**Hierarchical Cluster Miner**

Metodi avanzati di Programmazione – Caso di studio a.a. 2023-2024

Progetto realizzato da: Blanco Lorenzo - Cannito Antonio

**Indice**

1. Introduzione
2. Guida di installazione
3. Casi di test
4. Documentazione

**1) Introduzione**

Il progetto consiste nella creazione di:

* Un applicativo server, il quale include funzionalità di data mining per la scoperta di un dendrogramma di cluster di dati con algoritmo di clustering agglomerativo, interagendo con la base di dati e comunica con il client;
* Un applicativo client: costituisce l’interfaccia utente, veicola le richieste dell’utente verso il server

Dati una collezione di transazioni dove, ogni transazione è un vettore di valori misurati per una collezione di attributi numerici, ed un intero k, lo scopo del clustering è partizionare la collezione di transizioni in k insiemi di transizioni tali che

Immagine che contiene testo, Carattere, bianco, linea

Descrizione generata automaticamente

Assumendo che un esempio è un vettore di numeri reali, la distanza tra due esempi è calcolata tramite la distanza Euclidea. Per partizionare la collezione di transizioni si possono considerare due tipi di distanze tra cluster:

* Distanza single link, la distanza minima tra tutti gli elementi dei due cluster;
* Distanza average link, la distanza media delle distanze tra tutti gli elementi dei due cluster

Per la costruzione del cluster agglomerativo, per ogni livello vengono uniti i due cluster con distanza minima, con distanza una a scelta tra quelle precedenti, fino al livello specificato dall’utente.

Il client una volta caricato il dataset contenuto nel database, può decidere se apprendere il dendrogramma da database oppure caricare il dendrogramma da file.

**2) Guida di installazione**

Prima installazione: eseguire lo script “setup.bat” presente nella cartella Base/Script/Bat il quale si occupa di creare l’utente, il database, le tabelle e di inserire dei valori di esempio su Mysql, infine esegue l’applicativo server e l’applicativo client ( se la password dell’utente “root” per il server Mysql è diversa da “root”, modificare il file setup.bat sostituendo la password corretta ).

Per le esecuzioni successive, eseguire gli script Base/Script/Bat/avvia\_server.bat e Base/Script/Bat/avvia\_client.bat in questo ordine.

**3) Casi di test**

***Leggenda****:* Ogni caso di test è composto da delle pre-condizioni, che descrivono i passaggi necessari per raggiungere la sezione del programma da verificare, il flusso di esecuzione che rappresenta il normale funzionamento del codice, le post-condizioni che indicano la parte del programma che si dovrebbe raggiungere in condizioni normali di operatività, infine vi sono gli scenari alternativi che rappresentano alcune possibili situazioni di errore che potrebbero verificarsi e il comportamento previsto del programma in tali circostanze.

***CASO DI TEST 1*** **:** Creazione della connessione client-server

Pre: E’ stata effettuata la prima installazione

Flusso di esecuzione: Eseguire lo script avvia\_server.bat, successivamente eseguire lo script avvia\_client.bat

Post: Entrambi gli applicativi sono stati avviati correttamente.

Sull’applicativo server si visualizza:

Immagine che contiene testo, schermata, nero, Carattere

Descrizione generata automaticamente

Sull’applicativo client si visualizza:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

***Scenario alternativo*** : La porta occupata dal server risulta già in uso per un altro processo.

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

In tal caso per liberare la porta 8080 da altri processi è possibile usare lo script: Base/Script/bat/free8080.bat.

***CASO DI TEST 2*** **:** Caricamento del dataset

Pre: Portare a termine con successo CT1

Flusso di esecuzione: il client invia al server il nome della tabella di cui recuperare il dataset ; il server costruisce il dataset recuperando i dati dal database dalla tabella il cui nome è stato dato dal client.

Post: Se il nome della tabella è valido si visualizza sul client la scelta del caricamento da effettuare

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

Sul server invece

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

***Scenario alternativo*** 1) La tabella con nome specificato dal client è inesistente.

Sul client si visualizza

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

Sul server si visualizza

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

Al client viene richiesto di inserire un altro nome di tabella fino a quando il nome è valido.

***Scenario alternativo*** 2) La tabella con nome specificato dal client ha 0 attributi oppure non contiene nessun Esempio.

Sul client si visualizza

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Sul server si visualizza

Immagine che contiene testo, Carattere, schermata, nero

Descrizione generata automaticamente

Al client viene richiesto di inserire un altro nome di tabella fino a quando il nome è valido.

***CASO DI TEST 3*** **:** Il client deve scegliere se caricare il dendrogramma da file o apprendere il dendrogramma da database.

Pre: Portare a termine con successo CT2

Flusso di esecuzione: il client inserisce “1” per caricare il dendrogramma da file o “2” per apprendere il dendrogramma da database

Post: viene chiesto all’utente il nome dell’archivio dal quale recuperare il dendrogramma memorizzato.

Se il client sceglie 1, sul client si visualizza

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Se il client sceglie 2, sul client si visualizza

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

***Scenario alternativo :*** Il client inserisce una scelta non valida

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

In tal caso gli viene richiesto di inserire una nuova scelta fin quando non è corretta.

***CASO DI TEST 4*** **:** Caricamento da file

Pre: Portare a termine con successi CT3 e aver scelto 1

Flusso di esecuzione: Il client inserisce l’archivio da cui recuperare il dendrogramma

Post: il dendrogramma recuperato da file viene stampato a video e l’esecuzione del client termina.

Immagine che contiene testo, schermata

Descrizione generata automaticamente

Sul server si visualizza

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

***Scenario alternativo*** 1) il client inserisce il nome di un archivio non valido

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

Sul server si visualizza

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

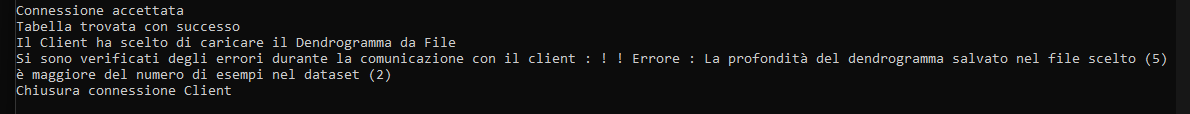
In tal caso la connessione con quello specifico client si interrompe ma il server rimane disponibile per altre richieste da altri client.

***Scenario alternativo*** 2) il client inserisce il nome di un archivio in cui è salvato un dendrogramma con profondità maggiore del numero di esempi del dataset caricato.

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Sul server si visualizza



In tal caso la connessione con quello specifico client si interrompe ma il server rimane disponibile per altre richieste da altri client.

***CASO DI TEST 5*** **:** Apprendimento da database, introduzione della profondità

Pre: Portare a termine con successo CT3 e aver scelto 2

Flusso di esecuzione: Il client inserisce la profondità del dendrogramma da costruire.

Post: il dendrogramma viene costruito lato server e viene chiesto all’utente di scegliere il tipo di distanza che desidera utilizzare per il clustering.

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Sul server si visualizza

Immagine che contiene testo, schermata, Carattere

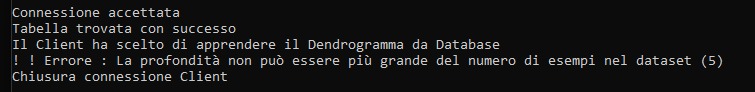
Descrizione generata automaticamente

***Scenario alternativo :*** il client inserisce una profondità maggiore del numero di esempi del dataset

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

Sul server si visualizza



In tal caso la connessione con quello specifico client si interrompe ma il server rimane disponibile per altre richieste da altri client.

***CASO DI TEST 6*** **:** Scelta del tipo di distanza

Pre: Portare a termine con successo CT5

Flusso di esecuzione: Il client sceglie la distanza tra Single Link (1) e Average Link (2)

Post: viene stampato il dendrogramma, a seconda della distanza scelta, e viene chiesto all’utente il nome dell’archivio dove effettuare il salvataggio.

Se il client sceglie 1 si visualizza

Immagine che contiene testo, schermata, menu, Carattere

Descrizione generata automaticamente

Se il client sceglie 2 si visualizza

Immagine che contiene testo, schermata, menu, Carattere

Descrizione generata automaticamente

Sul server si visualizza

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

***Scenario alternativo*** : Il client sceglie una distanza non valida

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

In tal caso viene richiesto al client di inserire una nuova distanza fino a quando non inserisce una valida.

***CASO DI TEST 7*** **:** Salvataggio su file

Pre: Portare a termine con successo CT6

Flusso di esecuzione: Il client inserisce il nome dell’archivio dove vuole salvare il dendrogramma.

Post: viene salvato il dendrogramma su file, la connessione tra client e server viene terminata ma il server rimane disponibile per altre richieste da altri client.

Sul client si visualizza

Immagine che contiene testo, schermata, menu, Carattere

Descrizione generata automaticamente

Sul server si visualizza

Immagine che contiene testo, schermata, Carattere, nero

Descrizione generata automaticamente

**4) Documentazione**

Nella cartella Base/Documentazione/javadoc è presente la documentazione javadoc.

Nella cartella Base/Documentazione/uml sono presenti i diagrammi delle classi.