2SC7990 – What you unwittingly say: decryption and automatic analysis of nonverbal behaviors

Instructors: Catherine Soladie
Department: DOMINANTE - MATHÉMATIQUES, DATA SCIENCES, DOMINANTE - INFORMATIQUE ET NUMÉRIQUE, DOMINANTE - VIVANT, SANTÉ, ENVIRONNEMENT
Language of instruction: ANGLAIS
Campus: CAMPUS DE RENNES
Workload (HEE): 80
On-site hours (HPE): 48,00

Description

What you unwittingly say: decryption and automatic analysis of nonverbal behaviors.

Letters, words, sentences: the algorithms we have today are more and more effective in decrypting our grammar, and understanding what we can say. And yet this only covers a tiny part of our communication.

Joy, resignation, irony: only our body and the tone of our voice reveal our deep intentions, our real message, and the automated understanding of such human behaviors and emotions is a big challenge.

To take it up, every year, the audio, video and machine learning communities gather around international challenges of research on the automatic analysis of human behaviors: emotions, depression, mood, motion detection, ... (ex: http://sspnet.eu/avec2017/).

Through this project, you will be able to face one of these challenges. You will have a large dataset representing people in action and will have to automatically determine their behavior and emotions.

Each project team focuses on a particular study (eg the voice, the face, ...), and all teams will gather their work to compare your results to those of the competitors of the international challenge.

Quarter number
ST7

Prerequisites (in terms of CS courses)

Statistics et machine learning.
Signal processing

Computer science :

- Algorithms
- Programming langages (basics)

Syllabus

Background (5%)

- Introduction by the research team.
- Group organization.
- Provision of the challenge data.

State of the art (20%)

- Research and understand research papers on the subject.
- Reproduce a selected subset of state-of-the-art methods (they will serve as a basis for your work).

Pre-processing, understanding and visualization of data (40%)

- Depending on the chosen topic (voice, face, ...), extract the interesting features for your analysis
- Explore the relevant visual representation modes
- Use these representations to guide your analysis strategy

Statistical analysis and learning (20%)

- Choose and build your analysis and learning models
- Quantify your results and compare them to the state of the art

Visibility points and final presentation (15%)

- 3 daily feedback will have to be carried out to present your progress as the project progresses (a different member of the team each time).
- Structure your presentations with the objectives, the state of the art, the architecture diagram, the results tables.
- At the end of the project, present as a team your results to our industrial and academic partners.
- Provide a scientific report

Class components (lecture, labs, etc.)

- Immersion in the FAST research team: supervision by researchers, PhD students and post-docs.
- Organization in teams of 2 to 5 students. If possible, coordination of the different teams for the production of a single overall final result.
- Presentation of the results to our partners.

Grading

Individual daily feedback: 1/4 of the mark
Defense in front of the partners: 1/4 of the mark
Scientific results (system performance): 1/4 of the mark
Scientific report: 1/4 of the mark

Resources

Teaching team:

- Catherine SOLADIE
- Renaud SEGUIER
- Simon LEGLAIVE
- PhD students of AIMAC research team

Software tools and number of licenses needed:

- TensorFlow or equivalent (free)

Learning outcomes covered on the course

At the end of this project, you will be able to:

- Specify or redefine the need (C4.1)
- Navigate among the research papers of a subject, read them and understand them (C2.4)
- Reproduce a selected subset of state-of-the-art methods in signal processing and / or machine learning (C3.2)
- Mix skills from signal processing, statistical analysis and machine learning to analyze data (C2.2)
- Explore visual representation modes that are relevant to your data (C6.3)
- Use these representations to guide your analysis strategy (C3.3)
- Choose and build your analysis and learning models (C1.2, C6.1)
- Quantify your results and compare them to the state of the art (C2.4, C3.3)
- Conduct a large-scale scientific project in a group (C8.1)
- Decrypt a set of non-verbal messages during human interactions (C7.4)
- Argue your scientific approach (C7.1)

Description of the skills acquired at the end of the course

- C2 Jalon 2
    - C2.4 **Données** : Exploiter un ensemble cohérent de données et réaliser un état de l'art exhaustif avec un esprit critique
- C6 Jalon 2
    - C6.3 **Traiter des données** : Mettre en œuvre des algorithmes traitant ou utilisant des données massives (intelligence artificielle, clustering)
- C4 Jalon 2
    - C4.1 **Besoin client** : Identifier avec le client les autres dimensions ne figurant pas dans la formulation initiale : techniques, économiques, humaines, etc.
- C7 Jalon 2
    - C7.1 **Convaincre sur le fond** : Adapter le fond et son argumentation en fonction d'interlocuteurs ou de contextes élargis, « avoir du répondant » pour défendre sa solution (maîtrise du sujet des interlocuteurs, valeurs, engagements, disponibilité, attention, etc.).
- C8 Jalon 2
    - C8.1 **Travailler en équipe** : Associer chaque membre de l'équipe en fonction de ses forces

ST7 – 80 – SOURCE SEPARATION FOR OPTIMAL SIGNAL PROCESSING

Department : MDS (Mathematics, Data Science)
Language of instruction: Français
Campus : Campus de Metz

Engineering problematic

In order to increase productivity, to reduce ecological impact or to improve the quality of services, manufacturers and research centers must respond to the issues of data analysis made up of overlapping signals. Indeed, in many applications, sensor signals are mixtures of several sources that emit at the same time, and the information carried by these signals can only be exploited if we manage to separate what comes from each source. For example, in oil exploration, companies are now carrying out acquisition campaigns in simultaneous source mode (several sources emit at the same time). It is then necessary to distinguish in the signals of the geophones the echoes due to the various sources to obtain an image of the subsoil. Or in a "hi-tech" store, a robot that has started a conversation with a person may be disturbed by other customers or by ambient noise, while it must continue its conversation with the same person. Or in electrocardiograms, to separate the signal of the fetus' heart from that of its mother. Or for sound recording in a conference or round table with several speakers, rather than giving a microphone to each speaker, it would be practical to separate what each speaker says from recordings of microphones placed in fixed positions. The solutions to these problems are based on the same mathematical concepts:

- The knowledge of a data representation space adapted to the problem, where data have a sparse representation (in other words the data can be represented by a vector which has a few non-zero coordinates, in a space of great dimension);
- the minimization of a criterion that depends on the data; and
- a learning and testing stage to verify that the obtained solution generalizes and thus avoid overfitting.

Necessary prerequisites

First year courses in probability, statistics, signal processing and algorithms; good skills in a programming environment (Matlab, Python, ...).

Context and issues modules: These modules include an introductory lecture on the theme, presentations on environmental and societal issues and on

technological and scientific challenges, as well as a presentation of related projects.

Specific course (60 HEE): *Sparse representations of signals*

- **Brief description:** This course presents the mathematical tools of signal analysis and their properties (complements on the Fourier transform, subsampling/oversampling, harmonic signal, STFT, multiresolution analysis, Paley-Littlewood and bi-orthogonal wavelet decompositions, perfect reconstruction filter banks) as well as signal decomposition methods (Matching Pursuit, Basis Pursuit, Independent Component Analysis).

Projet n°1 : *Tracking a speaker by a robot*

- **Associated partner:** ORANGE, Cognitive Computing
- **Campus**: Campus de Metz
- **Brief description:** Robots are increasingly present in our environment. When a robot has started a conversation with a speaker, the problem is to keep the focus on the interlocutor while several people are talking around the robot, or another interlocutor is talking to it. ORANGE wants to solve this problem by using a monophonic audio signal recorded by the robot, without adding other modalities.

  The issue is therefore to find one or more data representation spaces well adapted to the problem of speaker tracking; to learn from a small number of samples (i.e. over a small recording duration) the features of the speaker to be tracked; to avoid overfitting which may occur if the learned features depend on the words spoken by the speaker.

Projet n°2 : *Separation of sound sources from recordings of several microphones*

- **Associated partner:** CentraleSupélec, audio plateform of the « smartroom
- **Campus:** Campus de Metz
- **Brief description:** There are many concrete situations in which you want to capture a sound so you can either record it for replay or amplify it live so that all participants have a good perception. Dans certaines de ces situations, plusieurs sources peuvent intervenir (par exemple, une conférence ou table ronde dans laquelle plusieurs locuteurs sont présents).

  In order to allow good intelligibility in the case of several speakers, a microphone is usually placed in front of each speaker, or, in the theater, a radio transmitter microphone is placed directly on the actors. A significant improvement could be to use a fixed microphone array and to have a

processing algorithm to separate the sources, thus giving the illusion of having an individual microphone per speaker or actor.

The problem is therefore, starting from a fixed microphone array and assuming a finite number of sources (e.g. speakers), to ensure the separation of the sources and to provide an output channel per source (speaker), which channel would also contain the position information of this source. For simplicity, we will start by assuming that the speakers are at fixed positions known in advance. This is the case, for example, for speakers at a round table. In this case, the system could be calibrated beforehand using white noise generators, and then could operate using the fixed parameters determined in this way.

Projet n°3 : *Non-invasive foetal electrocardiogram extraction*

- **Associated partner:** INSERM
- **Campus**: Campus de Metz
- **Brief description:** Non-Invasive foetal electrocardiography (NI-FECG) represents an alternative foetal monitoring technique to traditional Doppler ultrasound approaches, that is non-invasive and has the potential to provide additional clinical information. However, despite the significant advances in the field of adult ECG signal processing over the past decades, the analysis of NI-FECG remains challenging and largely unexplored. This is mainly due to the relatively low signal-to-noise ratio of the FECG compared to the maternal ECG, which overlaps in both time and frequency.

The issue is to find one or several data representation spaces well adapted to the problem of foetal ECG extraction. It will be a question of applying and testing one or several methods starting from an article that reviews recent advances in NI-FECG research including: publicly available databases, NI-FECG extraction techniques for foetal heart rate evaluation and morphological analysis, NI-FECG simulators and the methodology and statistics for assessing the performance of the extraction algorithms..