



Michel Povlovitsch Seixas Machine Learning Data Scientist

> Son parcours

Michel Povlovitsch Seixas a fait ses études à l'École Polytechnique de l'Université de São Paulo, où il a obtenu une maîtrise d'ingénierie mécatronique, robotique et automation en 2008, suivie d'un doctorat en Recherche Opérationnelle en 2013. Son sujet de sa thèse : "Méthodes heuristiques et exactes appliquées à un problème de routage et de planification de véhicules".

Dans le même temps, il découvre les data science au sein de l'unité Business Analytics d'IBM, en tant que consultant Optimisation et Logistique.

De 2013 à 2015, il est Machine Learning Data Scientist chez Cnova. Il occupera les deux années suivantes le même poste chez Cdiscount, avant de rejoindre en juillet 2017 Continental Digital Services France. Il collabore notamment avec le département Powertrain Innovation France, appartenant aujourd'hui à Vitesco Technologies, entité du Groupe Continental.

Data Scientist : l'Indiana Jones de la donnée

« Utiliser les observations du monde autour de soi pour répondre à des questions et prendre de meilleures décisions ». C'est ainsi que Michel Povlovitsch Seixas expliquerait à un enfant son domaine, la Data Science. En rejoignant Continental à Toulouse il y a près de trois ans, il a pu fusionner trois passions : les mathématiques statistiques, le machine learning, et l'automobile. C'est pour cette industrie qu'un brevet porte sa signature depuis décembre 2019, dans un contexte pour le moins inattendu quand on pense données et intelligence artificielle, le pot catalytique.

Le métier de data scientist a trois composantes essentielles : mathématiques/statistique, informatique et business. On s'en doute, la première est nécessaire à toute modélisation mathématique. La seconde consiste à savoir coder (essentiellement dans les langages R et Python) pour la mise en œuvre algorithmique de ses stratégies et l'analyse de résultats. La dimension business est certainement celle qui surprend le plus, or elle est fondamentale. « Le métier est centré sur des problématiques business que m'exposent des ingénieurs en toutes sortes de dis-

ciplines, ce qui nécessite une grande flexibilité intellectuelle et une bonne compréhension des enjeux. »

Un pédagogue de l'algorithme

On imagine aisément qu'en mathématiques et statistique, le data scientist doit avoir un niveau flirtant avec celui des lauréats de la Médaille Fields. Il ne doit pas pour autant être enfermé dans un monde de chiffres et d'équations, bien au contraire. « Pour bien comprendre ce qui se passe et

répondre au mieux aux attentes du monde de l'automobile, ma formation d'ingénieur mécatronique m'aide beaucoup », explique Michel P. Seixas. Cela suppose une curiosité intellectuelle, doublée d'une forte capacité d'apprentissage, qui peut aller jusqu'à suivre des formations dans des disciplines éloignées des mathématiques. Mais, la prédisposition intellectuelle majeure est sans conteste sa capacité à être pédagogue, voire vulgarisateur. « Les ingénieurs automobiles ne maîtrisent pas forcément les approches des Data Science/machine learning.

Afin d'y adhérer, ils me posent toujours des questions et veulent comprendre autant les résultats que les méthodologies utilisées, poursuit Michel P. Seixas, il faut savoir vulgariser toute l'approche algorithmique. »

Ne pas confondre Data Scientist et Data Engineer !

Si le data scientist participe à tous les stades d'une Data chain (définition d'une problématique, collecte des données, nettoyage des bases, analyse, contrôle qualité, etc.) ses actions croisent souvent celles du data engineer. Néanmoins, il y a une distinction entre les deux postes, que Michel P. Seixas décrit ainsi : « Le data scientist est le chercheur qui va concevoir la solution, définir la modélisation mathématique, déterminer l'algorithmie et établir les stratégies de contrôle qualité, d'échantillonnage et de mises à jour des modèles. Le data engineer est celui qui maîtrise les techniques informatiques et les technologies autour de la collecte, du traitement, du stockage et de la distribution de la donnée en grande échelle. »

Par ailleurs, on est en droit de se demander ce qui différencie un projet data science de tout autre projet informatique « classique » ? La différence est juste énorme. « Un projet

logiciel lambda est déterministe, on connaît les points d'entrée et on sait ce qu'on attend en sortie. Dans un projet data science, il y a la composante statistique en plus, ce qui change tout dans l'appréhension et l'obtention des résultats. On se doit d'intégrer des intervalles de confiance, d'émettre des hypothèses sur la performance et la qualité du système conçu. » Il s'agit là d'une des plus épineuses particularités des Data Science à appréhender par les ingénieurs automobiles, habitués à obtenir des résultats répondant à des spécifications drastiques, voire cruciales comme en Safety (cf. les normes ISO26262 et SOTIF) . « Il faut se baser sur une sensibilité au risque acceptable – ou pas – et donner des indices de confiance », ajoute Michel P. Seixas. On comprend pourquoi la philosophie des Data science est difficile à intégrer pour certains spécialistes, notamment des systèmes embarqués.

« L'intelligence artificielle human-like n'existe pas encore »

Michel P. Seixas suit avec assiduité les publications de deux groupes de recherches : Google DeepMind et OpenAI. En effet, il s'intéresse de près à l'apprentissage par renforcement (reinforcement learning), qui imite les techniques cognitives des êtres vivants. « Aujourd'hui, un supercalculateur

approche les capacités de mémoire et de calcul d'un être humain, affirme-t-il, mais une intelligence artificielle équivalente à celle de l'homme n'existe pas encore et n'est pas prête d'exister. » Cependant, le data scientist s'enthousiasme devant les stratégies inédites et insoupçonnées qu'avait trouvées un agent autonome codé par Google DeepMind, renversant les plus grands spécialistes du jeu de GO il y a quelques années.

Quand on pense IA et automobile, le premier réflexe est de répondre voiture autonome, ce que concède Michel P. Seixas. Mais il mentionne d'autres secteurs où les data science ont de beaux jours à venir : manufacturing 4.0, supply chain, marketing prédictif, expérience client, services de mobilités, ingénierie et bien sûr, R&D.

Des algorithmes en guise de machette

Le titre de cet article prend maintenant tout son sens. « Être défricheur des utilisations des data science et du machine learning est l'une des parties les plus fun de mon métier », avoue notre data scientist. Et pour cause ! Il a validé en décembre 2019 le brevet intitulé « Apprentissage automatique prédictif pour la prédiction d'une fréquence de résonance d'un catalyseur de réduction sélective des oxydes d'azote ». Pour simplifier, le machine learning identifie les paramètres qui impactent les mesures d'un capteur dans un pot catalytique. Ainsi, Michel P. Seixas a conçu des modèles qui permettent de déduire les mesures d'un capteur virtuel à partir de celles d'autres capteurs, mais bien réels, eux. Grâce à cette « redondance », le capteur virtuel peut, en cas de panne, suppléer un ou plusieurs capteurs physiques, voire les remplacer, ou en indiquer l'état d'usure.

« Amener de l'innovation en explorant des problématiques existantes définies par les ingénieurs automobiles, et leur proposer de nouvelles solutions qui peuvent être hybrides (machine learning et experts), c'est ce qui me motive », conclut Michel Povlovitch Seixas. Qui manie l'algorithme et l'incertitude statistique comme Indiana Jones la machette et l'archéologie... ●

Ali Hammami



Ndlr : La communauté d'experts IA et Big Data de la SIA avec notamment son président Paul Schimmerling et Olivier Flebus, représentant de Continental, travaille de pied ferme sur ces nouvelles compétences que la mutation scientifique par l'intelligence artificielle engendre dans toute l'industrie automobile.