Proyecto Final

Daniel Felipe Rambaut Lemus
Miguel Gutierrez Vidal
Felipe Guzman

Escuela de Ingeniería, Ciencia y Tecnología
Universidad del Rosario
Bogotá D.C.

Abril 2021

1. Introducción

Un clasificador de género musical es un programa de software que predice el género de una pieza musical en formato de audio. Estos dispositivos se utilizan para tareas como etiquetar música automáticamente para distribuidores como Spotify y iTunes y determinar la música de fondo adecuada para eventos.

Actualmente, los seres humanos realizan manualmente la clasificación de géneros aplicando su comprensión personal de la música. Esta tarea aún no ha sido automatizada por enfoques algorítmicos convencionales, ya que las distinciones entre géneros musicales son relativamente subjetivas y están mal definidas. Sin embargo, la ambigüedad de la clasificación de géneros hace que la inteligencia artificial sea adecuada para esta tarea. Con suficientes datos de audio, de los cuales se pueden obtener fácilmente grandes cantidades de música disponible gratuitamente en internet, el aprendizaje automático puede observar y hacer predicciones utilizando estos patrones definidos. El objetivo de este proyecto es construir un clasificador de géneros musicales de prueba de concepto utilizando un enfoque de aprendizaje profundo que pueda predecir correctamente el género al cual pertenece un audio de entrada.

2. Metodología

Los coeficientes cepstrales de frecuencia Mel MFCC representa un conjunto de características del espectro de potencia a corto plazo del sonido y se ha utilizado en técnicas de reconocimiento y categorización de sonido de última generación. Modela las características de la voz humana. Esta característica es una gran parte del vector de características final tamaño variante. El método para implementar esta función vemos que existen diferentes librerías en Python que permite la obtención de estos valores. Como hemos mencionado anteriormente, esta forma de representar cada uno de los audios y así realizar la clasificación de cada uno de los mismos.

Además utilizaremos Resnet34, que es una red neuronal convolucional de 34 capas que se puede utilizar como modelo de clasificación de imágenes de última generación. Este es un modelo que se ha entrenado previamente en el conjunto de datos ImageNet, este conjunto de datos tiene más de 100,000 imágenes en 200 clases diferentes. Sin embargo, es diferente de las redes neuronales tradicionales en el sentido de que toma residuos de cada capa y los usa en las capas conectadas subsiguientes (similar a las redes neuronales residuales que se usan para la predicción de texto).

3. Resultados

Para este proyecto utilizamos el modelo de clasificación Resnet34 y Resnet18. Los resultados que obtuvimos a lo largo del desarrollo del proyecto fueron los siguientes:

1. En el primer paso realizamos la importación de las imágenes de la base de datos de Kaggle, sin embargo a lo largo del desarrollo del proyecto, se evidenció que las imágenes traen un borde blanco el cual puede afectar el entrenamiento del modelo de clasificación y por ello se procedió a crear un nuevo data set con las nuevas imágenes ya recortadas.
2. En el segundo paso, realizamos la construcción del modelo de clasificación y para ello utilizamos lo visto en el tutorial de Fastai. El modelo que se selecciono fue el Resnet34 ya que nos proporciona un accuracy alto en comparación con la Resnet18. Además debemos mencionar que no se agregaron funciones extras como lo es la normalización entre otras, debido a los resultados obtenidos bajo estas funciones, los cuales no fueron muy buenos.
3. En la parte final de nuestro proyecto, teniendo nuestro modelo de clasificación ya entrenado, decidimos no guardarlo debido a que la el tiempo de entrenamiento es relativamente corto y de esta forma procedemos a implementar una interfaz en donde el usuario pueda realizar diferentes pruebas. Para ello utilizamos librerías de Colab para permitirle al usuario cargar un audio y nosotros atravesé de diferentes funciones de la libraría de librosa obteníamos su respectivo espectrograma.
4. Con los modelos se pudo llegar a una precisión cerca de 0.73 lo cual es bueno para el numero de tipos de clases que se analizaron.

4. Conclusión

Finalmente, podemos obtener un modelo de clasificación el cual funciona dado un audio que tiene una duración de 30 segundos. Además se pude evidenciar una función en donde podemos calcular el respectivo espectrograma del audio entregado por el usuario. Realizando diferentes pruebas pudimos ver que el modelo se comporta bien para diferentes tipos de canciones, sin embargo cuando intentamos predecir canciones que tiene influencia por otro tipo de genero hay casos en donde no clasifica correctamente. Como opción de mejora podemos construir un data set en donde podamos encontrar las imágenes del mel espectrograma y aplicando diferentes técnicas de visión computador para obtener mas información sobre las canciones y así el modelo pueda identificar con mayor precisión.

Referencias