

VOICHELP

Diana Torres
Luis Jimenez
Maria Aparicio

Prof. Fabio Martinez Carrillo



Ingengería de Sistemas
16 de Julio de 2019

Problema

Actualmente el 2% de la población Colombiana tiene problemas auditivos y se les dificulta para comunicarse con el resto de la sociedad.

Asimismo es una herramienta que podrá ser utilizada para el aprendizaje del idioma español.

¿Qué es?

Voichelp es una herramienta principalmente dedicada a personas con discapacidad de audición, la cual le ayudará a traducir de audio a texto las palabras que se requiera al hablar con una persona, de manera sencilla.

Dataset

Contamos con un dataset propio de 40 clases, 1215 audios por cada uno de ellas; es decir en total 48556; Generados mediante grabaciones de personas, librosa y Cross_over.

MFCC

Los Mel Frequency Cepstral Coefficients (Coeficientes Cepstrales en las Frecuencias de Mel) o MFCCs son coeficientes para la representación del habla basados en la percepción auditiva humana. Estos extraen características de las componentes de una señal de audio que sean adecuadas para la identificación de contenido relevante, así como obviar todas aquellas que posean información poco valiosa como el ruido de fondo.

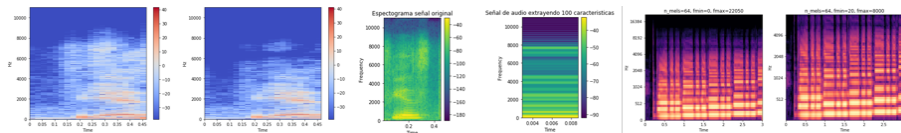


Figura 1: filtros utilizados

Clasificadores usados

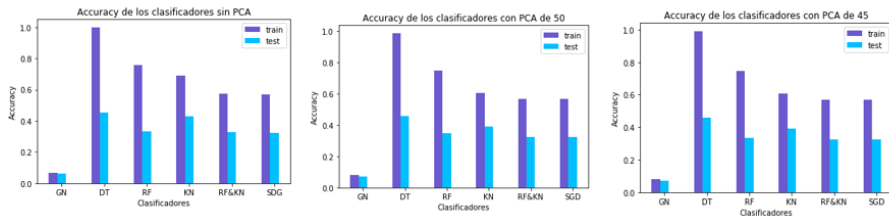


Figura 2: Comparación Clasificadores

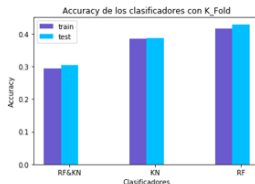
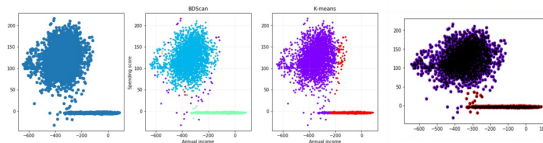


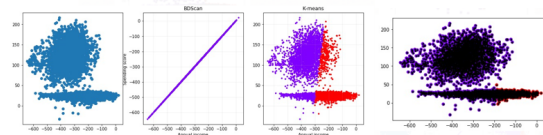
Figura 3: Comparación utilizando K_{Fold}

Aprendizaje no supervisado

- Primero se permutaron los audios grabados de las personas lo cual genera ruido. Haciendo uso de MFCC de dos se genero un csv para el ruido y para los audios originales, y se uso Kmeans,DBScan y Brinch para 2 clusters.



- Posteriormente se quiso observar que también dividia los audios originales con los audios producidos mediante librosa y cross-over



Conclusiones

- 1 Se requieren más audios de cada clase debido a que tiende a sobreaprender un poco, sin embargo da un accuracy bueno.
- 2 El mejor clasificador fue KNeighbor con un número de vecinos igual a siete.
- 3 Se hizo de uso de MFCC debido a que era el tratamiento de datos que generaba un mejor accuracy.
- 4 Los métodos de conjunto nos dieron un mejor resultado debido a que combinan múltiples algoritmos en un solo modelo predictivo para disminuir la varianza, disminuir el sesgo o mejorar las predicciones
- 5 El algoritmo de Birch fue el que mejor nos dio resultado en el aprendizaje no supervisado separando el audio permutado (ruido) con los audios originales obtenidos de las grabaciones de personas.
- 6 El aprendizaje no supervisado fue una mala opción para identificar entre audios generados por las funciones de cross-over y el audio original, debido a que apesar de que se cambia el tono de la voz MFCC extrae las características más relevantes por lo cual van a ser muy semejantes ambos audios.

REFERENCIAS

- 1 <https://arxiv.org/abs/1803.01094>
- 2 <https://www.ijcsmc.com/docs/papers/March2015/V4I3201545.pdf>
- 3 https://scholar.google.com.co/scholar?q=speech+recognition+classification+python&hl=es&as_sdt=0&as_vis=1&oi=scholar
- 4 <https://www.code-learner.com/python-speech-recognition-introduction-and-practice/>
- 5 <http://dkopczyk.quantee.co.uk/speech-nn/>
- 6 <https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8>