

Time Series and Anomaly Detection

ECTS : 6

Description du contenu de l'enseignement :

This lecture is thought as an introduction to the analysis of complex data, and particularly to that having a temporal component. Methods aimed at exploring and modelling time series, longitudinal data and graphs with temporal components will be addressed. The issues of detecting patterns, breakpoints, changes of regimes, and anomalies will be at the core of the different approaches.

The first chapters will be devoted to hidden Markov models. After having briefly recalled some definitions and properties of Markov processes, we will define hidden Markov processes, illustrate them with several examples and give some of their properties. Inference techniques using the EM algorithm and Bayesian approaches will be presented and illustrated in practice. We will particularly focus on some specific models which are extremely useful for segmenting time series stemming from the economics field, such as autoregressive Markov switching models.

The second part of the lecture will tackle the issue of change-point detection methods. We will start by introducing the change-point detection issue. More specifically, we will consider several frameworks and derive inference procedures for computing and locating change-points : online vs. offline strategies, single vs. multiple change point detection, known vs. unknown number of change points, parametric vs. non-parametric approaches.

The third chapter will be aimed at introducing the issue of anomaly detection in the context of temporal data. After having defined what an anomaly is, we will start by assessing whether and how hidden-Markov models and change-point analysis may be useful for detecting anomalies. Then, we will compare these two approaches with other techniques, stemming either from the field of computational statistics, or from that of machine learning. During this chapter, we will also consider the questions of detecting patterns and clustering temporal data.

The fourth chapter will address data that can be modelled as a graph or a temporal graph. We will start by introduce some definitions and summaries for characterising the network (degree distribution, centrality indices, ...). Afterwards, we will tackle the questions of community detection and graph clustering. Eventually, we will address the issues of random networks and associated tests for randomness. Models and methods introduced in this lecture will be practiced using existing implementations in R and "real-life" datasets.

Compétence à acquérir :

Gain some background and perspective of time series analysis from a data science point of view.

Be able to handle temporal data subject to anomalies and change-points.

Have some basic knowledge about graphs and temporal graphs mining.

Mode de contrôle des connaissances :

Data challenge. A project to be done individually or by two, analysing real life data.