Ανάπτυξη Λογισμικού για Αλγοριθμικά Προβλήματα

2η Προγραμματιστική Εργασία

Υλοποίηση του αλγορίθμου συσταδοποίησης K-medoids στη γλώσσα C/C++

Σταματέλου Ισμήνη- Χριστίνα 1115201200171  
Μπόκος Δημήτρης 1115201300117

Στην εργασία αυτή ζητείται να υλοποιήσουμε τον αλγόριθμο K-medoids για τη συσταδοποίηση δεδομένων που ανήκουν σε σύνολο το οποίο αποτελεί μετρικό χώρο, χρησιμοποιώντας τους 8 συνδυασμούς από τις εξής παραλλαγές:

**Initialization**

1. K-medoids++

2. Concentrate (Park-Jun)

**Assignment**

1. PAM assignment (simplest approach)

2. Assignment by LSH/DBH: Reverse Approach

**Update**

1. Update a la Lloyd’s (improved PAM)

2. CLARANS Update.

Ακόμα ζητείται η υλοποίηση των αλγορίθμων CLARA και Silhouette. Ο τελευταίος χρησιμοποιείται για να εξεταστεί η κάθε συσταδοποίηση.

Στην εργασία έχουν υλοποιηθεί όλα τα ζητούμενα.

Επίσης, χρησιμοποιείται η κατάλληλη βιβλιοθήκη και γίνεται εκτέλεση ορισμένων συναρτήσεων μέσω unit testing.

Διάρθρωση των αρχείων:

Ευκλείδεια μετρική:

structs.h: Σε αυτό το αρχείο επικεφαλίδας βρίσκονται οι δηλώσεις όλων τον τύπων που χρησιμοποιούμε καθώς και τα πρωτότυπα όλων των συναρτήσεων. Η δομή που κρατιούνται τα στοιχεία όταν τα εισάγουμε από το αρχείο είναι μία λίστα κάθε κόμβος της οποίας αποτελείται από:  
το όνομα του στοιχείου, έναν μονοδιάστατο πίνακα μεγέθους d με τις συντεταγμένες του στοιχείου και ένας αριθμός που δηλώνει αν έχουμε επισκεφτεί τον κόμβο ή όχι.

structs.c: Περιέχονται οι υλοποιήσεις όλων των συναρτήσεων που χρησιμοποιούνται στην ευκλείδεια μετρική καθώς και όλες οι δομές.

Μετρική cosine

struct\_cos.h: Όπως και στην ευκλείδεια έτσι και στη μετρική cosine για κάθε στοιχείο αποθηκεύεται το όνομά του, ένας πίνακας με τις συντεταγμένες του και ο αριθμός που δείχνει αν το στοιχείο έχει επισκεφθεί ή όχι. Στο αρχείο βρίσκονται οι δηλώσεις των δομών και των συναρτήσεων

struct\_cos.c Υλοποίηση συναρτήσεων και οι δομές.  
  
Μετρική Hamming  
  
struct\_ham.h: Εδώ, σε κάθε κόμβο της λίστα περιέχεται το όνομα του στοιχείου, ο αριθμός για την επίσκεψη του κόμβου και ένας unsigned long long int που αναπαριστά την δυαδική συμβολοσειρά στο δεκαδικό σύστημα. Αυτό γίνεται για ευκολία των υπολογισμών καθώς και εξοικονόμηση μνήμης. Το αρχείο περιέχει τις δηλώσεις των δομών και των συναρτήσεων.  
  
struct\_ham.c: Υλοποίηση συναρτήσεων και δομών.

Distance Matrix:  
  
struct\_ma.h: Σε κάθε κόμβο, φυλάσσεται το όνομα του στοιχείου, η θέση του στον πίνακα αποστάσεων καθώς και ένας πίνακας με την απόστασή του από τα άλλα σημεία. Δηλώσεις δομών και συναρτήσεων.  
  
struct\_ma.c: Υλοποιήσεις των παραπάνω.

Main.c  
Στη main διαβάζεται η πρώτη και αν χρειαστεί και η δεύτερη σειρά του αρχείου και επιλέγεται η μετρική που αναγράφεται στο αρχείο. Ανάλογα τη μετρική, καλείται μία συνάρτηση που είναι υπεύθυνη για την κλήση των επιμέρους συναρτήσεων που χρειάζονται. Έτσι έχουμε επιτύχει να μικρύνουμε αισθητά τη main και να κάνει τα τελείως απαραίτητα, δίνοντας το ρόλο σε άλλες συναρτήσεις.  
  
functions.h και functions.c  
Εδώ υπάρχουν όλες οι συναρτήσεις που δεν ανήκουν σε κάποια μετρική ή χρησιμοποιούνται από δύο μετρικές όπως για παράδειγμα η υλοποίηση της gaussian κατανομής που χρησιμοποιείται από την μετρική cosine και την ευκλείδεια μετρική.  
  
  
Σε όλους τις μετρικές χρησιμοποιείται το ίδιο μοτίβο για εισαγωγή και αναζήτηση στοιχείων στο και από το hash table και όλες οι δομές είναι ίδιες προσαρμοσμένες προφανώς στις ανάγκες κάθε μεθόδου. Τα βήματα επομένως που γίνονται είναι πανομοιότυπα κάτι που φαίνεται και από την ομοιομορφία της main.c

Μεταγλώττιση:  
Έχουμε υλοποιήσει makefile, έτσι λοιπόν ο χρήστης για να μεταγλωττίσει το πρόγραμμα αρκεί να πληκτρολογήσει την εντολή "make".

Αν θέλει να τρέξει το test πρέπει να πατήσει make test και στη συνέχεια την εντολή εκτέλεσης ./utest  
  
Εκτέλεση:  
Ο χρήστης πρέπει υποχρεωτικά στην εντολή εκτέλεσης να πληκτρολογήσει τις σημαίες και τα ονόματα των αρχείων εισόδου και του αρχείου εξόδου. Στην περίπτωση που δεν τα δώσει, το πρόγραμμα τερματίζει με εμφάνιση μηνύματος λάθους.

Ενδεικτική εντολή: ./medoids -d DataEuclidean\_5\_500x100.csv -c configuration.csv -o results.csv