



UNIVERSITÀ DI PISA

Corso di Laurea in Business Informatics

Visual Analytics

Multidimensional graphical analysis of musical features

A.A. 2015-2016

Tommaso Furlan (538049)

Maria Teresa Rossi (535893)

Abstract

Presentazione dataset

Il progetto è basato sul Million Song Dataset, una collezione di informazioni riguardanti circa 700 000 brani, disponibile online. Essendo in formato hdf5, è stato utilizzato uno script in python ([link](#)), trovato in rete ma opportunamente modificato, in grado di estrarre i dati riguardanti ogni singolo brano. In questo modo si è ottenuto un file .csv, dove ogni riga corrisponde a uno dei brani sopra citati.

Dopo un'analisi preliminare sui dati ottenuti è emersa la mancanza di informazioni utili e significative, a partite da questa fonte, quindi per ogni artista presente nel dataset, sono stati scaricati dati relativi a 10 brani dalle API di Spotify, ottenendo così un dataset di 500 000 righe.

Delle informazioni che abbiamo mantenuto nel nostro file possiamo considerare 17 attributi:

NOME	TIPO DI ATTRIBUTO	BREVE DESCRIZIONE
ArtistName	String	Nome dell'artista che ha composto la canzone.
Energy	Float	Misura compresa tra 0.0 e 1.0 dell'intensità della canzone.
Liveness	Float	Probabilità che la canzone sia stata eseguita in presenza di un pubblico.
Tempo	Float	Il tempo in BPM.
Speechiness	Float	Misura la presenza di parlato nella canzone.
Uri	String	L'URI di Spotify per la canzone.
Acousticness	Float	Misura compresa tra 0.0 e 1.0 di quanto la canzone è acustica.
Instrumentalness	Float	Indica se una canzone contiene o meno dei vocalizzi.
TimeSignature	Int	Il ritmo della canzone.
Danceability	Float	La ballabilità compresa tra 0.0 ed 1.0 della canzone.
Key	Int	Tonalità della canzone.
Duration_ms	Float	Durata della canzone in millisecondi.
Loudness	Float	Rumorosità in decibel.
Valence	Float	Misura compresa tra 0.0 ed 1.0 della positività della canzone.
Id	String	ID di Spotify per la traccia.
Mode	Int	Modo della canzone (1 maggiore, 0 minore).
Genre	String	Genere della canzone (Blues, Country, Electronic, International, Jazz, Latin, Pop/Rock, R&B, Rap, Reggae).

Il genere è stato aggiunto considerando una risorsa trovata online (<http://web.cs.miami.edu/home/yajiehu/resource/genre/>), che include il genere dei brani contenuti nel dataset. È opportuno specificare che i generi sono stati associati alle canzoni tramite una combinazione di classificatori basati su confidenza, pertanto l'affidabilità non è del 100%.

Essendo troppo vasto ai fini dell'analisi, dal dataset è stato estratto un brano per ogni artista, ottenendo così una collezione di circa 50 000 record.

Finalità dell'analisi

La nostra analisi sarà contestuale ad ogni genere per cui analizzeremo da un lato gli attributi strutturali del brano, come la chiave, il ritmo, il modo ed il tempo, dall'altro vorremmo osservare l'andamento dell'energia e della positività(valence) rispetto agli altri attributi, al fine di estrarre dei pattern significativi e chiari all'utente attraverso l'uso di opportune visualizzazioni.



Figura 1 - Esempio di una possibile analisi
(fonte http://sobigdata.borgheresi.it/music_in_numbers)