

Elementy Bioinformatyki

Temat 15. Filogeneza dystansowa. Typ A. Dla danej macierzy odległości między liśćmi utworzyć ważone drzewo ukorzenione metodą UPGMA oraz nieukorzenione metodą NJ. Wylistowanie obu drzew oraz możliwość automatycznej weryfikacji zgodności ich topologii (tj. z pominięciem długości krawędzi)

1. Opis problemu

Naszym zadaniem było utworzenie programu, który na wejściu otrzymuje macierz odległości między liśćmi, na wyjściu zaś zwraca ważone drzewo ukorzenione metodą UPGMA oraz nieukorzenione metodą NJ. W naszym projekcie nie zaimplementowano mechanizmu automatycznej weryfikacji zgodności topologii otrzymanych drzew.

2. Algorytmy

- **Algorytm UPGMA**

Pseudokod:

```
clusters [|L|]
while ( clusters . length > 1):
    calculate distances between clusters
        (sum of distances between cluster members
         divided by product of cluster cardinalities)
    find the lowest distance
    merge the closest clusters
```

- **Algorytm NJ**

Data: ultrametric matrix d for set L .

Q - matrix: $Q(i, j) = (n - 2)d(i, j) - \sum_{k=1}^n d(i, k) - \sum_{k=1}^n d(j, k)$

Distance from the pair members to the new node:

$$d'(f, u) = \frac{1}{2}d(f, g) + \frac{1}{2(n-2)}(\sum_{k=1}^n d(f, k) - \sum_{k=1}^n d(g, k))$$

$$d(g, u) = d(f, g) - d'(f, u)$$

Pseudokod:

```
clusters [|L|]
while ( number of clusters > 2):
    calculate Q- matrix
    find the lowest q- distance
    merge the q- closest clusters
    update distances
merge last 2 clusters
```

3. Przykładowe dane wejściowe/wyjściowe

Wejście programu:

Ultrametryczna macierz odległości

```
a,b,c,d,e
0 ,8 ,8 ,5 ,3
8 ,0 ,3 ,8 ,8
8 ,3 ,0 ,8 ,8
5 ,8 ,8 ,0 ,5
3 ,8 ,8 ,5 ,0
```

Wyjście programu:

Drzewo utworzone przy pomocy metody UPGMA

```
[[8.0]]
|-[[5.0]]
| |-[[3.0]]
| | |---- [[a]]
| | |`---- [[e]]
| | `---- [[d]]
| `---- [[3.0a]]
|   |---- [[b]]
|   `---- [[c]]
```

Drzewo utworzone przy pomocy metody NJ (pomimo faktu, że algorytm NJ zwraca drzewo nieukorzenione, dla łatwiejszej wizualizacji jest ono zaprezentowane jako drzewo ukorzenione z pewnym węzłem wybranym jako korzeń tegoż drzewa)

```
[[g]]
|---- [1.5-<-[a]]
|---- [1.5-<-[e]]
`-[1.75-<-[h]]
|---- [1.0-<-[d]]
|   `-[4.0-<-[f]]
|       |---- [1.5-<-[b]]
|       `---- [1.5-<-[c]]
```