

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

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

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.

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 *soundscape*, 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 *pattern recognition* (PR) e il *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 *soundscape* 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.

Quinn *et alii* identificano i *soundscape* 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 *soundscape*. 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 *soundscapes* 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.

Capitolo 2

Background

L'obiettivo di questo capitolo consiste nel fornire al lettore 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, trattasi 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

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.

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 questo 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.

2.2 Features

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*

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.

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 la percezione del timbro. Si suddividono in:

- *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]

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

- *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]

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 alte frequenze [11].

- *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:

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

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.

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

oggetto, ne elabora le *features* per identificare il *pattern*, e restituisce un valore discreto che determina a quale categoria appartiene.

In generale gli approcci si suddividono in generativi e discriminativi.

Il generativo mira a definire un modello per ogni categoria, o classe, da discriminare nel *dataset*. Questa forma presenta una struttura più flessibile in grado di adattarsi a nuove classi, è più rapida nell'addestramento e ottiene una migliore capacità descrittiva per la singola classe.

L'approccio discriminativo, invece, si basa sulla ricerca del migliore confine decisionale per separare le classi nello spazio. Per sua natura è più rapido nella scelta quindi velocizzando quindi la fase di test. Risulta molto efficace, ed è flessibile nell'eventualità di un modello errato.

Un'ulteriore suddivisione nei classificatori si basa sulla loro natura parametrica o non parametrica. La parametrica si caratterizza per l'assunzione della forma di distribuzione dei dati per ogni classe (es. la distribuzione normale), e si concentra nella stima dei parametri della funzione che genera tale distribuzione. Diversamente, la non parametrica non assume nessuna forma di distribuzione, ma viene stimata direttamente dal *training set*. Risulta più dispendiosa in termini computazionali ma non basandosi su assunzioni determina un modello maggiormente flessibile e adattabile al contesto.

2.4.1 Apprendimento supervisionato

Nella PR spesso si utilizza il paradigma dell'*apprendimento da esempi*. Questo metodo può essere visto come l'apprendimento di un bambino che sperimenta e acquisisce conoscenza da esempi, dall'esperienza. In particolare, nel contesto della classificazione, si utilizza un approccio *supervisionato* in cui la conoscenza viene acquisita tramite dati conosciuti, il *training set*, dotato di categorie, o etichette, note. Conoscendo la reale classe di appartenenza degli oggetti, il modello può addestrarsi e affinare gradualmente la sua capacità decisionale. In questo processo è importante che il sistema non “impari a memoria” i dati, il cosiddetto *overfitting*, un eccessivo adattamento ai dati di addestramento. Per evitare ciò il modello deve essere in grado di generalizzare quindi di classificare correttamente anche oggetti sconosciuti, mai visti. In tal modo dimostra di non essersi specializzato sui dati di addestramento e di aver appreso le caratteristiche essenziali.

2.3.2 Validazione

Una volta costruito il modello è necessario verificare la qualità del classificatore. Per tale scopo si utilizza il *testing set*, un *dataset* che presenta elementi sconosciuti al sistema, quindi diversi anche da quelli usati in addestramento, ma dotato di categorie note da confrontare con il risultato. Viene eseguito il modello su tali dati e si valuta la predizione in base all'errore ottenuto. Questo valore pone in rapporto le previsioni errate rispetto al numero totale di oggetti analizzati. Una previsione errata consiste in una falsa valutazione del classificatore, quindi un valore diverso dalla reale categoria di appartenenza.

Nella costruzione di un classificatore di solito si dispone di un unico *dataset* che si deve suddividere in due parti, il *training set*, per l'addestramento, e il *testing set* per i test e la validazione. Un metodo ideale sarebbe poter usare tutti dati di esempio per il *training* ed estrarre altri esempi dal problema per testare il modello, ma nella realtà potrebbe essere non fattibile o troppo dispendioso. Si preferisce quindi suddividere il *dataset* in due parti uguali, una proporzione 50/50, nella modalità definita come "*train-test split*". Tuttavia, una metodica comune consiste nella *cross validation* che permette di ottenere una valutazione più valida e consistente. Esistono diverse varianti, ognuna con le sue caratteristiche. La forma più semplice è la *holdout*, che distribuisce casualmente i dati in due insiemi di uguale dimensione. Un'alternativa simile è l'*average holdout*, che per essere indipendente dalle partizioni effettua più *holdout* e calcola l'errore come media dei risultati ottenuti in tutti i casi.

Infine, una delle più utilizzate è la *Leave One Out (LOO)*, una variante particolare che ottiene ottimi risultati in termini di affidabilità, e soprattutto con *dataset* ristretti. Come suggerisce il nome, consiste nell'effettuare l'addestramento con tutti gli oggetti del *dataset* meno uno, x_i , che viene invece utilizzato per validare il modello. Si ripete il procedimento per ogni elemento x_i del *dataset*, e al termine si media il risultato ottenuto. Presenta un costo computazionale maggiore ma garantisce indipendenza dalla partizione e dai dati scelti del *training set* e *testing set*.

2.5 Anomaly Detection

L'*anomaly detection (AD)* consiste nell'identificare fenomeni ed eventi, che presentano un comportamento anomalo rispetto al resto dei *dataset*. Tali fenomeni, denominati *outlier*, si confondono tra gli *inlier*, il resto dei dati normali, per la loro caratteristica natura e la scarsa numerosità.

2.5.1 Gli outlier

Individuare gli *outlier*, pur essendo complicato, è una necessità. Ciò consentirebbe di inferire informazioni preziose oppure rappresentare una prevenzione per eventuali criticità. Contribuisce in modo significativo anche nel *data cleaning*, la “pulizia dei dati”, essenziale per la preparazione dei dati per l’analisi e altre svariate applicazioni. Si tratta di un insieme di processi che servono a rimuovere duplicati, uniformare e filtrare i dati. In questo contesto rimuovere gli *outlier* semplificherebbe le fasi successive di analisi migliorando la qualità del risultato.

Si distinguono varie tipologie di *outlier*:

- i *point*, che sono identificati tali indipendentemente se si trovano da soli o in gruppo;
- i *collective*, considerati o. solo se rilevati in gruppo, altrimenti rientrano negli *inlier*;
- i *contextual*, dove il contesto determina la loro forma, normale o anomala;

2.5.2 Le applicazioni

L’AD ricopre un ruolo importante in molteplici campi. Per esempio, possiamo citare:

- le intrusioni di rete, dove l’attaccante deve mantenere un basso profilo per non essere individuato all’interno del sistema;
- l’ambito sanitario, dove rilevare un segnale anomalo nelle diagnosi potrebbe indicare malattie gravi, come il cancro;
- i sistemi automatizzati, in cui prevenire un avaria o, riuscire a intervenire in tempo nel caso si manifestasse, è essenziale;
- nel processamento di immagini o testi, come rilevare *fake news*;
- nel rilevamento di frodi, all’interno delle innumerevoli transazioni bancarie prodotte ogni giorno.

Queste casistiche sono accomunate da un enorme quantità di dati dove sistemi troppo rigidi e specifici non potrebbero adattarsi al continuo mutamento delle variabili in gioco. Le problematiche nell’AD consistono non solo in dataset di grandi dimensioni, ma anche dall’alto costo computazionale derivato per elaborarli. Inoltre, un’eccessiva numerosità di *features* potrebbe portare alla *curse of dimensionality*, un fenomeno dove un’alta dimensionalità comporta una drastica riduzione delle prestazioni.

2.5.3 I metodi

In generale la suddivisione si basa sul metodo utilizzato per rilevare anomalie. I vari metodi possono essere:

- Metodi basati sulla statistica: essi ricercano elementi che non rispecchiano la distribuzione dei dati. Una bassa probabilità di appartenenza alla forma di distribuzione determina un'alta probabilità di essere un outlier.
- Metodi basati sul *clustering*: essi suddividono i dati per somiglianza in gruppi, i *cluster*, e le anomalie risultano evidenti poiché troppo diverse dal loro gruppo di appartenenza. Oppure gli outlier definiscono gruppi malformati con altre anomalie.
- Metodi basati sull'apprendimento: vengono utilizzati algoritmi di apprendimento automatico specializzati a modellare il comportamento normale e individuare le anomalie.
- Metodi basati sulla distanza o sulla densità: viene considerata la distanza tra gli elementi, o la densità locale. Nel primo caso gli *outlier* si troveranno distanti dagli altri punti, nel secondo saranno in zone a bassa densità.
- Metodi basati su insiemi, o *ensemble*: queste tipologie combinano i risultati di metodi diversi, o gli stessi con parametri differenti, per ottenere una previsione più accurata.

La maggioranza degli algoritmi rappresenta il proprio risultato mediante un valore, denominato *anomaly score*, che quantifica quanto un elemento è probabile che sia un anomalia.

In letteratura esistono differenti algoritmi di AD, ma di seguito ne saranno descritti solo tre, quelli utilizzati nello studio. I metodi sono L'*IForest* (IF), o *Isolation forest*, il *Local Outlier Factor* (LOF) e l'*Ocsvm* (OCSVM), o *One Class Support Vector Machine*.

IF è un algoritmo basato su *ensemble*, e si sviluppa su alberi decisionali. Per albero decisionale si intende una struttura ad albero dove ogni nodo rappresenta una decisione e i rami le possibili alternative. Procedendo dall'alto verso il basso del modello, i dati vengono suddivisi nelle varie decisioni fino alle foglie. In IF viene costruita una foresta di alberi decisionali attraverso una selezione casuale dei dati. Ciascun albero cerca di isolare i dati mediante suddivisione, e le anomalie si troveranno più in alto poiché avranno bisogno di meno divisioni rispetto ai dati normali. In sostanza, più risulta facile isolare un oggetto e con maggiore probabilità sarà un anomalia. Questo algoritmo risulta molto scalabile, veloce e con una bassa tendenza all'*overfitting*.

LOF è un metodo basato sulla densità. La densità di un oggetto dipende dal suo vicinato, ovvero dalla numerosità degli elementi che gli sono vicini. Il confine per definire vicino un oggetto dipende dalla metrica scelta nell'implementazione. La più comune è la distanza euclidea, che misura la lunghezza del segmento tracciato tra due punti. In sostanza, il funzionamento definisce che minore è la densità locale più alta è la probabilità che in un determinato oggetto potrebbe esserci un'anomalia. Questo metodo è efficace con dati ad alta dimensionalità e nel rilevamento di anomalie, ma al contrario, risulta molto dispendioso in termini computazionali.

Infine, OCSVM si basa sull'apprendimento. Si tratta di una variante delle *Support Vector Machine*, un approccio discriminativo applicato solitamente a problemi binari. In questa forma, a singola classe, l'algoritmo viene addestrato per definire un confine che racchiude al suo interno solo i dati normali. Così facendo i dati anomali, che risultano esterni alla distribuzione, emergono e sono quindi rilevabili. Spesso viene utilizzato il trucco del *kernel*, che consiste nel proiettare i dati in una dimensione superiore, dove può risultare più facile separare i dati. Questo metodo presenta un buon adattamento anche a dati non lineari, grazie alla funzione *kernel*. In svantaggio si presenta meno scalabile per *dataset* molto grandi.

Capitolo 3

Dataset

Lo studio condotto si è basato sui dati raccolti nell'articolo della rivista scientifica “*Biota Colombiana*” pubblicata dall'*Instituto de Investigación de Recursos Biológicos Alexander von Humboldt*, che ha sviluppato un'analisi di monitoraggio acustico passivo nella *Riserva Naturale Los Yátaros*, nel dipartimento di *Boyacá* in *Colombia* [2]. Con il termine passivo si identifica una modalità di osservazione del paesaggio incentrata sulla registrazione di un particolare luogo e solo successivamente prevede un'analisi approfondita, diversamente da quella attiva, dove si osserva e si analizza il fenomeno in tempo reale.

La riserva è composta da querceti e foresta subandina in diversi stadi di rigenerazione naturale, e presenta una biodiversità acustica molto particolare. Il progetto mirava a profilare l'impronta acustica della riserva campionando suoni nello spettro udibile e negli ultrasuoni. Sono stati predisposti tre siti, denominati *YAT*, lungo il sentiero principale, distanti 150 m, con due sensori acustici *AudioMoth* ciascuno, per le due forme di suono desiderate, posti ad altezze diverse. Il periodo di campionamento si è svolto dall'1 marzo al 2 maggio 2020, registrando un minuto di audio ad intervalli di trenta minuti, quindi 48 giornalieri, e per un totale di 9055 audio nell'udibile e 3392 nell'ultrasuono. Nel contesto applicativo, sono stati considerati solo i dati nello spettro udibile, per permettere l'ascolto del contenuto.

3.1 Dataset prima fase

Nella prima fase dello studio è stato utilizzato il *dataset* completo (*DATA1*) che presenta l'insieme originale dei dati. Il gruppo si presenta con una suddivisione per i tre siti (*YAT1*, *YAT2*, *YAT3*) con quantità leggermente differenti. Gli audio sono 3018 per *YAT*, a parte il primo con 3019. Ogni sito presenta 1482 file per il mese di marzo (1483 solo per *YAT1*), 1440 per il mese di aprile e 96 per il mese di maggio.

Data l'ingente quantità di dati disponibile, non si è potuto analizzare il contenuto, ossia ascoltare l'intero insieme di registrazioni. In un primo sondaggio, analizzando diversi audio in momenti diversi della giornata e del mese, si è esplicitato che tutti e tre i luoghi risultano molto caratterizzati dal suono del fiume e della cascata vicina. Anche se tale suono fosse stato ad una distanza maggiore avrebbe sortito lo stesso effetto. Farina *et alii* sostengono che “la geofonia può essere rilevata anche a grandi distanze in base all'ampiezza della sorgente

sonora” [2]. A questo si aggiungono ulteriori problematiche dovute a periodi piovosi, il cui rumore sovrasta in diverse occasioni i suoni ambientali naturali. Entrambi gli elementi appena descritti determinano un ambiente umido che potrebbe influire anche sulla capacità del sensore.

3.2 Dataset seconda e terza fase

Il *dataset* della seconda e terza fase dello studio (DATA2) consiste in un sottoinsieme del primo appena analizzato. Per potere inferire maggiori informazioni dal contesto si è stabilito che era necessaria una descrizione più accurata del contenuto. Quindi si è ristretto l'insieme ad un campione di dati minore che potesse essere ascoltato e studiato nel dettaglio. L'analisi ha ridefinito il *dataset* originale in 186 audio, focalizzandosi sul sito YAT1, nel mese di marzo con finestre temporali a intervalli di due ore, partendo dalle due del mattino, quindi nelle ore: 02:00, 06.00, 10:00, 14:00, 18:00, 22:00. Con questa modalità si poteva ottenere una visione abbastanza generale della varietà sonora presente nella giornata, includendo i due momenti fondamentali alba e tramonto, caratterizzati da picchi di attività acustica.

L'interpretazione manuale del *dataset* ha classificato il contenuto assegnando delle etichette ai vari elementi distinti. In relazione all'ANT è stato individuato un unico suono, appartenente al rumore di veicoli (classe V). Nella BIO è stato individuato il verso degli uccelli e dei grilli (classi U e G). Per la GEO si è rilevato il precedentemente menzionato rumore del fiume/cascata, la pioggia e i tuoni (classi C, P e T). Infine, sono stati identificati i rumori relativi alle interferenze del sensore (classe I), ed eventuali elementi uditi ma non interpretati, purtroppo sconosciuti (classe S). Rispetto a quanto specificato nell'introduzione, all'interno della GEO si è integrato anche l'insieme dei suoni relativi alla quiete. Tale scelta è derivata da una maggiore semplicità nella trattazione, ma specialmente per l'impossibilità nel poterli classificare correttamente.

Nel grafico 3.1 è possibile visualizzare la distribuzione degli elementi descritti nelle fasce analizzate. Da una prima osservazione risulta evidente la significativa presenza dell'elemento C, come già esplicitato nel paragrafo precedente. Lo accompagna il suono dell'elemento U, risultato meno attivo solo nella parte centrale della giornata. Il resto è relativamente distribuito, a parte I e S diffusi con bassa intensità. Si può notare come S sia presente solo nelle due fasce pomeridiane. Dal grafico 3.2, possiamo avere una visione del mese di marzo, per il quale è lecito esplicitare le medesime considerazioni annotate.

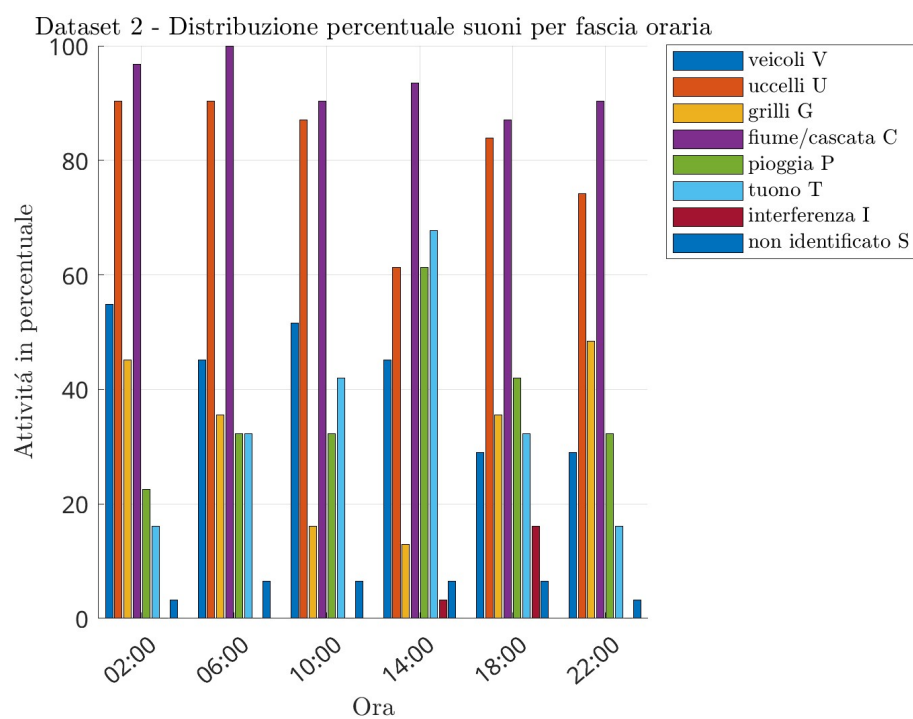


Grafico 3.1

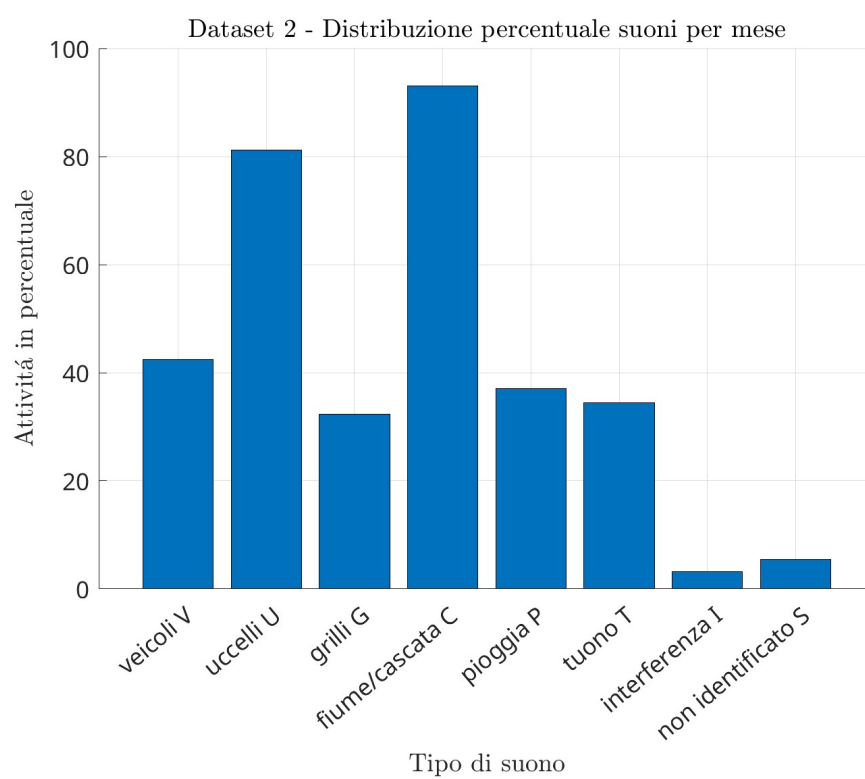


Grafico 3.2

Capitolo 4

Classificazione

In questo capitolo si desidera analizzare le prime due fasi dello studio basate sulla classificazione: la prima riguardante la rappresentazione su categorie note, la seconda esplorativa su categorie semantiche. Innanzitutto saranno definiti i gruppi di *features* utilizzati negli esperimenti, descrivendo poi la tipologia di classificatore scelto. A seguire, saranno illustrati nel dettaglio i problemi di classificazione disegnati, ed infine si andrà ad analizzare i risultati ottenuti nei due esperimenti.

L'obiettivo è determinare quale finestra risulta più funzionale, quali *features* sono più efficaci al nostro contesto, e in quali problemi ottengono i risultati migliori. Questo studio considera gli esperimenti di classificazione precedentemente condotti dai colleghi Ilaria Ballerini e Andrea Piazza.

Inoltre, è stata valutata una fase di filtraggio per alcune frequenze, per capire se potesse rappresentare un miglioramento alla qualità dei dati in *input*. L'idea consiste nel rimuovere alcuni elementi rumorosi preponderanti su ogni audio, dovuti in maggior parte alle problematiche intrinseche del contesto, ma anche alla sensibilità dello strumento di misurazione.

4.1 Configurazioni features

Prima di poter costruire un modello di classificazione in grado di spiegare il contesto esaminato occorre identificare le caratteristiche necessarie che possano descrivere gli oggetti studiati. Nello studio sono state considerate sia le *features* nella loro forma originale, come descritta nel capitolo 2.2, sia raggruppate in insiemi. Gli elementi di gruppi sono stati derivati attraverso concatenazioni delle caratteristiche originali o delle loro medie matematiche. Si premette che quelle identificate come originali in realtà sono descritte da un numero definito di componenti che per il momento verranno indicate con la variabile N , che ne indica la numerosità: maggiori dettagli sulla numerosità di N saranno forniti alla fine di questo paragrafo. Ogni *feature* originale è stata descritta da un vettore riga di $1 \times N$ componenti.

In particolare sono stati identificati 19 gruppi di *features*:

- 11 ORIGINALI (o ORIG), per ognuna delle 11 *features*, utilizzate singolarmente, $1 \times N$ elementi ciascuna;

- CONCATENAZIONE ORIGINALI (o CONC.ORIG.), formata dalla concatenazione orizzontale delle 11 *features* originali, ottenendo un totale di $11 \times N$ componenti;
- CONCATENAZIONE SPETTRALI (o CONC.SPE.), ottenuta dalla concatenazione delle componenti delle 5 *features* spettrali, quindi $5 \times N$ componenti;
- CONCATENAZIONE TONALI (o CONC.TON.), come la precedente ma considerando le 3 *features* della tonalità, ovvero $3 \times N$ componenti;
- CONCATENAZIONE TEMPORALI (o CONC.TEM.), come la precedente ma con le 3 *features* temporali, $3 \times N$ componenti.
- CONCATENAZIONE MEDIE ORIGINALI (o CONC.MED.ORIG.), formata dalla concatenazione orizzontale delle medie delle 11 *features* originali, quindi 11 componenti;
- CONCATENAZIONE MEDIE SPETTRALI (o CONC.MED.SPE.), ottenuta dalla concatenazione orizzontale delle medie delle componenti delle 5 *features* spettrali, quindi solo 5 componenti;
- CONCATENAZIONE MEDIE TONALI (o CONC.MED.TON.), come la precedente ma utilizzando le 3 *features* tonali, quindi solo 3 componenti;
- CONCATENAZIONE MEDIE TEMPORALI (o CONC.MED.TEM.), come la precedente ma utilizzando le 3 *features* temporali, in totale 3 componenti;

Ai 19 gruppi sopracitati si è tenuto conto anche della relativa versione standardizzata, ovvero ottenuta dalle componenti processate con la tecnica di standardizzazione *Z Score*, definendo quindi 38 gruppi: 19 puri e 19 standardizzati. In questo modo, si dispone anche di una rappresentazione con una scala comune indipendente dalle misurazioni originali.

In aggiunta, i dati sono stati estratti in due forme diverse (ottenendo quindi 78 gruppi di *features*) basate su diverse finestre temporali scelte per il campionamento nel calcolo dello spettrogramma: la prima si basa su un tempo minore di 1 secondo, che indicheremo con FS0X, a 32768 campioni; la seconda, FS1, invece si basa su un tempo uguale a 1 secondo, ovvero a 48000 campioni.

La precedente variabile definita come N è stata introdotta per indicare il numero di componenti per ogni *feature*. Per poterla definire era necessario introdurre il concetto relativo alle due finestre di intervallo misurate. Nel caso di FS1 sono state estratte 120 componenti ($N=120$): il segnale audio analizzato presenta una lunghezza temporale di 60 secondi e, per costruire lo spettrogramma lo si analizza con un intervallo di 1 secondo alla volta, ottenendo 60 finestre. Inoltre, considerando un passo pari alla metà dell'intervallo, 0.5 secondi, si

ottengono altre 60 finestre, per un totale di 120 campionamenti. Alla stessa modo, è stato fatto per la forma *FS0X* dove, considerando una finestra più breve, si è ottenuto un maggior numero di componenti ($N=176$).

La tabella 4.1 riassume i vari gruppi di *features* considerati (il simbolo # indica la numerosità).

GRUPPI DI FEATURES	DESCRIZIONE FEATURES	# COMPONENTI PER FEATURE	# N x FEATURE CASO FS0X	# N x FEATURE CASO FS1
CONC.ORIG.	Concatenazione originali	11	1320	1936
CONC.SPE.	Concatenazione spettrali	5	600	880
CON.TON.	Concatenazione tonali	3	360	528
CONC.TEM.	Concatenazione temporali	3	360	528
CONC.MED.ORIG.	Concatenazione medie originali	11	11	11
CONC.MED.SPE.	Concatenazione medie spettrali	5	5	5
CONC.MED.TON.	Concatenazione medie tonali	3	3	3
CONC.MED.TEM.	Concatenazione medie temporali	3	3	3
ORIG	Spectral Centroid	1	120	176
ORIG	Spectral Crest Factor	1	120	176
ORIG	Spectral Decrease	1	120	176
ORIG	Spectral Flatness	1	120	176
ORIG	Spectral Flux	1	120	176
ORIG	Spectral Roll off	1	120	176
ORIG	Spectral Spread	1	120	176
ORIG	Spectral Tonal Power Ratio	1	120	176
ORIG	Time Zero Crossing Rate	1	120	176
ORIG	Time Acf Coeff	1	120	176
ORIG	Time Max Acf	1	120	176

Tabella 4.1

4.2 Dettagli classificatore

In questo studio è stato utilizzato il classificatore *K Nearest Neighbor (KNN)*, un approccio supervisionato generativo non parametrico, semplice e intuitivo: il metodo si basa sul classificare un punto, un oggetto, assegnandogli la classe che più frequentemente ritroviamo tra i k oggetti più vicini. Il concetto di vicinanza è relativo alla forma scelta di

specializzazione dell'algoritmo: nel nostro caso la distanza euclidea, una delle misure più utilizzate. Come risulta chiaro, la scelta del valore di k è cruciale. E' stato utilizzato il valore più semplice, con k uguale a 1, facile da implementare e da comprendere, in grado di catturare dettagli molto fini nel *dataset* poiché nella decisione della classe si basa unicamente sull'elemento più vicino (comunemente viene indicata con solo 1-NN, o solo NN). A suo svantaggio, un valore troppo piccolo, come nel nostro caso, lo rende molto sensibile al rumore, determinando risultati errati o incongruenze, influenzando l'accuratezza del modello. Per ovviare al problema, è stata integrata una validazione incrociata *Leave One Out* (LOO), per migliorare la robustezza e l'affidabilità del risultato. Per semplicità l'insieme dei due metodi sarà indicato con *LOO KNN*.

4.3 Problemi di classificazione disegnati

Le due fasi di classificazione di questo studio affrontano problemi diversi che illustreremo nei prossimi paragrafi, ma in sostanza si differenziano dalla tipologia di etichette scelte: la prima rappresenta i dati utilizzando delle etichette note, la seconda invece delle etichette semantiche.

4.3.1 Rappresentazione con etichette note

La prima fase propone uno studio sul *dataset* DATA1. Essendo privo di annotazione, procedere già dall'inizio con la classificazione supervisionata non era possibile; inoltre optare per un'etichettatura manuale, identificando i vari suoni all'interno, non era considerabile, sia per l'eccessivo tempo necessario che per la mancanza di risorse. Tuttavia, come specificato nell'introduzione, i limiti più ostici consistono da una parte nella difficoltà oggettiva intrinseca di discriminare elementi all'interno di un *soundscape* e, dall'altra, nella competenza tecnica necessaria a identificare la biodiversità presente. Per questi motivi, si è deciso di proporre dei problemi affrontabili basati su etichette note, cioè su informazioni deducibili dal contesto dell'oggetto, invece che dal suo contenuto: si è tenuto conto del luogo di registrazione, e della temporalità, come l'ora del giorno, o una fase della giornata, o del mese.

Sono stati individuati i seguenti problemi:

- PR-1.1 YAT. Riguarda il luogo di registrazione, gli *yat*. Si tratta di una classificazione multiclasse, nel nostro caso 3, relative alle 3 zone in cui sono stati collocati i microfoni. La cardinalità delle classi vede un 33% di presenza per ognuno.

- PR-1.2 G/N. Un problema binario che consiste nel discriminare le due fasi astronomiche principali del giorno, classe G e della notte, classe N. La fascia oraria di riferimento del giorno si presenta tra le ore 6 e le 17, estremi compresi. Le ore restanti sono assegnate alla rispettiva classe che identifica la notte. Le classi presentano un rapporto tra le quantità di circa il 50%.
- PR-1.3 AT/R. Un tipo di classificazione binaria che considera in un classe i suoni presenti all'alba e al tramonto, classe AT, e nella seconda le ore rimanenti, classe R. La fascia oraria considerata per alba è stata identificata tra le ore 5 e le ore 7, estremi compresi, per il tramonto invece tra le ore 18 e le 20, estremi compresi. L'ipotesi alla base di questo problema riteneva che nella prima classe potessero esserci dei suoni caratterizzanti e simili rispetto al resto della giornata, riconoscibili nell'alba come il risveglio della natura e nel tramonto come il calare della quiete. Quinn *et alii* evidenziano l'attività della biofonia in tali fasce della giornata [1]. In questa casistica la distribuzione si trova in parte sbilanciata sulla classe R, presente per un 70%.
- PR-1.4 A/T/G/N. Si tratta di una classificazione multiclasse, composta da 4 classi, discriminando alba, classe A, tramonto, classe T, e le ore rimanenti diurne e notturne, classi G e N. Le fasce orarie per l'alba e il tramonto sono le medesime presentate sopra. Vediamo le classi A e T in una percentuale di distribuzione minore, con circa un 15% cadauna, rispetto alle classi G e N, che presentano un valore di 40% e 30%.
- PR-1.5 M. Problema a 3 classi, che distingue i relativi mesi in cui sono stati registrati i dati. Per il terzo mese, maggio, è presente un forte squilibrio dovuto alla mancanza di dati, che si identifica con solo un 4% dei dati, rispetto a marzo e aprile con il 49% e 47%.
- PR-1.6 MM. Classificazione binaria che discrimina i mesi in due parti, la prima da inizio mese fino al quindicesimo giorno, la seconda dal sedicesimo giorno fino a fine mese. La distribuzione è ottimamente bilanciata con circa il 50% per entrambe le classi.

In alcune casistiche i dati presentano parti sbilanciate che potrebbero inficiare sulla qualità del risultato: riguarda maggiormente il problema PR1.5 che soffre di una componente poco rappresentata, il mese di maggio. Pur essendo a conoscenza del problema, non si è potuto prendere in considerazione delle tecniche conosciute in letteratura per compensare il divario. Si sottolinea tuttavia, che in tale situazione i risultati ottenuti presentano comunque una valenza considerevole.

4.3.2 Rappresentazione con etichette semantiche

La seconda fase utilizza il *dataset* DATA2, sottoinsieme del *dataset* principale. Come precedentemente descritto nel capitolo 3, riducendo il set a un numero di dati censibile si è potuto procedere con una rilevazione dei suoni principali classificando i relativi gruppi. Con una conoscenza più specifica dell'oggetto si è potuto caratterizzare lo studio mediante delle rappresentazioni basate su etichette semantiche, considerando quindi il contenuto, invece che il contesto (come nella prima fase).

Sono stati disegnati tre gruppi di problemi di classificazione, suddivisi in base alle classi di cui sono composti:

- Gruppo PR-2.1. Consiste in una classificazione binaria positivo/negativo, ovvero il sistema distingue la presenza o l'assenza di una determinata classe, mediante le etichette con i valori 1 e 0. In questo gruppo sono stati descritti quattro problemi, con il relativo rapporto percentuale di distribuzione:
 - PR-2.1.1 V. Riguarda la classe V del veicolo, in rapporto 42/58;
 - PR-2.1.2 G. Identifica il suono G dei grilli, con 32/68;
 - PR-2.1.3 P. Identifica la classe P che caratterizza la pioggia, con un bilanciamento di 37/63;
 - PR-2.1.4 T. Riguarda la classe T che definisce i tuoni, in rapporto 34/66;
- Gruppo PR-2.2. Concerne una classificazione binaria tra due classi distinte, con etichette 1 e 2. Nel seguente gruppo sono stati disegnati i problemi:
 - PR-2.2.1 V/P. Vengono messe in corrispondenza ANT/GEO, con le due classi V e P, in rapporto 55/45;
 - PR-2.2.2 V/G. Si confronta ANT/BIO, con le classi V e G, in rapporto 60/40.
 - PR-2.2.3 G/T. Si relaziona BIO/GEO, con le classi G e T, in rapporto 48/52.
 - PR-2.2.4 G/P. In una seconda contrapposizione BIO/GEO, con le classi G e P, in rapporto 46/54.
- Gruppo PR-2.3. Si tratta di una classificazione ternaria, coinvolge quindi tre classi distinte, identificate da tre etichette, 1, 2 e 3. Verranno descritti i seguenti problemi che mettono in relazione ANT/BIO/GEO:
 - PR-2.3.1 V/G/P. Relazionando le classi V, G e P, in un rapporto di 35/30/35;
 - PR-2.3.2 V/G/T. Relazionando le classi V, G e T, in un rapporto di 32/30/38;

Si sottolinea che la circostanza del problema è multi-etichetta, ovvero che ogni audio è caratterizzato da più suoni, quindi più etichette assegnate allo stesso oggetto. Pertanto, per

proseguire con la stessa metodologia applicata nella prima fase e poter disegnare dei problemi di classificazione, per i gruppi PR-2 e PR-3, vi era la necessità di assegnare ad ogni classe solo gli oggetti che la presentassero univocamente all'interno del problema: per esempio in una classificazione binaria, in cui vi sono due classi distinte, non è possibile accettare nel *dataset* un audio con entrambe le etichette, poiché sarebbe assegnabile ad entrambi. Dunque sono state estratte le combinazioni delle classi dove vi era una distribuzione accettabile dei dati e derivare così una casistica interessante da sottoporre. Seguendo questa considerazione, alcune etichette sono state escluse da tutti i problemi a causa di una distribuzione non accettabile, poiché si sarebbe creato uno sbilanciamento in grado di falsare il risultato:

- le etichette C e U, relative al suono del fiume/cascata e degli uccelli, si trovano in una percentuale molto alta, di circa il 94% e l'81%;
- le etichette I e S, relative invece alle interferenze e ai suoni non riconosciuti, diversamente, presentano un valore troppo poco rappresentativo, del 3% e 5%.

4.4 Valutazione del filtraggio di frequenze

In questo paragrafo sarà illustrata la valutazione preliminare effettuata per esplorare l'effetto di un possibile filtraggio delle frequenze dello spettrogramma sul contesto dello studio. Il filtro consiste nell'esclusione delle componenti superiori o inferiori a determinate soglie di frequenza dallo spettrogramma, rendendo più chiare e definite le rimanenti. Nel dettaglio, l'approccio applica filtri passa-basso e passa-alto, che rispettivamente accettano solo determinate frequenze sotto o sopra un limite definito. Applicandoli poi in combinazione si ottiene un filtro passa-banda, che lascia passare delle frequenze intermedie, filtrando in sincronia le superiori e inferiori. Con questi metodi, si dovrebbero eliminare elementi che potrebbero essere causa di rumore, o che, più semplicemente, risultano irrilevanti per lo studio. Tale effetto determina un miglioramento della qualità dei dati, e di conseguenza, perfeziona la capacità discriminativa del classificatore.

Come metodo di valutazione è stato impiegato lo stesso processo utilizzato per la classificazione, quindi valutando la percentuale di errore del modello, alimentato però con i dati filtrati, studiando poi il comportamento di tutte le features nelle due finestre considerate. Lo studio è stato condotto mediante etichette note del caso *PR-1.1 YAT*, ma con un *dataset* più ristretto: si è considerato il mese di marzo, con sei audio al giorno ogni quattro ore, a partire dalle due del mattino, per un totale di 556 file. Si è proceduto in maniera graduale applicando varie combinazioni di filtri, aumentando progressivamente prima dalla parte superiore, viceversa poi dalla parte inferiore, proseguendo in convergenza.

L'esito verrà poi esposto nel capitolo dei risultati, tuttavia si accenna che non è stato osservato nessun miglioramento. Pertanto, la classificazione è stata poi condotta senza l'utilizzo di filtri.

4.5 Risultati

Prima di esporre e analizzare i risultati ottenuti nei problemi di classificazione, si discuterà l'esito del tentativo di filtraggio delle frequenze condotto. Sebbene non abbia portato un effetto applicabile alle casistiche sottoposte, è possibile comunque trarre delle conclusioni interessanti.

4.5.1 Filtraggio delle frequenze

Come descritto precedentemente, sono state testate varie combinazioni di filtri utilizzando come metodo di valutazione il risultato della classificazione stessa. Data la quantità di risultati ottenuti si è proceduto sintetizzando con l'esposizione della media matematica su tutte le *features*, come descritto nel paragrafo 4.1 della classificazione.

Nei grafici 4.1 e 4.2 è possibile vedere il risultato del filtraggio. Sull'asse delle ascisse si presenta il valore in frequenza utilizzato per filtrare le due zone dello spettrogramma, nelle varie combinazioni, con un ordinamento diverso tra i due grafici: il primo per filtro inferiore ascendente, il secondo per il superiore. Sull'asse delle ordinate è esposta la percentuale di errore del classificatore. Per potere evidenziare la differenza rispetto alla versione priva di filtraggio, nel grafico si è aggiunta la voce corrisponde 0-0.

Come si evince la differenza tra il caso non filtrato e quelli filtrati è notevole, e tale contrasto denota un peggioramento significativo del risultato all'aumentare dei filtri impiegati.

La frequenza che ottiene i risultati migliori, indifferentemente dal filtro, è la *FS1.STD*: dal secondo grafico è possibile notare la combinazione di filtri migliore del test derivata dall'utilizzo del solo filtro inferiore nei valori 5, 10 e 15; all'opposto invece la frequenza *FS0X.STD* ottiene i valori migliori sempre con gli stessi valori, ma attivi sul filtro superiore (visibile nel grafico 1). La versione *FS0X.NOR* ottiene i risultati peggiori, distaccandosi di più di un centesimo dalle versioni *FS0X.STD* e la *FS1.NOR*.

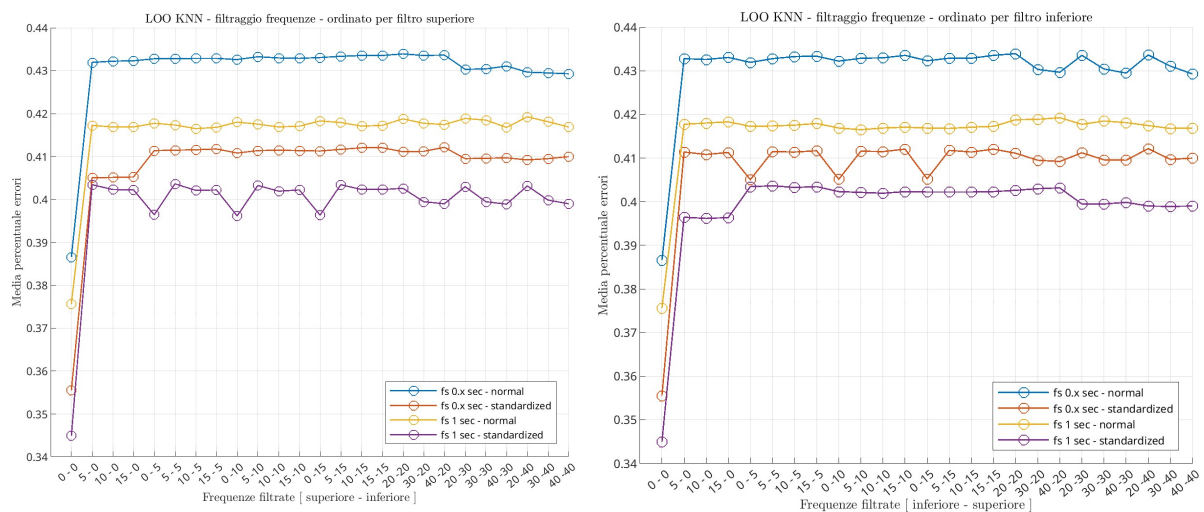


Grafico 4.1 e 4.2

4.5.2 Risultati rappresentazione con etichette note

In questa sezione si riportano i risultati della prima fase di classificazione, basata su rappresentazione tramite categorie note. Per questo studio è giusto tenere conto che il risultato è molto sensibile al fattore considerato nel disegno del problema ovvero il luogo o il tempo.

Nel grafico 4.3 si può visualizzare l'esito dei casi sottoposti: sull'asse delle ordinate troviamo l'errore medio ottenuto dal classificatore, su quello delle ascisse il tipo di problema. Il miglior risultato lo si individua nel caso *PR-1.1 YAT*. Tale caso è l'unico tra quelli con categorie note che tratta della discriminazione del luogo. Sicuramente ha goduto, oltre alle peculiarità della zona in sé, persino delle caratteristiche dello stato dello strumento di misurazione. Si può ipotizzare che elementi come le interferenze o le condizioni meteorologiche, colpendo indifferentemente i vari sensori, abbiano marchiato il prodotto segnando maggiormente uno tra questi rispetto agli altri per un determinato periodo temporale. Quindi hanno caratterizzato una parte del registrato e contribuito ad accrescere le già notabili differenze, definite dall'ambiente circostante, con il resto dei dispositivi. Non di meno, un altro dettaglio considerabile riguarda il suono rappresentante dell'elemento C, che è quasi onnipresente, poiché lo si ritrova nella totalità delle registrazioni, ma in versioni eterogenee tra i dispositivi, dato che sono posizionati a distanze differenti dalla sorgente del rumore. E' possibile affermare che una discriminazione basata sul luogo ha sicuramente beneficiato di queste condizioni rispetto alle altre casistiche.

D'altra parte, anche le tipologie sviluppate sulla temporalità hanno ottenuto complessivamente dei buoni risultati. L'idea alla base dei casi PR-1.2 G/N e PR-1.3 AT/R di valorizzare i periodi astronomici dell'alba e del tramonto, ha portato delle conclusioni interessanti, in particolare nel secondo, che si basava proprio sul discriminare questi due momenti principali rispetto al resto della giornata. Pur sapendo che in un contesto naturale presentano delle caratteristiche sonore uniche, si rende noto, che, diversamente dalle aspettative, non è stato così efficace per il problema PR-1.4 A/T/G/N, risultato con la maggiore percentuale di errori. A sua discolpa si può ipotizzare che il numero delle classi, quattro in questa situazione, superiore rispetto agli altri, ha sicuramente influito sulla prestazione del modello. Questo può determinare una notevole complessità quindi una minore capacità di generalizzazione. Inoltre può generare un aumento della variabilità all'interno dei dati con conseguente difficoltà per il modello nel discriminare le classi. Infine, è possibile sostenere che il risultato sia stato determinato anche dalla riduzione dei dati di addestramento da cui estrarre un modello identificativo della classe stessa: avendo un numero finito di campioni all'aumentare della classi diminuisce il numero di oggetti per descriverle.

Di pari importanza, la capacità delle finestre considerate. Si evidenzia dal grafico, mediante i colori delle barre, come la FS1.STD abbia ottenuto le migliori prestazioni in ogni sfida. Si può ipotizzare che la finestra più ampia riesca a cogliere maggiori dettagli di interesse nel contesto. Una finestra con maggiore risoluzione frequenziale sembra risultare vantaggiosa per lo studio. A tale risoluzione anche il rumore potrebbe risultare meno influente, non facendo quindi risaltare determinati elementi irrilevanti per la classificazione. Allo stesso modo, anche l'altra versione standardizzata FS0X.STD ha riportato un esito accettabile e abbastanza simile alla finestra appena descritta: entrambe le versioni standardizzate hanno conseguito un errore minore rispetto alle controparti regolari. Si considera però che, la finestra più corta, ma con dati standardizzati, FS0X.STD, nella maggioranza dei casi ha ottenuto dei risultati migliori della finestra più ampia ma con dati normali, FS1S.NOR. Quindi si può dedurre che per il contesto la scelta corretta sia una combinazione tra la finestra più ampia e dati standardizzati.

La medesima situazione è osservabile sotto un'altra prospettiva offerta dal grafico 4.4, dal punto di vista delle *features*, ottenuta con le medie dei valori trasversalmente alle casistiche. Si osserva che nei gruppi con le *features* singole ORIG le capacità per tipologia di finestra si discostano di poco, diversamente che nelle forme combinate dove le versioni standardizzate sono maggiormente incisive. Lo si può notare nel gruppo CONC.MED.ORIG., che ottiene i migliori risultati. Riassume le caratteristiche essenziali dei tre gruppi SPE/TON/TEM,

servendosi di un numero ristretto di *features*, undici in questo caso. Un insieme di componenti ridotto, riduce la complessità del modello, rendendolo più robusto, aumentando la sua interpretabilità e ottenendo una migliore generalizzazione.

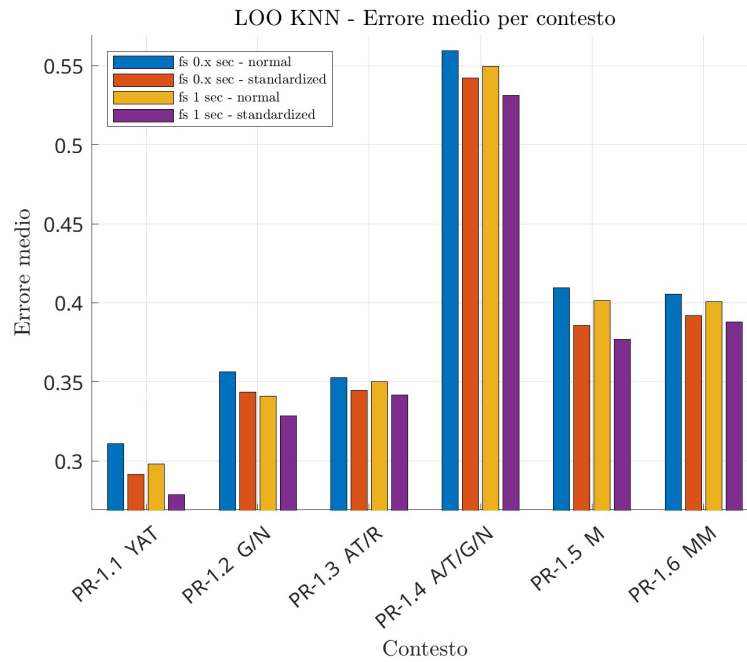


Grafico 4.3

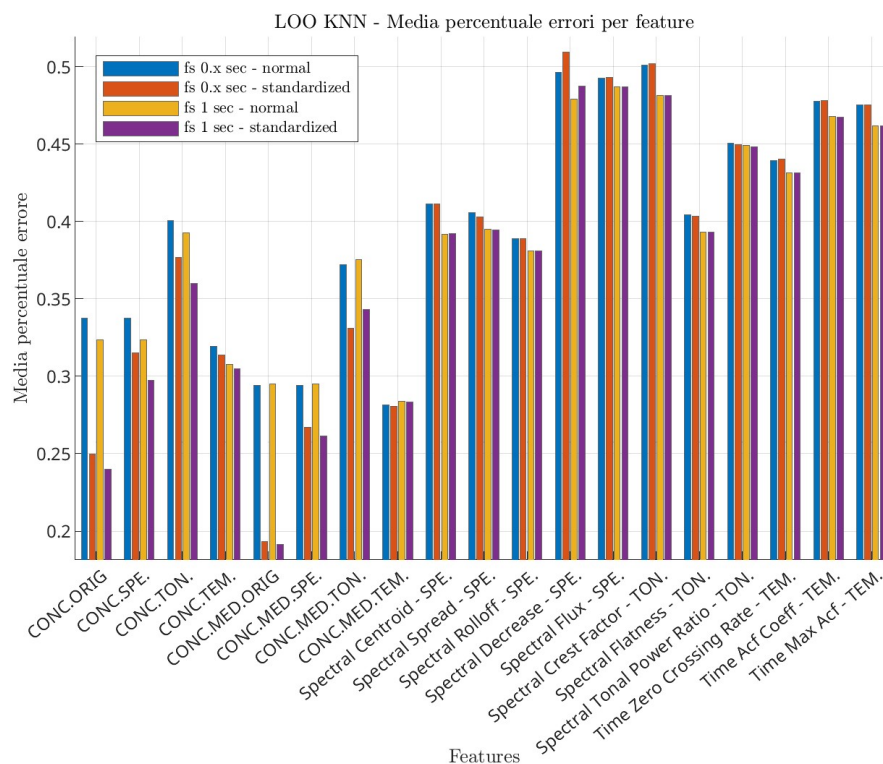


Grafico 4.4

4.5.3 Risultati rappresentazione con etichette semantiche

Nel corrente paragrafo andremo ad analizzare la seconda fase di classificazione che analizza i risultati dei problemi rappresentati da categorie semantiche. Come precedentemente affermato, la situazione in cui si sono disegnate queste casistiche è un condizione multi-etichetta. È possibile presupporre che ogni risultato sia stato influenzato negativamente dalla presenza di altri suoni non considerati nel problema in esame, come, allo stesso tempo, la quantità limitata di dati a disposizione per classe potrebbe conseguire un scarso addestramento condizionando la capacità di generalizzazione del discriminatore. Tuttavia, è comunque possibile trarre delle conclusioni interessanti.

In generale possiamo sostenere che le finestre con la versione standardizzata ottengono una media di errore minore rispetto alla controparte normale, nel dettaglio, la versione FS1.STD primeggia costantemente. Lo si può osservare nel grafico 4.5 che riassume gli esiti dei problemi disegnati, esponendo l'errore medio ottenuto trasversalmente alle *features*. È chiaro che il valore medio in queste analisi permette di trarre una visione generale dell'insieme. Tuttavia, si sottolinea che nei dati ottenuti sono presenti delle *features* che, a causa del loro scarso risultato, sbilanciano in modo negativo l'esito. Per questo verranno comunque evidenziati i dati che spiccano rispetto al gruppo, ma che potrebbero discostarsi in alcuni casi dai quelli esposti nel grafico. Si evidenzia che, siccome i casi di classificazione ternaria hanno avuto un errore fuori scala rispetto alla media generale, si è deciso per comprimere una parte dell'asse delle ordinate per ottenere una migliore visione dell'insieme. A seguire saranno descritti nel dettaglio i tre gruppi elencati nel capitolo 4.4.2.

In merito ai casi di classificazione PR-2.1 (binari) che discriminano la presenza/assenza di una determinata classe, i risultati mostrano un errore medio moderato. È possibile notare che la classe V ottiene un lieve miglioramento con la finestra FS0X.NOR, diversamente dagli altri casi, dove invece vediamo sempre protagonista la versione FS1.STD.

In questo gruppo i problemi con i migliori risultati riguardano i suoni appartenenti alla GEO, delle classi T e P nei problemi PR-2.1.3 P e PR-2.1.4 T. Si evidenzia il risultato delle feature CONC.MED.SPE, che primeggia in entrambe le frequenze, con errori pari al 25% nel primo e 27% nel secondo, e in un posizione molto vicina anche la feature CONC.MED.ORIG.. Il loro risultato lo si può notare nel grafico 4.6, che mette in evidenza il comportamento delle *features* trasversalmente ai casi sottoposti, con una media sui valori. È possibile notare che anche CONC.SPE. e CONC.ORIG, hanno ottenuto in generale un buon risultato. In questo gruppo la BIO sembra essere stata caratterizzata dalle *feature* spettrali.

Per quanto concerne il gruppo PR-2.2, classificazione binaria ma con due classi distinte, su tre dei quattro problemi, i risultati sono soddisfacenti per il nostro contesto. Le classi si trovano in una condizione molto più bilanciata e il sistema ne ha goduto, seppur lievemente. Questi problemi hanno permesso di valutare gli insiemi ANT/BIO/GEO in contrapposizione a coppie, evidenziando delle differenze rilevanti. È possibile notare l'ANT in contrasto con la GEO, nel problema PR-2.2.1 V/P, dove ottiene risultati molto peggiori rispetto al confronto con la BIO, nel problema PR-2.2.2. V/G: è ipotizzabile supporre che il suono dei grilli, essendo discontinuo come il suono del veicolo, ha permesso al sistema di poterli discriminare più facilmente, a differenza invece, del suono della pioggia, che si mantiene persistente. Allo stesso modo, lo ritroviamo nella correlazione di BIO con ANT, in PR-2.2.1 V/P, e BIO con GEO, in PR-2.2.4 G/P: si nota nel primo una maggiore difficoltà, come appena evidenziato. La minima presenza del suono di veicoli all'interno del dato ha sicuramente influito sul modello, quindi è possibile ipotizzare che il suono dei grilli con la pioggia ha permesso un confronto equo, essendo entrambi molto percepibili. Differentemente, potendo valutare due casi relativi allo stesso confronto BIO/GEO, in PR-2.2.3 G/T e PR-2.2.4 G/P, è osservabile un comportamento simile, con valori leggermente migliori dei casi sopra esposti. Si ipotizza che l'origine simile, per T e P, pur essendo suoni diversi, converga in un intervallo di frequenze simili, che rispecchiano la GEO. Da un punto di vista delle features, si evidenzia che, in maniera simile al primo gruppo i risultati migliori sono dati dalla versione standardizzata per CONC.ORIG e CONC.SPE per quanto riguarda il problema PR-2.2.4 G/P, con un 20% medio di errore. Invece con un 25% negli altri casi per le stesse versioni ma con le medie concatenate in *CONC.MED.ORIG* e *CONC.MED.SPE*.. Tali valori si possono notare nel grafico 4.6, che espone i risultati dalla prospettiva delle features.

Nei casi ternari, PR-2.3.1 V/G/P e PR-2.3.2 V/G/T, i risultati si sono rivelati più deludenti delle aspettative. I suddetti problemi, pur avendo classi con dati bilanciati, si trovavano in quantità significativamente minori. Inoltre, il numero di classi, pari a tre, ha complicato ulteriormente il modello, unitamente ai possibili disturbi causati dagli altri suoni non inerenti al problema. I valori si discostano di almeno il 15% dalle altre casistiche, collocandosi nell'intorno del 50esimo percentile medio di errore. Dal punto di vista delle features invece si nota un comportamento diverso dai casi precedenti. Infatti, nel primo caso PR-2.3.1 V/G/P otteniamo i migliori risultati con le feature CONC.MED.ORIG, nella versione standardizzata in particolare, invece nel secondo caso PR-2.3.2 V/G/T, vediamo protagonista la CONC.MED.TON con un errore del 40% circa, per entrambe le finestre ed entrambe le

forme normali e standardizzate. Il grafico 4.8 ci propone i dati del gruppo dal punto di vista delle features.

In sintesi, possiamo evidenziare che i tutti i casi analizzati le *feature* CONC.MED.ORIG si dimostrano tra le migliori per descrivere i vari casi esposti, seguite da CONC.MED.SPE. nel particolare dei casi binari, e da CONC.MED.TON. per i casi ternari. In particolare, esprimono al meglio il loro potenziale nella versione con la finestra *FS1.STD*, che è risultata superiore nella maggioranza delle casistiche.

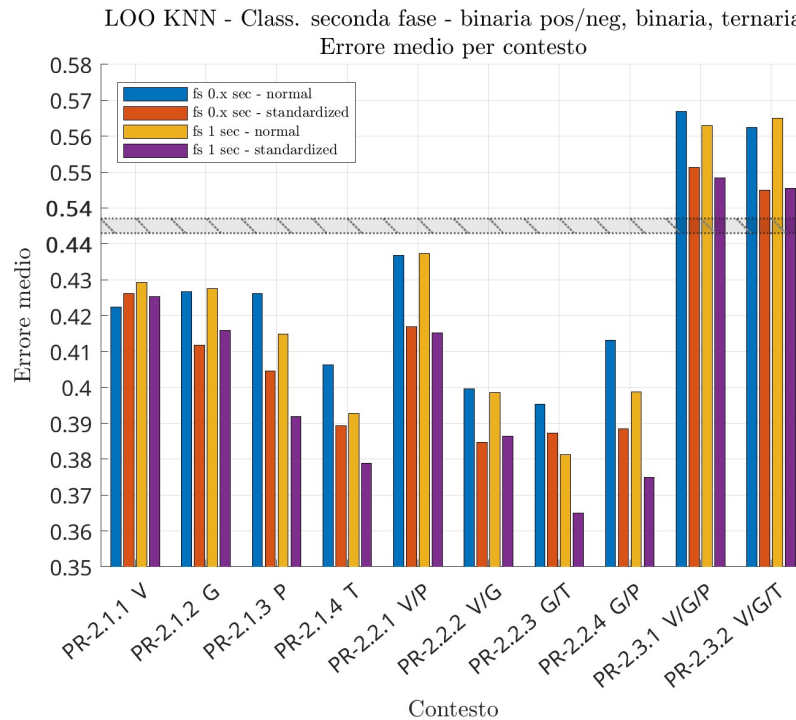


Grafico 4.5

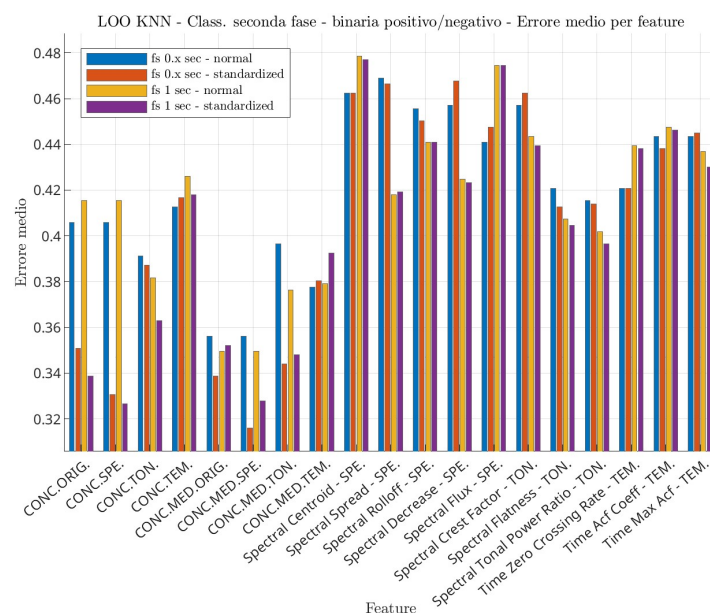


Grafico 4.6

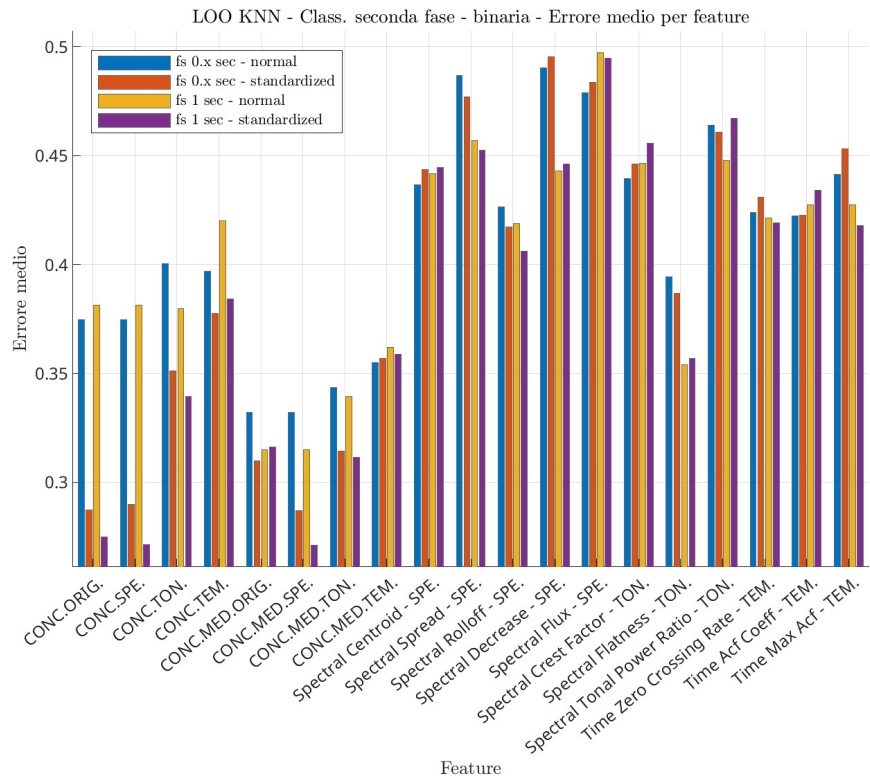


Grafico 4.7

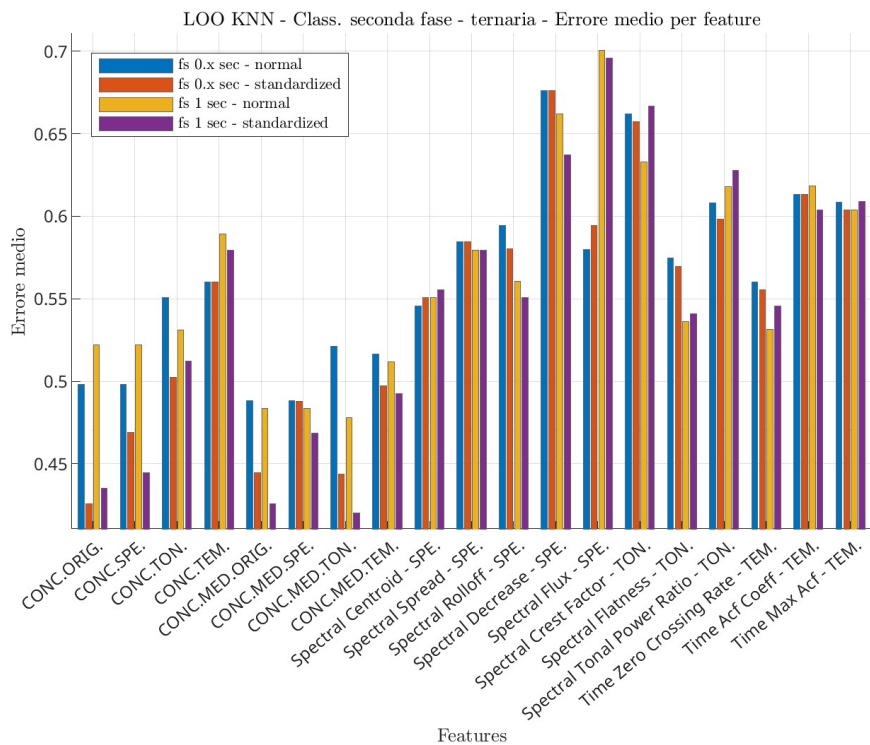


Grafico 4.8

Capitolo 5

Anomaly Detection

In questo capitolo si andrà ad esporre la terza fase dello studio, che si occupa di applicare algoritmi di A.D. per individuare eventuali oggetti anomali. Nel contesto dei *soundscape* gioca un ruolo molto interessante, per identificare suoni o eventi insoliti che potrebbero indicare cambiamenti ambientali, rumori estranei o fenomeni singolari: per esempio, la presenza di un cacciatore, oppure un animale non appartenente alla fauna del luogo. Nella circostanza dell'analisi, si è deciso di sperimentare l'utilizzo di questa metodologia applicando le conoscenze ottenute dalle fasi precedenti, per inferire informazioni ed esaminare quali elementi vengono classificati come anomali.

5.1 Configurazione *features* e metodi

Nel seguente paragrafo saranno descritte le *features* utilizzate in questo studio e le configurazioni dei metodi esposti nel paragrafo 2.3. Come menzionato sopra, la conoscenza ottenuta con lo studio di classificazione ci ha permesso di individuare *feature* che riescono ad illustrare chiaramente il *dataset*. Inoltre, dai risultati si è potuto verificare quale delle due finestre sia in grado di definire le caratteristiche più rilevanti. Si è quindi utilizzato il set di *feature* della versione FS1 di entrambe le forme di dati, per un totale di 38 *features*, 19 normali (FS1.NOR) e 19 standardizzate (FS1.STD).

Per quanto riguarda i metodi, le versioni utilizzate presentano delle configurazioni considerabili come standard, quindi non sono state fatte particolari ottimizzazioni. IF utilizza 100 alberi decisionali, un valore largamente utilizzato. La scelta di un numero elevato determina una maggiore robustezza del modello, ma incrementa notevolmente anche il tempo di calcolo. Il numero di campionamenti presi per ogni albero è di 256 elementi, un valore comune per mantenere una buona prestazione. LOF, invece, che discrimina un'anomalia basandosi sulla densità locale definita dai k vicini, si è tenuto per questi ultimi il valore di venti elementi, e come metrica di distanza, quella euclidea, già descritta precedentemente. L'algoritmo OCSVM, utilizza come *kernel* il modello gaussiano, così da proiettare i dati in uno spazio superiore dove poterli separare. Infine, come fattore di contaminazione, ovvero la soglia decisionale per definire quali elementi trattare come anomalie, è stata impostata a zero per tutti e tre gli algoritmi. Con tale valore il sistema tratta tutti gli elementi come normali e

non come anomalie, assegnandogli solo un punteggio di anormalità. Si è considerato questo valore poiché lo studio non ha valutato tale soglia, ma solo le posizioni massime di punteggio ottenute.

5.2 Validazione dei risultati

In questo paragrafo verrà esposta la modalità eseguita per applicare A.D. nel contesto dello studio. Gli algoritmi sopra descritti fanno parte della tipologia di approcci non supervisionati, ovvero non necessitano quindi di etichette per poter funzionare. Però per poter dedurre informazioni rilevanti, è stato comunque impiegato il *dataset DATA2*, che è etichettato, così da potersi servire delle conoscenze sul contenuto per verificare il risultato.

La modalità del processo prevede 30 esecuzioni dell'algoritmo alimentato con i dati di una *feature*. Dai punteggi di anomalia ottenuti viene calcolata la media, in maniera tale da essere indipendenti dal valore della singola esecuzione. L'approccio viene replicato per ciascuna *feature*, e per tutti gli algoritmi. Infine, da ogni esecuzione si sono considerati solo i primi tre valori in classifica con il punteggio maggiore: quindi 38 (*features*), per tre risultati, e per tre algoritmi. Per valutare ogni metodo, sono state estratte le prime cinque occorrenze più numerose da ognuno. Per avere visione complessiva, invece, sono state estratte le prime cinque occorrenze migliori di tutti e tre i set di risultati.

5.3 Risultati

Dopo aver esposto la modalità di applicazione dell'analisi si andrà a descrivere i risultati ottenuti. Nella tabella 5.1 è possibile avere una visione dei risultati come descritti nel paragrafo precedente: le maggiori occorrenze per i tre algoritmi, e poi sul totale, con le relative percentuali di distribuzione.

Dal punto di vista dei singoli suoni, vediamo come il rumore dei veicoli è stato rilevato per un 40% su IF e O, per un 60% invece per L. Il suono degli uccelli U è sempre presente per un 60% e con una distribuzione simile in tutte e tre le casistiche, diversamente invece, da quello dei grilli G, che è emerso presente per un 40% e solo in L. Il suono del fiume/cascata C, come prevedibile, appare molto diffuso tra i risultati, ma si può ipotizzare che non abbia nulla di particolare e che la sua presenza sia dovuta unicamente alla sua ampia distribuzione nei dati.

Anche il suono della pioggia si è notevolmente osservato, per un 80%, e si può affermare che valgono le stesse conclusioni appena fatte per il caso precedente. Invece, le interferenze e i suoni sconosciuti non hanno realizzato le aspettative. Si poteva prevedere che sarebbero

emersi data la loro minima presenza, quanto alla loro particolarità, ma invece sono stati rilevati solo per un caso, e solo da O.

Come è possibile notare, dalla prospettiva dei metodi, si vede che IF e O hanno espresso delle distribuzioni simili, diversamente da L. che ha dimostrato una sensibilità diversa nella maggioranza dei casi.

Invece, se si osserva dal punto di vista degli insiemi di suoni ANT/BIO/GEO, si può notare come la GEO sia molto presente, per almeno un 80%. Per una parte si possono trarre le stesse conclusioni definite prima sulla distribuzione dei suoni *C* e *P*, ma per quanto riguarda quello di *T*, si può ipotizzare che non lo sia. Infatti, la sua distribuzione nel *dataset* di riferimento è di circa del 60%, e in rapporto con la sua bilanciata presenza tra i risultati si può ritenere che l'informazione è molto interessante. Similmente, per *ANT* e *BIO* è possibile sostenere l'ipotesi che il risultato ottenuto sia rilevante: per il primo sulla bassa numerosità dei dati in input, per il secondo invece, data la sua grande distribuzione, perché non si è propagato similmente a quanto osservato nei suoni più rilevati della *GEO*.

Infine, da una prospettiva temporale, si manifesta che l'80% dei risultati sul totale, ma anche nei singoli algoritmi, si concentra alle ore 14:00, nella seconda parte della giornata. Lo stesso per l'elemento identificato maggiormente come anomalia che risulta invece verso le ore 18:00.

			ANT	BIO		GEO			ALTRO	
GRUPPO	AUDIO	#	V	U	G	C	P	F	I	S
TOTALE	20200323_180000.WAV	43	x	x			x			
	20200303_140000.WAV	25					x	x		
	20200304_140000.WAV	21				x	x	x		
	20200311_140000.WAV	18				x	x	x		
	20200318_140000.WAV	14	x	x		x				
			40%	40%	0%	80%	80%	60%	0%	0%
IFOREST	20200323_180000.WAV	22	x	x		x	x			
	20200303_140000.WAV	18					x	x		
	20200304_140000.WAV	12				x	x	x		
	20200309_060000.WAV	8	x	x		x				
	20200316_220000.WAV	8		x		x	x			
			40%	60%	0%	80%	80%	40%	0%	0%
LOF	20200311_140000.WAV	15				x	x	x		
	20200323_180000.WAV	15	x	x		x	x			
	20200318_140000.WAV	10	x	x		x				
	20200301_180000.WAV	9		x	x	x		x		
	20200325_020000.WAV	8	x		x	x				
			60%	60%	40%	100%	40%	40%	0%	0%
OCSVM	20200323_180000.WAV	6	x	x		x	x			
	20200309_060000.WAV	5	x	x		x				
	20200314_180000.WAV	4	x	x		x	x	x		x
	20200328_140000.WAV	4				x	x	x		
	20200303_140000.WAV	3					x	x		
			60%	60%	0%	80%	80%	80%	0%	20%

Tabella 5.1

Capitolo 6

Conclusione

In questo studio sono state applicate tecniche di *Pattern Recognition* e *Machine Learning* per caratterizzare mediante classificazione e *anomaly detection* un *soundscape*.

Come è già noto nella letteratura sui *soundscape*, il contesto dello studio presenta notevoli difficoltà intrinseche. Si deve tenere conto che lo stesso cervello umano, nonostante le sue straordinarie capacità, fatica a distinguere e classificare determinati suoni.

Entrambe le fasi di classificazione hanno dimostrato che la *feature* migliore consiste nella concatenazione delle medie dei gruppi di tipologie di *features*, nella versione standardizzata e in combinazione con la dimensione della finestra più ampia. Tra i problemi con categorie note la discriminazione sul luogo ha ottenuto risultati nettamente superiori rispetto al resto. Nelle casistiche con categorie semantiche i problemi binari con due classi distinte occupano la prima posizione distaccandosi di poco da quelli binari che discriminano la presenza/assenza della classe. I problemi ternari, essendo un contesto multi etichetta, hanno affrontato le maggiori difficoltà e i risultati si sono espressi con una predizione media corretta minore della metà degli oggetti. Nel particolare caso dei binari le feature spettrali hanno ottenuto risultati interessanti, per i ternari lo sono state quelle tonali.

L'*Anomaly detection* non ha rilevato particolari elementi anomali e non è emerso nessuno schema particolare tra i dati. Le aspettative riguardanti i suoni relativi a interferenze e a quelli sconosciuti non si sono rivelate, non sono emersi come anomalie. I suoni relativi alla geofonia sono stati riscontrati frequentemente, in particolare quello del tuono ha ottenuto un risultato interessante. Si è potuto notare che la presenza di dati anomali ricade nelle fasce orarie pomeridiane.

Il tentativo di filtraggio sperimentato non ha portato nessun vantaggio dal punto di vista della qualità dell'audio, ma i risultati ottenuti sono comunque validi. La stessa combinazione di caratteristiche vincenti emerse dalla classificazione si sono mantenute tali anche rimuovendo parte delle frequenze.

Noti i vantaggi derivati dalla comprensione dei *soundscape*, si termina questo studio con la speranza che la ricerca possa avanzare e trovare soluzioni migliori per approfondire la conoscenza del nostro ecosistema.

Bibliografia

- [1] Colin A. Quinn et alii. *Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data*. Ecological Indicators, Vol. 138, 2022, 108831, ISSN 1470-160X, <https://doi.org/10.1016/j.ecolind.2022.108831>
- [2] Farina Almo et alii, *Sonotopes reveal dynamic spatio-temporal patterns in a rural landscape of Northern Italy*. Frontiers in Ecology and Evolution, vol.11 ,2023, 2296-701X, <https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2023.1205272>
- [2] Tovar Garcia, J.D. and Acevedo-Charry, O. *Dataset of passive acoustic monitoring at the Nature Reserve Los Yátaros, Gachantivá, Boyacá, Colombia*. 2021. Biota Colombiana. 22, 1 (Jan. 2021), 200–208. <https://doi.org/10.21068/c2021.v22n01a13>
- [4] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.3
- [5] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.4
- [6] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.1
- [7] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.5
- [8] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.3.2
- [9] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.4.1.1
- [10] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 3.4.3
- [11] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 5.3.3.1
- [12] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 5.3.3.2
- [13] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. 2012. IEEE Press. par. 5.3.3.6