

**position filled**

### **Internship position**

**Subject:** Unsupervised Learning of Probabilistic models for functional/structural data clustering

**Profile:** We are looking for a master 2 student or a final year Engineering student in the fields of computer science, information processing engineering, machine learning and statistical data analysis, signal and image processing, wishing to make an internship in an area related to statistical learning, classification, data analysis and pattern recognition.

**Subject description:** In many areas, such as information retrieval, signal processing, image processing including functional imaging, genomics or diagnosis of complex systems, etc., the data to be analyzed are functions (eg, curves, images), rather than vectors with reduced dimension. This type of data are often very structured. The vectorial approaches considering that an individual is a vector in  $\mathbb{R}^p$ , are thus limited as they do not explicitly take into account the structure of individuals (which are functions in this case). The framework particularly suitable to this context is that of functional data analysis (FDA) which aroused great interest of the scientific community in last years, particularly through learning statistical models.

The objective of this internship project is to propose a generic and innovative model and implement it via an efficient learning algorithm for the clustering/segmentation of functional data.

More specifically, we will propose a new nonparametric Bayesian modeling latent variable (or latent process) model that operates in an unsupervised way. This model, due to its latent variable formulation, is particularly suitable to the unsupervised context for clustering. The nonparametric Bayesian framework for the approach will also offer more flexibility through the use of general random processes as priors on the model, rather than strong assumptions as in classical Bayesian analysis.

**Laboratory:** The internship will be conducted within the team Information Dynamics (Dyni) team (rated A<sup>+</sup> by the Evaluation Agency for Research and Higher education (AERES)) of the LSIS laboratory UMR CNRS 7296, in Toulon, France.

**Required skills:** Statistical Modeling, Statistical estimation (Maximum Likelihood, etc), statistical inference

Matlab, R

Scientific English

**Desired skills:** Learning / classification, Bayesian inference, functional data analysis  
Skills in the EM algorithm(s) will be a major asset

**Internship gratification:** ~ 436 € / month

**Duration:** 4 months

**Supervisor :** Dr. Faicel Chamroukhi, MCF USTV/LSIS

**How to apply:** Please send your detailed CV and a letter stating your professional project to [faicel.chamroukhi@univ-tln.fr](mailto:faicel.chamroukhi@univ-tln.fr) by putting in your email subject : [Stage M2/Ing 2013] application of "your name"

A French version is available in the next page

## Stage attribué

### Offre de stage de fin d'études Master 2 /Ingénieur

**Titre:** Apprentissage non-supervisé de modèles probabilistes pour le clustering de données fonctionnelles.

**Profil recherché :** Étudiant(e) en Master 2 recherche ou en 3ème année du cycle ingénieur de formation dans les domaines informatique, technologies de l'Information, apprentissage statistique et analyse de données, traitement des signaux et des images, etc., désirant effectuer un stage de fin d'études traitant des thèmes apprentissage statistique, classification, analyse de données et reconnaissance de formes.

**Description du sujet :** Dans de nombreux domaines, comme la recherche d'information, le traitement de signal, l'imagerie notamment fonctionnelle, ou encore la génomique ou l'aide au diagnostic etc, les données à analyser se présentent sous la forme de fonctions (e.g., courbes, images), plutôt que de vecteurs de dimensions réduites. Ces données sont souvent très structurées, et les approches d'analyse vectorielles classiques, considérant qu'un individu est un vecteur dans  $\mathbb{R}^p$ , se trouvent ainsi limitées du fait de la non prise en compte explicite de la structure des individus (fonctions dans ce cas). Le cadre particulièrement adapté à ce contexte est celui de l'analyse de données fonctionnelles (Functional Data Analysis (FDA) ) qui a suscité un grand intérêt de la communauté scientifique ces dernières années, en particulier via l'apprentissage de modèles statistiques.

L'objectif de ce projet de stage est donc de proposer une modélisation innovante et générique et la mettre en œuvre via un algorithme d'apprentissage efficace pour la classification non-supervisée (clustering/segmentation) de données fonctionnelles.

Plus spécifiquement, nous proposons une nouvelle modélisation Bayésienne non-paramétrique à variable latente (ou processus latent) qui opère dans un cadre non-supervisé. Cette modélisation, de part sa formulation à variable latente est particulièrement adaptée au contexte non-supervisé pour la classification automatique. Le cadre Bayésien non paramétrique de l'approche offre également plus de flexibilité à travers l'utilisation de processus aléatoires généraux comme distributions a priori sur le modèle, plutôt que d'hypothèses fortes comme en inférence Bayésienne classique.

**Laboratoire d'accueil :** Le stage sera effectuée au sein de l'équipe Dynamique de l'Information (DYNI) (évaluée A+ par l'AERES) du laboratoire LSIS UMR CNRS 7296, site de Toulon.

**Compétences requises :** Modélisation statistique, Estimation statistique (Maximum de Vraisemblance), statistique interférentielle  
Programmation Matlab, R  
Maîtrise de l'anglais scientifique

**Compétences souhaitées :** apprentissage/classification, inférence bayésienne, analyse de données fonctionnelles

Des compétences en algorithme(s) EM seront un atout majeur

Gratification de stage : ~ 436 € / mois

Durée : 4 mois

**Encadrant :** Dr. Faicel Chamroukhi, MCF USTV/LSIS

**Candidature :** Merci d'envoyer votre CV détaillé (mentions, contenu des cours, etc) et une lettre de motivation précisant votre projet professionnel à [faicel.chamroukhi@univ-tln.fr](mailto:faicel.chamroukhi@univ-tln.fr) en mettant l'objet de votre e-mail sous cette forme : [Stage M2/Ing 2013] : candidature de votre nom