Course guides
230623 - SLPDL - Speech and Language Processing with Deep Learning

Unit in charge: Barcelona School of Telecommunications Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications.

Degree: MASTER'S DEGREE IN TELECOMMUNICATIONS ENGINEERING (Syllabus 2013). (Optional subject).
MASTER'S DEGREE IN ADVANCED TELECOMMUNICATION TECHNOLOGIES (Syllabus 2019). (Optional subject).

Academic year: 2020  ECTS Credits: 5.0  Languages: English

LECTURER
Coordinating lecturer: Rodriguez Fonollosa, Jose Adrian
Others: Rodriguez Fonollosa, Jose Adrian
Bonafonte Cavez, Antonio Jesus
Ruiz Costa-Jussa, Marta

PRIOR SKILLS
Calculus and Linear Algebra
Probability and Statistics
Algorithmics and programming
Signal Processing
Machine Learning
Deep Learning

REQUIREMENTS
Large programming experience in Python, Matlab or C++.

TEACHING METHODOLOGY
Lectures (slides) and an external invited talk (mandatory attendance)
Theoretical and practical assignments grouped in subjects
Research project, presented in written and oral form

LEARNING OBJECTIVES OF THE SUBJECT
The course is focused on the study of the technologies for the development of advanced applications of spoken and written language. It focuses on new technologies based on deep learning and its application to automatic speech recognition, text to speech, and machine translation.
The final project gives students additional information about a particular topic, and also aims to help boost their own skills in the development of applications or in research.
STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours large group</td>
<td>39,0</td>
<td>31.20</td>
</tr>
<tr>
<td>Self study</td>
<td>86,0</td>
<td>68.80</td>
</tr>
</tbody>
</table>

Total learning time: 125 h

CONTENTS

1. Introduction to speech processing

Description:
Introduction to speech technology and applications.
Fundamentals of speech perception and speech production.
Speech spectrum: STFT, Spectrogram.
Features: Cepstrum, MFCC, Pitch.
Techniques: Vector Quantizers, Gaussian Mixture Models.

Full-or-part-time: 6h
Theory classes: 6h

2. Deep Learning

Description:
Introduction to Deep Learning techniques and their application to speech and language processing.

Specific objectives:
Introduction to deep learning, neural nets, learning algorithms, momentum and rmsprop, regularization, word embeddings, recursive neural networks. Applications to Natural Language Processing.

Full-or-part-time: 9h
Theory classes: 9h

3. Automatic speech recognition

Description:
Pattern matching. Dynamic time warping
Hidden Markov models. Isolated word recognition
Large vocabulary continuous ASR: Acoustic modeling. Language modeling
Deep Learning for language modelling and automatic speech recognition.

Toolkits

Full-or-part-time: 10h
Theory classes: 10h
4. Speech synthesis

Description:
Linguistic processing. Prosody modeling
Waveform generation.
Concatenation and statistical methods.
Deep learning in speech synthesis.

Full-or-part-time: 7h
Theory classes: 7h

5. Machine Translation

Description:
Introduction to Machine Translation. Statistical Machine Translation
Neural Machine Translation
Speech translation.

Full-or-part-time: 7h
Theory classes: 7h

GRADING SYSTEM

Midterm exam: 25%, assignments: 30%, research work: 45%

BIBLIOGRAPHY

Basic: