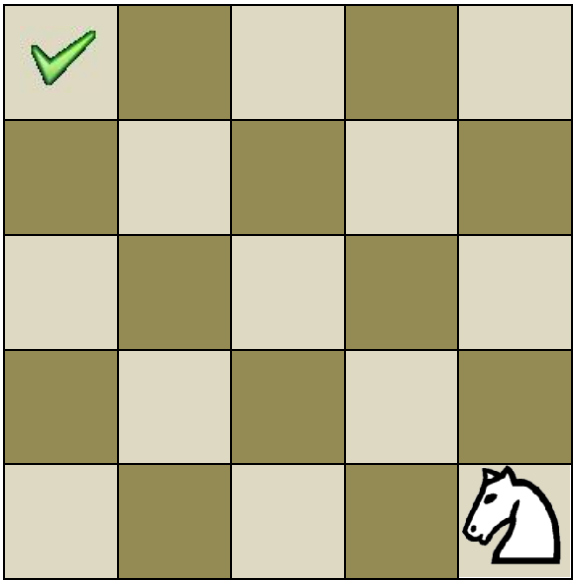


Рисунок А4

Агент може пересуватися по клітинках за правилами гри в шахи (агент - фігура кінь).

Варіант 5

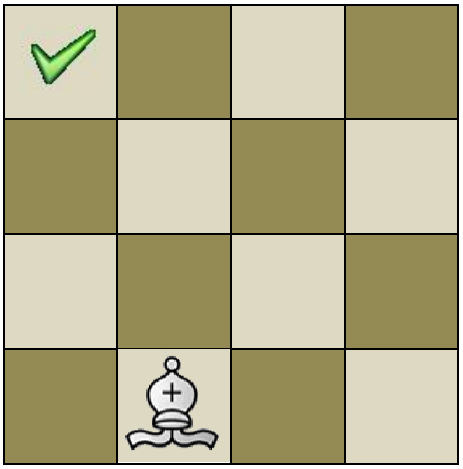
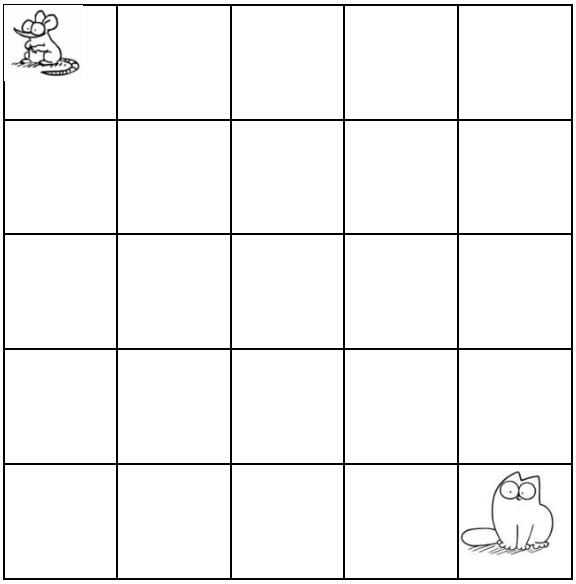


Рисунок А5

Агент може пересуватися по клітинках за правилами гри в шахи (агент - фігура слон).

Варіант 6



Агент може пересуватися по клітинках наступним чином:

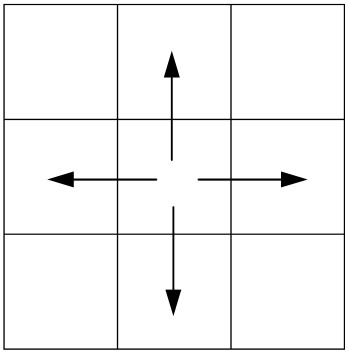
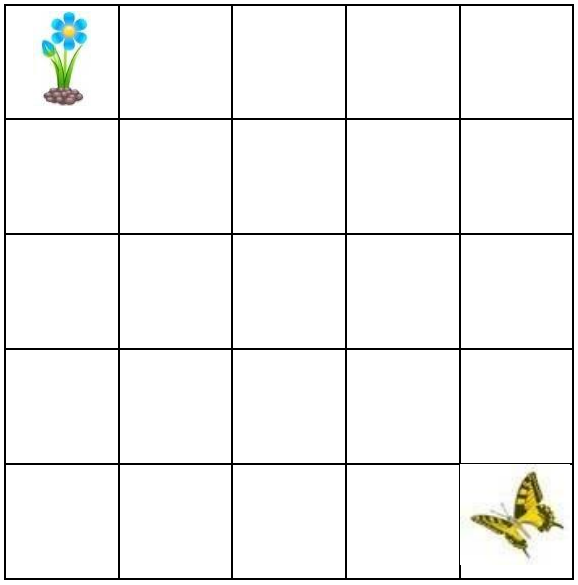


Рисунок А6

Варіант 7



Агент може пересуватися по клітинках наступним чином:

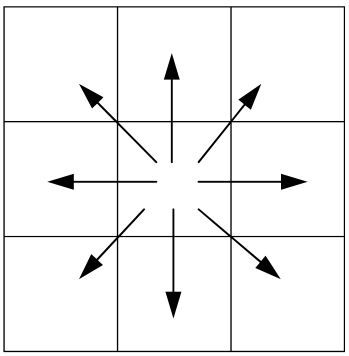
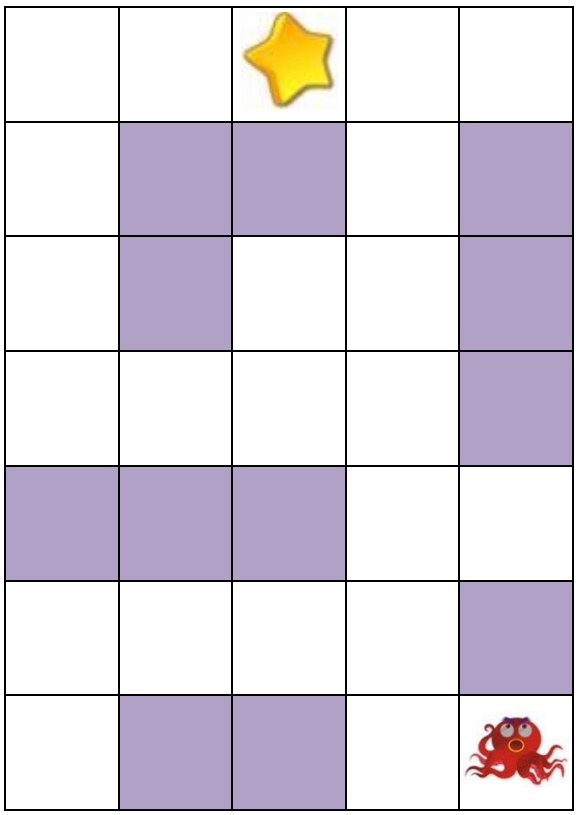


Рисунок А7

Варіант 8



Клітинки фіолетового кольору - перешкоди.

Агент може пересуватися по клітинках наступним чином:

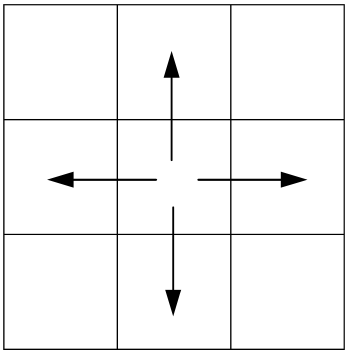
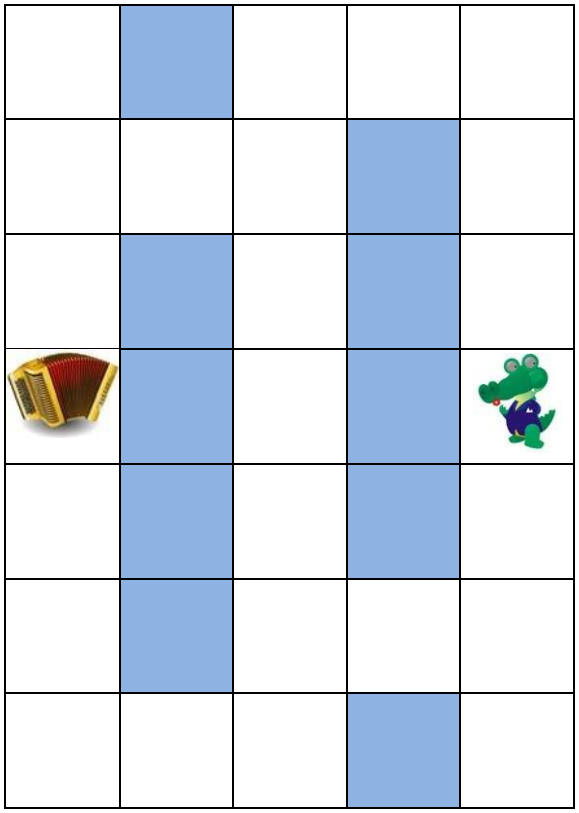


Рисунок А8

Варіант 9



Клітинки синього кольору - перешкоди.

Агент може пересуватися по клітинках наступним чином:

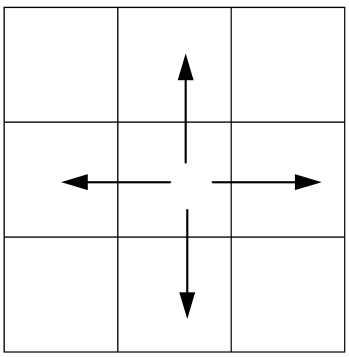
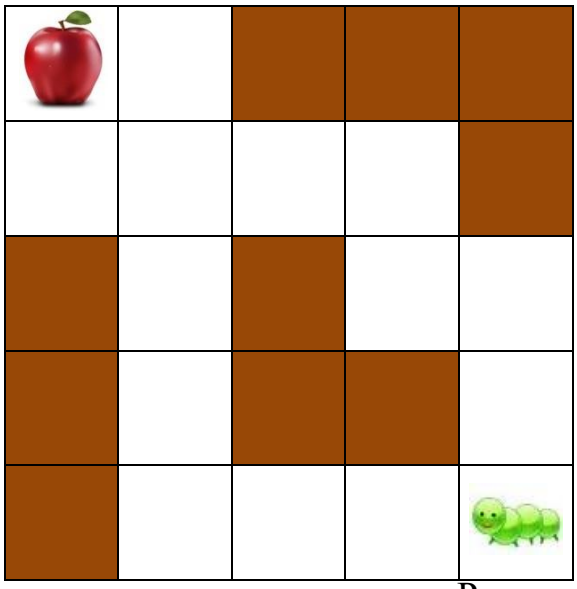


Рисунок А9

Варіант 10



Клітинки коричневого кольору - перешкоди. Агент може пересуватися по клітинках наступним чином:

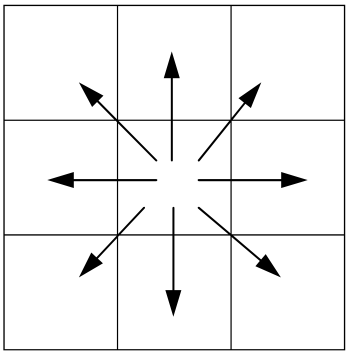


Рисунок А5

Контрольні питання:

1. Що таке інтелектуальний агент?
2. Які бувають інтелектуальні агенти?
3. У чому полягає алгоритм Q-навчання?
4. Для чого застосовується матриця R в алгоритмі Q-навчання?
5. Для чого застосовується матриця Q в алгоритмі Q-навчання?
6. Яким чином представляється зовнішнє середовище інтелектуального агента в алгоритмі Q-навчання?

Професор кафедри ІПЗЕ
доктор технічних наук, доцент Андрій МУСІЄНКО