



RECONOCIMIENTO DE EMOCIONES PARA LA DETERMINACIÓN DE LA VALIDEZ DE LLAMADAS DE EMERGENCIA



Introducción

Los servicios de emergencia alrededor del mundo existen para auxiliar a las personas ante situaciones que amenacen su integridad física y mental, entre muchos de estos servicios se encuentran las líneas telefónicas de emergencia, las cuales atienden miles de llamadas por año de personas que requieren asistencia.

TOTAL DE LLAMADAS DE EMERGENCIA: COMPARATIVO ANUAL

Enero-diciembre 2016-2019

Tipo de llamada	Enero - Diciembre 2016	Enero - Diciembre 2017	Enero - Diciembre 2018	Enero - Diciembre 2019	Variación	
					Absoluta	Relativa (%)
Total de llamadas	122,181,509	112,460,767	95,016,135	75,222,524	-19,793,611	-20.83%
Procedentes (reales)1/	12,666,827	13,937,674	16,143,177	16,044,138	-99,039	-0.61%
Seguridad	8,678,679	9,036,588	9,848,493	10,001,943	153,450	1.56%
Médico	1,684,462	2,030,217	2,124,722	2,216,769	92,047	4.33%
Asistencia*	797,658	768,766	2,462,535	1,981,922	-480,613	-19.52%
Protección Civil	746,015	853,844	876,338	1,039,940	163,602	18.67%
Otros Servicios*	603,962	1,055,881	677,866	642,585	-35,281	-5.20%
Servicios Públicos	156,051	192,378	153,223	160,979	7,756	5.06%
Improcedentes 2/	109,514,682	98,523,093	78,872,958	59,178,386	-19,694,572	-24.97%
Llamada muda	15,148,151	16,790,954	31,251,440	27,963,305	-3,288,135	-10.52%
Llamada incompleta	37,866,991	19,379,852	14,814,830	9,765,203	-5,049,627	-34.08%
Llamada de broma por niños	33,645,167	37,166,004	16,046,815	8,596,115	-7,450,700	-46.43%
Otras llamadas de no emergencia	21,052,699	23,766,473	12,486,891	8,977,889	-3,509,002	-28.10%
Jóvenes / Adultos jugando	NA	NA	2,228,515	1,403,322	-825,193	-37.03%
Transferencia de llamada	1,501,401	1,261,339	1,346,361	1,490,349	143,988	10.69%
Insultos por adultos / llamada obsena	NA	NA	574,364	803,392	229,028	39.88%
Llamada de prueba	300,273	158,471	123,742	178,811	55,069	44.50%

Motivación

Ayudar a hacer un descarte rápido de llamadas falsas basado en la emoción relacionada a la tonalidad de la voz del audio de la llamada.



Objetivos

- Identificación de emociones humanas a partir de grabaciones de voz.
- Exploración de métodos y metodologías sobre el tratamiento de audio.
- Implementación de una red neuronal o métodos de aprendizaje de máquina que realizan una buena predicción.

Planteamiento del problema

El reconocimiento de las emociones humanas por medios computacionales siempre ha sido un gran desafío, esto con el fin de que la interacción humano-máquina sea lo más humana posible, la interpretación de las diferentes señales humanas que denotan una emoción y que pueden ayudar a detectar, en este caso la validez de las llamadas de emergencia.

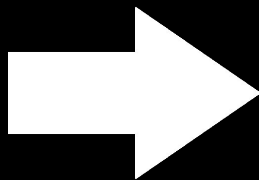
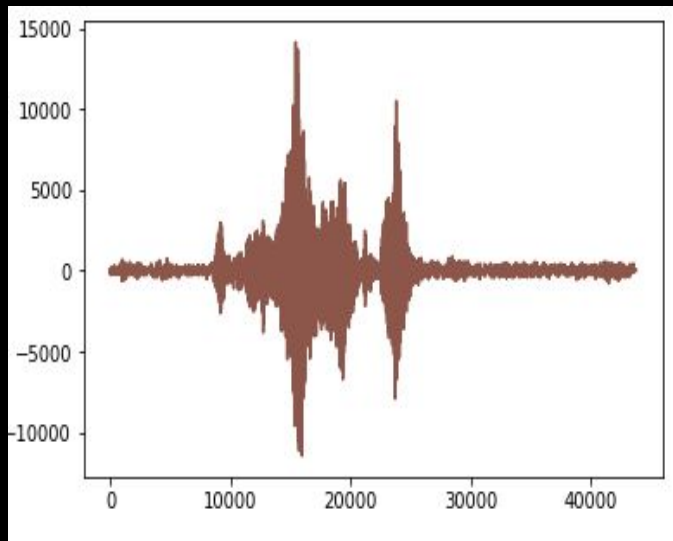
Desafíos

- Encontrar los puntos de inicio y fin de palabra es fundamental en procesamiento de voz.
- Filtrar el ruido proveniente del dispositivo que graba el audio y el ambiente.
- Subjetividad en definir qué rasgos del audio definen una emoción u otra.

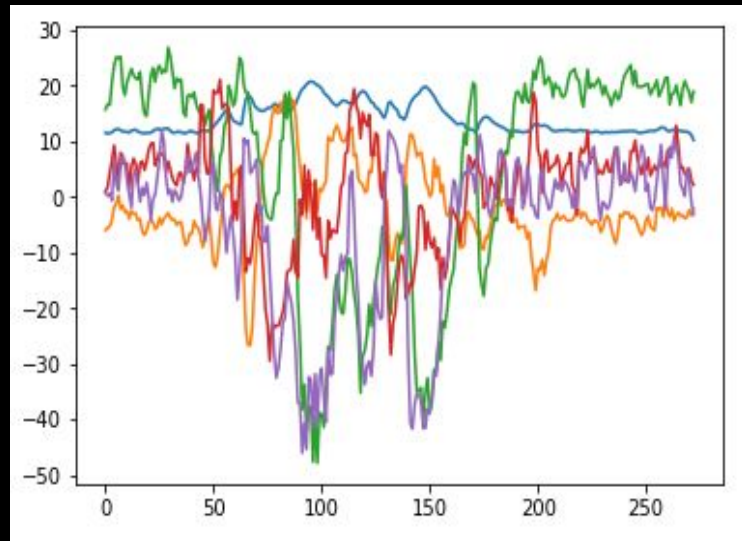
Coeficientes Cepstrales de Mel

MFCCs se calculan comúnmente de la siguiente forma:

- Separar la señal en pequeños tramos.
- A cada tramo aplicarle la Transformada de Fourier discreta y obtener la potencia espectral de la señal.
- Aplicar el banco de filtros correspondientes a la Escala Mel al espectro obtenido en el paso anterior y sumar las energías en cada uno de ellos.
- Tomar el logaritmo de todas las energías de cada frecuencia mel Aplicarle la transformada de coseno discreta a estos logaritmos.



MFEC



Dataset

Para reconocer emociones en la tonalidad de la voz humana se usó el dataset CREMAD4 en el que se tienen 7442 audios de 91 actores de voz (aparentemente tomados de películas) y 6 emociones para hombres y mujeres.

Consta de 7442 archivos de audio en formato .wav, el cual es muy usado para este tipo de audios.

No tiene un archivo asociado donde se puedan hallar las etiquetas a los audios, viene implícito en el nombre de cada archivo.

El dataset esta masomenos bien balanceado entre sus clases, como se ve a continuación:

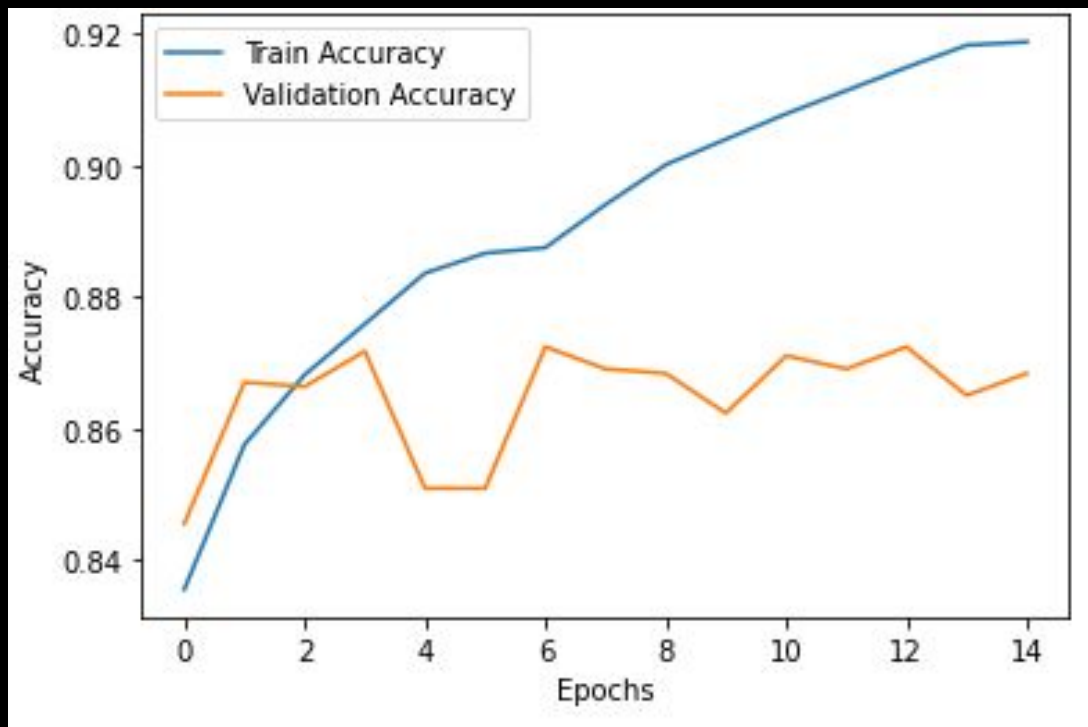
male_happy	671
male_fear	671
male_sad	671
male_angry	671
male_disgust	671
male_neutral	575

female_happy	600
female_fear	600
female_sad	600
female_angry	600
female_disgust	600
female_neutral	512

Modelo con mejores resultados

Layer (type)	Output Shape	Param #
=====		
flatten_15 (Flatten)	(None, 3328)	0
dense_40 (Dense)	(None, 2048)	6817792
dense_41 (Dense)	(None, 1024)	2098176
dense_42 (Dense)	(None, 512)	524800
dense_43 (Dense)	(None, 128)	65664
dense_44 (Dense)	(None, 128)	16512
dense_45 (Dense)	(None, 12)	1548
=====		
Total params: 9,524,492		
Trainable params: 9,524,492		
Non-trainable params: 0		

Resultados

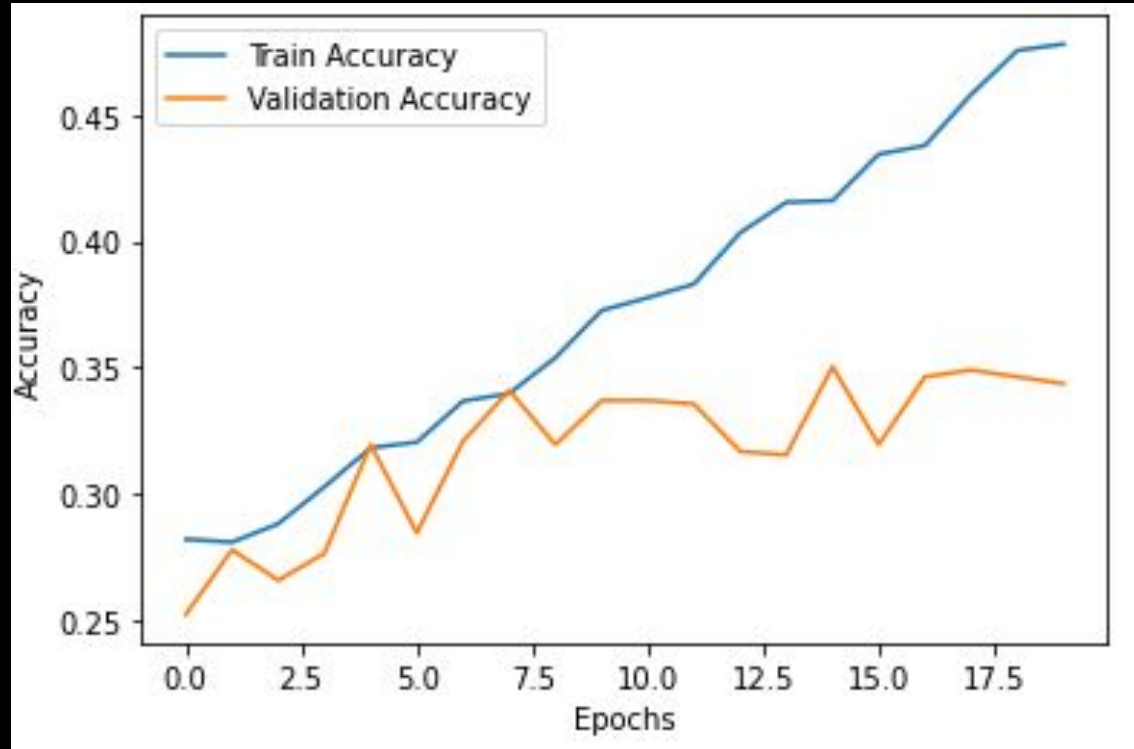


Modelo usando convoluciones

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 13, 16)	160
dropout (Dropout)	(None, 256, 13, 16)	0
conv2d_1 (Conv2D)	(None, 256, 13, 8)	1160
dropout_1 (Dropout)	(None, 256, 13, 8)	0
conv2d_2 (Conv2D)	(None, 256, 13, 4)	132
max_pooling2d (MaxPooling2D)	(None, 128, 6, 4)	0
flatten (Flatten)	(None, 3072)	0
dropout_2 (Dropout)	(None, 3072)	0
dense (Dense)	(None, 128)	393344
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8256
dense_2 (Dense)	(None, 12)	780
Total params: 403,832		
Trainable params: 403,832		
Non-trainable params: 0		

Otros resultados



Resultados con aprendizaje de máquina

Super Vector
Machine (SVC)

20.04%

Random Forest
Classifier (RFC)

21.9%

Gracias por su atención