Topic 15: Query by content for physiological signals

Submitter: Oudre Laurent
Laboratory: Centre Borelli

Query by content is a popular topic that is the core of several online applications such as Shazam or Google Images. Given a query image or sound, it consists in browsing a large database to retrieve the closest images or sounds up to a predefined measure of fit. The aim of this project is to build similar tools, but for physiological signals.

The Centre Borelli is a pluridisciplinary laboratory combining mathematicians and neuroscientists. In this context, we have gathered and recorded a large number of databases of physiological signals such as accelerometer data, ECG, EEG... in several protocols (gait analysis, anesthetization...). We want to build a system that, given a shape of interest (for instance the 3D shape of a step in gait analysis recorded with motion sensors, or a typical heartbeat in ECG data), is able to retrieve all occurrences of this shape in the database (or more specifically, all subsequences that are somehow close to the query). The main question is linked to the choice of the measure of fit, that must be accurate, relevant for the task and easy to compute. Several proposals have been made in the literature including normalized Euclidean distance [1] or Dynamic-Time-Warping [2]. Very efficient algorithms have been developed in this context and can retrieve shapes in databases composed of more than several trillions time series [2].

However, several problems still exist, in particular for physiological data. First, we want to retrieve multivariate shapes instead of simple univariate shape: for instance, 3D motion capture data only makes sense when patterns are discovered and compared in the original 3D space: unfortunately, fast computations for multivariate data are not available for now. Second, the comparisons we want to make may involve time series of different lengths (e.g. compare two recordings of different lengths or compare a query to subsequences of a time series) : although some very recent works provided an attempt to answer this question, it is still an open problem [3]. Third, the shapes we want to retrieve necessitate a time-consuming preprocessing step (e.g. if we want to compare a heartbeat to every heartbeat in a database, it implies that these heartbeats have been segmented beforehand) : we would like to be robust with respect to segmentation errors [4] and if possible, to remove the need for this preprocessing step.

One possible approach consists in starting from a model described by a jump process on a graph containing the complete information of the contacts and constructing a reduced population dynamics model integrating the size of the structures for each level of contacts selected, by reducing complexity in two related ways: reduction of the process dimension and approximation by deterministic dynamics. After a first step focused on defining the framework for model formalization, one or more relevant approximations of the initial complete model will be proposed. Their ability to approach the complete model dynamics will be analyzed at different temporal granularities and for various typologies (a "type" here defining an individual to whom a health state and a type of structure for each level of contacts considered are associated), while respecting a parsimony constraint (number of state variables or "types" that can be numerically managed). The reduced model(s) thus constructed can be used to explore, by simulation, various scenarios for controlling an epidemic, by preferentially targeting certain types of individuals or by specifically altering the interactions at different levels of the contact structure. This (these) model(s) can also be used to identify pertinent dynamical indicators of the epidemic evolution.

A reflection has already been carried out in the teams proposing this subject, a first prototype model with two levels of contact structures being currently under construction and analyzed. According to the profile and interest of the candidate, several directions are possible for the development of this project: on a theoretical side (approximation of epidemic processes on graphs...
with several levels of contacts), on a numerical side (comparison of scenarios, control, optimization) or concerning statistical aspects (model calibration from data).

This project is part of a collaboration between Elisabeta Vergu (MaIAGE, INRAE; epidemiological modeling) and Vincent Bansaye (CMAP, Ecole Polytechnique; probabilities), who will both co-supervise the post-doctorate.

References


