Κουγιανός Νικόλαος 1115201100060

Σιούστη Ηλιάνα-Χάρις 1115201100107

1/12/2016

Ανάπτυξη Λογισμικού για

Αλγοριθμικά Προβλήματα

2η Προγραμματιστική Εργασία

Υλοποίηση του αλγορίθμου συσταδοποίησης K-medoids στη γλώσσα C

Αρχεία Κώδικα:

**Functions.h:** Περιέχει τις δηλώσεις των συναρτήσεων του αρχείου Functions.c

**Functions.c:** Περιέχει όλες τις συναρτήσεις που χρησιμοποιούνται στο αρχείο k\_medoidsFINAL.c

**k\_medoidsFINAL.c:** Περιέχει τη main συνάρτηση στην οποία γίνεται το διάβασμα του αρχείου εισόδου, πραγματοποιούνται όλοι οι δυνατοί συνδυασμοί initialization-assignment-update και στο τέλος η αποδέσμευση της μνήμης και το κλείσιμο των ανοιχτών αρχείων.

Έχει υλοποιηθεί ο αλγόριθμος K-medoids για τη συσταδοποίηση δεδομένων σε σύνολο που αποτελεί μετρικό χώρο. Συγκεκριμένα, υλοποιήθηκαν οι συναρτήσεις k-medoids++ και concentrate όσον αφορά στο initialization μέρος της εργασίας, η συνάρτηση PAM assignment όσον αφορά στο assignment μέρος και οι συναρτήσεις update a la Lloyd’s, CLARANS και CLARA όσον αφορά στο update μέρος της εργασίας. Επιπλέον, έχει γίνει υλοποίηση της συνάρτησης Sihlouette προκειμένου να αξιοποιηθεί το clustering. Η συνάρτηση assignment by LSH/DBH δεν υλοποιήθηκε λόγω έλειψης χρόνου. Αναλόγως τη μετρική που δίνεται στο αρχείο εισόδου, υπολογίζονται οι αποστάσεις Euclidean, Cosine και Matrix, ενώ έχει υλοποιηθεί και η απόσταση Manhattan η οποία διδάχθηκε σε διάλεξη. Ο σχεδιασμός του κώδικα επιτρέπει την εύκολη επέκτασή του προγράμματος σε διαφορετικούς χώρους.

Αρχικά, δίνονται σαν παράμετροι στη γραμμή εντολών:

* το input file από το οποίο λαμβάνονται τα αντικείμενα, το πλήθος, οι διαστάσεις τους, καθώς και η μετρική που χρησιμοποιείται την εκάστοτε φορά
* το configuration file από το οποίο λαμβάνονται δεδομένα όπως ο αριθμός Κ των clusters, ο αριθμός των επαναλήψεων στην περίπτωση της clarans κτλ. Αν θέλουμε να δώσουμε τις default τιμές, βάζουμε στο αντίστοιχο πεδίο το 0.
* το output file στο οποίο θα εκτυπωθούν τα τελικά αποτελέσματα

Στη συνέχεια, γίνονται όλοι οι πιθανοί συνδυασμοί initialization- assignment-update των συναρτήσεων που έχουν υλοποιηθεί και εκτυπώνονται τα αποτελέσματα αυτών στο αρχείο εξόδου που έχει δοθεί ως παράμετρος στη γραμμή εντολών.

Σε όλη την έκταση του προγράμματος γίνονται έλεγχοι λάθους στη δέσμευση μνήμης και προβλήματος στο άνοιγμα των αρχείων, ενώ υπάρχει και σχετική ενημέρωση του χρήστη που ακολουθείται από τερματισμό του προγράμματος.

Τέλος, γίνεται αποδέσμευση της μνήμης και κλείσιμο των ανοιχτών αρχείων.

Το πρόγραμμα είναι πλήρως παραμετρικό.

Οδηγίες μεταγλώττισης:

/\* Makefile \*/

OBJS = k\_medoidsFINAL.o Functions.o

SOURCE = k\_medoidsFINAL.c Functions.c

HEADER = Functions.h

OUT = medoids

CC = gcc

CFLAGS = -g -c

all: $(OBJS)

$(CC) -g $(OBJS) -o $(OUT) -lm

k\_medoidsFINAL.o: k\_medoidsFINAL.c

$(CC) $(CFLAGS) k\_medoidsFINAL.c

Functions.o: Functions.c

$(CC) $(CFLAGS) Functions.c

clean:

rm -f $(OBJS) $(OUT)

count:

wc $(SOURCE) $(HEADER)

Το πρόγραμμα μεταγλωττίζεται με την εντολή make –f Makefile.

Ενδεικτικά, το πρόγραμμα θα μπορούσε να εκτελεστεί με κάποιον από τους παρακάτω τρόπους, καθώς τα ορίσματα μπορούν να δωθούν με οποιαδήποτε σειρά στη γραμμή εντολών:

./medoids –d DataHamming.csv –c config.txt -o output.txt

./medoids –c config.txt –d DataEuclidean.csv -o output.txt

./medoids -o output.txt –c config.txt –d DistanceMatrix.csv

./medoids –d DataHamming.csv –c config.txt -o output.txt -complete