2017

Συσταδοποίηση μοριακών διαμορφώσεων

Ανάπτυξη Λογισμικού για Αλγοριθμικά Προβλήματα
Μέρος 3ο

Κανελλόπουλος Στέφανος

(Α.Μ. 1115201200050)

Χανιωτάκης-Ψύχος Χαρίδημος

(Α.Μ. 1115201200194)

Περιγραφή προγράμματος

Αρχικά, το πρόγραμμα μας διαβάζει τα δεδομένα από το input αρχείο και τα αποθηκεύει σε έναν πίνακα μεγέθους όσο το πλήθος των conformations. Στην συνέχεια κατασκευάζουμε έναν πίνακα αποστάσεων, ο οποίος περιλαμβάνει τις αποστάσεις για κάθε conformation από όλα τα υπόλοιπα conformations. Με αυτό τον τρόπο υπολογίζουμε μόνο μια φορά τις αποστάσεις μειώνοντας τον χρόνο εκτέλεσης του προγράμματος.

Το επόμενο βήμα είναι η εκτέλεση της διαδικασίας του clustering τύπου k-medoids με τους πιο αποτελεσματικούς αλγορίθμους, οι οποίοι είχαν υλοποιηθεί στο προηγούμενο κομμάτι του project. Δοκιμάζουμε clustering με διαφορετικό αριθμό clusters, ξεκινώντας από το ¼ του συνόλου των conformations και σε κάθε βήμα κάνουμε clustering για το μισό πλήθος clusters από το προηγούμενο. Αυτό συνεχίζεται για clustering με αριθμό clusters μεγαλύτερο του δυο.

Η διαδικασία του clustering τερματίζεται υπό την συνθήκη ότι τα κέντρα είτε δεν μετατοπίστηκαν καθόλου είτε μετατοπίστηκαν ελάχιστα (συμβολική σταθερά THRESHOLD). Αφού τερματιστεί η διαδικασία, υπολογίζεται το silhouette για το σύνολο των clusters και το clustering με την καλύτερη τιμή silhouette αποθηκεύεται σε ένα αρχείο με όνομα ‘crsmd.dat’ ή ‘frechet.dat’.

Αρχεία κώδικα και σύντομη περιγραφή τους

- makefile: Μεταγλώττιση προγράμματος και παραγωγή εκτελέσιμου αρχείου
- main.c: Δέχεται τα ορίσματα του χρήστη, δημιουργεί τον πίνακα με τα δεδομένα και καλεί τις συναρτήσεις για clustering για διαφορετικό αριθμό clusters κάθε φορά
- structs.c: Περιλαμβάνει όλες τις δομές που ήταν απαραίτητες για την υλοποίηση του προγράμματος
- functions.c: Περιλαμβάνει βοηθητικες συναρτήσεις που είναι απαραίτητες για την εκτελεση του clustering
- functions.h: Περιλαμβάνει όλες τις standard libraries, όπως και τις συμβολικές σταθερές
- cRMSD.c: Περιλαμβάνει την υλοποιηση της συναρτησης c-RMSD με την χρήση των βιβλιοθηκών lapacke και cblas
- cRMSD.h: Αρχείο επικεφαλίδας του cRMSD.c
- metric_functions.c: Περιλαμβάνει τις συναρτήσεις για τον υπολογισμό της απόστασης frechet μεταξύ δύο conformations.
- metric_functions.h: Αρχείο επικεφαλίδας του metric_functions.c
- input_functions.c: Περιλαμβανει τις συναρτησεις για το διαβασμα των δεδομενων από το input file
- input_functions.h: Αρχείο επικεφαλίδας του input_functions.c
- output_functions.c: Περιλαμβάνει την συναρτήση για την εξαγωγή των αποτελεσμάτων στο output file
- output_functions.h: Αρχείο επικεφαλίδας του output_functions.c
- quicksort.c: Περιλαμβάνει τις συναρτήσεις για την υλοποίηση της quicksort για ταξινόμηση μονοδιάστατου πίνακα από integers
- quicksort.h: Αρχείο επικεφαλίδας του quicksort.c
- clustering_init.c: Περιλαμβάνει την υλοποίηση της συνάρτησης k-means++
- clustering_init.h: Αρχείο επικεφαλίδας του clustering_init.c
- clustering_assignment.c: Περιλαμβάνει την υλοποίηση της συνάρτησης lloyd
- clustering_assignment.h: Αρχείο επικεφαλίδας του clustering_assignment.c
- clustering_update.c: Περιλαμβάνει την υλοποίηση της συνάρτησης PAM
- clustering_update.h: Αρχείο επικεφαλίδας του clustering_update.c

- silhouette.c: Περιλαμβάνει την υλοποίηση της μετρικής αποδοτικότητας του clustering
- silhouette.h: Αρχείο επικεφαλίδας του silhouette.c

Μεταγλώττιση και χρήση του προγράμματος

Για την μεταγλώττιση του προγράμματος περιλαμβάνεται αρχείο makefile και πραγματοποιείται με την εντολή "./make" στον τρέχοντα κατάλογο. Με την εντολή "./make clean" γίνεται ο καθαρισμός των αρχείων αντικειμένων.

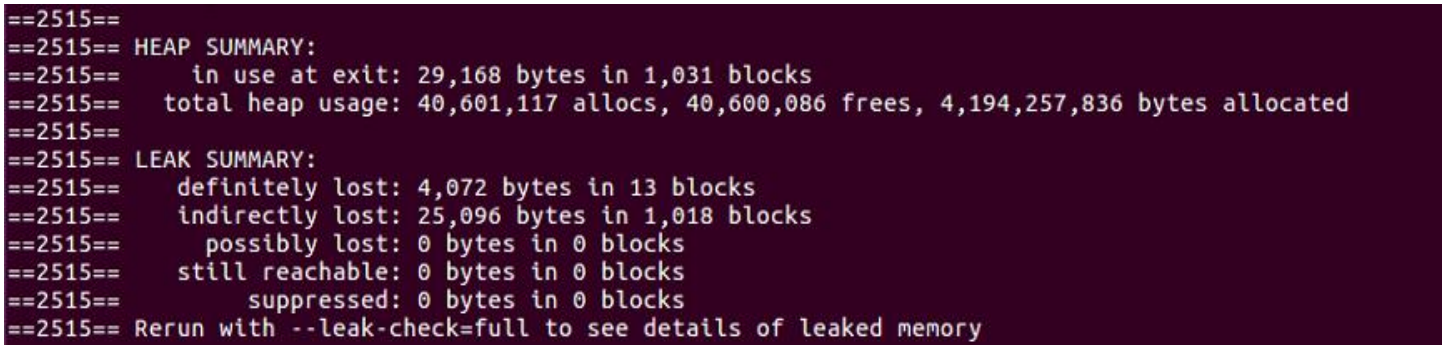
Το πρόγραμμα εκτελείται με την παρακάτω εντολή:

```
./cluster –i <input file> –frechet [optional]
```

Σε περίπτωση που δεν δοθεί το input file στο command line, ο χρήστης έχει την δυνατότητα σε runtime να το εισάγει.

Valgrind – Memory leaks

Έγινε έλεγχος με valgrind για memory leaks και τα αποτελέσματα φαίνονται παρακάτω.



Παρατηρούμε ότι το συγκεντρωτικό memory leak του προγράμματος ανέρχεται στα 29.168 bytes από τα συνολικά 4.194.257.836 bytes, που αντιστοιχούν στο 0,000007% της συνολικής μνήμης που χρησιμοποιήθηκε.