

Entrega 1 : Clasificación de Audio

Firmin Cadot

29 Septiembre 2023

1. Contexto de Aplicación

Este proyecto se centra en la clasificación de señales de audio utilizando modelos de aprendizaje profundo con dos tipos de modelos : redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN). El objetivo es desarrollar un modelo que pueda identificar y clasificar diferentes sonidos, lo que podría tener aplicaciones en varios campos como la música, la salud y la seguridad.

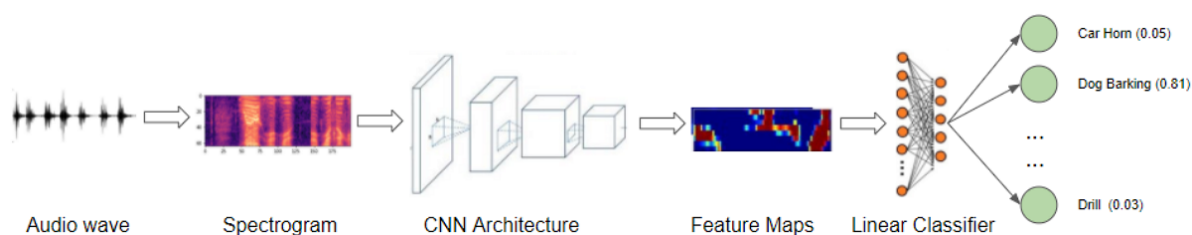


Figure 1: Ejemplo de un CNN para un problema de clasificación de audio

2. Objetivo de Machine Learning

El objetivo principal es predecir la etiqueta de clasificación de audio para los archivos de audio proporcionados, basándose en sus características acústicas extraídas.

3. Dataset

Tipo de Datos: Archivos de audio (.wav) y un archivo CSV con metadatos.

Tamaño:

- Número de Datos: 300 archivos de audio.
- Tamaño en Disco: 300 MB para los archivos de audio; el tamaño del archivo CSV es mínimo en comparación.
- Dataset Original: 18,877 archivos de audio (10.37 GB) y un archivo CSV correspondiente de Kaggle.

Descripción del CSV: El archivo CSV acompaña a los archivos de audio y contiene dos columnas:

- **fname:** El nombre del archivo de audio correspondiente.
- **label:** La etiqueta de clase del sonido, por ejemplo, 'Saxophone'.

Este archivo actúa como un mapeo entre los archivos de audio y sus respectivas etiquetas de clase, permitiendo un etiquetado y acceso eficiente a las clases durante el proceso de entrenamiento y evaluación del modelo.

Ejemplo de entradas en el archivo CSV:

```
fname,label
5388d14d.wav,Saxophone
078f7ac0.wav,Hi-hat
229be2be.wav,Snare_drum
```

Enlace al Dataset: [Freesound Audio Tagging](#)

Distribución de las Clases: El dataset consiste en 10 clases diferentes, con aproximadamente 30 ejemplos para cada clase, lo que indica una distribución equilibrada.

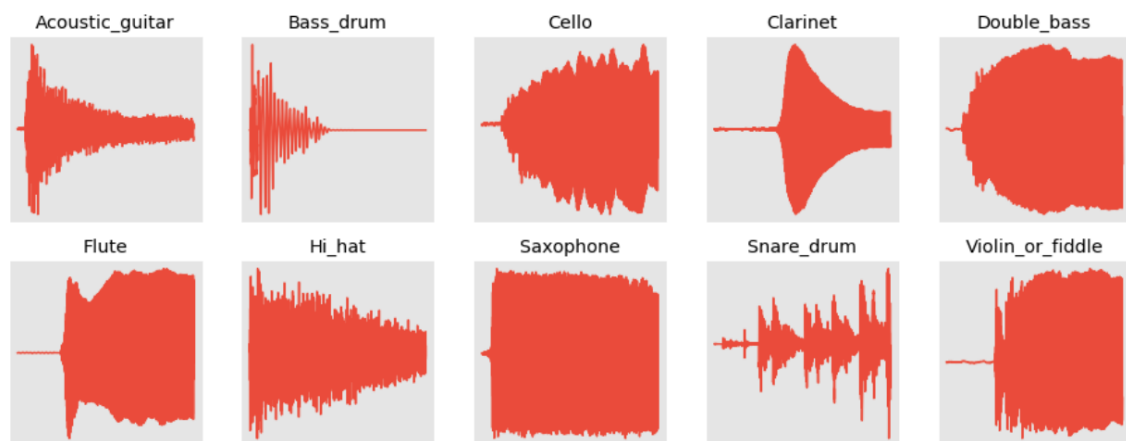


Figure 2: Series temporales de cada clase

4. Métricas de Desempeño

En este proyecto, se utilizarán principalmente dos métricas para evaluar el rendimiento de los modelos de clasificación de audio: la Exactitud (Accuracy) y la Curva ROC (Receiver Operating Characteristic).

4.1. Exactitud (Accuracy)

La exactitud es una métrica esencial en modelos de clasificación multiclase, la cual mide el ratio de predicciones correctas realizadas por el modelo con respecto al total de predicciones. Aunque la exactitud puede proporcionar una visión general rápida de la eficacia del modelo, es crucial considerar su uso en contextos donde las clases están desbalanceadas, ya que podría dar una visión errónea de la eficiencia del modelo.

4.2. Curva ROC y Área Bajo la Curva (AUC) para Multiclase

En el contexto de clasificación multiclase, la Curva ROC y el AUC pueden extenderse para evaluar el desempeño del modelo en cada clase contra el resto, proporcionando una visión detallada del comportamiento del modelo en diferentes situaciones. Cada clase se considera como la clase positiva contra todas las otras clases. El valor del AUC en un contexto multiclase puede ayudarnos a entender cómo el modelo se desempeña en general, en todas las clases, y un valor de AUC más cercano a 1 indica un mejor rendimiento del modelo.

Estas métricas permitirán una evaluación integral del modelo, facilitando la identificación de posibles problemas y áreas de mejora, y ayudarán en la optimización continua del modelo para alcanzar un rendimiento superior en tareas de clasificación de audio.



Figure 3: Gráfico de la Curva ROC

5. Referencias

- Freesound Audio Tagging - Kaggle
- Jon Nordby. (2019). *Practical introduction to Audio Classification using Deep Learning*. Presentado en EuroPython 2019, Basel. Enlace al Video. Grabación realizada por el equipo de EuroPython.
- Valerio Velardo - The Sound of AI. *Deep Learning (for Audio) with Python: Course Overview*. Enlace a la Serie de Videos.
- Haytham Fayek. (2016). *Speech Processing for Machine Learning*. Recuperado de <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>