'Bubble Sort'

George Bubble , François Bulle, Luca Bolla

Resum

Resultats: La relació obtinguda entre T i N es pot expressar mitjançant l'equació:

$$T = -3.945 \cdot 10^{-3} + 3.158 \cdot 10^{-6} \cdot N^2$$

Objectiu: Obtenir una model mitjançant un anàlisi estadístic que ens permeti predir el temps que triga l'algorisme d'ordenació en estudi, Bubble Sort, en ordenar vectors que tinguin unes grandàries que oscil·lin entre 500 i 1000 números.

Com s'ha fet: Es van generar 51 vectors de grandàries equiespaiades entre 500 i 1000 valors contenint números enters distribuïts uniformement entre 0 i 10,000. Es va aplicar l'algorisme d'ordenació a tots els vectors i es va modelar el temps d'execució (T) en segons en funció de la grandària (N) a través de regresió lineal simple.

Conclusió: Es pot predir amb gran precisió (R^2 = 99.9%) el temps que trigarà l'algorisme en fer l'ordenació.

Introducció

Bubble Sort és un senzill algorisme d'ordenació que funciona comparant cada element de la llista de valors amb el seu succesor, intercanviant-los de lloc si estan desendresats. Es necessari recórrer varies vegades tota la llista fins que no es necesitin més intercanvis, fet que indicarà que la llista està endreçada. El nom d'aquest algorisme prové de la forma que pujen els elements per la llista durant els intercanvis, como si fosin petites "bombolles". També es coneix com el mètode de l'intercanvi directe. Es considera l'algorisme més senzill de implementar.

L'ordre d'operacions que realitza aquest algorisme és de l'ordre $O(N^2)$.

Objectiu

L'objectiu d'aquest estudi és fer unes simulacions amb l'ordinador, posteriorment analitzar-les per estudiar quina és la relació entre N i T (al aplicar l'algorisme Bubble

Sort, que és el que estem estudiant) i, posteriorment, un cop ja ho es tingui tot lo anterior resolt, llavors determinar la precisió amb la que T es pot arribar a predir en funció de N.

Com s'ha fet

Es van generar uns quants vectors de grandàries entre cinc-cents i mil contenint números enters entre 0 i 10,000. Es va permetre que hi hagués repeticions.

Posteriorment, es van endreçar aquests vectors amb l'algorisme

Els experiments es van dur a terme utilitzant el programari R v.2.11.1 en un ordinador amb un microprocessador Intel Core 2 de 2.66GHz i 3.46 Gb de memòria RAM. El sistema operatiu emprat va ser el Microsoft Windows XP. El temps total d'execució de la simulació va ser de 94 segons.

Les variables recollides van ser la grandària del vector a endreçar (variable predictora) i el temps d'execució (variable resposta).

Anàlisi estadística

S'han calculat els estadístics descriptius de la variable temps. En l'anàlisi principal es va estudiar la relació entre el temps d'execució i la grandària del vector a través de regresió lineal simple. Posteriorment, es va fer una verificació de les premisses de linealitat, homoscedasticitat, normalitat i independència mitjançant un anàlisis dels residus. Si el model no era validat, llavors es provava de fer transformacions sobre les variables.

Resultats

Descriptiva

Aquesta taula conté la descriptiva dels temps d'execució en segons.

Mínim	1r quartil	Mediana	3r quartil	Màxim	Mitjana	Desviació Estàndard
0.8	1.211423	1.765	2.422321	3.2	1.84	0.707

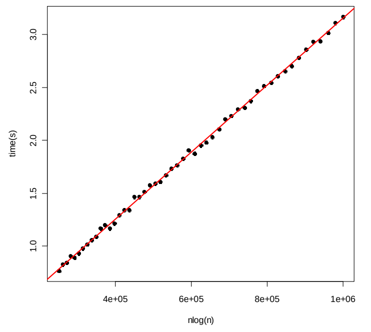
Anàlisi principal

L'Annex II mostra que el model amb la variable N com a predictora no compleix la premissa de linealitat. Sabent que el nombre d'operacions d'aquest algorisme és de l'ordre $O(N^2)$ es modela T en funció de N^2. La sortida obtinguda és la següent:

```
Call:
lm(formula = Bubble ~ N2, data = dades)

Residuals:
      Min        1Q    Median        3Q       Max
-0.042474 -0.014269 -0.002407  0.016572  0.054251

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.945e-03  9.065e-03  -0.435    0.665
N2           3.158e-06  1.451e-08 217.669   <2e-16
***
---
Residual standard error: 0.02297 on 49 degrees of
freedom
Multiple R-squared: 0.999,      Adjusted R-
squared: 0.9989
F-statistic: 4.738e+04 on 1 and 49 DF,  p-value: <
2.2e-16
```



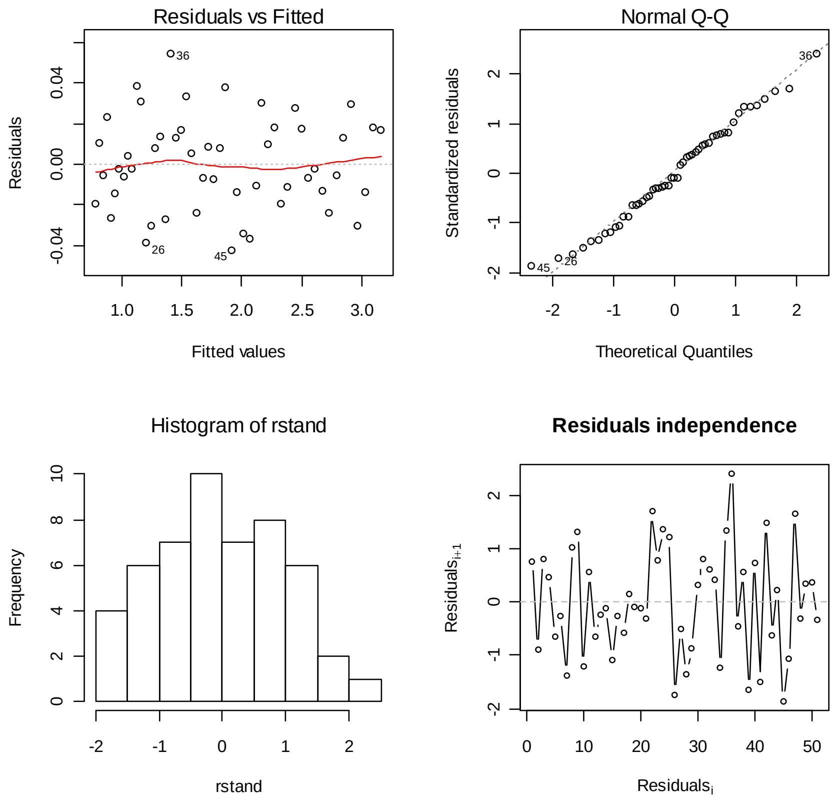
La R^2 ha augmentat de 99.1% a 99.9% elevant al quadrat la N.

El model resultant és:

$$T = -3.945 \cdot 10^{-3} + 3.158 \cdot 10^{-6} \cdot N^2$$

<u>II. Comprovació de les premisses</u>

La figura de més avall mostra els quatre gràfics de validació. Els residus en front dels valors predits mostren que amb la transformació de la variable predictora s'aconsegueix el compliment de la premissa de linealitat i es continua validant la homoscedasticitat. El QQ-Norm i l'histograma dels residus mostren resultats contradictoris, mentre el QQ-Norm plasma que els residus són Normals, l'histograma té una forma força irregular que posa en dubte la premissa de Normalitat. La premisa d'independència entre observacions consecutives sembla correcte, ja que no s'observa cap patró que indiqui relació entre els residus i els residus retardats.



Discussió

Els resultats confirmen la teoria ja coneguda de que el nombre d'operacions necessàries per endreçar una llista de valors són de l'ordre de N^2. Sembla raonable, doncs, que el temps d'execució també estigui relacionat amb l'ordre quadràtic de la grandària.

El coeficient de $3.15 \cdot 10^{-6}$, $IC_{95\%}$ = [$3.13 \cdot 10^{-6}$ a $3.19 \cdot 10^{-6}$] de la variable explicativa s'interpreta com que un increment de 100 unitats en la grandària del vector comporta un increment de 0.032 ($=3.15 \cdot 10^{-6} \cdot 100^2$).

La R^2 representa la proporció de variabilitat explicada pel model. En aquest cas, la previsió que podem fer és molt acurada. Per exemple, el valor esperat del temps emprat per una llista de 750 valors serà 1.77 s. $IC_{95\%}$ =[1.73 a 1.82]. Per altra banda, el temps estimat per l'endreçament per un vector concret d'aquesta mateixa longitud serà 1.773 s. $IC_{95\%}$ =[1.766 a 1.779].

Hem demostrat la relació entre el temps que trigarem en endreçar qualsevol llista de valors amb aquest algorisme o amb qualsevol altre algorisme amb un nombre d'operacions $O(n^2)$. Els resultats són extrapolables a totes les grandàries de vector (no hi ha cap motiu per pensar que no sigui així) i a qualsevol computador (avui en dia, tots els ordinadors són molt semblants)

Per la recerca futura, es recomana centrar-se en altres temes, perquè aquest ja ha quedat completament resolt.