Project of PhD thesis in Computer Science: "Learning clinical decision processes from EHR: application to common condition diagnosis"

Antoine Neuraz (AP-HP, Université de Paris) and
Adrien Coulet (Inria Paris)

May 2021

Keywords: AI in healthcare, Electronic health records, supervised machine learning, clinical practice guidelines, diagnosis, RNN, explanatory AI, reinforcement learning, evaluation
Titre en français : Apprentissage de processus de décision diagnostique: Expérimentations pour le diagnostic des affections fréquentes à partir de Dossier Patients Informatisés

1 Background

A clinical guideline (or CPG for Clinical Practice Guideline) is a document with the aim of guiding decisions and criteria regarding diagnosis, management, and treatment in specific areas of healthcare. In the case of diagnostics, guidelines may have the form of an algorithm with a set of questions to answer to be able to diagnose a patient with a specific condition. Classically, guidelines were established based on the medical literature, expert consensus, but also evidence from molecular biology or data-driven studies such as randomized clinical trials or observational studies.

However, CPG are rarely available in machine-actionable forms that would allow their integration in decision support systems and then provide recommendations along the care flow. Previous works developed formal representation of clinical guidelines, such as those done with the GLIF3 standard to produce what is named Computer-Interpretable Guidelines (or CIGs) [2, 12]. These CIGs may be obtained manually, following for instance cognitive approaches involving either medical expert, knowledge engineer or both [13]; or obtained partially automatically from narrative guidelines or the medical literature [16, 15, 11]. So far, only a few works have focused on data-driven approaches that would learn CIGs from patient data. Most of them rely on unsupervised approaches, such as process mining [5], or sequence mining [3, 18] that were leveraged to learn clinical pathways, or decision processes, i.e., ordered sets of events of patients that can be, in turn, compared to CIGs.

This thesis project is motivated by the fact that we think that modern EHR may be a source to learn clinical decision processes in a supervised setting and that various decision processes may be aggregated to be compared to CPG. In particular, we think that domain knowledge, such as those represented in CIGs and biomedical ontologies, may guide both learning and comparison processes. Being able to learn and compare CPG would be highly impactful, since this may serve to update CPG, and to guide clinical decisions.

2 Objective

In this thesis project, we propose to study how formal or semi-formal representations of diagnostic guidelines may be automatically learned from EHR data. To this aim we will (1) investigate two families of approaches to infer this type of process from large clinical data warehouses, and (2) will define a framework to evaluate and compare formal guidelines.

The initial hypothesis of this project is that starting with a large set of EHRs, it is feasible to train a deep neural network that provides a good prediction for events that are rather frequent in the population. Indeed, recent results show that even if predictive neural network models present many challenges (e.g., ensuring fairness, providing explanations or portability), they enable to predict with good performances clinical outcomes that may be observed in many patients, such as re-admission, length of stay, mortality, whereas sporadic events such as uncommon disease diagnostics have mitigated results [10, 14, 17].

As today, we propose two methodological directions, but depending on the candidate affinities and on results, the thesis may pursue only one, or two, of them.

(1a) The first direction is to experiment RNN models, because RNN are adapted to sequential decision tasks (such as speech recognition, language modelling), and to use explanatory methods over RNN to produce pieces of a diagram that summarizes decisions made for the models for a set of representative instances. Regarding explanatory approaches (i.e., approaches that provide elements of interpretation of hard-to-interpret machine learning models), TimeSHAP extends the general framework named KernelSHAP to the setting of recurrent models [9, 1]. The particularity of TimeSHAP is not only to identify what are the most important features of a recurrent model, but also to identify which previous events had the largest impact on a current prediction. If TimeSHAP may constitute a useful brick, many challenges remain towards learning a decision process, such as training a recurrent model that is adapted to specific diagnostic tasks, turning TimeSHAP explanatory elements in the form of instances of decisions, and aggregating those instance-level explanations in a unique diagram. One particular challenge we propose to explore is the use of existing CIGs to guide the recurrent model, in a knowledge-guiding manner. For instance, one may experiment on using existing CIGs to constrain a "decision process learner" on focusing on some features, or sequence of events, that would facilitate a latter comparison between existent CIGs and learned process.

(1b) Because the learning of sequential decision is also a task that may be achieved by Reinforcement Learning (RL), we also propose to consider investigating this family of approaches [8]. In particular, Deep Reinforcement Learning (DRL) adopts deep neural networks to learn the optimal policy that is central to RL and facilitate handling cases where the number of states is potentially large [7]. Following this direction, many challenges are to face: how to define the reward space in such a manner that decisions made in previous states are considered recursively in the general decision process and general policy? What is the optimal policy we aim at reaching in the case of a diagnosis decision? And how can this be evaluated? Can we consider states with incomplete knowledge on patients, as it is the case with EHR? Among various possible approaches, we may consider "simultaneous learning", which consists in decomposing the reward function into a sum of meaningful sub-reward, forcing the system to learn at the same time a policy and an explanation [4, 6]. For instance, each laboratory test ordered toward a final diagnosis may be a subgoal that serves to compose a global explanation for the final diagnosis. Such an explanatory approach for DRL produces diagrams that may be suitable with our goal. However, this kind of approaches has been applied to game playing, robotics, but never for decision support in healthcare.

(2) Independently from the "process learner", we propose to set up a framework for the evaluation of CIG with regards to a population (as one defined by a set of EHR for instance). In particular, this framework would aim at assessing CIGs regarding their coverage of the considered population, their robustness to incomplete data (common for many healthcare settings), their error rate, and potential additional features such as economical cost or time to decision. This framework would enable to compare results of potential CGI-learner.

3 Case study and data

As a start we will consider diagnoses of general conditions such as anemia and chronic kidney disease, and propose to use data from the AP-HP healthcare data warehouse (*l'Entrepôt de Données de Santé de l'AP-HP*).

4 Context of the thesis

The candidate will be part of the HeKA Inserm-Inria research team (`https://team.inria.fr/heka/`), located in Paris, and will be co-supervised by Antoine Neuraz and Adrien Coulet. The candidate will register to the Doctoral School ED386 at the University of Paris (`http://ed386.sorbonne-universite.fr/fr/index.html`).

Contacts

Antoine Neuraz, `antoine.neuraz@aphp.fr`
Adrien Coulet, `adrien.coulet@inria.fr`

References

[1] João Bento, Pedro Saleiro, André F Cruz, Mário AT Figueiredo, and Pedro Bizarro. Timeshap: Explaining recurrent models through sequence perturbations. *arXiv preprint arXiv:2012.00073*, 2020.

[2] Aziz A Boxwala, Mor Peleg, Samson Tu, Omolola Ogunyemi, Qing T Zeng, Dongwen Wang, Vimla L Patel, Robert A Greenes, and Edward H Shortliffe. Glif3: a representation format for sharable computer-interpretable clinical practice guidelines. *Journal of biomedical informatics*, 37(3):147–161, 2004.

[3] Elias Egho, Nicolas Jay, Chedy Raïssi, Dino Ienco, Pascal Poncelet, Maguelonne Teisseire, and Amedeo Napoli. A contribution to the discovery of multidimensional patterns in healthcare trajectories. *Journal of Intelligent Information Systems*, 42(2):283–305, 2014.

[4] Alexandre Heuillet, Fabien Couthouis, and Natalia Díaz-Rodríguez. Explainability in deep reinforcement learning. *Knowledge-Based Systems*, 214:106685, 2021.

[5] Zhengxing Huang, Xudong Lu, Huilong Duan, and Wu Fan. Summarizing clinical pathways from event logs. *Journal of biomedical informatics*, 46(1):111–127, 2013.

[6] Zoe Juozapaitis, Anurag Koul, Alan Fern, Martin Erwig, and Finale Doshi-Velez. Explainable reinforcement learning via reward decomposition. In *IJCAI/ECAI Workshop on Explainable Artificial Intelligence*, 2019.

[7] Siqi Liu, Kee Yuan Ngiam, and Mengling Feng. Deep reinforcement learning for clinical decision support: a brief survey. *arXiv preprint arXiv:1907.09475*, 2019.

[8] Siqi Liu, Kay Choong See, Kee Yuan Ngiam, Leo Anthony Celi, Xingzhi Sun, and Mengling Feng. Reinforcement learning for clinical decision support in critical care: comprehensive review. *Journal of medical Internet research*, 22(7):e18477, 2020.

[9] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*, 2017.

[10] Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6(1):1–10, 2016.

[11] Emilie Pasche, Patrick Ruch, Douglas Teodoro, Angela Huttner, Stephan Harbarth, Julien Gobeill, Rolf Wipfli, and Christian Lovis. Assisted knowledge discovery for the maintenance of clinical guidelines. *PloS one*, 8(4):e62874, 2013.

[12] Mor Peleg. Computer-interpretable clinical guidelines: a methodological review. *Journal of biomedical informatics*, 46(4):744–763, 2013.

[13] Mor Peleg, Lily A Gutnik, Vincenza Snow, and Vimla L Patel. Interpreting procedures from descriptive guidelines. *Journal of Biomedical Informatics*, 39(2):184–195, 2006.

[14] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, et al. Scalable and accurate deep learning for electronic health records. *npj Digital Medicine*, 1(18), 2018.

[15] Daniel R Schlegel, Kate Gordon, Carmelo Gaudioso, and Mor Peleg. Clinical tractor: A framework for automatic natural language understanding of clinical practice guidelines. In *AMIA Annual Symposium Proceedings*, volume 2019, page 784. American Medical Informatics Association, 2019.

[16] Radu Serban, Annette ten Teije, Frank van Harmelen, Mar Marcos, and Cristina Polo-Conde. Extraction and use of linguistic patterns for modelling medical guidelines. *Artificial intelligence in medicine*, 39(2):137–149, 2007.

[17] Yuqi Si, Jingcheng Du, Zhao Li, Xiaoqian Jiang, Timothy Miller, Fei Wang, W Jim Zheng, and Kirk Roberts. Deep representation learning of patient data from electronic health records (ehr): A systematic review. *arXiv preprint arXiv:2010.02809*, 2020.

[18] Yiye Zhang, Rema Padman, and Nirav Patel. Paving the cowpath: Learning and visualizing clinical pathways from electronic health record data. *Journal of biomedical informatics*, 58:186–197, 2015.