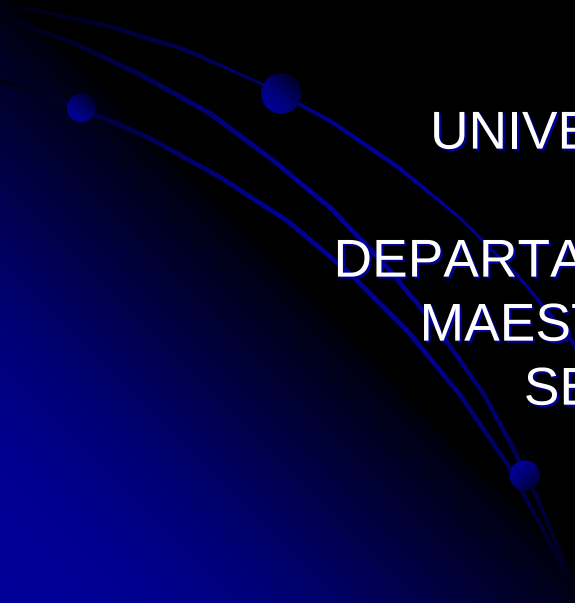


RECONOCIMIENTO AUTOMÁTICO DE VOZ (ESTADO DEL ARTE)

DEIBY ALEXANDER FANDIÑO R



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS
MAESTRÍA EN INGENIERÍA DE SISTEMAS
SEMINARIO DE INVESTIGACIÓN I

AGENDA

- Justificación
- Definición del problema
- El reconocimiento de voz
- **Técnicas de diseño**
 - Dynamic time warping
 - Modelos ocultos de markov
 - Redes neuronales
- Aplicaciones
- Conclusiones

Justificación

- Interfaces hombre-máquina controlados por la VOZ
- Por medio del reconocimiento de habla el ordenador sería capaz de extraer información de los comandos orales del operador o usuario, para realizar todo tipo de tareas y procesamiento de datos.
- Hacer la comunicación más rápida, y más agradable para los nuevos usuarios

Definición del problema

- *El Locutor.* Es quizás el aspecto que introduce mayor variabilidad en la forma de onda entrante.
- *La forma de hablar.* Es el segundo factor que determina la complejidad de un reconocedor de habla
- *El Vocabulario.* Se conoce por tal el número de palabras diferentes que debe reconocer el sistema
- *La Gramática.* Es el conjunto de reglas que limita el número de combinaciones permitidas de las palabras del vocabulario
- *El Entorno físico.* Es una parte tan importante como las anteriores para definir el reconocedor

El reconocimiento de voz

- Apropiaada interacción hombre-máquina a través de órdenes habladas
- áreas de trabajo:
 - Proceso de la señal de voz.
 - Técnicas de reconocimiento de patrones.
 - Diferentes estilos de habla.
 - Dependencia del locutor.
 - Vocabulario de reconocimiento.
 - Tarea de reconocimiento.
 - Bases de datos para entrenamiento y reconocimiento.

Técnicas de diseño(1)

- Técnicas topológicas: Dynamic Time Warping (DTW), basado en el cálculo y comparación de distancias.
- Técnicas probabilísticas: Modelos ocultos de Markov (HMM), que son modelos generativos de las palabras del vocabulario.
- Redes neuronales.
- Sistemas basados en el conocimiento: reconocedores por reglas o sistemas expertos.

Técnicas de diseño(2)

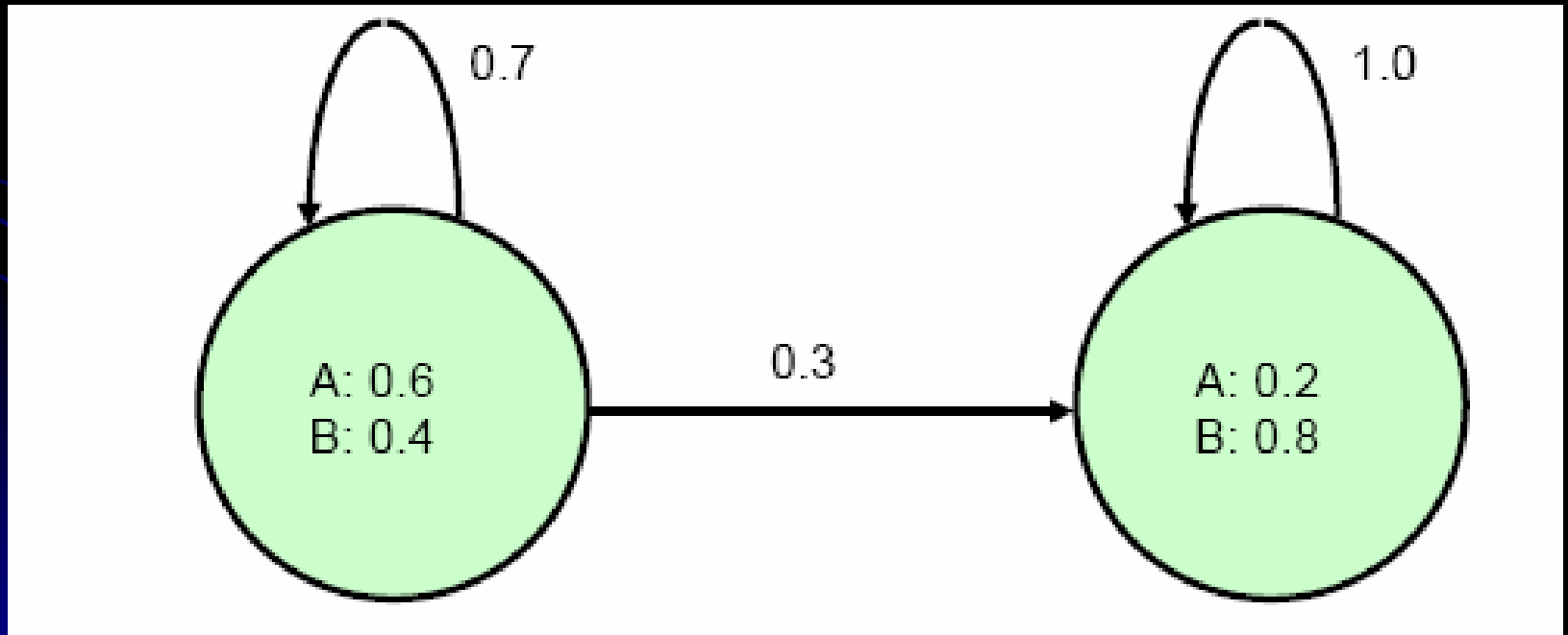
- Dynamic Time Warping (DTW)
 - se parametriza la señal de voz a reconocer
 - Se extrae un conjunto de parámetros (que pueden ser acústicos o coeficientes espectrales)
 - Las plantillas obtenidas se compararan en el proceso de reconocimiento con las plantillas de referencia.

Técnicas de diseño(3)

- La manera más fácil de reconocer una palabra aislada es compararla con un conjunto de plantillas previamente almacenadas y determinar cual es la que proporciona un mejor encaje.
- Sin embargo, este objetivo se complica por dos factores:
 - La duración de la palabra no tiene que ser la misma que la de las plantillas.
 - El ritmo con el que pronuncia esa palabra no tiene por que ser constante.

Técnicas de diseño(4)

- Modelos ocultos de Markov (HMM)
- Es una colección de estados conectado por transiciones.



Técnicas de diseño(5)

- El proceso comienza en un estado/s diseñado/s para tal efecto.
- En cada paso de ejecución se toma una transición a un nuevo estado y se genera un símbolo de salida asociado a ese estado.
- La elección de la transición y del símbolo es aleatoria y está gobernada por distribuciones de probabilidad.
- Un modelo oculto de Markov puede considerarse como un caja negra donde la secuencia de símbolos de salida generados a lo largo del tiempo es visible, pero la secuencia de estados por los que se ha pasado para generar la anterior serie se desconoce.
- Esto es por lo que se llaman modelos ocultos.

Técnicas de diseño(6)

- **Cuando se aplican los modelos ocultos de Markov al reconocimiento de la voz:**
 - Los estados se interpretan como modelos acústicos, indicando las ocurrencias de sonidos que son más probables durante los correspondientes segmentos de habla.
 - Las transiciones incorporan restricciones de tipo temporal acerca de cómo son las secuencias de aparición de esos sonidos.

Técnicas de diseño(7)

- Redes neuronales

- Los modelos computacionales basados en redes neuronales surgieron hace ya relativamente bastante tiempo, pero se abandono su estudio por no disponer de algoritmos eficientes de entrenamiento. Ahora ya no existe esa dificultad, y se ha demostrado ampliamente su enorme potencia computacional.

Técnicas de diseño(8)

- Se debe tener una base de conocimiento
- El espectro de los fonemas se clasifican según sus características (espectro, ancho de banda), y se obtienen vectores con pesos determinados.
- Se entrena la red
- la salida de la red es comparada con la salida esperada (y conocida de antemano), calculándose un error.

Aplicaciones

- Entrada de datos. Por ejemplo, datos en una inspección de fabrica, o el llenar un informe medico.
- Control de sistemas. El control de una central telefónica por medio de la voz, o el panel de mandos de un avión, son ejemplos de estas aplicaciones.
- Comunicaciones: marcación por voz como la más elemental, o la sustitución de una operadora como la más compleja.

Conclusiones

- Es una área en la cual todavía hay mucho por investigar.
- No existe un reconocedor de propósito general, el cual sea eficiente en cualquier contexto lingüístico
- Aunque en la actualidad los desarrollos en el área son significativos, por la naturaleza del problema las soluciones son temporales.