



université
PARIS-SACLAY



Internship Position in Metalearning for Healthcare Exploring Hierarchical Representations with Poincaré Variational Auto-Encoders

This project will lead to concrete machine learning empirical insights intended to be exploited in the larger context of IHU Prometheus, a multi-year large-scale research and training institute for understanding sepsis disease. The student will also have the opportunity to get involved in ongoing collaboration with the IMES at MIT.

Machine learning is increasingly used in healthcare applications to assist medical staff in diagnosing their patients and providing tailored medications. Many real datasets in healthcare are hierarchically structured. However, traditional machine learning models map the data to Euclidean latent space, which cannot efficiently handle tree-like structures. This is the case with variational auto-encoders (VAEs), a powerful type of machine learning model that is widely used for generative purposes. In particular, they allow capturing explainable factors of variation, an important property that we seek in healthcare applications. In this project, we are interested in exploring the benefits of hierarchical representations for healthcare applications with a particular kind of VAEs which are embedded in *hyperbolic* spaces [1, 2].

As a starting point, we will take a look at this paper "*Continuous Hierarchical Representations with Poincaré Variational Auto-Encoders*" [1] and adapt it to a healthcare application. The goal is to leverage meta-knowledge [3] about the learning problem, for example, the hierarchical structure of the labels to improve the learning process. In particular, we are interested in assessing what happens for the learning process and the learned representations when choosing appropriate parameterizations (here, leveraging the hierarchical structure of the target labels) compared to when we simply map data to the Euclidean latent space, i.e., flatten the target labels and ignoring their hierarchical structure. Other VAEs that feature hyperbolic embedded spaces, such as [2, 4, 5] will also be explored and compared to basic Euclidean embedded space in terms of performances, representational properties, and learning bounds.

The experiments will be performed on the MIMIC-IV dataset [6], a freely available real-world dataset encompassing electronic health records of patients admitted to intensive care units. The idea is to adapt one of the healthcare learning problems featured in MIMIC-IV dataset to a hierarchical representation learning problem.

We seek to publish the obtained results in machine learning and machine learning for healthcare related workshops or conference venues.

Supervision team and Contact

Massinissa HAMIDI
Maître de conférences
IBISC Laboratory, Univ. Évry Paris-Saclay
massinissa.hamidi@univ-evry.fr

Li-wei H. LEHMAN
Research Scientist
Massachusetts Institute of Technology
lilehman@mit.edu

Information

Internship period: 6 months starting from January/February/March 2025

Funded internship

Possibility to pursue a PhD program after the internship

Location: Institut de Biologie Génétique et Bioinformatique (IBGBI), Évry, France

Students in MSc. Computer Science programs or engineering schools with specialization in data science, statistics, or artificial intelligence are encouraged to submit their application.

To apply, please send a resume, a motivation letter, and the copy of your MSc. transcripts to massinissa.hamidi@univ-evry.fr.

Bibliography

- [1] Mathieu E, Le Lan C, Maddison CJ, Tomioka R, Teh YW. Continuous hierarchical representations with poincaré variational auto-encoders. *Advances in neural information processing systems*. 2019;32.
- [2] Bose, Joey, et al. "Latent variable modelling with hyperbolic normalizing flows." *International conference on machine learning*. PMLR, 2020.
- [3] Hamidi, Massinissa. *Metalearning guided by domain knowledge in distributed and decentralized applications*. Diss. Université Paris-Nord-Paris XIII, 2022.
- [4] Davidson, Tim R., et al. "Hyperspherical variational auto-encoders." *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018. Association For Uncertainty in Artificial Intelligence (AUAI)*, 2018.
- [5] Cho, Seunghyuk, Juyong Lee, and Dongwoo Kim. "Hyperbolic VAE via latent Gaussian distributions." *Advances in Neural Information Processing Systems* 36 (2024).
- [6] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, et al. *Mimic-iv, a freely accessible electronic health record dataset*. *Scientific data*, 10(1):1, 2023.