



RECONOCIMIENTO DE INDIVIDUOS POR SU VOZ USANDO REDES NEURONALES PROFUNDAS

Karina Mariela Figueroa Mora¹, Jonathan García² y José Antonio Camarena Ibarrola³

1 Fac. de cs físico-matemáticas, 2 Fac. de cs. Físico-matemáticas, 3 Fac. de Ingeniería Eléctrica.
karina.figueroa@umich.mx

El reconocimiento de individuos es una tarea necesaria en muchos contextos de nuestras vidas, por ejemplo para restringir el acceso a ciertas zonas, el poder autorizar realizar operaciones importantes. En particular, se sabe que la voz de un individuo es una característica que permitiría la identificación del mismo. Existen sistemas de seguridad que ya usan la voz para identificar a un individuo por ejemplo, algunos sistemas bancarios te piden leer un texto específico para grabar la voz de ese individuo. Cuando se solicita el acceso, el individuo debe leer exactamente el mismo texto y el sistema deberá decir si es la misma persona o no. En casos como éste solo se lleva a cabo una validación del texto leído en comparación con los audios guardados al momento de contratar el servicio.

En este tipo de sistemas de reconocimiento existen básicamente dos esquemas: texto dependiente y no texto dependiente. En el esquema texto dependiente el individuo debe pronunciar un texto específico y el sistema deberá a qué individuo se parece más. En el caso del sistema no-texto-dependiente, una persona no tiene que pronunciar el mismo texto y aún así, reconocer de quién se trata.

Por otro lado, recientemente las redes neuronales profundas han alcanzado desempeños competitivos en distintos problemas. La idea general de este tipo de redes es originada en una neurona. Una neurona es un sistema con entradas y cada una con un peso (valor) asociado, la suma de las entradas multiplicadas por los pesos genera remarcar las características de la entrada (en este caso la voz). Una red neuronal es una configuración de muchas de ellas para conseguir la identificación. Este tipo de redes requiere dos etapas, la de entrenamiento (las neuronas deben aprender) y la consulta.

En este trabajo presentamos el uso de redes neuronales profundas para la identificación de un individuo por su voz. Para la fase del entrenamiento se usaron una base de datos de audio con 20 individuos diciendo 30 palabras diferentes. En este trabajo se mostrará se alcanzan hasta un 75% de efectividad.