Streszczenie w języku polskim

Wykorzystanie rozkładu zapamietanego doświadczenia w uczeniu ze wzmocnieniem

W niniejszej rozprawie przedstawiamy cykl pięciu publikacji dotyczących rozwoju algorytmów uczenia ze wzmocnieniem off-policy poprzez wykorzystanie zależności między próbkami wynikających z działania tych algorytmów.

W pierwszej części pracy skupiamy się na zależnościach czasowych wynikających z metod eksploracji, które nie są niezależne czasowo. Najpierw zwracamy uwagę, że typowym podejściem do wymuszania podobieństwa kolejnych akcji jest użycie autokorelowanego szumu. Przedstawiamy proces, który produkuje taki szum, zachowując równocześnie stałą korelację między kolejnymi jego wartościami oraz rozkład graniczny szumu w każdym kroku czasowym. Następnie przedstawiamy algorytm, który wykorzystuje właściwości skorelowanego w czasie szumu akcji, aby dokładniej wyznaczać gęstości prawdopodobieństwa sekwencji akcji. Przedstawiamy wyniki eksperymentów, które pokazują, że nasze rozwiązanie otrzymuje znacząco wyższe nagrody niż zarówno bazowe, jak i powszechnie używane algorytmy uczenia ze wzmocnieniem. Ponadto przedstawiamy podejście do wykonywania akcji ze zmiennym, losowym czasem trwania. Pokazujemy, że w ramach tego podejścia możliwym jest obliczenie ilorazu gęstości prawdopodobieństw sekwencji akcji i wprowadzamy algorytm, który wykorzystuje stopniowo zmniejszające się oczekiwane czasy trwania akcji, aby poprawić wydajność treningu w symulowanych środowiskach robotycznych.

W drugiej części tej pracy poruszamy problem wyznaczania precyzji osiągania pośredniego celu w hierarchicznym uczeniu ze wzmocnieniem. Pokazujemy, że w dotychczasowych publikacjach wartości parametrów zadających precyzję osiągania celów pośrednich nie zawsze są optymalne. Prezentujemy algorytm, który na podstawie rozkładu odległości między poprzednimi osiągniętymi stanami, a zadanymi celami pośrednimi automatycznie wyznacza precyzję, z którą cele pośrednie mają być osiągane. Pokazujemy, że dla większości badanych algorytmów i środowisk nasza metoda pozwala osiągnąć lepsze wyniki niż używanie stałej zadanej precyzji osiągania celu.

Cała nasza praca wprowadza trzy algorytmy uczenia ze wzmocnieniem, które wykorzystują zależności występujące w zebranych próbkach.