PhD Proposal

Title Model-based reinforcement learning for the control of partially observable piecewise deterministic semi-Markov decision processes

Location The thesis will take place mainly at the Université de Montpellier but stays of varying lengths will be organized at INRAE in Toulouse depending on the mobility of the candidate.

Supervision The thesis will be supervised by Alice Cleynen (alice.cleynen@umontpellier.fr, CR CNRS at université de Montpellier), Benoîte de Saporta (benoite.de-saporta@umontpellier.fr, Professor at université de Montpellier) and Régis Sabaddin (regis.sabbadin@inrae.fr, DR INRAE at MIAT at INRAE Toulouse).

Funding This thesis project is part of the ANR Hidden Semi Markov Models: INference, Control and Applications (HSMM-INCA) project, led by N. Peyrard (DR INRAE-MIAT). The ANR project will start in January 2022 and will be funded for a period of 4 years.

Data The application is addressed within the framework of a long-standing collaboration with the CRCT in Toulouse between Alice Cleynen and Hervé Avet-Loiseau, allowing data to be obtained and discussions with oncologists.

Application Anyone interested is encouraged to contact the three supervisors for any further information. Applicants are invited to send a brief CV as well as a cover letter to the three email addresses mentioned above.

SCIENTIFIC CONTEXT

General framework The long-term treatment of human diseases such as cancer is generally based on monitoring the dynamics of variables (markers in the organism) over time, modeled by a series of continuous trajectories defined from a “mode” (conjunction of a stage of the disease and a treatment). The transitions between modes depend both on intrinsic characteristics and on the treatments applied. The transitions between stages and the sojourn times can be modeled by semi-Markov kernels (depending on the treatments applied). Optimal treatment of a disease, in the ideal case where markers and stages are continuously observed and dynamics models are known, amounts to optimizing a strategy based on the patient's condition, for a semi-Markov decision process. In reality, (i) the stages of the disease are not observed, (ii) the markers are only observed during sampling, the dates of which are to be decided together with the treatments and (iii) the models of dynamics are not known (one will make the assumption of a parameterized form, of which the parameters are unknown). The objective of this thesis is to propose a framework of representation and optimization algorithms for these problems. Their unifying characteristic is a piecewise deterministic dynamic which should allow the development of specific approaches, more efficient than the general framework of partially observable semi-Markov decision processes. On the application level, we will focus on the problem of cancer monitoring and treatment, for which the members of the consortium have both monitoring data and collaborate with experts.

Mathematical framework Piecewise deterministic Markov processes (PDMP) form a class of processes particularly suitable for modeling [D93]. If the literature regarding the theoretical aspects of the control of these processes is abundant, the numerical aspects are much less studied, in particular when the jump dates are not observed [CdS18]. We can reduce the problem to a continuous state-space partially observable Markov decision process (POMDP), for which there are few solving approaches [dSDN16], [Z13], Z10]. We do not know of any approach for controlling PDMPs in the partially observable and unknown model case. However, [DSD20] recently became interested in the application of Bayesian Reinforcement Learning (BRL) methods using ODEs for the control of continuous-time semi-Markov decision processes. In addition, some BRL approaches have been developed to solve POMDPs [GMPT15]. However, these approaches (i) do not deal with the semi-Markov case and (ii) ignore the case of “piecewise deterministic” dynamics. Yet it is likely that this type of dynamics, very frequent in the models used in

biology or medicine, admits interesting control methods (probably approximate). We wish to provide the first contributions to the control of piecewise deterministic semi-Markov processes, partially observed and with poorly known model.

RESEARCH PROGRAM

After a bibliographic study devoted to familiarization with piecewise deterministic (semi-) Markov process models and their decisional counterparts, reinforcement learning methods, including Bayesian, and the handling of medical data in the problematic of monitoring and treatment of long-term illnesses, the thesis will focus on sub-families of models of increasingly complexity.
We will be particularly interested in solving methods of increasing difficulty:

- Case where the markers are perfectly observed at discrete dates, and only the parameters of the deterministic dynamics are unknown.
- Case where the observations of the markers at discrete dates are noisy and only the parameters of the deterministic dynamics are unknown.
- The most complex case, where the dates of observations of the markers are "rare" and optimized by a control strategy. This case will require the development of Bayesian RL methods implemented in a “batch RL” context (that is to say, exploitation of the complete patient follow-up dataset, to build an online strategy adapted to a particular patient).

The thesis will focus on constant feedback to the application, with, if possible, the construction of treatment strategies for patients intelligible to physicians.

EXPECTED SKILLS OF THE CANDIDATE

This thesis has a strong methodological component, at the crossroads between statistics (process models) and artificial intelligence (decision, learning). Skills in one of these areas are therefore required. The doctoral student will acquire new methodological skills during the thesis and will thus present, at the end of the thesis, a highly-valued multidisciplinary methodological profile.
She or he will also be required to implement the control and learning algorithms developed (R, Python...), and must be familiar with at least one of these programming languages. Finally, she or he will have to evolve in a multidisciplinary environment, by being in regular contact with our biologist and physician partners. She or he will progress in understanding multidisciplinary issues and communicating research results to different communities. An open personality and curious about the issues of related scientific fields is therefore required.

REFERENCES

[BGPS14] M. Bonneau, S. Gaba, Na. Peyrard, R. Sabbadin. Reinforcement learning-based design of sampling policies under cost constraints in Markov random fields: Application to weed map reconstruction. Computational Statistics & Data Analysis, Vol. 72, 2014.
[CdS18] A. Cleynen and B. de Saporta. Change-point detection for piecewise deterministic Markov processes. Automatica 97, pp. 234–247, 2018.
[D93] MHA. Davis. Markov models and optimization, volume 49 of Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1993
[dSDN16] B. de Saporta, F. Dufour, and C. Nivot. Partially observed optimal stopping problem for discrete-time markov processes.. 4OR, 2016.
[DFD20] J. Du, J.Futoma, F.Doshi-Velez. Model-based Reinforcement Learning for Semi-Markov Decision Processes with Neural ODEs. NeurIPS 2020.
[GMPT15] M. Ghavamzadeh; S. Mannor; J. Pineau; A. Tamar, Bayesian Reinforcement Learning: A Survey , now, 2015.
[LGR12] S. Lange, T. Gabel, M. Riedmiller. Batch Reinforcement Learning. Book Chapter, “Reinforcement Learning: State of the Art,” 2012.
[Yu06] H.Yu. Approximate solution methods for partially observable markov and semi-markov decision processes. PhD Thesis, Massachusetts Institute of Technology, 2006.
[Yus81] A.A. Yushkevich. On semi-Markov controlled models with an average reward criterion. Theory of Probability and Its Applications. 26: 796–802, 1981.
[Z13] E. Zhou. Optimal stopping under partial observation: Near-value iteration. Automatic Control, IEEE Transactions on, 58(2):500{506, 2013.
[Z10] E. Zhou, M.C. Fu, and S.I. Marcus. Solving continuous-state pomdps via density projection. Automatic Control, IEEE Transactions on, 55(5):1101{1116, 2010.