Alec Mabhiza Chirawu - 亚历克上 - M202161029

Advanced Coding
Assignment 1

亚历克上 M202161029 (Ex 1) → 3, 6, 7, 10
Page 38, 39

Assignment 1

3) First-order entropy

a) $-4\left(\frac{1}{4}\right)\log_2\left(\frac{1}{4}\right) \Rightarrow 2$ bit/symbol

b) $-1\left(\frac{1}{2}\right)\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\frac{1}{4} - \frac{2}{8}\log_2\frac{1}{8} \Rightarrow \frac{7}{4} = 1{,}75$ bit/symbol

c) $-0{,}505\log_2 0{,}505 - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - 0{,}12\log_2 0{,}12 \Rightarrow 1{,}739818$ bit/symbol

6) a) Probability of the sequence ⇒

$A = 21 \Rightarrow \frac{21}{84} \quad P(A) = \frac{21}{84}$

$T = 23 \Rightarrow \frac{23}{84} \quad P(T) = \frac{23}{84}$

$G = 16 \Rightarrow \frac{16}{84} \quad P(G) = \frac{4}{21}$

$C = 24 \Rightarrow \frac{24}{84} \quad P(C) = \frac{6}{21}$

The first order entropy $\Rightarrow H = -\sum_{i=1}^{n} P(i)\log_2(P(i))$

$$-\frac{21}{84}\log\frac{21}{84} - \frac{23}{84}\log_2\frac{23}{84} - \frac{4}{21}\log_2\frac{4}{21} - \frac{6}{21}\log_2\frac{6}{21} \Rightarrow 1{,}9837 \text{ bit/symbol}$$

The second order entropy
$$\Rightarrow -\frac{3}{42}\log_2\frac{3}{42} - \frac{9}{42}\log_2\frac{9}{42} - \frac{12}{21}\log_2\frac{12}{21} - \frac{8}{42}\log\frac{8}{42} - \frac{1}{21}\log\frac{1}{21}$$
$$\Rightarrow 1{,}87436 \text{ bit/symbol}$$

Third order entropy ⇒

$ATG = 1/28$

$CTT = 7/28 \Rightarrow 1/4$

$AAC = 6/28 = 3/14$

$CTG = 8/28 \Rightarrow 2/7$

$AAG = 5/28$

$CCG = 2/28 \Rightarrow 1/14$

$$\Rightarrow -\tfrac{1}{28}\log_2 \tfrac{1}{28} - \tfrac{1}{4}\log \tfrac{1}{4} - \tfrac{3}{14}\log_2 \tfrac{3}{14} - \tfrac{2}{7}\log \tfrac{3}{7} - \tfrac{5}{28}\log_2 \tfrac{5}{28} - \tfrac{1}{14}\log \tfrac{1}{14}$$

$\Rightarrow$ 2,3801 bit/symbol

Forth order entropy ⇒

~~ATCG~~

$ATGC = 1/21$

$TTAA = 2/21$

$CCTG = 1/21$

$CTTA = 2/21$

$ACCT = 2/21$

$GAAG = 3/21$

$CTTC = 1/21$

$GCTT = 1/21$

$CGCT = 1/21$

$AACC = 3/21$

$TGCT = 1/21$

$GAAC = 1/21$

$CCGC = 1/21$

$GCTG = 1/21$

$TTCT = 1/21$

$$\Rightarrow -10(\tfrac{1}{21})\log_2 \tfrac{1}{21} - 4(\tfrac{2}{21})\log_2 \tfrac{2}{21} - \tfrac{1}{7}\log \tfrac{1}{7}$$

$\Rightarrow$ 3,78494 bit/symbol

6b) Based on the results of the above entropies, most igerence approaches on alignment and the quality of a multiple sequence alignment can have a strong diversity and the original sequence can be fully constructed.

7) a) The algorithm is in the file.

b) The output is shown on the image.

- The entropy has increased from 6.34506 to 7.7450.
- The difference is slim due to the color combinations of the image I used.

c) First entropy => 6,34506 bit/symbol
and The second entropy => 7,745 bit/symbol

I have discovered that the work is obtained from ordered ~~s~~ sequence, the amount of entropy is also a measure of the pixel disorder or randomless. The concept of entropy provides deep insight into the direction of spontaneous change for many everyday phenomena.

```
captaincheq@captaincheq-ThinkPad-11e ~/Desktop/test $  cd /home/cap
/python3 /home/captaincheq/.vscode/extensions/ms-python.python-2021
launcher 41841 -- /home/captaincheq/Desktop/test/Image_entropy.py
6.34506244304
7.74500367689
captaincheq@captaincheq-ThinkPad-11e ~/Desktop/test $
```

10) Determining ~~[illegible]~~ whether the following are uniquely decododable or not :-

Simple method :-

1) Take a look on the group codewords
2) If the group has no prefix found then its uniquely decodable code else if prefix found then add D_s (dangling suffix).
3) If D_s = some codeword then its not uniquely decodable else if all D_s are over then its uniquely decodable

∴ most of the time all prefix code are uniquely decodable

a) $\{0, 01, 11, 111\}$

There is prefix 0 ~~0~~ 1

0 is a prefix

1 is a dangling suffix

The new list is :-

$\{0, 01, 11, 111, 1\}$

Since the new value is equal to the second dangling suffix, we conclude that :-

It is not uniquely decodable.

b) $\{0, 01, 110, 111\}$

There is prefix 0 ~~0~~ 1

First 0 is a prefix

1 is a D_s

The new list is :-

$\{0, 01, 110, 111, 1\}$

The new value = to the new dangling suffix

then :-

It is **not** uniquely decodable

c) $\{0, 10, 110, 111\}$

All values are prefix

So we conclude that :-

Its uniquely decodable

d) $\{1, 10, 110, 111\}$

There is prefix 1 0 0

The first prefix is 1

D_s is 0

New list :-

$\{1, 10, 110, 111, 0\}$

The new dangling suffix = to 0.

So we conclude that :-

Its not uniquely decodat