

Cadre méthodologique de spécification formelle d'un simulateur par apprentissage machine profond pour assurer sa validation

Christophe Denis^{*,**,***,****}

*EDF R&D, Département PRISME,
6 quai Watier, 78400 Chatou

**Saclay INdustrial Collaborative Laboratory for Artificial Intelligence Research
Boulevard Gaspard Monge, 91120 Palaiseau, France.

***Sorbonne Université, UMMISCO
32 Avenue Henri Varagnat, 93143 Bondy Cedex, France

**** Université Panthéon-Sorbonne, IHPST
13 Rue du Four, 75006 Paris, France

Résumé. Cette contribution présente un cadre méthodologique pour formaliser les spécifications d'un simulateur basée sur de l'apprentissage machine profond. Ce travail de recherche, mêlant épistémologie de la modélisation et informatique théorique, pour objectif de mettre en place des techniques informatiques de vérification et validation pour des systèmes critiques nécessitant la puissance prédictive d'une modélisation par apprentissage machine.

Le Groupe EDF exploite des matériels et des ouvrages à longue durée de vie comme des barrages hydroélectriques et des centrales nucléaires et mène historiquement au sein de sa R&D une forte activité de simulation numérique dans différents domaines scientifiques, allant de la mécanique des structures à la mécanique des fluides en passant l'électromagnétisme. Les codes industriels de simulation se basent historiquement sur une modélisation hypothético-déductive qui a permis d'obtenir des avancées significatives sur la prédiction et la compréhension de phénomènes. Le principe de cette modélisation est de formaliser le phénomène d'intérêt sous la forme d'un système d'équations mathématiques exprimées tout d'abord sur des supports continus puis discrétisées sous la forme d'un schéma numérique spatio-temporel.

La modélisation hypothético-déductive n'est pas toujours la panacée pour questionner un phénomène. En effet, il n'est pas toujours possible de trouver le modèle mathématique d'un phénomène mal connu dans l'état de connaissance scientifique actuel, d'une part, et la simulation de ce modèle peut produire des prédictions peu précises ou nécessiter des temps de calcul ou de consommation énergétique jugés trop importantes, d'autre part. Les techniques d'apprentissage machine profond rendent possible la mise au point de simulateurs sur des phénomènes pour lesquels on ne dispose pas de modèle théorique satisfaisant. Il peut être tentant faute de mieux d'adapter pour l'apprentissage machine profond la méthodologie éprouvée de spécification et de validation utilisée pour l'approche hypothético-déductive. Cette solution ne se justifie pas puisque la spécification et la mise au point d'une simulation par d'apprentissage machine sont différentes de celles mises en œuvre pour une approche hypothético-déductive.

Cadre méthodologique de spécification d'un simulateur par apprentissage machine profond

Les spécifications d'un simulateur basé sur de l'apprentissage machine se déterminent implicitement dans une boucle de rétroaction visant à réduire une métrique d'erreur. L'optimisation des paramètres est agnostique par rapport au phénomène d'intérêt pouvant induire un glissement implicite de spécification par rapport à celle souhaitée dans le cahier des charges. Ce glissement de spécification peut être détecté en appliquant des techniques d'interprétabilité *post-hoc* lors d'analyses de dysfonctionnements du simulateur Christophe Denis (2022). Cette stratégie n'est pas acceptable pour des applications industrielles suffisamment critiques.

Nous présentons dans cette contribution un cadre méthodologique pour formaliser les spécifications d'un simulateur basé sur de l'apprentissage machine profond. Ce travail de recherche, mêlant épistémologie de la modélisation et informatique théorique, a pour objectif comme mentionné dans Girard-Satabin (2021) de "*réconcilier l'abondance des techniques de vérification de programmes classiques et l'absence de garanties sur les réseaux de neurones pour permettre aux logiciels critiques de conserver leur haut niveau de confiance*". Nous faisons l'hypothèse que l'absence de spécification formelle empêche de mettre en place une démarche rigoureuse de validation. Cette contribution constitue le fondement épistémologique d'une démarche systématique de validation ne reposant pas sur l'hypothèse peu crédible et non utile d'une compréhension fine des mécanismes internes du simulateur. Il s'agit d'explicitier les spécifications formelles du simulateur, son modèle sous-jacent, en adaptant pour l'apprentissage machine profond la théorie de la Modélisation et de la Simulation Zeigler et al. (2000). Ce cadre méthodologique est en cours de mise à l'épreuve sur un cas d'usage prédisant un phénomène physique en mécanique des fluides et sur la production d'un résumé de texte avec l'agent conversationnel ChatGPT. Enfin, nous poursuivons un travail de clarification épistémologique pour définir la notion d'algorithme d'apprentissage machine associé au simulateur en nous basant sur des nouveaux modèles de calcul en informatique théorique.

Références

- Christophe Denis, F. V. (2022). Interprétabilité et explicabilité de phénomènes prédits par de l'apprentissage machine. *Revue Ouverte d'Intelligence Artificielle* 3, 679–696.
- Girard-Satabin, J. (2021). *Verification and validation of Machine Learning techniques*. Ph. D. thesis, Université Paris-Saclay.
- Zeigler, B. P., H. Praehofer, et T. G. Kim (2000). *Theory of Modeling and Simulation—Integrating Discrete Event and Continuous Complex Dynamic Systems* (2nd ed.). San Diego, CA : Academic Press, Inc.

Summary

This contribution presents a methodological framework for formalising the specifications of a simulator based on deep machine learning. The aim of this research, which combines modelling epistemology and theoretical computer science, is to develop verification and validation techniques for critical systems requiring the predictive power of machine learning modelling.