

dgl studi
Università di Verona

Corso di Laurea triennale in Bioinformatica
Elaborato finale

Analisi esplorativa di soundscapes con approcci di *Pattern Recognition* e *Machine Learning*

Versione n.1

31/08/2024

Ricettore: M. Biagi

QQ 23/26

Pomari Giambattista, VR38062

Sommario	3
Introduzione	4
1.1 Pattern recognition e machine learning per l'analisi di soundscapes e in generale di audio per l'ecologia	4
1.2 Obiettivo della tesi: classificazione e analisi preliminare di anomaly detection	6
Background	7
2.1 Dall'analisi del suono allo spettrogramma	7
2.2 Features	8
2.3 Standardizzazione dei dati	10
2.4 Classificazione	10
2.4.1 Apprendimento supervisionato	11
2.3.2 Validazione	12
2.5 Anomaly Detection	12
2.5.1 Gli outlier	13
2.5.2 Le applicazioni	13
2.5.3 I metodi	14
Dataset	16
3.1 Dataset prima fase	16
3.2 Dataset seconda e terza fase	17
Classificazione	19
4.1 Configurazioni features	19
4.2 Dettagli classificatore	21
4.3 Problemi di classificazione disegnati	22
4.3.1 Rappresentazione con etichette note	22
4.3.2 Rappresentazione con etichette semantiche	24
4.4 Valutazione del filtraggio di frequenze	26
4.5 Risultati	27
4.5.1 Filtraggio delle frequenze	27
4.5.2 Risultati rappresentazione con etichette note	28
4.5.3 Risultati rappresentazione con etichette semantiche	31
Anomaly Detection	35
5.1 Configurazione features e metodi	35
5.2 Validazione dei risultati	36
5.3 Risultati	36
Conclusione	39
Bibliografia	40

Sommario

Per soundscape si intende un

Il seguente studio si colloca nell'ambito della *Pattern Recognition* e del *Machine Learning*, in particolare alla sua applicazione nell'analisi di soundscape, ossia l'ambiente sonoro composto da suoni naturali e artificiali,

L'analisi si propone di caratterizzare un dataset di soundscape mediante classificazione, per

ricercare quali caratteristiche rappresentano con maggiore precisione il contesto, definire quali delle casistiche proposte realizza i migliori risultati e quale dimensione della finestra nel calcolo dello spettrogramma si adatta meglio. Si è proceduto su due fasi principali: la prima che propone dei casi basati su categorie note, la seconda su categorie semantiche.

Con tali conoscenze derivate, si prosegue in un terza fase sugli stessi dati, effettuando un'analisi esplorativa mediante algoritmi di anomaly detection per inferire informazioni e analizzare eventuali singolari pattern emersi.

Inoltre, è stata condotta un'analisi preliminare di filtraggio per verificare se potesse offrire eventuali vantaggi nelle analisi appena descritte.

Non esistono:

tecniche di

1) TROPPO SPECIFICO.

Io dico "abbiamo implementato una pipeline di PL".

- il segnale audio è stato contenuto in un database denile tracer d'attivazione delle features - o caratteristica (ad esempio features legate al contenuto spettrale, o alla forma, o al timbro)

- mediante tecniche di classificazione si è provato ad ~~classificare~~ misurare le specifiche dimensioni di queste caratteristiche e classificare diverse categorie, con ad esempio giorno/ notte oppure la presenza di un temporale o no.

ANCHE UNA FORMA + DIRE:

- la pipeline proposta è stata fatta su

(anno el. 2018)³

- i risultati sono buoni / incogniti ...

Capitolo 1

Introduzione

In questo capitolo sarà introdotto il contesto applicativo, il quale si compone delle nozioni fondamentali per comprendere l'ambito su cui si è sviluppato questo studio. Saranno definiti i concetti relativi a *pattern recognition* e *machine learning*, al loro utilizzo nell'analisi di *soundscapes*, e in generale nell'analisi di audio nel campo dell'ecologia. Nel paragrafo successivo, si andrà ad esplicitare il fine che ha suggerito lo sviluppo di questo studio.

1.1 Pattern recognition e machine learning per l'analisi di soundscapes e in generale di audio per l'ecologia

la (?)

[1]

La *pattern recognition* (PR) e *machine learning* (ML) rappresentano una branca fondamentale dell'intelligenza artificiale, e definiscono un insieme di tecniche utilizzate per estrarre informazioni dai dati tramite il riconoscimento automatico di specifici schemi, definiti *pattern*. In modo approssimato, si può dire che sono equivalenti, poiché condividono obiettivi, strumenti e approcci.

Il loro impiego è noto in molteplici ambiti: dal riconoscimento vocale o di immagini, all'elaborazione del linguaggio naturale, dai sistemi di raccomandazione, al monitoraggio in tempo reale e molti altri. Tra questi emerge un contesto poco analizzato, che in letteratura si presenta come una sfida ancora aperta: l'analisi di *soundscape*. Prima di esaminare nel dettaglio come i *soundscapes* sono stati affrontati nella PR/ML, per chiarezza, si desidera spiegare cosa si intende con tale definizione, le motivazioni per cui merita attenzione e le varie problematiche annesse.

et al. Quinn *et alii* identificano i *soundscapes* come una "particolare combinazione di suoni in un paesaggio" considerandola come "una caratterizzazione ecologica dei paesaggi" [1]. Gli autori ritengono che la composizione di un *soundscape* si divide in quattro elementi principali: l'antropofonia (ANT: indica l'attività antropogenica), la biofonia (BIO: intesa come le vocalizzazioni della fauna selvatica), la geofonia (GEO: descrive i suoni dei fenomeni meteorologici) [1] e infine la quiete (indicata come il suono dell'ambiente).

A tal proposito, è molto interessante la caratterizzazione fornita da Farina *et alii* [2]. L'argomentazione descritta espone una visione alternativa più mirata e strutturata: separa il concetto di *sonoscape* da *soundscapes*. Con *sonoscape* intende "il mosaico di tutte le non

interpretate informazioni sonore all'interno di un *landscape*" [2]. Da questa definizione si deduce per esclusione l'interpretazione che l'autore attribuisce al *soundscape*, ossia "un *sonoscape* che è stato cognitivamente interpretato in un mosaico di categorie di ANT, BIO e GEO semioticamente interpretate da un organismo" [2]. Un'ulteriore suddivisione separa gli elementi in quelle che definisce unità sonore, i *sonotope* per gli *sonoscape*, e i *soundtope* per i *soundscape*. Rispettivamente vengono definiti dall'autore come una *patch* spazialmente unica di suoni non interpretati, per i *sonotope*, e di suoni di ANT, BIO e GEO semioticamente interpretati da un organismo [2], per i *soundtope*. Rispetto ad un'umana suddivisione in ANT/BIO/GEO [2], questi concetti appena espressi consentirebbero una classificazione con maggiore dettaglio e specificità. Ciononostante, i termini *sonotope* e *soundtope* sono tuttora relegati a mere speculazioni a causa di una scarsità di evidenze empiriche [2].

Sebbene la definizione di *soundscape* possa risultare complessa, molto chiara è invece la sua importanza. Il ruolo che ricopre nell'ambiente naturale rappresenta un segnale della salute dell'ecosistema. Tale segnale può essere utilizzato per studi ecologici [2], diviene un significativo approfondimento della biodiversità e dell'impatto umano [1], può evidenziare cambiamenti negli habitat dove la qualità acustica è fondamentale per la dimensione vitale e il rumore umano risulta deleterio sulla biodiversità [1].

I vantaggi appena descritti supportano e incoraggiano l'analisi degli *soundscape*. *PR/ML* possono dare un grande contributo in tale processo. Sviluppare sistemi automatici mediante tecniche di *PR/ML* permetterebbe di supportare le sfide riguardanti l'analisi dei dati e il monitoraggio in tempo reale dell'ecosistema. La classificazione di *soundscape* consente l'identificazione automatica di suoni indesiderati su grandi quantità di dati [1], inoltre permette di modellare gli effetti e le interazioni di suoni diversi [1], e utilizzare poi tali modelli per identificare pattern spazio-temporali nell'attività sonora [1]. Introdurre un automatismo in grado di monitorare la presenza o l'abbondanza di particolari specie potrebbe prevenire la scomparsa di tale specie in una determinata zona, o addirittura evitarne l'estinzione. Allo stesso modo, può essere utile per prevenire situazioni di pericolo come il bracconaggio. I dati ricavati, sarebbero fonte di studio per molti comportamenti animali in specifici periodi dell'anno, come il corteggiamento.

Questa innovazione nell'analisi degli *soundscape* non è priva di problematiche. Pochi studi di ecoacustica hanno provato a classificare *soundscape*s utilizzando intere categorie sonore come ANT/BIO/GEO e quiete [1]. Tale difficoltà si sviluppa su due elementi. In primo piano, l'identificazione manuale delle sorgenti sonore è altamente dispendiosa in termini di tempo [1]. Ciò è dovuto all'enorme quantità di dati da visionare manualmente che servono a censire

un *dataset* di addestramento per i sistemi di *PR/ML*. Tanto più il *dataset* risulta ampio e dettagliato, maggiore sarà la qualità del sistema sviluppato. Oltre al tempo impiegato si possono sottintendere anche i costi di tale opera. Il secondo punto riguarda le competenze specifiche del settore. Infatti, per censire i dati di addestramento è richiesta una conoscenza della vocalizzazione degli animali del contesto, determinando per necessità la scelta di sviluppare *dataset* di addestramento di piccole dimensioni [1]. Tali *dataset* non riescono a spiegare nel complesso il problema, limitando così la qualità dei sistemi realizzabili.

1.2 Obiettivo della tesi: classificazione e analisi preliminare di anomaly detection

Il seguente studio si propone di caratterizzare un *soundscape*, tramite tecniche di classificazione applicate a PR/ML, per ricercare quali caratteristiche rappresentano con maggiore precisione il contesto, definire quali delle casistiche proposte realizza i migliori risultati e quale dimensione della finestra nel calcolo dello spettrogramma si adatta meglio. Con tali conoscenza si propone successivamente di effettuare un'analisi esplorativa mediante *anomaly detection* per inferire informazioni e analizzare eventuali pattern emersi.

come
come la
prima pagina

Capitolo 2

IN questa lezione
monteremo:
infine bibliografia;
oggi up.

Background

L'obiettivo di questo capitolo consiste nel fornire all'utente le conoscenze necessarie per poter comprendere l'analisi affrontata nei capitoli successivi e una conoscenza utile da portare con sé. Esse comprendono l'introduzione all'analisi di un segnale (definendo il concetto di spettrogramma), la standardizzazione dei dati, la sua rappresentazione mediante le caratteristiche selezionate, la classificazione ed infine l'*anomaly detection*.

2.1 Dall'analisi del suono allo spettrogramma

Il suono nasce dalla vibrazione o oscillazione di un corpo sonoro. Queste vibrazioni creano delle onde sonore, cioè variazioni di pressione del mezzo che le propaga, per esempio l'aria. L'onda è definita da tre caratteristiche: l'ampiezza, la frequenza e il timbro.

L'ampiezza (o intensità) dell'onda è associata a quanto il suono è percepito intenso, del volume, ed è misurata in Decibel.

La frequenza (o altezza) identifica il numero di oscillazioni in un secondo, esprime un valore minore o maggiore in base a che il suono risulti più grave o più acuto, determinando così il tono, ed è calcolata in Hertz.

Il timbro, infine, caratterizza un'onda a parità di ampiezza e frequenza, ne definisce quindi il profilo, lo spettro, è composto dalle onde armoniche, ed è in sostanza ciò che ci permette di distinguere due strumenti musicali che stanno eseguendo la stessa nota.

Nel quotidiano utilizzo del mondo digitale, è comune visualizzare un segnale audio come un'onda, senza sapere che questa prospettiva rappresenta graficamente l'andamento dell'ampiezza (sull'asse delle ordinate) in funzione del tempo (l'asse delle ascisse). Questo tracciato è il risultato di un processo, il campionamento, effettuato sul segnale analogico, ovvero sulla forma originale del suono rilevato, che trasforma i campioni, ad intervalli regolari, in segnale digitale. Gli intervalli di campionamento sono come delle istantanee che misurano e registrano digitalmente il valore dell'ampiezza del segnale in precisi istanti di tempo. La struttura digitale non è in grado di cogliere il segnale nella sua forma continua, nella sua reale interezza, ma la registra in forma discreta, ovvero traccia un valore numerico, discreto, a intervalli regolari. Questi sono in grado di descrivere e riprodurre con un certa

Ripeti quello sopra,
io lo integrerò sopra
(me devo ripetere)

MM → scrivere anglo

precisione il suono nel suo formato digitale. Maggiore è la frequenza di campionamento quindi, il numero di campioni analizzati per ogni secondo, maggiore è la qualità del risultato.

Le forme appena analizzate appartengono al *dominio del tempo*, cioè lo spazio che misura la variazione dell'ampiezza rispetto al tempo. Questo punto di vista è molto utile per evidenziare la durata dei suoni, la durata delle pause e la struttura temporale del segnale.

Tuttavia, per poter analizzare nel dettaglio le componenti frequenziali, si deve effettuare un cambio di prospettiva. Si applica la *Trasformata di Fourier*, una funzione in grado di suddividere l'onda complessa nelle sue sottocomponenti sinusoidali, permettendo quindi di visualizzare il segnale nel *dominio delle frequenze*, dove la variazione dell'ampiezza è in rapporto con le frequenze.

occhio di può avere confusione

Se invece di eseguire la trasformata sull'intero segnale, lo si suddivide in blocchi, o finestre temporali, e su ognuna si applica separatamente la funzione, si ottengono più spettri di frequenza, uno per ogni intervallo. Questi spettri, combinati in un'unica rappresentazione, formano lo *spettrogramma*, un grafico che mostra l'andamento delle frequenze in funzione del tempo. Tale prospettiva ci permette di cogliere in combinazione le informazioni temporali e frequenziali. In questa processo è importante la dimensione della finestra temporale e del passo, che definisce quanto si devono sovrapporre le finestre consecutive. Il numero di campioni analizzati per gruppo determina la dimensione della finestra. Maggiore è il numero di campioni considerati, minore sarà il numero di finestre utilizzate nel calcolo dello spettrogramma. Una finestra maggiore determina una migliore risoluzione delle frequenze, ma una peggiore risoluzione temporale. Il passo, invece, viene impostato ad un valore uguale alla metà del numero di campioni utilizzati per la finestra.

Estrazione delle *Potresti mettere 1 figura* *di esempio* *tipico* *tempo*
2.2 Features *Spettro* *Spazio*

Distinguere due oggetti qualsiasi, come una bottiglia e una mela, può sembrare una capacità comune, per nulla speciale. Questa abilità è frutto di un meccanismo che il nostro cervello sviluppa attraverso l'esperienza e la conoscenza. Per ogni oggetto con cui interagiamo, la mente elabora un insieme di caratteristiche in grado di descriverlo e lo esegue con una tale velocità che nemmeno ce ne accorgiamo. Il cervello estrae elementi in grado di definire l'oggetto, come il colore, la lunghezza e la forma, e con ogni senso del corpo. L'oggetto è da intendersi anche come un profumo, un suono, un'immagine o qualsiasi altra percezione.

Al fine di insegnare questa capacità ad una macchina, è necessario identificare ciò che è rilevante, discriminante e misurabile nei dati: le caratteristiche, o *features*. Le *features*

*una frase è il concetto di generalizzazione
"capace di riconoscere anche oggetti
mai visti"*

forniscono le informazioni necessarie per costruire il modello in grado di riconoscere e generalizzare i pattern nei dati.

In questo studio, che tratta di *soundscape*, l'oggetto da analizzare è un segnale audio. *Per conoscere,*

Le feature utilizzate che lo descrivono si basano sui concetti illustrati nel paragrafo precedente.

Si possono distinguere tre gruppi principali di *features*: spettrali (SPE), tonali (TON) e temporali (TEM). []

Le SPE caratterizzano la forma dello spettro e influenzano le percezioni del timbro. Si suddividono in: *specifiche se vengono calcolate sui segmenti time, frequenze o ampiezza*

- *Spectral Centroid*: consiste nella media pesata delle frequenze nel segnale e indica il centroide, ovvero il centro di massa dello spettro. Valori più elevati indicano un suono più brillante [4]. Per brillante si intende che la maggioranza delle armoniche si trova su alte frequenze.
- *Spectral Spread*: misura la dispersione delle frequenze attorno al centroide [5]. Un basso valore indica una concentrazione maggiore delle frequenze attorno al centroide.
- *Spectral Rolloff*: misura la frequenza al di sotto della quale si trova una percentuale specifica dell'energia totale dello spettro. Caratterizza la distribuzione dell'energia delle frequenze [6].
- *Spectral Decrease*: misura quanto l'energia spettrale cala rapidamente all'aumentare delle frequenze. Una curva ripida indica una diminuzione rapida dell'energia spettrale, quindi un blocco ricco di basse frequenze e povero di alte frequenze [7].
- *Spectral Flux*: rileva il numero di cambiamenti nella forma dello spettro. Identifica variazioni rapide e significative nel contenuto del segnale [8].

*una
dico*

Le TON misurano le componenti tonali del segnale rispetto al rumore, ed indicano qualità del suono. Sono composte da:

non è mai definito

- *Spectral Crest Factor*: compara il massimo della magnitudo dello spettro con la sua media quadratica [9]. La magnitudo indica una frequenza molto presente nel segnale.
- *Spectral Flatness*: indica quanto lo spettro è uniforme. Un valore alto suggerisce un segnale con poca struttura tonale, quindi molti rumori.
- *Spectral Tonal Power Ratio*: rapporta l'energia tonale con l'energia totale. Un valore alto indica che l'energia si concentra in componenti tonali, basso sui rumori [10].

*è il
monino
che es
mette*

Infine le TEM, che descrivono come il segnale varia rispetto al tempo:

- *Time Zero Crossing Rate*: identifica il numero di volte in cui il segnale cambia di segno quindi quando ha valore zero. Indica una forte presenza di alta frequenze [11].

un valore alto

- *Time Acf Coeff*: quantifica la correlazione tra il segnale e una versione ritardata dello stesso (funzione di autocorrelazione). Utile per identificare pattern ripetitivi, [12].
- *Time Max Acf*: indica il valore massimo dell'autocorrelazione. Un valore alto può esprimere una forte periodicità del segnale, [13].

2.3 Standardizzazione dei dati

La standardizzazione è un'attività di pre-processamento dei dati in grado di trasformarli in una forma indipendente dalla scala utilizzata. Per scala si intende il metodo utilizzato per confrontare i dati, ossia l'intervallo di valori. Infatti, una certa misurazione può avere un rapporto diverso con gli altri dati a seconda del metodo di confronto. Immaginiamo di avere i risultati di due esami scolastici diversi, fatti da due gruppi di studenti. Il primo gruppo ha ottenuto i risultati in centesimi, un intervallo da 1 a 100, invece il secondo gruppo, in trentesimi, da 1 a 30. Se volessimo confrontare i risultati per capire la media totale dei punteggi, il valore ottenuto sarebbe distorto. Si deve quindi riportare i dati sullo stesso piano di misura, indifferente alla scala utilizzata. Allo stesso modo nella standardizzazione, quando vengono rimossi i riferimenti alla scala e alla locazione, si ottiene una versione standard dei dati, "senza dimensionalità".

Una tecnica molto utilizzata è lo *Z-score standardization*, qui specificata con la formula:

$$x_{ji}^* = \frac{x_{ji} - \bar{x}_j}{\sigma_j}$$

di standardizzazione dell'oggetto i dell'oggetto j

Si definisce x_{ji}^* la j-esima *feature* standardizzata, x_{ji} la j-esima *feature* prima della standardizzazione, \bar{x}_j la media lungo la direzione j ed infine σ_j la deviazione standard che identifica la dispersione dei dati. Si consideri ora una matrice composta sulle righe dagli oggetti in esame e in colonna i valori delle *feature*. Per ogni elemento x in riga j e posizione i viene sottratta la media calcolata in riga j , e il valore ottenuto si divide con la deviazione standard estratta dalla riga j . L'effetto prodotto è caratterizzato da media uguale a 0 e deviazione standard a 1.

dopo la standardizzazione
ogni feature ha

2.4 Classificazione

La classificazione è un ambito molto studiato nella PR e nel ML. Si tratta di un sistema decisionale in grado di assegnare una categoria, o etichetta, ad un oggetto sulla base di un modello costruito mediante le *features*. In sostanza, è una funzione che prende in input un vettore (vector) di *features* e produce un output (output) di classe. I dati di input sono spesso rappresentati come vettori numerici, mentre i dati di output sono generalmente categoriali.