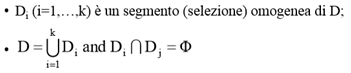
Descrizione del progetto

Il lavoro prevede la progettazione e la realizzazione di un sistema client-server denominato “QT”, dove:

* Il server è un’applicazione che include funzionalità di data mining per la scoperta di cluster di dati.
* Il client è un’applicazione Java che consente di usufruire del servizio di scoperta remoto e visualizza i cluster individuati.

Il *clustering* (o *segmentazione*) individua gruppi con elementi omogenei all’interno di un gruppo e diversi da gruppo a gruppo. Il *clustering* ha come obbiettivo il partizionamento di un insieme di transazioni D, dove ogni transazione è una tupla, cioè un vettore di item (coppie attributo-valore), in k insiemi di transazioni D1, D2, …, Dk tale che:



dove k è un intero e D è una collezione di transazioni.

Eseguiamo il cluster mediante l’algoritmo di Quality-Threshold (QR), dove i cluster sono rappresentati con descrizione intensionale mediante i centroidi del cluster e il raggio, con lo scopo di calcolare e memorizzare i centroidi contenuti nei cluster.

L’algoritmo di Quality-Threshold ha come parametri di input D e un raggio, e come output un clusterSet. Quest’ultimo è un insieme di cluster Di che a loro volta sono insiemi di transazioni di D.

Il primo passo è determinare i cluster candidati associati a ciascuna transazione mediante la costruzione del cluster candidato C per ogni transazione t in D; successivamente C raggruppa le transazioni t’ di D tale che: distanza (t, t’) <= raggio.

Il secondo passo è la scelta del cluster candidato più popoloso come cluster definitivo da aggiungere a clusterSet e la rimozione delle transazioni in esso incluse da D.

L’algoritmo poi controlla se esistono ancora transazioni in D, e torna al primo passo.

I cluster e/o centroidi sono utilizzati nelle applicazioni reali perché hanno il vantaggio di:

* compattare in termini di spazio di memoria, perché viene memorizzata una singola transazione invece di un insieme di transazioni.
* Individuare il segmento a cui appartiene una nuova transazione scegliendo il cluster più vicino mediante l’uso dei centroidi dei cluster.

Il progetto differisce dalla versione iniziale per l’aggiunta di un’interfaccia utente, basata su libreria JavaFX, che permette di eseguire in maniera interattiva le operazioni previste dal progetto.