Politechnika Wrocławska Wydział Elektroniki Kierunek Teleinformatyki

Inżynierska praca dyplomowa

Projekt sztucznej inteligencji

Michał Żarejko

Dyplomant:	**Michał Żarejko**
Nr albumu:	**249374**
Promotor:	**dr inż. Paweł Zyblewski**

Wrocław. 2021

Spis treści

Rozdział 1

Wstęp

Duży rozwój nauki w ostatnich latach związany z uczeniem maszynowym, spowodował powstanie wielu nowoczesnych algorytmów i technologi, które pomagają dzisiaj w codziennych czynnościach lub zastępują ludzi w odpowiedzialnych procesach.

Przykładem mocno rozwijanych narzędzi używanych w życiu prywatnym są asystenci głosowi, tłumacze maszynowe, modele wyświetlające elementy na stronach internetowych na podstawie gustu użytkownika, gry wideo, inteligentne samochody lub przetwarzanie obrazów. Dodatkowo uczenie maszynowe jest mocno wykorzystywane w większości firmach, między innymi na halach produkcyjnych, w transporcie, medycynie, cyberbezpieczeństwie.

Po mimo tak wielu możliwości i zastosowań, sztuczna inteligencja zyskuje największą popularność medialną przez gry rywalizacyjne, gdzie głównym zadaniem jest pokazanie przewagi algorytmów względem ludzi. W ostatnich 20 latach można spotkać się z dużą ilością wydarzeń gdzie profesjonalni gracze muszą stoczyć pojedynek z wytrenowaną sztuczną inteligencją. Między innymi w 2016 roku zorganizowano mecz między 18 mistrzami świata w grze "Go", mieli oni za zadanie pokonać algorytm ĄlphaGo"utworzony przez zespół "DeepMind". Dużym osiągnięciem była wygrana modelu z wszystkimi graczami, gra była uważana powszechnie za skomplikowaną i trudną do rozwiązania.

W 2019 roku zespół ÓpenAI"utworzył grupę współpracujących "botów"w grze komputerowej "Dota 2". W 4 dniowym wydarzeniu odbywającym się internetowo modele rwalizowały z 5 osobowymi grupami. Maszyny AI zdołały pokonać dużą część graczy. Wydarzenie było mocno omawiane z powodu poziomu skomplikowania gry. Dla porównania środowisko gry "Goźawiera 150 możliwych ruchów na turę, "Dota 2"może posiadać ich 20 000 przez czas 45 minut.

Takie wydarzenia pokazały, że w dzisejszych czasach sztuczna inteligencja może przewyrzszać myśleniem strategicznym człowieka. Często w tworzeniu takich progra-

mów dużym wyzwaniem jest poziom skomplikowania gry, zależy to między innymi to od tego, czy jest to środowisko deterministyczne, czy stochastyczne, czy głównym zadaniem jest rywalizacja czy, współpraca, ile modeli ma zawierać gra. Aktualnie jednak jednym z największych problemów takich programów jest niedostateczny zakres dostępnych informacji o środowisku. Sztuczna inteligencja, aby zwyciężać musi zostać nauczona grać, więc potrzebuje dużej ilości danych wejściowych, które są rozróżnialne. Przykładem gry, która jest pozbawiona tego problemu są szachy. Sztuczna inteligencja wykonuje ruchy bazując na informacjach w jaki sposób są ułożone pionki w danym momencie i na historii dotychczasowej gry. Zmiana stanu środowiska jest zauważalna przez gracza co pozwala na natychmiastową reakcję i prostsze sposoby na uczenie maszynowe. Przykładem gry ciężkiej do rozwiązania, w której występuje nie pełny zestaw informacji jest Poker Texas Holdem, po mimo wiedzy o kartach w ręce i na stole, gracz nie posiada wiedzy o kartach przeciwników, w takim przypadku dwa pozornie identyczne stany środowiska w rzeczywistości mogą się różnić. Z powodu takich cech większość popularnych algorytmów jak "DQN", Ąctor-Criticłub ĄlphaZeroštaje sie bezużyteczna i nie daje dobrych rezultatów.

W niniejszej pracy przedstawiono sposób możliwego rozwiązania takiego problemu przy pomocy algorytmu o nazwie "Deep CFR". Pierwszy i drugi rozdział dokładnie opisuje cele pracy, zakres projektu, zagadnienia wymagane aby zrozumieć działanie algorytmu. W trzecim rozdziale skupiono się na implementacji, czwarty pokazuje uzyskane wyniki.

1.1 Cel i zakres pracy

Głównym celem pracy jest implementacji algorytmu "Deep CFR"do gry "Heads Up Limit Texas Poker Hold'em". Jest to popólarna wersja rozgrywki 2-osobowej gdzie uczestnicy nie mogą wybrać samodzielnie kwoty podbicia stawki, jest ona ograniczona przez ustaloną wartość. Takie środowisko minimalizuje możliwe ruchy do 3 akcji. Używając omawianego algorytmu wytrenowano 5 modeli rozpoznawania, które zostały następnie wykorzystane do rozegrania turnieju składającego się na wszystkie kombinacje rozgrywek modeli po 10 powtórzen gry. Taki proces pozwoli określić, który model najlepiej gra w "Heads Up Limit Texas Poker Hold'em".

1.2 Przegląd dostępnych rozwiązań

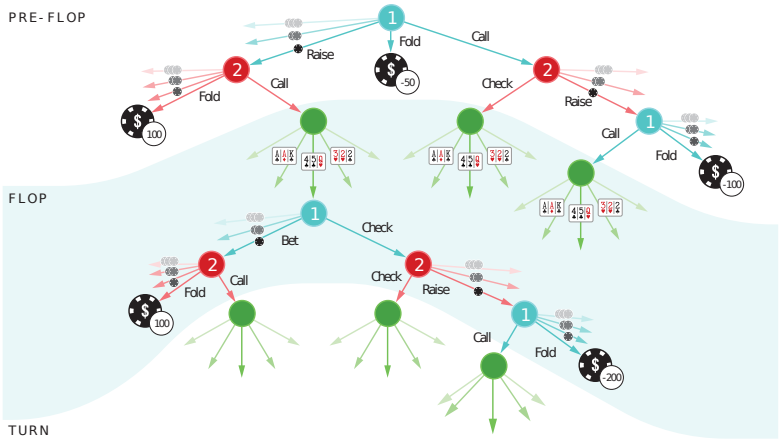
W ciągu ostatnich 10 lat powstało wiele algorytmów rozwiązujących różne werjse gry Poker. Miedzy innymi Extensive-Form Fictitious Play", Ńeural Fictitious Self-Play", Żegret Policy Gradients", Ćounterfactual Regret Minimization". Z wymienionych algorytmów popólarnym aktualnie rozwiązaniem jest ĆFRżozszerzony o sieci neuronowe, zwany "Deep CFR". Daje on najszyszą zbierzność [3].

1.2.1 Teoria Gier

Aby zrozumieć działanie wymienionych algorytmów nalerzy zapoznać się działem matematyki o nazwie "Teoria Gier"[1]. Bada on optymalne zachowanie w grach hazardowych przez dobieranie odpowiedniej strategii bazującej na zasadzie Żówności Nashaóraz opisie środowiska jako gry w postaci ektensywnej.

Gra w postaci ekstensywnej

Aby rozwiązać skąplikowane grę należy ją przeanalizować przy pomocy uproszczonego opisu. Gry w formie ekstensywnej można przedstawić jako drzewo decyzyjne gdzie każdy węzeł rozgałęzia się na możliwe akcje oraz identyfikuje aktualny stan gracza przez zestaw informacji I, ostatnie węzły to stany końcowe gdzie określony gracz zyskuje nagrode lub ją traci. Ponieważ drzewo przedstawia grę wielu graczy, dlatego każdy z węzłów należy przypisać odpowiedniemu uczestnikowi.



Rysunek 1.1: Drzewo decyzyjne gry Ńo-limit Poker Texas Hold'em". [4]

Równość Nasha

W grach postaci normalnej [6] to twierdzenie określa perfekcyjny stan gry gdzie wszyscy gracze wykorzystują najlepszy zestaw strategii, którego zmiana przyniesie tylko straty. Oznacza to, że nie jest możliwe zwiększenie uzyskanej nagrody bedąc w tym stanie [1].

Podział gier

Dodatkowo dział nauki zakłada, że omawiane środowiska można podzielić na gry o sumie stałej, zmiennej oraz zerowej. Wszystkie wymienione formy opisują różnice między wygraną i przegraną graczy. W pracy założono, że "Heads Up Poker Texas Hold'em"jest grą 2-osobową o sumie zerowej, czyli wygrana jednego uczestnika oznacza całkowitą porażkę oponenta w wyskokości wygranej stawki w taki sposób, że suma wygranej i przegranej wynosi zero. W takiej formie można zaimplementować "Deep CFR", który ma szanse zbiec się do stanu bliskiego Żówności Nasha"[3]. Algorytm będzie przez wiele iteracji eksplorował drzewo decyzyjne i dobierał odpowiednie strategie aż trafi na takie, które dają najlepsze rezultaty.

Metryki

Popólarnym sposobem badania postępów modelu względem bliskości Równości Nasha jest metryka Ęxploitabilitćzyli różnica między najlepszą możliwą strategią w grze, a aktualnym modelem.

1.2.2 Historia modeli Poker Texas Hold'em

Bazując na teorii gier oraz różnych algorytmach powstało wiele rozwiązań różnych wersji gry Poker. Pierwsze dokumenty naukowe badające grę zaczęły powstawać w 2005 roku omawiające bardzo proste środowiska jak "Poker Kuhn", na pierwszy duży sukces trzeba było czekać 10 lat. W 2015 roku utworzono sztuczną inteligencje Ćepheuszozwiązującą problem "Heads Up Limit Texas Hold'em"wykorzystujące algorytm CFR+. Po tym osiągnięciu rozpoczęto prace nad algorytmem mogącym rozwiązać problem gry "Heads Up No-limit Texas Hold'em". Zajeło to 2 lata od Ćepheus'a", model nazwano "DeepStack", mieszał on techniki znane z CFR z siecimi neuronowymi. Przetestowano go na 33 profesjonalnych graczach w wielu iteracjach gry. Algorytm w większości przypadków wygrał [4]. Była to pierwsza wygrana AI z człowiekiem w najtrudniejszej wersji gry Poker.

Rozdział 2

CFR

2.1 Regret Matching

Jest to metoda uczenia polegająca na minimalizacji żalu używana w algorytmie CFR. Opisuje się to jako sposób na liczenie wektorów wag o długości równej liczbie możliwych akcji A , korztstając z u^t czyli nagrody uzyskanej w stani t oraz z dystrybucji ruchów p^t [2]. Posiadając te informacje algorytm iteracyjnie aktualizuje wagi wzorem 2.1.

$$p_i^t(a) = \begin{cases} \frac{R^{t-1,+}(a)}{\Sigma_{a' \in A} R^{t-1,+}(a')} & \text{if } \Sigma_{a' \in A} R^{t-1,+}(a') > 0; \\ \frac{1}{|A|} & otherwise. \end{cases} \tag{2.1}$$

Gdzie R^t(a) jest równe formule 2.2, a $R^{t,+}(a)$ oblicza się jak w 2.3.

$$R^t(a) = \frac{1}{T}\sum_{t=1}^{T} u^t(a) - \sum_{a \in A} p^t(a) u^t(a) \tag{2.2}$$

$$R^{t,+}(a) = max(R^t(a), 0) \tag{2.3}$$

Podsómowując pwyżesze wzory, dla każdej akcji wektora w danym stanie wylicza się $R^t(a)$, gdzie należy skorzystać z sumy przyszłych wartości $u^t(a)$ oraz $p^t(a)$ dla wybrania danych ruchów. Całość obrazuje jak bardzo gracz żałuje wybranych akcji w czasie od t do T.

2.2 Regret Minimization

W przypadku algorytmu CFR i gry Heads Up Limit Texas Hold'em gracz ma do zynienia z wyborem strategi σ_i^t dlatego wylicza się wartość zwaną 'Average overall regret', która określa jak gracz i będzie żałował wybór danej strategi aż do czasu T [2].

$$R_i^T = \frac{1}{T} \max_{\sigma_i^* \in \Sigma_i} \sum_{t=1}^{T} \left(u_i \left(\sigma_i^*, \sigma_{-i}^t \right) - u_i(\sigma^t) \right) \tag{2.4}$$

Głównym zadaniem algorytmu CFR jest minimalizacja tej wartości dla wszystkich graczy. Dodatkowo dla każdego gracza oblicza się średnią strategię dla każdego stanu I oraz akcji A [2].

$$\sigma_i^{-t}(I)(a) = \frac{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I) \sigma^t(I)(a)}{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)} \tag{2.5}$$

Jesli wszyscy gracze będą dążyć do minimalizacji wartości 'Average overall regret', wtedy gra powinna osiągnąć po wielu iteracjach stan bliski Równości Nasha [2].

2.3 Counterfactual Regret

2.4 Monte Counterfactual Regret Minimization

Bibliografia

[1] Myerson, Roger B. Game theory. Harvard university press, 2013.

[2] Zinkevich, Martin, et al. Żegret minimization in games with incomplete information.Ądvances in neural information processing systems 20 (2007): 1729-1736.

[3] Brown, Noam, et al. "Deep counterfactual regret minimization."International conference on machine learning. PMLR, 2019.

[4] Moravčík, Matej, et al. "Deepstack: Expert-level artificial intelligence in heads-up no-limit poker.Ścience 356.6337 (2017): 508-513.

[5] Davis, Trevor, Neil Burch, and Michael Bowling. "Using response functions to measure strategy strength."Twenty-Eighth AAAI Conference on Artificial Intelligence. 2014.

[6] Heinrich, Johannes, Marc Lanctot, and David Silver. "Fictitious self-play in extensive-form games."International conference on machine learning. PMLR, 2015.