



Planificaciones

8653 - Procesamiento del Habla

Docente responsable: ESTIENNE CLAUDIO FRANCISCO

OBJETIVOS

El principal objetivo es introducir al alumno en las técnicas usadas actualmente en procesamiento de la voz, dedicando particular atención en el problema del reconocimiento del habla. El contenido del curso se divide en tres partes. En la primera se describen los métodos básicos del análisis y modelado de la señal acústica de habla. En la segunda se estudian los principales métodos de modelado estadístico del habla basados en el paradigma de los Modelos Ocultos de Markov. Los mismos permiten la implementación de los sistemas de reconocimiento de habla actuales. En la tercer parte se cubren diferentes tópicos avanzados, algunos de ellos relacionados con la mejora de los sistemas de reconocimiento de voz y otros relacionados con otras áreas del procesamiento del habla. Como objetivo anexo se plantea que el alumno aplique los conceptos de procesamiento de señales y procesamiento estadístico vistos en otras asignaturas a un problema concreto de ingeniería como es el reconocimiento de habla.

CONTENIDOS MÍNIMOS

PROGRAMA SINTÉTICO

Teoría Acústica de la producción del habla. Fisiología de la fonación. Modelización del tracto vocal. Fonos y fonemas. Representación de la señal de habla. Espectrograma. Percepción auditiva. Modelización del sistema auditivo. Extracción de características del habla: Predicción lineal, Cepstrum y Predicción Lineal Perceptual. Clasificación de patrones. Modelos Ocultos de Markov. Algoritmo forward-backward. Algoritmo de Viterbi. Estimación de parámetros. Algoritmo EM. Mezclas de gaussianas. Modelos de lenguaje. Modelos de N-gramas. Reconocimiento de palabras aisladas y continuas. Modelización fonética. Modelos contexto dependientes. Suavizado de parámetros. Sistemas de gran vocabulario. Optimización del algoritmo de Viterbi. Síntesis de voz. Identificación y verificación del hablante. Adaptación. Ruido. Detección de palabras claves.

PROGRAMA ANALÍTICO

TEORÍA ACÚSTICA DE LA PRODUCCIÓN DEL HABLA

Fisiología de la fonación. Mecanismo básico de la fonación. Modelización del tracto vocal. El modelo de tubo sin pérdidas. Ecuación de onda, efectos de radiación y excitación. Modelo digital. Relación con los modelos de filtros AR y ARMA.

NOCIONES DE FONÉTICA ACÚSTICA

Categorías lingüísticas. Fonos y fonemas. Clasificación de los fonemas. Características particulares de los fonemas en español y en español rioplatense.

REPRESENTACIÓN DE LA SEÑAL DE HABLA

Análisis de Fourier. Transformada de Fourier de corto tiempo. Espectrogramas. Caracterización de los sonidos del habla mediante espectrogramas. Percepción auditiva, fisiología de la audición. El sistema auditivo como banco de filtros. Percepción de la entonación. Nociones de psicoacústica. Enmascaramiento.

EXTRACCIÓN DE CARACTERÍSTICAS DEL HABLA

Caracterización de la señal de habla mediante coeficientes de predicción lineal (LPC). Relación con el modelo acústico del tracto vocal. Determinación de la trayectoria de los formantes y de la frecuencia glótica mediante coeficientes LPC. Aplicaciones a codificación. Vocoders. Análisis Cepstral. Aplicación a la señal de habla. Extracción de la frecuencia glótica mediante coeficientes Cepstrum. Representación de la señal de habla mediante coeficientes cepstrum en escala Mel. Determinación de coeficientes mel-Cepstrum (MFCC). Otras formas de caracterizar el habla basadas en modelos perceptuales. Predicción Lineal Perceptual (PLP).

CLASIFICACIÓN DE PATRONES

Técnicas de clustering. Método K-means. Clustering jerárquico. Métodos divisivos y aglomerativos. Clasificadores paramétricos y no paramétricos. Clasificadores de mínimo error. Clasificadores MAP y clasificadores de máxima verosimilitud. Clasificación no supervisada. El algoritmo EM. Reducción de la dimensionalidad. Test de significancia.

MODELOS OCULTOS DE MARKOV

Definición de un Modelo Oculto de Markov (Hidden Markov Models, HMM). Los tres problemas de los HMM. Evaluación de la probabilidad de observación. El algoritmo forward-backward. Determinación de la secuencia

oculta (decodificación) mediante alineamiento temporal dinámico (Dynamic Time Warping, DTW) y mediante el algoritmo de Viterbi. Estimación de los parámetros de un HMM mediante el algoritmo Expectación- Maximización (EM).

MODELIZACION ACUSTICA DEL HABLA I

Reconocimiento de habla usando HMM. Modelización de la probabilidad de observación mediante mezclas de gaussianas. Modelos semi-continuos. El problema de la inicialización en la estimación de parámetros. Diferentes tipos de topología. Suavizado de parámetros.

MODELOS DE LENGUAJE

Modelos de N-gramas. Suavizado de modelos de N-gramas. Métodos de backoff y métodos interpolados. Método de Kneser-Ney. Métodos de Máxima Entropía.

BASES DE DATOS DE ENTRENAMIENTO

Análisis de diferentes tipos de tarea de reconocimiento de habla. La base TIMIT. Bases de datos en español. La base Latino 40. Bases de dictado y de habla espontánea. Incorporación del ruido. La base NOISEX.

MODELIZACIÓN ACÚSTICA DEL HABLA II

Modelización de palabras aisladas. Modelización de habla continua mediante concatenación de modelos fonéticos. Modelos de monofonos. Modelos contexto dependientes. El problema de la insuficiencia de datos. Técnicas de suavizado. Sistemas de pequeño y gran vocabulario. Reconocimiento de palabras aisladas. Reconocimiento de habla continua. Combinación de modelo acústico y modelo de lenguaje. Optimización del algoritmo de Viterbi para sistemas de gran vocabulario. Técnicas de beam-searching. Estimación de listas de N-mejores hipótesis.

TÓPICOS AVANZADOS

Normalización de Features.

Adaptación al hablante: métodos MAP y MLLR.

Técnicas de procesamiento robusto. Adaptación al ruido. Substracción Espectral, PMC. Técnicas de normalización del hablante y del canal.

Identificación y Verificación del hablante.

Detección de tópicos y detección de palabras claves. Diarización.

Síntesis de voz. Sistemas de conversión de texto a voz.

Sistemas de diálogo.

Sistemas de medidas de confianza.

Prosodia y detección de emociones.

BIBLIOGRAFÍA

Bibliografía Básica

Huang, Acero, and Hon. Spoken Language Processing. Upper Saddle River, NJ: Prentice-Hall, 2001. ISBN: 0130226165.

Jelinek. Statistical Methods for Speech Recognition. Cambridge, MA: MIT Press, 1998. ISBN: 0262100665.

Rabiner & Juang. Fundamentals of Speech Recognition. Upper Saddle River, NJ: Prentice-Hall, 1993. ISBN: 0130151572.

Speech and Audio Signal Processing: Processing and Perception of Speech and Music by Ben Gold, Nelson Morgan and Dan Ellis. 2011

Digital Processing of Speech Signals by Lawrence R. Rabiner and Ronald W. Schafer (1978)

Bibliografía complementaria

Duda, Hart, and Stork. Pattern Classification. New York, NY: Wiley & Sons, 2000. ISBN: 0471056693.

Pattern Recognition and Machine Learning (Information Science and Statistics) by Christopher M. Bishop (2007)

Stevens. Acoustic Phonetics. MIT Press, 1998. ISBN: 0262692503.

Artículos de publicaciones y conferencias especializadas

IEEE, Transactions on Audio, Speech and Language Processing

IEEE, Proceedings of the IEEE

Computer Speech and Language (Elsevier)

Speech Communication (Elsevier)

Proceedings de ICASSP (IEEE)

Proceedings de INTERSPEECH (ISCA).

RÉGIMEN DE CURSADA

Metodología de enseñanza

La materia consta de dos clases semanales de tres horas de duración cada una, que se desarrollan en un laboratorio equipado con computadoras. La primera hora y media se dedicará a conceptos teóricos que se van complementado con ejemplos computacionales. Durante la siguiente hora y media se trabaja sobre ejercicios mayormente de simulación sobre los temas desarrollados en forma teórica. Las simulaciones se realizan sobre dos tipos de plataforma. La primera es un lenguaje tipo MATLAB, octave o pylab. Sobre dicha plataforma se trabajará en la primera parte del curso enfocada al análisis de la señal de habla. En la segunda se utilizará el toolkit HTK, licenciado en forma gratuita por Microsoft y desarrollado y mantenido por la Universidad de Cambridge, para el diseño y construcción de sistemas de reconocimiento de habla.

Modalidad de Evaluación Parcial

La modalidad de evaluación de la materia será mediante la realización individual de cuatro trabajos prácticos. Los mismos se evaluarán en fechas establecidas a lo largo del cuatrimestre. En todos los casos el alumno será evaluado en forma individual en la forma que el docente lo crea conveniente. Es condición para mantener la regularidad de la materia la asistencia al 75 por ciento de las clases (teóricas y prácticas) y la aprobación de los trabajos prácticos en las fechas estipuladas. El último trabajo práctico tiene carácter de integrador, por lo que la aprobación del mismo, y el cumplimiento de los requisitos de asistencia y aprobación del primer trabajo práctico, son suficientes para la aprobación de la materia.

Nota: Esta materia no contempla la modalidad de examen final o coloquio integrador. Si bien el segundo último práctico tiene carácter integrador en el sentido de integración de conocimientos adquiridos, el mismo deberá ser aprobado en las fechas estipuladas dentro del período de cursada de la materia.

CALENDARIO DE CLASES

Semana	Temas de teoría	Resolución de problemas	Laboratorio	Otro tipo	Fecha entrega Informe TP	Bibliografía básica
<1> 09/03 al 14/03	Introducción general de la materia Teoría acústica de la producción del habla		Repaso de procesamiento digital de señales			
<2> 16/03 al 21/03	Modelo de predicción lineal Fonética		Análisis espectral. Análisis de parámetros del habla mediante espectrogramas			
<3> 23/03 al 28/03	Fisiología de la audición, análisis cepstral. Predicción lineal perceptual, estimación de pitch y formantes.		Análisis de parámetros del habla mediante espectrogramas Análisis de fenómenos perceptuales			
<4> 30/03 al 04/04	Clasificación de patrones, clasificación no supervisada, clustering Clasificación de patrones, clasificación paramétrica		TP1 explicación			
<5> 06/04 al 11/04	Clasificación de patrones, algoritmo EM. Programación dinámica DTW.		TP1 evaluación			
<6> 13/04 al 18/04	Clasificación de patrones, algoritmo EM. Programación dinámica DTW.		TP2. Explicación			
<7> 20/04 al 25/04	Modelos ocultos de Markov (HMM). Presentación del modelo HMM Algoritmo Forward-Backward.		TP2 evaluación			
<8> 27/04 al 02/05	Modelos ocultos de Markov (HMM). Presentación del modelo HMM Algoritmo Forward-Backward.		Introducción al uso del sistema HTK para modelización de HMMs.			
<9> 04/05 al 09/05	HMM Estimación de parámetros, algoritmo EM HMM Estimación de la secuencia óptima, Algoritmo de Viterbi		TP3 explicación			
<10> 11/05 al 16/05	HMM Estimación con mezcla de gaussianas. Modelización acústica. Topología de modelos		TP3 evaluación			
<11> 18/05 al 23/05	Modelización acústica. Modelos de monófonos Modelización acústica. Modelos contexto dependiente		Implementación de un sistema de reconocimiento de 10 palabras continuas			
<12> 25/05 al 30/05	Sistemas de reconocimiento de palabras aisladas. Modelos de lenguaje		Implementación de un sistema de reconocimiento de habla continua (5000 palabras)			
<13> 01/06 al 06/06	Sistemas de reconocimiento de habla continua de gran vocabulario.		Implementación de un sistema de reconocimiento de habla continua de gran vocabulario.			
<14> 08/06 al 13/06	Sistemas de reconocimiento de habla continua de gran vocabulario.		Implementación de un sistema de reconocimiento de habla continua de gran vocabulario.			
<15> 15/06 al 20/06	TP especial 4. Realización		TP especial 4. Realización			
<16>	TP especial 4		TP especial 4. Realización			

Semana	Temas de teoría	Resolución de problemas	Laboratorio	Otro tipo	Fecha entrega Informe TP	Bibliografía básica
22/06 al 27/06	. Realización					

CALENDARIO DE EVALUACIONES

Evaluación Parcial

Oportunidad	Semana	Fecha	Hora	Aula
1º				
2º				
3º				
4º				
Observaciones sobre el Temario de la Evaluación Parcial				
Los trabajos prácticos especiales mencionados en el calendario y en la modalidad de evaluación, equivalen a la evaluación parcial y coloquio integrador. Las fechas de realización y evaluación se hallan estipuladas en el calendario.				