ACADÉMIE DES TECHNOLOGIES
POUR UN PROGRÈS RAISONNÉ, CHOISI ET PARTAGÉ

MÉTROLOGIE APPLIQUÉE AU NUMÉRIQUE ET À L'INNOVATION

Thomas Grenon
Membre de l'Académie des technologies

Séance du 10 janvier 2024

Résumé

Face aux innovations numériques, la métrologie doit relever le défi d'établir de nouveaux étalons de mesure, indispensables pour l'adoption de ces technologies par l'industrie. Le programme MetriQs-France, dédié au développement de moyens de mesure de référence pour les technologies quantiques, coordonné par le LNE, a ainsi été conçu comme un soutien à l'innovation et à l'établissement de la filière industrielle du quantique. Le projet BACQ, à travers un benchmarking applicatif développé dans ce cadre, vise à répondre à des questions simples, telles que : quels progrès permettent les ordinateurs quantiques par rapport aux ordinateurs classiques? Quels sont leurs atouts pour des applications données?

Il existe également une forte demande de moyens fiables d'évaluer les performances de l'intelligence artificielle. Dans ce domaine, plutôt qu'une démonstration mathématique, le LNE privilégie la réalisation d'essais, par exemple avec le banc d'essai Immersion combinant des essais physiques, réalisés avec un robot, et de la simulation projetée autour de ce dernier, qui permet de tester les capacités des capteurs tout en prenant en compte les contraintes physiques du robot.

L'un des défis de ces nouveaux outils de mesure est d'inclure l'évaluation de leur propre incertitude. Dans le domaine de l'apprentissage supervisé, celle-ci peut provenir de multiples sources, comme l'échantillonnage et son éventuelle corruption, le choix de l'architecture du modèle, l'entraînement du réseau de neurones, etc. Les travaux du LNE se concentrent sur la modélisation de deux sources d'incertitudes particulièrement déterminantes, celle des données et celle de la répétabilité.

Intervenants

Félicien Schopfer
Directeur du programme « Technologies quantiques » du Laboratoire national de métrologie et d'essais (LNE)

Agnès Delaborde
Responsable du département « Évaluation de l'intelligence artificielle » au LNE

Nicolas Fischer
Responsable du département « Sciences des données et incertitudes » au LNE

Sommaire

Introduction 2

Mesurer les performances des technologies quantiques pour accélérer leur développement et leur adoption 2

Considérations métrologiques pour l'évaluation et la certification de l'IA 5

Incertitude, méthodes d'apprentissage et métrologie 6

Débats 9

Introduction

Thomas Grenon

La séance d'aujourd'hui est consacrée aux besoins de métrologie liés à la transition numérique, question que nous aborderons avec trois chercheurs du Laboratoire national de métrologie et d'essais (LNE). Félicien Schopfer évoquera la mesure des performances des technologies quantiques dans la perspective de leur développement par l'industrie. Agnès Delaborde traitera de l'évaluation et de la certification de l'intelligence artificielle. Enfin, Nicolas Fischer reviendra sur une notion consubstantielle à la métrologie, l'incertitude, et à son calcul appliqué aux méthodes d'apprentissage telles que l'intelligence artificielle.

Mesurer les performances des technologies quantiques pour accélérer leur développement et leur adoption

Félicien Schopfer

Les technologies quantiques résultent de l'exploitation des propriétés quantiques de la lumière et de la matière à l'échelle des atomes. La première génération de ces technologies a donné lieu aux lasers et aux transistors, par exemple. Aujourd'hui, elles connaissent un nouvel essor grâce aux progrès de l'ingénierie des systèmes quantiques ainsi qu'au contrôle des objets quantiques individuels (atomes, électrons, photons) et des propriétés de superposition d'états et d'intrication.

Ces nouveaux développements permettent de réaliser des systèmes quantiques élémentaires baptisés qubits (ou bits quantiques). À terme, ceux-ci doivent permettre de résoudre des problèmes difficiles comme la simulation des matériaux et de leurs propriétés, d'accélérer les calculs d'optimisation ou encore de factorisation, mais aussi de sécuriser le transfert des données via la cryptographie ou d'effectuer des mesures avec une sensibilité et une exactitude sans précédents. Ces innovations représentent des enjeux non seulement socio-économiques mais également de défense et de souveraineté. C'est pourquoi elles font l'objet d'une très vive compétition internationale.

Quels rapports entre technologies quantiques et métrologie?

Les technologies quantiques ont été exploitées depuis de longues années en métrologie pour développer des étalons de mesure reproductibles et « universels » en tant que liés à des constantes physiques (horloges atomiques, étalons électriques quantiques…), puis pour produire des capteurs ultra-sensibles, capables de détecter des particules uniques (photons, électrons, atomes) ou d'effectuer des mesures de haute sensibilité à l'échelle nanométrique. Les progrès de l'ingénierie permettent aussi d'envisager une miniaturisation de ces dispositifs ainsi qu'un fonctionnement simplifié à plus faible coût.

Inversement, le développement des technologies quantiques s'accompagne de besoins de mesures fiables de leurs caractéristiques et performances afin de démontrer leur éventuel avantage par rapport aux technologies existantes, d'accélérer les innovations et de favoriser leur adoption par l'industrie et par les utilisateurs. Ces mesures sont aussi un prérequis à la normalisation, qui joue également un rôle important dans le processus d'industrialisation.

Pour répondre à ces besoins, la métrologie doit relever plusieurs défis. Elle doit, notamment, définir de nouveaux référentiels pour mesurer, par exemple, le taux d'erreur des qubits ou les performances des calculateurs quantiques, qui sont des systèmes complexes, ou encore démontrer, par une chaîne ininterrompue de comparaisons, la traçabilité des mesures par rapport au Système international d'unités.

Le programme MetriQs-France

Le programme national MetriQs-France vise à développer, exploiter et promouvoir des moyens de mesure de référence validés et harmonisés pour caractériser et évaluer les performances des technologies quantiques avec fiabilité, impartialité et comparabilité. Il a été conçu comme un soutien à l'innovation et à l'établissement de la filière industrielle des technologies quantiques. Ce programme est coordonné par le LNE, qui met à profit son expertise en métrologie, essais, certification, normalisation, règlementation et en technologies quantiques, et bénéficie du statut de tiers de confiance.

Le premier volet de MetriQs-France comprend des projets de R&D destinés à développer des référentiels de mesure en partenariat avec l'ensemble des acteurs publics et privés du domaine : CEA, CNRS, INRIA, AFNOR, Thales, Air Liquide, Eviden, etc., ainsi que de nombreuses startups. Un premier projet, BACQ, concerne le benchmark applicatif des calculateurs quantiques. D'autres porteront sur la caractérisation des composants quantiques et des technologies permettant de faire fonctionner les systèmes quantiques.

Le deuxième volet est le déploiement d'une infrastructure de test et de mesure, comprenant une plateforme de métrologie quantique opérée par le LNE et ses partenaires du Réseau national de la métrologie française, LNE-SYRTE et LNE –CNAM, ainsi qu'un réseau de plateformes « accréditées ». Cette infrastructure doit être experte et indépendante. La plateforme traitera des sujets tels que la caractérisation des qubits à l'état solide et de l'électronique de contrôle associée, la thermométrie à très basse température, la gravimétrie quantique ou la caractérisation des horloges atomiques. À moyen terme, elle sera ouverte à des utilisateurs extérieurs.

La mesure des performances des calculateurs quantiques

Ces dernières années, des grands acteurs tels que Google, IBM ou encore des entreprises chinoises ont annoncé des avancées présentées comme décisives dans le développement de l'ordinateur quantique. Ces informations sont-elles fiables, sachant que la caractérisation des performances est effectuée par les équipes de développement elles-mêmes? Comment comparer les résultats exposés, alors que les technologies ne sont pas les mêmes, que le nombre de qubits mobilisés n'est pas suffisant, que leur qualité peut différer? Et surtout, comment, à partir de critères très techniques, apprécier les applications et l'utilité réelles de ces ordinateurs quantiques pour le monde industriel?

Ces interrogations nous ont conduits à envisager de définir des benchmarks applicatifs pour les calculateurs quantiques. Il s'agit d'un véritable défi, compte tenu de la multiplicité des plateformes matérielles utilisées pour réaliser les ordinateurs quantiques (photons, spins électroniques, atomes froids, ions piégés, circuits supraconducteurs), de la diversité des modalités de calculs selon que les machines sont analogiques ou à portes, du niveau inégal de maturité des technologies (qui pose le problème de la durabilité du benchmark) ou encore du nombre d'applications possibles.

Différents types de benchmarks

On peut distinguer trois niveaux de benchmarks, selon qu'ils portent sur les composants et la caractérisation de leurs propriétés physiques clés ; sur les systèmes constitués de ces composants et les opérations qu'ils sont capables d'effectuer ; ou sur les applications.

Par exemple, IBM a défini un benchmark appelé QuantumVolume, visant à caractériser les processeurs par la taille des plus grands circuits d'opérations qu'ils sont capables d'exécuter. Non seulement ce benchmark reste très technique mais il n'est adapté qu'aux processeurs à portes, comme ceux développés par IBM.

Le consortium américain QED-C (Quantum Economic Development Consortium) travaille sur un jeu de benchmarks fondés sur l'exécution d'algorithmes élémentaires, mais les métriques relatives à la fidélité d'exécution restent, là encore, très techniques.

Eviden-Atos a mis au point un benchmark, Q-score, qui concerne la résolution d'un problème d'optimisation combinatoire, le MaxCut. Son mérite est de porter sur une application concrète, ce qui offre plusieurs avantages. Les benchmarks applicatifs permettent une évaluation quels que soient les particularités technologiques et visent à évaluer la machine dans son ensemble ; ils permettent d'apprécier son utilité pour des applications très concrètes et industrielles ; ils contribuent à soutenir le développement des usages et des applications.

Le projet BACQ

C'est l'option que nous avons retenue pour notre projet BACQ, développé dans le cadre du programme MetriQs-France. L'ambition est de développer un instrument de mesure objectif et durable qui puisse servir de référence commune pour l'évaluation des performances pratiques des ordinateurs quantiques, à partir d'un jeu de benchmarks proches des applications et répondant aux questions que peuvent se poser les utilisateurs finaux : quel progrès représentent les ordinateurs quantiques par rapport aux ordinateurs classiques ? Quels sont leurs atouts pour des applications données ?

Le projet BACQ, d'une durée de trois ans, a démarré en septembre 2023, avec six partenaires : Thales (qui pilote le projet), Eviden, CEA, CNRS, Teratec et le LNE. Il mobilise 7 équivalents temps plein par an et son budget est de 3,9 millions d'euros, dont 2,5 millions d'euros de subventions. L'objectif est d'aboutir à un jeu de benchmarks validés et à un modèle de notation multicritère des ordinateurs quantiques qui seront maintenus au fil des ans et mis à disposition par le LNE.

Nous devons commencer par choisir des problèmes de référence à résoudre, puis déterminer les métriques techniques témoignant de la résolution de ces problèmes. Celles-ci seront ensuite agrégées dans une analyse multicritère permettant d'offrir des indicateurs opérationnels compréhensibles par les utilisateurs. L'objectif est de pouvoir évaluer n'importe quelle machine, analogique ou à portes, fonctionnant en régime bruité (NISQ) ou avec correction d'erreur (FTQC).

La démarche du projet est inspirée du benchmark Limpack utilisé dans le domaine du calcul haute performance, mais nous considérons ici plusieurs problèmes de référence à résoudre : des problèmes de simulation en physique, d'optimisation combinatoire, de résolution de systèmes linéaires, de factorisation en nombres premiers.

Nous avons établi une première liste des métriques adaptées, qui comprend, par exemple, la taille du problème traité, la probabilité de résoudre le problème, le temps de calcul, l'exactitude du résultat en fonction d'un certain niveau de bruit, la performance énergétique ou encore le Q-score.

L'agrégation des critères techniques s'inspirera du modèle Myriad, développé par Thales pour d'autres applications, car il a l'avantage d'être complètement transparent, avec des étapes parfaitement explicables et permet la prise en compte des préférences des utilisateurs.

La notation finale permettra de comparer différents ordinateurs quantiques, soit en considérant l'ensemble des problèmes résolus, soit en se focalisant sur un problème donné.

Le statut de tiers de confiance du LNE devrait permettre de favoriser l'adoption du jeu de benchmarks que nous développons dans le projet, véritable outil d'évaluation objectif et impartial.

Nous avons commencé à échanger avec des startups travaillant sur différentes technologies, qu'elles soient françaises, européennes ou internationales (États-Unis, Canada, Australie, Japon), ainsi qu'avec des équipes impliquées dans les travaux de R&D sur les benchmarks (Fraunhofer en Allemagne, TNO et TU-Delft aux Pays-Bas, QED-C aux États-Unis). Nous prévoyons également de mener des actions dans le domaine de la normalisation.

Au-delà, un véritable écosystème mondial se met en place pour accompagner le développement et l'industrialisation des technologies quantiques sur le plan de la métrologie, pour des référentiels de mesure partagés (EURAMET/EMN-Q), projet européen Qu-Test, etc.). Le programme MétriQs-France et le LNE entendent y jouer un rôle de premier plan.

Considérations métrologiques pour l'évaluation et la certification de l'IA

Agnès Delaborde

L'actualité de l'intelligence artificielle (IA) est très riche, avec, notamment, l'adoption de l'AI Act et d'autres règlementations européennes sur les données ou la responsabilité juridique autour des systèmes autonomes, mais également de règlementations sectorielles (robotique, médical, aviation…) ; les besoins de certification qu'entraînent ces nouvelles réglementations, en particulier la perspective d'un marquage CE pour les systèmes d'IA ; les questions d'acceptabilité, de confiance, d'éthique, d'impact sociétal, largement débattues dans les médias ; des évolution technologiques très rapides, telles que l'émergence de l'IA générative ; ou encore la généralisation du recours à l'IA, que ce soit dans la médecine, l'aéronautique, l'agriculture, la finance, ou encore les ressources humaines.

Dans ce contexte, des besoins pressants s'expriment pour des moyens fiables d'évaluer la performance de l'IA. Encore faut-il savoir ce que l'on entend par performance : s'agit-il de la performance fonctionnelle du système (est-il capable de distinguer un chien d'un chat ?), opérationnelle (le système respecte-t-il les contraintes de métier ?), sécuritaire (résiste-t-il aux attaques, aux pannes ?), économique (améliore-t-il réellement la productivité de l'entreprise, sa compétitivité ?), ou encore énergétique ? S'agit-il de la fiabilité du système, de sa réactivité, de son interopérabilité, de sa convivialité ?

Dans ses travaux, le LNE souhaite prendre en compte l'ensemble de ces aspects.

Applicabilité des principes métrologiques à l'IA

La question suivante consiste à savoir quels outils de mesure utiliser pour évaluer les performances de l'IA, ou, en d'autres termes, quelle est l'applicabilité des principes métrologiques à l'IA.

Il n'existe pas encore d'unité de mesure répertoriée par le Système international des unités de mesure pour évaluer les performances de l'IA. Les critères de mesure déjà utilisés sur les différents types de performance que j'ai évoqués ne font pas consensus. L'OCDE (Organisation de coopération et de développement économiques) a entrepris un louable travail de recensement des différentes métriques admises par la communauté, mais celles-ci ne fonctionnent pas dans tous les cas d'application et leurs résultats ne sont pas nécessairement faciles à interpréter.

Pour mesurer la performance d'un système, on peut recourir à une démonstration formelle, mathématique, dont l'inconvénient est de relever actuellement de travaux de recherche qui pourraient ne pas répondre entièrement aux besoins de la réglementation industrielle. On peut également recourir à des essais, qui ont l'avantage de produire un score de performance valide.

Le département dont je fais partie étant spécialisé dans les essais, c'est cette voie que nous avons choisie. Encore faut-il, pour cela, étalonner le banc d'essai, ce qui est plus facile pour des composantes physiques d'un banc mécanique, par exemple, que pour un modèle de simulation.

Un autre principe métrologique fondamental est la traçabilité. Toute la communauté n'est pas d'accord sur les méthodes de référence à utiliser, compte tenu de la diversité des applications et des contextes d'utilisation.

Toute démarche métrologique digne de ce nom doit également estimer l'incertitude de ses propres mesures, concept dont Nicolas Fischer vous parlera de façon plus approfondie. Dans certains cas, cette incertitude peut être quantifiée (par exemple, on sait que les conditions de luminosité influent fortement sur la performance d'une caméra intelligente) mais, la plupart du temps, on ne sait pas exactement quels facteurs affectent la performance et il est donc difficile de quantifier l'incertitude de mesure.

En matière de précision et de justesse, il est également compliqué de définir des valeurs de référence. On peut parfois démontrer que la réponse apportée à une question par ChatGPT est fausse mais il est délicat d'établir qu'une réponse est bonne.

Autre difficulté, la métrologie doit s'adapter à l'évolution technologique de l'objet qu'elle souhaite mesurer. Or les technologies d'IA évoluent très vite et la métrologie manque de recul à leur sujet.

Dans ce contexte, et compte tenu du besoin de l'industrie de certifier la conformité des produits, on peut être tenté de s'appuyer sur les bonnes pratiques existant en dehors du domaine de l'IA, dans la mesure où elles paraissent suffisamment généralisables. Dans le cadre de la règlementation sur la sécurité des machines, par exemple, il existe des méthodes pour certifier des systèmes à base de circuits de composants programmables. Nous avons tenté l'expérience d'appliquer ces normes à des composants de sécurité à base d'IA pour savoir s'ils étaient certifiables, et notre réponse est oui… et non.

En appliquant ces normes de façon « bête et méchante », on peut répondre que ces composants sont certifiables mais, si l'on tient compte des spécificités de l'IA, on constate que de nombreux points de contrôle manquent.

Les essais réalisés au LNE

Depuis 2020, le département des essais du LNE a mis en place un programme d'investissements pour acquérir des équipements permettant de réaliser le type d'essais que l'on peut raisonnablement envisager pour les systèmes d'IA.

Nous disposons déjà d'un banc d'essais par simulation dédié, à l'origine, au véhicule autonome, que nous utilisons aussi pour de la robotique médicale ou agricole.

Un autre banc d'essai baptisé Immersion va être inauguré, en 2024, sur notre site de Trappes. Il a la particularité de combiner des essais physiques, réalisés avec un robot, et de la simulation projetée autour de ce dernier, ce qui permet de tester les capacités du capteur tout en prenant en compte les contraintes physiques du robot.

Par ailleurs, le LNE réalise, depuis des décennies, des tests entièrement physiques sur des produits de l'industrie placés dans des situations précises et contrôlées, que ce soit en environnement réel ou en laboratoire, pour lesquels ces tests sont considérés comme valides du point de vue règlementaire. Nous estimons que certains aspects de l'IA embarquée pourraient être testés de cette façon.

Enfin, nous réalisons depuis quinze ans environ des essais sur base de données, en nous appuyant sur des méthodes de comparaison que nous maîtrisons parfaitement et qui permettent des analyses statistiques fines.

Pour garantir la qualité de ces moyens d'essais, nous menons des travaux avec différents organismes notifiés, autorités notifiantes et pouvoirs publics, et ce dans divers secteurs critiques tels que la vidéosurveillance, le médical, la sécurité machines, les véhicules autonomes, les radars routiers, etc. Nous avons, par exemple, travaillé avec le ministre de l'Intérieur sur la conformité à la loi des systèmes de vidéosurveillance qui seront utilisés dans le cadre des Jeux olympiques, ou sur les radars routiers qui vont être déployés sur les voies réservées.

Nous devons également vérifier l'adaptation de nos offres d'essais aux besoins du marché, notamment par rapport à l'AI Act et, enfin, renforcer la fiabilité et la qualité de nos essais en visant l'accréditation 17025 pour certains d'entre eux.

Incertitude, méthodes d'apprentissage et métrologie

Nicolas Fischer

Les deux piliers de la métrologie sont la traçabilité et l'incertitude de mesure, cette dernière quantifiant le doute ou la méconnaissance qui peuvent demeurer sur la valeur d'une grandeur après son mesurage. Ceci signifie, implicitement, que le résultat d'une mesure n'est pas une valeur unique, mais un ensemble de valeurs numériques qui ne sont pas toutes également probables – d'où le recours à des outils statistiques pour déterminer l'incertitude de mesure en métrologie.

Pourquoi et comment évaluer une incertitude de mesure?

L'incertitude de mesure est l'indicateur de qualité d'une mesure. Elle permet de comparer une valeur à une valeur de référence, un seuil réglementaire, une limite de conformité, mais également de comparer deux mesures, deux laboratoires, deux produits industriels entre eux.

Le Bureau international des poids et mesures publie et met à jour, depuis 1993, le Guide pour l'expression de l'incertitude de mesure, dont les règles sont reprises par des dizaines de milliers de normes à travers le monde. Ce guide décrit une démarche standardisée consistant à modéliser mathématiquement les processus de mesures physiques en quantifiant les sources d'incertitude des grandeurs d'entrée, qui peuvent être nombreuses (opérateurs, milieu environnemental, matériaux, chaîne de mesure des appareils…), puis à propager ces différentes incertitudes, à travers des méthodes de cumul quadratique ou de simulation Monte-Carlo, jusqu'aux grandeurs de sortie. Cette méthode est utilisée quotidiennement dans de très nombreux domaines, comme on va le voir à travers quelques exemples.

La mesure de l'acidification des océans

Les océans sont des puits de carbone et contribuent à la limitation du réchauffement climatique mais, en contrepartie, ils subissent un phénomène d'acidification. Les océanographes participant au GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) ont besoin de mesurer le pH des océans en différents points du globe de façon extrêmement fiable, afin de pouvoir étudier l'évolution de cette acidification. Des échantillons d'eau sont prélevés et analysés avec un spectrophotomètre, qui mesure leur absorbance. La température et la salinité des échantillons étant connues, il est alors possible de calculer le pH de l'eau de mer, avec une incertitude résultant de celle de chacun des paramètres (absorbance, température, salinité). Sur cette base, nous sommes en train d'élaborer un outil de calcul qui sera directement exploitable par les océanographes.

L'amélioration des performances des matériaux thermoélectriques

Les matériaux thermoélectriques, de plus en plus utilisés, présentent une grande conductivité électrique, ce qui est un atout, mais également thermique, ce qui est un inconvénient. Il est possible de réduire la seconde tout en préservant la première, en travaillant sur des matériaux ayant différentes propriétés physiques à l'échelle nanométrique. Cette démarche nécessite d'évaluer très précisément la conductivité thermique des matériaux testés, avec des mesures traçables et présentant un niveau d'incertitude connu.

Le département des matériaux du LNE vient de développer une méthodologie permettant de réaliser la première mesure traçable de conductivité thermique à l'échelle nanométrique, à l'aide d'un STHM (Scanning Thermal Microscopy) d'une résolution inférieure à 100 nanomètres. Les chercheurs ont développé un modèle mathématique à partir d'étalons connus sur la conductivité thermique, avec un certain niveau d'incertitude, puis ont appliqué ce modèle à des échantillons inconnus, en utilisant des techniques bayésiennes permettant d'obtenir une distribution associée à la conductivité thermique ainsi qu'une incertitude. C'est une première européenne, sinon mondiale.

Métrologie des infrasons

Les infrasons sont des ondes acoustiques de très basse fréquence, inférieure aux fréquences audibles par l'oreille humaine, qui peuvent avoir des origines naturelles (explosions volcaniques, tremblements de terre…) ou non naturelles (essais nucléaires, lancements de fusées…).

Dans le cadre du réseau mondial IMS (International Monitoring System) du CTBTO (Comprehensive Nuclear-Test-Ban Treaty Organization), le LNE contribue, en étroite collaboration avec le CEA DAM, à l'amélioration de la traçabilité des mesures de l'activité infrasonore à travers l'élaboration d'une méthodologie d'évaluation de l'incertitude associée à la mesure sur site.

Le principe de mesure est le suivant : une onde infrasonore atteint un détecteur composé de micro-baromètres et de systèmes de réduction du bruit dont l'organisation spatiale (en étoile, en pyramide…) permet l'estimation de la vitesse et de l'azimut de l'onde par la mesure de la variation de pression induite par la propagation de l'onde au niveau des micro-baromètres. Les données recueillies sont destinées à l'identification de la nature et de l'origine géographique du signal.

Dans un premier temps, nous avons développé une méthodologie de propagation de l'incertitude associée à l'étalonnage sur site. Nous avons reproduit des données issues de l'Observatoire de Haute-Provence, nous les avons décrites et modélisées, puis nous les avons bruitées volontairement et nous avons représenté la chaîne d'étalonnage avec les incertitudes associées. Dans un second temps, nous avons travaillé sur un cas d'étude réel (la station IS26, située en Bavière) pour étendre la méthodologie à la prise en compte de l'incertitude sur la totalité de la chaîne de traçabilité, en prenant en compte l'étalon primaire situé au LNE, l'étalon secondaire du CEA, l'étalon de référence sur site, les micro-baromètres qui servent de capteurs, ainsi que toutes les sources d'incertitude (les bruits environnementaux, le léger dénivelé entre les micro-baromètres, etc.). L'objectif était d'atteindre, dans les meilleures conditions de rapport signal sur bruit, une incertitude limitée à quelques degrés pour l'azimut et à quelques pourcents pour la vitesse.

L'application du concept d'incertitude à l'apprentissage supervisé

Les réseaux de neurones artificiels sont capables, une fois optimisés par une méthode d'apprentissage, d'extraire des informations de toutes sortes de jeux de données (données textuelles, images géospatiales, enregistrements audio, débits sanguins, etc.).

Le concepteur du réseau de neurones détermine le nombre de couches de neurones et de neurones par couche, les paramètres de l'algorithme de descente de gradient, le critère d'évaluation, la taille des batchs de données utilisées, etc.

Les différentes entrées sont pondérées et affectées de biais, puis sommées. On leur applique alors des fonctions d'activation non linéaires. Cette opération est reproduite n fois pour essayer d'identifier des sous-structures de signal intéressantes.

L'apprentissage supervisé se caractérise par l'utilisation de jeux de données étiquetées qui entraînent des algorithmes afin de leur apprendre à classer des données ou de prédire des résultats avec précision. Au fur et à mesure que les données en entrée sont introduites dans le modèle, celui-ci affine ses pondérations jusqu'à ce que le modèle soit correctement ajusté. Le modèle peut alors être appliqué à de nouvelles données pour établir des prévisions.

Au LNE, cette méthode d'apprentissage supervisé a été utilisée, par exemple, pour segmenter des images en microscopie. Compte tenu de soupçons sur la toxicité du dioxyde de titane employé sous forme nanoparticulaire dans des produits alimentaires ou cosmétiques, la règlementation européenne exige que des échantillons des produits soient analysés pour vérifier, sur plusieurs centaines de particules, si leur taille est en moyenne inférieure ou supérieure à 100 nanomètres. Au départ, la segmentation se faisait manuellement ou à l'aide de logiciels de traitement d'image plus ou moins performants. Cette opération étant très fastidieuse, en raison des chevauchements et agglomérations de particules, l'idée a germé de développer des outils d'apprentissage capables de reconnaître les particules et de réaliser la segmentation automatiquement en effectuant des mesures de diamètre.

Le LNE a également utilisé cette méthode pour automatiser la détection d'anomalies dans les électrocardiogrammes, avec un double objectif d'aide à la décision pour les cardiologues et de gain de temps d'analyse leur permettant de se consacrer aux cas les plus complexes.

Le LNE participe également à un projet européen qui vient de démarrer autour de la photopléthysmographie (PPG). Cette technologie permet de déterminer la tension artérielle à l'aide d'un petit appareil placé au bout du doigt qui transmet des ondes infrarouges à la peau puis détecte les signaux qui lui sont renvoyés. L'objectif du projet est de construire un réseau de neurones alimenté avec des données PPG et de lui apprendre à les comparer avec des données recueillies à l'aide d'un tensiomètre classique, afin d'améliorer la qualité des résultats obtenus par la PPG. Il sera alors possible de rendre l'utilisation de cette dernière accessible au grand public et de favoriser ainsi la détection précoce de pathologies.

Le pôle Énergie, environnement et combustion du LNE réalise souvent des calculs de CFD (Computational Fluid Dynamics) destinés à simuler des opérations de désenfumage dans de grands bâtiments. Ces codes comportent des milliers de paramètres et leur utilisation est très coûteuse en temps et en ressources, certains calculs pouvant durer jusqu'à un ou deux mois. Nous entraînons des réseaux de neurones à réaliser très rapidement une première estimation du processus de désenfumage, en se basant à la fois sur des simulations incendies précédentes et sur les lois de la physique.

Intégrer l'incertitude dans les méthodes d'apprentissage

Ce type d'exercice soulève de nombreuses difficultés pour les industriels, liées au nombre de données à prévoir pour l'entraînement des réseaux de neurones (des millions d'exemples, pour les modèles les plus complexes), aux ressources matérielles nécessaires (cartes graphiques, ressources en mémoire), à la robustesse du modèle face à des données corrompues ou trop bruitées, ou encore à l'incertitude des résultats, très difficile à évaluer dans la mesure où un modèle comprend des millions de paramètres.

Cette incertitude peut avoir de nombreuses origines : l'échantillonnage (dans l'exemple du dioxyde de titane, une base de données de 2 000 particules peut-elle être considérée comme suffisante?), l'outil et le processus de mesure, le choix de l'architecture et des hyperparamètres du modèle, l'entraînement du réseau, l'utilisation du modèle sur de nouvelles données qui peuvent être corrompues ou hors du domaine, etc. Notre travail consiste à réaliser une cartographie de ces sources d'incertitudes et à proposer une modélisation pour deux d'entre elles, qui nous paraissent les plus déterminantes : l'incertitude des entrées et celle de la répétabilité.

Face à une base de données comportant des images corrompues ou bruitées, l'approche classique consiste à modéliser l'incertitude prédictive comme une gaussienne dont on cherche à estimer les moments mais, en pratique, cette approche conduit à des résultats insatisfaisants. On pourrait alors être tenté de recourir à la méthode Monte-Carlo mais, s'agissant de réseaux profonds, celle-ci peut s'avérer trop coûteuse. Nous avons développé une approche alternative consistant à modéliser l'incertitude prédictive induite par l'incertitude des entrées comme une mixture de gaussiennes construite itérativement, ce qui permet de capter les non-linéarités d'un réseau de neurones.

L'incertitude de répétabilité est l'incertitude prédictive associée au réentraînement du réseau. La solution consistant à réentraîner le réseau n fois et à analyser la variabilité des résultats est envisageable pour un acteur comme Google, mais pas pour des industriels, sachant que chaque entraînement peut durer plusieurs jours et que, dans l'idéal, il faudrait procéder à mille ou dix mille entraînements pour bien évaluer l'incertitude.

Comment retrouver, à moindre coût, la variabilité estimée par la méthode Monte-Carlo? L'approche proposée consiste à utiliser la connectivité des modes de la distribution des sorties et à effectuer une recherche locale autour d'un mode obtenu par Monte-Carlo, ce qui permet d'estimer l'incertitude de répétabilité en se dispensant de réentraîner un grand nombre de fois le réseau.

Il nous reste à développer un cadre méthodologique prenant en compte toutes les sources d'incertitudes que j'ai évoquées pour fournir un indicateur de qualité des prévisions issues des réseaux de neurones. Ceci permettra, par exemple, lorsqu'on analysera l'IRM d'une tumeur, de savoir quelle est la probabilité associée à la prédiction selon laquelle cette tumeur est maligne ou bénigne.

Enfin, les solutions techniques proposées pour intégrer l'incertitude dans les méthodes d'apprentissage devront pouvoir être mises en œuvre dans un temps raisonnable, si l'on veut qu'elles soient industrialisables.

Débats

Une technologie aux résultats très incertains

Les bases sur lesquelles reposent les technologies quantiques sont très fragiles. Aucun ordinateur quantique ne fonctionne encore et personne n'est capable de dire quand ces recherches aboutiront. L'existences de minuscules startups travaillant en concurrence avec des grands groupes sur ces sujets prouve que personne n'a la solution. On ne peut donc attendre aucune retombée industrielle avant une quinzaine d'années. Quel peut être votre rôle pour introduire un peu de rationalité dans toute cette démarche ?

Félicien Schopfer : Le rôle de la métrologie est d'accompagner tous ces développements mais, comme vous le soulignez, la route est longue et incertaine… Notre objectif est de développer des outils permettant d'évaluer de façon plus précise la performance des machines quantiques, même si elles en sont encore à des niveaux de maturité très faibles, bien loin des espérances que l'on peut placer dans ces technologies.

L'étalonnage des mesures dans les technologies quantiques

J'ai une idée précise de ce qu'est un étalon dans le domaine de la physique. S'agissant de la mesure des performances de technologies quantiques, quelle est la nature de l'étalon ?

Félicien Schopfer : Pour les calculateurs quantiques, il s'agit, par exemple, d'algorithmes de référence, tels que ceux que nous cherchons à développer dans le cadre du programme BACQ. L'évaluation de capteurs quantiques, comme les gravimètres à atomes froids, met en jeu, de manière plus traditionnelle, des instruments de référence, bien caractérisés et validés par comparaison avec d'autres.

Les enjeux de souveraineté

Comment des projets généralement financés par l'Europe pourraient-ils servir des enjeux de souveraineté nationale ? Parmi vos partenaires, je note l'omniprésence de l'Allemagne. Les Instituts Fraunhofer ne risquent-ils pas de tirer la couverture à eux ?

Félicien Schopfer : Il existe effectivement des financements européens mais le programme MetriQs-France est financé par le plan France 2030.

Thomas Grenon : La recherche se déroule à l'intérieur de plusieurs cercles : un cercle international, un cercle européen, et un cercle français très confidentiel. Toute la difficulté est de trouver un équilibre entre concurrence féroce et nécessité de partager les résultats…

L'AI Act

Pour pouvoir mesurer ou certifier l'IA, il faudrait commencer par définir ce que c'est…

Agnès Delaborde : Bonne question !

Pour l'instant, l'IA se réduit à de l'apprentissage, donc à de l'imitation, sur la base d'un certain nombre d'échantillons. Votre approche ne serait-elle pas plus efficace si vous vous concentriez sur la mesure de la qualité de cette imitation ?

Agnès Delaborde : Dans un premier temps, c'est le choix que nous avions fait, mais nous devons travailler dans le cadre de l'AI Act, qui a sa propre définition de l'IA.

Une définition qui n'est basée que sur des rêves ! On est dans l'irrationnel.

Agnès Delaborde : Je suis d'accord avec vous, mais nous devons nous adapter.

J'imagine que l'AI Act va donner des prescriptions sur la fiabilité, la traçabilité, la transparence, etc. Toutes ces prescriptions donnent-elles lieu à des éléments mesurables ?

Agnès Delaborde : Non, mais il s'agit d'un texte règlementaire qui n'a pas vocation à tout expliquer. La Commission européenne a mandaté un organisme pour la définition de normes harmonisées qui, elles, rentreront dans le détail.

Quelle application aux modèles économétriques ?

Les exemples que vous avez pris concernant l'incertitude portent essentiellement sur des mesures physiques, chimiques ou biologiques. J'ai remarqué que lorsqu'on aborde la question de l'incertitude avec un économiste, il trouve cette notion étonnante. Vous arrive-t-il d'analyser l'incertitude à propos de mesures économétriques ?

Nicolas Fischer : Dans les mesures physiques, chimiques ou biologiques, le modèle est connu et les incertitudes le concernant sont très faibles. En électricité, par exemple, on peut s'appuyer la loi d'Ohm, selon laquelle « la tension aux bornes d'une résistance est proportionnelle à l'intensité du courant qui la traverse ». Dans le domaine financier, le modèle est souvent probabiliste, ce qui change tout…

Mots-clés : apprentissage supervisé, calculateurs, incertitude, intelligence artificielle, Laboratoire national de métrologie et d'essais, MetriQs-France, métrologie, technologies quantiques

Citation : Thomas Grenon, Félicien Schopfer, Agnès Delaborde & Nicolas Fischer. (2024). *Métrologie appliquée au numérique et à l'innovation*. Les séances thématiques de l'Académie des technologies. @

Retrouvez les autres parutions des séances thématiques de l'Académie des technologies sur notre site

Académie des technologies. Le Ponant, 19 rue Leblanc, 75015 Paris. 01 53 85 44 44. academie-technologies.fr
Production du comité des travaux. Directeur de la publication : Patrick Pélata. Rédacteur en chef de la série : Hélène Louvel. Auteur : Élisabeth Bourguinat. n° ISSN : 2826-6196.

Les propos retranscrits ici ne constituent pas une position de l'Académie des technologies et ils ne relèvent pas, à sa connaissance, de liens d'intérêts. Chaque intervenant a validé la transcription de sa contribution, les autres participants (questions posées) ne sont pas cités nominativement pour favoriser la liberté des échanges.