

Załącznik do sprawozdania

Sprostowanie i uzupełnienie do zadania 5

Needleman-Wunsch, a Smith-Waterman

Można wyszczególnić różnicę między tymi algorytmami porównując je do siebie, wiedząc iż jeden jest do globalnych dopasowani, a drugi do lokalnych. Trzema głównymi różnicami są:

	<i>Algorytm Needleman-Wunsch'a</i>	<i>Algorytm Smith-Waterman'a</i>
<i>Inicjalizacja</i>	1 kolumna i wiersz podlegają tzw. "karze".	1 kolumna i wiersz ustawione na wartość zero.
<i>Punktowanie</i>	Wynik może być ujemny	Wynik negatywnej punktacji ustawiony na 0.
<i>Śledzenie tzw. „traceback”</i>	Rozpoczęcie od komórki w prawym dolnym rogu macierzy i zakończenie na lewej górnej komórce.	Rozpoczęcie od najwyższego wyniku i zakończenie na wartości 0.

Macierz

Funkcja ta przyjmuje 3 argumenty:

1. 1 Sekwencja
2. 2.Sekwencja
3. Znak pustego znaku

W funkcji deklarujemy iż mimo braku przekazania będzie ona pracowała na zmiennej globalnej ALN_PATHWAYS.

Zapisujemy długość łańcuchów gdzie użyjemy tych wartości do utworzenia macierzy i generujemy macierz punktacji o wymiarach odpowiednio 1+długość każdej sekwencji. Wszystkie elementy pierwszego wiersza i pierwszej kolumny są ustawione na 0. Jednocześnie pierwszy wiersz i kolumna pozwala na dopasowanie jednej sekwencji względem drugiej w dowolnej pozycji, a ustawienie ich na 0 powoduje, że luka końcowa jest wolna od kary.

Punktacja

Oceniamy każdy element od lewej do prawej, od góry do dołu macierzy, biorąc pod uwagę wyniki zamiany (punktacja po przekątnej) lub dodania luk (punktacja pozioma i pionowa). Jeśli żaden z wyników nie jest pozytywny, to element otrzymuje 0. W przeciwnym razie używany jest najwyższy wynik i zapisywane jest źródło tego wyniku. Uzyskany wynik zwracamy wraz z uzyskaną macierzą zapełnioną liczbami.

Wyszukanie ścieżek, tzw. „traceback”

Począwszy od elementu z najwyższą punktacją, śledzimy wstecz na podstawie źródła każdej punktacji za pomocą rekurencji, aż do napotkania 0. W tym procesie generujemy segmenty, które mają najwyższy wynik podobieństwa w oparciu o dany system punktacji. W celu uzyskania następnego najlepszego dopasowania stosujemy znów proces wyszukiwania/śledzenia/traceback'u zwał jak zwał. Zaczynamy od drugiego najlepszego wyniku.

Spis zdefiniowanych funkcji(wartość wzrosła względem wykazu na sprawozdaniu wysłanym wcześniej)

```
def print_alns_only(ALIGNMENTS)
```

Wytlumaczona w 1 prezentacji.

```
def find_each_patch(c_i,c_j,path = '')
```

Jest to funkcja, której celem jest znalezienie ścieżek. Opisane pod nagłówkiem "Wyszukanie ścieżek, tzw. „traceback”"

```
def point(SEQUENCE_1,SEQUENCE_2,GAP_CHARACTER)
```

Wyliczenie punktacji oraz zapełnienie macierzy wartościami. Opisane pod nagłówkiem „Punktacja”

```
def check_seq(SEQ_1,SEQ_2)
```

Funkcja na wyrost z celem ,ale jednak sprawdza czy zadane sekwencje są sekwencjami aminokwasowymi.

```
def save(ALIGNMENTS)
```

Zapisuje uzyskaną wartość punktową jak również przyrównania które zostały uzyskane. Ponowne uruchomienie programu powoduje zastąpienie starego pliku nowym.

Czas

Poprawiono zliczanie czasu działania funkcji rekurencyjnej. Wywołanie punktu startowego i końcowego dla zegara następują w sąsiedztwie do wywołania funkcji „find_each_patch(i,j)”. Uzyskany czas obliczeń jest ukazywany

Przerobienie ścieżki na sekwencje

Ostatnia pętla w kodzie iterując po zawartość listy, ALN_PATHWAYS ,z zawartymi ścieżkami odtwarza sekwencje, co zostało napomniane również w nagłówku „Wyszukanie ścieżek, tzw. „traceback”". Uzyskane przyrównania zostają dopisywane do listy ALIGNMENTS, którą na końcu działania programu odczytuje wywołana funkcja „print_alns_only(ALIGNMENTS)”.

Kończąc

Opis działania, jak również problemy go trapiące są zawarte w PDF'ie oraz w tym pliku.

Działania jednego ze studentów: Dominik.

Jednocześnie prace nad tym zagadnieniem będą kontynuowane w czasie wolnym wraz z próbą implementacji kodu do jako wykonawczy do aplikacji WPF, czyli stworzenia pełnej aplikacji desktop'owej.

Więcej będzie tu: https://github.com/DesertFoxFenek/SW_Agorithm-for-BioInf

2021-06-09

Bioinformatyka lic. Semestr IV
Dominik Lisiecki – Desert_Fox_Fenek
Przemysław Wieczorek – Ciemny99