

Report

B05505004 朱信禕

1. 我的 thread 是開在 training decision tree 的時候，我會建立一個 pool 然後將開好使用者指定的 thread，接者再將 training 每棵 decision tree 視為是一個 job，由空閒的 thread 去負責 train。而這個 pool 是利用 condition variable 實做。

2. Sampling number: 300 Tree number: 20

| thread 數 | time(s) | 正確率(%) |
|----------|---------|--------|
| 1 | 31.57 | 85.67 |
| 2 | 31.36 | 84.57 |
| 3 | 31.85 | 84.39 |
| 4 | 31.85 | 85.18 |
| 5 | 33.45 | 85.14 |
| 6 | 32.16 | 85.24 |
| 10 | 30.30 | 85.41 |
| 20 | 32.16 | 84.84 |
| 30 | 33.54 | 85.39 |

由於系上工作站是 16 核心的機器，而從以上的數字我們可以發現 thread 數字開在 16 附近的時間最快，但當數字大於 16 後時間卻不會繼續下降，而是又會開始上升。另外一個原因是因為在這個實驗中我訓練的 tree 數量是 20 棵，所以當 thread 的數量大於 tree 的數量時，並不會有任何幫助，同時還會浪費時間在開 thread 上。

3. Sampling number: 1000 Tree number: 20

| thread 數 | instruction 數 |
|----------|---------------|
| 1 | 218252 |
| 2 | 218160 |
| 3 | 218162 |
| 4 | 218162 |
| 5 | 218160 |
| 6 | 218161 |
| 10 | 218161 |
| 20 | 218251 |
| 30 | 218161 |
| 100 | 218253 |
| 500 | 218161 |

4. Sampling number: 1000 Thread number: 50

| Tree 數 | instructions:u |
|--------|----------------------|
| 10 | 218162 802810 802810 |
| 20 | 218160 843457 843457 |

| | |
|-----|------------------------|
| 30 | 218161 1170230 1170230 |
| 40 | 218161 771720 771720 |
| 100 | 218161 966836 966836 |
| 200 | 218160 896293 896293 |
| 500 | 218161 844510 844510 |

5.我測試的時候我發現 sampling 的數量並不會對於正確率有很大的影響，但會大幅的影響程式的執行時間。然後當 thread 開的數量大過 cpu 數量時，時間就不太會顯著的下降，反而會持平。