Model ewolucyjny

w oparciu o Geometryczny Model Fishera (GMF)

Wprowadzenie

Pomimo nieustająco zachodzących zmian w szeroko rozumianym środowisku naturalnym często jesteśmy w stanie zaobserwować sukcesywną adaptację organizmów/populacji do nowych warunków. Taka dynamika jest wypadkową wielu złożonych procesów, które będziemy starali się uchwycić i zamodelować w naszym pierwszym zadaniu. Celem tego ćwiczenia jest opisanie ewolucji populacji P, składającej się z osobników o, opisanych ich fenotypem, reprezentowanym jako wektor w n-wymiarowej przestrzeni cech (tj. $P = \{o \in \mathbb{R}^n\}$), w zadanym, zmiennym środowisku.

Założenia

Populacja P składa się z N osobników o, opisanych ich **fenotypem** $p \in \mathbb{R}^n$, który bezpośrednio określa cechy organizmu.

- **Liczba osobników**: N (stała populacja).
- **Przestrzeń fenotypowa**: Każdy organizm ma fenotyp p.
- **Mutacja**:
  - Każdy osobnik w populacji ma prawdopodobieństwo μ na wystąpienie mutacji w danej generacji.
  - Jeśli osobnik ulega mutacji, każda cecha fenotypu p_i (dla $i = 1, \ldots, n$) ulega mutacji niezależnie z prawdopodobieństwem μ_c.
  - Współrzędne fenotypu, które uległy mutacji, zmieniają się zgodnie z rozkładem normalnym:

    $$\Delta p_i \sim \mathcal{N}(0, \xi^2)$$

    gdzie ξ^2 określa wariancję mutacji.
  - Nowy fenotyp organizmu po fazie mutacji, to: $p^\star = p + \Delta p$
- **Funkcja fitness**:

  $$\phi_\alpha(p) = \exp\left(-\frac{||p - \alpha||^2}{2\sigma^2}\right)$$

  gdzie α to optymalny fenotyp, a σ^2 kontroluje siłę selekcji (większe σ^2 oznacza słabszą selekcję). Fitness $\phi_\alpha(p)$ określa prawdopodobieństwo przetrwania osobnika, a selekcja przebiega poprzez:
  - **Model proporcjonalny:** prawdopodobieństwo przeżycia osobnika (lub wydania przez niego potomstwa) jest proporcjonalne do $\phi_\alpha(p)$.
  - **Model progowy:** osobniki o fitness poniżej pewnego progu są eliminowane.
- **Zmiana środowiska**:

  $$\alpha(t) = \alpha(t-1) + \mathcal{N}(c, \delta^2 I)$$

  gdzie:
  - $c \in \mathbb{R}^n$ to wektor kierunkowej zmiany środowiska, określający średni kierunek ewolucji optymalnego fenotypu.
  - $\delta^2 I$ reprezentuje fluktuacje losowe, przy czym każda cecha fenotypu zmienia się niezależnie zgodnie z $\mathcal{N}(c_i, \delta^2)$.
  - Jeśli $c = 0$, środowisko podlega jedynie losowym zmianom, bez określonego trendu.
  - W niektórych implementacjach można ograniczyć maksymalną zmianę α w pojedynczym kroku.

Parametry Populacji

Populację P charakteryzują następujące cechy:

- liczebność: N ($|P| = N$)
- liczba cech fenotypowych osobnika o: n (tj. $o \in \mathbb{R}^n$)
- parametry związane z wystąpieniem losowej mutacji dla pojedynczego organizmu

Parametry Środowiska

Populacja żyje w środowisku, które jest opisane następującymi regułami:

- W środowisku istnieje optymalny fenotyp, który zmienia się w każdym pokoleniu $\alpha : \mathbb{N} \to \mathbb{R}^n$. Przykładowym scenariuszem zmiany optymalnego fenotypu jest ocieplanie klimatu, które zrealizujemy jako:
$$\alpha(t) = \alpha(t-1) + c,$$
gdzie $c \in \mathbb{R}^n$ określa kierunkową zmianę optymalnego fenotypu, a dodatkowe fluktuacje losowe mogą być modelowane jako:
$$\alpha(t) = \alpha(t-1) + \mathcal{N}(c, \delta^2 I).$$
- Każdy osobnik o posiada swoją wartość fitness określoną przez funkcję $\phi_\alpha : \mathbb{R}^n \to \mathbb{R}$, parametryzowaną optymalnym fenotypem α.
- Prawdopodobieństwo posiadania potomstwa zależy od wartości fitness organizmu i może być realizowane na kilka sposobów:
  - **Model proporcjonalny**: Liczba potomków osobnika o jest proporcjonalna do jego wartości fitness:
  $$P_{reprodukcji}(o) = \frac{\phi_\alpha(o)}{\sum_{o' \in P} \phi_\alpha(o')}$$
  co zapewnia, że organizmy o wyższym fitness częściej przekazują swoje cechy do następnej generacji.
  - **Model progowy**: Osobniki z fitness poniżej pewnej wartości krytycznej ϕ_{min} nie przekazują swoich cech do następnego pokolenia.
  - **Model losowy**: Liczba potomków jest determinowana przez funkcję fitness, ale z dodatkowym składnikiem losowym, np.:
  $$k(o) \sim \text{Poisson}(\lambda \phi_\alpha(o))$$
  gdzie $k(o)$ to liczba potomków osobnika o, a λ jest współczynnikiem kontrolującym średnią liczbę potomków w populacji.

Model ewolucji

Populacja ewoluuje w sposób iteracyjny. Każda iteracja składa się z czterech kroków ewolucyjnych.

- mutacja - każdy osobnik w populacji ulega losowym mutacjom zgodnie z przyjętymi założeniami;
- selekcja - każdy osobnik podlega selekcji w oparciu o jego miarę fitness;
- reprodukcja - każdy osobnik zgodnie z miarą fitness wprowadza osobników do następnego pokolenia;
- zmiana środowiska - zmiana optymalnego fenotypu zgodnie z przyjętym scenariuszem.

Charakterystyka (wciąż) zawiera wiele dowolności, choć częściowo dowolność ograniczyliśmy na ćwiczeniach.

- selekcja - przebiega dwuetapowo, po pierwsze środowisko eliminuje niektóre osobniki na podstawie ich miary fitness (ponownie jak poprzednio - ϕ_α dostarcza miary prawdopodobieństwa), a następnie jeśli rozmiar populacji przekracza N redukuje liczbę osobników do liczby N. Niekoniecznie muszą to być najsilniejsze jednostki, bo różnorodność genetyczna też jest ważna, a osobniki alfa mogą się eliminować na podstawie silnej konkurencji;
- reprodukcja - decyzja projektowa polega na rozstrzygnięciu czy chcemy opisywać populację płciową czy bezpłciową. W przypadku populacji bezpłciowej możemy przyjąć, że organizm rozmnaża się przez klonowanie. W przypadku rozmnażania płciowego wśród cech organizmu należy wyróżnić cechę płci (przyjmujemy, że nie podlega ona mutacji). Dodatkowo można rozważyć wprowadzenie diploidalności tj. $n = 2m$ i wprowadzić mechanizm crossing-over.
- zmiana środowiska - podstawowe scenariusze jakie chcielibyśmy rozważyć to: (i) uderzenie meteorytu: co T pokoleń następuje radykalna zmiana α; (ii) globalne ocieplenie: w każdym pokoleniu α zmienia się o stałą, niską wartość.

Cele

Celem zadania jest analiza podstawowych własności ewolucyjno-adaptacyjnych populacji.

- Przy zadanym scenariuszu zmian środowiskowych (np. liniowe ocieplenie klimatu) jakie jest optymalne prawdopodobieństwo wystąpienia losowej mutacji oraz efekt mutacji, które pozwalają na adaptację populacji?
- Wizualizacja ewolucji populacji w czasie. Chcemy przygotować serię wykresów (np. co k pokoleń), które na płaszczyźnie 2D zobrazują populację oraz optymalny fenotyp w danym pokoleniu.
- Jak wpływa siła selekcji σ^2 na adaptację populacji? Co dzieje się w przypadku braku selekcji ($\sigma \to \infty$)? Czy przy założonym σ^2 da się ustalić odpowiednie parametry mutacyjności organizmów, aby adapracja była możliwa, a populacja nie wymiera?
- Jakie mutacje się propagują? Czy da wskazać się jakie zmiany powodują przeżycie organizmu, a jakie sprawiają, że jest eliminowany?
- Wszelkie inne pomysły na ocenę przebiegu ewolucji w kontekście przyjętych założeń/parametrów są mile widziane.

Możliwe rozszerzenia modelu

Rozszerzenie 1: Modelowanie adaptacji przestrzennej

Wprowadzamy dodatkową strukturę przestrzenną, w której każdy osobnik o posiada swoje położenie w płaszczyźnie (x, y), obok fenotypu $p \in \mathbb{R}^n$. Osobniki rozmnażają się w pobliżu swojego aktualnego położenia, a wartość ich fitness zależy nie tylko od fenotypu, ale również od położenia w środowisku.

- Każdy osobnik ma położenie $o = (x, y, p)$.
- Optimum fenotypowe zależy od położenia w przestrzeni:

  $$\alpha(x, y) = \alpha_0 + g(x, y)$$

  gdzie $g(x, y)$ odpowiada na pytanie jaki jest oczekiwany fenotyp w danym punkcie przestrzeni (x,y). Dla prostoty możemy przyjąć liniowy gradient:

  $$g(x, y) = \beta x$$

  gdzie β określa siłę zmiany środowiska.
- Reprodukcja: potomkowie są umieszczani w pobliżu rodzica zgodnie z rozkładem normalnym:

  $$(x', y') = (x, y) + \mathcal{N}(0, \sigma_d^2 I)$$

  gdzie σ_d określa zakres losowej dyspersji potomka.

- Fitness osobnika o określamy jako:

$$\phi(o) = \exp\left(-\frac{\|p - \alpha(x,y)\|^2}{2\sigma^2}\right)$$

Efekty modelu:

- Osobniki powinny stopniowo przemieszczać się w kierunku bardziej korzystnych rejonów przestrzeni.
- Powstaje struktura zależna od położenia: w mniej korzystnych rejonach przetrwają tylko te osobniki, które zmienią fenotyp w odpowiednim kierunku.
- Zależność przestrzenna może prowadzić do powstania lokalnych nisz ekologicznych.

Rozszerzenie 2: Populacja płciowa

Wprowadzamy populację płciową, gdzie rozmnażanie wymaga dwóch osobników (samca i samicy), a fenotyp potomstwa jest wynikiem rekombinacji cech fenotypowych rodziców.

- Każdy osobnik ma płeć $s \in \{M, F\}$ (samiec lub samica).
- Do rozmnażania wymagane jest znalezienie partnera o przeciwnej płci.
- Fenotyp potomstwa p_{child} jest kombinacją fenotypów rodziców p_m i p_f:

  $$p_{child} = \lambda p_m + (1-\lambda)p_f + \mathcal{N}(0, \xi^2)$$

  gdzie $\lambda \sim \mathcal{U}(0,1)$ określa proporcję dziedziczonych cech.
- Fitness osobnika określa jego szanse na znalezienie partnera: osobniki o wyższym fitness częściej dobierają się w pary.

Efekty modelu:

- Populacja ewoluuje wolniej niż w modelu bezpłciowym, ponieważ nie każde osobniki mogą się rozmnażać.
- Możliwa jest silniejsza selekcja naturalna – tylko osobniki z wysokim fitness skutecznie rozmnażają się.
- Rekombinacja może umożliwiać szybsze dostosowanie do zmieniającego się środowiska.

Rozszerzenie 3: Wiele optimum i specjacja

Wprowadzamy wiele różnych optimum fenotypowych w przestrzeni fenotypowej, co może prowadzić do powstawania oddzielnych subpopulacji.

- Zamiast jednego optimum α, definiujemy zbiór S różnych optimum:

  $$S = \{\alpha_1, \alpha_2, ..., \alpha_k\}.$$

- Fitness osobnika zależy od najbliższego optimum:

  $$\phi(o) = \max_{\alpha_i \in S} \exp\left(-\frac{\|p - \alpha_i\|^2}{2\sigma^2}\right).$$

- Jeśli osobnik znajduje się w pobliżu jednego z optimum, jego potomkowie również mają większe prawdopodobieństwo pozostania w tym obszarze fenotypowym.
- Możemy wprowadzić mechanizmy izolacji reprodukcyjnej: osobniki zbyt różne fenotypowo nie mogą się krzyżować.

Efekty modelu:

- Populacja może podzielić się na oddzielne subpopulacje, każda dostosowana do innego optimum.
- Jeśli optimum zmieniają się dynamicznie, możliwe są **przejścia ewolucyjne**, gdzie część populacji migruje między optymalnymi punktami.
- Powstają warunki do **specjacji sympatrycznej**, gdzie populacja dzieli się na dwie lub więcej grup bez geograficznej izolacji.

Problem rozwiązujemy w grupach 2/3-osobowych. Zaliczenie projektu składa się z zaprezentowania wyników podczas zajęć oraz umieszczenie kodu rozwiązania na GitHub. Wszelkie niejasności i rozwój Państwa rozwiązań będziemy dyskutowali na zajęciach, ale zachęcam aby w miejscu gdzie widzimy rodzaj dowolności podejmować decyzje projektowe (wraz z ich zdroworozsądkowym uargumentowaniem, uwzględniając zasadę, że wszelka decyzja jest lepsza niż jej brak). Dobrej zabawy!

Oddawanie rozwiązań Zaliczenie projektu polega na założeniu repozytorium na GitHub, które będzie zawierało:

- kody źrodłowe zaimplementowanego modelu
- fragment koodu, który pozwoli na przykładowe uruchomienie modelu
- raport z podsumowaniem realizacji projektu, tj. przyjęte założenia programistyczne i modelowe, zrealizowane/zaimplementowane zagadnienia, przedstawione w formie graficznej i przedyskutowane wyniki, które udało się osiągnąć za pomocą modelu.

i dodanie jako współtworzącego repozytorium (contributor) użytkownika o id: `storaged`. Termin oddawania rozwiązań: NN.NN.NNNN, godz. 12:00 (południe).