CEA LIST

Internship

Co-clustering of time series

Context

The French Atomic Energy and Alternative Energy Commission (CEA) is a public research organization that is a major actor in the European Research field with a growing international presence.

Within CEA Tech, the CEA LIST Institute focuses its research on intelligent digital systems. With major economic and societal challenges, its R&D programs focus on advanced manufacturing (robotics, virtual & augmented reality, non-destructive testing, vision), embedded systems (safety & security, software engineering and systems, computing architectures), ambient intelligence (sensors, instrumentation & metrology, communication & sensory interfaces, data processing & multimedia). By developing advanced technologies whose applications cover the transport, security/defense, manufacturing, energy and health sectors, CEA LIST contributes to the industrial competitiveness of its partners through innovation and technology transfer (www-list.cea.fr).

Within the CEA LIST Institute, the student will work in the Data and Decision Sciences Laboratory (LS2D), which includes about thirty people.

Topic

Co-clustering is an unsupervised machine learning method that aims to identify the homogeneous block structure of a data table from a joint classification of rows and columns. Since 1965, this problem has been developed in several variants but its interest has considerably increased in recent years with the arrival of many applications such as text data analysis, marketing analysis, genomics, recommendation systems and energy data study. This type of approaches simultaneously organizes the rows and columns of a table to discover homogeneous blocks aligned to provide a simplified reading of the data and/or to extract characteristics (feature engineering) used in machine learning models as prediction models. Among the methods developed, two types of approaches can be distinguished: matrix reconstruction based and model-based clustering.

The objective of the internship is to study the particular case of co-clustering for time series. In this context, recent approaches propose to first transform these data into functions to take into account the notion of order due to temporality. Then a co-clustering method is applied on these transforms. For example, an approach consists of applying the latent block model (model-based clustering) on the linear coefficients of the different functions. Another one employs this model on the projected coefficients of these functions.

During the internship, the student will study and empirically test these two approaches. The student will complete her/his study with a state of the art on the co-clustering of time series. Similarly to the two methods already developed, it may be interesting to propose an alternative solution coupling signal processing (Besoz transforms for example) and co-clustering. The internship can be followed by a PhD in which an alternative method, based on matrix reconstruction based, will be developed.

Keys words: co-clustering, machine learning, time series, signal processing

Environment and Prerequisites

- **Location:** The internship will take place at CEA Saclay, in the DIGITEO building
- **Duration:** 6 months. As the formalities required to recruit the candidate are quite long, it is recommended to start the process at least 3 months before the start of the internship.
- **Salary:** from 600€ according to profile
- **Prerequisites:** The candidate is in M2 specialized in machine learning and signal processing.
- **Contacts :**
  - Aurore LOMET, aurore.lomet@cea.fr
  - Antoine Souloumiac, antoine.souloumiac@cea.fr