Indice

Scansionando i codici QR accanto alle immagini è possibile visionare i grafici dinamici elaborati dalle autrici. Per i grafici più complessi, dopo aver scansionato i codici, potrebbe essere necessario cliccare sulla barra dell'url e aggiornare la pagina.

Introduzione

1. Ligabue: non dovete badare al cantante

Questo libro nasce dall'incontro tra due persone con le stesse due passioni di vita: la statistica e Ligabue. E chi lo ha detto che nella vita non si può unire la passione musicale con la vita lavorativa quotidiana? Karl Pearson diceva «Statistics is the grammar of Science» e chi può non definire la musica come Scienza? La musica che incontra la scienza dà origine a creatività e bellezza, e in questa libro si andrà alla ricerca di questa connessione. Attraverso l'uso di sofisticate tecniche statistiche viaggeremo tra le caratteristiche musicali dei brani di Luciano Ligabue, studiandone i cambiamenti temporali e gli elementi di maggiore interesse per chi lo ascolta ma anche per chi lo critica. Ma, oltre alla musica, quello che colpisce di Ligabue sono anche i testi delle sue canzoni. Tempo fa mi è capitato di leggere su un blog un commento che ho condiviso totalmente e che suonava così: «Ligabue è quel cantante che, mentre tu credi non saprà mai della tua esistenza, sta già scrivendo un testo che ti conoscerà più di te stesso». Da quel momento ho pensato che sarebbe stato interessante cercare di far venir fuori, con l'aiuto di metodologie statistiche, i messaggi nascosti dietro le parole usate da Ligabue nei testi; da qui l'interesse per un'analisi testuale delle canzoni, sia in termini di distribuzioni di frequenza delle parole usate sia in termini di emozioni trasmesse agli ascoltatori.

È chiaro che le procedure statistiche utilizzate in questa tesi possono essere applicate alla musica e ai testi di qualsiasi altro

cantautore, pertanto, riferendomi a coloro che leggeranno queste pagine per puro interesse statistico, con una citazione dello stesso Ligabue, dico: «Non dovete badare al cantante», considerate la discografia di Ligabue solo un'applicazione statistica di metodologie di valenza più generale a un caso che, personalmente, ritengo interessante e stimolante; e poi chi lo sa, magari potreste scoprire anche voi che i testi di Ligabue vi trasmettono qualche sensazione ("sentiment"), alla quale non avevate dato mai peso prima, e diventate anche voi dei fan, «Mai dire mai».

Capitolo 1

Metti in circolo i tuoi dati

1.1. Music Data Mining

Questo libro mira a comprendere, attraverso un'analisi sistematica e quantitativa della discografia di Ligabue, quali siano le "dimensioni" musicali e semantiche che maggiormente influenzano la percezione di una canzone da parte del pubblico e che, certamente, contribuiscono a determinarne il successo. L'approccio utilizzato in questo capitolo e nei seguenti per analizzare in modo rigoroso testi e musica del cantautore emiliano ha natura generale e può essere facilmente adattato all'analisi di altri autori e generi musicali.

La definizione di data mining ha origini legate alle attività di estrazioni minerarie. La radice del verbo inglese rimanda al lavoro di rimozione di grandi quantità di materiale all'interno delle miniere alla ricerca di filoni di materiale nobile. L'analogia con l'attività di ricerca ed estrazione dei dati nella Rete è decisamente aderente.

L'utilizzo di strumenti informatici sempre più avanzati ha caratterizzato e reso possibile l'enorme sviluppo di questa attività svolta attraverso un ecosistema di strumenti tecnici e metodologie consolidate. Con il 49.2% di dilagazione nella popolazione mondiale, pur con tutte le differenziazioni tra le diverse latitudini, Internet e le tecnologie a esso connesse, occupa il posto centrale tra gli strumenti di cambiamento sociale e culturale oggi esistenti. Analizzare i dati che delineano questo stato di cose è necessario per poter formulare delle ipotesi, confortate da un'analisi esausti-

va, circa la centralità che va sempre più assumendo il data mining nelle società occidentali; per polarizzare vasti settori della ricerca, pubblica, privata e universitaria, convogliare enormi capitali finalizzati allo sviluppo di sempre più potenti e sofisticati algoritmi specifici per la ricerca, lo stoccaggio, la catalogazione e l'analisi di dati. Tutto proveniente dalla sola Rete.

Ciascun utente utilizzando la Rete, oltre ad attingere enormi quantità di informazioni, immette altrettante quantità di dati che vengono reperiti incessantemente dagli algoritmi di acquisizione attraverso tutto il Cyberspazio. I dati a disposizione della rete sono sempre più diversificati: dati bancari, transazioni finanziarie e commerciali, acquisti e vendite, e-Learning, intrattenimento, immagini, video, musica, file eterogenei, documenti, e-Books.

Il data mining può essere dunque definito come l'insieme di tecniche e metodologie che partendo da informazioni "criptiche", disseminate senza ordine apparente in un database (testuale, multimediale, di dati misti, ecc.), consente di arrivare a una conoscenza sfruttabile e quindi a un utilizzo scientifico di questo sapere. L'intero processo viene chiamato KDD (acronimo di Knowledge Discovery in Databases) e in realtà non si esaurisce con la procedura di data mining vera e propria. La sequenza di KDD infatti, conta più passi, i principali dei quali sono:

1. identificazione dell'obiettivo che si vuole raggiungere;
2. preselezione dei dati utili a raggiungerlo;
3. pulizia dei dati e preelaborazione: ulteriore separazione fra dati validi e inutili, scelta di come trattare i campi incompleti o vuoti, selezione definitiva delle informazioni fondamentali per il modello ideale di riferimento;
4. trasformazione: il formato con il quale sono rappresentati i dati è valido per essere dato in pasto ai software di analisi? Se la risposta è no, i dati devono essere convertiti;
5. data mining: è naturalmente il passo più importante. Viene scelto il software migliore per il singolo caso, il quale scandaglia il data warehouse in modo selettivo per fornire la risposta

cercata. Il data mining solitamente si compone di più sottopassaggi, anche ripetuti diverse volte, per affinare la procedura e verificare man mano i risultati raggiunti;
6. interpretazione dei risultati: si valuta se l'obiettivo è raggiunto, e se la risposta è no si procede con la reiterazione (ed eventuale modifica) del passo precedente e talvolta anche di altri;
7. visualizzazione dei risultati in un formato comprensibile.

Le tecniche e le strategie applicate alle operazioni di data mining sono per larga parte automatizzate, consistendo in specifici software e algoritmi adatti al singolo scopo. A oggi, in particolare, si utilizzano reti neurali, alberi decisionali, clustering e analisi delle associazioni.

I settori di applicazione del data mining sono innumerevoli, ma raggruppabili in alcune macrocategorie. Le principali sono: marketing, economia e finanza, scienza, tecnologie dell'informazione e della comunicazione (ICT), statistica e industria. Tra le tecnologie della comunicazione la Musica oggi assume un ruolo predominante. La musica, infatti, è parte integrante delle nostre vite, è quel linguaggio che ci aiuta a esprimerci e descrivere le nostre emozioni quando nessuna parola ci riesce. Ma oggi la musica sta subendo una svolta epocale: la musica è divenuta digitale e il digitale è sinonimo di Social Media. E se da un lato questo rappresenta un motivo di crisi per le industrie discografiche, dall'altro l'analisi costante delle interazioni sui social è diventata fondamentale per la raccolta di dati che aiutino gli operatori del settore a capire e a orientare il mercato. Diverse aziende, come ad esempio la Universal Music, hanno quindi sviluppato, in collaborazione con società specializzate nei Big Data, sofisticati software allo scopo di creare un *predictive profiling* del fan musicale, non solo per comprenderne personalità e gusti ma, soprattutto, per studiarne le affinità con prodotti di largo consumo allo scopo di scegliere, tra gli artisti, quello più amato e utilizzarlo come testimonial in operazioni di marketing. Anche Shazam è un'altra fonte ricca di dati, contenendo 25 milioni di tracce che creano interazioni tra più di 450 milio-

ni di persone a livello globale. Di queste circa 90 milioni ogni mese taggano 17 milioni di canzoni, generando oltre il 7% delle vendite di Apple Music. E grazie a questo immenso bacino di informazioni, gli analisti di Big Data sono in grado di capire, con mesi di anticipo sul mercato, quali canzoni funzioneranno e quali no, quali saranno i prossimi successi e persino dove si diffonderanno. Se una canzone piace, gli algoritmi prediranno che la prossima hit potrebbe essere una canzone simile, con un rischio implicito di ottenere musica sempre più simile. In base ai risultati di un recente studio fatto dal sito SeatSmart, i testi delle canzoni americane di maggior successo degli ultimi dieci anni appaiono infatti semplici e ripetitivi, utilizzando un lessico di sole 300 parole. E tuttavia, per ora, questo dato sembra importare poco, poiché nel mondo della musica la domanda che assilla i discografici è unica: quale sarà la prossima canzone che la gente vorrà ascoltare?

Così, oltre a Twitter, Pandora e Shazam, anche Spotify e Genius divengono proficue fonti di Big Data "musicali" per riuscire a individuare chi sarà la nuova, brillante stella internazionale e cosa dovrà cantare. Tuttavia, per fortuna, anche se i Big Data sono una risorsa nella crisi del mercato discografico, la qualità delle canzoni e la raffinatezza degli artisti restano non misurabili con formule matematiche. Gli algoritmi, infatti, ottimizzano l'esistente e non lasciano spazio a variabili fuori controllo, mentre la creatività è per sua natura impalpabile, non misurabile e soprattutto non prevedibile. E siamo quindi certi che, anche nell'universo ordinato dei Big Data, ciascuna nota all'apparenza impazzita, magari in un sound "incontenibile" riuscirà ancora a incrinare l'inflessibilità delle statistiche, nella sonorità di un talento imprevedibile.

Queste pagine pertanto andranno oltre la mera ricerca della canzone di Luciano che più piace: si tenterà di capire il motivo che sta dietro il successo ma soprattutto di determinare in che modo le sue canzoni riescono a colpire le emozioni di chi le ascolta, suscitando reazioni di felicità, tristezza o in generale di carica (sia positiva che negativa). Queste domande troveranno una risposta

a partire dai dati estratti con l'aiuto dei programmi musicali più usati del momento, Spotify e Genius.

1.2. Music Metadata

La musica gioca un ruolo importante nella vita di tutti i giorni per molte persone, e con la digitalizzazione, si formano grandi raccolte di dati musicali che tendono a essere ulteriormente accumulate dagli appassionati di musica. Questo ha portato a raccolte musicali, non solo sullo scaffale sotto forma di dischi audio o video e Dischi di dominio, ma anche sul disco rigido e su Internet, per crescere al di là di ciò che in precedenza era fisicamente possibile. Con l'avvento delle nuove tecnologie, È diventato impossibile per un singolo individuo tenere traccia della musica e delle relazioni tra le diverse canzoni. Le tecniche di data mining e di apprendimento automatico possano aiutare la navigazione all'interno del mondo della musica.

Le strategie di data mining sono spesso basate su due problemi principali: la tipologia di dato che si ha a disposizione e l'uso che se ne vuole fare. Lo stesso vale per il data mining musicale. Che tipo di dato è la musica? Una raccolta di brani musicali è costituita da vari tipi di dati; ad esempio, si possono avere dati costituiti da file audio musicali o da metadati come titolo e artista. Che tipo di analisi si possono condurre? Il data mining musicale prevede metodi specifici per rispondere alle più svariate domande: ad esempio la classificazione di genere, l'identificazione di artisti/cantanti, il rilevamento di umore/emozione, il riconoscimento dello strumento, la ricerca della somiglianza musicale, la sintesi musicale e così via.

I metadati musicali contengono varie informazioni relative a determinati brani musicali. In generale, file musicali supportano una struttura nota come ID3, progettata per memorizzare i metadati musicali quali il nome dell'artista, il titolo della traccia, la descrizione della musica e il titolo dell'album. Pertanto, risulta non

troppo complicato estrarre il contenuto informativo dal formato dati ID3.

La forma più semplice per reperire metadati musicali online richiede l'esecuzione di questi mediante l'API (Application Programming Interface). Alcune delle più note applicazioni di base per la gestione di metadati musicali, ad esempio All Music Guide e FreeDB, rappresentano delle piattaforme flessibili per gli appassionati di musica per cercare, caricare e gestire i metadati musicali.

I risultati riportati in questo libro si basano sui metadati estratti da Spotify e sui testi delle canzoni ottenuti da Genius.