

Ανάπτυξη λογισμικού για αλγοριθμικά προβλήματα
Χειμερινό εξάμηνο 2016-2017
3 Προγραμματιστική εργασία
Υλοποίηση αλγορίθμων υπόδειξης (Recommendation) - Συσταδοποίηση μοριακών διαμορφώσεων

Σίγας Χρήστος A.M.1115201200159
Σταμάτη Βασιλική A.M.1115201200172

Η εργασία υλοποιήθηκε σε σύστημα LINUX.

Κατάλογος των αρχείων:

- 1.main.c
- 2.nnlsh.c Recommendation lsh και οι συναρτήσεις που το υλοποιούν
- 3.nnlsh.h
- 4.clrec.c Recommendation clustering και οι συναρτήσεις που το υλοποιούν
- 5.clrec.h
- 6.rmsd.c Υλοποίηση crmsd καθώς και της συνάρτησης clara για τη συσταδοποίηση
- 7.rmsd.h
- 8.drmsd.c Υλοποίηση drmsd καθώς και της συνάρτησης k-means για τη συσταδοποίηση
- 9.LAPACKE Βιβλιοθήκες και header file
- 10.ενδεικτικά output για κάθε μέθοδο
- 11.makefile

Η υλοποίηση πληρεί όλες τις απαιτήσεις της εκφώνησης.

Έχει γίνει έλεγχος με unit testing.

Έγινε έλεγχος με valgrind για memory leaks.

Οδηγίες μεταγλώττισης:

Για να μεταγλωττιστεί το πρόγραμμα αρκεί η εντολή make και οι εντολές εκτέλεσης είναι αυτές της εκφώνησης.

Για να εκτελεστεί η συσταδοποίηση μοριακών διαμορφώσεων αρκεί η εντολή:

```
./recommendation bio_small_input.dat
```

Για να εκτελεστεί το recommendation αρκεί η εντολή:

```
./recommendation -d yahoo_music_small.dat -o output.dat
```

```
./recommendation -d yahoo_music_small.dat -o output.dat -validate
```

//υπολογισμός MAE

Υλοποίηση αλγορίθμων υπόδειξης (Recommendation)

Παρατηρούμε ότι αν το k είναι ίσο με N/P το πρόγραμμα αργεί πολύ και δεν βελτιώνεται η silhouette.

Άρα καταλήγουμε στο συμπέρασμα ότι το k πρέπει να είναι μικρότερο του N/P .

Ενδεικτικά αποτελέσματα για τον αριθμό σφάλματος:

Cosine MAE~1.8

Hamming MAE~1.7

Euclidean MAE~0.8

Συσταδοποίηση μοριακών διαμορφώσεων

Παρατηρούμε ότι για $r=N(N-1)/2$ που το T απενεργοποιείται δεν έχουμε καλή silhouette.

Για r μικρά παρατηρούμε ότι για τις περιπτώσεις που παίρνουμε τυχαίες αποστάσεις η silhouette

είναι λίγο μικρότερη από τις περιπτώσεις που παίρνουμε τις μεγαλύτερες και τις μικρότερες αποστάσεις.

Επομένως καταλήγουμε στο συμπέρασμα ότι r μικρότερες και r μεγαλύτερες αποστάσεις έχουν καλύτερα

αποτελέσματα. Τα συμπεράσματα αυτά προέκυψαν μετά από πολλές δοκιμές.