

Streszczenie

Rozprawa doktorska

Metody Monte Carlo Tree Search oraz Reinforcement Learning w wieloetapowej strategicznej grze karcianej

mgr inż. Konrad Godlewski

Rozprawa przedstawia metody uczenia maszynowego wykorzystane do stworzenia agentów do gry karcianej "The Lord of the Rings: The Card Game" (LOTRCG). Cechą wyróżniającą grę jest nierywalizacyjny i złożony charakter rozgrywki. Gracz podejmuje decyzje w pięciu, ściśle określonych momentach podczas rundy. Akcje mają charakter sekwencyjny, czyli decyzja podjęta w jednej fazie gry wywiera wpływ na decyzję w etapie kolejnym. Taki mechanizm rundy wymaga strategicznego zarządzania kartami. Dodatkowo pomiędzy decyzjami występują zdarzenia losowe, które uniemożliwiają łączenie kilku momentów decyzyjnych.

LOTRCG jest grą typu collectible card game, co oznacza szeroki wybór kart dostępnych dla gracza. Karty posiadają różne statystyki oraz zdolności specjalne profilujące je do konkretnej fazy gry. Duża różnorodność kart przekłada się na dużą popularność tego tytułu wśród miłośników gier karcianych. Ten element decydujący o atrakcyjności gry stanowi zarazem duże wyzwanie dla agentów AI.

Złożony charakter rundy jest pierwszym obiektem badań tej rozprawy. W celu analizy poszczególnych faz gry zaimplementowano agenta strategii losowej. Uzyskane przy jego pomocy wyniki pozwoliły określić kluczowe momenty decyzyjne. Ponadto agent losowy posłużył do wykonywania losowego próbkowania drzewa algorytmu MCTS.

Rozprawa wykorzystuje dwie rodziny technik inteligencji obliczeniowej. Pierwszą z nich jest Monte Carlo Tree Search (MCTS). Metoda ta opiera się na wyszukiwaniu heurystycznym korzystającym z losowego próbkowania stanów gry. MCTS przechowuje rozgrywkę w postaci drzewa, które rozbudowywane jest w sposób iteracyjny. Podstawowa wersja algorytmu MCTS została rozbudowana o redukcję akcji, która pozwala na ograniczenie rozmiarów drzewa. Modyfikację tę oparto o wiedzę ekspercką, która pozwala na eliminację kart o niskiej użyteczności. Dodatkowo przeprowadzono optymalizację hiperparametrów MCTS nie tylko względem skuteczności, ale także pod kątem czasu obliczeniowego. W ten sposób uzyskano optymalny zestaw agentów MCTS, który osiągnął współczynnik 82.8% wygranych.

Drugą rodziną analizowanych technik jest uczenie ze wzmocnieniem (Reinforcement Learning - RL). Uczenie ze wzmocnieniem opiera się na metodzie prób i błędów, w którym agent jest w interakcji ze środowiskiem. Agent otrzymuje obserwację stanu gry, następnie wybiera akcję, która zostaje wykonana w środowisku. Środowisko odpowiada nagrodą oraz następną obserwacją. Celem agenta jest maksymalizacja sumy nagród na przestrzeni całego epizodu. Z uwagi na działanie agenta w środowisku o zmiennej liczbie akcji, selekcję algorytmów RL oparto na sposobie kodowania akcji. W pracy stworzono dwa typy kodowania. Pierwszy z nich to makroakcje, które pozwalają na uzyskanie stałej liczby akcji. Drugim typem są akcje bezpośrednie, czyli zagrywanie poszczególnych kart przez agenta. Dla makroakcji zaimplementowano algorytmy Q-Learning oraz Actor-Critic (AC). Q-Learning to szeroko stosowane rozwiązanie w dziedzinie RL aproksymujące wartość użyteczności stanu gry. Actor-Critic korzysta z aproksymacji zarówno funkcji użyteczności jak i strategii. Agent AC zaprogramowano do makroakcji jak i akcji bezpośrednich. W wyniku przeprowadzonej optymalizacji najlepszy zestaw agentów RL wykazał się skutecznością na poziomie 95.3%.

W pracy zestawiono najlepszych agentów MCTS oraz RL. Przeprowadzone eksperymenty wskazują na istotną przewagę uczenia ze wzmocnieniem. Jednak rozwiązanie to wymaga dużego nakładu obliczeniowego na etapie uczenia. Czas uczenia jest silnie zależny od dostępnej mocy obliczeniowej. MCTS ma pozornie mniejsze wymagania, jednak czas decyzji jest znacznie większy.