270225 - POE - Spoken and Written Language Processing

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 739 - TSC - Department of Signal Theory and Communications
Academic year: 2019
Degree: BACHELOR'S DEGREE IN DATA SCIENCE AND ENGINEERING (Syllabus 2017). (Teaching unit Compulsory)
ECTS credits: 6  Teaching languages: Spanish

Prior skills
College Calculus, Linear Algebra
Basic Probability and Statistics
Large programming experience in Python
Machine Learning.
Introduction to Deep Learning

Degree competences to which the subject contributes

Basic:
CB4. That the students can transmit information, ideas, problems and solutions to a specialized and non-specialized public.
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy

Specific:
CE5. Design and apply techniques of signal processing, choosing between different technological tools, including those of Artificial vision, speech recognition and multimedia data processing.
CE6. Build or use systems of processing and comprehension of written language, integrating it into other systems driven by the data. Design systems for searching textual or hypertextual information and analysis of social networks.

Generic:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CG4. Identify opportunities for innovative data-driven applications in evolving technological environments.
CG5. To be able to draw on fundamental knowledge and sound work methodologies acquired during the studies to adapt to the new technological scenarios of the future.

Transversal:
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Teaching methodology
Lectures presenting new theoretical material and practical examples.
Theoretical and practical assignments grouped in subjects.
Research project, presented in written and oral form by the students.
Learning objectives of the subject

1. Know the most important deep learning technologies of interest in the processing of oral and written language.
2. The student must know the most important applications of speech and language technology.
3. The student must be able to select the most appropriate speech and language technology for a particular task or application.
4. Develop innovative applications that use speech technology appropriately.
5. El alumno debe ser capaz de identificar los parámetros fundamentales de la voz en el dominio temporal y frecuencial.
6. The student must know the most important mathematical and machine learning tools for the analysis of the voice as vector quantification (VQ), Gaussian mixture models (GMM) and hidden Markov models (HMM).
7. The student must know the techniques for statistical language modeling.

Study load

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Hours large group: 45h</th>
<th>30.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours small group: 15h</td>
<td>10.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 90h</td>
<td>60.00%</td>
</tr>
</tbody>
</table>
# Content

## Introduction to language and speech technologies and applications

**Degree competences to which the content contributes:**

**Description:**
- Applications of oral and written language processing. Social impact.
- Main blocks of a natural language processing system: speech recognition, natural language processing, text to speech conversion.
- Language as a sequence of words. Vector representation of words. One-hot encoding versus continuous-space representations.
- Word2vec: Continuous bag-of-words (CBOW) versus Continuous skip-gram. GloVe vectors. Structures and analogies in word vector representations.

## Language Modeling

**Degree competences to which the content contributes:**

**Description:**
- Statistical modeling based on N-grams.

## Contextual language representations

**Degree competences to which the content contributes:**

**Description:**
- General purpose language representations.
- Unsupervised training. Unidirectional and bidirectional systems.
- Main architectures: ULMfit, OpenAI GPT, ELMo, BERT, XLM. Applications.

## Neural Machine Translation

**Degree competences to which the content contributes:**

**Description:**
- Introduction to Machine Translation. Automatic quality evaluation: BLEU
- Neural Machine Translation.

## Introduction to automatic speech recognition

**Degree competences to which the content contributes:**

**Description:**
- Pattern matching. Dynamic time warping.
- Hidden Markov models. Isolated word recognition.
- Large vocabulary continuous ASR: Acoustic modeling, Language modeling, Search.
Speech synthesis

Degree competences to which the content contributes:

Description:
- Linguistic processing.
- Prosody modeling.
- Waveform generation.
- Concatenation methods.
# Planning of activities

| Topic development: Introduction to speech and language technology and applications | Hours: 18h  
Theory classes: 6h  
Practical classes: 0h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 10h |
|---|---|
| Description:  
Introduction to speech and language technology and applications.  
Word vectors |  |
| Specific objectives: | 2, 3 |

| Topic development: Language Modeling | Hours: 18h  
Theory classes: 6h  
Practical classes: 0h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 10h |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific objectives:</td>
<td>6</td>
</tr>
</tbody>
</table>

| Topic development: Automatic Speech Recognition | Hours: 20h  
Theory classes: 9h  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 1h  
Self study: 10h |
|---|---|
| Description:  
Automatic Speech Recognition |  |
| Specific objectives: | 5, 6, 7 |

| Topic development: Speech Synthesis | Hours: 16h  
Theory classes: 6h  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 10h |
|---|---|
| Description:  
Speech Synthesis |  |
### Specific objectives:

#### Topic development: Contextual language representations

**Hours:** 28h  
- Theory classes: 9h  
- Practical classes: 0h  
- Laboratory classes: 2h  
- Guided activities: 1h  
- Self study: 16h

#### Specific objectives:

1

#### Topic development: Neuronal Machine Translation

**Hours:** 18h  
- Theory classes: 6h  
- Practical classes: 0h  
- Laboratory classes: 2h  
- Guided activities: 0h  
- Self study: 10h

**Description:**  
Neuronal Machine Translation

#### Final project

**Hours:** 38h  
- Theory classes: 3h  
- Practical classes: 0h  
- Laboratory classes: 7h  
- Guided activities: 4h  
- Self study: 24h

**Description:**  
Final project

**Specific objectives:**  
3, 4
Qualification system

Course evaluation is based on three aspects:

- Two exams, a midterm exam and the final exam, to assess the knowledge acquired but the student on the topics worked on in theory and practice sessions (40%)
- Evaluation of laboratory assignments: (30%)
- Evaluation of the final project (30%)

Only the 40% of the grade corresponding to the exams can be reassessed. The new grade will replace the grade obtained in the two exams taken during the course, with the same weigh in the final grade.

Bibliography

Basic:


Complementary: