# Meetrapport titel

## Namen en datum

Jeroen Huisen & Hendrik Cornelisse

6-4-2014

## Doel

Het doel van dit experiment is bepalen van toenamen van tijd in verhouding tot de toename van het aantal clusters. De vraag is dan: Hoeveel extra tijd kost het kiezen van 1 extra cluster bij kmeans clustering?

## Hypothese

Wij verwachten dat de extra tijd te bepalen valt via een lineaire verband. Omdat de berekeningen alleen meerdere keren moeten worden uitgevoerd. Het werkt alleen als de clusters niet gevonden worden maar dankzij de iteraties wordt gestopt. Hiervan gaan wij dan ook uit. Wanneer het eerder gevonden wordt kan dit namelijk andere waardes opleveren. Hier gaan we dus niet van uit! Ook zou de computer hardware zorgen voor verschillen maar we testen op 1 computer.

## Werkwijze

Het programma wordt opgestart en de afbeelding met k waarde is gekozen. De aangeleverde basetimer klasse houdt de totale tijd bij. De totale tijd wordt genoteerd en per k waarde 5x getest. Er wordt gebruik gemaakt van 3 afbeeldingen met alle 3 een andere aantal pixels.



1024 x 768 281 x 211 500 x 375

De k waardes worden getest met 3, 5, 7 en 20 zoals genoemd in de opgave. De tijd van de k waardes wordt van elkaar afgehaald. Deze worden vergeleken met elkaar en dan gekeken welk verband kan ontstaan.

Gebruikte hardware:

Intel i7 2670qm, 2.2 ghz

8 gb ram 1333 mhz

Windows 7 64bit

## Resultaten

Apple 1024 x 768

20 iteraties.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| K | 3 | 5 | 7 | 20 |
| Meting 1 | 9.79 s | 13.24 s | 16.49 s | 38.68 s |
| Meting 2 | 9.72 s | 13.23 s | 16.57 s | 38.65 s |
| Meting 3 | 9.78 s | 13.10 s | 16.49 s | 38.54 s |
| Meting 4 | 9.80 s | 13.05 s | 16.45 s | 38.79 s |
| Meting 5 | 9.76 s | 13.12 s | 16.51 s | 38.64 s |
| Gemiddelde | 48.85/5 = 9.77 s | 65.74 / 5 = 13.15 s | 82.51/5 = 16.50 s | 193.3/5=38.66 s |

Art 281 x 211

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| k | 3 | 5 | 7 | 20 |
| Meting 1 | 0.747 s | 1.00 s | 1.25 s | 2.93 s |
| Meting 2 | 0.751 s | 1.00 s | 1.26 s | 2.91 s |
| Meting 3 | 0.754 s | 1.01 s | 1.26 s | 2.88 s |
| Meting 4 | 0.745 s | 1.00 s | 1.26 s | 2.93 s |
| Meting 5 | 0.742 s | 1.00 s | 1.25 s | 2.90 s |
| Gemiddelde | 3.739/5 = 0.748 s | 5.01 / 5 = 1.00 s | 6.28/5 = 1.26 s | 14.55/5=2.91 |

Path 500 x 375

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| k | 3 | 5 | 7 | 20 |
| Meting 1 | 2.35 s | 3.16 s | 3.98 s | 9.26 s |
| Meting 2 | 2.35 s | 3.16 s | 3.97 s | 9.29 s |
| Meting 3 | 2.35 s | 3.33 s | 4.00 s | 9.31 s |
| Meting 4 | 2.36 s | 3.25 s | 3.99 s | 9.28 s |
| Meting 5 | 2.35 s | 3.16 s | 3.99 s | 9.35 s |
| Gemiddelde | 11.76/5=2.35 | 16.06/5=3.21 | 19.93/5=3.99 | 46.49/5=9.30 |

## Verwerking

Apple

13.15 – 9.77 = 3.38 3.38/(5-3) = 1.69 seconden toename per cluster 5 - 3

16.50 - 13.15 = 3.35 3.35/(7-5) = 1.68 seconden toename per cluster 7 - 5

38.66 – 16.50 = 22.16 22.16/(20-7) = 1.70 seconden toename per cluster 20 - 7

Art

1.00 – 0.75 = 0.25 0.25/(5-3) = 0.13 seconden toename per cluster 5 - 3

1.26-1.00 = 0.26 0.26/(7-5) = 0.13 seconden toename per cluster 7 - 5

2.91-1.26 = 1.65 1.65(20-7) = 0.13 seconden toename per cluster 20 - 7

Path

3.21-2.35 =0.86 0.86/(5-3) = 0.43 seconden toename per cluster 5 - 3

3.99-3.21 = 0.78 0.78/(7-5) = 0.39 seconden toename per cluster 7 - 5

9.30-3.99 = 5.31 5.31/(20-7) = 0.41 seconden toename per cluster 20 - 7

Toename clusters in verhouding met tijd is een linear verband. Dit kan afgeleid worden aan de gegevens hierboven. Het is niet helemaal precies maar dit ligt aan de metingen. Hiervan hadden er meer moeten verricht worden voor precieze waardes.

|  |  |  |
| --- | --- | --- |
| format | Total pixels | Toename in seconden |
| 281 x 211 | 59291 | 0.13 |
| 500 x 375 | 187500 | 0.41 |
| 1024 x 768 | 786432 | 1.69 |

0.41-0.13 = 0.28 0.28/(187500-59291) = 2.18 \* 10^-6 seconden toename per pixel en per cluster

1.69-0.41 = 1.28 1.28/(786432-187500) = 2.14 \* 10^-6 seconden toename per pixel en per cluster

1.69-0.13 = 1.56 1.56/(786432-59291) = 2.15 \* 10^-6 seconden toename per pixel en per cluster

Het is mogelijk dat zelfs per pixels een linear verband bestaat. Maar dit zijn te weinig metingen om dit met zekerheid te kunnen zeggen.

Als er vanuit gegaan wordt dat dit een linear verband is kan er een formule worden opgesteld welke luidt:

Nu een klein testje van de formule. De gegevens hieronder zijn van de apple afbeelding

|  |  |  |
| --- | --- | --- |
| 3 | 5 | 7 |
| 48.85/5 = 9.77 s | 65.74 / 5 = 13.15 s | 82.51/5 = 16.50 s |

Er wordt de waarde van 3 clusters genomen, 9.77. Hierbij wordt de toename berekent bij 2 extra clusters.

Dat is 2.16\*10^-6 \* 786432\*2. Dit leidt dan tot:

9.77+2.16\*10^-6 \* 786432\*2 = 13.17 s

13.17 seconden dit klopt ongeveer met de gemeten waarde. 13.15 zit namelijk niet ver van 13.17 af zeker niet omdat er meetonnauwkeurigheden zijn. Zoals eerder gezegd de formule is niet compleet accuraat omdat er te weinig metingen gedaan zijn. Deze formule werkt ook alleen met dezelfde code en met dezelfde hardware als gebruikt is tijdens het testen. Maar de hardware zou geen invloed mogen hebben op of het een lineaire verband is of niet.

## Conclusie

De tijd die het kost om 1 extra cluster te gebruiken, is variabel. Het is afhankelijk van de grootte van de afbeelding(aantal pixels). Natuurlijk ook van het aantal clusters maar het totaal aantal maakt niet uit om 1 extra cluster toe te voegen. De extra tijd het kost om k aantal clusters toe te voegen is lineair.

De opgestelde formule lijkt te voldoen maar deze zal in een volgend experiment degelijk getest moeten worden.

## Evaluatie

De toename is lineair net zoals voorspeld in de hypothese. De onderzoeksvraag is beantwoord alleen niet zo gedetailleerd als wij wouden. Dit komt door de meetnauwkeurigheden. Verschillende metingen lagen een beetje uit elkaar om het preciezer te maken hadden we mogelijk wel 100 testen per keer moeten doen inplaats van 5. Ook moeten we meer afbeeldingen formaten gaan gebruiken indien het aantal pixels getest moeten worden. Hieraan hadden wij niet goed genoeg gedacht tijdens het experimenteren.