Esquisses philosophiques autour de la compréhension de phénomènes complexes avec des outils de prédiction basés sur de l'apprentissage machine

Christophe Denis

Esquisses philosophiques autour de la compréhension de phénomènes complexes avec des outils de prédiction basés sur de l'apprentissage machine

Christophe DENIS*,**

*Laboratoire d'Informatique (LIP6)
Sorbonne Université
christophe.denis@lip6.fr,
**Équipe de Recherche Interdisciplinaire sur les Aires Culturelles (ERIAC)
Université de Rouen-Normandie

Résumé. Les outils de prédiction basés sur de l'apprentissage machine sont utilisés pour prédire des phénomènes complexes dans de nombreux secteurs de la société dont la recherche scientifique. Ces outils sont notamment utilisés pour contrecarrer la faiblesse de l'approche hypothético-déductive lorsque l'état de la connaissance n'est pas suffisant pour paramétriser correctement le phénomène sous la forme d'équations mathématiques. Ces outils soulèvent la problématique du prédire sans expliquer pour lequel il est important d'aborder la question philosophique du rapport entre la technique et la connaissance. Nous présentons dans cet article une esquisse de travaux philosophiques menés sous la direction de Franck Varenne. Il s'agit d'étudier les conditions d'utilisation de ces outils pour améliorer la connaissance scientifique dans un paradigme différent ou non de l'approche hypothético-déductive.

Introduction

La survie puis le développement de l'humanité, démunie de qualités physiques comparables à d'autres espèces animales, s'est construit grâce à l'utilisation de techniques de plus en plus sophistiquées. L'appropriation de la technique puis de la connaissance par l'humanité peut être symbolisée par le mythe grec de Prométhée. Deux frères titans, Prométhée, le prévoyant, et Epiméthée, l'étourdi, sont en charge de distribuer les ressources terrestres entre les différentes espèces animales. La dernière étape de la répartition révèle l'étourderie Epiméthée : plus rien n'est disponible pour les êtres humains car toutes les ressources ont déjà été distribuées par ce Titan aux autres espèces animales. Prométhée corrige l'erreur de son frère en dérobant le feu à Zeus et le fer à Athéna, vol qui sera ensuite puni par Zeus en lui infligeant une condamnation cruelle et renouvelée chaque jour. Bien avant l'âge du Fer, les premiers outils techniques conçus par l'humanité, les outils en pierre taillée, datent d'environ 2,7 millions d'années durant la période du Paléolithique. Les travaux des archéologues montrent l'existence d'une technologie, dénommée technologie lithique, regroupant un ensemble de techniques et

de connaissances, partagé, et sans cesse amélioré pour concevoir ces premiers outils. La notion de chaîne opératoire a été proposé par l'archéologue et l'ethnologue André Leroi-Gourhan pour étudier les productions matérielles préhistoriques (Leroi-Gourhan (1945)). Faisons un pas de géant dans l'histoire de l'humanité, magistralement symbolisé par le célèbre fondu enchainé de Stanley Kubrick dans son film "2001, l'Odyssée de l'Espace" [1]. L'humanité est désormais à la croisée des chemins condamnée pour assurer sa survie, sur une planète à bout de souffle, de résoudre rapidement des problèmes qu'elle a elle-même engendrée par un excès de technique industrielle (IPCC (2021)). Suite à la première révolution industrielle, qui a eu lieu en Angleterre au milieu du dix-neuvième siècle, la technologie est devenue une discipline à part entière, qui en reprenant le mythe de Prométhée, assurerait le progrès de l'humanité dans une production toujours plus accrue et efficace de marchandises. Un champ philosophique nouveau a vu le jour, la philosophie de la technologie, pour étudier l'impact sur la société et la culture, en relation étroite avec les sciences sociales. Une relecture du mythe de Prométhée est donc possible dans laquelle la prétendue étourderie d'Epiméthée serait finalement un acte délibéré de sa part pour établir une meilleure harmonie entre l'humanité et la nature.

Depuis l'arrivée des premiers ordinateurs dans les années 1940, une révolution numérique de nos sociétés s'est progressivement opérée avec une complexification technologique toujours croissante. Ainsi, en plus d'analyser l'étude de l'impact de ces technologies sur la société, une branche de la philosophie de la technologie analyse plus précisément le principe de conception d'un dispositif basé ces technologies et la nature des éléments produits par celui-ci, illustré par le conseil avisé de Bernard Stiegler «*Toute technologie est un pharmakon, c'est à la fois ce qui permet de prendre soin, et ce dont il faut prendre soin - au sens où il faut y faire attention : c'est une puissance curative dans la mesure et la démesure où c'est une puissance destructrice* ». La discipline de la cybernétique se développe à partir des travaux du mathématicien Norman Wiener publiés en 1948 pour définir un cadre théorique unifié pour la technologie de l'information et de la communication ainsi que du rétrocontrôle. La cybernétique définit l'intelligence comme une capacité d'adaptation par rapport à un objectif donné d'une trajectoire d'un système dynamique en boucle fermée. L'élément clé est la présence d'un boucle de rétroaction modifiant les paramètres de contrôle pour réduire un écart par rapport à l'objectif donnée. Dans le cadre de la cybernétique, Frank Rosenblatt met au point en 1957 le perceptron, un réseau de neurones possédant une seule couche et une sortie binaire, pour discriminer des données dans deux classes. Ce type de réseau de neurones ne permet pas de traiter des problèmes non-linéaires réduisant son intérêt pratique. Reposant sur le même principe cybernétique que le perceptron, les réseaux de neurones profonds peuvent contenir des millions de neurones répartis en plusieurs dizaines de couches en raison principalement de :

— la conception de l'algorithme de rétropropagation du gradient ;
— l'augmentation de la puissance de calcul (cartes graphiques par exemple) et de la disponibilité d'un volume toujours croissant de données.

Il est ainsi possible de prédire des frontières de décision complexes, non linéaires, augmentant en conséquence ses capacités prédictives au prix d'une opacité de son fonctionnement interne. Les travaux pionniers de la cybernétique nécessitent d'être analysés pour donner des clés de lecture par rapport à l'utilisation des méthodes d'apprentissage machine profonds. Toutefois

1. Ce fondu enchainé transforme l'os jété par un singe dans le ciel de la savane préhistorique en un vaisseau spatial dans l'espace.

comme le mentionne si justement, le philosophe Mathieu Triclot «*l'objectif n'est pas tant de prétendre retrouver dans l'antique cybernétique une clé de lecture du contemporain que de mesurer les écarts et de localiser ce qui peut être irrémédiablement déplacé et périmé dans le discours de la cybernétique* [2]».

L'apprentissage machine profond est une clé de voute technologique pour assurer la mise en place de nouveaux paradigmes dans de nombreux secteurs de notre société. C'est particulièrement le cas dans le domaine de la santé, dans l'espoir de continuer à délivrer des soins de qualité dans un contexte d'une démographie croissante, d'un vieillissement de la population et d'une chronicisation des maladies. Cependant le recours à l'apprentissage machine dans le domaine de la santé pose légitimement des interrogations sur le plan juridique, scientifique et éthique. Le legislateur français a par exemple imposé aux concepteurs de dispositifs médicaux utilisant de l'apprentissage machine d'assurer l'explicabilité de son fonctionnement à ses utilisateurs [3] ce qui reste aujourd'hui un verrou scientifique. Bien que les notions de compréhension, d'interprétabilité et d'explicabilité soient utilisées dans de nombreux travaux de recherche en apprentissage machine, il n'existe pas de consensus sur ces notions. La section 1 présente un travail de clarification epistémologique autour de ces notions mené avec Franck Varenne. Il est nécessaire d'obtenir une connaissance à partir de la conception technique de la méthode d'apprentissage machine pour apporter aux professionnels de santé les raisons, compréhensibles pour eux par rapport à leurs propres connaissances médicales, pour lesquelles le dispositif médical a prédit ou non un risque de maladie, par exemple. La section 2 présente la relation ambivalente entre la technique et la connaissance dans le contexte d'utilisation d'un dispositif utilisant l'apprentissage machine.

Notre hypothèse de travail philosophique, présentée en section 3, est que l'apprentissage machine peut être également un dispositif technique permettant d'améliorer la connaissance scientifique de phénomènes, comme la turbulence d'écoulement ou le comportement de phénomènes chaotiques, que l'approche hypothético-déductive peine à prédire. Il est pour cela nécessaire de définir un nouveau paradigme scientifique différent ou complétant celui utilisé depuis les travaux fondateurs de Galilée sur la mathématisation de la physique, dans le but de ne pas incorporer des artefacts statistiques dans la connaissance scientifique.

1 Proposition de définitions sur la compréhension, l'interprétabilité et l'explicabilité

Les notions de compréhension, d'interprétabilité et d'explicabilité sont utilisées dans de nombreux travaux de recherche en apprentissage machine. Il n'existe pas de consensus sur ces notions, et les notions d'interprétabilité et d'explicabilité sont souvent utilisées confusément. Nous avons mené avec Franck Varenne un travail de clarification epistémologique autour de ces notions (Denis et Varenne (2019)) . Commençons tout d'abord par définir la notion de compréhension proposée dans (Varenne (2013),Varenne (2018))

2. Conférence de Mathieu Triclot, «Le monde cybernétique est-il advenu ?», https ://issue-journal.ch/focus-posts/mathieu-triclot-le-monde-cybernetique-est-il-advenu, IRAD, 2018.

3. Article 17 de la loi n° 2021-1017 du 2 août 2021 relative à la bioéthique, https ://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043884384/

Définition 1 *Compréhension d'un phénomène, ou d'un calcul (du latin cum-prehendere : saisir, appréhender ensemble) : il y a compréhension d'un phénomène quand notre esprit dispose de la possibilité d'en saisir l'ensemble et d'en unifier les manifestations successives ou diverses sous une représentation à la fois unique et aisée à concevoir et à rappeler à l'esprit.*

La définition suivante de l'interprétabilité a été proposée par (Mittelstadt et al. (2019), Lisboa (2013), Miller (2019), Molnar (2019)) :

Définition 2 *« L'interprétabilité réfère au degré de compréhensibilité humaine d'un modèle de type boîte noire ou d'une décision »*

Cette définition de l'interprétabilité mobilise la notion problématique de compréhension, elle-même non préalablement définie. Et elle inverse le rapport habituel de détermination interprétabilité $\rightarrow$ explicabilité $\rightarrow$ compréhension. L'interprétabilité, au sens où nous l'entendons est fondamentalement liée à la possibilité de relier des symboles individuels à des entités ou propriétés d'entités possibles, fictives, c'est-à-dire, simplement pensées, ou encore réelles. Définissons donc maintenant l'interprétabilité d'un modèle.

Définition 3 *Interprétabilité d'un modèle : propriété que possède un ensemble de symboles ou un modèle (une structure de symboles) de se voir composé d'éléments (signes, symboles, figures, concepts, données, etc.) qui ont chacun un sens [c'est-à-dire un référent possible] pour un sujet humain ou un groupe d'humains.*

On a donc obtenu une définition sémantique de l'interprétabilité d'un modèle : un modèle est interprétable quand tous ses symboles le sont. Nous pouvons proposer une définition de l'explicabilité pour un modèle : un modèle est explicable quand son fonctionnement donne lieu à des représentations des articulations entre ses états qui sont interprétables.

Définition 4 *L'explicabilité d'un modèle pourvu d'un algorithme est la capacité de déploiement et d'explicitation de cet algorithme ou de ses sorties en série d'étapes reliées entre elles par ce qu'un être humain peut interpréter sensément comme des causes ou des raisons.*

2 Relation entre technique d'apprentissage machine et connaissance

Prenons pour illustrer nos propos la conception d'un dispositif médical basé sur de l'apprentissage machine profond dont l'objectif est de prédire un niveau de diabète d'un patient à l'aide d'images de son visage. On suppose pour faciliter la compréhension de l'exposé que la classification recherchée est binaire : absence ou présence de risque de diabète en fonction de l'analyse d'un visage. Une fonction de classification retournant comme valeur « absence de risque » ou « présence de risque » pour une image doit être déterminée. La méthode d'apprentissage supervisée a pour objectif de déterminer empiriquement la fonction de classification à l'aide d'images annotées, par exemple après l'examen de patients, sur le risque de diabète pour chaque image. La mise au point de la méthode d'apprentissage profond est effectuée par une personne experte en sciences des données qui utilise à la fois sa connaissance scientifique et son savoir-faire technique acquis par l'expérience pour :

— choisir empiriquement la topologie du réseau de neurones profonds (nombre de neurones, nombre de couches, etc.) et l'algorithme d'optimisation utilisé pour pondérer les neurones du réseau ;
— traiter d'éventuels biais dans les données ;
— effectuer le prétraitement des images, dans un contexte de valeurs tabulaires, les données sont toujours prétraitées (remplacement des valeurs manquantes ou des valeurs aberrantes) puis parfois même combinées pour construire des caractéristiques plus aisément interprétables.

Les concepteurs d'une méthode d'apprentissage sont confrontés à deux problèmes majeurs concernant la validation et l'explicabilité de celle-ci en raison :

1. d'un problème de généralisation ou de sur-apprentissage : dans ce cas, la fonction de classification fournit de très bons résultats sur les images utilisées par la méthode d'apprentissage mais obtient des résultats médiocres sur de nouvelles images. On peut estimer la capacité de généralisation d'une méthode d'apprentissage en mettant en place une technique de validation croisée ;
2. de la topologie des réseaux de neurones profonds pouvant contenir des millions de neurones répartis en plusieurs dizaines de couches. Il n'est pas possible pour un esprit humain de comprendre et de valider le fonctionnement du réseau de neurones en analysant les valeurs de chaque neurone : on parle d'effet boite noire. L'enchevêtrement de neurones permet de prédire des frontières de décision complexes, non linéaires augmentant en conséquence ses capacités prédictives au prix d'une opacité de son fonctionnement. La communauté informatique, en particulier travaillant au niveau du génie logiciel, ont mis au point des méthodes de validation de boites noires mais celles-ci, contrairement à celles des méthodes d'apprentissage machine, sont spécifiées fonctionnellement : il est donc possible de les valider en comparant à l'aide de tests les résultats de ces boites avec les spécifications fonctionnelles.

Pour faciliter l'explicabilité, une solution serait d'utiliser des méthodes d'apprentissage plus simples dont on comprend le fonctionnement comme les régressions linéaires ou les arbres de classification pour éviter l'effet boite noire (point 2) mais le traitement du phénomène de sur-apprentissage (point 1) reste entier. Dans le cas du traitement d'images, ces méthodes auraient toutefois des performances statistiques trop faibles pour être utilisées en pratique.

Les relations entre la connaissance et la technique ont depuis longtemps accaparé des penseurs qu'il est important de se remémorer pour se poser aujourd'hui les bonnes questions concernant l'utilisation de l'apprentissage machine. Les sophistes grecs ne considéraient pas de différence entre la morale et la technique à la grande différence des premiers philosophes grecs comme Platon. Nous retrouvons encore ces positionnements dans nos sociétés, en notant bien entendu qu'il serait faux de considérer la pensée de Platon comme technophobe, dans un rejet pur et simple de la technique. La technique n'a pour Platon pas de valeur en soi, et comme rappelé par André Vergez, il n'existe pas pour lui *"de technique hors de l'intelligence, hors du sens des proportions et grâce auquel les arts produisent tous leurs chefs-d'œuvre"* (Vergez (1956)). Platon imagine un démiurge qui fabrique imparfaitement pour l'humanité, dans le monde sensible, une instanciation des Idées venant du monde intelligible pour lequel le démiurge possède l'accès. Par exemple, le monde intelligible peut contenir l'Idée de cercle qui sera toujours imparfaitement reproduit dans le monde humain, sensible. Nous émettons l'hypo-

thèse que réciproquement au démiurge, un outil d'apprentissage machine permet de faire une partie du chemin inverse : à partir des données et des observations du monde sensible, l'outil permet d'établir une représentation intermédiaire comme clé d'accès au monde intelligible. Seulement, contrairement à l'approche hypothético-déductive, les explications d'un résultat obtenus par apprentissage machine peuvent conduire à plusieurs hypothèses possibles qu'un raisonnement humain, abductif devra trancher. En d'autres termes, il sera nécessaire d'ajouter des connaissances par l'intermédiaire de l'expertise de l'utilisateur pour augmenter ou réduire la vraisemblance des hypothèses, afin d'en choisir une. Il est à noter que le raisonnement abductif permet d'élargir le rayon d'action personnel de l'utilisateur par rapport à l'outil et de préserver son autonomie décisionnelle, ce qui est un enjeu éthique important notamment, mais pas seulement, en médecine.

Pour étayer cette hypothèse, notre travail en cours consiste à analyser le rapport établi entre la technique et la connaissance par plusieurs écoles de pensée philosophique. Pour étudier la connaissance apportée par la conception d'un dispositif utilisant de l'apprentissage machine, notre méthodologie consiste à utiliser le concept de chaine opératoire, telle qu'utilisée par André Leroi-Gourhan [4], pour mettre en lumière les connaissances apportées par la conception technique d'un dispositif basé sur de l'apprentissage machine.

3 Apprentissage machine et amélioration de la connaissance scientifique

Une première approche pour questionner et comprendre la nature consiste à l'observer pour être en mesure d'accéder à sa vérité (*naturo veritas*). Ainsi, la cosmologie définie par Aristote formule que par essence les objets les plus lourds tombent plus rapidement que les objets moins lourds. En utilisant une expérience de pensée, Galilée montre une contradiction logique dans la loi formulée par Aristote. Il propose ensuite une nouvelle loi indiquant que la vitesse d'un objet est indépendante de sa masse dans le vide. Il s'agit de la première loi physique établie à l'écart du monde sensible dans lequel le vide n'existe pas naturellement. En 1623, Galilée inscrit la physique dans un paradigme mathématique en indiquant dans son ouvrage l'Essayeur : *"La philosophie est écrite dans cet immense livre qui se tient toujours ouvert devant nos yeux, je veux dire l'Univers, mais on ne peut le comprendre si l'on ne s'applique d'abord à en comprendre la langue et à connaître les caractères avec lesquels il est écrit. Il est écrit dans la langue mathématique et ses caractères sont des triangles, des cercles et autres figures géométriques, sans le moyen desquels il est humainement impossible d'en comprendre un mot. Sans eux, c'est une errance vaine dans un labyrinthe obscur"*. On retrouve cette abstraction de la physique dans le Discours de la Méthode lorsque Descartes indique qu'une connaissance exhaustive des mécanismes de la nature aussi précise que celle acquise sur les métiers des artisans permettra de rendre l'humanité *"comme maîtres et possesseurs de la nature"*.

Les travaux fondateurs de Galilée portant sur la mathématisation de la physique a permis d'accroitre considérablement la connaissance scientifique. Depuis les années 1970, les avancées de l'algorithmique numérique et du calcul haute performance ont augmenté la précision des simulations numériques dans l'espoir d'expliquer des phénomènes de plus en plus complexes. Le paradigme de modélisation basée sur une approche hypothético-déductive at-

4. cf. introduction de l'article

teint cependant ses limites, en termes de validation, de précision et même d'explication, pour prédire des phénomènes complexes et couplés, tels que ceux entrant en jeu dans l'étude du changement climatique. L'approche hypothético-déductive consiste à émettre des hypothèses, à recueillir des données, puis à tester les résultats obtenus pour réfuter ou appuyer les hypothèses. Il s'agit de décrire par l'usage d'un modèle représentant d'une manière ou d'une autre la physique effective du phénomène (ses lois) et utilisant souvent des équations mathématiques tirées de celles-ci mêmes lois. Le paradigme de modélisation basé sur une approche hypothético-déductive n'arrive pas à prédire ou à reproduire avec précision certains comportements physiques complexes en raison :

- d'un manque de connaissance scientifique sur les phénomènes, rendant difficile l'établissement d'équations mathématiques pour les décrire ;
- d'une difficulté à fixer les paramètres des équations ;
- d'une difficulté à discrétiser les équations mathématiques et à les résoudre précisément sur un ordinateur [5].

Le concept de paradigme que nous utilisons ici a été proposé par Thomas S. Kuhn dans son ouvrage *La structure des révolutions scientifiques* . Le paradigme regroupe un ensemble de connaissances et de techniques *qui servent de cadre de référence à la communauté des chercheurs de telle ou telle branche scientifique*" (Kuhn (1972)) durant une période longue de "science normale", puis est remplacé lors d'une crise scientifique par un autre paradigme à la suite d'une période moins longue de "révolution scientifique". Ce concept permet de concilier l'interaction entre la technique et la connaissance scientifique, dans une relation de de plus en plus proche entre la philosophie des sciences et la philosophie de la technologie. Notre hypothèse de travail est que les différents paradigmes de modélisation ont un lien commun : la caractérisation de modèle formulée par Marvin Minsky, un des chercheurs fondateurs de la discipline de l'Intelligence Artificielle "*Pour un observateur B, un objet A* est un modèle d'un objet A s'il permet à B de répondre à une question qu'il se pose sur A*" (Minsky (1965)) . Le paradigme de modélisation vise donc à construire le modèle A* pour répondre à des questions à un observateur B sur sur un objet A. Prenons l'exemple de la prédiction de la hauteur d'eau maximum d'un écoulement d'un fleuve. Cette prédiction est nécessaire au concepteur d'un barrage utilisée pour dimensionner un barrage protégeant un site industriel localisé près du fleuve. Dans ce contexte :

- l'observateur B est le concepteur du barrage ;
- l'objet A est le fleuve pour lequel le concepteur cherche à prédire sa hauteur d'eau maximum ;
- l'objet A* dépend du paradigme de modélisation utilisé :
  - il peut s'agir d'un programme informatique qui résout les équations mathématiques, par exemple les équations de Navier-Stokes, choisies pour représenter le phénomène d'intérêt, en l'occurence l'écoulement à surface du fleuve.
  - Il est aussi possible de construire un programme informatique dont les paramètres sont fixés à partir des données mesurées sur le phénomène d'intérêt, ou à partir de données synthétiques, en utilisant une méthode d'apprentissage machine.

5. En particulier,l'utilisation de l'arithmétique flottante, modèle de l'arithmétique réelle, ne permet pas de résoudre précisément certaines équations en raison de la propagation des erreurs d'arrondi et peut fournir un résultat ne possédant aucun chiffre significatif c'est-à-dire du bruit numérique. Ce problème est amplifié dans la résolution de problèmes chaotiques.

Depuis la conférence scientifique ECCV-2012 [6], les capacités prédictives des réseaux de neurones convolutifs, ont bouleversé en profondeur la discipline du traitement et de la reconnaissance d'images. Ces résultats peuvent conforter en première approche la fin prophétisée de l'usage de modèle dans la discipline scientifique puisqu'il serait possible d'analyser les données sans formuler d'hypothèses les concernant Anderson (2008). Les failles de cette argumentation ont été présentées par plusieurs travaux de recherche, en particulier dans le cas des sciences sociales (Venturini Tommaso et Cointet. (2014)). Pour autant, il est indéniable que de nombreuses disciplines scientifiques notamment computationnelles développent des activités de recherche orientées vers l'apprentissage machine. Au delà d'un possible effet d'aubaine, ces activités de recherche n'ont pas la même finalité. Il peut s'agir de :

1. diminuer la puissance de calcul nécessaire pour mener des simulations numériques, en construisant un métamodèle ;
2. augmenter les capacités prédictives d'un processus déjà basé sur une approche statistique comme par exemple pour l'étude du climat (Jebeile et al. (2020)) ou en biologie moléculaire Gómez-Bombarelli et al. (2018) ;
3. combler le manque de connaissance d'un phénomène modélisé selon une approche hypothético-déductive, c'est-à-dire décrit par un modèle représentant d'une manière ou d'une autre la physique effective du phénomène (ses lois) et recourant souvent à des équations mathématiques tirées de ces mêmes lois.

Les deux premières finalités n'introduisent pas, selon notre hypothèse , un changement epistémologique tandis que la troisième finalité est au cœur de notre projet de recherche. Il s'agit de montrer que le processus d'explicabilité de l'apprentissage machine diffère épistémologiquement de celle mise en place dans le cadre de la modélisation mathématique et causale d'un phénomène physique.Un des objectifs de la science est d'accéder à des signes qualifiant la propriété d'un objet d'étude à travers les signaux quantifiant une interaction entre cet objet et un dispositif de mesure. Le force d'une technique d'apprentissage est de prédire une ou plusieurs propriétés de l'objet d'étude à partir de signaux le concernant directement ou non. La difficulté est que la technique d'apprentissage machine ne donne généralement pas accès aux signes de l'objet. Nous avons émis avec Franck Varenne dans (Denis et Varenne (2019)) la proposition qu'une technique d'apprentissage machine permet d'obtenir une représentation intermédiaire, entre signal et signe, de la seule structure d'information du système cible. Un modèle explicatif est donc nécessaire pour accéder aux signes de l'objet à partir de la structure d'information du système cible. Ce modèle explicatif ne doit pas seulement être validé dans sa capacité à reproduire certains comportements du système cible. Il est nécessaire d'évaluer sa capacité à représenter pas à pas, de manière correcte, c'est-à-dire approximativement réaliste, non seulement les états successifs du système cible mais aussi chaque opération du processus lui-même.

6. European Conference on Computer Vision

Conclusion et perspectives

Nous avons présenté dans cet article une esquisse de travaux philosophiques menés sous la direction de Franck Varenne. Nos travaux de thèse portent sur l'utilisation éthique et conviviale d'une machine apprenante utilisant des réseaux de neurones profonds. Reposant sur le même principe cybernétique que le perceptron, les réseaux de neurones profonds peuvent contenir des millions de neurones répartis en plusieurs dizaines de couches. L'enchevêtrement de neurones permet de prédire des frontières de décision complexes, non linéaires augmentant en conséquence ses capacités prédictives au prix d'une opacité de son fonctionnement.

L'apprentissage machine profond est une clé de voute technologique pour assurer la mise en place de nouveaux paradigmes dans de nombreux secteurs de notre société, notamment en santé, dans l'espoir de continuer à délivrer des soins de qualité dans un contexte d'une démographie croissante, d'un vieillissement de la population et d'une chronicisation des maladies. Cependant le recours à l'apprentissage machine dans le domaine de la santé pose légitimement des interrogations sur le plan juridique, scientifique et éthique. Le législateur français a par exemple imposé aux concepteurs de dispositifs médicaux utilisant de l'apprentissage machine d'assurer l'explicabilité de son fonctionnement à ses utilisateurs, ce qui reste aujourd'hui un verrou scientifique. Bien que les notions de compréhension, d'interprétabilité et d'explicabilité soient utilisées dans de nombreux travaux de recherche en apprentissage machine, il n'existe pas de consensus sur celles-ci. Nous avons dans la première section de cet article une proposition de définitions alternatives, sémantiques, de ces concepts.

La section 2 présente la relation entre la technique et la connaissance lors de la conception d'un dispositif utilisant l'apprentissage machine. Notre méthodologie consiste à utiliser le concept de chaine opératoire, telle qu'utilisée par André Leroi-Gourhan, pour caractériser les connaissances apportées lors de la conception technique d'un outil basé sur de l'apprentissage machine.

La dernière section de l'article a présente notre hypothèse de recherche portant sue la possibilité que l'apprentissage machine peut permettre d'améliorer la connaissance scientifique de phénomènes complexes et couplés, comme la turbulence d'écoulement ou le comportement de phénomènes chaotiques, que l'approche hypothético-déductive peine à prédire. Il est pour cela nécessaire de définir un nouveau paradigme scientifique différent ou complétant celui utilisé depuis les travaux fondateurs de Galilée sur la mathématisation de la physique, dans le but garantir une intégrité scientifique, évitant ainsi de produire des artefacts statistiques.

Au delà du besoin d'explication, il existe un problème éthique et sociétal d'aliénation par l'utilisation d'outils techniques normatives de l'humanité en la privant de son autonomie et en lui imposant ses besoins. Nous menons un travail de caractérisation philosophique et mathématique de propriétés de préservation de l'autonomie décisionnelle de l'utilisateur et d'élargissement de son rayon d'action personnel.

Remerciements

L'auteur remercie vivement les remarques et les commentaires des relecteurs.

Références

Anderson, C. (2008). The end of theory : The data deluge makes the scientific method obsolete. *Wired 16*(7).

Denis, C. et F. Varenne (2019). Interprétabilité et explicabilité pour l'apprentissage machine : entre modèles descriptifs, modèles prédictifs et modèles causaux. Une nécessaire clarification épistémologique. In *National (French) Conference on Artificial Intelligence (CNIA) - Artificial Intelligence Platform (PFIA)*, Toulouse, France, pp. 60–68.

Graham, G. (2005). Episteme and techne. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2007 ed. ed.).

Gómez-Bombarelli, R., J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, et A. Aspuru-Guzik (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science 4*, 268–276.

IPCC (2021). *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Pean, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekci, R. Yu, and B. Zhou (eds.).

Jebeile, J., V. Lam, et T. Raz (2020). Understanding climate change with statistical downscaling and machine learning. *Synthese*.

Kuhn, T. (1972). *La structure des révolutions scientifiques*. Flammarion.

Leroi-Gourhan, A. (1945). *Milieu et Technique*. Albin Michel.

Lisboa, P. J. G. (2013). Interpretability in machine learning – principles and practice. In F. Masulli, G. Pasi, et R. Yager (Eds.), *Fuzzy Logic and Applications*, Cham, pp. 15–21. Springer International Publishing.

Miller, T. (2019). Explanation in artificial intelligence : Insights from the social sciences. *Artificial Intelligence 267*, 1–38.

Minsky, M. (1965). Matter, mind and models. In *Proc. of the International Federation of Information Processing Congress*.

Mittelstadt, B. D., C. Russell, et S. Wachter (2019). Explaining explanations in AI. In *FAT* '19 : Proceedings of the Conference on Fairness, Accountability, and Transparency*.

Molnar, C. (2019). *Interpretable Machine Learning*. `https://christophm.github.io/interpretable-ml-book/`.

Simondon, G. (1958). *Du mode d'existence des objets techniques*. Éditions Aubier-Montaigne.

Thom, R. (1991). *Prédire n'est pas expliquer*. Flammarion.

Varenne, F. (2013). Modèles et simulations dans l'enquête scientifique : variétés traditionnelles et mutations contemporaines. In F. Varenne et M. Silberstein (Eds.), *Modéliser & Simuler. Épistémologies et pratiques de la modélisation et de la simulation, Tome I*. Matériologiques.

Varenne, F. (2018). *From Models to Simulations*. Routledge.

Venturini Tommaso, D. C. et J.-P. Cointet. (2014). Présentation. *Réseaux 188*(6).

Vergez, A. (1956). *Technique et morale chez Platon*. Presses Universitaires de France.

Summary

Prediction tools based on machine learning are used to predict complex phenomena in many sectors of society including scientific research. These tools are notably used to counter the weakness of the hypothetical-deductive approach when the state of knowledge is not sufficient to correctly parameterize the phenomenon in the form of mathematical equations. These tools raise the problem of predicting without explaining for which it is important to address the philosophical question of the relationship between technique and knowledge. In this article, we present a sketch of philosophical work carried out under the direction of Franck Varenne. The aim is to study the conditions of use of these tools to improve scientific knowledge in a paradigm different or not from the hypothetico-deductive approach.