УДК 004.852, 004.942

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

ІГРОВА ЗАДАЧА ВЗАЄМОДІЇ ЕЛЕМЕНТІВ МУЛЬТИАГЕНТНИХ СИСТЕМ

Досліджено ефективність взаємодії елементів мультиагентних систем на основі моделей та методів стохастичних ігор з обміном інформацією про поточні виграші у межах локальних коаліцій гравців. Метою кожного гравця є максимізація середнього виграшу сформованої навколо нього локальної коаліції гравців. Глобальна мета гри полягає у досягненні однієї із ситуацій рівноваги між її учасниками. Синтезовано адаптивні ігрові методи з обміном інформацією та визначено умови їхньої збіжності до оптимальних за Парето стратегій.

The efficiency of interaction of elements of multi-agent systems on the basis of models and methods of stochastic games with information interchange within the limits of local coalitions of the players is investigated. The purpose of each player is increase of an average prize of the local coalition, generated in his vicinity, of the players. The global purpose of game consists in achievement of a situation of balance between its participants. The synthesis of adaptive game methods with information interchange is executed and the conditions of their convergence to Pareto-optimum strategies are specified.

Вступ

Мультиагентні системи (МАС) представляють сучасний напрямок розвитку розподілених систем штучного інтелекту. Сьогодні це перспективна сфера досліджень і застосувань, яка поєднує поняття, ідеї та результати багатьох дисциплін, включаючи штучний інтелект, інформатику, теорію керування, соціологію, економіку, філософію [1, 2]. Базовими поняттями МАС є агент та середовище. Агент – це автономна апаратна або програмна сутність, яка діє в інтересах досягнення сформульованих користувачем цілей, виходячи із своїх внутрішніх уявлень про середовище. Агент є активним об’єктом, який може отримувати, опрацьовувати, передавати інформацію, навчатися та приймати самостійні рішення. Агенти також мають здатність спілкуватися – отримувати та відправляти повідомлення один одному. Агенти існують у середовищі, взаємодіють з середовищем та між собою і можуть змінювати середовище.

Під взаємодією розуміють співпрацю (узгодження, координацію дій) множини агентів, яка забезпечує оптимальні режими роботи багатофакторної або розподіленої системи прийняття рішень. З теоретико-ігрової точки зору співпрацю агентів можна розглядати як проблему пошуку рівноваги у колективній повторюваній грі [3].

Співпраці елементів МАС в основному можна досягти такими способами: реалізаціями ігрових стратегій з врахуванням попередніх домовленостей; поточним навчанням агентів під час повторювальної гри; спілкуванням між агентами під час гри.

Здатність до навчання є необхідною умовою адаптації та самоорганізації мультиагентної системи в умовах невизначеності. Навчання агентів, як правило, здійснюється методом проб і помилок під час взаємодії з динамічним середовищем. Методи навчання агентів – це надзвичайно активна галузь сучасних досліджень, тісно пов’язана з машинним навчанням, системами штучного інтелекту, нейромережами, психологією, статистикою та ін. [4, 5]

Спілкування полягає в обміні інформацією між агентами перед реалізацією чергового кроку гри. Для спілкування агенти використовують певну сигнальну систему – спеціалізовану мову діалогу. Спілкування регламентують протоколи комунікації, які мають декілька рівнів. Найнижчий рівень протоколу визначає метод взаємного зв’язку між агентами; середній рівень визначає формат, або синтаксис переданої інформації; верхній рівень визначає значення (семантику) інформації.

Агенти-гравці можуть обмінюватися інформацією про поточні виграші або обрані стратегії. Спілкування повинно допомогти агентам уникати координаційних помилок, оптимізувати власну або колективну поведінку. Попередньо можна припустити, що обмін інформацією зменшує невизначеність системи, прискорює навчання та призводить до зростання ефективності дій агентів.

Координації роботи агентів досягають також за неявної комунікації, коли агенти формують власні дії на основі спостережень за діями інших, або на основі логічних висновків про можливі раціональні дії інших агентів у поточній ситуації.

Надалі вважають, що всі гравці є абсолютно надійними, мають ідентичні інтереси, а у грі з обміном інформацією відсутні спотворення у каналах передавання даних.

Обмін інформацією між гравцями забезпечує досягнення оптимальних станів рівноваги, які описуються колективними стратегіями гравців, наприклад, таким є стан оптимальності за Парето. За відсутності обміну інформацією між гравцями стани рівноваги описують індивідуальними стратегіями гравців, наприклад, таким є стан рівноваги за Нешем [6].

Розв’язок гри є оптимальним за Парето, якщо не існує колективної стратегії, за якою можна збільшити виграш одночасно для всіх гравців. Розв’язок гри є рівноважний за Нешем, якщо для кожного окремого гравця не існує індивідуальної стратегії, за якою можна збільшити його виграш за умови, що усі інші гравці дотримуються своїх оптимальних стратегій.

У цій роботі досліджено вплив навчання та передавання інформації між гравцями на колективну поведінку в координаційних іграх. Мета дослідження полягає в отриманні відповіді на запитання: за яких умов обмін інформацією забезпечує оптимальну рівновагу стохастичної гри.

Формулювання ігрової задачі в умовах невизначеності

Нехай D – непорожня множина гравців, які взаємодіють між собою у випадковому середовищі з апріорі невідомими імовірнісними характеристиками. Кожен гравець $i \in D$ змінює стани середовища за допомогою чистих статегій $U^i = (u^i(1), u^i(2), ..., u^i(N_i))$, які вибирають у дискретні моменти часу $n = 1,2,...$.

Для кожного гравця задано непорожню локальну підмножину сусідніх гравців $D_i \subseteq D \; \forall i \in D$. Колективні стратегії, утворені декартовим добутком чистих стратегій гравців з множини D_i,

$$u^{D_i} \in U^{D_i} = \underset{j \in D_i}{\otimes} U^j \tag{1}$$

визначають випадкову величину поточного виграшу $\xi_n^i = \xi_n^i(u_n^{D_i}, \omega)$, яка є функцією елементарних випробувань $\omega \in \Omega$. Позначимо через $\tilde{D}_i$ множину гравців, виграші яких залежать від стратегій *i*-го гравця.

Середовище прийняття рішень задають системою розподілів випадкових виграшів з матрицями математичних сподівань $\left[v^i(u^{D_i})\right]_{\forall u^{D_i} \in U^{D_i}}$ та дисперсій $\left[d^i(u^{D_i})\right]_{\forall u^{D_i} \in U^{D_i}}$ $\forall i \in D$. Нехай гра агентів відбувається у стаціонарному знакододатному середовищі, для якого

$$v^i(u^{D_i}) = const, \; d^i(u^{D_i}) = const, \; \min_{i \in D} \min_{u^{D_i}} v^i(u^{D_i}) > 0 \quad \forall u^{D_i} \in U^{D_i} \; \forall i \in D. \tag{2}$$

Приймемо такі припущення: 1) послідовності випадкових величин $\{\xi_n^i\}$ незалежні $\forall u_n^{D_i} \in U^{D_i}$, $\forall i \in D$, $\forall n = 1,2,...$; 2) для будь-яких $i \in D$, $u^{D_i} \in U^{D_i}$, $n = 1,2,...$ математичні сподівання $M\{\xi_n^i(u^{D_i}, \omega)\} = v(u^{D_i}) = const$ не відомі та мають обмежений другий момент

$\sup_n M\{[\xi_n^i(u^{D_i},\omega)]^2\} = \sigma_i^2(u^{D_i}) < \infty$; 3) гравці вибирають чисті стратегії незалежно один від одного: $P\{u_n^{D_i} \mid u_t^{D_i}, \xi_t^j (t = \overline{1,n-1}; \forall j \in D_i)\} = \prod_{j \in D_i} P\{u_n^j \in U^j \mid u_t^j, \xi_t^j (t = \overline{1,n-1})\} \quad \forall i \in D$.

Залежно від можливостей комунікації між гравцями розрізняємо два види ігор: без обміну та з обміном інформацією.

В <u>іграх без обміну інформацією</u> кожен гравець незалежно максимізує власні виграші ξ_n^i; гравці взаємодіють лише за параметрами середовища.

В <u>іграх з обміном інформацією</u> гравці обмінюються між собою даними про результати взаємодії з середовищем або про стан, у якому вони перебувають. Наприклад, гравці можуть повідомляти один одного про поточні виграші, значення чистих або змішаних стратегій.

Надалі вважатимемо, що кожен *i*-й гравець в моменти часу $n = 1,2,...$ повідомляє гравців з множини D_i про величину отриманого виграшу ξ_n^i. З врахуванням локальної визначеності гри *i*-й гравець отримає таку інформацію від множини гравців $\tilde{D}_i$, виграші яких залежать від його стратегій. Отримані поточні виграші зважуються додатними коефіцієнтами, які можуть визначати ступінь важливості гравця або міру довіри до отриманої інформації. Тоді для локально визначеної гри з обміном інформацією поточні виграші i-го гравця обчислюють так:

$$\xi_n^i = \sum_{k \in \tilde{D}_i} \alpha^k \xi_n^k \quad \forall i \in D, \tag{3}$$

де $\alpha^k > 0 \quad \forall k \in \tilde{D}_i$.

Виграші (3) визначаються колективними стратегіями

$$u^{K_i} \in U^{K_i} = \bigcup_{k \in \tilde{D}_i} \bigotimes_{j \in D_k} U^j \tag{4}$$

об'єднання множин гравців

$$K_i = \bigcup_{k \in \tilde{D}_i} D_k. \tag{5}$$

Обчислені поточні виграші гравці використовують для ефективного вибору варіантів рішень надалі. Ефективність вибраних варіантів оцінюють поточними середніми виграшами гравців.

Для гри без обміну інформацією поточні середні виграші i-го гравця визначаються послідовністю спільних стратегій $\{u_t^{D_i} \mid t = \overline{1,n}\}$ гравців з локальної множини D_i:

$$\Phi_n^i(\{u_n^{D_i}\}) = \frac{1}{n}\sum_{t=1}^{n} \xi_t^i \quad \forall i \in D. \tag{6}$$

Для гри з обміном інформацією поточні середні виграші визначаються спільними стратегіями $\{u_t^{K_i} \mid t = \overline{1,n}\}$ гравців з множини K_i:

$$\Phi_n^i(\{u_n^{K_i}\}) = \frac{1}{n}\sum_{t=1}^{n} \sum_{k \in \tilde{D}_i} \alpha^k \xi_t^k \quad \forall i \in D \tag{7}$$

Метою кожного гравця є максимізація середніх виграшів:

$$\underline{\lim}_{n \to \infty} \Phi_n^i \to \max \; \forall i \in D. \tag{8}$$

Розв'язки задачі векторної оптимізації (8) треба шукати у множині компромісних рішень, наприклад, рівноваги за Нешем або оптимальності за Парето [6–8].

Оскільки імовірнісні характеристики випадкових величин ξ_n^i не відомі, то в умовах невизначеності стохастична гра повинна складатися з необмеженої послідовності ходів $\{u_n^i\}$ $\forall i \in D$ з метою забезпечення достовірного виконання однієї з умов асимптотичної оптимальності, наприклад:

1) ε -рівноваги за Нешем:

$$\forall i \in D \quad \underline{\lim}_{n\to\infty} \left[\Phi_n^i(\{u_n^{D_i}\}) - \Phi_n^i(\{\hat{u}_n^{D_i}\}) \right] \ge \varepsilon\,; \tag{9}$$

2) ε -оптимальності за Парето:

$$\begin{cases} \forall i \in D \quad \underline{\lim}_{n\to\infty} \left[\Phi_n^i(\{u_n^{D_i}\}) - \Phi_n^i(\{\tilde{u}_n^{D_i}\}) \right] \ge \varepsilon \\ \exists i \in D \quad \underline{\lim}_{n\to\infty} \left[\Phi_n^i(\{u_n^{D_i}\}) - \Phi_n^i(\{\tilde{u}_n^{D_i}\}) \right] > \varepsilon \end{cases}, \tag{10}$$

де нерівності (9) та (10) виконуються з ймовірністю 1, $u_n^{D_i}, \tilde{u}_n^{D_i} \in U^{D_i}$; $\hat{u}_n^{D_i} = u_n^{D_i} \setminus u_n^i + \tilde{u}_n^i \in U^{D_i}$; $u_n^i, \tilde{u}_n^i \in U^i$; $\varepsilon \ge 0$.

Якщо $\varepsilon = 0$, то з (9) та (10) відповідно отримуємо умову абсолютної рівноваги за Нешем та оптимальності за Парето.

Враховуючи описані моделі середовища прийняття рішень, визначення структури гри, моделі ігрової взаємодії гравців, критерії та умови оптимальності, сформулюємо ігрову задачу в умовах невизначеності. Гравці повинні за спостереженнями власних поточних виграшів ξ_n^i та під час обміну ними в межах локально визначених коаліцій D_i $\forall i \in D$ незалежно вибирати свої чисті стратегії $u^i \in U^i$ так, щоб сформована послідовність варіантів рішень $\{u_n^i\}$ $\forall i \in D$ за $n \to \infty$ з імовірністю 1 задовольнила умови асимптотичної оптимальності (9) або (10).

Для досягнення асимптотичних цільових рішень стохастична гра в умовах невизначеності повина складатися з реалізацій чистих стратегій, отриманих за допомогою генератора випадкових величин, з динамічним у часі, адаптивним законом розподілу [8].

Адаптивною, або самонавчальною називають стохастичну гру в умовах невизначеності середовища прийняття рішень, у якій гравці на основі аналізу інформації про передісторію реалізують оптимальні стратегії поведінки, які у середньому призводять до зменшення ймовірності прийняття неефективних рішень і забезпечують виконання однієї з умов асимптотичної оптимальності.

Детермінована матрична ігрова задача

Нехай чисті стратегії вибирає незалежно кожний гравець на основі векторів змішаних стратегій $p^i \in S^{N_i}$, визначених на одиничних симплексах

$$S^{N_i} = \left\{ p^i \,|\, \sum_{j=1}^{N_i} p^i(j) = 1;\, p^i(j) \ge 0\,;\, j = \overline{1, N_i} \right\}. \tag{11}$$

Для розв'язування сформульованої ігрової задачі в умовах невизначеності розглянемо тісно пов'язану з нею безкоаліційну детермінованому матричну гру у змішаних стратегіях з полілінійними функціями середніх виграшів

$$V^i(p^{D_i}) = \sum_{u^{D_i} \in U^{D_i}} v^i(u^{D_i}) \prod_{j \in D_i; u^j \in u^{D_i}} p^j(u^j)\,, \tag{12}$$

заданими на опуклих симплексах $S^{D_i} = \prod_{j \in D_i} S^{N_j}$ змішаних стратегій.

З теореми Гліксберга для нерухомої точки, або більш часткових результатів теорем Брауера та Неша [9] випливає, що матрична гра має хоча б одну точку рівноваги у змішаних стратегіях $p^* = (p^{i*} \mid \forall i \in D)$, для якої виконується одна з умов:

1) ε -рівновага за Нешем:

$$\forall i \in D \quad V^i(p^{D_i *}) - V^i(p^{D_i \setminus i*}, p^i) \geq \varepsilon ;$$

2) ε -стійкість за Парето:

$$\begin{cases} \forall i \in D \quad V^i(p^{D_i *}) - V^i(p^{D_i}) \geq \varepsilon \\ \exists i \in D \quad V^i(p^{D_i *}) - V^i(p^{D_i}) > \varepsilon \end{cases},$$

де $\varepsilon \geq 0$; $V^i(p^{D_i *})$ – функція середніх виграшів, визначена у точці розв'язку $p^{D_i *} \in S^{D_i}$; $V^i(p^{D_i})$ – функція середніх виграшів, визначена у будь-якій точці симплексу S^{D_i}; $V^i(p^{D_i \setminus i*}, p^i)$ – функція середніх виграшів, визначена на симплексі S^{D_i} за будь-якого відхилення стратегії i -го гравця від точки розв'язку.

Враховуючи компактність та опуклість просторів S^{N_i} $\forall i \in D$, для існування рівноважних за Нешем або оптимальних за Парето рішень достатньо неперервності функцій середніх виграшів (12) [6, 7].

Розв'язок матричної гри з локальними зв'язками є ε-рівноважним за Нешем, якщо будь-яке відхилення від оптимальної стратегії p^{i*} кожного гравця $i \in D$ не може принести йому збільшення функції середнього виграшу $V^i(p^{D_i})$ більше ніж на $\varepsilon \geq 0$, якщо гравці з локальної підмножини D_i дотримуються точки рівноваги $p^{D_i \setminus i*}$. При $\varepsilon = 0$ розв'язок, рівноважний за Нешем, відображає приватні інтереси кожного гравця, які можна висловити твердженням: "не існує індивідуальної стратегії, покращання індивідуального рішення, якщо гравці з локально зв'язаної підмножини дотримуватимуться ситуації рівноваги".

У загальному випадку у точці Неша можуть існувати домінуючі колективні стратегії p^{D_i} $\forall i \in D$, які покращують рішення одночасно для всіх гравців. На відміну від цього, стійкість за Парето визначає множину непокращуваних колективних рішень.

Розв'язок матричної гри з локальними зв'язками є ε -стійким (інакше ε -оптимальним) за Парето, якщо в точці розв'язку не існує колективних стратегій p^{D_i} таких, щоб збільшити функцію середнього виграшу $V^i(p^{D_i})$ більше ніж на $\varepsilon \geq 0$ $\forall i \in D$. При $\varepsilon = 0$ розв'язок, оптимальний за Парето, відображає колективні інтереси гравців: "не існує колективної стратегії, за якою можна покращати рішення одночасно для всіх гравців".

У загальному випадку рівновага за Нешем та оптимальність за Парето є компромісними розв'язками для групи гравців, оскільки діючи індивідуально у цьому ж середовищі, вони можуть досягти кращих для себе результатів.

Особливість рівноваги за Нешем та оптимальності за Парето для гри з локальними зв'язками полягає в тому, що для i -го гравця ці умови виконуються і тоді, коли гравці з доповняльної множини $D \setminus D_i$ не дотримуються обов'язкових домовленостей, тобто локальна визначеність гри призводить до покращання стійкості її розв'язку.

Доведення асимптотичної адекватності детермінованої та стохастичної ігрової задачі виконано в [8].

Визначення розв'язків на основі умови доповняльної нежорсткості

Розв'язки, які задовольняють задані умови оптимальності, можна визначити за умовою доповняльної нежорсткості та властивостей вирівнювальних стратегій.

За умовою доповняльної нежорсткості, для матричної гри без обміну інформацією у точках рівноваги за Нешем у змішаних стратегіях виконується система рівностей [7]:

$$\nabla_{p^i} V^i = V^i e^{N_i}, \; p^i \in S^{N_i}, \; \forall i \in D, \tag{13}$$

де $\nabla_{p^i} V_n^i$ – градієнт функції середніх виграшів; $e(u^{N_i}) = (1_j \mid j = \overline{1, N_i})$ – вектор, всі компоненти якого дорівнюють 1.

Розв'язок системи полілінійних рівнянь (13) за відомих невироджених матрицях виграшів визначає точку рівноваги за Нешем.

Змішані стратегії, для яких справедлива умова доповняльної нежорсткості (13), називають вирівнювальними.

У загальному випадку вирівнювальні стратегії є підмножиною повної множини рівноважних за Нешем рішень.

З теорії векторної оптимізації відомо [10], що оптимальний за Парето розв'язок задачі можна отримати шляхом максимізації згортки увігнутих (вгору) критеріальних функцій V^i, заданих на опуклих областях їхнього визначення:

$$W = \sum_{i \in D} \alpha^i V^i (p^{D_i}) \to \max_{\forall p^i}, \tag{14}$$

де $\alpha^i > 0 \;\; \forall i \in D$ – вагові коефіцієнти.

Коефіцієнти α^i у (14) можуть визначати ступінь важливості критеріїв оптимальності. Здебільшого їх задають у вигляді нормованого вектора $\alpha = (\alpha^i \mid \alpha^i \in (0,1); \sum_{i \in D} \alpha^i = 1)$. Змінюючи елементи вектора α, можна знайти всі парето-оптимальні розв'язки задачі. Для рівноважливих критеріїв необхідно задати $\alpha^i = 1/|D|$.

Із врахуванням локальної визначеності функцій середніх виграшів V^i у базисі змішаних стратегій гравців з множини D_i, задача (14) зводиться до умовної максимізації системи локальних згорток таких функцій за параметрами p^i:

$$W^i = \sum_{k \in \tilde{D}_i} \lambda^i(k) V^k (p^{D_k}) \to \max_{p^i}, \quad p^i \in S^{N_i} \quad \forall i \in D, \tag{15}$$

де $\tilde{D}_i$ – множина гравців, функції середніх виграшів яких залежать від стратегії *i*-го гравця, а коефіцієнти $\lambda^i(k) \forall k \in \tilde{D}_i$ є елементами вектора:

$$\lambda^i = \left(\lambda^i(k) \forall k \in \tilde{D}_i \,\middle|\, \lambda^i(k) = \alpha^k \Big/ \sum_{j \in \tilde{D}_i} \alpha^j > 0; \; \sum_{k \in \tilde{D}_i} \lambda^i(k) = 1 \right). \tag{16}$$

У частковому випадку, якщо $\lambda^i(i) = 1$, а $\lambda^i(k) = 0 \; \forall k \in \tilde{D}_i, k \neq i$, тобто за відсутності обміну інформацією між гравцями, з (15) отримаємо $W^i = V^i \;\; \forall i \in D$.

Функції середніх виграшів $V^i \;\; \forall i \in D$ визначено у просторі варіантів U^{D_i} спільного вибору гравців з множини D_i. Функції W^i визначено у просторі варіантів U^{K_i} (4), що визначається стратегіями гравців з множини K_i (5).

Розв'язуючи задачу умовної оптимізації (15) методом Лагранжа [10], отримаємо систему рівнянь

$$\nabla_{p^i} W^i(p^{K_i}) = e^{N_i} W^i(p^{K_i}) \quad \forall i \in D, \tag{17}$$

де $\nabla_{p^i} W^i$ – градієнт функції W^i за вектором p^i; $e^{N_i} = (1_j \mid j = \overline{1, N_i})$ – вектор, всі компоненти якого дорівнюють 1. Система (17) описує умову рівноваги за Нешем для локальних згорток функцій середніх виграшів W^i. Отже, оптимальні за Парето розв'язки гри без обміну інформацією задовольняють рівновагу за Нешем для гри з обміном інформацією (у межах запропонованої моделі гри).

Вираз (17) означає таке: оптимальним розв'язком гри є набір змішаних стратегій p^{i*} $\forall i \in D$, визначених на непорожніх спектрах чистих стратегій $\tilde{U}^i = \{ j \mid p^{i*}(j) > 0\} \subseteq U^i$, які забезпечують виконання системи рівностей $\nabla W^i(j) = W^i$ $\forall j \in \tilde{U}^i$. Якщо ж $\exists k \in U^i$, що $\nabla W^i(k) \neq W^i$, то $k \notin \tilde{U}^i$ та $p^{i*}(k) = 0$.

З вигляду функцій середніх виграшів (12) та (15) та означення умов оптимальності випливає, що для отримання рівноважного за Нешем розв'язку ігрової задачі не вимагається обміну будь-якою інформацією між гравцями, а для отримання отимального за Парето розв'язку є необхідним обмін інформацією між гравцями про ефективність реалізованих стратегій.

Методи розв'язування задачі

В умовах апріорної невизначеності матриць виграшів оптимальні розв'язки не можна отримати явно з систем рівнянь (13) або (17). Для досягнення одного із станів оптимальності необхідно застосувати процедуру його цілеспрямованого пошуку на системі одиничних симплексів.

Нехай елементи векторів змішаних стратегій є умовними імовірностями $p_n^i(j) = P\{u_n^i = u^i(j) \mid u_t^i, \xi_t^i (t = \overline{1, n-1}\}$, $j = \overline{1, N_i}$, значення яких змінюється у дискретні моменти часу $n = 1,2,...$ і залежить від передісторії випадкових процесів $\{u_t^i\}$ та $\{\xi_t^i\}$. Поточні значення $p_n^i(u_n^i)$ визначають вибір чистих стратегій $u_n^i = u^i \in U^i$ у момент часу n. Послідовність векторів $\{p_n^i\}$ повністю визначає властивості послідовностей чистих стратегій $\{u_n^i\}$ $\forall i \in D$, $n = 1,2,...$.

Переміщення векторів змішаних стратегій у межах одиничного симплексу виконаємо за допомогою рекурентного методу, побудованого на основі матричного формулювання ігрової задачі та умови доповняльної нежорсткості (17). Для цього для кожного гравця сформуємо вектор похибки виконання умови доповняльної нежорсткості, покомпонентно зважений елементами змішаної стратегії p^i:

$$diag(p^i)(e^{N_i} W^i(p^{K_i}) - \nabla_{p^i} W^i(p^{K_i})) = 0 \quad \forall i \in D, \tag{18}$$

де $diag(p^i)$ – квадратна діагональна матриця порядку N_i, побудована з елементів вектора p^i.

Зважування похибки доповняльної нежорсткості (18) необхідне для врахування можливих розв'язків на межі одиничного симплексу.

Побудуємо вектор руху рекурентного методу так, щоб у середньому він визначав виконання умови доповняльної нежорсткості (18):

$$M\{R_n \mid p_n^i = p^i\} = diag(p^i)[W^i e^{N_i} - \nabla_{p^i} W^i] = M\{\sum_{k \in \tilde{D}_i} \lambda^i(k) \xi_n^k [p_n^i - e(u_n^i)] \mid p_n^i = p^i\}, \tag{19}$$

де $p^i \in S^{N_i}$; $e(u_n^i)$ – одиничний вектор-індикатор вибору чистої стратегії $u_n^i \in U^i$.

На основі (19) методом стохастичної апроксимації [11] для гри з обміном інформацією отримано такий рекурентний метод:

$$p_{n+1}^i = \pi_{\varepsilon_{n+1}}^{N_i} \left\{ p_n^i - \gamma_n \sum_{k \in \tilde{D}_i} \lambda^i(k) \xi_n^k \left[p_n^i - e(u_n^i) \right] \right\}, \tag{20}$$

де $\gamma_n > 0$ – крок методу; λ^i – вектор коефіцієнтів, елементи якого визначаються згідно з (16); $\pi_{\varepsilon_{n+1}}^{N_i}$ – оператор оптимального проектування на ε-симплекс

$$S_\varepsilon^{N_i} = \left\{ p^i \mid p^i \in S^{N_i};\ p^i(j) \ge \varepsilon\ \ (j = \overline{1, N_i}) \right\},\ \varepsilon \in (0, \min_i N_i^{-1}). \tag{21}$$

Оператор $\pi_{\varepsilon_{n+1}}^{N_i}$ задовольняє умови:

$$\left\| p^i - \pi_\varepsilon^{N_i}\{q^i\} \right\| \le \left\| p^i - q^i \right\|;\ \pi_\varepsilon^{N_i}\{q^i\} \in S_\varepsilon^{N_i} \quad \forall i \in D, \forall p^i \in S_\varepsilon^{N_i}, \forall q^i \in R^{N_i}.$$

Роботу методу (20) напрямлено на максимізацію згортки функцій середніх виграшів, або, інакше, середнього виграшу коаліції гравців $\tilde{D}_i$. Працездатність методу (20) забезпечується умовою псевдоградієнтності

$$\rho_n = \left\langle diag(p_n^i)(e^{N_i} W_n^i - \nabla W_n^i),\ p_n^i - \tilde{p}_n^i \right\rangle = W_n^i \Delta_n^i \ge 0 \tag{22}$$

його вектора руху відносно функції Ляпунова

$$\Delta_n = \sum_{i \in D} \Delta_n^i = \sum_{i \in D} \left\| p_n^i - \tilde{p}_n^i \right\|^2, \tag{23}$$

де $\tilde{p}_n^i = diag(p_n^i) \nabla_{p_n^i} W_n^i \big/ W_n^i$ – зважена змішана стратегія *i*-го гравця, яка визначає поточне відхилення від виконання умови доповняльної нежорсткості. Умова (22) достовірно виконується для методу (20) у знакододатному середовищі.

Верхня оцінка відхилення від оптимального розв'язку для методу (20) має вигляд:

$$M\left\{ \Delta_{n+1} \middle| \mathsf{F}_n \right\} \le (1 - 2\gamma_n w_{\min}) \Delta_n + C(\mu_n + \gamma_n^2), \tag{24}$$

де $\mathsf{F}_n = \left\{ u_t^i, \xi_t^i \mid t = \overline{1, n-1}; \forall i \in D \right\}$ – σ-алгебра, яка визначає передісторію подій; $w_{\min} = \min_{i \in D} \{ \sum_{k \in \tilde{D}_i} \lambda^i(k) v_{\min}^k \}$, $v_{\min}^k = \min_{u^{D_k} \in U^{D_k}} v(u^{D_k}) > 0$, $C \sim |D| N_{\max} (w_{\max} - w_{\min}) \sigma_{\max}^2 > 0$, $N_{\max} = \max_{i \in D} N_i$, $w_{\max} = \max_{i \in D} \{ \sum_{k \in \tilde{D}_i} \lambda^i(k) v_{\max}^k \}$, $v_{\max}^k = \max_{u^{D_k} \in U^{D_k}} v(u^{D_k}) > 0$, $\sigma_{\max}^2 = \max_{i \in D} (\sigma_{\max}^i)^2$, $(\sigma_{\max}^i)^2 = \max_{u^{D_i} \in U^{D_i}} [\sigma^i(u^{D_i})]^2$, $\mu_n = |\gamma_{n-1} - \gamma_n| + |\varepsilon_n - \varepsilon_{n+1}|$.

Усереднюючи (24) за F_n, отримаємо:

$$M\{\Delta_{n+1}\} \le (1 - 2\gamma_n w_{\min}) M\{\Delta_n\} + C(\mu_n + \gamma_n^2). \tag{25}$$

Твердження 1. Нехай виконано припущення про незалежність випадкових величин, а також при $n \to \infty$ $\gamma_n > 0$; $\gamma_{n+1} < \gamma_n$; $\sum_{n \to \infty} \gamma_n = \infty$; $\varepsilon_n \in (0, \min_{i \in D} N_i^{-1})$; $\varepsilon_{n+1} < \varepsilon_n$; $\lim_{n \to \infty} \varepsilon_n = 0$. Тоді проекційний ігровий метод (20) з обміном інформацією між гравцями для будь-якого початкового наближення $p_1^i \in S_{\varepsilon_1}^{N_i}; \forall i \in D$ за фіксованих векторів λ^i $\forall i \in D$ (15) забезпечує виконання умови доповняльної нежорсткості (18) у знакододатному середовищі $w_{\min} > 0$:

1) з ймовірністю 1, якщо $\sum_{n=1}^{\infty} (|\varepsilon_n - \varepsilon_{n-1}| + \gamma_n^2) < \infty$;

2) у середньоквадратичному, якщо $\lim_{n \to \infty} (|\varepsilon_n - \varepsilon_{n-1}| \gamma_n^{-1} + \gamma_n) = 0$.

Збіжність з імовірністю 1 отримано з оцінки (24) та наслідків теореми Роббінса–Сігмунда [8]. Збіжність у середньоквадратичному отримано на основі середніх оцінок (25) та результатів теореми про рекурентні числові нерівності [12].

Збіжність методу (20) дослідимо у класі монотонно спадних послідовностей $\{\gamma_n\}$ та $\{\varepsilon_n\}$ вигляду

$$\gamma_n = \gamma(n+a)^{-\alpha}; a > 0; \ \varepsilon_n = \varepsilon(n+b)^{-\beta}; b > 0. \tag{26}$$

Збіжність методу (20) існує:

1) з ймовірністю 1, якщо $\alpha \in (0.5;1]$; $\beta > 0$;

2) у середньоквадратичному, якщо $\alpha \in (0,1]$; $\beta > 0$.

Оцінювання асимптотичного порядку швидкості збіжності виконано для послідовностей величин (26) методом моментів Чжуна [13]:

$$\overline{\lim_{n\to\infty}} n^{\theta} M\{\varDelta_n\} \le \vartheta, \tag{27}$$

де θ – параметр порядку, ϑ – величина швидкості збіжності. Більшому θ та меншому ϑ відповідає більша швидкість збіжності ігрового методу. Для цього, враховуючи (26), перепишемо оцінку (25) у вигляді:

$$M\{\varDelta_{n+1}\} \le \left(1 - 2w_{\min}\frac{\gamma}{n^{\alpha}}\right) M\{\varDelta_n\} + C\frac{m}{n^{\rho}}, \tag{28}$$

де $m = \varepsilon\beta\chi(1+\beta \le 2\alpha) + \gamma\chi(\alpha = 1 \wedge \beta \ge 1) + \gamma^2\chi(1+\beta \ge 2\alpha)$; $\rho = \min(1+\beta, 2\alpha)$.

Твердження 2. Якщо виконано умови збіжності методу (20) у середньоквадратичному для послідовностей $\{\gamma_n\}$, $\{\varepsilon_n\}$ вигляду (26), зокрема, $\alpha \in (0,1]$, $\beta > 0$, і при $\alpha = 1$ виконується нерівність $2\gamma w_{\min} > \min(1, \beta)$, тоді

$$\overline{\lim_{n\to\infty}} n^{\theta} M\{\varDelta_n\} \le \frac{Cm}{2\gamma w_{\min} - \min(1,\beta)\chi(\alpha = 1)} < \infty, \tag{29}$$

де $\theta = \min(1+\beta-\alpha, \alpha)$ – порядок швидкості збіжності.

В оцінці (29) точне значення константи C не відоме, тому отримати кількісне значення величини швидкості збіжності ϑ не можливо. Тим не менше, можна проаналізувати якісний вплив параметрів методу на величину швидкості збіжності та отримати їхні початкові значення, наближені до оптимальних. Оптимальними будуть такі значення параметрів, які мінімізують вираз ϑ в оцінці (29).

Враховуючи, що $C \sim |D| N_{\max}(w_{\max} - w_{\min})\sigma^2_{\max} > 0$, з (29) випливає, що із зростанням кількості гравців $|D|$, кількості стратегій $N_{\max}$, міри неоднорідності середовища $w_{\max} - w_{\min}$ та значення другого моменту $\sigma^2_{\max}$ випадкових виграшів ξ величина швидкості збіжності ігрового методу до вирівнювальних стратегій зменшується. Зростання величини $w_{\min}$ призводить до зростання швидкості збіжності методу.

Враховуючи, що $[w_{\min}, w_{\max}] \in [v_{\min}, v_{\max}]$, де $v_{\min} = \min_{i\in D}\min_{u^{D_i}\in U^{D_i}} v(u^{D_i})$, $v_{\max} = \max_{i\in D}\max_{u^{D_i}\in U^{D_i}} v(u^{D_i})$, з (29) випливає, що у середньому швидкість збіжності методу (20) з обміном інформацією є більшою від швидкості збіжності методу без обміну інформацією.

Якщо нормування коефіцієнтів $\lambda^i(k)$ згорток функцій W_n^i відсутнє $\forall k \in \tilde{D}_i$, $\forall i \in D$ (в загальному випадку воно не є обов'язковим), тобто якщо функції W_n^i задати у початковому вигляді (14), то за однакових значень коефіцієнтів згорток $\alpha^i = 1/|D|$, отримаємо оцінки $w_{\min} = |D|^{-1}|\tilde{D}|_{\min} v_{\min}$ та $w_{\max} = |D|^{-1}|\tilde{D}|_{\max} v_{\max}$, де $|\tilde{D}|_{\min} = \min_{i\in D}|\tilde{D}_i|$, $|\tilde{D}|_{\max} = \max_{i\in D}|\tilde{D}_i|$. Тоді з (29) видно, що зростання $|\tilde{D}|_{\min}$, тобто розширення потужності множини гравців, задіяних в обміні інформацією, призводить до зростання швидкості збіжності ігрового методу. Нерівномірний розподіл на коаліції гравців, що виникає із зростанням $|\tilde{D}|_{\max}$ та із зменшенням $|\tilde{D}|_{\min}$, призводить до зменшення швидкості збіжності. Серед можливих фіксованих розподілів на коаліції максимальна швидкість збіжності досягається для рівновеликих коаліцій гравців при $|\tilde{D}|_{\min} = |\tilde{D}|_{\max}$. Найбільшої швидкості збіжності досягають для повнозв’язної гри, коли обмін інформацією відбувається між усіма гравцями $|\tilde{D}|_{\min} = |\tilde{D}|_{\max} = |D|$. З останнього та оцінки (29) випливає, що для гри без обміну інформацією при $|\tilde{D}|_{\min} = |\tilde{D}|_{\max} = 1$ досягають найменшої швидкості збіжності.

Оскільки у цій моделі гри множини $\tilde{D}_i$ та D_i пов’язані між собою, то розширення множин $\tilde{D}_i$ можливе за рахунок розширення множин D_i, а отже, за рахунок збільшення кількості спільних стратегій гравців, що визначають виграші i-го гравця $\forall i \in D$. Зростання ж кількості стратегій призводить до зменшення швидкості збіжності гри. Тому про ефективність обміну інформацією можна говорити тільки для фіксовано визначених коаліцій гравців.

Наслідок. В умовах твердження 2 параметр θ, який задає порядок швидкості збіжності методу (20), задовольняє нерівність $\theta \le 1$. Максимальний порядок швидкості збіжності методу (20) становить $\theta = 1$ і досягається при $\alpha = 1$, $\beta \ge 1$.

Висновки

Отримані теоретичні результати визначають достатні умови асимптотичної збіжності рекурентного ігрового методу з обміном інформацією до оптимальних за Парето стратегій. Встановлено, що швидкість збіжності ігрового методу значною мірою визначається розмірністю ігрової задачі, початковими значеннями регульованих параметрів методу та характеристиками випадкового середовища прийняття рішень. Із збільшенням кількості гравців, стратегій та значення дисперсії випадкових виграшів швидкість збіжності ігрового методу зменшується. В отриманих оцінках не виявлено впливу закону розподілу випадкових виграшів та початкового наближення змішаних стратегій на одиничному симплексі на умови та швидкість збіжності ігрового методу.

Встановлено, що обмін поточними виграшами в межах локальних коаліцій гравців призводить до зростання величини швидкості збіжності (зменшення її константної складової). В загальному, методи з обміном інформацією мають більшу величину швидкості збіжності, ніж методи без обміну інформацією.

Перспективним розвитком цього напрямку дослідження є вивчення інших способів взаємодії гравців, визначення станів рівноваги та умов збіжності ігрових методів з обміном інформацією у межах моделі гри з відмовами гравців та завадами передавання інформації.

1. Gerhard Weiss and Sandip Sen, editors. Adaptation and Learning in Multiagent Systems. Springer Verlag. – Berlin, 1996. 2. Wooldridge M. An Introduction to Multiagent Systems. John Wiley & Sons. – Chichester, England, 2002. 3. Доманский В.К. Стохастические игры // Математические вопросы кибернетики. – 1988. – № 1. – С. 26–49. 4. Fudenberg D., Levine D.K. The Theory of Learning in Games. – MIT Press, 1998. 5. Stone P. Layered Learning in Multiagent Systems. – MIT Press, 2000. 6. Воробьев Н.Н. Основы теории игр: Бескоалиционные игры. – М.: Наука, 1984. 7. Мулен Э. Теория игр с примерами из математической экономики. – М.: Мир, 1985. 8. Назин А.В., Позняк А.С. Адаптивный выбор вариантов: Рекуррентные алгоритмы. – М.: Наука, 1986. 9. Опойцев В.И. Равновесие и устойчивость в моделях коллективного поведения. – М.: Наука, 1977. 10. Батищев Д.И. Методы оптимального проектирования. – М.: Радио и связь, 1984. 11. Вазан М. Стохастическая аппроксимация. – М.: Мир, 1972. 12. Невельсон М.Б., Хасьминский Р.З. Стохастическая оптимизация и рекуррентное оценивание. – М.: Наука, 1972. 13. Цыпкин Я.З., Позняк А.С. Рекуррентные алгоритмы оптимизации в условиях неопределенности // Итоги науки и техники. Сер. Техническая кибернетика. – 1989. – Т. 16. – С. 3–70.