

## Streszczenie

Aproksymacja stanu równowagi Stackelberga w grach wielokrokowych o sumie niezerowej z niepełną informacją z użyciem metod Monte Carlo

Rozprawa opisuje dwie metody, będące oryginalnym wynikiem, które pozwalają przybliżyć strategię lidera ze stanu Równowagi Stackelberga w grach wielokrokowych o sumie niezerowej z niepełną informacją. Hipotezą badawczą stawianą w rozprawie jest twierdzenie, że możliwe jest zbudowanie efektywnej metody poszukiwania Równowagi Stackelberga w tych grach w oparciu o próbkowanie Monte Carlo.

Gry wielokrokowe, to rodzaj gier, w których gracze mają wiele punktów decyzyjnych, w każdym z tych punktów muszą wykonać akcję. W teorii gier często do reprezentacji gier wielokrokowych używa się postaci ekstensywnej – drzewa, gdzie węzłami są stany gry, a krawędziami ruchy możliwe do wykonania w tych stanach. Gry z niepełną informacją charakteryzują się tym, że gracze nie mają pełnej informacji o całym stanie gry, w szczególności o działaniach przeciwnika, a mogą obserwować jedynie projekcję stanu udostępniającą wybrane informacje. Dodatkowo, w tej rozprawie rozważane są tylko gry z doskonałą pamięcią, to znaczy takie, gdzie udostępniana graczowi informacja zawsze uwzględnia rozróżnienie stanów na podstawie wszystkiego co gracz dotąd zaobserwował, w tym wykonanych przez niego akcji. Suma niezerowa oznacza, że w grze dla dwóch graczy wypłaty otrzymane na koniec gry przez graczy nie muszą sumować się do zera.

Równowaga Stackelberga jest pojęciem z obszaru teorii gier. W Grze Stackelberga bierze udział dwóch asymetrycznych graczy, lider i naśladowca. Lider wybiera strategię mieszaną jako pierwszy, następnie upublicznia ją. Naśladowca wybiera swoją strategię znając już strategię lidera. Model Stackelberga zakłada pełną racjonalność naśladowcy, czyli zachowanie gdzie naśladowca zawsze wybierze strategię, która daje mu najlepszą możliwą wypłatę. Stanem Równowagi Stackelberga nazywamy układ strategii lidera i naśladowcy, gdzie strategia naśladowcy jest optymalną odpowiedzią na strategię lidera, a strategia lidera daje liderowi najwyższą możliwą wypłatę spośród wszystkich układów strategii spełniających warunek optymalnej odpowiedzi naśladowcy. Rozprawa wskazuje pozycje w literaturze, które mówią o praktycznym zastosowaniu Równowagi Stackelberga w sytuacjach związanych z interakcją pomiędzy siłami bezpieczeństwa (np. policją, strażą graniczną), a łamiącymi prawo (przemysłnikami, terrorystami).

Po zdefiniowaniu potrzebnych pojęć z teorii gier zaprezentowane są istniejące w literaturze podejścia do obliczeniowego wyznaczania Równowagi Stackelberga. Duża część z istniejących w literaturze podejść jest dedykowana bardzo szczególnym podklasom wielokrokowych Gier Stackelberga o sumie niezerowej z niepełną informacją. Są to metody, które wykorzystują specyficzne cechy w strukturze gry, aby znacznie przyspieszyć obliczenia. Tych metod nie da się uogólnić na całą klasę gier rozważaną w rozprawie. Zaprezentowane są również metody z literatury dedykowane całej wspomnianej klasie gier. Wszystkie prezentowane metody, zarówno te specyficzne dla danej klasy, jak i ogólne wykorzystują programowanie liniowe jako istotny

element rozwiązania. Część metod konstruuje program liniowy, który wylicza stan równowagi, w przypadku części metod wyliczanych jest wiele programów liniowych, a sama strategia jest uzyskiwana z pomocą fragmentów działających poza programowaniem liniowym. Analiza metod z literatury wskazuje powtarzające się elementy, które można wykorzystać przy budowie nowych metod rozwiązujących Gry Stackelberga. Są to: technika przeglądania wszystkich strategii naśladowcy i dobierania do każdej z nich strategii lidera, metoda generowania kolumn, metoda podwójnej wyroczni. Oprócz samych metod wskazany jest też zbiór Search Games wykorzystywany przez niektóre prace opisujące te metody, który można wykorzystać do ewaluacji eksperymentalnej metod rozwiązujących Gry Stackelberga.

Następnie prezentowany jest główny wkład autora rozprawy w dziedzinę. Pierwszym elementem jest rodzina gier z niepełną informacją rozgrywanych na grafach, która została wykorzystana do eksperymentalnej ewaluacji metod. W skład tej rodziny wchodzi trzy różne zbiory gier testowych. Drugim elementem jest opis dwóch metod do poszukiwania strategii lidera, która będzie dobrym przybliżeniem strategii lidera z Równowagi Stackelberga. Obie proponowane metody wykorzystują popularną metodę rozwiązywania gier z pełną informacją, na przykład gier planszowych, nazywaną Upper Confidence Bound applied to Trees (UCT). Metoda UCT w trakcie swojego działania buduje drzewo statystyk na temat ruchów, nazywane drzewem UCT. Pierwsza z metod, nazwana Mixed-UCT opiera się o wielokrotne uruchamianie metody UCT w grze dla jednego gracza, gdzie lider wybiera swoje ruchy, a ruchy naśladowcy pochodzą z wcześniej ustalonej strategii. Następnie statystyki zebrane w ten sposób w drzewie UCT są przekształcane w strategię mieszaną lidera. Ten proces jest powtarzany iteracyjnie, a po każdym uzyskaniu strategii lidera, aktualizowana jest ustalona strategia naśladowcy, przeciwko której rozgrywane są symulacje UCT, tak aby uwzględnić fakt, że strategia naśladowcy powinna być najlepszą odpowiedzią na strategię lidera. Metoda Mixed-UCT jest następnie uruchomiona na zbiorze testowym gier, które są bliskie sumie zerowej i porównana z metodami z literatury, które da się stosować do ogólnej klasy gier. Wyniki tych eksperymentów pokazują, że Mixed-UCT jest znacznie szybsza od metod z literatury dla dużych gier testowych, a otrzymane strategie są tylko nieznacznie gorsze od strategii optymalnych. Potrzebuje też dużo mniej pamięci operacyjnej. Niestety Mixed-UCT nie działa zbyt dobrze dla gier o sumie dalszej od sumy zerowej. Druga metoda, nazwana O2-UCT, nie stosuje podejścia UCT bezpośrednio. Główna zasada działania tej metody to wielokrotne próbkowanie strategii naśladowcy, a następnie dobieranie do niej strategii lidera tak, aby spełniony był warunek, że strategia naśladowcy jest najlepszą odpowiedzią na tę strategię lidera, a w drugiej kolejności, żeby wypłata lidera była możliwie duża. W tej metodzie podejście UCT wykorzystywane jest do ukierunkowanego próbkowania strategii naśladowcy tak, aby preferować strategie dla których da się budować strategię lidera o dużej wypłacie. Sama metoda dobierania strategii lidera opiera się o koncepcję podwójnej wyroczni. Naprzemiennie poprawiana jest strategia lidera i poszukiwana jest najlepsza odpowiedź naśladowcy. W zależności od tego jaka odpowiedź została znaleziona, zmienia się kierunek poprawy strategii lidera. Metoda O2-UCT została przetestowana eksperymentalnie

na trzech zbiorach gier testowych i porównana z metodami z literatury. Dla wszystkich zbiorów testowych, dla dużych instancji gier testowych O2-UCT jest szybsza od metod z literatury. Wartości wypłat uzyskiwane przez strategie są bardzo bliskie optymalnym dla wszystkich gier, dla których udało się policzyć rozwiązania dokładne.

Wyniki eksperymentalne badające zaproponowane w tej rozprawie metody potwierdzają hipotezę badawczą postawioną w rozprawie.

**Słowa kluczowe:** Równowaga Stackelberga, UCT, MCTS

A handwritten signature in blue ink, reading "Jan Komorowski". The signature is written in a cursive, flowing style with a long horizontal stroke at the end.

