

Séance 43

Le champ de l'*Explainable AI* : des machines pour expliquer les machines?

23 novembre 2022

Présentation

L'avènement des réseaux de neurones en apprentissage machine a opéré un changement de paradigme dans de nombreux domaines, des sciences appliquées aux usages de la vie de tous les jours. Nous sommes désormais en interaction quotidienne avec des réseaux de neurones : sur nos smartphones, en naviguant sur internet ou bien par la présence d'objets connectés.

Néanmoins, leur fonctionnement interne demeure assez obscur, y compris pour les communautés qui les développent, et il est généralement admis qu'il n'existe à l'heure actuelle pas de formalisme mathématique satisfaisant pour décrire leur processus d'apprentissage. Pourtant, une compréhension sur le plan des mathématiques seules serait-elle suffisante? Étant donné l'immense impact de ces technologies en dehors des laboratoires, un nouvel impératif (fondé sur différentes dimensions sociales, légales, économiques) apparaît : il faut produire des explications des résultats des réseaux de neurones pour les usagers.

Dans cette présentation, je proposerai une analyse préliminaire des réponses apportées à cette question par la recherche en *Explainable AI*, tout en discutant les conditions dans lequel ce très jeune champ s'est constitué.

Intervenant :

Nicolas Berkouk est chercheur post-doctoral au sein du laboratoire pour la topologie et les neurosciences de l'EPFL (Lausanne). Diplômé de l'École polytechnique et de l'Imperial College en mathématiques pures, il a effectué une thèse à l'INRIA Saclay sur les fondements théoriques de l'analyse de données topologique. Il s'intéresse maintenant à étudier les réseaux de neurones comme des objets topologiques, et collabore en parallèle avec Pierre François et Laurence Barry à l'étude du champs récemment né de l'*Explainable AI*.

*

La chaire PARI (programme sur l'appréhension des risques et des incertitudes), portée par l'ENSAE et Sciences Po, a pour objectif d'identifier et comprendre (i) le champ de pertinence de nos outils d'appréhension des risques, et (ii) leurs conditions d'émergence et d'utilisation. Créée début 2015, elle organise un séminaire de recherche mensuel de 2h pour présenter et échanger autour de ses travaux et des thématiques connexes. Le deuxième cycle de la chaire porte sur les enjeux du big data pour l'assurance.