

Άσκηση 2

Παράλληλα & Κατανεμημένα Συστήματα Υπολογιστών

21 Νοεμβρίου 2018

Να υλοποιήσετε σε MPI¹ κατανεμημένο αλγόριθμο κατασκευής της δενδρικής δομής vantage-point για ένα σύνολο δεδομένων \mathcal{X} . Η δομή χρησιμοποιείται στην εύρεση των k κοντινότερων γειτόνων (k nearest neighbor search – k NN) των σημείων του vantage-point tree [1].

Ακολουθήστε τα παρακάτω βήματα στο πρόγραμμά σας:

1. Μετατρέψτε τον κώδικα² ώστε να δουλεύει i) με αποστάσεις σε float και ii) με χωριστά υποσύνολα διαδικασιών που είναι ίσα σε μέγεθος και πάντα δύναμη του 2 (δείτε παρακάτω πως χρησιμοποιείται).
2. Δημιουργήστε p διεργασίες (*processes*) με την εκκίνηση του προγράμματος:
 - Κάθε διεργασία έχει ένα N/p σημεία από ένα σετ δεδομένων $\mathcal{X} \in \mathbb{R}^{N \times D}$.
 - επαναλάβετε τα παρακάτω l φορές.
 - Ο αριθμός της επανάληψης προσδιορίζει σε ποια υποομάδα κάθε διαδικασία ανήκει, κι ποια διεργασία είναι ο “αρχηγός” για την τρέχουσα επανάληψη. Δηλαδή, στη πρώτη επανάληψη, αρχηγός είναι η 0 για την ομάδα $0 : p - 1$, στη δεύτερη επανάληψη, η 0 είναι ο αρχηγός της ομάδας $0 : p/2 - 1$ και η $p/2$ της ομάδας $p/2 : p - 1$, στην τρίτη επανάληψη, αρχηγοί είναι οι 0, $p/4$, $p/2$ και $3p/4$ στις αντίστοιχες υποομάδες, κ.ο.κ.
 - Ο αρχηγός ανακοινώνει το vantage point που διαλέγει τυχαία από τα σημεία του σε όλους της ομάδας του. Η κάθε διεργασία μετράει την απόσταση των σημείων της από το vantage point.
 - Υπολογίζεται η διάμεση απόσταση για τα στοιχεία κάθε υποομάδας με χρήση του κώδικα (1).
 - Κάθε διεργασία χωρίζει και μετρά τα σημεία με αποστάσεις που είναι μικρότερες ή ίσες και μεγαλύτερες από την διάμεση απόσταση και ανακοινώνουν το πλήθος τους στον αρχηγό της υποομάδας.
 - Στην συνέχεια τα στοιχεία αναδιανέμονται στην υποομάδα με την βοήθεια του αρχηγού, ώστε οι πρώτες μισές διεργασίες να έχουν τα στοιχεία με απόσταση μικρότερη ή ίση της διαμέσου από το vantage point και οι υπόλοιπες με απόσταση μεγαλύτερη της διαμέσου, αντίστοιχα.
3. Υπολογίστε όλο το vantage-point tree. Μετά από l επαναλήψεις, οι διαδικασίες δουλεύουν μόνο με τοπικά δεδομένα χωρίς να χρειάζονται επικοινωνίες. Ποιο είναι το l ;
4. Στο τέλος, όλες οι διεργασίες πληροφορούνται τα vantage point από ολόκληρο το κοινό δέντρο και φυσικά γνωρίζουν ποιος έχει ποιο υποδένδρο.
5. Υπολογίστε το all- k -NN.
 - Υπολογίστε τους k κοντινότερους γείτονες όλων των σημείων, με $k = 2^{[1:8]}$.
 - Όσα σημεία έχουν τον k γείτονα πιο μακρινό από την απόσταση που χωρίζει το υποδένδρο, ομαδοποιούνται και στέλνονται στον ιδιοκτήτη του υποδένδρου. Στη συνέχεια, η διεργασία αναμένει τις απαντήσεις.

Υλοποιείτε τον αλγόριθμο vantage-point όπως περιγράφεται παραπάνω και όχι κάτι άλλο που τυχόν βρήκατε στο διαδίκτυο ³.

Παραδώστε:

¹<https://computing.llnl.gov/tutorials/mpi/>

²<https://github.com/akintsakis/quickselectMpi>

³<http://stevehanov.ca/blog/index.php?id=130>

- Αναφορά έως και 3 σελίδων που να περιγράφει τη μέθοδο του παραλληλισμού καθώς και τους ελέγχους ορθότητας που χρησιμοποιήσατε.
- Σχόλια/διαγράμματα για την ταχύτητα της κατασκευής και εύρεσης γειτόνων συγκριτικά με τη σειριακή υλοποίηση που σας δίνεται. Τα συμπεράσματά σας θα πρέπει να στηρίζονται σε μετρήσεις στο Hellasgrid. Δεν συμπεριλαμβάνουμε στον χρόνο σύγκρισης το χρόνο ανάγνωσης του αρχικού συνόλου \mathcal{X} και το χρόνο επιβεβαίωσης των αποτελεσμάτων. Χρησιμοποιείστε δύο `MPI_Barrier` για να μετρήσετε σωστά το χρόνο στο `PID == 0`.

Μόνο το pdf στο elearning!

Δεοντολογία: Εάν χρησιμοποιήσετε κώδικες από το διαδίκτυο ή αλλού, να αναφέρετε την πηγή και τις αλλαγές που κάνατε.

Σημείωση: Ομαδικές εργασίες γίνονται δεκτές. Ο μέγιστος αριθμός φοιτητών που μπορούν να συνεργαστούν σε μία ομάδα είναι δύο. Η ίδια εργασία με τα δύο ονόματα πρέπει να κατατεθεί. Αντιγραφές ή ευρύτερες συνεργασίες ακυρώνουν την εργασία όλων των συνεργαζόμενων.

Ημερομηνία παράδοσης: 12 Δεκεμβρίου 2018.

Αναφορές

- [1] Peter N Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. In *Fourth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics Philadelphia*, volume 93, pages 311–321, 1993.