ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΝΑΛΥΣΗ ΑΛΓΟΡΙΘΜΩΝ

Ενότητα 7β: Όρια Αλγόριθμων Ταξινόμησης

Μαρία Σατρατζέμη
Τμήμα Εφαρμοσμένης Πληροφορικής

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Μακεδονίας» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Όρια Αλγόριθμων Ταξινόμησης

- Μέχρι στιγμής εξετάσθηκαν μέθοδοι ταξινόμησης με πολυπλοκότητα της τάξης $\Theta(n^2)$ ή $\Theta(n\log_2 n)$.
- Τι εκφράζει η κάθε μία από αυτές τις τάξεις πολυπλοκότητας;
- Η μέθοδος της ταξινόμησης με εισαγωγή, της ταξινόμησης με επιλογή και της ταξινόμησης με ανταλλαγή (bubblesort) είναι οι λεγόμενες **ευθείες** (*straight*) μέθοδοι που είναι απλές στην κατανόηση, την κωδικοποίηση και την ανάλυση αλλά αποτελούν μία οικογένεια αργών αλγορίθμων ταξινόμησης. Για τις μεθόδους αυτές ισχύει η επόμενη θεώρηση.
- Δοθέντος ενός πίνακα A με *n* διακριτά στοιχεία λέγεται ότι υπάρχει μία **αντιστροφή** (*inversion*) αν ισχύει A[i] > A[j] για κάποια i < j.

  Με απλά λόγια, στην ακολουθία S = 3, 14, 1, 5, 9 οι αντιστροφές είναι (3, 1), (14, 1), (14, 5) και (14, 9).

Πρόταση.

Δεδομένου ενός πίνακα που περιέχει μία τυχαία διάταξη με n στοιχεία, <u>το άθροισμα των αποστάσεων</u> που διανύονται από τα στοιχεία κατά την ταξινόμησή τους είναι (n^2 - 1)/3.

Απόδειξη.

Αν υποθέσουμε ότι ο πίνακας *A* περιέχει μία τυχαία διάταξη των τιμών 1, 2, ... , *n*. Δοθέντος ενός τυχαίου στοιχείου *A[j],* η απόσταση που θα διανύσει είναι *A[j] j.* Επομένως θεωρώντας το σύνολο των στοιχείων ισχύει ότι η μέση τιμή της απόστασης ισούται με:

$$E[A[j]-j]=\frac{|1-j|+|2-j|+\ldots+|j-j|+|n-j|}{n}$$

$$=\frac{|1-j|+|2-j|+\ldots+|(j-1)-j|+|j-j|+|j+1-j|+|j+2-j|+\ldots+|n-j|}{n}$$

$$=\frac{|1-j|+|2-j|+\ldots+1+0+1+2+\ldots+|n-j|}{n}=\frac{1}{n}\left(\sum_{i=1}^{j-1}i+\sum_{i=1}^{n-j}i\right)=\frac{1}{n}\left(\frac{j(j-1)}{2}+\frac{(n-j)(n-j+1)}{2}\right)$$

Συνεπώς η μέση διανυόμενη απόσταση είναι:

$$\sum_{j=1}^{n}\left(E[A[j]-j]\right)=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{j(j-1)}{2}+\frac{(n-j)(n-j+1)}{2}\right)$$

Οι όροι εντός του αθροίσματος είναι ίδιοι για συμπληρωματικές τιμές του *j* και ισχύει

$$\sum_{j=1}^{n}\left(E[A[j]-j]\right)=\frac{1}{n}\sum_{j=1}^{n}j(j-1)=$$

$$\sum_{j=1}^{n}\left(E\left[A\left[j\right]-j\right]\right)=\frac{1}{n}\sum_{j=1}^{n}j\left(j-1\right)=\frac{1}{n}\sum_{j=1}^{n}(j^2-j)=\frac{1}{n}\left[\sum_{j=1}^{n}j^2-\sum_{j=1}^{n}j\right]$$

$$=\frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}-\frac{n(n+1)}{2}\right]=\frac{1}{n}\left[\frac{n(n+1)(2n+1)-3n(n+1)}{6}\right]=\frac{1}{n}\left[\frac{n(n+1)(2n+1-3)}{6}\right]$$

$$=\frac{(n+1)(2n-2)}{6}=\frac{(n+1)2(n-1)}{6}=\frac{(n+1)(n-1)}{3}=\frac{n^2-1}{3}$$

Επομένως ισχύει

$$\sum_{j=1}^{n}\left(E\left[A\left[j\right]-j\right]\right)=\frac{1}{n}\sum_{j=1}^{n}j\left(j-1\right)=\ldots=\frac{n^2-1}{3}$$

Συνεπώς ένας αλγόριθμος ταξινόμησης που μετακινεί τα στοιχεία κατά ένα σταθερό αριθμό θέσεων σε κάθε βήμα απαιτεί $O(n^2)$ βήματα.

Ερώτημα "Πόσο γρήγορα μπορεί μία μέθοδος να ταξινομήσει έναν πίνακα;"

Θα αποδειχθεί ότι όταν μία μέθοδος ταξινόμησης στηρίζεται μόνο σε συγκρίσεις και ανταλλαγές κλειδιών, τότε στη χειρότερη περίπτωση δεν μπορεί να ταξινομήσει σε χρόνο κατώτερο του $\Omega(n\log_2 n)$

Η απόδειξη της σχετικής πρότασης θα περιγραφεί με τη διαδικασία της ταξινόμησης με **δένδρο αποφάσεων** (*decision tree*).

Ένα μονοπάτι του δένδρου δείχνει μία πιθανή διαδοχή από υπολογισμούς που ο αλγόριθμος πιθανώς να ακολουθήσει.

Για παράδειγμα, ας υποτεθεί ότι πρέπει να ταξινομηθούν τρία κλειδιά: τα k_1, k_2 και k_3 αντίστοιχα.

Η σειρά εισόδου των κλειδιών είναι k_1, k_2 και k_3 που φαίνεται στο δένδρο αποφάσεων ως (1, 2, 3).

Αρχικά συγκρίνονται τα κλειδιά k_1 και k_2.

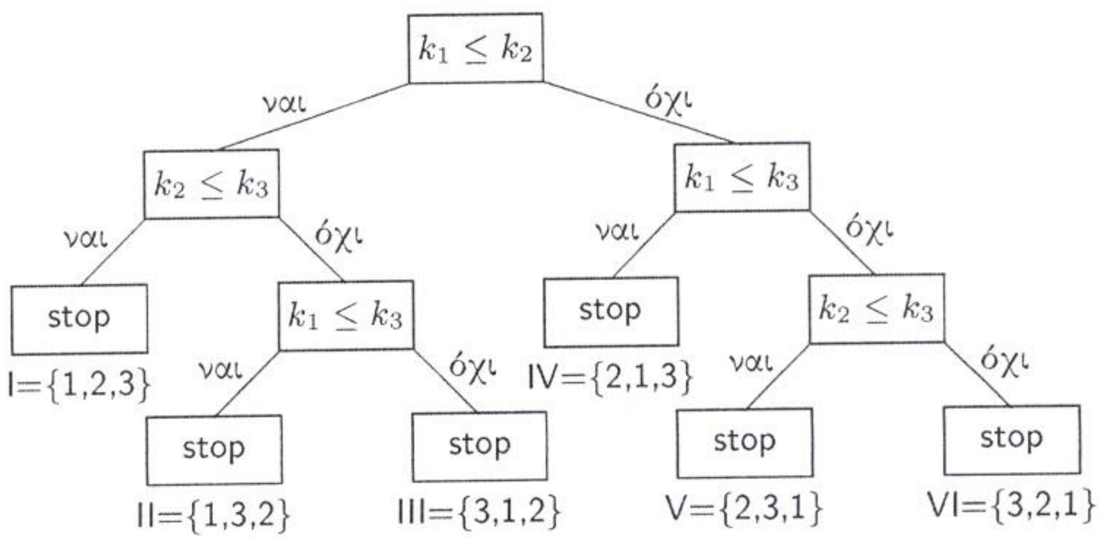
Αν $k_1 < k_2$ τότε η σειρά (1, 2, 3) δεν αλλάζει, αλλιώς γίνεται (2, 1, 3).

Στο επόμενο σχήμα φαίνονται όλα τα πιθανά ενδεχόμενα που μπορούν να συμβούν σε έναν αλγόριθμο ταξινόμησης ανάλογα με τη σειρά σύγκρισης των κλειδιών στο δένδρο.

Από το σχήμα αυτό φαίνεται ότι το δένδρο αποφάσεων είναι ένα δυαδικό δένδρο.

Τα φύλλα του δένδρου είναι αριθμημένα από I ως VI και αποτελούν τα μοναδικά σημεία τερματισμού κάθε αλγορίθμου ταξινόμησης.

Το δένδρο αυτό έχει 3!=6 φύλλα που εξασφαλίζει ότι ο αλγόριθμος βρίσκει πάντοτε τη διάταξη εκείνη που αντιστοιχεί στην ταξινομημένη σειρά.



Πρόταση

Κάθε δένδρο αποφάσεων που ταξινομεί n διακεκριμένα κλειδιά έχει ύψος τουλάχιστον $\log_2 n!$.

Απόδειξη

Ένα δένδρο αποφάσεων έχει *n!* φύλλα που αντιστοιχούν σε κάθε μία από τις *n!* διατάξεις των *n* κλειδιών.

Κάθε δυαδικό δένδρο ύψους *h* έχει το πολύ 2^h φύλλα

Οπότε συνεπάγεται ότι $n! \leq 2^h$.

Λογαριθμώντας και τα δύο μέλη της σχέσης (η συνάρτηση $\log_2 n$ είναι μονότονα αύξουσα) προκύπτει:

$$\log_2 n! \leq \log_2 2^h = h$$

Πρόταση *(Θεώρημα 8.1 CLRS)*

Κάθε αλγόριθμος που ταξινομεί κάνοντας μόνο συγκρίσεις έχει στη χειρότερη περίπτωση χρόνο τάξης $\Omega(n\log_2 n)$.

Απόδειξη

Η εκτέλεση ενός αλγορίθμου διάταξης ισοδυναμεί με την ανίχνευση ενός μονοπατιού από τη ρίζα προς ένα φύλλο ενός δένδρου απόφασης.

Αρα ο χειρότερος χρόνος ενός αλγόριθμου αντιστοιχεί στο μέγιστο μήκος μονοπατιού, δηλαδή στο ύψος του ισοδύναμου δένδρου απόφασης

Ετσι αν προσδιοριστεί ένα κάτω όριο για τα ύψη των δένδρων αποφάσεων έχει βρεθεί ταυτόχρονα και ένα κάτω φράγμα για τους συγκριτικούς αλγόριθμους ταξινόμησης.

Από την προηγούμενη πρόταση προκύπτει ότι κάθε μονοπάτι είναι μήκους $\log_2 n!$

Απόδειξη

$$
\begin{aligned}
2^h \geq n! \Rightarrow h &\geq \log(n!) \\
&= \log(n(n-1)(n-2)\cdots(2)) \\
&= \log n + \log(n-1) + \log(n-2) + \cdots + \log 2 \\
&= \sum_{i=2}^{n} \log i \\
&= \sum_{i=2}^{n/2-1} \log i + \sum_{i=n/2}^{n} \log i \\
&\geq 0 + \sum_{i=n/2}^{n} \log \frac{n}{2} \\
&= \frac{n}{2} \cdot \log \frac{n}{2} = \frac{n}{2}(\log n - \log 2) = \frac{n}{2}\log n - \frac{\log 2}{2} n \\
&= \Omega(n \log n)
\end{aligned}
$$

(*Απόδειξη με κανόνα ορίου*)

$\forall i, 2 \leq i \leq \frac{n}{2} \quad \log i \succ 0 \quad (1)$

Αντικαθιστώ όπου log *i με 0*

$\forall i, \frac{n}{2} \leq i \leq n \quad \log i \geq \log \frac{n}{2} (2)$

Αντικαθιστώ όπου log*i με* $\log \frac{n}{2}$

Άρα $\Omega(n\log_2 n)$ είναι ένα κάτω φράγμα για τη χειρότερη περίπτωση για συγκριτικές ταξινομήσεις.

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ