Learning methods for the spatiotemporal analysis of longitudinal data

Mots clés :
- Directeur de thèse : Stanley DURRLEMAN
- Co-encadrant(s) :
- Unité de recherche : Centre de recherche de l'Institut du cerveau et de la moelle épinière
- Ecole doctorale : Ecole Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Divers

Résumé du projet de recherche (Langue 1)

Keywords: statistical learning, spatiotemporal data, longitudinal data sets, personalization, classification, prediction, model of disease progression, brain imaging, neurological diseases Context: Longitudinal data sets are often acquired in biological and medical sciences to capture variable temporal phenomena, which are due for instance to growth, ageing or disease progression. They consist in the observation of several individuals, each of them being observed at multiple points in time. The statistical exploitation of such data sets is notably difficult since data of each individual follow a different trajectory of changes and at its own pace. This difficulty is further increased if observations take the form of structured data like images or measurements distributed at the nodes of a mesh, and if the measurements themselves are normalized data or positive definite matrices for which usual linear operations are not defined. Our team has contributed to the definition of a generic theoretical and algorithmic framework for learning typical trajectories from longitudinal data sets [1, 2, 3]. This framework is built on tools from the Riemannian geometry to describe trajectories of changes for any kind of data and their variability within a group. The inference is based on a stochastic EM algorithm coupled with Markov Chain simulation methods. The framework has been used so far to describe the dynamics of a set of biomarkers, namely unstructured data sets, which evolve in a sequential manner. Topic: The goal of the thesis is to extend this theoretical framework to deal with much more challenging data sets, called iconic-geometric data sets. These data take the form of measurements that are spatially distributed on a sub-manifold of any dimension (curve, surface or volume). In a discrete form, such a data takes the form of a measurement value, such as an intensity, a symmetric definite positive matrix or a ratio for instance, which is attached at each node of a geometric mesh. In brain imaging, the mesh may represent the cortical surface, where nodes represent precise anatomical locations, and signals may be the thickness of the cortical ribbon or a functional signal measuring glucose consumption at each anatomical location. The first objective of the thesis is to define a computational framework to track changes in such data sets. The main difficulty is that both the shape of the geometric support and the values of the measurements mapped onto the mesh vary at the same time. To this end, the framework of functional currents seems particularly interesting [4,5], as it defines a metric between iconic-geometric data sets built on a diffeomorphic deformation of the mesh to match the two geometric supports and a square integrable gradient of the signal to match the measurements. The deformation and variation together follow a geodesic path on a certain Riemannian manifold of high-dimension. This framework needs to be adapted though, since, in our case, the variation of the measurement may not be defined by the addition of a square integrable function. Measurements like positive definite matrices live also on a Riemannian manifold and the variation should then take the form of a geodesic on this manifold. Furthermore, we will introduce structured variations, so that signal variations propagate along neighboring nodes, and not independently from each other. The second objective of the thesis is to include this computational model for iconic-geometric data set into the generic statistical model for longitudinal data that we have defined in [3]. In principle, this step is straightforward given that we provide the space of iconic-geometric data with a Riemannian structure. Nevertheless, several approximations will be investigated to make the inference tractable. In particular, we would need to derive an efficient numerical scheme to compute parallel transport on a manifold by borrowing ideas from [6], and an efficient computation of geodesic distances in a mesh that is continuously deformed. All in one, this computational and statistical framework will allow us to estimate typical trajectories of data changes from iconic-geometric data sets acquired in several individuals at multiple points in time. Extensions of this framework will include mixture models, which will allow us to estimate a family of typical trajectories indicative of clusters of individuals with similar spatiotemporal patterns. The thesis also aims to personalize the typical scenarios to data of new individuals in order to position them along a typical scenario and to predict their future evolution. References: [1] S. Durrleman, X. Pennec, A. Trouve?, J. Braga, G. Gerig, N. Ayache, Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data, International Journal of Computer Vision 103 (1), 22-59, 2013 [2] J.-B. Schiratti, S. Allassonnie?re, A. Routier, O. Colliot, S. Durrleman, A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data, Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science (LNCS) 9123, 564-575, 2015 [3] J.-B. Schiratti, S. Allassonnie?re, A. Routier, O. Colliot, S. Durrleman, Learning scenarios of disease progression with spatiotemporal mixed effect models for manifold-valued data, Neural Information Processing System (NIPS), 2015 [4] N. Charon and A. Trouve?, Functional currents : a new mathematical tool to model and analyse functional shapes, Journal of Mathematical Imaging and Vision, vol. 48, p. 413–431, 2013 [5] B. Chartier, N. Charon and A. Trouve?, The iShape framework for the variability analysis of functional shapes, Fundations of Computational Mathematics, doi : 10.1007/s10208-015-9288-2, 2015 [6] V. I. Arnold, Mathematical Methods of Classical Mechanics, Springer-Verlag (1989)

Résumé du projet de recherche (Langue 2)
The thesis will make several important methodological contributions in the blooming topic of statistical learning for manifold-valued data. This work may be apply to several imaging data sets acquired in patients with different types of neurodegenerative diseases such as the Alzheimer’s disease or fronto-temporal dementia, to which we have access via our clinical collaborations. The developed methodology will estimate digital models of brain aging displaying simultaneously the anatomical deformation of the brain and the propagation of functional and metabolic alterations in the brain during the course of the disease. To the best of our knowledge, this will be the first time that the combination of anatomical and functional data will be displayed in such a dynamical fashion, thus offering clinicians a unique tool to deepen our understanding of the biological mechanisms underlying the onset of the disease. Unsupervised clustering by the mean of mixture models will also exhibit several pathological sub- groups within heterogeneous diseases like Alzheimer’s. Personalization of the digital model to new patients data will estimate the most probable trajectory of progression for each individual. Such tool will be used to derive computer-aided diagnosis and prognosis systems. Being able to robustly predict clinical symptoms only few months before their onset is a crucial public health challenge, which will open up the possibility to test treatments at the stage when they have the highest chance of success.

Informations complémentaires (Langue 1)

The PhD will take place within the ARAMIS lab (www.aramislab.fr), which is a common lab between INRIA (the national French institute dedicated to applied mathematics and computer science) and the Brain and Spine Institute (ICM). Located within an institute dedicated to basic and translational neurosciences, the ARAMIS lab develops advanced statistical and computational methods to analyze imaging and electrophysiological data. The development of statistical tools for the exploitation of longitudinal data set is among the top priorities of the team for the forthcoming years. This research aims to theoretical breakthroughs, the development of innovative technologies and their transfer to the clinic and pharmaceutical companies. The candidate will interact with our engineers and clinical collaborators in Europe and the United States.