Optymalizacja systemu sygnalizacji świetlnej w oparciu o przepływowy model ruchu pojazdów.

Michał Lis

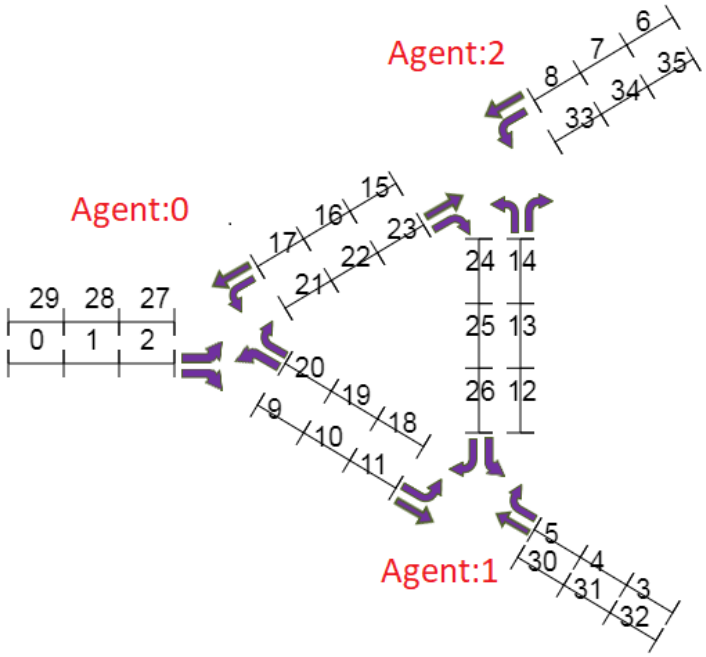
29 czerwca 2019

Spis treści

Rozdział 1

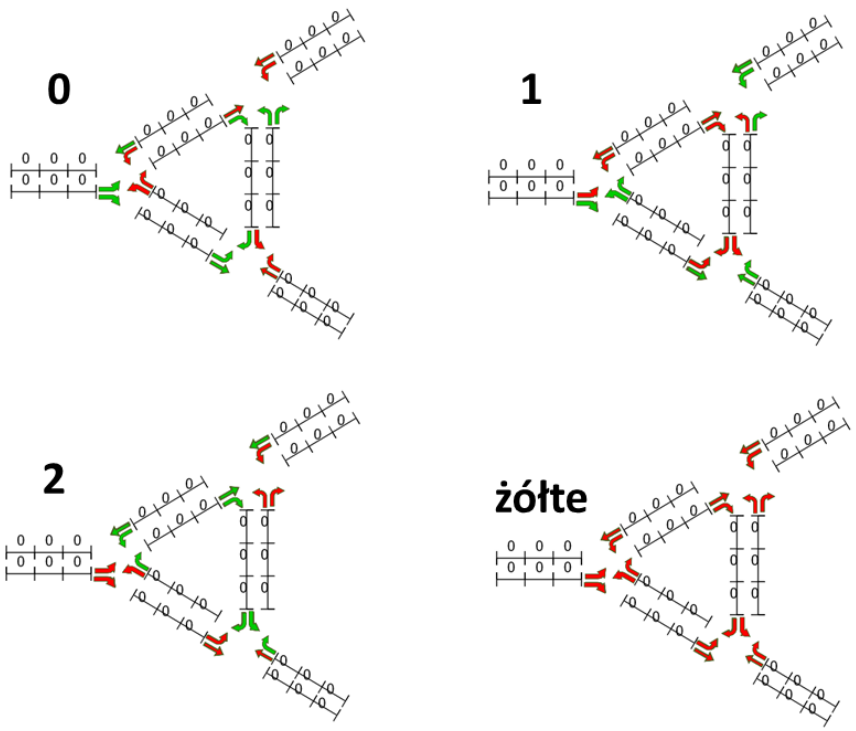
Środowiska symulacyjne i ich nauka

1.1 Środowisko 4



Rysunek 1.1: środowisko 4

Środowisko posiada 12 jednokierunkowych dróg. Każda droga ma 3 odcinki co daje w sumie 36 odcinków (są numerowane od 0 co widać na rysunku 1.1). W sieci dróg znajdują się 3 skrzyżowania. Do każdego z nich jest przypisany agent, który odpowiada za sterowanie sygnalizacją świetlną. Pojazdy w jednym interwale czasowym pokonują jeden odcinek. Na skrzyżowaniach w przypadku zielonego światła przejeżdża maksymalnie 10 pojazdów w jedną stronę.



Rysunek 1.2: środowisko 4 - fazy świateł

Fazy świetlne: Każde skrzyżowanie posiada 4 fazy świetlne przedstawione powyżej. Fazy 0, 1 i 2 posiadają pewne zielone światła. Agent podejmuje decyzję o zmianie tych trzech faz. Zmiana faz świateł nie jest natychmiastowa i następuje dopiero po 2 interwałach czasowych fazy żółtych świateł. Agent może podjąć akcję a należącą do [0,1,2] w przypadku gdy obecna faza f należy do [0,1,2]. W pozostałym przypadku agent jest zobowiązany do przekazania akcji 'żółte'.

1.2 Uczenie Środowiska 4

1.2.1 Podejście 1

Każdy z trzech agentów jako stan przyjmuje 10 elementowy wektor. 9 elementów to ilości pojazdów na odcinkach będących przed skrzyżowaniem przypisanym do agenta. Wektor uzupełnia wartość obecnej fazy (0,1,2 lub 'żółte'). Nagrody sa przyznawane jako suma pojazdów, które przejechały przez skrzyżowanie w trakcie najbliższych 4 interwałów czasowych. Początkowo przeprowadzana jest symulacja 100 epizodów z czego każdy trwa 90 interwałów czasowych. Ma ona na celu wygenerowania danych do treningu. Do nauki agent zapamiętuje jedynie te stany, których faza to 0, 1 lub 2. Nieistotne w procesie uczenia są stany z fazą 'żółte' gdyż agent ma tylko 1 możliwą decyzję do podjęcia. Do danych zapisywane są wartości stanu oraz nagród które zostały przydzielone dla pary stan-akcja. Następnie trenowana jest sieć neuronowa przyjmująca na wejście stan - 10 elementowy wektor. Sieć na wyjściu zwraca

3-elementowy wektor określający przewidziane nagrody dla akcji podjętej w zadanym stanie. Podsumowując dla wybranego agenta:

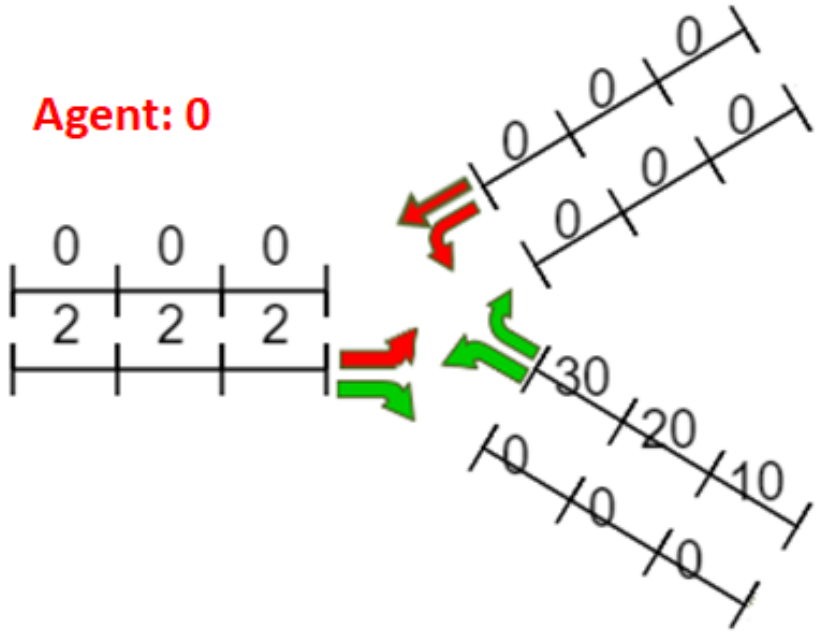
- **Stanem** są ilości pojazdów przed skrzyżowaniem oraz aktualna faza świetlna
- **Nagrodą** w chwili t jest suma pojazdów, które przejechały przez skrzyżowanie w trakcie najbliższych 4 interwałów czasowych czyli do momentu t+4.
- **Dane** są generowane poprzez przeprowadzenie 100 symulacji (każda ma 90 interwałów czasowych).
- **Sieć neuronowa** na podstawie wygenerowanych danych przewiduje najlepszą akcję dla obecnego stanu
- **Końcowa symulacja** zostaje przeprowadzona wedle przewidzianych przez sieć neuronową najlepszych akcji

Algorytm jest następujący:

1. Utworzeni zostają 3 agenci dla poszczególnych skrzyżowań. Każdy z nich posiada sieć neuronową z 10 elementową warstwą wejściową i 3 elementową warstwą wyjściową.
2. Przeprowadzone zostaje 50 losowych symulacji. Agenci zapamiętują dane z każdej chwili symulacji. Na te dane składa się stan oraz 3 elementowa tablica, która pod indeksem akcji przechowuje przydzieloną nagrodę. Pozostałe wartości tej 3 elementowej tablicy są przewidywane przez sieć neuronową.
3. Następuje trening sieci neuronowej. Parametry sieci neuronowej zostały wybrane w procesie walidacji krzyżowej. Zbiór walidacyjny to 0.2 wygenerowanego zbioru w punkcie 2. Pozostała część przeznaczona jest do treningu. Warunkiem stopu jest zwiększenie się błędu dla zbioru walidacyjnego.
4. Agenci zapominają dane z przeprowadzonych 100 symulacji. Natomiast pozostawiają w pamięci wagi wytrenowanej siei neuronowej.
5. Opcjonalnie w celu monitorowania postępów nauki przeprowadzona zostaje symulacja. Jej akcje są wybierane w sposób zachłanny. Dla aktualnego stanu zostaje zawsze wybrana akcja o najwyższej przewidzianej nagrodzie przez sieć.
6. Powrót do kroku 2

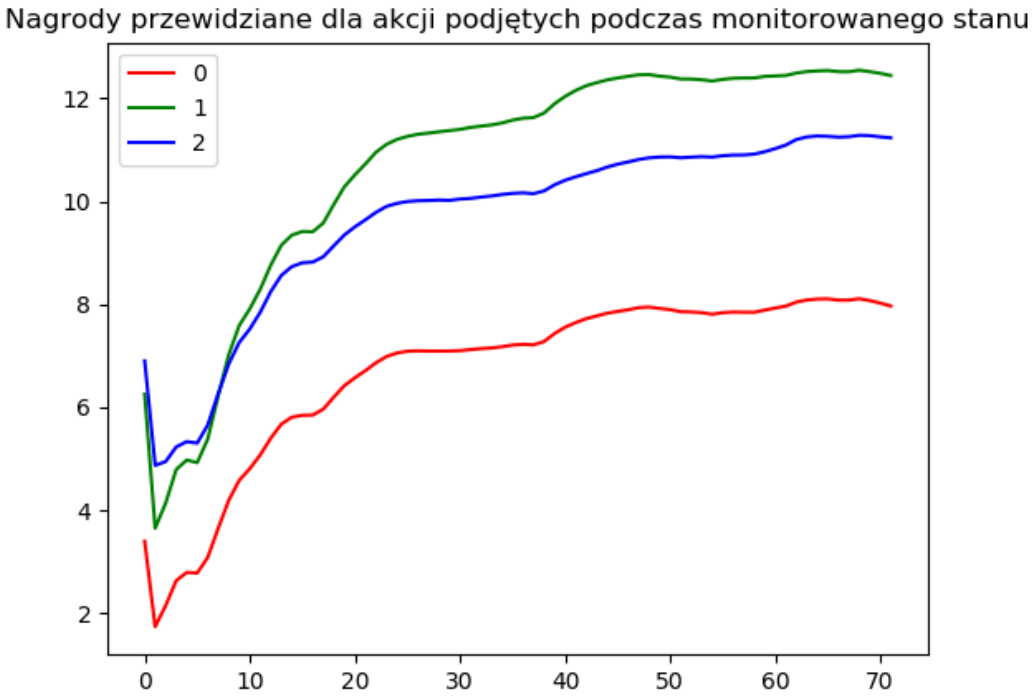
1.2.2 Monitorowanie decyzji dla wybranego stanu

W tej sekcji zostanie przedstawione jak zmieniały się rekomendowane przez agenta 0 akcje dla pewnego szczególnego stanu. Ilości pojazdów przed skrzyżowaniem to [2,2,2,30,20,10,0,0,0]. Aktualna faza świetlna to 1. Warto zadać sobie pytanie jaka akcja jest najbardziej opłacalna dla takiego stanu. Z pewnością warto podtrzymać aktualną fazę świetlną - gdyż przed tymi światłami jest najwięcej pojazdów. By osiągnąć ten cel należy wybrać akcję 1. Porządanym zjawiskiem jest, aby agent 0 dla tego stanu przewidywał największe nagrody dla akcji 1. Po



Rysunek 1.3: Monitorowany stan

każdej sesji uczenia sieci neuronowej zapisane zostały spodziewane nagrody. Ich wykres jest poniżej (każda pełna sesja algorytmu daje podobny wykres).



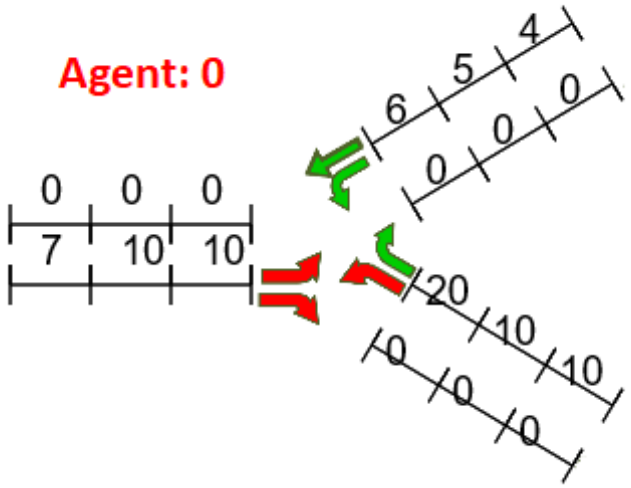
Zgodnie z przypuszczeniami z upływem czasu agent zauważa, że najlepszą akcją jest akcja 1. Jest to jak najbardziej dobry prognostyk. Niewiele gorszy wynik ma akcja 2 - zauważalnie lepszy od akcji 0. Ma to uzasadnienie w tym, że faza świetlna 2 ma zielone światło dla prawoskrętu przed najbardziej zatłoczoną drogą.

1.3 Analiza wyników uczenia

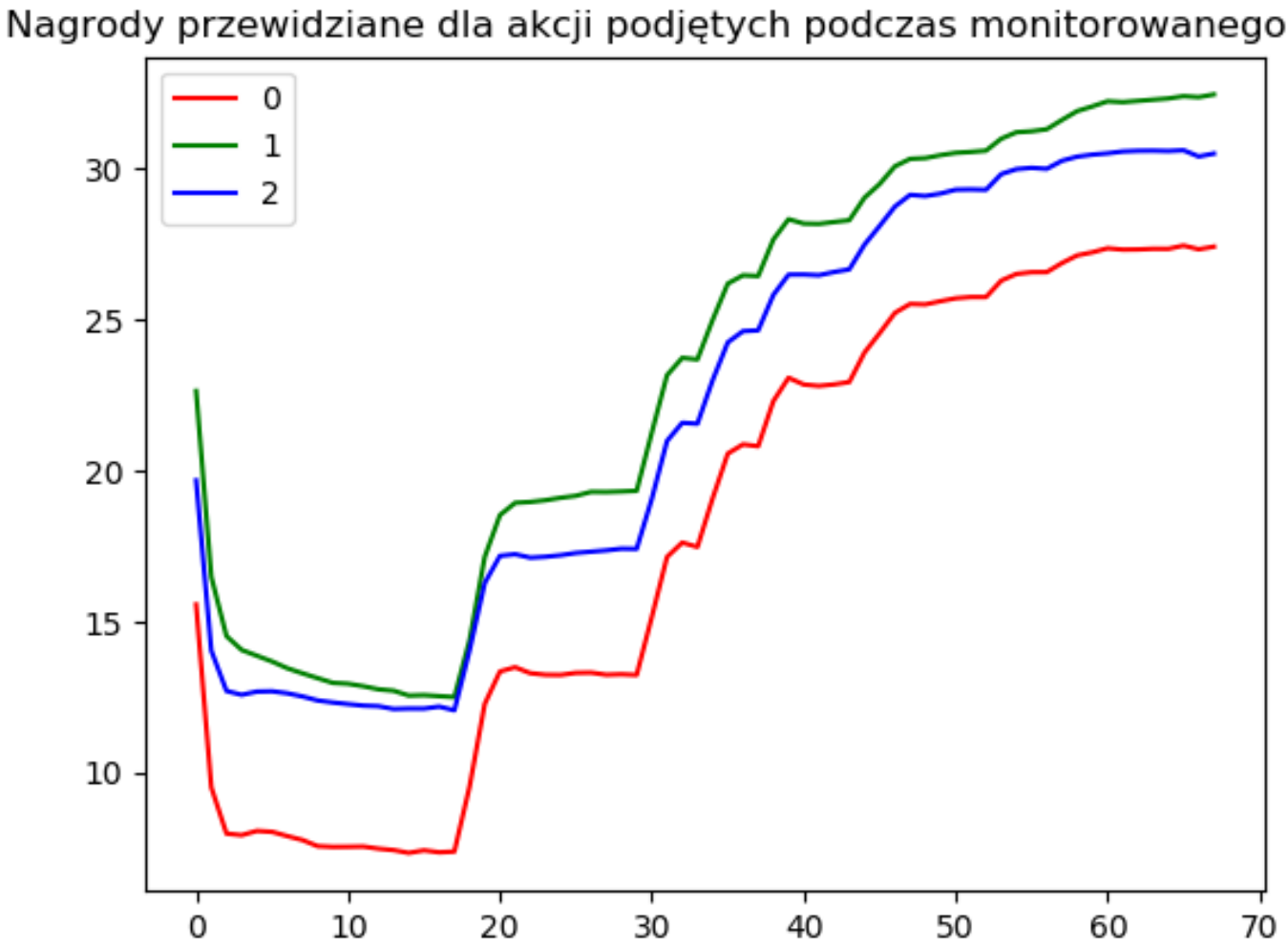
Wykres ilości pojazdów, które opuściły układ nie wygląda jakby wraz z upływem uczenia wynik się polepszał. Podobnie wygląda analogiczny wykres dla strategii losowych akcji. Należy zwrócić jednak uwagę iż średni wynik dla wyuczonej strategii to 1080. Do tego wyniku nie zbliżają się nawet najszczęśliwsze rezultaty dla losowych symulacji.

1.4 Monitorowanie kolejnego stanu

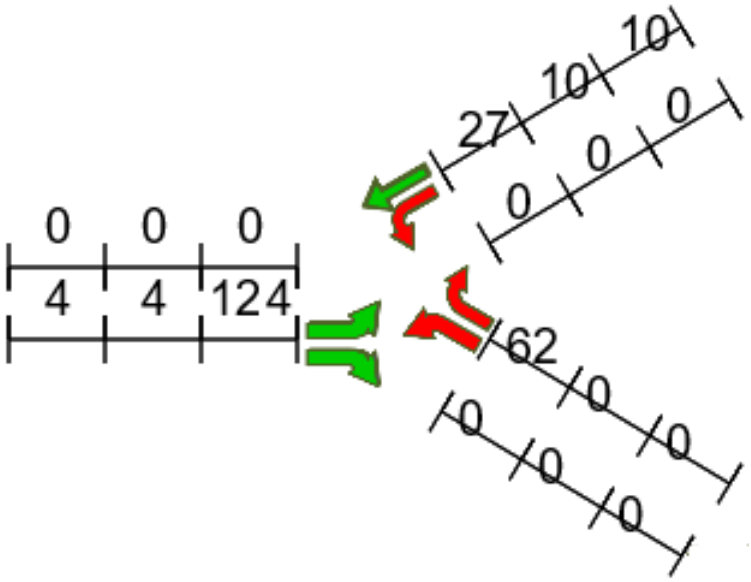
Dla poprzednio monitorowanego stanu ukazanego w (1.2.2) sieć bardzo dobrze przewidywała najlepszą akcję. Był to jednak trywialny przykład z oczywistą do przewidzenia akcją. Należy prześledzić wyniki uczenia dla innego stanu. Z pewnością trudniej agentowi będzie podjąć poprawną decyzję dla następującego stanu.



Przewidzenie najlepszej akcji (którą wydaje się być 2) nie sprawia agentowi większego kłopotu. Przeprowadzone zostało 5 pełnych sesji uczenia (każda po 10 minut) i wszystkie wyniki są bardzo podobne. Jedynie w jednej z nich były nieliczne momenty podczas których agent uznawał akcję 2 za najbardziej optymalną. Wykres przewidywanych nagród z tej sesji jest przedstawiony poniżej.



Ten losowo wybrany stan, który wydaje się być wymagający w kwesti doboru najlepszej decyzji okazuje się nie sprawiać problemów dla agenta. Należy zatem prześledzić symulację, aby odnaleźć stany w których są podejmowane nieoptymalne akcje. Przykładowy stan z którym agent nie radzi sobie jest przedstawiony poniżej.


0
0
0
4
4
124
27
10
10
0
0
0
62
0
0
0
0
0