

Streszczenie

September 29, 2017

W jednej z najgłośniejszych w ostatnich latach prac dotyczących uczenia ze wzmocnieniem, algorytm *Q-learningu* działający na podstawie surowych danych obrazowych w środowisku 2D nauczył się grać w klasyczne gry Atari, dla wielu z nich osiągając wyniki znacznie przewyższające ludzkie. Mimo odniesionych sukcesów *Q-learning* ma wadę: uczenie w ten sposób trwa bardzo długo, a przez znaczną część początkowego działania agent zachowuje się zbyt chaotycznie, by pozwolić mu na działanie w prawdziwym środowisku.

Odpowiedzią na ten problem są metody *uczenia przez demonstrację*, które opierają się na uczeniu agenta na podstawie zaprezentowanych mu zachowań eksperta. Klasyfikator odtwarzający zachowania eksperta może być uzyskany w krótkim czasie, a wynikowy agent może osiągnąć zadowalające zachowanie przed pierwszym kontaktem ze środowiskiem.

Stworzone w ramach tej pracy za pomocą metod *uczenia przez demonstrację* i głębokich sieci neuronowych agenty przejawiają wysoką świadomość środowiska 3D w którym się znajdują, poprawnie rozpoznając i zachowując się w stosunku do ruchomych i nieruchomych obiektów oraz w stosunku do przeciwników. Agenty uzyskują wyniki zbliżone do wyników uzyskanych za pomocą *Q-learningu*, przy ponad stukrotnie krótszym czasie nauki, co potwierdza użyteczność metod *uczenia przez demonstrację* dla problemów uczenia ze wzmocnieniem w środowisku opartym na informacji obrazowej 3D.

One of the most influential papers about reinforcement learning in recent years presents a *Q-learning* algorithm capable of learning to play classical Atari games, and exceed humans in many of them, based only on raw 2D visual data input. Despite this success, *Q-learning* has a drawback: learning is slow and for a good part of the training agent's behavior is too chaotic to let it perform in real environment.

The solution are *imitation learning* methods. *Imitation learning* focus on learning expert's actions shown to the agent, and is capable of creating a behavior cloning classifier in very little time compared to *Q-learning*. Agents based on this classifier behave feasibly even before first encounter with environment.