

Guia docent

230623 - SLPDL - Processament de la Veu i el Llenguatge amb Aprenentatge Profund

Última modificació: 24/05/2024

Unitat responsable: Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

Unitat que imparteix: 739 - TSC - Departament de Teoria del Senyal i Comunicacions.

Titulació: MÀSTER UNIVERSITARI EN ENGINYERIA DE TELECOMUNICACIÓ (Pla 2013). (Assignatura optativa).
MÀSTER UNIVERSITARI EN TECNOLOGIES AVANÇADES DE TELECOMUNICACIÓ (Pla 2019). (Assignatura optativa).

Curs: 2024

Crèdits ECTS: 5.0

Idiomes: Anglès

PROFESSORAT

Professorat responsable: FRANCISCO JAVIER HERNANDO PERICAS

Altres: Segon quadrimestre:
ANTONIO JESUS BONAFONTE CAVEZ - 30
FRANCISCO JAVIER HERNANDO PERICAS - 30

CAPACITATS PRÈVIES

Càlcul i Àlgebra Lineal
Probabilitat i Estadística
Algorísmia i Programació
Processat del Senyal
Aprenentatge automàtic
Aprenentatge profund
Àmplia experiència en programació en Python, Matlab o C++.

METODOLOGIES DOCENTS

Classes magistrals i un convidat extern.
Assignació de treballs teòrics i pràctics agrupats per temàtica.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

El curs està enfocat a estudiar les tecnologies que permeten el desenvolupament d'aplicacions avançades del llenguatge parlat i escrit. Es posa especial atenció en les noves tecnologies basades en l'aprenentatge profund i la seva aplicació al reconeixement de la paraula, conversió de text a veu i traducció automàtica.
El projecte final dona als estudiants una informació addicional sobre un tema en particular, i també té com a objectiu ajudar a potenciar les seves pròpies habilitats de desenvolupament d'aplicacions o investigació.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANT

Tipus	Hores	Percentatge
Hores grup gran	39,0	31.20
Hores aprenentatge autònom	86,0	68.80

Dedicació total: 125 h

CONTINGUTS

1. Introducció al processament de la parla

Descripció:

Introducció a la tecnologia de la parla i aplicacions.

Fonaments de percepció i producció de veu.

Espectre de veu: STFT, Espectrograma.

Característiques: Cepstrum, MFCC, Pitch.

Tècniques: Quantificació vectorial, Gaussian Mixture Models.

Dedicació: 6h

Grup gran/Teoria: 6h

2. Aprenentatge Profund

Descripció:

Introducció a les tècniques d'aprenentatge profund i la seva aplicació al processat de la veu i del llenguatge.

Objectius específics:

Introducció a l'aprenentatge profund, xarxes neuronals, algoritmes d'aprenentatge, momemtum i rmsprop, regularització, representacions vectorials de paraules, xarxes neuronals recursives. Aplicacions en el processament del llenguatge natural.

Dedicació: 9h

Grup gran/Teoria: 9h

3. Reconeixement de veu

Descripció:

Pattern matching. Dynamic time warping

Models ocults de Markov: reconeixement de paraules aïllades

Reconeixement de parla contínua: modelat acústic. Modelatge de llenguatge

Aprenentatge profund en modelatge de llenguatge i reconeixement de veu

Eines

Dedicació: 10h

Grup gran/Teoria: 10h

4. Síntesi de veu

Descripció:

Processador lingüístic. Modelatge de la prosòdia.

Generació de formes d'ona.

Mètodes de concatenació i estadístics.

Aprenentatge profund en síntesi de veu.

Dedicació: 7h

Grup gran/Teoria: 7h



5. Traducció automàtica

Descripció:

Introducció a la traducció automàtica. Traducció Automàtica Estadística.
Traducció automàtica amb xarxes neuronals.
Traducció de veu.

Dedicació: 7h

Grup gran/Teoria: 7h

SISTEMA DE QUALIFICACIÓ

Examen parcial: 20%, Examen final: 20%, Treballs: 60%

BIBLIOGRAFIA

Bàsica:

- Huang, X.; Acero, A.; Hon, H-W. Spoken language processing: a guide to theory, algorithm and system development. Upper Saddle River: Prentice Hall, 2001. ISBN 0130226165.
- Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning [en línia]. Cambridge, Massachusetts: MIT Press, 2016 [Consulta: 22/06/2016]. Disponible a: <http://www.deeplearningbook.org/>. ISBN 9780262035613.