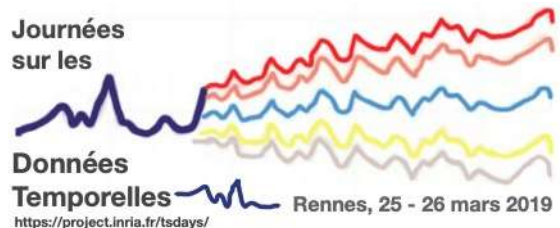# Journées données temporelles

Lundi 25 Mars 2019

14h30 [Themis Palpanas] *Indexation de séries temporelles*

There is an increasingly pressing need, by several applications in diverse domains, for developing techniques able to index and mine very large collections of sequences, or data series. Examples of such applications come from social media analytics and internet service providers, as well as from a multitude of scientific domains. It is not unusual for these applications to involve numbers of data series in the order of hundreds of millions to billions, which are often times not analyzed in their full detail due to their sheer size. However, no existing data management solution (such as relational databases, column stores, array databases, and time series management systems) can offer native support for sequences and the corresponding operators necessary for complex analytics. In this talk, we argue for the need to study the theory and foundations for sequence management of big data sequences, and to build corresponding systems that will enable scalable management and analysis of very large sequence collections. We describe recent efforts in designing techniques for indexing and mining truly massive collections of data series that will enable scientists to easily analyze their data. Finally, we present our vision for the future in big sequence management research, including the promising directions in terms of storage, distributed processing, and query benchmarks.

15h20 [Joyce-Madison Giacofci] *Functional data analysis: modelling curves variations*

Owing to the constant evolution of technologies many scientific studies lead nowadays to the collection of large amounts of data that are usually modelled as functional data. Functional data are characterized by their complexity, their underlying structure and their high dimensionality. After brief recall on functional data analysis, we will focus on the modelling of inter-individual curves deviations through functional mixed models and deformation models.

16h10 Pause

16h40 [Albert Bifet] *Machine Learning for Data Streams*

Big Data and the Internet of Things (IoT) have the potential to fundamentally shift the way we interact with our surroundings. The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. Advanced analysis of big data streams from sensors and devices is bound to become a key area of data mining research as the number of applications requiring such processing increases. Dealing with the evolution over time of such data streams, i.e., with concepts that drift or change completely, is one of the core issues in stream mining. In this talk, I will present an overview of data stream mining, and I will introduce some popular open source tools for data stream mining.

Mardi 26 Mars 2019

9h40 [Pierre-François Marteau] *Kernelized time elastic averaging of time series*

In the light of regularized dynamic time warping kernels, we re-considers the concept of time elastic centroid for a set of time series. More precisely, we will present an algorithm based on a probabilistic interpretation of kernel alignment matrices. This algorithm expresses the averaging process in terms of a stochastic alignment automata. It uses an iterative agglomerative heuristic method for averaging the aligned samples, while also averaging the times of occurrence of the aligned samples. We will demonstrate its use in some applications, namely dataset simplification, bootstrapping and time series denoising.

10h30 Pause

10h50 [José Lozano] *Time series data mining challenges*

Time series have gained much interest in the last decade. They appear naturally in industrial, medical or economical environments to name a few. Time series mining refers to the activities related with the extraction of knowledge from time series databases. Particularly, typical machine learning activities such as supervised classification or clustering are carried out from this kind of data. Furthermore, the timely nature of the data allows considering new problems. An example is the early classification of time series, where the objective is to classify the series as early as possible a before its end. In this talk we will review time series mining algorithms and pointed out to new avenues to do research in the area.

11h40 Minute madness (présentation posters)

- Learning Interpretable Shapelets for Time Series Classification through Adversarial Regularization, Wang Yichang, Irisa/Inria
- Detection of false data injection in transport systems using Machine Learning, Chevrot Antoine, FEMTO-ST / DISC
- Clustering de courbes de charge du réseau de distribution de l'électricité de SRD en vue de la minimisation des pertes, Tribak Mohamed, SRD Energies & LIAS Laboratory
- Robust Detection and Estimation of Change-Points in a Time Series of Multivariate Images, Mian Ammar, CentraleSupélec SONDRA
- Cyclone artificial intelligence prints investigation, Hochard Guillaume, QUANTMETRY
- Utilisation de données de logs dans les problèmes de Bandits Contextuels, Claeys Emmanuelle, IRMA/ICUBE
- Warp 10: The Most Advanced Time Series Platform, Vialatte Jean-Charles, SenX
- Comparaison de Modèles d'apprentissage Profond pour la prédiction de séries temporelles: application à l'effacement électrique dans les chambres froides, Mellouli Nédra, LIASD, Univiersité Paris8
- Classification of multivariate digital farm time-series, Tosic Tamara, Orange Labs

12h10 Session poster déjeunatoire (sur place)

14h00 [Fabien Lotte] *Traitement et classification d' EEG pour les interfaces cerveau-ordinateur*

Les Interfaces Cerveau-Ordinateur ou BCI (de l'anglais « Brain-Computer Interfaces ») sont des systèmes de communication permettant à leurs utilisateurs d'envoyer des commandes à un ordinateur en utilisant uniquement leurs signaux cérébraux, grâce à la mesure et au traitement de ces signaux. Puisque les BCI permettent de contrôler un ordinateur sans aucune activité physique, elles sont très prometteuses pour de nombreux domaines d'application, notamment les technologies d'assistance, par exemple pour contrôler des fauteuils roulants, et l'interaction homme-machine. Dans cette présentation, je présenterai un panorama des méthodes d'analyses et de classification des signaux électroencéphalographiques (EEG) utilisés pour mesurer l'activité cérébrale dans les BCI. J'illustrerai notamment les informations qu'on peut extraire de ces signaux, et comment les extraire et les classifier malgré la nature bruitée et non-stationnaire de ceux-ci.

14h50 [Romain Tavenard] *Weakly supervised Machine Learning for Time Series*

For a wide range of application settings (eg. remote sensing), large amounts of data are collected over time. However, collecting associated ground truth information can be a tedious task. In this talk, I will present some recent advances towards using less supervision to extract information from time series. The talk will cover the development of an open-source toolkit for machine learning with time series (tslearn), as well as the presentation of novel siamese models for unsupervised learning. Finally, I will illustrate connections between time series and other structured data such as graphs and the opportunities it can offer in terms of new methodologies.

15h40 Conclusions/Cloture