

Computación Blanda

Soft Computing

Autor: Alejandro Agudelo Toro

IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia

Correo-e: alejandro.agudelol@utp.edu.co

Resumen— Este documento presenta un resumen del problema de reconocimiento de archivos de audio, los cuales corresponden a las voces de personas con un reconocido prestigio público. El sistema de reconocimiento de voz se conoce como el proceso de convertir un archivo de audio en texto o también se puede ver como aquel capaz de reconocer automáticamente qué persona está hablando, reconociéndose a través de una base de datos de señales de audio clasificadas previamente.

Palabras clave— Voz, Reconocimiento, Clasificación, Base De Datos, Sistemas, Software, Inteligencia Artificial.

Abstract— This document presents a summary of the problem of recognition of audio files, which correspond to the voices of people with recognized public prestige. The voice recognition system is known as the process of converting an audio file into text or it can also be seen as the one capable of automatically recognizing which person is speaking, recognizing itself through a database of previously classified audio signals.

Key Word— systems, networks, artificial intelligence, software, computing, research, industry, genetic, learning.

I. INTRODUCCIÓN

Los estudios de inteligencia artificial se basan en el desarrollo de sistemas de procesamiento de datos que imitan el comportamiento de la mente humana, con sistemas de decisión tras un aprendizaje previo, en esta ocasión se trata imitar el sentido auditivo de los humanos, La voz humana es definida como el Sonido que el aire expulsado de los pulmones produce al salir de la laringe, haciendo que vibren las cuerdas vocales. El habla consiste en combinar las unidades fónicas que la voz humana es capaz de generar para formar una lengua. Estos símbolos no son idénticos de un individuo a otro, pero poseen características comunes que los hace descifrables dentro de una lengua o dialecto. Los sistemas sobre el reconocimiento de voz son útiles en numerosos campos o ámbitos los cuales pueden ser: Seguridad, Comercio, Educación entre otros.

En lo siguiente se mostrarán técnicas usadas para este problema de reconocimiento de voz y otros aspectos importantes.

I.1 PROCESANDO GRANDES ARCHIVOS DE AUDIOS

Cuando la entrada es un archivo de audio largo, la precisión del reconocimiento de voz disminuye. Sabemos que la API de reconocimiento de voz de Google no puede reconocer archivos de audio largos con buena precisión. Por lo tanto, necesitamos procesar el archivo de audio en fragmentos más pequeños y luego alimentar estos fragmentos a la API. Hacer esto mejora la precisión y nos permite reconocer grandes archivos de audio.

I.2 DIVIDIENDO EL AUDIO

Una forma de procesar el archivo de audio es dividirlo en trozos de tamaño constante. Luego podemos alimentar estos fragmentos a la API y convertir la voz en texto concatenando los resultados de todos estos fragmentos. Este método es inexacto. Dividir el archivo de audio en trozos de tamaño constante podría interrumpir las oraciones intermedias y podríamos perder algunas palabras importantes en el proceso.

I.3 DIVIDIENDO EL AUDIO BASADO EN EL SILENCIO.

La otra forma es dividir el archivo de audio según el silencio. Los humanos hacen una pausa corta por un corto período de tiempo entre oraciones. Si podemos dividir el archivo de audio en fragmentos basados en estos silencios, entonces podemos procesar el archivo oración por oración y concatenarlos para obtener el resultado. Este enfoque es más preciso que el anterior porque nos cortamos oraciones intermedias y el fragmento de audio contendrá la oración completa sin interrupciones. De esta manera, no necesitamos dividirlo en trozos de longitud constante.

I.4 PROCESAMIENTO DE LA SEÑAL DE VOZ

La primera operación que debe realizarse para reconocer la voz es procesar la señal de voz de entrada al sistema, con objeto de extraer la información acústica relevante para la tarea que debemos realizar. Los rasgos o características que se deben extraer de la señal de voz, son el resultado de un largo proceso de investigación sobre diferentes procedimientos de parametrización de la voz. Planteándose como solución actual más extendida una parametrización de la envolvente espectral que incluye consideraciones perceptuales a partir del funcionamiento del oído. La señal de entrada puede venir acompañada por efectos perturbadores, los cuales se desea sean eliminados, para ello se ha generado tres técnicas fundamentales:

- **Detección robusta de voz.**
- **Reducción de ruido.**
- **Cancelación de ecos.**

REFERENCIAS

Referencias en la Web:

[1]

https://e-archivo.uc3m.es/bitstream/handle/10016/26403/TFG_Raquel_Donos_Garcia_Del_Castillo.pdf?sequence=1&isAllowed=y

[2]

<https://www.clubdetecnologia.net/blog/2020/python-reconocimiento-de-voz-en-grandes-archivos-de-audio/>

[3]

<http://www.frsn.utn.edu.ar/tecnicas3/problemas/Reconocimiento%20de%20voz.pdf>