

Warszawa, 07.04.2020r.

ALHE

Algorytmy heurystyczne

Skład zespołu: Patryk Łempio, Łukasz Wolanin
Opiekun: mgr inż. Kamil Deja

Specyfikacja projektowa

“Projekt z koroną”

1. Treść zadania.

W ramach projektu z koroną Państwa zadaniem jest uratowanie świata przed zabójczym wirusem.

Szalony naukowiec odkrył lek na chorobę covid-19. Niestety ze względu na jego złożoność może go dystrybuować tylko osobiście lecząc na raz wszystkich chorych znajdujących się w tym samym państwie. Zadanie utrudniają mu zamknięte lotniska i porty. W związku z tym poruszać może się jedynie pieszo bądź kajakiem pomiędzy sąsiadującymi krajami (posiadającymi granicę lądową lub morską). Podróż taka zawsze zajmuje mu 1 dzień. Korzystając z najnowszych danych dotyczących zachorowań <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide> oraz z listy wszystkich granic pomiędzy krajami np: https://en.wikipedia.org/wiki/List_of_countries_and_territories_by_land_and_maritime_borders, opracuj możliwie najlepszą ścieżkę startującą z dowolnego kraju, tak by wyleczyć jak najwięcej ludzi. W ramach projektu proszę zaproponować i przetestować działanie różnych algorytmów.

2. Interpretacja i analiza danych.

Dane zawierają liczbę nowych przypadków i zgonów na dany dzień dla krajów w których monitorowany jest przebieg pandemii COVID-19. Dodatkowo dla każdego kraju dostępna jest informacja o liczebności populacji w 2018 roku. W danych występują spore rozbieżności jeśli chodzi o daty pierwszych wpisów dla poszczególnych krajów. Więc jeśli naukowiec zacznie leczyć wcześniej to istnieje szansa, że niektóre możliwe ścieżki będą prowadzić przez kraj lub kraje, dla których nie ma jeszcze danych. W takim wypadku przyjmujemy, że nikt w tych krajach nie zostanie uleczony. Przyjmujemy również, że nowe przypadki Sars-CoV-2 pojawiają się ponownie w kraju który został wyleczony przez doktora, zgodnie z dostępnymi danymi, oraz że algorytm może do wyboru kolejnego ruchu wykorzystywać dane z przyszłości. Dodatkowymi danymi pomocniczymi jest lista granic między państwami, na jej podstawie można ustalić jak naukowiec może dostać się z kraju A do kraju B. Najbardziej intuicyjną formą reprezentacji powyższych danych jest graf, w którym wierzchołki stanowią kraje, natomiast krawędziami są dostępne pomiędzy nimi granice wodne i lądowe. Dla każdego wierzchołka będzie zdefiniowana liczba zarażonych danego dnia trwania pandemii. Program na wejściu powinien otrzymywać datę w której naukowiec zaczyna leczenie. Proces leczenia kończy się, w ostatnim dniu dla którego są dostępne dane, lub gdy wszyscy zostaną wyleczeni.

3. Wstępna propozycja rozwiązania.

Cały projekt będzie napisany w języku Python. Do prezentacji wyników wykonania planujemy wykorzystać środowisko Jupyter Notebook. W projekcie zakładamy, w celu porównania wyników, wykorzystanie więcej niż jednego algorytmu.

a) Algorytm zachłanny

Chcielibyśmy, aby jednym z nich był algorytm zachłanny, który będzie miał na celu maksymalizację liczby wyleczonych osób, biorąc ciągle pod uwagę największą ilość przypadków, która zanotowana jest obecnie u sąsiadów państwa, w którym obecnie znajduje się szalony naukowiec. Przykładem takiego algorytmu może być m.in. *algorytm w głąb* lub *algorytm wszerz*.

b) Algorytm ewolucyjny

Drugim algorytmem będzie algorytm ewolucyjny ($\mu + \lambda$). Przestrzeń rozwiązań dla konkretnej instancji problemu stanowią ścieżki w wyżej zdefiniowanym grafie o długości równej liczby dni, którą będzie trwała podróż szalonego naukowca. Każdą ścieżkę można przedstawić jako sekwencję odwiedzonych państw. Wyznaczanie wartości funkcji celu polegałoby w zależności od definicji tej funkcji np. na policzeniu wyleczonych pacjentów. Do realizacji selekcji osobników można wykorzystać algorytm ruletkowy, natomiast krzyżowanie będzie możliwe jedynie gdy oboje rodzice będą posiadali wspólne wierzchołki na ścieżce. Wtedy należy wylosować

jeden wspólny węzeł i przeprowadzić krzyżowanie jak na przykładzie (w którym długość wszystkich ścieżek wynosi 5):

P1: 1-3-4-7-2 P2: 9-8-2-4-5

O1: 1-3-4-2-8 O2: 9-8-2-4-7

Jeśli przy zwykłym połączeniu ścieżek potomna ścieżka byłaby za krótka, to wtedy można albo ją odrzucić, albo zastosować połączenie gdzie niektóre węzły będą w odwrotnej kolejności niż były w rodzicu (jak na powyższym przykładzie O1). Mutacja mogłaby polegać na usuwaniu losowego wierzchołka ze ścieżki, a następnie dodaniem innego (niekoniecznie w tym samym miejscu) oczywiście tak żeby zachować sens danej ścieżki, taka mutacja powinna być raczej rzadka.

Oczywiście algorytm ten zatrzymywał by się po upływie zadanego czasu lub po otrzymaniu satysfakcjonującego wyniku.

Na tym etapie nie decydujemy się jeszcze na wybranie konkretnego. Decyzja ta zostanie podjęta wraz z późniejszą pracą nad projektem.

4. Definicja funkcji celu.

Funkcja celu wynika z samego polecenia. Jest nią suma wyleczonych ludzi przez naukowca poruszającego się po zadanej ścieżce. Oczywiście algorytmy będą miały za zadanie zmaksymalizować funkcję celu. Można także za funkcję celu obrać liczbę ofiar pandemii COVID-19.

5. Sposób mierzenia jakości rozwiązania.

Oczywistym miernikiem jakości rozwiązania jest, tak jak w przypadku funkcji celu, ilość wyleczonych przez naukowca pacjentów oraz to, w jakim czasie algorytm znalazł to rozwiązanie. Zależać to będzie od wyboru punktu startowego algorytmu, parametrów zadanych oraz od rodzaju samego algorytmu. Patrząc na zależności pomiędzy zadanymi parametrami i algorytmem będzie można zbadać ich wpływ na jakość stosowanego algorytmu i na tej podstawie spróbować wyciągnąć wnioski. W przypadku algorytmu ewolucyjnego parametrami, których zależność na jakość rozwiązania będziemy sprawdzać, będą: prawdopodobieństwo mutacji, wielkość populacji, z której będziemy reprodukować, i różne strategie selekcji osobników do następnego pokolenia. Oczywiście wymaga to dużego nakładu na testowanie, żeby wyniki były miarodajne.