

Exercise 1

Parallel & Distributed Computer Systems

Oct 24, 2024

V0

Να γραφεί υπορουτίνα σε C/C++ το οποίο να υπολογίζει τους k κοντινότερους γείτονες ενός συνόλου Q σημείων/ερωτημάτων (*query-set*) ως προς ένα σύνολο C δεδομένων σημείων (*corpus-set*), d διαστάσεων. Χρησιμοποιείστε το `[idx, dst] = knnsearch(C, Q, 'k', k)` της MATLAB ως το πρότυπο. Η υπορουτίνα αυτή υπολογίζει την ακριβή (όχι προσεγγιστική) απάντηση, με σειριακό τρόπο, κάνοντας χρήση της βιβλιοθήκης OpenBLAS για τον υπολογισμό του πίνακα των αποστάσεων

$$D = \sqrt{C^2 - 2CQ^T + Q^{2T}}.$$

Σημείωση, τα τετράγωνα και η τετραγωνική ρίζα υπολογίζονται *element-wise*.

Οι πιο κοντινοί γείτονες υπολογίζονται με `quick-select`.

Προφανώς η παραπάνω συνάρτηση δε μπορεί να χειριστεί μεγάλο αριθμό σημείων καθώς ο πίνακας αποστάσεων δε θα χωρά στη μνήμη ολόκληρος. Αλλά μπορούμε να τον υπολογίζουμε τμηματικά, ένα μπλοκ τη φορά.

Τα μπλοκ αποστάσεων μπορούν να υπολογιστούν και ταυτόχρονα.

Για την άσκηση αυτή θα υπολογίσουμε τους γείτονες για ένα πρόβλημα όπου $C == Q$, και η λύση μας θα είναι προσεγγιστική γιατί δεν θα υπολογίσουμε την απόσταση όλων με όλους.

Η ιδέα είναι ότι θα υπολογίσουμε τους γείτονες για δύο ή περισσότερα διακριτά υποσύνολα (αναδρομικά) και μετά θα “ενώνουμε” τις επιμέρους λύσεις με κάποιο αποδοτικό τρόπο.

V1

Υλοποιείστε την αναδρομή (και ένωση ίσως) παράλληλα κάνοντας χρήση του OpenMP, OpenCilk και PTHREADS.

Θα κάνουμε δοκιμές με δεδομένα που θα ανακοινωθούν. Ένας εύκολος τρόπος να δοκιμάσετε την ορθότητα του κώδικα σας είναι να χρησιμοποιήσετε την `knnsearch` της MATLAB για να υπολογίσετε τους γείτονες και να συγκρίνετε τα αποτελέσματα. Θα βοηθήσει να χρησιμοποιήσετε έτοιμη βιβλιοθήκη για ανάγνωση και εγγραφή `.mat` αρχείων από την C/C++ ώστε η ανάλυση κι η οπτικοποίηση των αποτελεσμάτων να γίνει στο περιβάλλον MATLAB.

Όλες οι αναφορές σε MATLAB, μπορούν να αντικατασταθούν από Octave, Python, Julia ή άλλη γλώσσα προγραμματισμού που σας είναι εύκολο να επιβεβαιώσετε ορθότητα σε ότι κάνετε.

Παραδοτέα

- Αναφορά 4 – 5 σελίδων που να περιγράφει τη μέθοδο, ειδικά του παραλληλισμού καθώς και τους ελέγχους ορθότητας που χρησιμοποιήσατε.
- Σχόλια για την ταχύτητα υπολογισμών, την επιτάχυνση που πετύχατε και τον αριθμό των νημάτων που χρησιμοποιήσατε για συγκεκριμένα σενολα δεδομένων που θα ανακοινωθούν. Σύντομη περιγραφή του υπολογιστικού συστήματος που χρησιμοποιήσατε για τα πειράματα.
- Ο κώδικας να μπει στο νέφος με Makefile και οδηγίες ώστε να μπορεί να δοκιμαστεί από τους διορθωτές/σχολιαστές.