Blackjack - reinforcement learning

Strategia deterministyczna

Na 5 000 000 gier próbnych strategia deterministyczna uzyskała:

- 43.29168% wygranych
- 47.57704% przegranych

Wszystkie testy strategii niedeterministycznych przeprowadziłem na 500 000 gier.

Monte Carlo Exploring Starts

gamma	**ilość epizodów**	**procent wygranych**	**procent przegranych**
1	50 000	42.8774 %	48.7686 %
1	500 000	42.79438 %	48.96288 %
1	5 000 000	42.7548 %	48.9928 %
0.5	50 000	42.6082 %	49.3564 %
0.5	500 000	42.9208 %	49.2102 %
0.5	5 000 000	42.913 %	48.8426 %
0.1	50 000	42.9002 %	49.3802 %
0.1	**500 000**	**42.9256 %**	**48.808 %**
0.1	5 000 000	42.893 %	48.859 %

On-policy first-visit MC control

epsilon	**gamma**	**ilość epizodów**	**procent wygranych**	**procent przegranych**
0.2	1	50 000	42.65 %	48.8458 %
0.2	1	500 000	43.2588 %	47.8012 %
0.2	**1**	**5 000 000**	**43.0416 %**	**48.2674 %**
0.2	0.2	500 000	42.953 %	48.2022 %
0.2	0.2	500 000	43.0828 %	48.334 %

Q-learning

alfa	epsilon	gamma	ilość epizodów	procent wygranych	procent przegranych
0.5	0.1	0.5	500 000	40.81 %	50.97 %
0.5	0.1	0.5	5 000 000	41.27 %	50.9 %
0.2	0.1	0.5	500 000	42.14 %	49.27 %
0.2	0.1	0.5	5 000 000	40.7178 %	51.1734 %
0.5	0.1	0.2	500 000	41.6054 %	50.0052 %
0.5	0.1	0.2	5 000 000	42.132 %	49.0892 %
0.2	0.1	0.2	500 000	42.199 %	48.657 %
0.01	**0.1**	**0.2**	**500 000**	**43.2352 %**	**47.6534 %**
0.01	0.1	0.2	5 000 000	42.8978 %	48.2704 %

Sarsa

alfa	epsilon	gamma	ilość epizodów	procent wygranych	procent przegranych
0.2	0.1	0.2	500 000	42.5 %	49.19 %
0.3	0.1	0.2	5 000 000	47.8282 %	49.58 %
0.01	**0.1**	**0.2**	**500 000**	**43.3232 %**	**47.8282 %**

Podsumowanie wyników

Najlepiej działającym agentem okazał się algorytm Sarsa, który różnił się od strategii deterministycznej na jedynie około 4 stanach. Bardzo podobne wyniki uzyskał Q-learning różniąc się na około 6-7 stanach. On-policy first-visit MC control uzyskał jeszcze gorszy wynik, cechując się większą skłonnością do stickowania niż algorytm deterministyczny, a ostanie miejsce zajął Monte Carlo Exploring Starts.

Mikołaj Twaróg