



Identificador de instrumentos con Machine Learning usando características extraídas con Librosa



Alejandra Calderón Alfonso, Emmanuel Botero Osorio

Instituto de Física, Universidad de Antioquia U de A Calle 70 No. 52-21, Medellín, Colombia

INTRODUCCIÓN

La música tiene una gran variedad de instrumentos de sonidos con diferentes propiedades.

¿Se pueden determinar características cuantificables del sonido a través del paquete de python Librosa?

Con estas características ¿se puede hacer un modelo de aprendizaje supervisado que prediga a que instrumento corresponde el sonido?

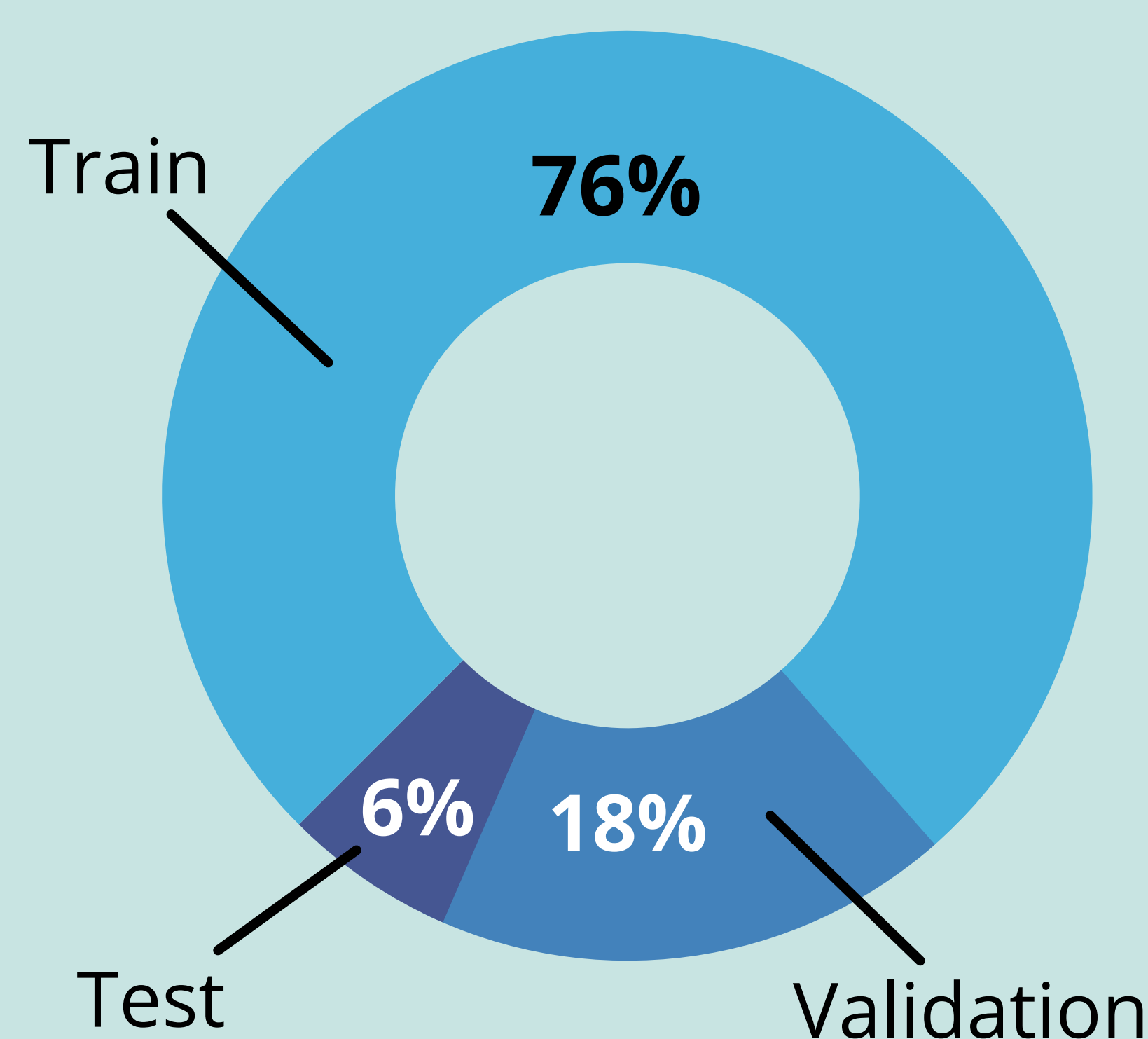
¿Cual algoritmo de Machine Learning da el mejor modelo y cómo se puede mejorar este?

MATERIALES Y MÉTODOS

Se usó "The NSynth Dataset". Se le hizo una extracción de características con Librosa:

- Coeficientes cepstrales en las Frecuencias de Mel
- Espectrograma de Mel
- Componentes armonicos
- Energía cromática
- Contraste espectral

Segmentación de los datos



Se hicieron 4 ejecuciones diferentes:

- 5k audios por cada instrumento (9-8 inst.)
- 1k audios por cada instrumento (8-7 inst.)

Se entrenó un modelo de aprendizaje supervisado con Random Forest para cada ejecución.

RESULTADOS

Al hacer la primera ejecución, se obtiene que el algoritmo Naive Bayes da un modelo con 13.21% de precisión, Random Forest con 57.69% y Support Vector Machine con 3.44%.

Tomando únicamente RF se obtienen los tiempos de extracción de características de los datos de entrenamiento t_{ext} y las precisiones de modelos expuestas en la siguiente tabla:

Ejecución	Precisión RF [%]	t_{ext} de train data
1	57.69	5h20m15s
2	63.30	5h11m12s
3	55.94	1h18m6s
4	66.25	1h15m53s

Matriz de confusión de RF: ejecución 4

Para las cuatro ejecuciones se obtienen matrices similares a la de este gráfico, el cual muestra los coeficientes de ajuste del modelo Random Forest para c/instrumento.

bajo	0.72	0.00	0.00	0.16	0.09	0.01	0.01
cobres	0.05	0.62	0.03	0.00	0.00	0.06	0.24
flauta	0.05	0.00	0.85	0.06	0.01	0.00	0.03
guitarra	0.39	0.06	0.01	0.46	0.04	0.03	0.01
marimba	0.15	0.10	0.06	0.03	0.54	0.05	0.07
cuerdas	0.07	0.01	0.01	0.06	0.01	0.82	0.00
voz	0.17	0.07	0.10	0.03	0.00	0.01	0.61
	bajo	cobres	flauta	guitarra	marimba	cuerdas	voz

ANÁLISIS

RF, al funcionar con una estructura similar a los arboles de decisión, asocia c/u de las características de audio a un instrumento en específico, por lo cual logra obtener una mejor predicción.

Reducir a 1/5 los datos, mejora la eficiencia: disminuye t_{ext} un 75.6 % sacrificando solo un 11.6% de la precisión.

Eliminar entradas del modelo aumenta la precisión pero no hace una gran diferencia en la efectividad del código.

CONCLUSIONES

- Características extraídas con Librosa buenas para ML.
- Mejor modelo: RF para la ejecución 4.
- Tomar menos instrumentos y archivos de forma estratégica permite optimizar el código.
- Los instrumentos con mayor similitudes entre si son los de peor precisión en la predicción.

BIBLIOGRAFÍA

*Kawwa, N. (2019) Can we guess musical instruments with machine learning?, Medium. Disponible en <https://medium.com/@nadimkawwa/can-we-guess-musical-instruments-with-machine-learning-afc8790590b8>

*The NSYNTH dataset (2017) Magenta. Disponible en <https://magenta.tensorflow.org/datasets/nsynth#description>

*IBM Cloud Education. (n.d.). What is supervised learning? IBM. Disponible en <https://www.ibm.com/cloud/learn/supervised-learning>

*Calderón Alfonso, A., & Botero Osorio, E. (2022, Noviembre). Identificador de instrumentos con ML. Identificador_ML_librosa. Disponible en https://github.com/ebo1703/Identificador_ML_librosa