

Le machine learning, un vrai roman d'anticipation

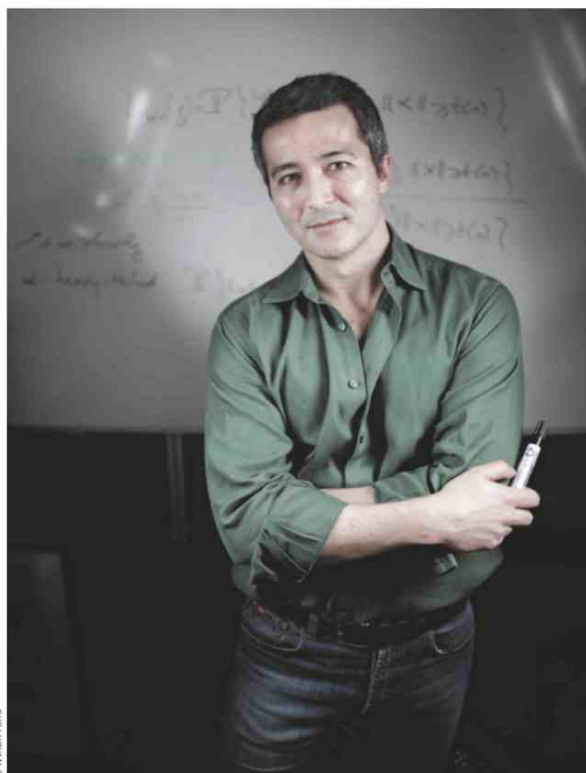
TECHNOLOGIE. Une fois par mois, un enseignant nous présente ses travaux en cours. Cette semaine, Stéphan Cléménçon, de Télécom-Paris Tech, et les enjeux liés au « machine learning », pour apprendre aux machines à décider.

CAPUCINE COUSIN @Capucine_Cousin

Comment rendre intelligible et lisible le big data, ces masses de données ? Comment trouver des méthodes automatiques permettant de classer des éléments et, même, de tenter de prédire un comportement à partir de différentes variables ? Le fameux algorithme de recommandation du service de vidéo à la demande Netflix repose précisément sur le traitement de datas complexes, car celles-ci expriment quelques préférences de la part d'un individu.

Le « machine learning », c'est « comment une machine peut-elle apprendre à décider toute seule ? » Comment compresser, représenter et prédire de l'information à partir de données choisies pour servir d'exemples ? L'enjeu est de parvenir à concevoir des algorithmes adaptés au traitement des données massives.

FAIBLE MARGE D'ERREUR. Selon Stéphane Cléménçon, professeur à Télécom-Paris Tech, de nouvelles méthodes s'imposent, face aux anciennes basées sur les statistiques traditionnelles, qui reposaient sur une modélisation a priori des données. Avec le big data, on dispose d'un grand nombre de « variables » pour chaque individu : des caractéristiques de localisations pour un internaute, de pages web visitées... Contrairement aux statistiques classiques, où les sondes disposent d'un échantillon représentatif, dans lequel plus il y a de personnes, plus la marge d'erreur est faible.



© Willem Paris

L'enjeu est de trouver des méthodes automatiques permettant de classer des éléments, d'identifier des traits récurrents ou constants... « Lorsque j'ai été recruté en 2007 sur le machine learning, expliquait-il, l'enjeu portait sur le traitement informatique de l'information, l'exploitation de nouvelles bases de données, dans différents champs d'action : médecine, véhicules connectés... Aujourd'hui, avec les smartphones et objets connectés, les capteurs sont partout, et les données sont prélevées automatiquement. Sans usage pré-

défini mais avec l'idée qu'elles recèlent une information précieuse à exploiter. L'algorithme va conduire à produire une règle. Et donc, par exemple, un arbre de décision, soit une combinaison de règles interprétables. D'où la classe d'algorithmes que j'ai introduite, qui a été brevetée par l'Institut Télécom-Paris Tech.

PRÉDIRE. Ces algorithmes permettent de prédire quels facteurs expliquent tel effet. On peut imaginer des usages dans des domaines divers, comme les risques de crédits bancaires, le scoring pour du ciblage comportemental (utile par exemple pour Criteo, spécialisée dans le retargeting publicitaire), un diagnostic médical... On va tenter de prédire un comportement à partir de différentes variables, ce qui équivaut à tracer la meilleure courbe passant par tous les paramètres. Par exemple, pour les réseaux sociaux, on a des graphes qui évoluent dans le temps, et une évolution de l'attachement préférentiel. Le machine learning doit permettre d'anticiper, prédire ces préférences des internautes. D'où l'intérêt de Facebook pour cette discipline... ♦

COMMENT s'orienter dans la data ? Le machine learning va informer le marketing du futur.

Le prof

Stéphan Cléménçon est professeur à Télécom-Paris Tech, Institut Mines-Télécom. Il effectue ses travaux de recherche en mathématiques appliquées au LTCI UMR Télécom ParisTech/CNRS N° 5141. Ses thématiques de recherche se situent principalement dans les domaines du machine learning, des probabilités et des statistiques. Il est responsable du master spécialisé « big data » à Télécom Paris Tech, du certificat d'études spécialisées « data scientist » et titulaire de la chaire industrielle « machine learning for big data ».