

Academic Year: (2024 / 2025)

Review date: 26-04-2024

Department assigned to the subject: Signal and Communications Theory Department

Coordinating teacher: PELAEZ MORENO, CARMEN

Type: Electives ECTS Credits : 6.0

Year : 4 Semester :

OBJECTIVES

Audiovisual data is of vital importance for the entertainment industry (digital media, television, radio, podcasts, video games, music, etc.) in which its combination with telecommunications has radically changed our lives, especially in the way we which we interact with said data (intelligent assistants such as Siri, Alexa, Google Assistant, etc).

In addition, these data are increasingly relevant in areas such as medicine where the increasing sophistication of sensing devices and even wearable devices such as smart watches, virtual and augmented reality glasses, etc., generate more and more types of data with immense potential for transforming society and creating new markets.

But the key technology that has become essential to transform this large amount of audiovisual data into useful information and knowledge is artificial intelligence or machine learning, including neural networks and deep learning.

Therefore, the objective of this subject is to provide students with theoretical and methodological knowledge on algorithms and methods for the analysis of audiovisual information, including retrieval and indexing of multimedia information for navigation and search, user profiling, opinion mining and positioning, personalization of recommendations, etc.

In addition, an eminently practical point of view will be adopted, providing the tools to put theoretical knowledge into practice in the laboratory, so that students end up being able to develop an audiovisual data analysis project based on machine learning.

Ultimately, this will allow connections to be made with the myriad of applications the business products and services they support (for example, various Google services, platforms like Twitter, Instagram, TikTok, Spotify, Netflix, YouTube, Twitch, Shazam, and more).

1. TRANSVERSAL/GENERIC COMPETENCES:

- 1.1. Personal work capacity.
- 1.2. Capacity for analysis and synthesis.
- 1.3. Ability to apply theoretical concepts in practical cases.
- 1.4. Skills related to group work and collaboration with other colleagues.
- 1.5. Skills related to making oral and written presentations.

2. SPECIFIC OBJECTIVES:

- 2.1. To understand the fundamentals of audio-visual data analysis and its applications.
- 2.2. To understand the basic methods of representation and description of speech, audio, image and video.
- 2.3. To understand the methods and technologies used for classification, detection, indexing, recovery, filtering, personalization or recognition of voice, audio, image or video.
- 2.4. Ability to design and implement the above methods and technologies in practical problems of automatic voice, audio, image and video analysis.

Audiovisual data is of vital importance for the entertainment industry (digital media, television, radio, podcasts, video games, music, etc.) in which its combination with telecommunications has radically changed our lives, especially in the way we which we interact with said data (intelligent assistants such as Siri, Alexa, Google Assistant, etc).

In addition, these data are increasingly relevant in areas such as medicine where the increasing sophistication of sensing devices and even wearable devices such as smart watches, virtual and augmented reality glasses, etc., generate more and more types of data with immense potential for transforming society and creating new markets.

But the key technology that has become essential to transform this large amount of audiovisual data into useful information and knowledge is artificial intelligence or machine learning, including neural networks and deep learning.

Therefore, the objective of this subject is to provide students with theoretical and methodological knowledge on algorithms and methods for the analysis of audiovisual information, including retrieval and indexing of multimedia information for navigation and search, user profiling, opinion mining and positioning, personalization of recommendations, etc.

In addition, an eminently practical point of view will be adopted, providing the tools to put theoretical knowledge into practice in the laboratory, so that students end up being able to develop an audiovisual data analysis project based on machine learning.

Ultimately, this will allow connections to be made with the myriad of applications the business products and services they support (for example, various Google services, platforms like Twitter, Instagram, TikTok, Spotify, Netflix, YouTube, Twitch, Shazam, and more).

1. TRANSVERSAL/GENERIC COMPETENCES:

- 1.1. Personal work capacity.
- 1.2. Capacity for analysis and synthesis.
- 1.3. Ability to apply theoretical concepts in practical cases.
- 1.4. Skills related to group work and collaboration with other colleagues.
- 1.5. Skills related to making oral and written presentations.

2. SPECIFIC OBJECTIVES:

- 2.1. To understand the fundamentals of audio-visual data analysis and its applications.
- 2.2. To understand the basic methods of representation and description of speech, audio, image and video.
- 2.3. To understand the methods and technologies used for classification, detection, indexing, recovery, filtering, personalization or recognition of voice, audio, image or video.
- 2.4. Ability to design and implement the above methods and technologies in practical problems of automatic voice, audio, image and video analysis.

DESCRIPTION OF CONTENTS: PROGRAMME

The modern information overload problem caused by the availability of enormous amounts of information through internet makes it necessary to design systems that allow us to find the information we search and filter or personalize the information according to our needs.

For this, it is essential to be able to automatically index not only textual content but also audio, voice, image or video, using methods based on the content itself or collaborative labeling such as that which takes place on social networks.

Topic 0. Overview of audiovisual data analysis.

Topic 1. Audiovisual data descriptors

Topic 2. Methods for the analysis of audiovisual data

Topic 3. Audiovisual data retrieval and filtering systems

Topic 4. Applications

LEARNING ACTIVITIES AND METHODOLOGY

Several types of learning activities are proposed: theoretical and practical lessons, lab assignments and final project.

Several methodologies will be adopted: theoretical lessons and problem-based learning (with different levels of supervision and guidance).

The following learning activities and methodologies are employed: Combined master and lab classes, flipped classes and final project.

Combined master and lab classes (3 ECTS):

Master classes provide an overview of the main theoretical & mathematical concepts of the representation and processing of audiovisual data along with the analytic tools employed for indexing and accessing audiovisual contents and for their profiling and automatic recommendation.

In these classes, lab examples will be introduced as part of the theoretical expositions: all the formative sessions (lab availability provided) will take place in the lab to imbricate practical examples within the explanations to add dynamism to the class. This is also beneficial to solve different background issues given the possibility to access this subject from all the degrees from the Telecommunications family.

Moreover, every unit will begin with a debate of its technological implications. For this purpose, flipped classroom methodologies will be employed. In particular, students will be provided with some selected videos in advance to motivate the debate together with a list of questions (sometimes controversial) that the instructor will not answer categorically to encourage discussions. In this way, we expect to awake the curiosity of the student on the materials that will be subsequently explained.

COLLABORATIVE LEARNING (1 ECTS)

A practical example of the use of collaborative methods learned in theory classes for content labeling will be built, to establish connections between applications and commercial services that benefit from the analysis of audiovisual data, as well as the milestones and failures of learned technologies. For this, collaborative concept mapping tools will be used.

FINAL PROJECT (2 ECTS)

Students will work on a project in which they will program a complete modular system of one of the tools explained in class. The students will be provided with some guidelines and some preparatory sessions by using problem-based learning.

ASSESSMENT SYSTEM

% end-of-term-examination:	0
% of continuous assessment (assignments, laboratory, practicals...):	100
Lab assignments and questionnaires (40%)	
Final Project (60%)	

BASIC BIBLIOGRAPHY

- C. D. Manning, P. Raghavan and H. Schultze Introduction to Information Retrieval, MIT press, 2008
- Rafael C. González and Richard E. Woods Digital Image Processing, Fourth Edition, Pearson, 2018

ADDITIONAL BIBLIOGRAPHY

- Ben Gold (Author), Nelson Morgan (Author), Dan Ellis (Author) Speech and Audio Signal Processing: Processing and Perception of Speech and Music, Wiley, 2011
- Dan Jurafsky and James H. Martin Speech and Language Processing (3rd ed.), Prentice Hall, 2018
- Li Deng (Editor), Yang Liu (Editor) Deep Learning in Natural Language Processing, Springer, 2018
- Ricardo Baeza-Yates, Berthier Ribeiro-Neto Modern Information Retrieval: the concepts and technology behind search, 2nd Edition, Pearson, 2011

- S. Theodoridis and K. Koutroumbas Pattern Recognition, 4th ed., Academic Press, 2008

- Wilhelm Burger and Mark J. Burge Principles of Digital Image Processing: Fundamental Techniques, Springer-Verlag, 2009

- Yu, Dong, Deng, Li Automatic Speech Recognition, Springer, 2015