Metaheurystyczna aproksymacja stanu równowagi Stackelberga w grach wielokrokowych z niepełną informacją

Jan Karwowski Jacek Mańdziuk

Zakład Sztucznej Inteligencji i Metod Obliczeniowych
Wydział Matematyki i Nauk Informacyjnych PW

2021

Cel badań

Zaproponowanie metody, która będzie w stanie przybliżać strategię lidera ze Stanu Równowagi Stackelberga dla dużych gier w postaci ekstensywnej z niepełną informacją o sumie niezerowej.

Skończona gra dwuosobowa

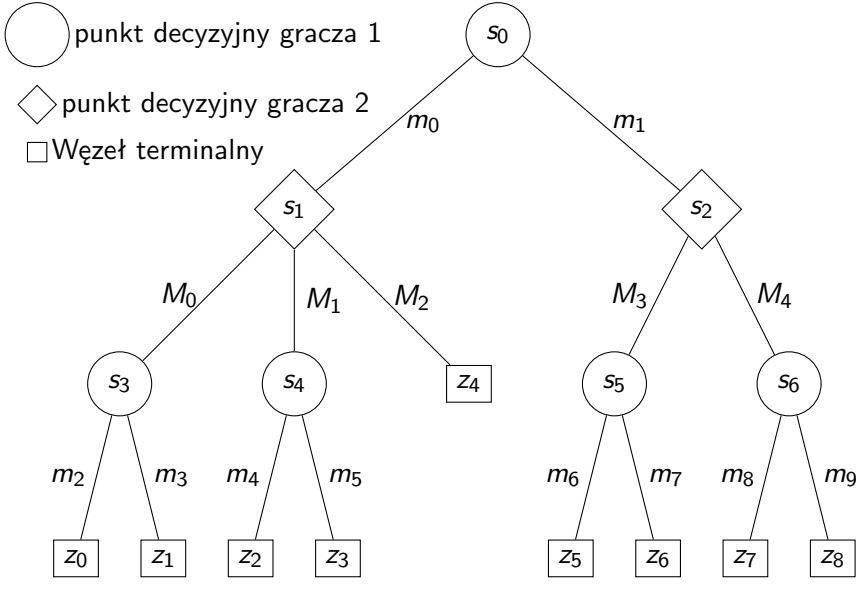
Macierze gry

	A	B	C
α	2	3.1	-1
β	1	2.5	1.8
γ	-1	-3	-2
δ	-5	10	-5

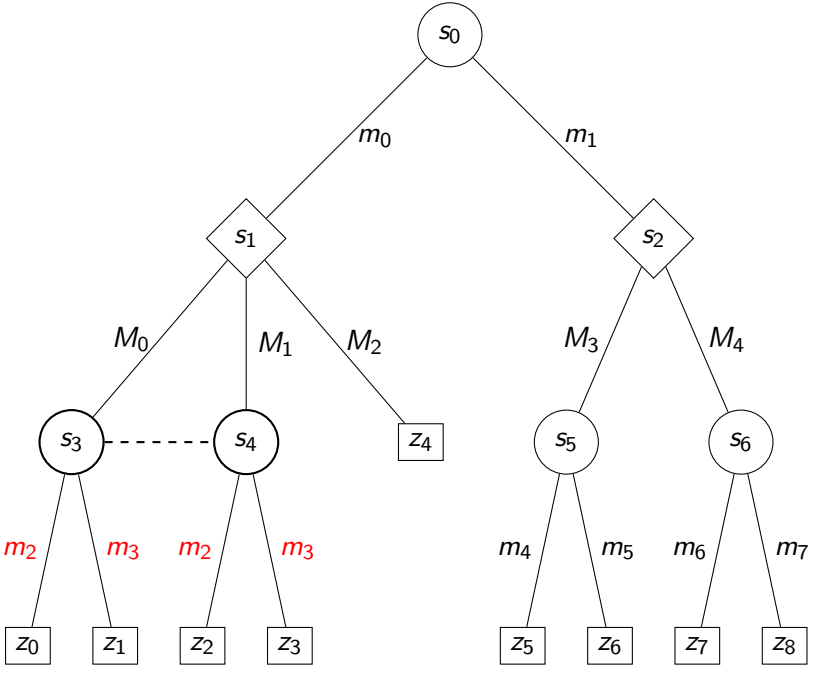
	A	B	C
α	-0.5	-1.2	-1
β	-1	-3	-5
γ	0	-2	-2
δ	-7	0	-12

- Dwóch graczy, jeden wybiera kolumnę, drugi wiersz
- Co jeden gracz wie o ruchach drugiego?
- Czy gracze wykonują ruchy jednocześnie?

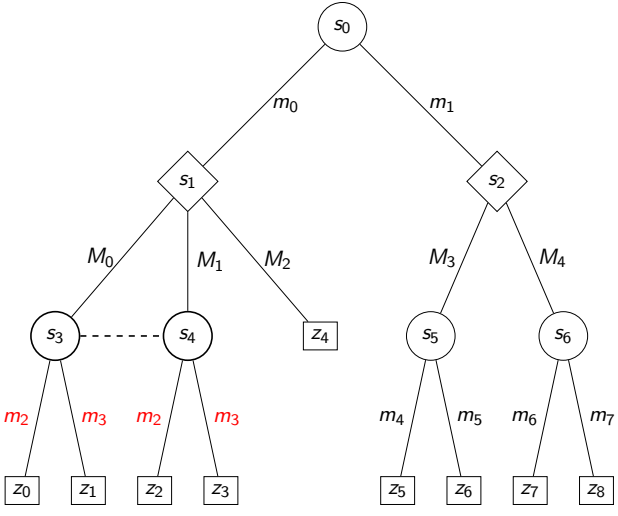
Gra w postaci ekstensywnej



Niepełna informacja (imperfect information)



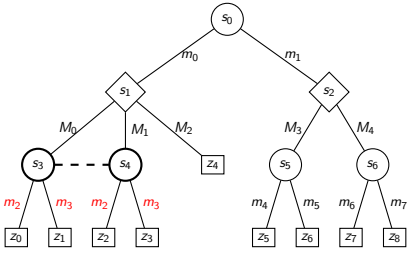
Niepełna informacja (imperfect information)



- Zbiory informacyjne – definiują jak dużo informacji ma gracz podejmując decyzję
- W zbiorze informacyjnym gracz zawsze ma ten sam zestaw ruchów.
- Dodatkowo: Doskonała pamięć (perfect recall) – potomkowie sekwencji rozróżnialnych są rozróżnialni.

Strategie I

Strategie w grze w postaci ekstensywnej



- Strategia prosta przypisanie każdemu ze zbiorów informacyjnych dokładnie jednego ruchu do zagrania.
- $I_1 = \{s_0\}$ $I_2 = \{s_3, s_4\}$ $I_3 = \{s_5\}$ $I_4 = \{s_6\}$ $I_5 = \{s_1\}$ $I_6 = \{s_2\}$
- Przykład:
  - Gracz 1: $I_1 \to m_1, I_2 \to m_3, I_3 \to m_4, I_5 \to m_7$
  - Gracz 2: $I_5 \to m_4, I_6 \to m_8$

Strategie II

Strategie mieszane (mixed strategy)

- Gracz decyduje o rozkładzie prawdopodobieństwa strategii prostych.
- W pojedynczej rozgrywce wykonywane są niezależne losowania strategii z wybranego rozkładu.
- Profil strategii (para strategii dwóch graczy) determinuje wartość oczekiwaną gry.
- Możemy traktować strategie mieszane jako nadzbiór prostych.

Cel badań

Zaproponowanie metody, która będzie w stanie przybliżać strategię lidera ze Stanu Równowagi Stackelberga dla dużych gier w postaci ekstensywnej z niepełną informacją o sumie niezerowej.

Stany równowagi I

Stan równowagi to taki profil strategii, w którym żadnemu z graczy, przy pewnych założeniach, nie opłaca się zmieniać decyzji.

Równowaga Nasha

- Najbardziej znana równowaga
- Symetryczni gracze
- Warunek: żaden z graczy nie może zyskać, jeśli zmieni swoją strategię, a pozostali gracze zostaną przy swoich strategiach.
- Jakie strategie rozważamy?

Stany równowagi II

Równowaga Nasha – strategie proste

	A	B	C
α	10	7	5
β	4	3	2
γ	4	9	4

	A	B	C
α	10	6	2
β	8	5	1
γ	1	4	1

- Profil strategii (A, α) jest Równowagą Nasha

Równowaga Nasha – strategie mieszane

	A	B
α	-1	1
β	1	-1

	A	B
α	1	-1
β	-1	1

- Nie ma równowagi Nasha w strategiach prostych
- Profil

$$(\{p(A) = 0.5, p(B) = 0.5\}, \{p(\alpha) = 0.5, p(\beta) = 0.5\})$$

jest równowagą Nasha

Równowaga Stackelberga

Definicja

- Asymetryczni gracze
  - Lider
  - Naśladowca (follower)
- Lider wybiera strategię i ujawnia ją naśladowcy
- Naśladowca wybiera strategię, która maksymalizuje jego wynik gry przy ustalonej strategii lidera (przypomnienie: gry o sumie niezerowej)
- Równowaga: taka strategia lidera, która maksymalizuje wynik lidera przy założeniu optymalnej odpowiedzi naśladowcy
- Zwykle w strategiach mieszanych

Równowaga Stackelberga

Definicja

- Asymetryczni gracze
  - Lider
  - Naśladowca (follower)
- Lider wybiera strategię i ujawnia ją naśladowcy
- Naśladowca wybiera strategię, która maksymalizuje jego wynik gry przy ustalonej strategii lidera (przypomnienie: gry o sumie niezerowej)
- Równowaga: taka strategia lidera, która maksymalizuje wynik lidera przy założeniu optymalnej odpowiedzi naśladowcy
- Zwykle w strategiach mieszanych

Nash

Symetria definicji: żaden z graczy nie może zyskać, jeśli zmieni swoją strategię, a pozostali gracze zostaną przy swoich strategiach.

Równowaga Stackelberga

Definicja

- Asymetryczni gracze
  - Lider
  - Naśladowca (follower)
- Lider wybiera strategię i ujawnia ją naśladowcy
- Naśladowca wybiera strategię, która maksymalizuje jego wynik gry przy ustalonej strategii lidera (przypomnienie: gry o sumie niezerowej)
- Równowaga: taka strategia lidera, która maksymalizuje wynik lidera przy założeniu optymalnej odpowiedzi naśladowcy
- Zwykle w strategiach mieszanych

Formalna definicja

Układ strategii (δ_l^*, δ_f^*) taki, że:

$$\begin{cases} \delta_l^* = \arg\max_{\delta_l \in \Delta_l} u_l(\delta_l, BR(\delta_l)) \\ \delta_f^* = BR(\delta_l^*) \\ BR(\delta_l) = \arg\max_{\delta_f \in \Delta_f} u_f(\delta_l, \delta_f) \end{cases}$$

$u_l()$, $u_f()$ – wartość oczekiwana wypłaty lidera/naśladowcy
$\delta_l, delta_f$ – strategie proste graczy
Δ_l, Δ_f – zbiory wszystkich strategii

Ważna obserwacja

W przypadku naśladowcy nie ma strategii mieszanych, które dałyby lepszy wynik niż strategie proste.

Silna Równowaga Stackelberga

Równowaga Stackelberga

Układ strategii (δ_l^*, δ_f^*) taki, że:

$$\begin{cases} \delta_l^* = \arg\max_{\delta_l \in \Delta_l} u_l(\delta_l, BR(\delta_l)) \\ \delta_f^* = BR(\delta_l^*) \\ BR(\delta_l) = \arg\max_{\delta_f \in \Delta_f} u_f(\delta_l, \delta_f) \end{cases}$$

Gra

L	A	B	C
α	2	5	1
β	1	4	3
γ	2	3	3

N	A	B	C
α	2	**4**	2
β	1	2	1
γ	1	**4**	3

Niejednoznaczność

W powyższej definicji *BR* jest źle zdefiniowana – niejednoznaczność, gdy lider wybierze B.

Silna Równowaga Stackelberga

$\overline{BR}$ – funkcja przypisuje zbiór najlepszych odpowiedzi

$$\begin{cases} \delta_l^* = \arg\max_{\delta_l \in \Delta_l, \delta_f \in \overline{BR}(\delta_l)} u_l(\delta_l, \delta_f) \\ \overline{BR}(\delta_l) = \{\delta_f \in \Delta_f | \forall \delta_f' \in \Delta_f \, u_f(\delta_l, \delta_f) \geqslant u_f(\delta_l, \delta_f')\} \end{cases}$$

G. Leitmann. "On generalized Stackelberg strategies". In: *Journal of Optimization Theory and Applications* 26.4 (Dec. 1978), pp. 637–643. ISSN: 1573-2878

W tym momencie naśladowca zawsze wybierze α.

Istniejące wdrożenia I

Security Games

Arunesh Sinha et al. "Stackelberg Security Games: Looking Beyond a Decade of Success". In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, July 2018, pp. 5494–5501

Christopher Kiekintveld et al. "Computing optimal randomized resource allocations for massive security games". In: *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. 2009, pp. 689–696

Manish Jain et al. "Software assistants for randomized patrol planning for the LAX airport police and the federal air marshal service". In: *Interfaces* 40.4 (2010), pp. 267–290

Bo An et al. "A deployed quantal response-based patrol planning system for the US Coast Guard". In: *Interfaces* 43.5 (2013), pp. 400–420

Istniejące wdrożenia II

Green Security Games

Fei Fang et al. "PAWS — A Deployed Game-Theoretic Application to Combat Poaching". In: *AI Magazine* 38.1 (Mar. 2017), p. 23. ISSN: 0738-4602

Matthew Paul Johnson, Fei Fang, and Milind Tambe. "Patrol Strategies to Maximize Pristine Forest Area.". In: *AAAI Conference on Artificial Intelligence*. 2012, pp. 295–301

Kai Wang et al. "Scalable Game-Focused Learning of Adversary Models: Data-to-Decisions in Network Security Games.". In: *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems*. 2020, pp. 1449–1457

Istniejące wdrożenia III

Cechy wspólne

- Stosunkowo małe gry lub duże gry o bardzo specyficznej strukturze
- Rozwiązania wyspecjalizowane do jednej klasy gier

Cel badań (raz jeszcze)

Zaproponowanie metody, która będzie w stanie przybliżać strategię lidera ze Stanu Równowagi dla dużych gier w postaci ekstensywnej z niepełną informacją o sumie niezerowej.

Postać normalna (macierzowa) – przypomnienie

	A	B	C
α	2	3.1	-1
β	1	2.5	1.8
γ	-1	-3	-2
δ	-5	10	-5

	A	B	C
α	-0.5	-1.2	-1
β	-1	-3	-5
γ	0	-2	-2
δ	-7	0	-12

Postać najłatwiejsza z punktu widzenia budowy metody obliczeniowego poszukiwania strategii w grach.

Podejścia do gier w postaci normalnej

Rozwiązywanie wielu programów liniowych

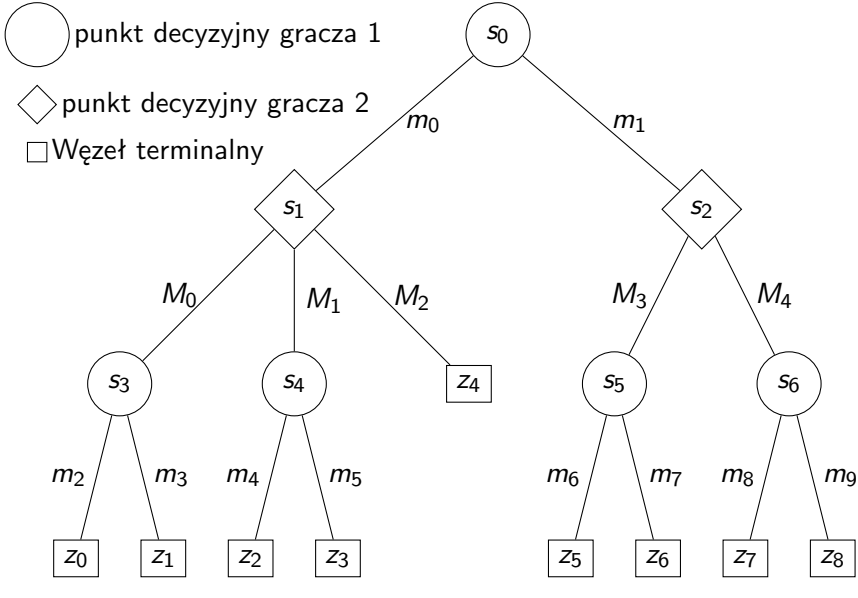
Vincent Conitzer and Tuomas Sandholm. "Computing the optimal strategy to commit to". In: *Proceedings of the 7th ACM conference on Electronic commerce*. ACM. 2006, pp. 82–90.

- Poszukiwanie Równowagi Stackelberga poprzez rozwiązywanie wielu programów liniowych
- Możliwe dzięki obserwacji, że wystarczy rozważyć strategie proste naśladowcy.
- Autorzy wykonują pełny przegląd wszystkich strategii naśladowcy i dla każdej z nich budują program liniowy, którzy znajduje najlepszą strategię lidera, dla której ta strategia naśladowcy jest optymalną odpowiedzią.
- Program liniowy budowany jest dla gry w postaci normalnej.

DOBSS – Jeden MILP zamiast wielu LP.

Praveen Paruchuri et al. "Playing games for security: an efficient exact algorithm for solving Bayesian Stackelberg games". In: *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*. 2008, pp. 895–902

Ale co z grami w postaci ekstensywnej?



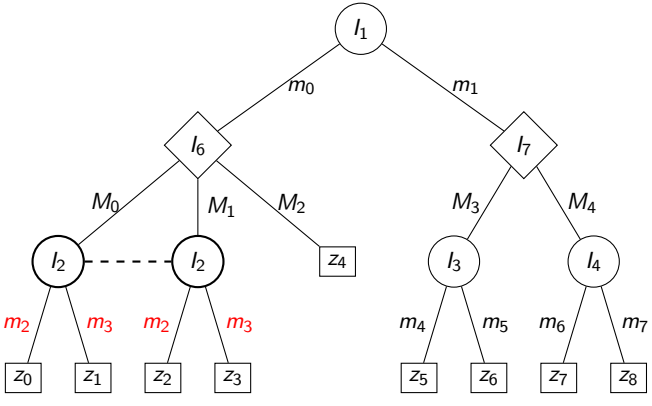
Transformacja z postaci ekstensywnej do postaci normalnej I

John C. Harsanyi. "Games with Incomplete Information Played by "Bayesian" Players, I–III Part I. The Basic Model". In: *Papers in Game Theory* (1982), pp. 115–138

- Każdej strategii prostej odpowiada wiersz/kolumna macierzy
- W macierzy wielokrotnie występują wyniki dla tych samych liści drzewa gry

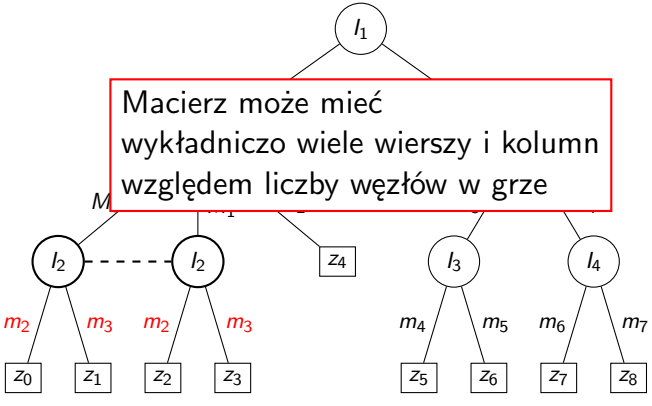
Transformacja do postaci normalnej – przykład

	$l_1 \to m_0, l_2 \to m_2$ $l_3 \to m_4, l_4 \to m_6$	$l_1 \to m_0, l_2 \to m_3$ $l_3 \to m_4, l_4 \to m_6$	$l_1 \to m_0, l_2 \to m_2$ $l_3 \to m_5, l_4 \to m_6$	...
$l_6 \to M_2, l_7 \to M_3$	z_4	z_4	z_4	
$l_6 \to M_2, l_7 \to M_4$	z_4	z_4	z_4	
...				



Transformacja do postaci normalnej – przykład

	$l_1 \to m_0, l_2 \to m_2$ $l_3 \to m_4, l_4 \to m_6$	$l_1 \to m_0, l_2 \to m_3$ $l_3 \to m_4, l_4 \to m_6$	$l_1 \to m_0, l_2 \to m_2$ $l_3 \to m_5, l_4 \to m_6$	...
$l_6 \to M_2, l_7 \to M_3$	z_4	z_4	z_4	
$l_6 \to M_2, l_7 \to M_4$	z_4	z_4	z_4	
...				



Postać sekwencyjna gry

- Jeszcze bardziej kompaktowa reprezentacja macierzowa.
- Macierz indeksowana potencjalnymi sekwencjami ruchów gracza
- Wymiary macierzy liniowe względem liczby węzłów w drzewie
- Tylko dla gier z doskonałą pamięcią

Metody wykorzystujące postać sekwencyjną

Branislav Bošanský and Jiří Čermak. "Sequence-Form Algorithm for Computing Stackelberg Equilibria in Extensive-Form Games". In: *29th AAAI Conference on Artificial Intelligence*. Ed. by Blai Bonet and Sven Koenig. AAAI Press, 2015, pp. 805–811. ISBN: 978-1-57735-698-1

Jiří Čermák et al. "Using Correlated Strategies for Computing Stackelberg Equilibria in Extensive-Form Games". In: *30th AAAI Conference on Artificial Intelligence*. 2016, pp. 439–445

Wymagają przechowania całej reprezentacji gry na raz w pamięci!

Metody przybliżone oparte o uproszczenie gry

Zwinięcie dużych fragmentów gry do jednopoziomowych poddrzew

Jakub Černý, Branislav Bošanský, and Christopher Kiekintveld. "Incremental Strategy Generation for Stackelberg Equilibria in Extensive-Form Games". In: *Proceedings of the 2018 ACM Conference on Economics and Computation, Ithaca, NY, USA, June 18-22, 2018*. Ed. by Éva Tardos, Edith Elkind, and Rakesh Vohra. ACM, 2018, pp. 151–168

Reprezentacja strategii z użyciem niedużego algorytmu skończeniestanowego

Jakub Cerný, Branislav Bosanský, and Bo An. "Finite State Machines Play Extensive-Form Games". In: *EC '20: The 21st ACM Conference on Economics and Computation, Virtual Event, Hungary, July 13-17, 2020*. Ed. by Péter Biró et al. ACM, 2020, pp. 509–533. ISBN: 978-1-4503-7975-5

Jak można usprawnić poszukiwanie równowagi Stackelberga

- Nie wszystkie strategie naśladowcy są istotne
- Niektóre obszary gry nigdy nie będą odwiedzone przez rozsądne strategie
- W postaci ekstensywnej często występuje korelacja pomiędzy wynikami w sąsiednich gałęziach
- Ukierunkowane próbkowanie drzewa gry, zamiast pełnego rozwinięcia od początku
- Metoda przybliżona

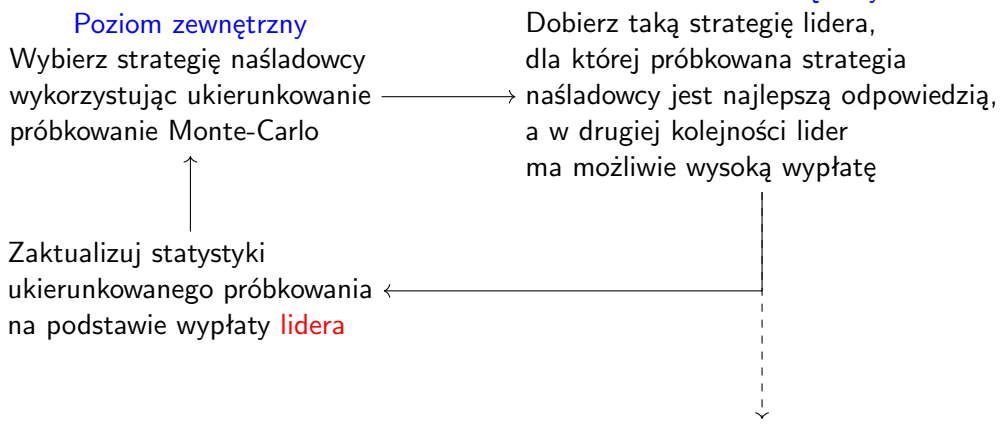
Jan Karwowski and Jacek Mańdziuk. "Double-oracle sampling method for Stackelberg Equilibrium approximation in general-sum extensive-form games". In: *34th AAAI Conference on Artificial Intelligence (AAAI-20)*. 2020

Podejście O2UCT Rozbicie na dwa poziomy

- Oddzielne rozpatrywanie poziomu wewnętrznego i zewnętrznego
- Optymalizacja zewnętrzna – strategie naśladowcy (strategie proste)
- Optymalizacja wewnętrzna – strategie lidera (strategie mieszane)
- Dla każdej rozważanej strategii zewnętrznej uruchomienie strategii wewnętrznej
- Poziom zewnętrzny: naśladowca (wewnętrzna optymalizacja w sformułowaniu SE)
- Poziom wewnętrzny: lider
- **O2UCT** = Double Oracle + UCT

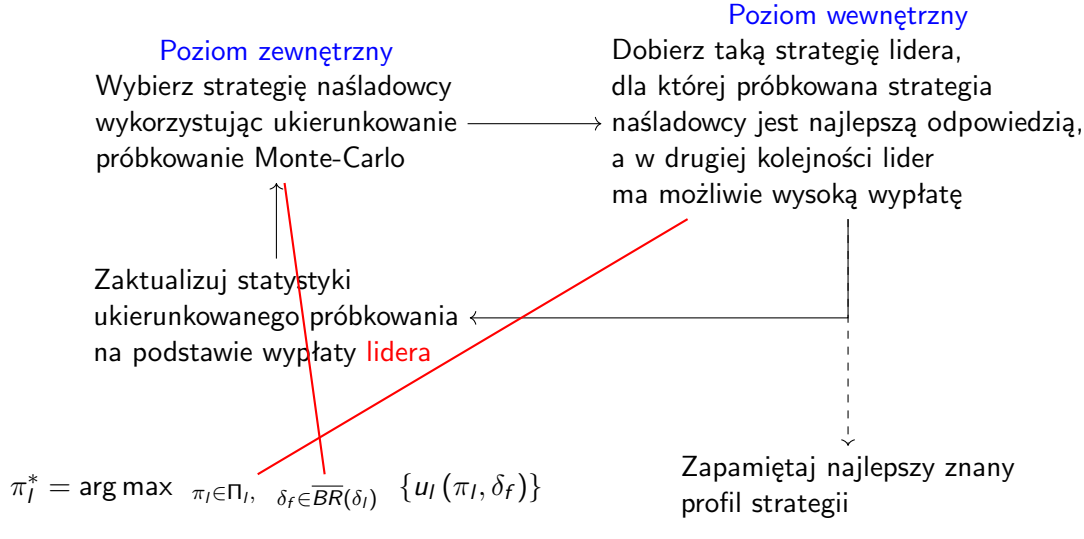
$$\begin{cases} \delta_l^* = \arg\max_{\delta_l \in \Delta_l, \delta_f \in \overline{BR}(\delta_l)} u_l(\delta_l, \delta_f) \\ \overline{BR}(\delta_l) = \{\delta_f \in \Delta_f | \forall \delta_f' \in \Delta_f u_f(\delta_l, \delta_f) \geqslant u_f(\delta_l, \delta_f')\} \end{cases}$$

O2UCT – ogólny schemat



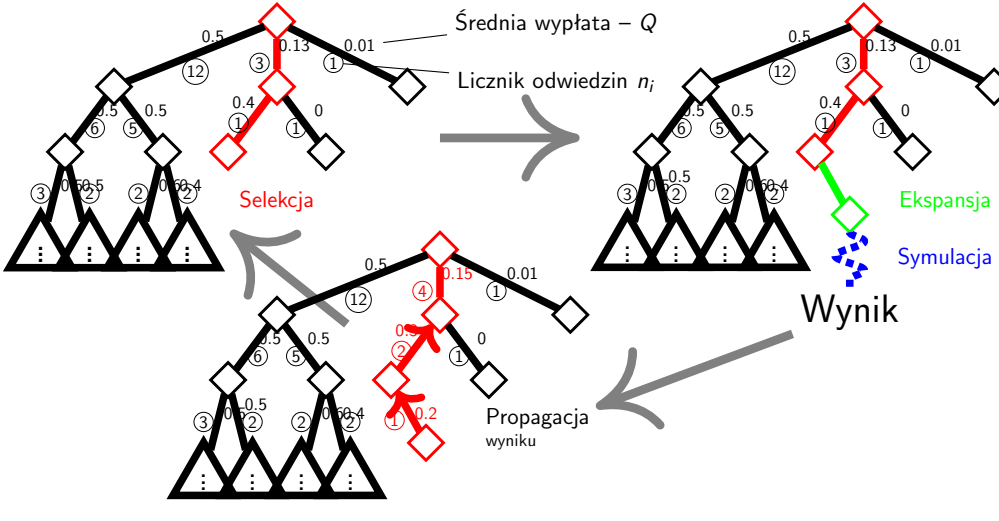
$$\pi_l^* = \arg\max_{\pi_l \in \Pi_l,\ \ \delta_f \in \overline{BR}(\delta_l)} \{u_l(\pi_l, \delta_f)\}$$

O2UCT – ogólny schemat



O2UCT – poziom zewnętrzny

Monte Carlo Tree Search (MCTS)



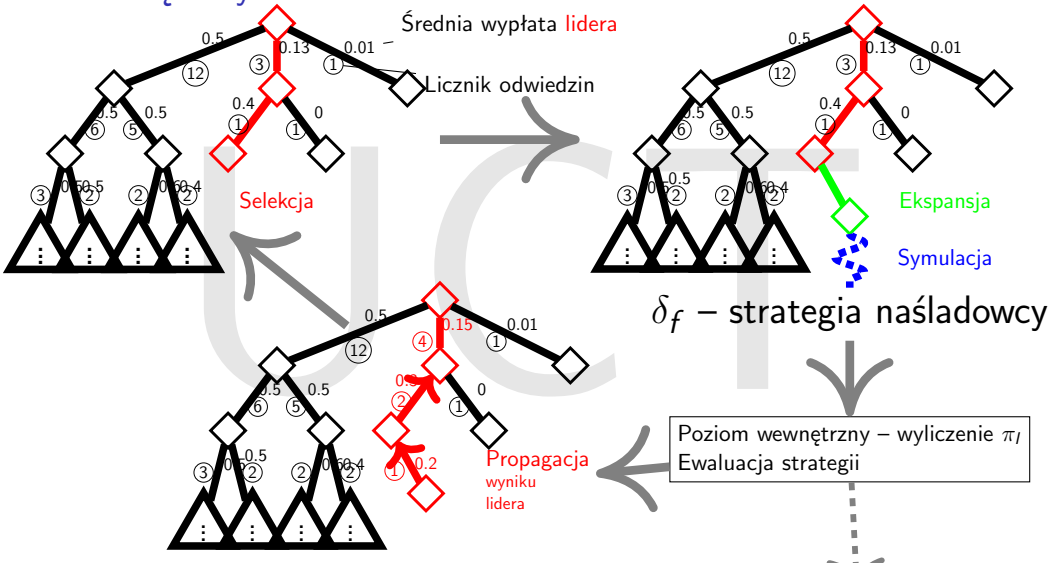
Upper confidence bound applied to trees (UCT)

Selekcja: $\arg\max_i \left\{ Q_i + c\sqrt{\log N / n_i} \right\}, \qquad N = \sum_i n_i$

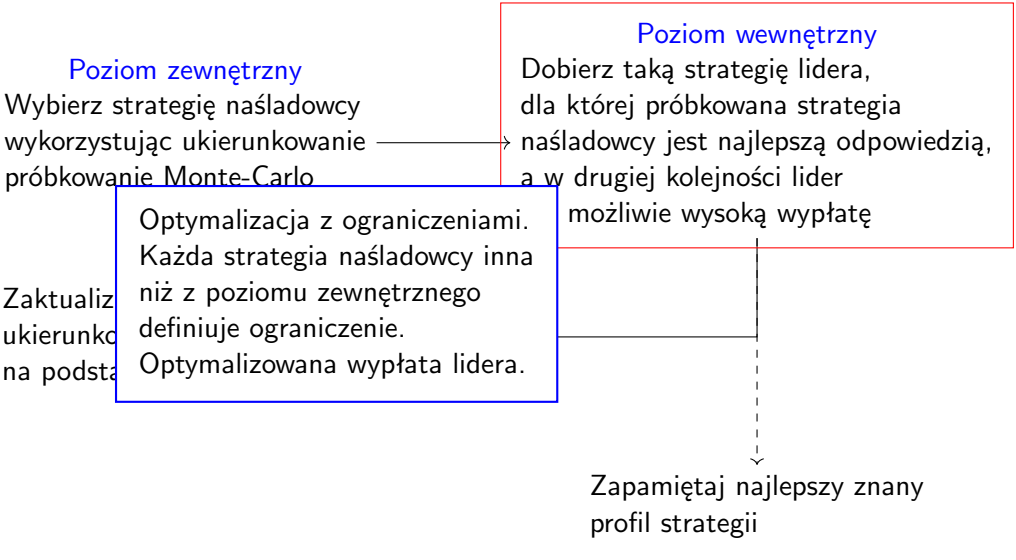
Poziom zewnętrzny I

- Budujemy grę pomocniczą dla jednego gracza, w której wynikiem jest zredukowana strategia prosta naśladowcy
- Z użyciem UCT wybieram kolejne strategie do przetworzenia w poziomie wewnętrznym
- To jeszcze nie daje wyniku do propagacji w UCT
- Obliczamy strategię lidera (poziom wewnętrzny)
- Propagujemy wartość wypłaty **lidera** z rozwiązania wewnętrznego

Poziom zewnętrzny II



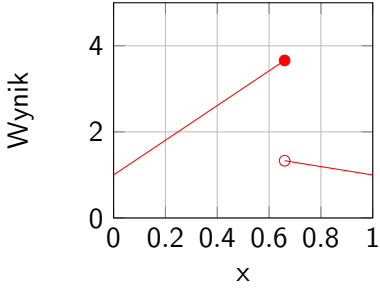
O2UCT – Poziom wewnętrzny



Ograniczenia – przykład

$$M_l = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 5 & 1 \\ B & 1 & 2 \end{array} \qquad M_f = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 1 & 2 \\ B & 3 & 1 \end{array}$$

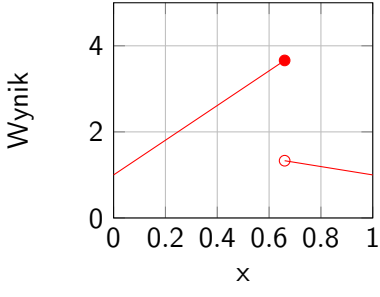
Strategia lidera parametryzowana x: $\{p(A) = x, p(B) = 1 - x\}$



Ograniczenia – przykład

$$M_l = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 5 & 1 \\ B & 1 & 2 \end{array} \qquad M_f = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 1 & 2 \\ B & 3 & 1 \end{array}$$

Strategia lidera parametryzowana x: $\{p(A) = x, p(B) = 1 - x\}$

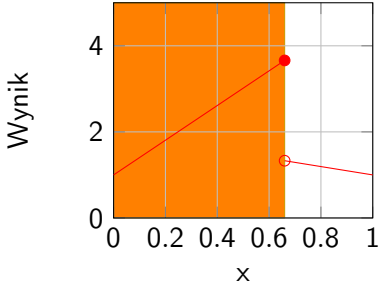


Wypłata lidera w zależności od strategii. Widoczny jest punkt nieciągłości, gdzie zmienia się odpowiedź naśladowcy.

Ograniczenia – przykład

$$M_l = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 5 & 1 \\ B & 1 & 2 \end{array} \qquad M_f = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 1 & 2 \\ B & 3 & 1 \end{array}$$

Strategia lidera parametryzowana x: $\{p(A) = x, p(B) = 1 - x\}$

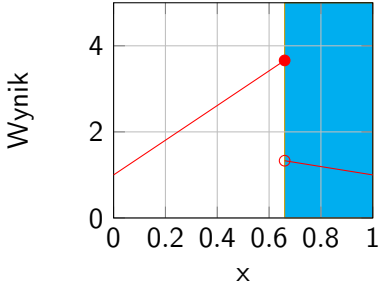


Gdy zewnętrzne próbkowanie zwróciło ruch α, obszar dopuszczalny $x \leqslant 2/3$

Ograniczenia – przykład

$$M_l = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 5 & 1 \\ B & 1 & 2 \end{array} \qquad M_f = \begin{array}{c|cc} & \alpha & \beta \\ \hline A & 1 & 2 \\ B & 3 & 1 \end{array}$$

Strategia lidera parametryzowana x: $\{p(A) = x, p(B) = 1 - x\}$



Gdy zewnętrzne próbkowanie zwróciło ruch β, obszar dopuszczalny $x > 2/3$

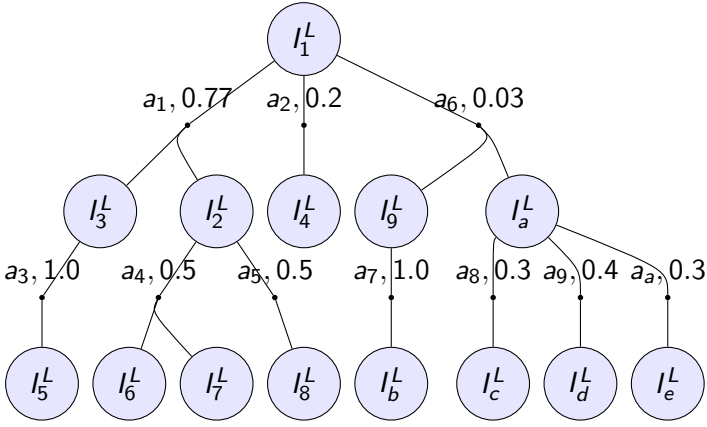
Podwójna wyrocznia (Double Oracle)

- Metoda polegająca na ustalaniu strategii w grze dla dwóch graczy poprzez:
  - Posiadanie wyroczni dającej optymalne zagranie pierwszego gracza przy ustalonej strategii drugiego
  - Posiadanie wyroczni dającej optymalne zagranie drugiego gracza przy ustalonej strategii pierwszego
  - Naprzemienne uruchamianie tych wyroczni aż do otrzymania stanu równowagi
- W O2UCT wyrocznia lidera jest bardziej skomplikowana
  - Nie daje optymalnego wyniku
  - Przyrostowo aktualizuje strategie

Poziom wewnętrzny

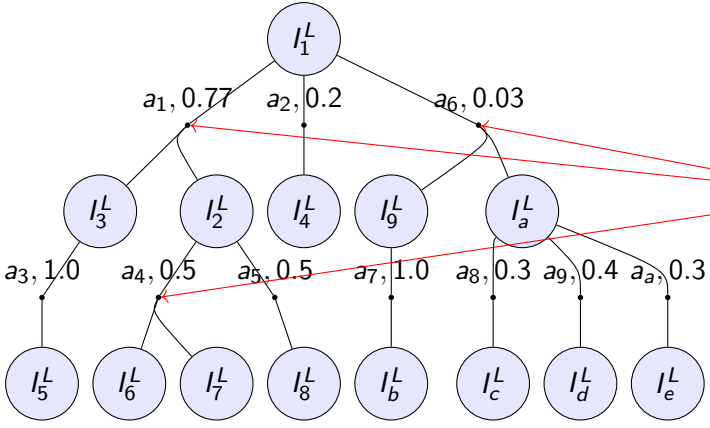
- Dla zadanej z zewnętrznego poziomu strategii naśladowcy zaproponuj strategię lidera
- Otrzymany profil strategii musi być dobrym kandydatem na równowagę
- Chcemy uniknąć materializacji całej gry w pamięci
  - Rozpoczynamy od analizy strategii prostej lidera
  - W trakcie, z użyciem próbkowana, dodajemy kolejne ruchy
  - Reprezentujemy strategię behawioralną jako drzewo

Strategia behawioralna (behavior strategy)



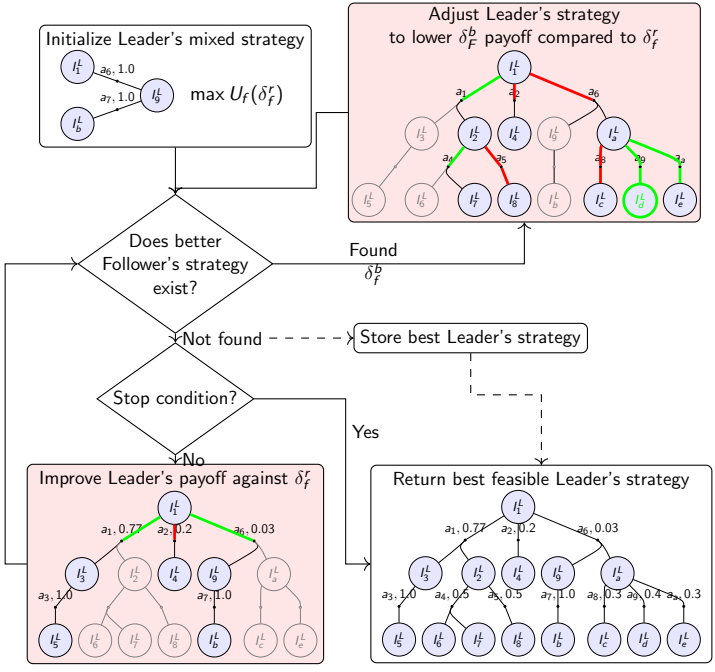
- Rozkład prawdopodobieństwa nad zbiorem akcji w każdym węźle decyzyjnym z osobna.
- Możliwe różne następstwa akcji w zależności od działań przeciwnika.
- Izomorficzne ze strategiami mieszanymi, jeśli gra z doskonałą pamięcią.

Strategia behawioralna (behavior strategy)

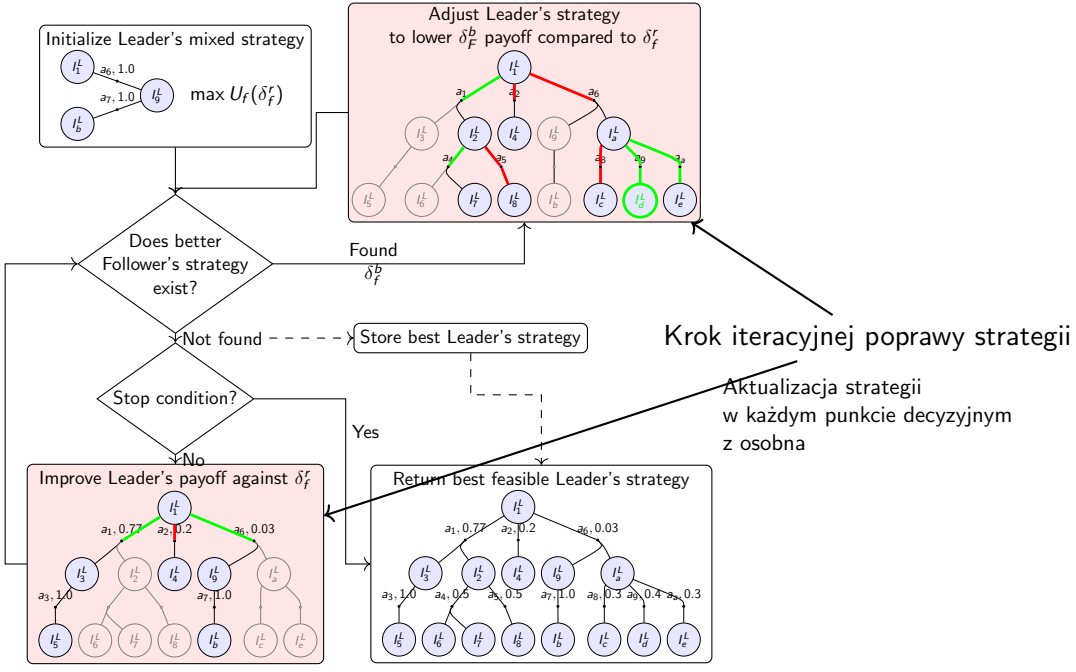


- Rozkład prawdopodobieństwa nad zbiorem akcji w każdym węźle decyzyjnym z osobna.
- Możliwe różne następstwa akcji w zależności od działań przeciwnika.
- Izomorficzne ze strategiami mieszanymi, jeśli gra z doskonałą pamięcią.

Poziom wewnętrzny – schemat działania

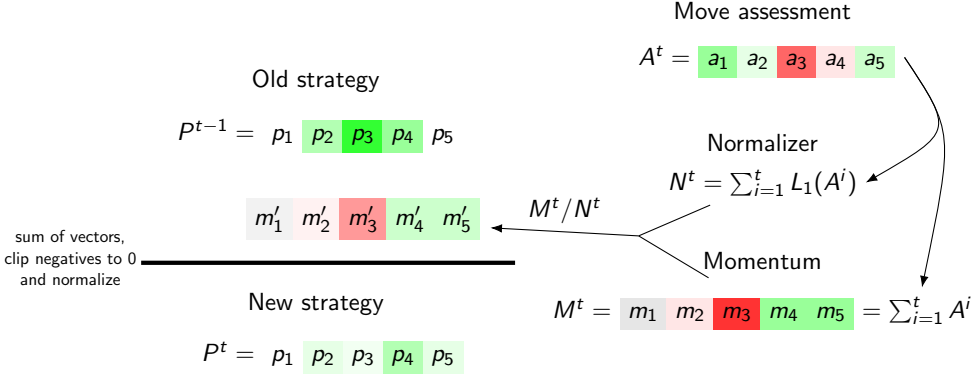


Poziom wewnętrzny – schemat działania



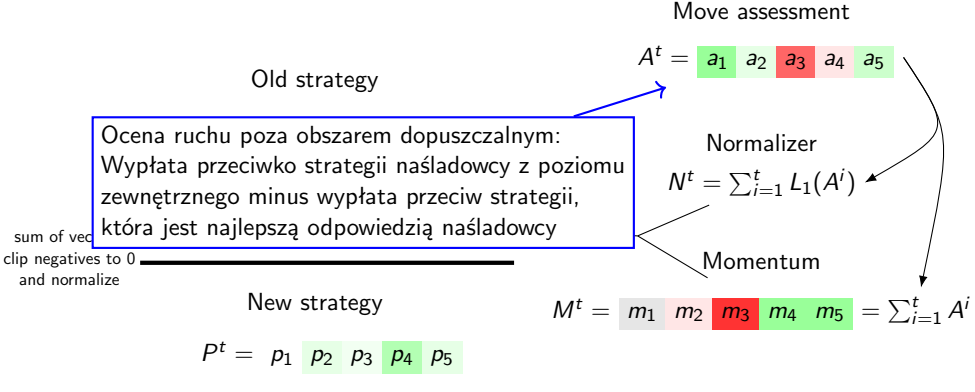
Aktualizacja strategii w węzłach decyzyjnych

- Utrzymujemy współczynnik normalizacji N, początkowo 0
- Utrzymujemy wektor zakumulowanego kierunku (moment) M
- W każdej iteracji aktualizujemy stosownie do oceny ruchów A w tym punkcie decyzyjnym



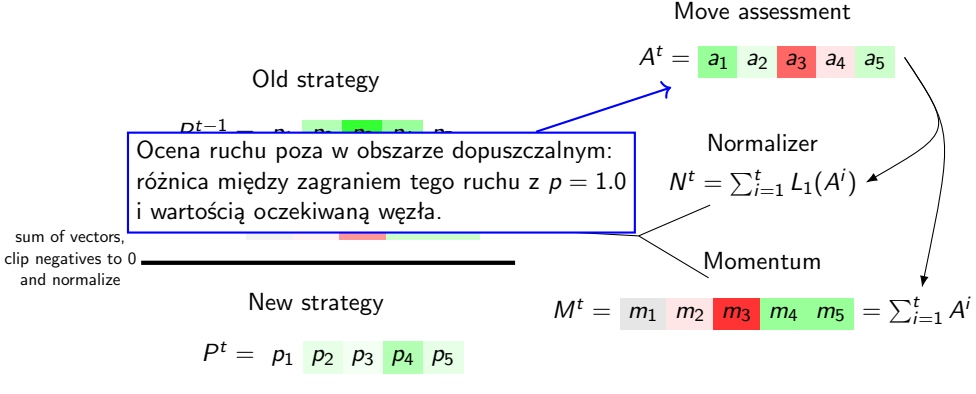
Aktualizacja strategii w węzłach decyzyjnych

- Utrzymujemy współczynnik normalizacji N, początkowo 0
- Utrzymujemy wektor zakumulowanego kierunku (moment) M
- W każdej iteracji aktualizujemy stosownie do oceny ruchów A w tym punkcie decyzyjnym



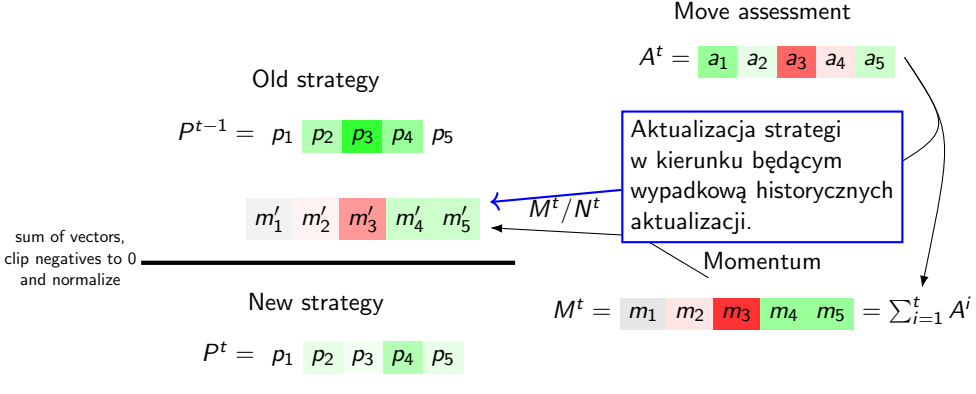
Aktualizacja strategii w węzłach decyzyjnych

- Utrzymujemy współczynnik normalizacji N, początkowo 0
- Utrzymujemy wektor zakumulowanego kierunku (moment) M
- W każdej iteracji aktualizujemy stosownie do oceny ruchów A w tym punkcie decyzyjnym



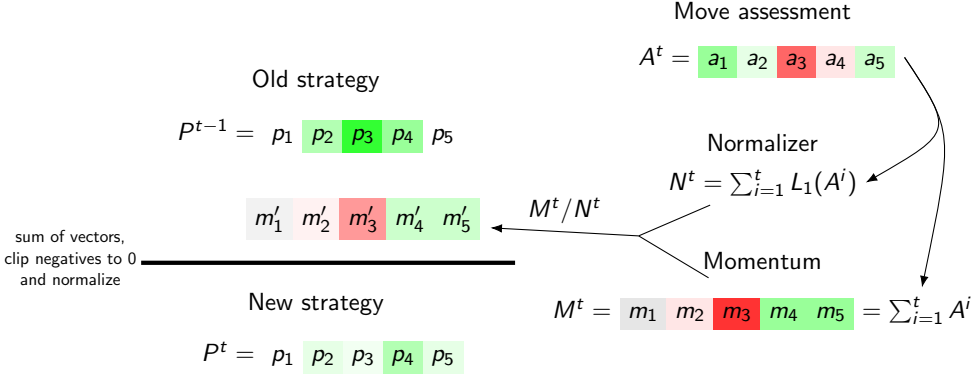
Aktualizacja strategii w węzłach decyzyjnych

- Utrzymujemy współczynnik normalizacji N, początkowo 0
- Utrzymujemy wektor zakumulowanego kierunku (moment) M
- W każdej iteracji aktualizujemy stosownie do oceny ruchów A w tym punkcie decyzyjnym



Aktualizacja strategii w węzłach decyzyjnych

- Utrzymujemy współczynnik normalizacji N, początkowo 0
- Utrzymujemy wektor zakumulowanego kierunku (moment) M
- W każdej iteracji aktualizujemy stosownie do oceny ruchów A w tym punkcie decyzyjnym

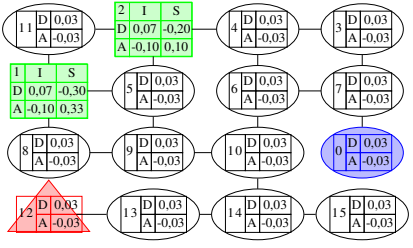


Różne usprawnienia

- Pamiętamy, które strategie naśladowcy były ostatnimi naruszonymi ograniczeniami i sprawdzamy je w pierwszej kolejności
- Aktualizacja tylko losowego poddrzewa strategii
- Odcinanie gałęzi, których prawdopodobieństwo zostało wyzerowane

Gry testowe I

Warehouse Games (WHG)



- Lider (obrońca, D) startuje w wierzchołku 0, naśladowca w trójkącie
- Naśladowca (atakujący, A) dostaje wypłatę za dotarcie do zielonych wierzchołków
- Lider może złapać naśladowcę w tym samym wierzchołku
- Gracze nie widzą siebie nawzajem, dopóki nie dojdzie do kontaktu
- Każdy wierzchołek ma przypisane kary za złapanie
- Losowe struktury grafów, nietrywialne
- Wypłaty bliskie sumie zerowej

Jan Karwowski and Jacek Mańdziuk. "A Monte Carlo Tree Search approach to finding efficient patrolling schemes on graphs". In: *European Journal of Operational Research* 277.1 (2019), pp. 255–268

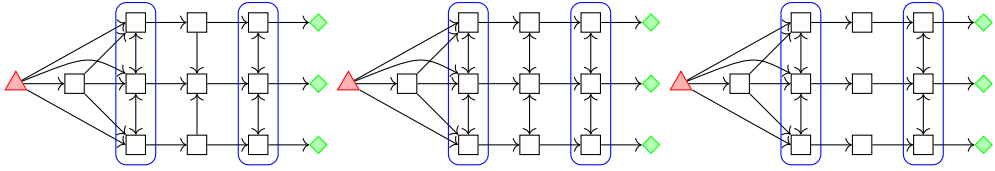
Gry testowe II

WHG non-zero sum (WNZ)

- Modyfikacja WHG, polegająca na zmianie rozkładów, z których są losowane wypłaty, tak aby zmniejszyć ujemną korelację między wypłatami lidera i naśladowcy.

Gry testowe III

Search Games (SEG)



- Naśladowca ma jedną jednostkę i zaczyna w skrajnym lewym wierzchołku
- Lider ma jedną jednostkę w każdym prostokącie
- Celem naśladowcy jest dojście do jednego z trzech celów z prawej
- Celem lidera jest złapanie naśladowcy w wierzchołku
- Lider nie może opuścić prostokąta
- Naśladowca zostawia ślady w wierzchołkach, które lider może zobaczyć

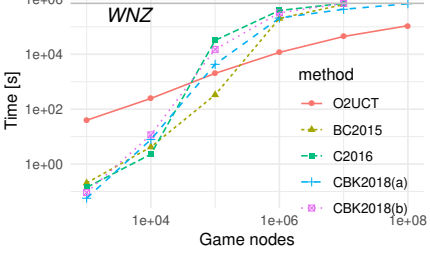
Branislav Bošanský et al. "An Exact Double-Oracle Algorithm for Zero-Sum Extensive-Form Games with Imperfect Information". In: *J. Artif. Intell. Res.* 51 (2014), pp. 829–866

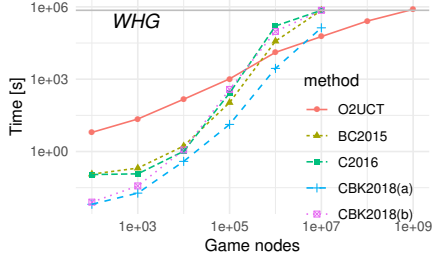
Metody

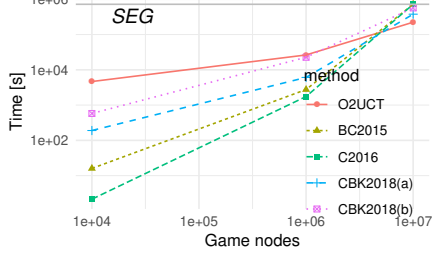
- BC2015, C2016 – metody dokładne, wykorzystujące postać sekwencyjną
- CBK2018 – metoda przybliżona, wykorzystująca uproszczenie gry, w dwóch wariantach parametryzacji
- O2UCT

Wyniki – czas

- Agregacja gier według liczby węzłów w postaci ekstensywnej
- Uśredniony czas
- Limit czasu wykonania każdego przebiegu 200h, po przekroczeniu limitu do średniej wchodzi 200h

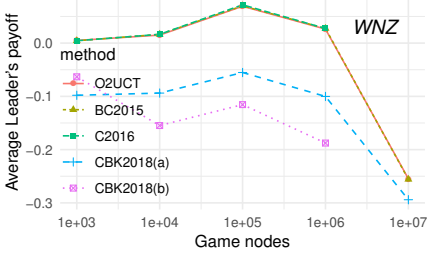


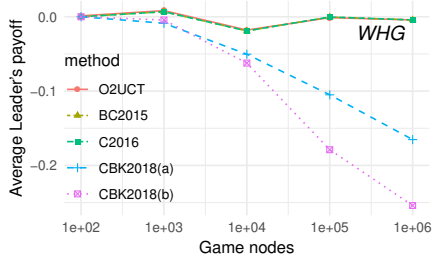


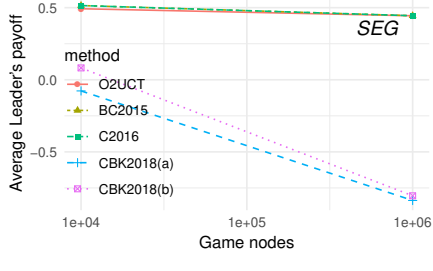


Wyniki – jakość strategii

- Agregacja gier według liczby węzłów w postaci ekstensywnej
- Uśredniona wypłata lidera
- Limit czasu wykonania 200h, gry, gdzie był przekroczony limit są usuwane z agregacji dla wszystkich metod







Uwagi o pamięci

- W przypadku gier w kubełku 10^8 węzłów dla większości gier metody wykorzystujące programy liniowe nie mieszczą się w 256GB pamięci.
- O2UCT jest w stanie rozwiązywać te gry wykorzystując $8GB$ pamięci.

Spostrzeżenia

- Dla małych gier jest słabo, dla dużych gier skaluje się dużo lepiej (typowa obserwacja w inteligencji obliczeniowej)
- Dużo mniejsze zapotrzebowanie na pamięć

Dalsze kierunki

- Metoda optymalizacji wewnętrznej jest nieskomplikowana. Zastąpienie jej czymś lepszym, np. wykorzystującym strukturę wypłat, może dać dużą poprawę
- Pełny przegląd strategii naśladowcy w optymalizacji wewnętrznej można zastąpić próbkowaniem losowym

Pytania?