УДК 004.825; 004.891

П. Кравець
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

МУЛЬТИАГЕНТНА ІГРОВА МОДЕЛЬ ПРИЙНЯТТЯ РІШЕНЬ З КОРЕЛЬОВАНИМИ СТРАТЕГІЯМИ

Досліджено проблему кооперативного прийняття рішень у мультиагентних системах на основі моделі стохастичної гри з корельованими стратегіями. Сформульовано ігрову задачу, розроблено метод та алгоритм для її розв’язування. Наведено та проаналізовано результати комп’ютерного моделювання стохастичної гри з корельованими стратегіями.

Ключові слова: кооперативне прийняття рішень, стохастична гра, корельовані стратегії, мультиагентна система.

The problem of co-operative decision-making in multiagent systems on the basis of stochastic game model with the correlated strategies is investigated. The formulation of a game problem is executed, the method and algorithm are developed for its solving. Results of computer modelling of stochastic game with the correlated strategies are described and analysed.

Key words: co-operative decision-making, stochastic game, correlated strategies, multiagent system.

Вступ

Сучасний розвиток мережних інформаційних технологій спрямований у напрямку розроблення та впровадження мультиагентних систем. Мультиагентні системи використовуються для розв’язування складних проблем розподіленого штучного інтелекту для вироблення та прийняття рішень в умовах невизначеності [1]. Агент – це автономна система з елементами штучного інтелекту, яка функціонує згідно з цілями, закладеними його розробником або власником, для чого може взаємодіяти з іншими агентами та людиною, використовуючи ресурси інформаційної мережі. Мультиагентна система (МАС) – це система, утворена декількома взаємодіючими інтелектуальними агентами з властивостями: автономності – агенти повністю або частково незалежні; спеціалізації та обмеженості знань – кожен з агентів виконує вузькоспеціалізовані функції і не має цілісного уявлення про систему; децентралізації – агенти контролюють локальні ділянки розподіленої системи.

Прийняття рішень в МАС забезпечується координацією дій агентів при розподіленому розв’язуванні ними спільної задачі. Координація (узгодження) дій ґрунтується на взаємодії агентів, яка може бути змодельована за допомогою гри. Ігрова модель визначає абстрактне формулювання цільових функцій, стратегій та результатів взаємодії агентів. Агенти діють раціонально та стратегічно, вони обирають стратегії, які оптимізують їхні цільові функції, враховуючи знання про інших агентів. Стратегії агентів визначають план гри – дії або імовірності вибору дій гравців у часі.

Для дослідження МАС, що функціонують в умовах невизначеності, використовують моделі стохастичних ігор. На відміну від детермінованих ігор, у стохастичних іграх матриці виграшів апріорі не відомі. Стохастична гра задається множиною гравців, множинами чистих стратегій та матрицями розподілів випадкових виграшів (або програшів). Стохастична гра – це повторювальна математична гра в умовах невизначеності матриць виграшів, у якій гравці після незалежного випадкового вибору чистих стратегій на основі побудованого на динамічних змішаних стратегіях імовірнісного механізму отримують випадковий виграш (або програш), який використовують для адаптивної модифікації змішаних стратегій так, щоб максимізувати функції середніх виграшів (або мінімізувати функції середніх програшів) [2].

Багатоагентна стохастична гра є зручною моделлю колективного прийняття рішень в умовах невизначеності, оскільки дозволяє дослідити процеси конкуренції, кооперації, координації, самоорганізації, навчання та адаптивної поведінки гравців.

Серед двох великих класів ігор, кооперативних та некооперативних, з погляду прийняття рішень кооперативні ігри мають більшу теоретичну привабливість та практичну цінність, оскільки дають змогу враховувати коаліційну узгодженість стратегій поведінки агентів для досягнення результату за менший час та з меншими зусиллями. У кооперативних іграх гравці домовляються про стратегії поведінки у межах створених ними коаліцій. Агенти однієї коаліції мають спільну глобальну мету, якої досягають діями окремих гравців коаліції. Тим самим коаліція накладає обмеження на стратегії поведінки її учасників. Побудова кооперативної гри агентів ґрунтується на обміні інформацією, коли гравці узгоджують власні дії, обмінюються значеннями поточних виграшів, чистими або змішаними стратегіями. У некооперативних безкоаліційних іграх агенти діють незалежно і не можуть укладати домовленостей між собою.

Відомі методи ігрового вибору варіантів рішень переважно ґрунтуються на відсутності обміну інформацією між гравцями, що спрощує математичний аналіз їх збіжності до станів колективної оптимальності. Врахування обміну інформацією між гравцями дає змогу вивчати проблеми кооперативного вироблення та прийняття рішень за рахунок укладання договорів та переговорів, у результаті яких встановлюються значення функцій виграшів, визначаються сценарії кооперативної поведінки, враховуються психологічні фактори під час вибору варіантів рішень та ін. Домовленості між гравцями можуть бути обов'язковими або необов'язковими. При обов'язкових домовленостях гравці суворо дотримуються результатів переговорів, а при необов'язкових – свобода вибору гравців ніяк не обмежується. Стабільність коаліції гравців при необов'язкових домовленостях забезпечується визначенням умов, порушення яких стає невигідним для гравців.

Під час некооперативних стохастичних ігор гравці вибирають стратегії випадково на основі незалежних розподілів імовірностей. У кооперативних іграх вибір стратегій є залежним у межах коаліції гравців. Така залежність може бути змодельована корельованими стратегіями – об'єднаним, спільним розподілом імовірностей вибору чистих стратегій гравців. Для цього до системи прийняття рішень вводиться арбітр, який на основі спільного розподілу знаходить рішення та пропонує його для виконання кожному гравцю. Для гравців це рішення не є обов'язковим – вони можуть генерувати власні незалежні від арбітра рішення. Відхилення стратегій гравців від рекомендованих арбітром значень оцінюється функцією штрафів.

Результати дослідження стохастичних ігор МАС з корельованими стратегіями не достатньо висвітлені у науковій літературі. Тому розроблення моделей стохастичних ігор для колективного вироблення і прийняття рішень, зокрема з корельованими стратегіями, є актуальною науковою та практичною проблемою, яка сьогодні привертає значну увагу дослідників МАС.

Метою роботи є побудова ігрового методу колективного прийняття рішень в умовах невизначеності з корельованими стратегіями гравців. Для досягнення мети побудовано модель стохастичної гри з корельованими стратегіями, розроблено метод, алгоритм та програмні засоби моделювання стохастичної гри для прийняття рішень в умовах невизначеності.

Модель мультиагентної стохастичної гри

Гра в умовах невизначеності задається кортежем $\Gamma_\zeta = (I, \{U^i\}_{\forall i \in I}, \{\zeta^i(U\}_{\forall i \in I})$ на множині гравців $I \neq \varnothing$, які взаємодіють зі стохастичним середовищем за допомогою скінченного набору чистих статегій $U^i = (u^i(1), u^i(2), ..., u^i(N_i))$, $N_i \geq 2$. Вибирають чисті стратегії $u_n^i = u^i$ гравці у дискретні моменти часу $n = 1, 2, ...$ випадково і незалежно на основі векторів змішаних стратегій $p_n^i \in S^{N_i}$, де S^{N_i} – N_i-вимірний одиничний симплекс з властивістю $\sum_{u^i \in U^i} p_n^i(u^i) = 1$, $p_n^i(u^i) \geq 0$ $\forall u^i \in U^i$. Значення комбінованої стратегії $u \in U = \underset{i \in I}{\times} U^i$ є колективним рішенням гравців, за яке кожен з них отримує випадковий програш $\xi_n^i(u_n)$.

Допустимо, що математичні сподівання програшів $M\{\xi_n^i(u)\} = v^i(u) = const$ не відомі та мають обмежений другий момент $\sup_n M\{[\xi_n^i(u)]^2\} = \sigma_i^2(u) < \infty$.

Нехай у системі колективного прийняття рішень присутній арбітр, який, діючи згідно з узагальненим розподілом $q_n \in S^N$, рекомендує гравцям обрати для реалізації дії, які утворюють комбінований варіант $a_n = (a_n^1, ..., a_n^L) \in U$. Тут S^N – N -вимірний одиничний симплекс ($N = \prod_{i \in I} N_i$) із властивістю $\sum_{u \in U} q_n(u) = 1,\ q_n(u) \geq 0\ \ \forall u \in U$, а $L = |I|$ – кількість гравців. Гравець з номером i отримує інформацію тільки про компоненту a_n^i комбінованого варіанта $a \in U$. Цей сигнал сприймається i -м гравцем як необов'язкова пропозиція виконати дію a_n^i . Кожен гравець таємно і незалежно вибирає у момент часу n дійсний варіант дії u_n^i , за що отримує поточний програш:

$$\zeta_n^i(u_n, a_n^i) = \lambda \xi_n^i(u_n) + (1-\lambda)\delta_n^i(u_n^i, a_n^i), \tag{1}$$

де $\xi_n^i(u_n)$ – поточний програш i -го гравця за вибір варіанта u_n^i , який залежить від колективного вибору гравців $u_n \in U$; $\delta_n^i(u_n^i, a_n^i)$ – штраф i -го гравця за відхилення вибору u_n^i від пропонованого арбітром рішення a_n^i ; $\lambda \in (0,1)$ – ваговий коефіцієнт.

У [3] показано, що найкращою відповіддю u_n^i гравців є дотримання рекомендацій арбітра a_n^i , оскільки його стратегії охоплюють ігрове поле, задане на декартовому добутку чистих стратегій усіх гравців.

Поточний штраф гравця за відхилення від рішення арбітра визначимо так:

$$\delta_n^i(u_n^i, a_n^i) = \left| ORD(u_n^i) - ORD(a_n^i) \right| / N_i ,$$

де $ORD(\cdot)$ – функція визначення порядкового номеру чистої стратегії; $\delta_n^i \in [0,1)$.

Ефективність поточних ходів гри оцінюється функціями середніх програшів гравців:

$$\mathrm{Z}_n^i = \frac{1}{n} \sum_{t=1}^{n} \zeta_t^i \quad \forall i \in I . \tag{2}$$

Метою кожного гравця є мінімізація функції середніх програшів (2):

$$\overline{\lim_{n \to \infty}} Z_n^i \to \min \ \ \forall i \in I . \tag{3}$$

Розв'язки гри шукатимемо у множині точок корельованої рівноваги (CE, Correlated Equilibrium), яка визначається об'єднаним розподілом стратегій гравців $q_n \in S^N$, коли кожен агент не має мотивацій, щоб відхилитися від домовленостей в односторонньому порядку. В умовах повної інформації маємо:

$$\sum_{a_{-i} \in A_{-i}} q^{CE}(a_{-i} \mid a_i) v^i(a_{-i}, a_i) \leq \sum_{a_{-i} \in A_{-i}} q^{CE}(a_{-i} \mid a_i) v^i(a_{-i}, \tilde{a}_i),\ \forall a_i, \tilde{a}_i \in U_i,\ \forall i \in I , \tag{4}$$

де $U_{-i} = \underset{j=1,\, j \neq i}{\overset{L}{\times}} U_j$, $U = U_{-i} \times U_i$, $a_{-i} \in U_{-i}$, $a = (a_{-i}, a_i) \in U$, $v^i(a) = M\{\zeta_n^i(a)\}$, $q^{CE}(a_i) = \sum_{a_{-i} \in A_{-i}} q^{CE}(a_{-i}, a_i)$, $q^{CE}(a_{-i} \mid a_i) = q^{CE}(a_{-i}, a_i) / q^{CE}(a_i)$, $q^{CE}(a_i) > 0$.

Домноживши обидві частини нерівності (4) на $q^{CE}(a_i)$, отримаємо еквівалентну нерівність:

$$\sum_{a_{-i} \in A_{-i}} q^{CE}(a_{-i}, a_i) v^i(a_{-i}, a_i) \leq \sum_{a_{-i} \in A_{-i}} q^{CE}(a_{-i}, a_i) v^i(a_{-i}, \tilde{a}_i) . \tag{5}$$

При $q^{CE}(a_{-i}, a_i) = \prod_{\substack{j \in I \setminus \{i\}; \\ a_j \in a_{-i}}} p_j^{NE}(a_j)$ з (5) отримуємо умову рівноваги за Нешем (NE, Nash Equilibrium). Усі точки рівноваги за Нешем є точками корельованої рівноваги [3]. Якщо $\forall i \in I$ $q^{CE}(a_{-i} \mid a_i) = q^{CE}(a_{-i} \mid \tilde{a}_i)$, $\forall a_i, \tilde{a}_i \in U_i$, $\forall a_{-i} \in U_{-i}$, $\forall q^{CE}(a_i), q^{CE}(\tilde{a}_i) > 0$, то корельована рівновага є також рівновагою за Нешем.

Множина точок СЕ-рівноваги є непорожньою, опуклою та компактною і в умовах повної інформації може бути обчислена за допомогою методів лінійного програмування. У випадку мінімізації сумарного програшу гравців задача лінійного програмування для знаходження q^{CE} може бути сформульована так:

$$\sum_{a\in U}\sigma(a)\sum_{i=1}^{L}v^i(a)\to\min_{\sigma},$$

$$\sum_{a_{-i}\in U_{-i}}q(a_{-i},a_i)\left(v^i(a_{-i},a_i)-v^i(a_{-i},\tilde{a}_i)\right)\le 0,$$

$$\forall i\in I,\forall a_i\in U_i,\forall\tilde{a}_i\in U_i,$$

$$q(a)>0\ \forall a\in U,$$

$$\sum_{a\in U}q(a)=1.$$

Значення середніх програшів $v^i(a)\ \ \forall a\in U$ апріорі не відомі гравцям в умовах невизначеності. Тому корельована рівновага визначається для випадкових послідовностей $\{u_n\}$ асимптотичною умовою:

$$\forall i\in I\qquad\overline{\lim_{n\to\infty}}\left[Z_n^i(\{u_n\})-Z_n^i(\{\hat{u}_n\})\right]\le 0, \tag{6}$$

де нерівності (6) виконуються з імовірністю 1, $\hat{u}_n=u_n\setminus u_n^i+\tilde{u}_n^i\in U$; $u_n^i,\tilde{u}_n^i\in U^i$.

Отже, в умовах невизначеності гравці повинні за спостереженнями власних поточних програшів ζ_n^i здійснювати незалежний вибір своїх чистих стратегії $u^i\in U^i$ так, щоб сформована послідовність варіантів рішень $\{u_n^i\}\ \ \forall i\in I$ при $n\to\infty$ задовольнила умову асимптотичної оптимальності (6).

Метод розв'язування ігрової задачі

Сформуємо послідовності $\{u_n^i\}$ з потрібними властивостями на основі самонавчального рекурентного методу зміни змішаних стратегій гравців [4]:

$$p_{n+1}^i=\pi_{\varepsilon_{n+1}}^{N_i}\left\{p_n^i-\gamma_n R\left(\zeta_n^i,p_n^i,u_n^i\right)\right\}\ \ \forall i\in I, \tag{7}$$

де $\pi_{\varepsilon_{n+1}}^{N_i}$ – проектор на одиничний ε-симплекс, який забезпечує нормалізацію вектора імовірностей вибору варіантів дій $p_{n+1}^i\in S_{\varepsilon_{n+1}}^{N_i}\subseteq S^{N_i}$; $\varepsilon_n>0$ – монотонно спадна послідовність величин, яка регулює розширення ε-симплексу; $\gamma_n>0$ – монотонно спадна послідовність величин, що регулює крок методу; $R\left(\zeta_n^i,p_n^i,u_n^i\right)$ – поточний крок методу; ζ_n^i – поточний програш i-го гравця.

Значення ζ_n^i обчислюється згідно з (1) із врахуванням поточної стратегії гравця та стратегії арбітра.

Вектор $R\left(\zeta_n^i,p_n^i,u_n^i\right)$ задає середній напрямок на оптимальний розв'язок стохастичної гри. В умовах невизначеності значення $R\left(\zeta_n^i,p_n^i,u_n^i\right)$ визначається на основі методу стохастичної апроксимації [4 – 6]. Так, при $R\left(\zeta_n^i,p_n^i,u_n^i\right)=\zeta_n^i/(e^{\mathrm{T}}(u_n^i)p_n^i)$ з (7) отримуємо градієнтний метод пошуку стратегій, які задовольняють умову рівноваги за Нешем [4]. При $R\left(\zeta_n^i,p_n^i,u_n^i\right)=\zeta_n^i(e(u_n^i)-p_n^i)$ з (7) отримуємо метод доповняльної нежорсткості, призначений для пошуку вирівнювальних стратегій [7]. Тут $e(u_n^i)$ позначає орт, що відповідає вибраній чистій стратегії $u_n^i\in U^i$.

Змішана стратегія арбітра q_n може визначатися стратегіями гравців, що дає змогу знаходити розв'язки гри, які задовольняють умову рівноваги за Нешем. Для цього гравці повідомляють арбітру значення поточних змішаних стратегій $p_n^i\ \ \forall i\in I$. Елементи вектора q_n є імовірностями

вибору комбінованих стратегій гравців $\forall u \in U$ і обчислюються добутком імовірностей відповідних чистих стратегій гравців $u_i \in u \ \ \forall i \in I$:

$$q_n = \left(\prod_{i \in I} p_n^i(u_i) \middle| u_i \in u, \forall u \in U \right),$$

де $U = \underset{i \in I}{\times} U^i$.

У загальному змішана стратегія q_n арбітра не залежить від стратегій гравців. Тоді в умовах невизначеності її можна задати подібно до (7):

$$q_{n+1} = \pi_{\varepsilon_{n+1}}^N \left\{ q_n - \gamma_n \Upsilon(\psi_n, q_n, u_n) \right\}, \tag{8}$$

де $N = \prod_{i \in I} N_i$; $\Upsilon(\psi_n, q_n, u_n)$ – поточний крок методу; ψ_n – отриманий від середовища поточний програш арбітра.

Середовище гри задається математичними сподіваннями програшів усіх гравців $v^i(a) \forall a \in U, \forall i \in I$.

Вважається, що арбітр генерує правильні (оптимальні у середньому) рекомендації щодо вибору стратегій гравців. Для цього використовується навчена змішана стратегія q^* або рекурентний метод (8) забезпечує більшу середньоквадратичну швидкість навчання арбітра, ніж метод навчання агентів (7).

Вибір чистої стратегії $a_n = (a_n^1, ..., a_n^L)$ арбітра на основі розподілу q_n може здійснюватись детерміновано або випадково.

Детермінована процедура полягає у визначенні максимального значення розподілу q_n:

$$a_n = \left\{ U(k) \middle| k = \arg \max_{j=1}^{N} q_n(j), N = \prod_{i \in I} N_i \right\}. \tag{9}$$

Випадкові реалізації чистих стратегій арбітра визначаються з імовірностями q_n:

$$a_n = \left\{ U(k) \middle| k = \arg \left(\min_k \sum_{j=1}^{k} q_n(j) > \omega \right), k = 1..N, N = \prod_{i \in I} N_i \right\}, \tag{10}$$

де $\omega \in [0,1]$ – випадкова величина з рівномірним розподілом.

Узагальнений розподіл q_n арбітра ґрунтується на знанні глобального стану гри і дає змогу згенерувати у середньому оптимальний комбінований сигнал a_n. На його основі визначаються рекомендовані для вибору гравцям чисті стратегії: $a_n = (a_n^1, ..., a_n^L) \in U$. У результаті, надісланий i-му гравцю сигнал a_n^i несе інформацію про оптимальний (у середньому) вибір поточної дії з врахуванням рішень інших гравців. Однак воля гравців ніяк не обмежується, і вони мають право на автономний незалежний вибір стратегій.

Власні чисті стратегії гравців вибираються аналогічно до (9) або (10) . Наприклад, випадковий вибір чистої стратегії здійснюється так:

$$u_n^i = \left\{ U^i(k) \middle| k = \arg \left(\min_k \sum_{j=1}^{k} p_n^i(j) > \omega \right), k = 1..N_i \right\} \ \ \forall i \in I. \tag{11}$$

Реалізоване колективне рішення $u_n = (u_n^1, ..., u_n^L)$ визначає поточні програші гравців $\xi_n^i(u_n) \ \forall i \in I$. Реальне значення поточних програшів ζ_n^i додатково враховує стратегію арбітра і обчислюється згідно (1).

Алгоритм розв'язування стохастичної гри

1. Задати початковий момент часу $n = 0$; множину I та кількість $L = |I|$ гравців; кількості чистих стратегій $N_i \ \ \forall i \in I$; значення чистих стратегій $U^i = (u^i(1), ..., u^i(N_i))$; початкові значення

матриць програшів $v^i(u)\ \ \forall u \in U$, $\forall i \in I$; початкові значення змішаних стратегій $p_n^i(u_i)=1/N_i$ $\forall i \in I$; точність обчислення $\varepsilon > 0$; максимальну кількість кроків повторювальної гри $n_{\max}$.

2. Обчислити значення вектора змішаної стратегії арбітра q_n згідно з (8).

3. На основі стратегії q_n виконати випадковий вибір дії арбітра $a_n = (a_n^1, ..., a_n^L)$ згідно з (10).

4. Згідно з (11) виконати випадковий вибір дій гравців u_i на основі змішаних стратегій p_n^i $\forall i \in I$.

5. Отримати поточні програші гравців $\zeta_n^i\ \ \forall i \in I$ згідно з (1).

6. Модифікувати вектори змішаних стратегій p_{n+1}^i згідно з (7).

7. Якщо $\left\| p_{n+1}^i - p_n^i \right\| < \varepsilon\ \forall i \in I$ (або $n < n_{\max}$), то задати $n := n+1$ і перейти на крок 2.

8. Вивести розраховані значення змішаних стратегій гравців $p_n^i\ \ \forall i \in I$. Кінець.

Контрольний приклад

Використовуючи розроблений алгоритм, виконаємо розв'язування стохастичної гри двох агентів з двома чистими стратегіями та матрицями програшів, заданими у таблиці.

Матриці програшів гравців

Стратегії	Перший гравець		Другий гравець	
×	$p^2(a_2[1])$	$p^2(a_2[2])$	$p^2(a_2[1])$	$p^2(a_2[2])$
$p^1(a_1[1])$	0.5	0.1	0.9	0.1
$p^1(a_1[2])$	0.3	0.2	0.1	0.9

Елементи $v^i(u)\ \ \forall u \in U$ матриць виграшів не відомі гравцям апріорі. Замість них гравці отримують нормально розподілені випадкові поточні програші

$$\xi^i = Normal(v^i(u), d^i(u))$$

з математичними сподіваннями $v^i(u)$ та дисперсіями $d^i(u)\ \ \forall u \in U$.

Нормально розподілені випадкові величини отримано з допомогою суми дванадцяти рівномірно розподілених випадкових чисел $\omega \in [0,1]$:

$$\xi_n^i(u_n, \omega) = v^i(u_n) + \sqrt{d(u_n)} \left(\sum_{j=1}^{12} \omega_j - 6 \right),$$

де $u_n \in U$.

Динаміка змішаних стратегій гравців визначається рекурентним перетворенням

$$p_{n+1}^i = \pi_{\varepsilon_{n+1}}^{N_i} \left\{ p_n^i - \gamma_n \zeta_n^i (e(u_n^i) - p_n^i) \right\}\ \ \forall i \in I, \tag{12}$$

отриманим за методом стохастичної апроксимації умови доповняльної нежорсткості:

$$V^i(U^{I\setminus\{i\}}, u^i) = V^i,\ \forall u^i \in U^i,\ \forall i \in I.$$

де $V^i = \sum_{u^i \in U^i} p^i(u^i) V^i(U^{I\setminus\{i\}}, u^i)$ – функція середніх виграшів гравців.

Параметри γ_n та ε_n є дійсними монотонно спадними величинами і можуть бути обчислені так:

$$\gamma_n = \gamma n^{-\alpha},\ \ \varepsilon_n = \varepsilon n^{-\beta}, \tag{13}$$

де $\gamma > 0$; $\alpha \in (0,1]$; $\varepsilon > 0$; $\beta > 0$.

Поточна змішана стратегія q_n арбітра визначається згідно з (8), або використовується попередньо навчена стратегія.

Збіжність стохастичної гри оцінюється такими параметрами:

1) еквклідовою нормою вектора змішаних стратегій арбітра гри:

$$\|q_n\| = \sqrt{\sum_{j=1}^{N} (q_n[j])^2}\ ; \tag{14}$$

2) усередненими за кількістю гравців середніми програшами:

$$\mathrm{Z}_n = \frac{1}{|I|} \sum_{i=1}^{|I|} Z_n^i\ ; \tag{15}$$

3) усередненою за кількістю гравців похибкою доповняльної нежорсткості:

$$\Delta_n = \frac{1}{|I|\, n} \sum_{t=1}^{n} \sum_{i \in I} \left\| p_t^i - \tilde{p}_t^i \right\|, \tag{16}$$

де $N = \prod_{i \in I} N_i$ – кількість комбінованих стратегій гравців; $|I|$ – кількість гравців; $\tilde{p}_n^i = p_n^i V_n^i (U^{i\setminus\{i\}}) / V_n^i$ – поточна оцінка вирівнювальної стратегії для умови доповняльної нежорсткості; V_n^i – поточне значення функції середніх виграшів i-го гравця.

Для визначення асимптотичного порядку швидкості збіжності використано метод моментів Чжуна [8]:

$$\overline{\lim_{n \to \infty}}\, n^{\theta} M\{\Delta_n\} \le \vartheta, \tag{17}$$

де θ – параметр порядку; ϑ – величина швидкості збіжності. Більшому θ та меншому ϑ відповідає більша швидкість збіжності ігрового методу. Довжина досліджуваної вибірки становить 10 тис. кроків.

Враховуючи (17), поведінка процесу Δ_n у часі апроксимована залежністю $\Delta_n = \vartheta / n^{\theta}$, де $\vartheta > 0$, $\theta \in (0,1]$, $n = 1,2,\ldots$. Після логарифмування отримаємо лінійне співвідношення:

$$\lg \Delta_n = \lg \vartheta - \theta \lg n\ . \tag{18}$$

Параметр $\theta = \lg \Delta_n / \lg n$ вказує на порядок швидкості збіжності стохастичної гри. Для його експериментального обчислення виконано апроксимацію випадкового процесу $\lg \Delta_n$ лінійною залежністю (18) на відрізку $\lg n \in [3,4]$ за методом найменших квадратів.

Згладжування випадкової складової швидкості збіжності та виділення порядку цієї швидкості виконано усередненням по реалізаціях випадкового процесу Δ_n.

Результати моделювання

Результати комп'ютерного моделювання стохастичної гри з корельованими стратегіями подано на рис. 1–5.

На рис. 1 у логарифмічному масштабі зображено графіки зміни у часі характеристик збіжності стохастичної гри з корельованими стратегіями. Графік 1 відповідає функції еквклідової норми вектора змішаних стратегій арбітра гри $\|q_n\|$ (14), графік 2 – усередненої по кількості гравців функції середніх програшів Z_n (15), графік 3 – усередненої за гравцями функції похибки доповняльної нежорсткості Δ_n (16). Результати отримано для таких значень параметрів гри: $L = |I| = 2$, $N_i = 2\ \forall i \in I$, $\lambda = 0.5$, $\alpha = 0.5$, $\beta = 2$, $\gamma = 1$, $\varepsilon = 0.999 / N_i$, $d = 0.01$.

Спрямування логарифмічної функції норми змішаних стратегій арбітра $\|q_n\|$ до нуля свідчить про досягнення розв'язку гри у вершині одиничного симплексу. Динаміка функції Z_n демонструє мінімізацію середніх програшів гравців. Зменшення Δ_n ілюструє наближення змішаних стратегій гравців до оптимальних значень у межах одиничного симплексу.

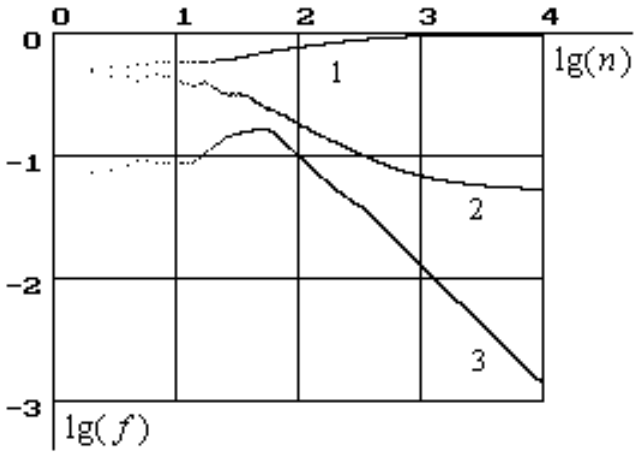
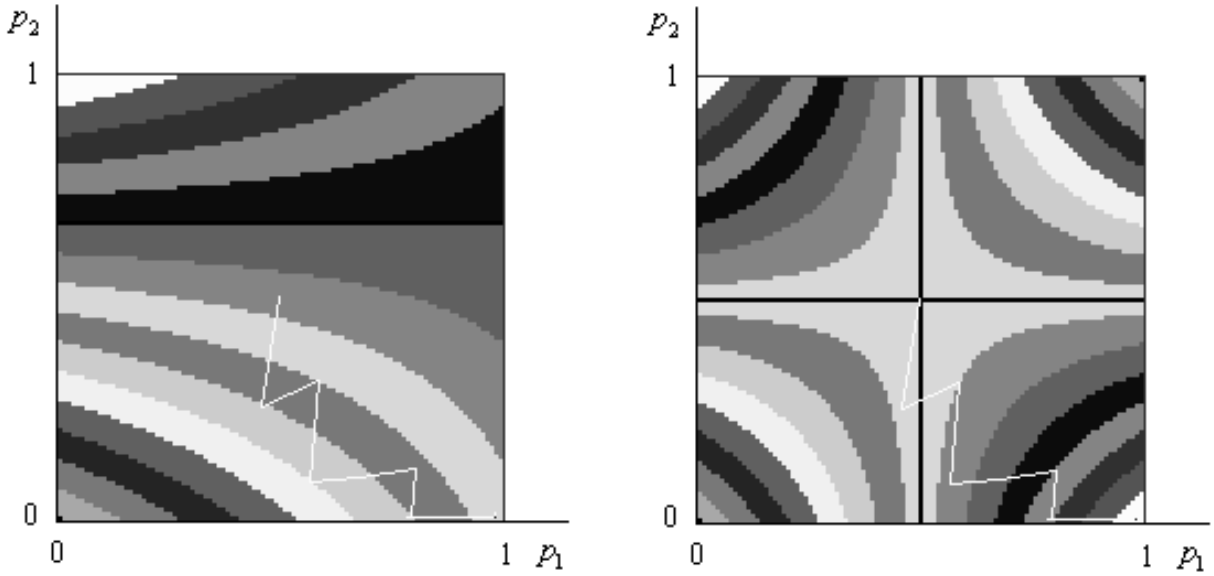


Рис. 1. Характеристики збіжності стохастичної гри з корельованими стратегіями

Загальний вигляд функцій середніх програшів гравців та отримана під час їх мінімізації траєкторія змішаних стратегій на одиничному симплексі зображені на рис. 2. Координатні осі позначають значення одного з елементів векторів змішаних стратегій $p^i = (p_i, 1 - p_i)$, $i = 1..2$ для гри 2×2. Метод (12) забезпечує досягнення оптимального розв'язку гри у точці (1, 0), що відповідає рівновазі за Нешем у чистих стратегіях.



Перший гравець

Другий гравець

Рис. 2. Динаміка стратегій та функції середніх програшів гравців

Швидкість збіжності стохастичної гри значною мірою визначається параметрами середовища гри та параметрами методу навчання векторів змішаних стратегій гравців.

Залежність функції Δ_n від дисперсії програшів d зображено на рис. 3. Із зростанням дисперсії зменшується величина ϑ швидкості збіжності. Значення дисперсії програшів гравців практично не впливає на порядок θ швидкості збіжності стохастичної гри.

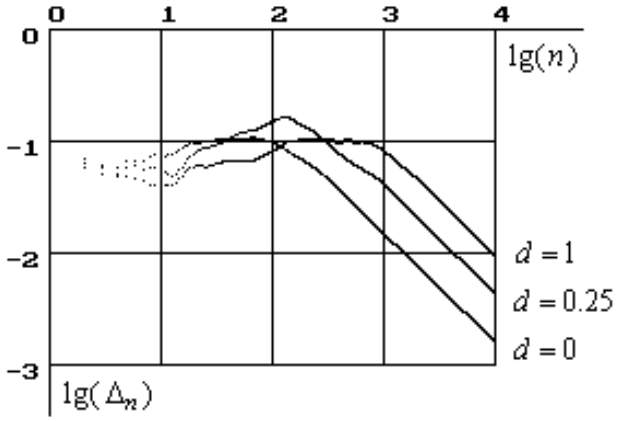


Рис. 3. Вплив дисперсії програшів d на збіжність стохастичної гри

Під час комп'ютерного моделювання виявлено нелінійну залежність швидкості збіжності стохастичної гри від параметра α методу (12). Відповідні графіки функції Δ_n для різних значень параметра α наведено на рис. 4.

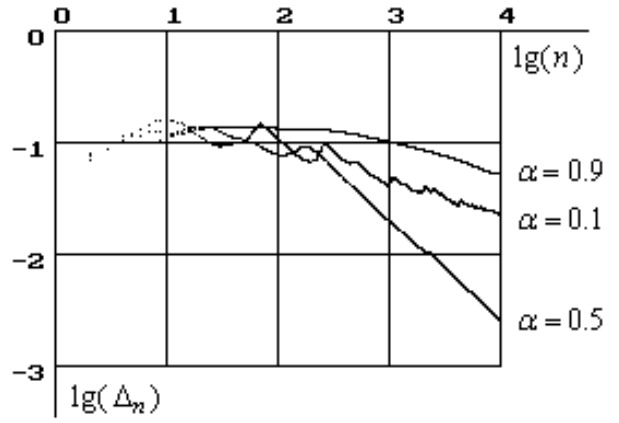


Рис. 4. Вплив параметра α на збіжність стохастичної гри

Встановлено, що для заданих матриць програшів найбільшого порядку швидкості збіжності ігрового методу з корельованими стратегіями досягається при $\alpha = 0.5$.

Швидкість збіжності стохастичної гри з навченим арбітром є більшою від гри без арбітра, що продемонстровано на рис. 5. Графік функції Δ_n, отриманий для значення вагового коефіцієнта $\lambda = 1$ поточних програшів (1), відповідає грі без арбітра, графік для коефіцієнта $\lambda = 0.5$ – грі з частковою довірою арбітру, а графік для коефіцієнта $\lambda = 0$ – грі з повною довірою арбітру.

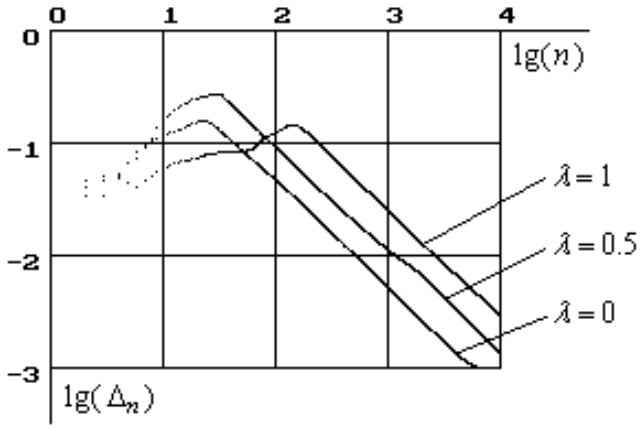


Рис. 5. Вплив параметра λ на збіжність стохастичної гри

Як видно на рис. 5, у грі з навченим арбітром у середньому забезпечується більша величина швидкості збіжності, ніж у грі без використання арбітра. Найбільша швидкість збіжності досягається у випадку, коли гравці абсолютно дотримуються рекомендацій навченого арбітра.

Висновки

Розглянута стохастична гра з корельованими стратегіями є продуктивною моделлю кооперативного прийняття рішень в МАС. Кооперація дій агентів задається узагальненим розподілом імовірностей, визначеним на декартовому добутку чистих стратегій коаліції гравців. Для цього у стохастичну гру введено додаткового гравця-арбітра, який виробляє статистично оптимальні рекомендаційні рішення, не обов'язкові для виконання коаліцією гравців. Зберігаючи незалежність, гравці отримують штраф за відхилення власного вибору від рекомендованого арбітром рішення. Дотримання гравцями визначених арбітром стратегій моделює коаліційну стійкість колективу гравців у процесі вироблення ними кооперативного рішення.

Розв'язування стохастичної гри в умовах невизначеності забезпечується динамічними змішаними стратегіми гравців, для цілеспрямованого формування яких запропоновано проекційний рекурентний метод, отриманий на основі стохастичної апроксимації вектора, що задає напрямок на оптимальний колективний розв'язок у межах одиничного симплексу. Апроксимований вектор залежить від значень поточних програшів гравців, величина яких визначається випадковою реакцією середовища на реалізовані стратегії та відхиленням стратегій від рекомендованого арбітром значення. У зв'язку з цим апроксимований вектор у середньому вказує на оптимальний колективний розв'язок стохастичної гри з корельованими стратегіями.

У результаті комп'ютерного моделювання встановлено, що за однакових початкових умов кооперативна стохастична гра з корельованими стратегіями у середньому має більшу швидкість збіжності, ніж некооперативна гра за рахунок того, що арбітр генерує оптимальні у середньому рішення. Дотримуючись раціональних рекомендацій арбітра, що враховується штрафами за відмову від їх реалізації, гравцям потрібна у середньому менша кількість кроків для знаходження розв'язку гри в умовах невизначеності.

1. Weiss G. Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence / G. Weiss, editor. – Springer Verlag, Berlin, 1996. – 643 pp. 2. Fudenberg D. The Theory of Learning in Games / D. Fudenberg, D.K. Levine. – Cambridge, MA: MIT Press, 1998. – 292 pp. 3. Greenwald A. Correlated Q-learning / A. Greenwald, K. Hall // Proceedings of the Twentieth International Conference on Machine Learning. – 2003. – P. 242–249. 4. Назин А.В. Адаптивный выбор вариантов: Рекуррентные алгоритмы / А.В. Назин, А.С. Позняк. – М.: Наука, 1986. – 288 с. 5. Граничин О.Н. Введение в методы стохастической аппроксимации и оценивания: Учеб. пособие / О.Н. Граничин. – СПб.: Изд-во СПб-го ун-та, 2003. – 131 с. 6. Вазан М. Стохастическая аппроксимация / М. Вазан. – М.: Мир, 1972. – 295 с. 7. Мулен Э. Теория игр с примерами из математической экономики / Э. Мулен. – М.: Мир, 1985. – 200 с. 8. Невельсон М.Б. Стохастическая оптимизация и рекуррентное оценивание / М.Б. Невельсон, Р.З. Хасьминский. – М.: Наука, 1972. – 304 с.