

Reconocimiento de archivos de audio

audio file recognition

Autor: **Aldahir Rojas Lancheros**

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

Correo-e: aldahir.rojas1@utp.edu.co

Resumen— El reconocimiento de voz es un tipo de inteligencia artificial que trata de establecer una comunicación entre el hombre y los ordenadores o dispositivos inteligentes, a través del lenguaje humano. El mecanismo de reconocimiento de voz hace que el procesador sea capaz de descifrar la información que contiene la voz humana. Los primeros pasos en este aspecto comenzaron en los años 50, pero ha sido en la última década cuando se ha hecho un avance gigantesco en el reconocimiento del lenguaje natural.

Palabras clave— sistemas, redes, inteligencia artificial, software, computación, investigación, industria, genético, aprendizaje.

Abstract— Voice recognition is a type of artificial intelligence that tries to establish communication between man and computers or intelligent devices, through human language. The speech recognition mechanism makes the processor capable of deciphering the information contained in the human voice. The first steps in this regard began in the 1950s, but it has been in the last decade that a gigantic advance has been made in the recognition of natural language.

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

I. INTRODUCCIÓN

El reconocimiento automático del habla consta de dos procesos de aprendizaje diferenciados:

Por un lado un aprendizaje deductivo que consiste en la transferencia de conocimientos del hombre a un sistema informático;

Y por otro lado un aprendizaje inductivo, aquí se trata de que el sistema sea capaz de obtener esos conocimientos a través de ejemplos. La evolución de los sistemas de reconocimiento de voz hacen que la máquina pueda interpretar como un “sí” no solo si oye esa palabra, sino también si escucha expresiones equivalentes. Lo que facilita el reconocimiento del lenguaje natural tal y como lo haría un ser humano, con una exactitud por encima del 90%

Los programas o sistemas de reconocimiento automático de voz también se clasifican según los siguientes parámetros:

Según su entrenabilidad, aquí entraría en juego si un sistema puede usarse directamente o requiere un entrenamiento previo del usuario antes de usarse.

Según la dependencia del hablante, sería si el sistema debe entrenarse o no para cada usuario diferente.

Según la continuidad, si el sistema reconoce el habla natural o requiere de pausas entre palabras.

Según la robustez, si funciona correctamente en ambientes ruidosos o con otras voces de fondo o no.

Y según el tamaño del dominio, si el sistema puede o no reconocer el lenguaje de un dominio reducido o extenso, por ejemplo si está diseñado para una compañía de telefonía y está adaptado a determinado número de palabras relacionadas.

Estos avances se han plasmado en la mejora de las diversas tecnologías relacionadas con los programas de reconocimiento de voz, como pueden ser los que se usan en el ámbito médico para el dictado de recetas o informes médicos. La tecnología del reconocimiento automático de voz, aplicado a ámbitos como la radiología o la anatomía patológica, aumenta considerablemente la eficiencia del trabajo ahorrando considerables cantidades de tiempo y

Por último, la calidad del micrófono que se utilice para interactuar con la máquina va a ser determinante en el resultado final, por lo que es necesario conocer cuáles son las características de buen micrófono.

Decodificador acústico-fonético

Las fuentes de información acústica, fonética, fonológica y posiblemente léxica, con los correspondientes procedimientos interpretativos, dan lugar a un módulo conocido como decodificador acústico-fonético (o en ocasiones a un decodificador léxico). La entrada al decodificador acústico-fonético es la señal vocal convenientemente representada; para ello, es necesario que ésta sufra un preproceso de parametrización. En esta etapa previa es necesario asumir algún modelo físico, contándose con modelos auditivos y modelos articulatorios.

Modelo del lenguaje

Las fuentes de conocimiento sintáctico, semántico y pragmático dan lugar al modelo del lenguaje del sistema. Cuando la representación de la Sintaxis y de la Semántica

tiende a integrarse, se desarrollan sistemas de RAH de gramática restringida para tareas concretas.

Reconocimiento de una gramática restringida

El reconocimiento de la gramática restringida trabaja reduciendo las típicas frases reconocidas a un tamaño más pequeño que la gramática formal. Este tipo de reconocimiento trabaja mejor cuando el hablante proporciona respuestas breves a cuestiones o preguntas específicas: las preguntas de "sí" o "no", al elegir una opción del menú, un artículo de una lista determinada, etc. La gramática especifica las palabras y frases más típicas que una persona diría como respuesta rápida y después asocia esas palabras o frases a un concepto semántico. Por ejemplo, un "sí" puede entenderse cuando se oye un "sip", "vale", "yes" o "okey", y un "no" con un "nop", "nada" o "en absoluto".

Si el hablante dice algo que gramaticalmente no tiene sentido, el reconocimiento fallará. Normalmente, si el reconocimiento falla, la aplicación incitará al usuario a repetir lo que ha dicho y el reconocimiento se intentará de nuevo. Si el sistema está correctamente diseñado y es repetidamente incapaz de entender al usuario (debido a que no se ha entendido bien la pregunta, un acento cerrado, interferencias o demasiado ruido alrededor), se retirará y desviará la llamada a otro operador. La investigación muestra que las llamadas a las que se las pide replantear la pregunta o cuestión una y otra vez, en poco tiempo se frustran y se agitan.

Los modelos del lenguaje más complejos necesitan para su correcto funcionamiento grandes corpora de voz y de texto escrito para el aprendizaje y la evaluación de los correspondientes sistemas. Gracias a ellos, se pueden abordar gramáticas más complejas y acercarse al Procesamiento de lenguajes naturales.

Características de los sistemas existentes

Los sistemas comerciales han estado disponibles desde 1990. A pesar del aparente éxito de estas tecnologías, muy pocas personas utilizan el sistema del reconocimiento del habla en sus computadoras. Parece ser que muchos de los usuarios utilizan el ratón y el teclado para guardar o redactar documentos, porque les resulta más cómodo y rápido a pesar del hecho de que todos podemos hablar a más velocidad de la que tecleamos. Sin embargo, mediante el uso de ambos, el teclado y el reconocimiento del habla, nuestro trabajo sería mucho más efectivo.

Este sistema donde está siendo más utilizado es en aplicaciones telefónicas: agencias de viajes, atención al cliente, información etc. La mejora de estos sistemas de reconocimiento del habla ha ido aumentando y su eficacia cada vez es mayor.

Clasificación

Los sistemas de reconocimiento de voz pueden clasificarse según los siguientes criterios:

Entrenabilidad: determina si el sistema necesita un entrenamiento previo antes de empezar a usarse.

Dependencia del hablante: determina si el sistema debe entrenarse para cada usuario o es independiente del hablante.

Continuidad: determina si el sistema puede reconocer habla continua o el usuario debe hacer pausas entre palabra y palabra.

Robustez: determina si el sistema está diseñado para usarse con señales poco ruidosas o, por el contrario, puede funcionar aceptablemente en condiciones ruidosas, ya sea ruido de fondo, ruido procedente del canal o la presencia de voces de otras personas.

Tamaño del dominio: determina si el sistema está diseñado para reconocer lenguaje de un dominio reducido (unos cientos de palabras p. e. reservas de vuelos o peticiones de información meteorológica) o extenso (miles de palabras).

Usos y aplicaciones

Aunque en teoría cualquier tarea en la que se interactúe con un ordenador puede utilizar el reconocimiento de voz, actualmente las siguientes aplicaciones son las más comunes:

Dictado automático: El dictado automático es, hasta hoy, el uso más común de las tecnologías de reconocimiento de voz. En algunos casos, como en el dictado de recetas médicas y diagnósticos o el dictado de textos legales, se usan corpus especiales para incrementar la precisión del sistema.

Control por comandos: Los sistemas de reconocimiento de habla diseñados para dar órdenes a un computador (p.e. "Abrir Firefox", "cerrar ventana") se llaman Control por comandos. Estos sistemas reconocen un vocabulario muy reducido, lo que incrementa su rendimiento.

Telefonía: Algunos sistemas PBX permiten a los usuarios ejecutar comandos mediante el habla, en lugar de pulsar tonos. En muchos casos se pide al usuario que diga un número para navegar un menú.

Sistemas portátiles: Los sistemas portátiles de tamaño reducido, como los relojes o los teléfonos móviles, tienen unas restricciones muy concretas de tamaño y forma, así que el habla es una solución natural para introducir datos en estos dispositivos.

Sistemas diseñados para discapacitados: Los sistemas de reconocimiento de voz pueden ser útiles para personas con discapacidades que les impidan teclear con fluidez, así como para personas con problemas auditivos, que pueden usarlos para obtener texto escrito a partir de habla. Esto permitiría, por ejemplo, que los aquejados de sordera pudieran recibir llamadas telefónicas.

Curiosidades

Los investigadores del grupo de reconocimiento de voz de Apple solían llevar una camiseta en la que se podía leer I helped Apple wreck a nice beach ("Yo ayudé a Apple a estropear una buena playa"), cuya pronunciación es prácticamente idéntica a I helped Apple recognize speech

("Yo ayudé a Apple a reconocer habla"). Esta broma ilustra la dificultad de desambiguar cadenas fonéticas.

REFERENCIAS

Referencias en la Web:

[1]

<https://computerhoy.com/reportajes/tecnologia/inteligencia-artificial-469917>

[2]

[https://es.wikipedia.org/wiki/Red_neuronal_artificial#:~:text=Las%20redes%20neuronales%20artificiales%20\(tambi%C3%A9n,entre%20s%C3%AD%20para%20transmitirse%20se%C3%B1ales.](https://es.wikipedia.org/wiki/Red_neuronal_artificial#:~:text=Las%20redes%20neuronales%20artificiales%20(tambi%C3%A9n,entre%20s%C3%AD%20para%20transmitirse%20se%C3%B1ales.)

[3]

<https://es.slideshare.net/mentelibre/logica-difusa-introduccion#:~:text=La%20l%C3%B3gica%20difusa%20es%20una%20t%C3%A9cnica%20de%20la%20inteligencia%20computacional,informaci%C3%B3n%20bien%20definida%20y%20precisa.>

[4]

https://es.wikipedia.org/wiki/L%C3%B3gica_difusa

[5]

<https://www.tecnologias-informacion.com/sistemas-expertos.html>

[6]

https://es.wikipedia.org/wiki/Algoritmo_gen%C3%A9tico#:~:text=En%20los%20a%C3%B1os%201970%2C%20de,y%20su%20base%20gen%C3%A9tica%20molecular.