

Le Lab IA de Criteo, raisons, ambitions & résultats

Compte rendu rédigé par ANDSI

En bref...

Romain LERALLUT est ingénieur des Ponts et Chaussées et titulaire d'un doctorat de l'école des Mines de Paris. Il interviendra au sujet du Lab IA qu'il dirige depuis cinq ans au sein de la société Criteo, et notamment sur l'alignement d'un modèle d'entreprise sur les capacités offertes par le machine learning, le recrutement et l'organisation des équipes, les problèmes ouverts dans la communauté et les opportunités dans ce domaine.

L'Association Nationale des Directeurs des Systèmes d'Information organise des débats et en diffuse des comptes rendus, les idées restant de la seule responsabilité de leurs auteurs. Elle peut également diffuser les commentaires que suscitent ces documents.

Introduction

L'entreprise Criteo est spécialisée dans la publicité *intelligente* diffusée en ligne au moyen de bannières personnalisées par rapport aux centres d'intérêt des internautes. L'optimisation est réalisée en temps réel. La société est cotée au Nasdaq et suit un milliard d'internautes. Elle exerce dans le BtoB, essentiellement dans le domaine de l'*advertising technology*. Le Criteo AI Lab emploie 25 chercheurs et 95 ingénieurs qui travaillent conjointement.

Fondamentaux de l'IA

Les fantasmes nourris autour de l'IA sont injustifiés dans la mesure où elle est développée sous le contrôle de chercheurs et d'ingénieurs. Elle constitue la prochaine révolution industrielle. De 1956 à 2012 a eu lieu une phase de développement qui a occasionné peu de réussites. Puis, en 2012, le deep learning a enregistré ses premiers succès grâce aux modèles mathématiques de réseau de neurones. Leur puissance de calcul est devenue à même de fournir des performances similaires à celle d'un être humain, si bien qu'aujourd'hui l'IA est présente dans tous les domaines.

Il n'existe pas de définition précise de l'IA. Il s'agit d'une branche de l'informatique et des mathématiques qui était à l'origine un système à base de règles : le machine learning. Désormais, l'IA générative permet la création artificielle d'images ainsi que d'un interlocuteur conversation.

L'IA pose deux problèmes principaux qui sont la classification et la régression. L'interprétation d'une image est une question de classification, tandis que la prédiction d'une valeur continue, à l'exemple du prix d'un appartement, est du ressort de la régression. L'apprentissage effectué par l'IA peut être supervisé ou non supervisé. Ainsi, pour connaître le prix d'un appartement, sont entrées les données relatives aux caractéristiques du bien, à partir desquelles le modèle active une régression. Toutefois, afin d'éviter de devoir entrer les données à la main, on s'est efforcé de rendre la machine capable de trouver elle-même les informations pertinentes. Une fois que le modèle a été entraîné, il est possible de modifier certains paramètres.

Alignement entre l'IA et le business

L'utilisation de l'IA nécessite la collecte préalable de données. En pratique, il est préférable de créer un système simple à partir d'une quantité restreinte de données. Pour l'affichage de bannières publicitaires, trois questions se posent.

- Faut-il afficher une publicité, et à quel prix, quand une requête est formulée par un utilisateur ?
- Quels sont les meilleurs produits à afficher pour cette personne ?
- De quelle façon les images doivent-elles être affichées, en fonction de la probabilité d'achat anticipée ?

À l'origine, le moteur Criteo tabulait en fonction de cinq variables qui permettaient de prévoir le taux de clics. En 2012, un véritable modèle de machine learning a été développé. Ensuite, l'entreprise a opéré un passage à un système de *big data* et à l'optimisation des ventes à partir de 2014. Les achats de l'internaute sont anticipés au moyen d'un modèle proche de ceux utilisés pour le langage. Le *target business model* constitue un changement en ceci que les sites de vente en ligne définissent leur cible afin que Criteo réalise une optimisation en conséquence. Ainsi, le modèle simple du début s'est complexifié au fil du temps.

Int : Criteo porte-t-elle une partie du risque ?

RL : Oui, elle facture seulement les clics dans le but d'éviter la publicité non pertinente. L'entreprise a fait le pari que la maximisation de la satisfaction client lui profiterait *in fine*.

Recrutement et organisation d'une équipe IA

Le fonctionnement d'une équipe IA s'apparente à un sport d'équipe. Les tâches qu'elle accomplit sont la collecte et la vérification de données, l'analyse, le déploiement, ainsi que le monitoring, c'est-à-dire la surveillance de la mise en œuvre et la gestion des imprévus. La data occupe une place essentielle. L'utilisateur est observé à l'occasion de ses consultations de sites marchands et de ses interactions avec les bannières. En 2011, le modèle à 5 variables était simple, mais performant. À l'heure actuelle, il en comporte 250. Les données sont nécessaires pour entraîner un modèle, et les clients indispensables pour obtenir des données.

En premier lieu, un data ingénieur et un data analyste doivent être recrutés pour observer les données et leurs corrélations. Par exemple, il peut être constaté que les ventes sont plus fréquentes durant le week-end que pendant la semaine. Un premier modèle peut ainsi être établi. À partir de 2015, il est devenu possible de prévoir le montant des ventes, et donc de proposer un paiement à la commission. Toutefois, une machine n'est pas douée de raison et se cantonne à susciter des clics. Le troisième recrutement qui doit être effectué est celui d'un *product manager* qui œuvre à l'adéquation entre le produit et le client.

L'apprentissage passe par des plateformes A/B test qui visent à comparer l'ancien et le nouveau système avec deux populations en grandeur nature. Par ailleurs, des tests hors ligne permettent de réaliser une analyse contrefactuelle des performances des deux dernières semaines. L'IA est une science expérimentale et l'intuition est utile pour observer les données. Elle doit ensuite être confrontée aux tests, mesures et résultats d'expérience qui sont conservés en vue d'une réutilisation ultérieure. Des ingénieurs logiciels et des spécialistes à même d'élaborer des plateformes d'expérimentation sont requis.

Le monitoring est indispensable. En effet, un système doit obligatoirement être surveillé. Une panne occasionne des pertes financières. Un changement peut être engendré par des causes externes. Il est donc nécessaire d'effectuer une observation globale, mais aussi individuelle pour chaque client ou fournisseur. Une modification des hypothèses initiales entraîne une transformation généralisée.

Le modèle du deep learning est puissant et évolué. Cependant, un modèle profond ne fonctionne pas nécessairement mieux. En pratique, un retard de vingt-quatre heures provoque déjà une marge d'erreur de 1 %.

Ainsi, le modèle de Google qui était théoriquement meilleur était trop lent pour être pertinent. Les performances en laboratoire sont différentes de celles qui sont enregistrées en production. Certains domaines tels que l'image, le texte, et la traduction bénéficient de vingt ans de recherche, et une architecture de référence est disponible. Hors de ces situations, il est préférable de commencer par un modèle simple et explicable au moyen d'arbres de décisions et de modèles linéaires qui permettent de rectifier les erreurs. Le traitement de la data implique le recrutement de *data engineers* qui se chargent de valider la qualité des données et leur correspondance avec l'objectif business. En définitive, une infrastructure robuste est essentielle. La construction d'un modèle peut être envisagée une fois que toutes les étapes précédemment décrites ont été franchies.

« Chat GPT » et plus généralement les derniers développements de l'IA n'apportent pas de changement significatif. Tandis que cette technologie devient une commodité, les autres aspects tels que le packaging ou la qualité du produit demeurent déterminants.

Les nouveaux défis

Si l'IA représente une avancée majeure, elle souffre de certaines limites. Ainsi, l'absence d'empathie d'un *chatbot* le rend peu à même de remplacer un enseignant. Air Canada a déployé un *chatbot* pour son support client. Or la compagnie a vu sa responsabilité mise en cause en raison d'une réponse erronée. Pour autant, l'IA pourrait constituer une opportunité d'amélioration. On pourrait ainsi décider de lui imposer des valeurs que les êtres humains ne respectent pas toujours, par exemple, en lui interdisant d'opérer des choix discriminatoires. L'Europe est le seul continent à avoir établi une réglementation. Cette dernière reste floue dans un contexte de lutte stratégique au niveau mondial.

Débat

Int : Comment vous positionnez-vous par rapport à vos concurrents en ligne ?

RL : Nos performances ont longtemps été meilleures que celles de Google et Amazon. Il est possible de faire mieux qu'un GAFAM avec une bonne équipe et en y mettant les moyens.

Présentation des orateurs

Romain Lerallut, est diplômé des Pont & Chaussées (ingénieur 2002) et des Mines de Paris (PhD 2006). Après son doctorat, il a enseigné aux machines la lecture de l'écriture manuscrite chez A2iA, entreprise parisienne de haute technologie. En 2011 il a rejoint la « licorne » Criteo, pour travailler sur le déploiement des premiers algorithmes de Machine Learning, dans le but de rendre la publicité en ligne pertinente et efficace. Depuis 2019, il dirige le lab IA de Criteo, regroupant une centaine d'ingénieurs et de chercheurs, pour apporter les meilleures solutions IA aux problèmes du e-commerce.