

Reconocimiento del habla

El **reconocimiento automático del habla** (RAH) o **reconocimiento automático de voz** es una disciplina de la inteligencia artificial que tiene como objetivo permitir la comunicación hablada entre seres humanos y computadoras. El problema que se plantea en un sistema de este tipo es el de hacer cooperar un conjunto de informaciones que provienen de diversas fuentes de conocimiento (acústica, fonética, fonológica, léxica, sintáctica, semántica y pragmática), en presencia de ambigüedades, incertidumbres y errores inevitables para llegar a obtener una interpretación aceptable del mensaje acústico recibido.

Un sistema de reconocimiento de voz es una herramienta computacional capaz de procesar la señal de voz emitida por el ser humano y reconocer la información contenida en ésta, convirtiéndola en texto o emitiendo órdenes que actúan sobre un proceso. En su desarrollo intervienen diversas disciplinas, tales como: la fisiología, la acústica, el procesamiento de señales, la inteligencia artificial y la ciencia de la computación.

Diseño de un sistema de RAH

Aprendizaje

Un aspecto crucial en el diseño de un sistema de RAH es la elección del tipo de aprendizaje que se utilice para construir las diversas fuentes de conocimiento. Básicamente, existen dos tipos:

- **Aprendizaje deductivo:** Las técnicas de Aprendizaje Deductivo se basan en la transferencia de los conocimientos que un experto humano posee a un sistema informático. Un ejemplo paradigmático de las metodologías que utilizan tales técnicas lo constituyen los Sistemas Basados en el Conocimiento y, en particular, los Sistemas Expertos.
- **Aprendizaje inductivo:** Las técnicas de Aprendizaje Inductivo se basan en que el sistema pueda, automáticamente, conseguir los conocimientos necesarios a partir de ejemplos reales sobre la tarea que se desea modelizar. En este segundo tipo, los ejemplos los constituyen aquellas partes de los sistemas basados en los modelos ocultos de Márkov o en las redes neuronales artificiales que son configuradas automáticamente a partir de muestras de aprendizaje.

En la práctica, no existen metodologías que estén basadas únicamente en el Aprendizaje Inductivo, de hecho, se asume un compromiso deductivo-inductivo en el que los aspectos generales se suministran deductivamente y la caracterización de la variabilidad inductivamente.

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

- Francisco Casacuberta Nolla. Congreso de la Lengua Española, Sevilla, 1992 ^[1]
- Asunción Moreno. Congreso de la Lengua Española, Sevilla, 1992 ^[2]
- <http://www.tldp.org/HOWTO/Speech-Recognition-HOWTO/>
- Tordera Yllescas, Juan Carlos (2011): "Lingüística computacional. Tratamiento del habla". Valencia: Universitat de València.

Enlaces externos

- Reconocedor y Analizador de Voz. Fernando Peralta, Anibal Cotrina. Universidad Nacional Mayor de San Marcos, Oct 2002 ^[3].
- Voice tools Project ^[4].
- Guía de programas de síntesis y reconocimiento de voz para Linux ^[5] (en inglés)
- *Proyecto de una base de datos acústicos de la lengua española*. Joaquim Llisterri, Dolors Poch. Congreso de la Lengua Española, Sevilla, 1992 ^[6]
- Configurar el reconocimiento de voz en Windows ^[7]
- Reconocimiento de voz de Windows Vista: ¿mejor, igual o peor que Dragon Naturally Speaking? ^[8] Comandos de voz De iOS ^[9] Siri

Referencias

- [1] http://cvc.cervantes.es/obref/congresos/sevilla/tecnologias/mesaredon_casacuberta.htm
 - [2] http://cvc.cervantes.es/obref/congresos/sevilla/tecnologias/mesaredon_moreno.htm
 - [3] http://rav.jorelweb.com/rav_ini_v.php
 - [4] <http://www.eclipse.org/vtp/>
 - [5] <http://linux-sound.org/speech.html>
 - [6] http://cvc.cervantes.es/obref/congresos/sevilla/tecnologias/ponenc_llisterripoch.htm
 - [7] <http://windows.microsoft.com/es-es/windows-vista/set-up-speech-recognition>
 - [8] http://medtrad.org/panacea/IndiceGeneral/n29_tribuna-Serrahima2.pdf
 - [9] <http://www.apple.com.mx/ios/siri>
-

Fuentes y contribuyentes del artículo

Reconocimiento del habla *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=75445663> *Contribuyentes:* Abr pr90, Acotrin, Ainara Torres, An17a, Biasoli, Carl dnf, Cinevoro, Comae, Cordwainer, Diamondland, Diego.viola, Diegusjames, Digigalos, Engoal, Enric Naval, Erwin, Especiales, F7peralta, FAR, Fadesga, Farisori, Flashlack, GermanX, HIPATIA2006, Ignacio Icke, Jynus, Kabri, Lagoset, LeCire, Lilo197, Lionelrovezzi, Lucien leGrey, Matdrodes, Nahia de miguel diez, Osmaga, Pilaf, Pólux, Queninosta, Santiagocapel, Sirpuppet, Taichi, Tano4595, Technopat, YonDemon, Zatl3r, 59 ediciones anónimas

Licencia

Creative Commons Attribution-Share Alike 3.0
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)