

# PROGETTO MAP - CLUSTERING THRESHOLD ALGORITHM

## NOMINATIVI E MATRICOLE COMPONENTI GRUPPO

N

### DESCRIZIONE

Il seguente programma permette di interagire con l'algoritmo che calcola i cluster in maniera tale da definire la distanza massima o la somiglianza minima tra due o più punti affinché vengano considerati parte dello stesso cluster.

Questi sono i passaggi impiegati dall'algoritmo QT:

1. **Inizializzazione dei cluster candidati**

Per ogni esempio  $x$  nel dataset si costruisce un cluster candidato  $C_x$  che contiene  $x$  e tutti gli altri esempi  $y$  tali che  $d(x, y) \leq R$ .

2. **Selezione del cluster migliore**

Tra tutti i cluster candidati, si sceglie quello con il **maggior numero di elementi** (criterio di qualità).

3. **Rimozione degli elementi coperti**

Si eliminano dal dataset tutti gli esempi già assegnati al cluster trovato.

4. **Ripetizione**

Si continua con i dati rimanenti fino a quando non ci sono più esempi che possano formare un cluster.

5. **Output**

L'algoritmo restituirà l'insieme dei cluster trovati.

Un valore alto può portare a pochi cluster grandi, mentre un valore basso genera molti cluster piccoli e precisi.

### MODALITA' DI AVVIO

Il programma può essere avviato tramite il file **start.bat**, che aprirà un terminale in cui verrà richiesta la password di MySQL. Dopo aver inserito la password, il file, qualora non siano già presenti, provvederà a creare il database e l'utente necessari per il progetto. All'interno del database verrà inoltre creata una tabella denominata Playtennis, che verrà popolata con i dati necessari all'esecuzione dell'algoritmo.

Successivamente, verranno aperti i terminali dedicati al server e al client. Di seguito è riportato un esempio del terminale relativo al server, che resterà sempre aperto:

```
Server in ascolto sulla porta 8080

Process finished with exit code 130
|
```

Il terminale relativo al client sarà invece quello con il quale l'utente dovrà interagire:

## COMANDI DISPONIBILI

Il programma offre due principali operazioni, utilizzabili tramite il terminale del client appena aperto:

```
Scegli una opzione:  
(1) Carica Cluster da File  
(2) Scopri Cluster da DB  
Risposta (1/2):
```

### 1. Carica Cluster da File:

- Permette di caricare un cluster precedentemente salvato da file;
- Richiede l'inserimento del nome del file (con o senza estensione), se il file non esiste, verrà mostrato un messaggio di errore;

### 2. Scopri Cluster da DB:

- Genera nuovi cluster a partire da dati presenti in un database MySQL già configurato.

Per procedere quindi con questa operazione il programma richiede:

#### 1. Nome della tabella da analizzare

L'utente inserisce il nome della tabella già presente nel database.

Se il nome della tabella è scritto in modo errato o la tabella non è raggiungibile (perché inesistente o per problemi di connessione al DB), viene mostrato un messaggio di errore chiaro e l'operazione si interrompe.

#### 2. Valore del raggio (threshold R)

Il programma chiede all'utente di inserire un numero reale maggiore di 0.

Se il raggio non è  $>0$  oppure è troppo grande (e di conseguenza ingloba tutti gli item in un unico cluster), viene mostrato un messaggio di errore e si richiede un nuovo valore.

#### 3. Salvataggio File

Una volta calcolati i cluster, il programma richiede il nome del file in cui salvare i cluster.

Se l'utente preme spazio + invio allora per salvare i dati si userà il nome di default, costruito automaticamente nella forma: NomeTabella.Raggio.dmp.

## ESEMPIO FLUSSO DI ESECUZIONE

1. All'avvio, il programma chiede di scegliere tra le due opzioni principali

2. Opzione "Scopri da DB": inserire prima il nome della tabella

```
Scegli una opzione:  
(1) Carica Cluster da File  
(2) Scopri Cluster da DB  
Risposta (1/2): 2  
Nome tabella: playtennis
```

2.2. Il programma mostra i dati presenti nella tabella;

2.3. Il programma chiede l'invio del raggio (maggiore di 0);

2.4. Il programma mostra i risultati della computazione;

Esempio di alcuni cluster scegliendo un raggio piccolissimo:

Dati tabella:

outlook,temperature,umidity,wind,play

1:sunny, 30.3, high, weak, no

2:sunny, 30.3, high, strong, no

3:overcast, 30.0, high, weak, yes

4:rain, 13.0, high, weak, yes

5:rain, 0.0, normal, weak, yes

6:rain, 0.0, normal, strong, no

7:overcast, 0.1, normal, strong, yes

8:sunny, 13.0, high, weak, no

9:sunny, 0.1, normal, weak, yes

10:rain, 12.0, normal, weak, yes

11:sunny, 12.5, normal, strong, yes

12:overcast, 12.5, high, strong, yes

13:overcast, 29.21, normal, weak, yes

14:rain, 12.5, high, strong, no

Inserisci raggio (>0):0.1

Numero cluster: 14

1: Centroid=(overcast 0.1 normal strong yes )

Examples:

[overcast 0.1 normal strong yes ] dist=0.0

AvgDistance=0.0

2: Centroid=(overcast 12.5 high strong yes )

Examples:

[overcast 12.5 high strong yes ] dist=0.0

AvgDistance=0.0

3: Centroid=(overcast 29.21 normal weak yes )

Examples:

[overcast 29.21 normal weak yes ] dist=0.0

AvgDistance=0.0

4: Centroid=(overcast 30.0 high weak yes )

Examples:

[overcast 30.0 high weak yes ] dist=0.0  
AvgDistance=0.0

5: Centroid=(rain 0.0 normal strong no )

Examples:

[rain 0.0 normal strong no ] dist=0.0  
AvgDistance=0.0

6: Centroid=(rain 0.0 normal weak yes )

Examples:

[rain 0.0 normal weak yes ] dist=0.0  
AvgDistance=0.0

7: Centroid=(rain 12.0 normal weak yes )

Examples:

[rain 12.0 normal weak yes ] dist=0.0  
AvgDistance=0.0

8: Centroid=(rain 12.5 high strong no )

Examples:

[rain 12.5 high strong no ] dist=0.0  
AvgDistance=0.0

9: Centroid=(rain 13.0 high weak yes )

Examples:

[rain 13.0 high weak yes ] dist=0.0  
AvgDistance=0.0

10: Centroid=(sunny 0.1 normal weak yes )

Examples:

[sunny 0.1 normal weak yes ] dist=0.0  
AvgDistance=0.0

11: Centroid=(sunny 12.5 normal strong yes )

Examples:

[sunny 12.5 normal strong yes ] dist=0.0  
AvgDistance=0.0

12: Centroid=(sunny 13.0 high weak no )

Examples:

[sunny 13.0 high weak no ] dist=0.0  
AvgDistance=0.0

13: Centroid=(sunny 30.3 high strong no )

Examples:

[sunny 30.3 high strong no ] dist=0.0  
AvgDistance=0.0

14: Centroid=(sunny 30.3 high weak no )

Examples:

[sunny 30.3 high weak no ] dist=0.0  
AvgDistance=0.0

4. Al termine, chiede il nome del file in cui salvare i risultati:

- Premere spazio + Invio per usare il nome di default , oppure inserire un nome file personalizzato.

```
Premi spazio e invio per usare il Nome di Default = playtennis0.1.dmp  
Nome file salvataggio: nome_file_personalizzato.dmp  
Salvataggio completato.
```

5. Infine, chiede se eseguire un'altra operazione (y/n). Scegliendo “n” il programma termina, scegliendo “y” l’utente viene riportato al menu principale, dove potrà scegliere di nuovo tra le opzioni presenti.

## SUGGERIMENTI

- I file salvati hanno estensione .dmp, ma è possibile usare qualsiasi altra estensione come .txt;
- I Javadoc sono collocati nelle cartelle doc:
  - Per il client: /codice\_con\_javadoc/client/doc/;
  - Per il server: /codice\_con\_javadoc/server/doc/.