

ЗАСТОСУВАННЯ АДАПТОВАНОГО АЛГОРИТМУ UTILITY AI ДЛЯ ГРУПОВОЇ ВЗАЄМОДІЇ ІНТЕЛЕКТУАЛЬНИХ АГЕНТІВ БЕЗ ОБМІНУ ІНФОРМАЦІЄЮ

Анотація: З метою вирішення завдання групової взаємодії інтелектуальних агентів в умовах відсутності комунікації, застосовано адаптований до середовища змагань RoboCup алгоритм, а саме, Utility AI. Описані та проаналізовані відомі рішення на дану тематику. Зроблено огляд оригінального Utility AI алгоритму та наданий опис модифікованого варіанту. Проведено поетапне тестування ігрового сценарію з трасуванням проміжних даних на прикладі гри в футбол. Представлено ключові елементи логіки прийняття рішення, інтерпретації навколишнього середовища, а також архітектуру інтелектуального агента.

Ключові слова: групова взаємодія, взаємодія без комунікації, інтелектуальний агент, Utility AI, середовище моделювання, моделі поведінки, методи прийняття рішень.

Вступ

Застосування груп роботів для вирішення спільних задач стає все більш поширеним. Для розвитку даного напрямку технологій, в 1997 році були засновані змагання по футболу – RoboCup [1]. Команди розробників здебільшого базують алгоритми взаємодії інтелектуальних агентів з урахуванням стійкої комунікації. Проте, існує окремий вид змагань під назвою «Drop-in player challenge» [2], що дає можливість застосовувати різних роботів в одній команді. Науковці з дослідницьких інститутів працюють над проблемою взаємодії різного роду інтелектуальних агентів, оскільки збільшення різноманіття спричиняє потребу в пошуку нових методів синхронізації. Та ж тенденція зберігається і для алгоритмів взаємодії в умовах відсутності зв'язку.

Існує багато різних підходів до процесу прийняття рішень інтелектуальним агентом, але одним із найпростіших і, в той же час, швидких, є Utility AI алгоритм на базі функції корисності. Даний підхід може бути використана для взаємодії в умовах відсутності зв'язку, адже дозволяє гнучко налаштовувати стратегію гри агента відповідно до гри союзників.

Метод Utility AI

Загальна концепція такого підходу полягає в тому, що кожна можлива дія оцінюється одночасно. Серед них вибирається та, яка набрала найбільшу кількість балів. Рішення може прийматися як для фактичної ситуації, так із застосуванням

коефіцієнтів ймовірностей для передбачення найкращого результату. Очікувана корисність EU (Expected Utility) визначається як [3]:

$$EU = \sum_{i=1}^n D_i P_i, \quad (1)$$

де D – корисність критерію; P – ймовірність того, що критерій матиме таке значення, n – кількість критеріїв, i – індекс критерію.

Зазвичай, дані для прийняття рішення є не однорідними і потребують нормалізації для згортання у загальній функції. Для вирішення даної проблеми в алгоритмі застосовуються формули для обчислення нормалізованої корисності (U), які можуть виражати лінійну залежність ($U = \frac{x}{m}$) (x – вхідне значення, m – максимальне значення), степеневу ($U = (\frac{x}{m})^k$), логістичну ($U = \frac{1}{1+e^{-x}}$) та інші. Для визначення загальної корисності певної дії, значення критеріїв (нормалізовані дані) можуть бути усереднені з урахуванням від'ємних значень, наприклад, для негативних факторів.

Перевагою алгоритму є широкий спектр параметрів моделі навколишнього світу. Саме тому алгоритм Utility AI використовується не тільки для того, щоб досягти кінцевої мети, а й зробити це оптимальним шляхом за певним критерієм. Ця особливість є однією з відмінностей цільових агентів від агентів на основі корисності.

У даному випадку, беруться до уваги не тільки критерії, що зв'язані без посередньо з метою, а й ті критерії, які дозволяють бути корисним у складі команди. Дана властивість алгоритму дає перспективу розвитку групової взаємодії в умовах неможливості комунікації.

Огляд існуючих методів командної взаємодії

Серед змагань, «Drop-in player challenge» [2] є найближчим по умовах до задачі взаємодії без комунікації. У даному типі змагань кожна команда програмує свого власного робота для координації з невідомими союзниками по команді. Команди просять не синхронізувати роботів завчасно задля того, щоб під час гри агенти брали участь у спеціальній командній роботі і аналізували, яким чином можуть бути корисні гравцям своєї команди. Та попри це, базова апаратна синхронізація та канали зв'язку передбачені. Головна мета кожного агента – виграти футбольний матч і отримати статус «найкращого гравця по команді».

Більш дотичним підходом є розробка науковців Університету Іллінойсу [4]. Запропонований метод використовує машинне навчання для вирішення синхронізації в умовах відсутності зв'язку та базується на алгоритмах навчання з підкріпленням. В дослідженні застосовуються локальні функції корисності, які забезпечують координацію під час навчання децентралізованих агентів. Функція сигналізує агенту, коли дія несе користь для команди і навпаки. Даний принцип використовується для відокремлення окремого впливу агента на глобальну функцію корисності від впливу всіх інших агентів.

Галузями застосування даного підходу є: військове спостереження, робота на складі, автономні транспортні засоби, керування електромережою тощо. У завданнях, які вирішує «іллійський метод», відсутні елементи гри. Тобто, він не придатний до вирішення протиборчих задач.

Також, можна зазначити певні труднощі в реалізації:

- реалізація методу вимагає значних об'ємів даних та обчислювальних ресурсів;
- внесення коректив поведінки потребує перенавчання, що вимагає часу та відповідних ресурсів.

Інший підхід до взаємодії агентів в умовах відсутності комунікації представлено в публікації [5]. Даний метод застосовується для підтримання ігрового процесу у випадку розриву комунікації. Розробники пропонують жорстко фіксувати ігрову поведінку та зони, в яких діє інтелектуальний агент (рис.1).

Attack Right 3-5-2	Agent number										
	1	2	3	4	5	6	7	8	9	10	11
pass to	2	5	4	2	6	11	9	10	0	11	0
Can Dribble	0	1	0	0	1	1	0	1	1	1	1
Dribble Range	0.0	3.0	0.0	0.0	2.0	2.0	0.0	1.0	0.5	0.5	0.5
Max X	0	32	-32	-32	-32	-32	15	25	-15	-15	-15
Max Y	0	32	-32	32	-32	-32	-15	25	-45	-25	15
Min X	-2	3	-3	1	-2	0	-2	-2	-2	0	1
Min Y	-1	1	-3	2	-1	2	-4	3	2	3	1

Рис. 1. Шаблони ігрових поведінок

На рис. 1 наведено сценарій атаки правою частиною поля. В нього включені такі навички:

- *Pass to*: схема передачі пасу між гравцями;
- *Can dribble*: здатність до дриблінгу;
- *Dribble range*: мінімальна відстань до суперника, коли агент може вести м'яч;
- *MaxY & MinY, MaxX & MinX*: якщо агент, який володіє м'ячем, повинен обійти поточного агента, він повинен перейти до області, координати якої вказані.

Основним недоліком підходу є його принципова жорсткість, що робить метод не універсальним. Також, при зміні параметрів середовища, агент вимагає зміни всіх конфігурацій.

Адаптація та застосування Utility AI в командній взаємодії

За класифікацією ФІФА існує три типи тактик: командна, групова та індивідуальна. Наразі **Utility AI** реалізується саме для групової взаємодії.

Адаптована форма алгоритму використовується разом з деревом прийняття рішень. Спочатку використовується Utility AI для прийняття рішення на рівні групової тактики, а потім застосовується дерево прийняття рішень для індивідуальної тактики.

Модифікований метод Utility AI використовується для аналізу поточної ситуації. Тобто $P = 1$ (1). Прогнозування майбутнього стану середовища відбувається тільки на рівні індивідуальної тактики для окремих сценаріїв (наприклад прораховування майбутньої позиції м'яча для перехвату).

Задля реалізації саме командної поведінки в методі, введені сталі значення корисності стратегій (табл. 1). Оскільки, константи не можуть бути обчислені через суб'єктивну природу, використовуються дані надані експертом. Поточна корисність CU (Current utility) визначається як:

$$CU = C + \sum_{i=1}^n D_i P_i, \quad (2)$$

де $0 \leq CU \leq 1$; $0 \leq C \leq 0.9$; $0 \leq \sum_{i=1}^n D_i P_i \leq 0.1$; $P = 1$.

Змінна C є найбільшим значенням зі сталих критеріїв D_i (табл. 1), які аналізуються для конкретної дії.

Беручи до уваги особливості алгоритму Utility AI, експерт може з легкістю змінювати поведінку інтелектуального агенту. Таким чином, константний кістяк створює суб'єктивний поведінковий шаблон.

Таблиця 1. Константи корисності стратегії

Стратегія		Умови	Корисність
А	Зайняття позиції на фланзі	Для флангового гравця на власній половині поля	0.7
		Для флангового гравця	0.888
Б	Ведення м'яча	Всі гравці окрім голкіпера та гравців в штрафних майданчиках	0.85
В	Захват м'яча і виконання удару по ворожим воротам	Для нападаючого	0.87
		Для будь якого гравця, якщо він знаходиться в радіусі R метрів	0.88
Г	Зайняття позиції між м'ячем і власними ворітьми	Для захисника та воротаря, якщо відстань до м'яча більше R метрів	0.87
		Для захисника та воротаря, якщо відстань менше R метрів	0.43
		Інші гравці	0.4
Д	Вибивання м'яча за межі ігрового поля	Для будь якого гравця, якщо м'яч знаходиться в зоні власних воріт	0.9
		Для будь якого гравця, якщо знаходиться поза межами воріт	0.43
Е	Самолокалізація гравця	Для гравців, які не самолокалізувалися.	0.89

* Для не зазначених випадків або ролей гравців константа корисності становить 0.0.

Коефіцієнтний каркас командної тактики компонується з обчисленням даних на рівні індивідуальної тактики. Таким чином, як це показано на (рис. 2), спершу застосовується грубий аналіз Utility AI, наприклад, таких даних, як роль гравця, позиція м'яча та воріт, після чого використовується індивідуальна тактика, яка визначає, як краще дістатися до м'яча, як захопити, як краще вдарити тощо.

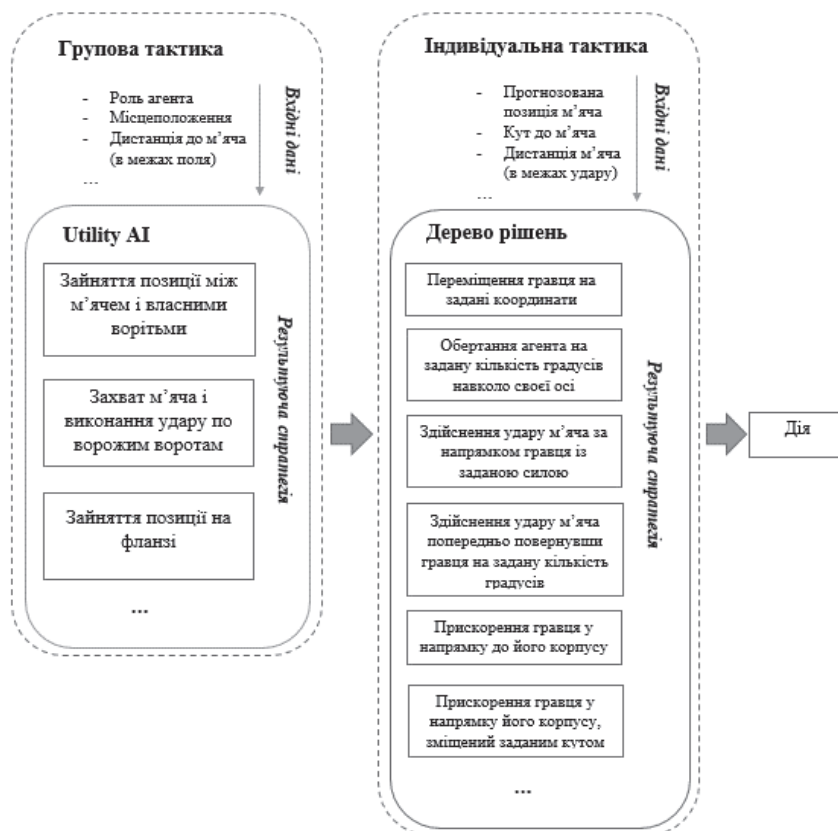


Рис. 2. Застосування Utility AI

Оптимальна групова тактика агента в i -й такт обирається за принципом максимальної корисності дії [6, с.166-167]:

$$W_k(X, Q) = \{x \in X \mid u(Q(x)) = \max_{y \in X} u(Q(y))\}, \quad (3)$$

де X – множина всіх стратегій; x – множина стратегій серед яких відбувається вибір; $u(Q(y))$ – функція корисності дії. Прийняття рішення щодо індивідуальної тактики здійснюється за допомогою дерева рішень.

Тестування результатів

Для оцінки роботи методу прийняття рішень та розроблених базових алгоритмів поведінки інтелектуального агента [7], зроблено тестовий пуск та наведено результати

роботи для тестового сценарію. Для випробувань обрано гравця з роллю нападаючого лівого флангу. Дані надходять з серверу, який симулює навколишнє середовище [8]. Для спрощення прикладу в описі будуть пропущені деталі обчислення $\sum_{i=1}^n D_i P_i$.

Сценарій матиме наступні етапи (E):

- E1. Ініціалізація (гравець команди починає гру зі спеціалізованого стартового місця);
- E2. Заняття позиції в формуванні перед грою;
- E3. Гра;
- E4. Виконання цілі тестового сценарію (Гол).

Центр поля відповідає точці з координатами (0,0). Решта параметрів, за правилами кубку RoboCup [8] мають наступні значення:

- довжина поля 104 м;
- ширина поля 68 м;
- ширина штрафного майданчика становить $S_x = 16.5$ м;
- довжина штрафного майданчика $S_y = 40.3$ м;

Етапи гри з проміжними даними для роботи методу прийняття рішень (МІР) описані в таблиці 2.

Таблиця 2. Трасування окремих тактів модифікованого Utility AI на прикладі «лівого нападаючого»

Такт t (сеанс обміну інформацією із сервером)	Етап гри	Позиція гравця	Позиція м'яча на полі	Візуальний контакт з воротями суперника	Зона пенальті	Дистанція до м'яча	Візуальний контакт з м'ячем	Можливість здійснити удар	Можливість рухатися на позицію у формуванні	Гравець на стартовій позиції	Вибрана стратегія
1	2	3	4	5	6	7	8	9	10	11	12
1	E1. Ініціалізація	X=-2.9 Y=-37.1	X=NaN Y=NaN	Так	Ні	-1	Ні	Ні	Так	Ні	Перед стартом гри
31	E2. Заняття позиції в формуванні	X=-21.5 Y=-10.4	X=2.8 Y=-5.2	Ні	Ні	24.8	Так	Ні	Ні	Так	А. Стратегія заняття позиції на фланзі

Закінчення табл. 2

1	2	3	4	5	6	7	8	9	10	11	12
83	Е3. Гра	X=-8.6 Y=-7.1	X=0.5 Y=0.9	Так	Ні	12.1	Так	Ні	Ні	Ні	В. Стратегія захоплення м'яча і виконання удару в сторону ворожих воріт
123	Е3. Гра	X=0.1 Y=0.2	X=-0.2 Y=-0.3	Ні	Ні	0.5	Так	Так	Ні	Ні	В. Стратегія захоплення м'яча і виконання удару в сторону ворожих воріт
189	Е3. Гра	X=1.9 Y=1.1	X = 9.9 Y=-2.8	Так	Ні	8.9	Так	Ні	Ні	Ні	А. Стратегія заняття позиції на фланзі
221	Е3. Гра	X=17.2 Y=0.3	X=29.6 Y= -1.9	Так	Ні	12.5	Так	Ні	Ні	Ні	В. Стратегія захоплення м'яча і виконання удару в сторону ворожих воріт
324	Е4. Виконання цілі тестового сценарію (Гол)	X=37.7 Y=-0.2	X=51.3 Y=0.9	Так	Так	13.6	Так	Ні	Ні	Ні	А. Стратегія заняття позиції на фланзі

Е1. Враховуючи те, що гравець повинен діяти відповідно до правил гри, режим симулятора завжди перевіряється на першому етапі аналізу. Інтелектуальний агент з кожним тактом гри отримує наступні константи, що позначають режими гри:

- before_kick_off – перед початком гри (всі агенти стоять на стартових позиціях);
- goal_r_ – гол в праві ворота;
- goal_l_ – гол в ліві ворота;
- free_kick_ – штрафний удар з місця;
- corner_kick_ – кутовий удар.

На кроці 1 (табл. 2) тестовий гравець отримує режим `before_kick_off`, а отже прямує на своє місце згідно формування перед стартом гри (рис. 3). Місце для гравця визначається координатами на полі відповідно.

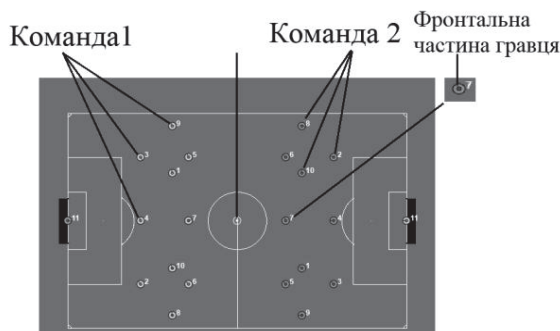


Рис. 3. Стартове формування

Е2. Ітеративно, на кожному кроці етапу 2, застосовується стратегія заняття позиції у формуванні. Порівнюється координата запланованого розташування та поточна. У випадку співпадіння, агент залишається на місці та вирівнює курс на м'яч. Для такої поведінки існує окремий алгоритм, який враховує розташування гравця та візуальний контакт із м'ячем. Допоки курс не визначено, агент здійснює оберти навколо своєї осі.

Е3. На третьому етапі МПР застосував стратегію «захоплення м'яча і виконання удару в сторону ворожих воріт». Під час аналізу ігрової ситуації, ключовими маркерами для вибору стратегії є наступні дані: роль гравця, позиція гравця та дистанція до м'яча.

Проведемо обчислення для визначення оптимальної стратегії (табл. 1) на основі даних для 3 етапу тестового сценарію 83-го такту гри (табл. 2).

А. Стратегія заняття позиції на фланзі

Беручи до уваги, що тестовий гравець є фланговим нападаючим гравцем на частині поля опонента, доцільність використання стратегії для 83-го такту складає 0.7. (табл. 1).

Б. Стратегія ведення м'яча

У випадку, якщо гравець є голкіпером, корисність дриблінгу складає 0. Якщо будь-який гравець є в штрафній зоні воріт опонентів, то корисність також складає 0 відсотків, оскільки є загроза перехвату м'яча опонентами. Таким чином, в разі локалізації на штрафному майданчику опонентів, гравець виконує удар, тобто застосовує «стратегію захоплення м'яча і виконання удару в сторону ворожих воріт».

Для 83-го такту гри, визначення локалізації позиції гравця в штрафному майданчику опонентів обчислюється наступним чином:

$$x \geq \text{ліва межа}, x \leq \text{права межа}, y \leq \text{нижня межа}, y \geq \text{верхня межа},$$

де x і y – координати гравця. Координати меж визначаються як:

– верхня межа: $(-\frac{S_y}{2}) = -40.3/2 = -20.15$;

– нижня межа: $S_y/2 = 40.3/2 = 20.15$;

– права межа: $z = 52.5$;

– ліва межа: $(z - S_x) = 52.5 - 16.5 = 36.0$,

де z – координата правої межі ігрового поля. Оскільки, для 83-го такту координати гравця становлять $x = -12.6, y = -10.1$, маємо:

$$-12.6 \geq 36.0 \quad i \quad -12.6 \leq 52.5 \quad i \quad -10.1 \leq 20.15 \quad i \quad -10.1 \geq -20.15.$$

Отже, умова не виконується.

Якщо агент має фізичний контакт із м'ячем, а також візуальний контакт з воротами опонента, то корисність складає 0.85. Для можливого фізичного контакту гравця з м'ячем повинні бути виконані наступні вимоги:

1. перебування в області ігрового поля;
2. наявність візуального контакту з м'ячем;
3. перебування на відстані до м'яча в межах радіусу R .

Алгоритм розрахунку першої вимоги аналогічний попередньому пункту. За результатами маємо підтвердження, що перша вимога виконується.

Візуальний контакт з м'ячем визначається як:

$$\Delta t \geq 0, \Delta t = t_b - (t_p - 1),$$

де t_b – такт з моменту візуального контакту з м'ячем; t_p – такт з моменту оновлення даних про навколишній світ. Підставивши дані, маємо:

$$-1 - (0 - 1) \geq 0$$

Оскільки нерівність виконується, то м'яч в полі зору агента. Третя вимога – це відстань до м'яча повинна бути в радіусі удару. Тобто:

$$S < r_y$$

де r_y – радіус удару, S – відстань до м'яча. Радіус обчислюється як:

$$r_y = d_{ym} - (d_z + d_m) = 0.7 - (0.3 + 0.085) = 0.315\text{м},$$

де d_z – радіусу корпусу гравця в проекції зверху, d_{ym} – зона досяжності удару по м'ячу, d_m – діаметр м'яча. Оскільки, відстань до м'яча на 83-ому кроці становить 12.1м (табл. 2), то умова не виконується. Як результат, одна з трьох вимог не виконується, тому на 83-ому такті гри корисність стратегії становить 0.

V. Стратегія захват м'яча і виконання удару по ворожим воротам

Дана стратегія передбачає нульову корисність для голкіпера у випадку, коли позиція м'яча невідома. Якщо гравець нападаючий, то корисність складає 0.87. Для такту 83, четвертого етапу тестового сценарію, розрахунок коефіцієнту корисності даної стратегії відбувається за формулою:

$$\min\left(0.88, \left(\frac{S}{10}\right)^{-1}\right) = \begin{cases} x = 0.88 & \text{якщо } x > 0.88 \\ x = (S/10)^{-1} & \text{якщо } x < 0.88 \end{cases}$$

Відповідно до даних таблиці 2 для 83-ого такту гри, маємо:

$$(12.1/10)^{-1} = 0.82 \Rightarrow x = 0.82$$

Отже, коефіцієнт корисності для даної стратегії на 83-ому такті гри становить 0.82.

Г. Стратегія зайняття позиції між м'ячем і власними воротями

Дана стратегія має «захисну» функцію, тому буде оптимальною для захисників та голкіпера. У випадку, якщо захисник або голкіпер перебуває на відстані більшій 5 метрів, то корисність складає 0.87 в іншому випадку 0.43. Для решти гравців корисність даної стратегії є 0.4. У випадку для 83-ого такту гри маємо 0.4.

Д. Стратегія вибивання м'яча за межі ігрового поля

Для запобігання прицільних ударів по воротах для всіх типів гравців, корисність даної стратегії становить 0.9 у випадку, якщо м'яч знаходиться у зоні власних воріт. Оскільки для 83-ого такту гри м'яч не знаходиться в зоні власних воріт, доцільність використання стратегії становить 0 відсотків.

Так як коефіцієнт корисності стратегії саме для «захоплення м'яча і виконання удару в сторону ворожих воріт» найвищий серед всіх, то саме вона буде вибрана як оптимальна (3).

Е4. Якщо удар був успішним, то зараховується результат і гра повертається з початкового формування. Цикл триває доти, доки не закінчиться час матчу.

Трасування показало (табл. 2), що алгоритм дає змогу агенту виконувати задачу в складі групи без використання шляхів комунікації. Така поведінка можлива завдяки константному шаблону поведінки (табл. 1).

Таблиця 3. Результати СУ для 83-го та 189-го тактів

Такт		83-ій	189-тий
Стратегія	А.	0.7	0.888
	Б.	0.0	0.0
	В.	0.82	0.88
	Г.	0.4	0.4
	Д.	0.0	0.0
СУ		0.82	0.888
Вибрана стратегія		В. Стратегія захоплення м'яча і виконання удару в сторону ворожих воріт	А. Стратегія заняття позиції на фланзі

Порівнюючи СУ тактів гри 83-тій і 189-тий (табл. 3), можна проаналізувати, що головні критерії для досягнення цілі (захоплення м'яча і виконання удару у ворота) ідентичні, але результуючі стратегії різні. Алгоритм врахував роль гравця, позицію гравця та м'яча і вибрав тактику заняття позиції на фланзі, хоча, одиночна гра повинна мати на меті рух до воріт.

Висновки

Створено архітектуру інтелектуального агента, на базі алгоритму Utility AI, із застосуванням дерева прийняття рішень та використанням констант корисності стратегій.

На основі результатів тестування алгоритму в офіційному симуляторі світових змагань RoboCup можна зробити наступні висновки:

- заздалегідь визначені коефіцієнти поведінки (табл.1) дозволили реалізувати командну тактику гри в умовах відсутності зв'язку між союзниками;
- запропонований інтелектуальний агент робить оцінку виходячи з поточної ситуації та не прораховує довготривалі стратегії, і тому може бути застосований в ситуаціях, де динамічно змінюються обставини;
- алгоритм дає можливість легко масштабуватися горизонтально за допомогою додавання нових поведінок та критеріїв вибору. Таке додавання виконується без зміни існуючої логіки;
- запропонований алгоритм, на відміну від алгоритмів машинного навчання, не вимагає великих обсягів даних і є більш гнучким для модифікацій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Robocup офіційний сайт [Електронний ресурс] – Режим доступу до ресурсу: <https://www.robocup.org/>.
2. Drop-in games at RoboCup / P.Stone, P. MacAlpine, K. Genter, S. Barrett. // AI Matters. – 2014. – P. 20–22.
3. Graham D. An Introduction to Utility Theory [Електронний ресурс] / David Graham – Режим доступу до ресурсу: [http://www.gameaiopro.com/ GameAIPro/ Chapter09/An_Introduction_to_Utility_Theory.pdf](http://www.gameaiopro.com/GameAIPro/Chapter09/An_Introduction_to_Utility_Theory.pdf).
4. Disentangling Successor Features for Coordination in Multi-agent Reinforcement Learning / S.Kim, G. Chowdhary, N. Stralen, H. Tran. // Carnel University. – 2022.
5. Amini A. Using Cooperation without Communication in a Multi-Agent Unpredictable Dynamic Real-Time Environment / Ali Amini. // Journal of Basic and Applied Scientific Research. – 2012. – №9. – P. 9705–9709.
6. Катренко А. В. Теорія прийняття рішень / А. В. Катренко, В. В. Пасічник, В. П. Пасько. – Київ: Видавнича група BVH, 2009. – 448 с. – (Видавнича група BVH). – (175).
7. Романенко Л. А. Базові методи прийняття рішення в задачах групового управління в середовищі симулювання футболу інтелектуальних агентів / Л. А. Романенко, О. Д. Фіногенов. // Адаптивні системи автоматичного управління. – 2022. – №40. – С. 118–128.
8. Robocup. Інструкція користувача [Електронний ресурс] – Режим доступу до ресурсу: <https://rcsoccersim.readthedocs.io/en/latest/>