Artificial Intelligence Term Project Spring 2017

0556087 / 資工碩一 / 謝禎原

想法說明

將同類(依 mutation 大小決定何為同類)的 patterns，建成一棵 Tree 的形式, Tree 的 Root(Initial State)為第一個 Sequence 的 Patterns，第二層為第二個 Sequence 的 Patterns … 以此類推．因此，最後取所有 Tree 中從 root 到 leaf 長度為 50 的 Path 即為一組可能解，將所有的可能解中的 Patterns 當做 Input，針對各個位置(1~15)進行投票分析，決定各個位置應該是哪個單字(A,T,C,G)，最後 Output 出 Significant Pattern .

搜尋設定

➢ State :

整個 forest(因為有多棵 Trees)當下的狀態.

➢ Operation :

- Add node

  當一個 Pattern 進來，會逐一跟每棵 Tree 的每個 leaves 比較是否相近(也就是是否符合 mutation 條件),當相近時，繼續沿著 path 往上比，直到比到 Root Node 為止，途中只要遇到一個 Node 為不相近時，就 break;反之，則成為該 leaf 的 child Node .

- Remove unchanged leaves

  當一個回合(也就是處理完一層 Sequence)結束時，會針對該回合沒有加入任何 child Node 的分枝進行 Pruning，來縮小之後的 Search Space.

- Remove unchanged trees

  一樣在每一個回合結束時,會針對該回合沒有加入任何 child Node(也就是說，在該回合這棵 Tree 的「所有」分枝都沒加入任何 child Node)的 Tree 進行刪除，來縮小之後的 Search Space.

- **Initial State :**

  初始狀態為 985 棵 Trees，各別以第一個 Sequence 的 Pattern 為 Root Node .

- **Goal Test :**
  - **是否有 Tree 的 height 是否已達到 50 (有 50 個 Sequence)？** 若是，就代表已經走完所有 Sequences, 取所有長度為 50 的 Path 上的 Pattern 進行分析 .
  - **或是否已經沒有任何 Tree 了 ？** 由於在判斷兩個 Pattern 是否相近的條件上設定的關係(考量到執行所需時間)，可能會導致還沒跑到 height 50 Tree 就已經被砍光了，因此只好退而求其次，取當下「深度達最深」的所有有 leaves 到 Root Node 的 Path 上的所有 Pattern 來進行分析.

補充

- **兩個 Pattern「相近」的條件**

  兩個 Pattern 相近的意思也就是這兩個 Pattern 是由同一個 Significant Pattern 所 mutate 而成的，也就是，兩個 Pattern「最多」會差 2 * mutation 個單字,

  e.g. (以最多 5 個 mutation 為例子)
  Significant Pattern : AAAAAAAAAAAAAA
  Pattern A : CCCCCAAAAAAAAA
  Pattern B : AAAAAAAAAAGGGGG

  但是，我在判斷兩個 Pattern 是否相近的條件下是以最多差 7 個單字來作為判斷(不管是 q2 還 q3)，也就是考量到執行時間的關係，越高的容忍度會囊括更多的 Patterns，讓所有 Tree 的分枝增加許多，導致處理時間拉長. 實際測試過，當我把容忍度調到 8 個單字時，在自己電腦所需處理的時間大概是容忍度 7 個單字的 60 倍 .

結果截圖

```
Question 1:
Target Pattern : ATTACCGTTAAGCTG
Seq 0 :{ "ATTACCGTTAAGCTG" ,554; }
Seq 1 :{ "ATTACCGTTAAGCTG" ,520; }
Seq 2 :{ "ATTACCGTTAAGCTG" ,540; }
Seq 3 :{ "ATTACCGTTAAGCTG" ,499; }
Seq 4 :{ "ATTACCGTTAAGCTG" ,157; }
Seq 5 :{ "ATTACCGTTAAGCTG" ,204; }
Seq 6 :{ "ATTACCGTTAAGCTG" ,710; }
Seq 7 :{ "ATTACCGTTAAGCTG" ,904; }
Seq 8 :{ "ATTACCGTTAAGCTG" ,893; }
Seq 9 :{ "ATTACCGTTAAGCTG" ,598; }
Seq 10 :{ "ATTACCGTTAAGCTG" ,442; }
Seq 11 :{ "ATTACCGTTAAGCTG" ,496; }
Seq 12 :{ "ATTACCGTTAAGCTG" ,907; }
Seq 13 :{ "ATTACCGTTAAGCTG" ,454; }
Seq 14 :{ "ATTACCGTTAAGCTG" ,153; }
Seq 15 :{ "ATTACCGTTAAGCTG" ,382; }
Seq 16 :{ "ATTACCGTTAAGCTG" ,746; }
Seq 17 :{ "ATTACCGTTAAGCTG" ,139; }
Seq 18 :{ "ATTACCGTTAAGCTG" ,284; }
Seq 19 :{ "ATTACCGTTAAGCTG" ,156; }
Seq 20 :{ "ATTACCGTTAAGCTG" ,446; }
Seq 21 :{ "ATTACCGTTAAGCTG" ,507; }
Seq 22 :{ "ATTACCGTTAAGCTG" ,36; }
Seq 23 :{ "ATTACCGTTAAGCTG" ,612; }
Seq 24 :{ "ATTACCGTTAAGCTG" ,843; }
Seq 25 :{ "ATTACCGTTAAGCTG" ,628; }
Seq 26 :{ "ATTACCGTTAAGCTG" ,200; }
Seq 27 :{ "ATTACCGTTAAGCTG" ,964; }
Seq 28 :{ "ATTACCGTTAAGCTG" ,622; }
Seq 29 :{ "ATTACCGTTAAGCTG" ,777; }
Seq 30 :{ "ATTACCGTTAAGCTG" ,572; }
Seq 31 :{ "ATTACCGTTAAGCTG" ,692; }
Seq 32 :{ "ATTACCGTTAAGCTG" ,715; }
Seq 33 :{ "ATTACCGTTAAGCTG" ,628; }
Seq 34 :{ "ATTACCGTTAAGCTG" ,761; }
Seq 35 :{ "ATTACCGTTAAGCTG" ,352; }
Seq 36 :{ "ATTACCGTTAAGCTG" ,164; }
Seq 37 :{ "ATTACCGTTAAGCTG" ,388; }
Seq 38 :{ "ATTACCGTTAAGCTG" ,78; }
Seq 39 :{ "ATTACCGTTAAGCTG" ,980; }
Seq 40 :{ "ATTACCGTTAAGCTG" ,561; }
Seq 41 :{ "ATTACCGTTAAGCTG" ,608; }
Seq 42 :{ "ATTACCGTTAAGCTG" ,753; }
Seq 43 :{ "ATTACCGTTAAGCTG" ,314; }
Seq 44 :{ "ATTACCGTTAAGCTG" ,829; }
Seq 45 :{ "ATTACCGTTAAGCTG" ,974; }
Seq 46 :{ "ATTACCGTTAAGCTG" ,683; }
Seq 47 :{ "ATTACCGTTAAGCTG" ,701; }
Seq 48 :{ "ATTACCGTTAAGCTG" ,141; }
Seq 49 :{ "ATTACCGTTAAGCTG" ,736; }
------
```

```
Question 2:
Target Pattern : CTGTTAGATCAACTG
Seq 0 :{ "CTGTTTGATCAGCTG" ,181; }
Seq 1 :{ "CTGCTAAATCAAGTG" ,292; }
Seq 2 :{ "ATGATAGATAAACTG" ,111; "CTGTCTGACCTAATG" ,505; }
Seq 3 :{ "CTGTTAGCTCAAACG" ,268; "CTTATAAGTGAACTG" ,877; }
Seq 4 :{ "ATGTCTGATCAACAG" ,679; }
Seq 5 :{ "CAGTTTGGTGATCTG" ,61; "CTGGCGCATCAACTG" ,428; "CCGTTAGGTCAATGG" ,585; }
Seq 6 :{ "TCGTTAGATCAACCG" ,940; }
Seq 7 :{ "CTGTGAGATCAATTG" ,623; }
Seq 8 :{ "CTTTTATATAAAGCG" ,242; "ATGTCAAATCGACTG" ,429; "CTGTTACCTTTATTG" ,959; }
Seq 9 :{ "ATGTTAGATCAACTT" ,70; }
Seq 10 :{ "ATTTTAAGTCAACCG" ,547; "CTGTTAATTCGACTG" ,983; }
Seq 11 :{ "CAGTGAGTTCTACTG" ,790; }
Seq 12 :{ "CTGTTAGTTCAGCTG" ,427; }
Seq 13 :{ "GTATTAGATCATCCA" ,69; "ATGTTAGATCACCGG" ,828; }
Seq 14 :{ "CTGGACGTCCAACTG" ,197; }
Seq 15 :{ "CGGTGATCTCAGCTG" ,363; }
Seq 16 :{ "CTGATAGATAAAGCG" ,376; }
Seq 17 :{ "CTTTTACATCAACTG" ,170; }
Seq 18 :{ "CTATTAGATCAACTT" ,935; }
Seq 19 :{ "CTGTCAAATTAACTG" ,154; "CTGTTCGATTTGCTC" ,499; }
Seq 20 :{ "CTCATAGATCAACTG" ,167; }
Seq 21 :{ "CCTTTATATCGCCTG" ,51; "TGGTCAGATCATCTG" ,917; }
Seq 22 :{ "CTATCAGATTAACGG" ,484; "CTCATAGATGAGCTG" ,543; "CTGTGGAATCAGATG" ,638; }
Seq 23 :{ "CCGTTAGTTGAAGTG" ,734; }
Seq 24 :{ "CGGTAATATCAAACG" ,583; "CTATTAGACCACCTA" ,933; }
Seq 25 :{ "CGGTTAGATCAACTA" ,677; }
Seq 26 :{ "CTGTCAGTTCCTCTG" ,497; }
Seq 27 :{ "CTTTTAGAGCACCTG" ,190; }
Seq 28 :{ "ATGTTAGTTCAACAG" ,560; }
Seq 29 :{ "ATGTCAGATCCAATG" ,509; }
Seq 30 :{ "CTGTTAGCTCAACTA" ,682; }
Seq 31 :{ "CTGATACATCGCCTA" ,94; "ATGTTCAGTCAACTG" ,377; "GTGTTGGTAGAACTG" ,630; }
Seq 32 :{ "CTGTTGGTTCGACTG" ,597; }
Seq 33 :{ "CTGTTTGATTCCCTC" ,420; }
Seq 34 :{ "CTCATAGATCAACTA" ,562; "CTGTTAGTTCAAACG" ,719; }
Seq 35 :{ "CTCTCAAATCATCTG" ,76; "GTATTAGATTGATTG" ,650; "CCGTAAGATAAAGTG" ,871; }
Seq 36 :{ "CTGTTAGACCAACAT" ,359; }
Seq 37 :{ "ATGTTTATTCAACCG" ,810; "CCGTTAGATCCACGC" ,853; "ATGTGCAATCAACTG" ,961; }
Seq 38 :{ "CTATGAGCTCTACTG" ,245; "GTGTCAGTTCAACTA" ,779; }
Seq 39 :{ "CTGTTCGTTGAACTG" ,36; "CTTATAGATCTGGTG" ,865; }
Seq 40 :{ "CTGCAAGATCACACG" ,346; "TTGTCATATCAATGG" ,379; "CTGTTAGACAAGTTG" ,568; }
Seq 41 :{ "CTGTTAGAACTTGTG" ,216; "GAGATAGAAAAACTG" ,849; }
Seq 42 :{ "CTGAAAGCTGAACGG" ,539; "CTGATGGATCGACCA" ,620; "CTATTAAATCAACCG" ,762; }
Seq 43 :{ "ATGTTTGGTCAGGTG" ,375; "CAGTCAGTTCTAATG" ,419; }
Seq 44 :{ "CAGTTAGATATCCTG" ,602; }
Seq 45 :{ "ATCTTATATCGACTC" ,42; "CTGTTAAATCCAATG" ,654; }
Seq 46 :{ "CTGTTAGATCAACTG" ,13; "CTTCTTGATCGCCTG" ,936; }
Seq 47 :{ "CTCTTAGAAATATTG" ,292; "CTGTTAGATCAAATT" ,842; }
Seq 48 :{ "CTCTTAGCTGAACTT" ,43; "TTGTTGGATCGCCTT" ,111; }
Seq 49 :{ "GTGTTAGATAGGCTG" ,449; "TTCTAATATCAACAG" ,748; }
```

(由於版面過大, Question 3 直截部分，可以直接 Run Code 看結果)

```
Question 3:
Target Pattern : ATGAATAAACCGGAA
Seq 0 :{ "ACGATATGCCTGGAA" ,9; "ATGCCTGGAACTGTA" ,14; "ATGTGTCGAGAGTAA" ,111; "ATTTAATACTTGGAA" ,187; "CTTAGTTTGCCGCAA" ,263; "AGGAATTAACTCAGT" ,287; "ATAATTCAACATGGA" ,332; "CCCAA
AATCTCGAT" ,360; "ATCGATAAAGTCGCC" ,446; "ATTACAAGGCCGGGC" ,469; "TTTAATAAAAACGGT" ,567; "TTAATAAAAACGGTC" ,568; "CGGAATTGATTGTAA" ,582; "AAGAAGAGAACCAAT" ,595; "GAGAACCAATTGGAT" ,
00; "GTGCATATAAAAGAA" ,617; "ATAAAAGAATCTAAG" ,623; "GTAAATGAATTGCAT" ,651; "ATGAATAAGCCGTAT" ,676; "ATTAAGGAATCTTAA" ,703; "AGGAATCTTAAGGGA" ,707; "CAGAGTCGACCGTAC" ,723; "ATGTTTT
TCCGTTA" ,814; "ATTGCGAAACCCGGT" ,892; "TTGCGAAACCCGGTC" ,893; "AAAAATCCTCCCGAA" ,982; }
Seq 1 :{ "TTGAAATAAGCGCCT" ,44; "ATGACGAGAACGAAA" ,94; "ACGAATAAAGCTGAA" ,175; "CTGAAATATTCGCAA" ,185; "TCGCAACAACCTGAG" ,194; "CTATACATCCTGGAA" ,237; "GGACAGACAGCGGAA" ,301; "ATTG
TTGGCCGAAG" ,417; "ACCAACCCAGTGGAA" ,448; "AGTTATACATCAGAT" ,628; "GAAGATAAAACAGTA" ,742; "AAGATAAAACAGTAG" ,743; "ATGAAAAAAAAAGGG" ,929; "TGAAAAAAAAAGGGA" ,930; "GAAAAAAAAAGGGAA"
931; "AAAAAAAAAGGGAAC" ,932; }
Seq 2 :{ "CTGGATTGATCGCTA" ,96; "ATCGCTAAGCCTCAT" ,104; "ACCAATCAATAGGCG" ,326; "ATCAATAGGCGTTTA" ,330; "ATTCATGCGCATGAA" ,387; "ATGAATTAGATTCTA" ,397; "AAGACATAACCCAGA" ,418; "AGA
ATAACCCAGAT" ,419; "CTCTATGACTCCGAA" ,435; "GTTGAAAAATCAGGA" ,534; "TTGAAAAATCAGGAA" ,535; "AGGAATGATTGGTTA" ,545; "CTGGGGAAAATGGTA" ,607; "CATTATAAACGGGAT" ,627; "ATGAGTCAATGGCGC"
,686; "AGGAAAATTGGTGAA" ,740; "ATACATTACCCAGGA" ,762; "ATTTAAAACATGGAA" ,824; "CTGCCTATAACCTAA" ,878; "ATAACCTAAAGGTAA" ,884; "AAGGTAACACCACAA" ,892; "CTGCAACATTCGGCA" ,970; }
Seq 3 :{ "ATTAAAGTTCCAGAT" ,34; "ATTGATGACGCGGGC" ,80; "TTAAATTATCCTGTC" ,109; "GAGAATCAATGGAAA" ,242; "ATGGAAATCCCGTTA" ,250; "GTTATTGAACTATAA" ,261; "AACTATAAACGGTCT" ,268; "CTAT
TGAATCGGTC" ,299; "AATGAAAAACATGGA" ,382; "ATGAAAAACATGGAC" ,383; "AGTATTCGGCCGGAT" ,426; "CGGAATTAGATGAAA" ,479; "TTGGACCCACCTTAA" ,500; "CTAAATAAGCGGCTT" ,621; "GTAAATTAATTGAAA"
641; "ATTAATTGAAAGAAA" ,645; "ATTACGAAGCGGGCT" ,659; "ATGAAGACACGGCAA" ,680; "AAGAATCTATAGGAT" ,780; "TTAAGCATACCCCAA" ,850; "AAGCATACCCCAAAG" ,852; "TTTGTTAAACATAAA" ,912; }
Seq 4 :{ "ATTAATGAGAAGGAT" ,33; "CAGAAGATAAAGAAA" ,58; "ATAAAGAAACAAGGC" ,64; "AACAAGGCACTGTAA" ,71; "CTTAATAATGTGGCT" ,210; "AGGTATACAGGTAAA" ,283; "ATGTATCGCCGTAAA" ,381; "TTCATT
ATGCGCAC" ,513; "TGGAGAAAATCGTGA" ,550; "GTGACTAAAATAGCA" ,561; "ATGGAGAAGACTCAA" ,588; "ATGAAGCATTCATTA" ,676; "ATTCATTACCCATCA" ,683; "TTGGATGATCCGTTT" ,718; "TTTCATACGAAGGAA" ,7
1; "GTGAATTTTGTGCAA" ,776; "TGTTATCAACCGGGC" ,801; "GTTATCAACCGGGCA" ,802; "ATCAAGAAACCGGAA" ,953; "AAGTAATAACCTTTA" ,966; "AGTAATAACCTTTAA" ,967; "ACCTTTAATATGGAA" ,974; "TGGAAAAA
TACGGA" ,984; "GGAAAAAATACGGAC" ,985; }
Seq 5 :{ "CTTCGTGAACTCGAA" ,44; "TTGAAATCATCCGAT" ,63; "TTACTTGTACCGGAT" ,112; "ATGAATCAAGGCGGT" ,144; "ACTTATAGACGGTAG" ,195; "CTTATTAGATCGTTA" ,215; "ACGGATGATTCTGTA" ,320; "ATTC
TCAAACAAAT" ,420; "CAAAATATACTGGTG" ,444; "GTCAATCACTAGAAA" ,485; "AGAAATAGGCACGGA" ,495; "ATCCAGACACGGGCT" ,542; "CTTTGTAAATCGCAA" ,605; "ATTAATATAGTGAAT" ,634; "ATGGAAAAATAAGCT"
653; "AAAAATAAGCTAGAA" ,657; "AGCTAGAAACGGGTA" ,664; "ACGGGTACAGCGAGA" ,672; "ACGTGTATCACGGCA" ,694; "AGGCATACACAGAAA" ,710; "CAGAAAACACTACAA" ,719; "TTAAATAATATATAA" ,758; "AATAAT
TATAATAA" ,761; "ATATAATAACGCGGA" ,767; "TATAATAACGCGGAG" ,768; "AAGATCCAACGTTAA" ,786; "CTGATCAGGCCCGAT" ,841; "AAGTACCAACCGGAG" ,861; "ACGTACAAACGTCCA" ,904; "ACGGGTCTAACGCAA" ,9
6; "ACTTGTAAACGGCTA" ,972; }
Seq 6 :{ "ATTTAGAGACCCAAC" ,4; "GAGACCCAACCAGCA" ,9; "ATTAATTAACCCACA" ,38; "TTAATTAACCCACAG" ,39; "ATGAGTTTACCTGAG" ,156; "TTGCATATCCGGTAC" ,209; "GTATCTAAAACCGAC" ,277; "ACACATAA
GCGCAG" ,290; "AAGGTAAAAGCGGTT" ,320; "ATTTTGGAGCCGGAG" ,424; "GTGGACGATCTTGAA" ,451; "TTGAACAGCGCGAAT" ,461; "ATGAATTATGCCCCT" ,505; "TCGCCTAAATCGATA" ,535; "GTTAATGACCAAAAA" ,554
"ACCAAAAACTCGTAG" ,561; "ATGGACATTCCAAAG" ,642; "ATTATTAGACTGGAA" ,683; "ATGATTAGAGTTCAC" ,707; "ATTATCACACAGAAT" ,753; "CAGAATACAACCTAA" ,762; "GTTAAAAATGCCGAT" ,868; "GCGATTACAT
TGAA" ,893; "ATCAATTTAGTGAAA" ,922; "ATGCATAGGTGGATA" ,983; }
Seq 7 :{ "ATATATCCACCCAAT" ,12; "ATTTTTCTACGGCAA" ,25; "CCGTGTAATTAGGAA" ,61; "CATACAAAGTCGGAA" ,153; "ATGCATTTACCGGAA" ,229; "GTGAAAAGTCCCTCA" ,281; "TTACATAAACTCTAT" ,432; "ACTAT
ATCGCAGAA" ,488; "GTGATTCATCCGAAT" ,615; "ATCAAGTAGCTGGGT" ,647; "CTGAGTGCATCGGGT" ,728; "CTCAATGAAAGGTTA" ,772; "ATCAGGACGCTAGAA" ,834; "AGGAACAATCGGTTG" ,935; "ATAAAGCAAGGGACA" ,
52; "ATCACATAAAAAGAA" ,982; }
```