Supervisors: Morgan MAGNIN (Prof.), Tony RIBEIRO (Dr.)
Host laboratory: Laboratoire des Sciences du Numérique de Nantes, CNRS UMR 6004
Institution: École Centrale de Nantes
Place: Nantes, France
Research project: Homi-lung
Funding: Horizon Europe Programme - health cluster, Destination- Tackling diseases and reducing disease burden
Research team: VELO (Vérification pour l'Environnement et le LOgiciel)
Doctoral school: MASTIC (https://ed-mastic.doctorat-paysdelaloire.fr/)

Context

The general goal of this PhD thesis is to bring new contributions to inductive logic programming that will benefit to the analysis of a major health issue.

More specifically, this PhD thesis is in line with the general aim of better understanding the causal links between respiratory tract infections (pneumonia) and the subsequent progression of cardiovascular and respiratory diseases (up to one year after disease).
It seeks to build longitudinal dynamic models of the evolution of these diseases over time, with the aim of (1) identifying diagnostic markers playing a key role in the evolution of complications of each disease subtype; (2) contributing to the prediction of disease progression and treatment effects.

This goal is part of the Homi-lung research project (Long-term alterations of host-microbiome interactions and cardiovascular and respiratory diseases progression after pneumonia). Homi-lung is funded by the European Commission (Horizon Europe Programme) and is coordinated by Professor Antoine Roquilly (Nantes University - CR2TI). This project consists in a consortium of 11 partners, including Centrale Nantes / LS2N (Laboratoire des Sciences du Numérique à Nantes).

Mixing inductive learning approaches and statistics-based methods, the work will consist in designing and producing new logic programs to analyze the dynamic behavior of certain patient variables with regard to their susceptibility to developing subsequent diseases.

Backgound

The team behind this thesis project has long been developing methods for symbolic learning and logical analysis of the dynamics of living systems. They are interested in the design and study of the properties of explainable dynamic models. Since a decade, their research focuses on the development of Learning From Interpretation Transition [Ino2014, Rib2021] (LFIT), an inductive logic programming paradigm that automatically builds a model of the dynamics of a system from the observation of its state-transitions. LFIT has the merits to output explainable dynamical models from time series without any prerequisite about the semantics of the model. Its applicability to biology has been assessed through various contributions, while triggering the needs for new developments. Figure 1 shows this learning process.

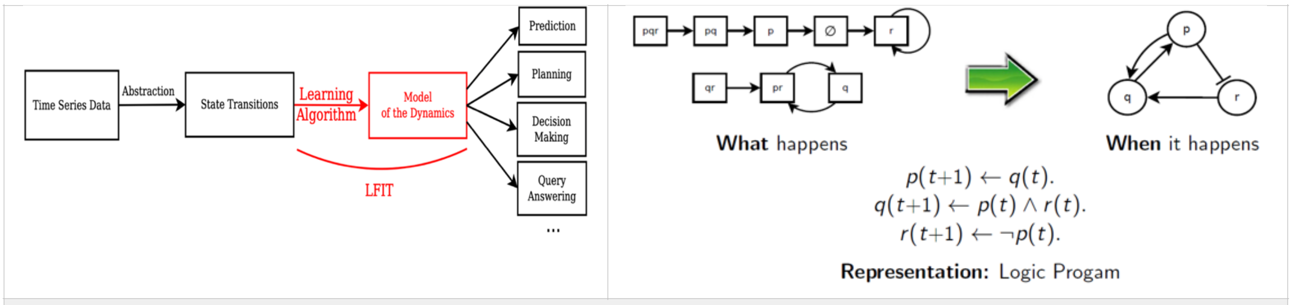

Figure 1: Left: General workflow for using the LFIT framework. Right: State transitions of a dynamics system (top left), the logic program learned by LFIT from the transitions (bottom) and the corresponding representation as Boolean network (top right) of the learned program

In the case of biological systems, this means that from time series data, LFIT builds a dynamic model expressed as a (comprehensible) set of logical rules.

This PhD thesis will contribute to the design of models of patients taking into account the dynamical evolution of various follow-up variables. The input will thus be time-course data of patients, obtained from data gathered by medical doctors. Using inductive learning techniques, the goal is to output logic programs that will allow analyzing dynamical behavior of some patients' variable with regard to their susceptibility to develop later illnesses.

To address this issue, we will take advantage of the most recent developments conducted on LFIT. In our most recent work [Rib2024], we addressed one of the major challenges for modelers, which is overcoming the absence of some data in existing cohorts. Some data may indeed be missing, because some variables were not accessible by experimental means (or were simply not considered of interest at the time of collecting data on real cohorts). To attain this goal we extended the LFIT framework to learn from transitions between partial states where some variable values are unknown. By modeling the unknown, we achieve an overestimation of the real system regarding both its dynamics and variables interactions.

There are two major scientific locks to overcome now.
The first lies in the design of a learning approach robust enough to a large heterogeneity of data collections. To do so, we need to integrate in our learning approach some statistics-based method. In the literature, there exists various ways to integrate statistics in an inductive logic programming framework [Ker2003].
Meanwhile, we also need to consider recent works that build bridges between logic-based approaches and deep learning methods. Several articles [Phu2019, Cro2020] published in in recent years give leads on how inductive logic programming could be combined with neural networks to improve both explainability and predictive power. Bridging these two approaches is all the more necessary for the applicability of these methods in health sciences.

Research subject, work plan:

The work will be decomposed into three main tasks: (1) Identify and analyse the few existing contributions linking logic programming and statistics/neural networks; (2) Propose learning algorithms that take time series data as input and output a model that can compete with the best competitors in terms of prediction of behaviour while being explainable (i.e., underlying rules about the dynamical behaviour of the system can be produced and made understandable to biological experts); (3) Assess the efficiency of the algorithms on real data coming either from the cohorts of the Homi Lung project.

The resulting algorithms will be implemented in a tool released under a free-software license with a GUI.

Candidate profile

To carry out this research work, the candidate will need solid skills in computer science, both theoretical (logic programming, formal methods related to model verification) and practical (implementation). The candidate should also have an interest in health sciences. Experience in Systems Biology would be a plus.
The candidate should have good writing skills and an aptitude for teamwork.

Contract

- Status : paid thesis under contract
- Fields: computer science, modeling, systems biology
- Contract duration: 36 months
- Desired start date: October 1st 2025

How to apply?

- Deadline for application: May, 15th 2025
- Send (1) a cover letter, (2) CV and (3) Master (1st and 2nd year) transcripts to :
  - Morgan MAGNIN (morgan.magnin@ls2n.fr)
  - Tony RIBEIRO (tony.ribeiro@ls2n.fr)

References:

[Cro2020] A. Cropper, Andrew, and R. Morel. "Learning programs by learning from failures." arXiv preprint arXiv:2005.02259 (2020).
[Ino2014] K. Inoue, T. Ribeiro, and C. Sakama. Learning from interpretation transition. Machine Learning, 94(1):51-79, 2014.
[Ker2003] L. De Raedt and K. Kersting. "Probabilistic logic learning." ACM SIGKDD Explorations Newsletter 5.1 (2003): 31-48.
[Pau2011] L. Paulevé, M. Magnin, O. Roux. Tuning Temporal Features within the Stochastic πCalculus. IEEE TSE, 37(6):858-871, 2011.
[Phu2019] Y.J. Phua and K. Inoue. "Learning Logic Programs from Noisy State Transition Data." International Conference on Inductive Logic Programming. Springer, Cham, 2019.
[Rib2015] T. Ribeiro, M. Magnin, K. Inoue, and C. Sakama. Learning Delayed Influences of Biological Systems. In Frontiers in Bioengineering and Biotechnology, 2, 81. 2015.
[Rib2024] T. Ribeiro, M. Folschette, M. Magnin, K. Okazaki, L. Kuo-Yen, A. Roquilly, J. Poschmann, K. Inoue. Learning From Interpretation Transitions with Unknowns. On the 4th International Joint Conference on Learning & Reasoning (IJCLR 2024). Nanjing, China, September 20-22, 2024.