# La musica si ascolta o si legge?

**Introduzione**

La musica, l’arte di produrre suoni armoniosi mediante l’uso di voce o strumenti musicali, accompagna da sempre le nostre vite, colorando le nostre giornate o accompagnandoci nei pensieri più tristi e difficili.

L’uomo ha da sempre avuto la necessità di esprimersi facendo musica e nel corso di migliaia di anni essa con lui si è evoluta differenziandosi in vari generi, classificati tramite una scelta arbitraria.

Se vi venisse chiesto di etichettare una canzone in un genere musicale provvedereste subito ad ascoltarla, non a leggerne misurazioni e dati su essa, ma difficilmente una macchina applicherebbe il vostro stesso metodo. Può quindi un computer senza udito e senza conoscenza musicale distinguere diversi brani in base al genere?

Questa è la domanda che ci siamo posti e per arrivare ad una conclusione abbiamo lavorato su R studio utilizzando un dataset contenente 1000 tracce audio ciascuna lunga 30 secondi e appartenente a un distinto genere musicale. Per ogni genere sono associate 100 tracce per un totale di 10 gruppi (blues, country, classica, disco, hip hop, jazz, metal, pop, reggae e rock).

Su ogni traccia sono misurate 28 variabili:

|  |  |
| --- | --- |
| Tempo | Velocità con cui viene riprodotto un brano |
| Beats | Il ritmo di ogni brano |
| Chorma\_stft | Fornisce una misura media del peso delle diverse frequenze nel segnale in esame, in tutta la sua durata |
| Rmse | Radice quadrata dell’ errore quadratico medio delle frequenze del suono |
| Spectral centroid | Indica dove si trova il centro di massa dello spettro sonoro |
| Spectral bandwidth | L’intervallo di lunghezze d’onda in cui una quantità spettrale irradiata non è inferiore alla metà del suo valore massimo |
| Rolloff | Il tasso di variazione dell'uscita del filtro rispetto alla frequenza |
| Zero crossing rate | La velocità con cui il segnale cambia da positivo a negativo o viceversa |
| Mfcc1/ Mfcc20 | Le frequenze che l’orecchio umano riesce a percepire |

Il primo passo è quindi quello di condurre un’analisi esplorativa per ridurre il numero di variabili.Immagine che contiene testo, schermata, diagramma, Diagramma

Descrizione generata automaticamente

Di fianco viene riportata la heatmap della matrice di correlazione del dataset. Si evince che alcune variabili come (Tempo e Beats) o come (Spectral Bandwidth/ centroid e roll off) siano altamente correlate fra di loro.

Immagine che contiene testo, schermata, diagramma, Carattere

Descrizione generata automaticamente

Non consideriamo tutte le variabili nel processo di clustering in quanto risulterebbe il problema di variabili ridondanti che possono creare problemi nelle analisi.

Tramite l’analisi delle componenti principali selezioniamo le 7 variabili che spiegano l’80% della varianza del dataset. Le variabili selezionate sono: mfcc8, rolloff, mfcc16, tempo, zero\_crossing\_rate, mfcc3, mfcc19.

Questa scelta è coerente in quanto prendiamo una variabile legata al ritmo (tempo), una alla velocità della frequenza del suono (zero\_crossing\_rate), una legata agli strumenti di ascolto (rolloff) e le ultime legate alla sensibilità dell’udito umano (mfcc3 per i bassi, mfcc8 e mfcc16 per i medi e mfcc19 per gli acuti).

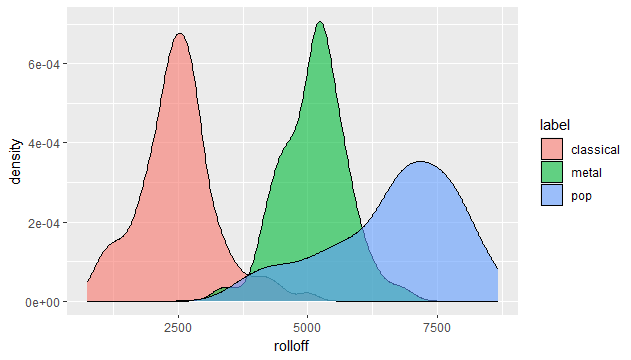
|  |  |  |
| --- | --- | --- |
| Immagine che contiene testo, diagramma, schermata, Policromia  Descrizione generata automaticamente | Immagine che contiene Policromia, schermata, linea  Descrizione generata automaticamente | |
| Immagine che contiene testo, schermata, diagramma, Policromia  Descrizione generata automaticamente | Immagine che contiene testo, schermata, diagramma, Policromia  Descrizione generata automaticamente | Immagine che contiene testo, schermata, diagramma, quadrato  Descrizione generata automaticamente |
| Immagine che contiene testo, diagramma, schermata, Policromia  Descrizione generata automaticamente | Immagine che contiene testo, diagramma, schermata, quadrato  Descrizione generata automaticamente | Immagine che contiene testo, diagramma, schermata, Policromia  Descrizione generata automaticamente |

I boxplot sono distinti per colore in base al genere musicale, le loro diverse posizioni sull’asse y mostrano come i valor medi relativi alle variabili selezionate nell’analisi precedente differiscano tra i gruppi. Alcuni generi tendono ad avere dei boxplot spessi, sintomo di un’alta variabilià. Per esempio il genere Metal (soprattutto per le variabili mfcc\_ e rollof) sembra essere ben definito e poco variabile rispetto al genere Blues. Quest’ultimo varia e non poco talvolta anche più del genere Pop che a volte ci sembra essere il genere più variopinto.

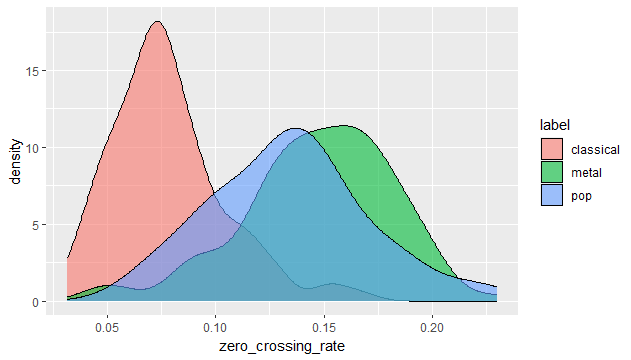
**MODEL BASED CLUSTERING**

Selezioniamo 3 generi ben distinti (Metal – Pop - Classica) con l’obiettivo di applicare tecniche di clustering, al fine di capire se l’algoritmo implementato, con le funzioni della libreria ‘mclust’, è in grado di distinguere bene i differenti stili musicali di 300 tracce selezionate. Nel dataset ad ogni osservazione è corrisposta una label (blues, country ….), quindi conoscendo il corretto etichettamente possiamo anche proporre una stima di bontà dell’algoritmo di classificazione.

Un problema importante da porsi è come si distribuiscono i valori delle variabili all’interno dei differenti gruppi, e se seguano leggi differenti.

Da questo grafico, che mostra la densità della variabile **rolloff**, si nota come la distribuzione cambi tra il genere classico, metal e pop.

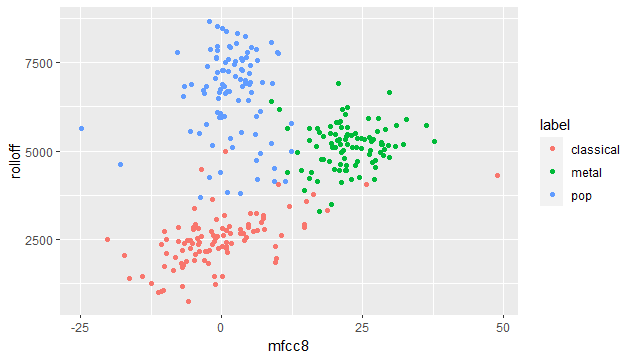
Il **rolloff** può essere diverso nella musica classica e pop a causa delle differenti preferenze estetiche e stili compositivi. La musica classica sfrutta una varietà di strumenti orchestrali che producono suoni complessi e armonici, per questo il rolloff è progettato in modo tale che ogni dettaglio sia udibile anche nelle frequenze più alte o più basse. Nel pop gli arrangiamenti possono essere più semplici e concentrarsi su suoni più immediati e facilmente udibili, con un'attenzione meno focalizzata sui dettagli complessi.



Questo grafico invece fa riferimento alla densità sulla variabile **zero crossing rate.**

Il genere metal o pop hanno dei tassi di attraversamento di zero tendenzialmente più alti e vari rispetto alla musica classica. Questo è sempre dovuto allo stile classico che tende ad avere sezioni con ritmi più lenti e con meno variazioni, di conseguenza non ci sarà un numero elevato di volte in cui il segnale audio attraversa l’asse temporale zero cambiando polarità. Al contrario il metal e il pop presentano ritmi più veloci e densi con rapide variazioni di frequenza che causano un maggior numero di attraversamenti.

GRAFICO DI DISPERSIONE VARIABILE ROLLOFF (ASSE Y) E VARIABILE MFCC8 (ASSE X)



Il grafico a dispersione evidenzia bene la distinzione in cluster dei differenti generi musicali in base ai valori rolloff e mfcc8. (**Musica Classica** – **Metal** – **Pop**)

Dall’analisi predente svolta sulle distribuzioni nei differenti cluster, supponiamo la presenza di una mistura all’interno della popolazione, applichiamo quindi la **Model Based Clustering**.

Nella model based clustering è fondamentale la scelta del modello che pone vincoli geometrici al fine di limitare la complessità della variabilità. Secondo il criterio ICL ( integrated complete likelihood) è stato scelto un modello ‘EEV’ (equal volume, equal shape, variable orientation). Abbiamo implementato un algoritmo di clustering, utilizzando la funzione ‘Mclust()’ della library mclust e, conoscendo il vero etichettamento delle tracce, con la funzione ‘classError()’ abbiamo individuato la percentuale totale e le osservazioni mal classificate.

Immagine che contiene testo, Carattere, schermata

Descrizione generata automaticamenteSono state scelte tracce appartenenti a 3 generi musicali molto differenti ma il risultato ottenuto è comunque sopra ogni aspettativa. L’accuracy dell’algoritmo è del 94,6% ma più della metà delle unità ‘misclassified’ (9/16) provengono dalla label pop.

Immagine che contiene testo, schermata, mappa, diagramma

Descrizione generata automaticamenteEvidenziate con colori leggermente diversi e con forma quadrata sono rappresentate le osservazioni che sono state classificate in modo sbagliato dall’algortimo. Si può notare come queste osservazioni siano quasi outlier e lontani dalle medie dei propri gruppi, infatti esse si trovano graficamente al centro.

CONFUSION MATRIX

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Classical | Metal | Pop | Tot |
| Classical | **96** | **1** | **1** | **98** |
| Metal | **4** | **97** | **8** | **109** |
| Pop | **0** | **2** | **91** | **93** |
|  |  |  |  |  |
| Tot | **100** | **100** | **100** | **300** |

Si nota che le osservazioni di genere pop sono quelle che sono state classificate peggio, l’algoritmo ha classificato 8 osservazioni pop nel genere metal, si è distinto molto meglio invece il genere classico.

Quantita cardine deducibili dalla confusion matrix:

* Accuracy: traccia/n = (96+97+91)/300 = 0,946 percentuale di osservazioni ben classificate
* Sensitivit: f33/f.3 = 91/100 = 0,91 percentuale di osservazioni pop ben classificate ( la più bassa tra i generi)

Tramite la divergenza di Kulback Leibler ci calcoliamo la distanza tra i gruppi:

|  |  |  |
| --- | --- | --- |
| distanza tra classico e metal | Distanza tra classico e pop | Distanza tra metal e pop |
| 45,81 | 114, 59 | 30,75 |

È curioso il risultato ottenuto, in quanto a primo ascolto penseremmo che artisti come Madonna siano più simile allo stile di Mozart rispetto a quello dei Metallica.

**MODEL BASED CLASSIFICATION**

L’intento è quello di applicare la model based classification. Dividiamo la parte di dataset che contiene i generi classico, metal e pop in training set (80% delle osservazioni) e test set (20% delle osservazioni) e applichiamo due famiglie di classificatori EDDA e MDA. La prima ipotizza la presenza di una mistura di normali d-dimensionali nella popolazione, l’ MDA invece introduce la flessibilità del classificatore supponendo una mistura di mistura di normali.

**EDDA**

Sul training set alleniamo il classificatore al fine di trovare il modello migliore. Una volta svolta la parte di **model choice** ipotiziamo di non conoscere le labels delle osservazioni del test set estratte casualmente e le classifichiamo sulla base del modello, scelto grazie alla funzione ‘mixmodPredict’.

Model Choice :

Si utilizza la funzione ‘mixmodLearn’ presente nella library ‘Rmixmod’ scegliendo il criterio in base alla cross validation con V pari a 10. Siccome la tecnica di cross validation lavora con l’estrazione di valor casuali è consigliabile svolgere l’algoritmo diverse volte all’interno di un ciclo.

Il modello scelto è il seguente : Gaussian\_p\_L\_Ck (equal volume free shape free orientation)

Classification:

Utilizzando la funzione ‘mixmodPredict’ sono state classificate le 60 osservazioni estratte casualmenete dalle 300 osservazioni di genere classico, metal e pop. L’accuracy è più che soddisfacente: 98,9% e osservazioni ben etichettate, di seguito viene riportata la confusion matrix.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Classical | Metal | Pop |
| Classical | 18 | 0 | 0 |
| Metal | 0 | 23 | 0 |
| Pop | 0 | 1 | 18 |

**MDA**

Applichiamo un secondo modello di classificazione ovvero l’MDA ( mixture discriminant anlysis) che assume che la normalità all’interno dei gruppi non vada bene, e che vada invece modelizzata assumendo una mistura di normali all’interno del gruppo stesso.

Abbiamo imposto che ogni mistura di normali per classe seguisse un modello EVV trovato grazie alla model choice eseguita in precedenza. La problematica è stata la scelta del numero di gruppi (e quindi di misture di normali) all’interno dei generi, ma siccome il criterio di cross validation non è implementato nella funzione ‘MclustDA’ lo abbiamo programmato noi manualmente.

Immagine che contiene diagramma, linea, Diagramma, testo

Descrizione generata automaticamenteil seguente grafico mostra l’andamento dell’ errore di classificazione in base al numero di gruppi all’interno dei generi.

Si nota appunto che per G=1 l’errore è minimizzato e applicando classificazione con ‘mclustDA’ con g=1 non vengono mal classificate nessuna osservzione nuova.

In questo caso la famiglia di classificazione MDA ha poco senso in quanto con G=1 stiamo ipotizzando che ci sia solo una mistura di normali all’interno di ogni gruppo che è la medesima ipotesi nell’EDDA.

**CONCLUSIONI**

La model based clustering e classification sono risultate utili ai fini di distinguere canzoni di diverso genere. Abbiamo si voluto lavorare con 3 generi ben distinti in modo da evidenziare maggiormente le differenze nei dati ma in ogni caso il nostro lavoro può essere ampliato anche a tutti i generi presenti nella musica.

Spotify utilizza appunto delle tecniche di Machine Learning, implementando algoritmi che sono in grado di riconoscere i tratti distintivi di ciascun genere e proporli all’ascoltatore.

Un tempo sceglievamo i brani secondo i nostri gusti musicali solo dopo averli ascoltati, ora invece ci vengono già proposte canzoni che sono di nostro interesse togliendoci la piacevole ricerca a noi. È notevole e a tratti affascinante la complessità che c’è dietro a questi algoritmi, ma può questo portarci in un ipotetico mondo in cui conosciamo solo ciò che ascoltiamo e quindi solo ciò che ci piace?