PhD proposal: An Artificial Intelligence framework to improve the learning of dynamic models from time series data

Supervisors : Pr. Morgan MAGNIN and Pr. Olivier ROUX (École centrale de Nantes, LS2N)
Web site: https://ls2n.fr/?lang=en
Email: morgan.magnin@ec-nantes.fr

Key words: artificial intelligence, logic programming, constraint programming, machine learning, bioinformatics

Background:
The research domain of this PhD thesis is the formal modelling and analysis of complex dynamical systems (specifically in biological systems). Such a topic is the area of expertise of **the MeForBio team** (acronym for "Formal Methods for Bioinformatics") of **LS2N**, **one of France's leading public research labs in digital sciences.** The MeForBio team is composed of 2 professors, 1 post-doc and 2 PhD students in Computer Science. It is much involved in local, national and international cooperations (especially with the National Institute of Informatics in Tôkyô). Even if the application domain of this thesis is primarily oriented towards biology, **no pre-requisite in life sciences are necessar**y. The main goal of the current proposal is to design innovative **machine learning algorithms that can capture – and provide explainability – to the behavior of large-scale dynamic systems.**

The scientific motivation of this research project lies in the fact that **a large amount of time series data can now be obtained quite easily**, with both (i) the spread of numerical tools in every part of daily life and (ii) the development of New Generation Sequencing methods (NGS) in biology. A critical question here is to attach a meaning to these data, i.e., build relevant models (a task that cannot be designed anymore by hands only) that are both meaningful (for the researcher to have a better understanding of the processes at stake) and predictive enough. It then becomes crucial to be able to connect the time series data with models to improve one's understanding of a targeted system. This means that we need to be able to learn the model from the input data, but also to analyze some key properties on these models. In other words, this implies either to formally prove that some properties are satisfied or to guarantee that these properties are not satisfied. And the designer then obviously needs some automotive help to control the system in such a way that the property may be satisfied. In recent years, we have investigated two complementary learning approaches to infer models from time series data: one is based on the use of Answer Set Programming [Pau2011, Ben2017], the other on inductive logic programming [Rib2015, Rib2017]. Thanks to these methods, we have been able to address systems with hundreds of interacting components. Both approaches however share the same drawback, which is to contain the inherent noise and/or imprecisions in the data. For example, when two identical observations at two different time steps lead to different behaviors, a key issue is to be able to understand why the same conditions result in different consequences.

To overcome this limit, and when several models are compatible with some temporal data series, we intend to **propose a (semi)-automatic way to decide, from a modeling point of view, which kind of additional experiments could be performed to choose the "real" model.** This is critical because, due to experimental constraints, several experiments cannot be implemented and it is often not possible to observe every dynamical step of a model. The goal of the PhD thesis is to contribute to a new methodological framework to (partly) automate the modeling of dynamic systems based on time series data. The resulting framework will be accompanied with efficient learning and analysis algorithms, and a practical implementation in a free-software tool.

Research subject & work plan:

The approach will be decomposed **into four main tasks**: (1) Obtain and curate data based on datasets provided at the beginning of the work, coming from DREAM Challenges (a popular reverse engineering challenge); (2) Learn the model (and its dynamics) based on time series data input. It will be necessary to assess the quality of its components based on background knowledge and deal with the conflicting data; (3) Check the validity of a set of dynamic properties that help to validate/invalidate the resulting model; (4) Design algorithms such that if the expected properties are not satisfied, it automatically provides help to (i) either suggest new experiments to raise the ambiguity in the initial data, (ii) or modify the model in such a way that the expected behaviour can be satisfied (note that such modification may not be possible for some properties) and validate the analysis w.r.t. physical consistency.

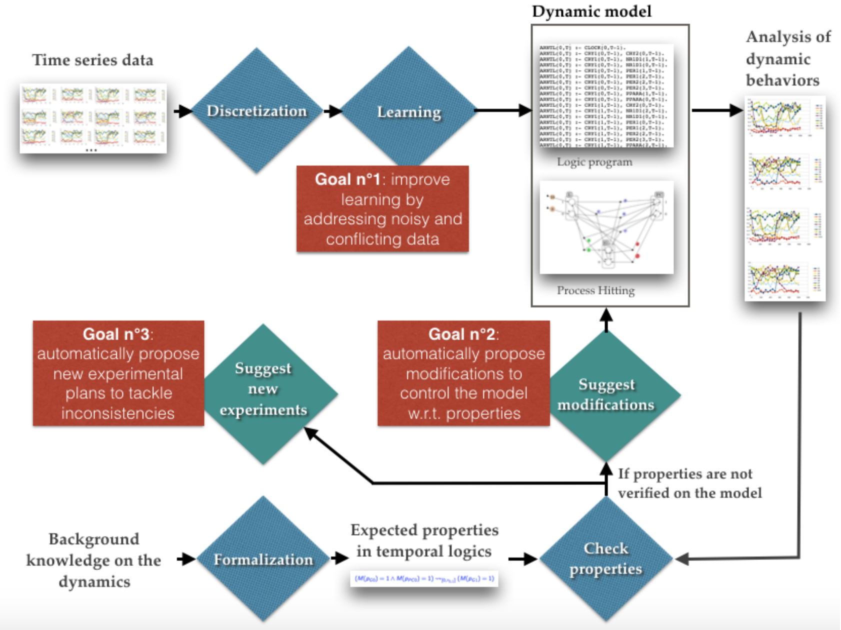
Time series data
Discretization
Learning
Dynamic model
Logic program
Process Hitting
Analysis of dynamic behaviors
Goal n°1: improve learning by addressing noisy and conflicting data
Goal n°3: automatically propose new experimental plans to tackle inconsistencies
Suggest new experiments
Goal n°2: automatically propose modifications to control the model w.r.t. properties
Suggest modifications
If properties are not verified on the model
Background knowledge on the dynamics
Formalization
Expected properties in temporal logics
Check properties

Figure 1: Research plan to improve the learning process

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

References:

[Ben2017] E. Ben Abdallah, T. Ribeiro, M. Magnin, O. Roux and Katsumi Inoue. Modeling Delayed Dynamics in Biological Regulatory Networks from Time Series Data. Algorithms, Volume 10, Number 1, 2017.

[Pau2011] L. Paulevé, M. Magnin, O. Roux. Tuning Temporal Features within the Stochastic π-Calculus. IEEE TSE, 37(6):858-871, 2011.

[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.

[Rib2017] T. Ribeiro, S. Tourret, M. Folschette, M. Magnin, D. Borzacchiello, P. Chinesta, O. F. Roux, K. Inoue. Inductive Learning from State Transitions over Continuous Domains. In International Conference on Inductive Logic Programming (pp. 124-139). Springer, 2017.