M2 Internship proposal

Identification and quantification of explanatory variables effects on latent NODE-based trajectories

Context:

In the context of biomedical research, we are frequently led to analyze high dimensional and complex datasets characterized in particular by:

- High-dimensional and heterogeneous observations, as the result of multiple data acquisition processes (e.g. images, clinical scores, biomarkers etc.)
- Longitudinal data, meaning that for each subject we may dispose of several observations distributed over time.

In addition, the time evolution of a specific medical condition can be highly heterogeneous across subjects due to several sources of variability. Between-subjects variability can be at least partially explained by taking into account available information on the patient itself (e.g., sex, age, initial diagnosis, etc.) and eventually on the treatment he/she is submitted to (if any).

(Non-)linear Mixed effects models [1] are a powerful statistical parametric model allowing to explicitly account for explained and unexplained (random) between-subject variability, with the great advantage of providing a straightforward interpretation of the included covariates in driving the observed response, while enabling personalized estimations. Well established algorithms and tools exist to estimate mixed effect models' parameters, when we are dealing with a variety of data types, including continuous, categorical, count, and time-to-event data. Nevertheless, these models are not well suited when we are dealing with high dimensional data, such as medical imaging.

Leveraging on recent works focused on the modeling and personalization of high dimensional longitudinal data ([2-5]), we aim at providing a framework to estimate population and individual parameters of latent dynamics: we will focus on the task of selection of variables of interest and quantification of their effect in the latent space (hence in the observed data), therefore on the interpretability of the final model.

Aims and methods:

The main goal of this internship project is to develop a framework to estimate mixed effect model's parameters on latent dynamics. We will start by considering a single modality: extension to multi-model data integration (e.g. [5]) will be considered afterwards.
Application to synthetically generated data and to real data extracted from publicly available clinical datasets will allow to challenge the obtained framework in correctly quantifying the effects of the included explanatory variables.

Required skills:

Competences in statistic, Bayesian learning and mathematical modeling are required. Programming skills (Python) are mandatory. Experience with machine learning and a taste for biomedical applications and interdisciplinary topics is recommended.

Location and hosting group:

The internship will take place at *Centre INRIA d'Université Côte d'Azur* (Sophia Antipolis), in the EPIONE (e-patient for e-medicine) research project.
During the internship the candidate will have the opportunity to interact with researchers and students from the EPIONE team and participate to the scientific life of the team and of the INRIA Sophia-Antipolis center.

Contacts: Irene Balelli - irene.balelli@inria.fr; Marco Lorenzi - marco.lorenzi@inria.fr

Funding: Funding will be provided by INRIA.

Durations: 6 months

References:

[1] Lavielle, M. (2014). *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press.

[2] Yildiz, C., Heinonen, M., & Lahdesmaki, H. (2019). ODE2VAE: Deep generative second order ODEs with Bayesian neural networks. *Advances in Neural Information Processing Systems*, *32*.

[3] Chung, I., Kim, S., Lee, J., Kim, K. J., Hwang, S. J., & Yang, E. (2020, April). Deep mixed effect model using Gaussian processes: a personalized and reliable prediction for healthcare. In *Proceedings of the AAAI Conference on Artificial Intelligence*(Vol. 34, No. 04, pp. 3649-3657).

[4] Nazarovs, J., Chakraborty, R., Tasneeyapant, S., Ravi, S. N., & Singh, V. (2022). Mixed Effects Neural ODE: A Variational Approximation for Analyzing the Dynamics of Panel Data. *arXiv preprint arXiv:2202.09463*.

[5] Abi Nader, Clément, et al. "Simulating the outcome of amyloid treatments in Alzheimer's disease from imaging and clinical data." *Brain communications* 3.2 (2021): fcab091.