



La data science au service de la performance industrielle

Dans l'industrie, la bonne compréhension et l'utilisation des données sont de formidables leviers de performance et d'innovation. Que ce soit pour l'anticipation des pannes, la réduction des temps d'arrêt, l'amélioration de la qualité, la réduction des temps d'exécution, la modélisation de la connaissance empirique... Pour autant, les industriels sont en retard sur l'utilisation de la data science par manque d'expertise interne et de structuration de leurs données.



Mehdi Brahimi
Lead data scientist

Docteur en Philosophie, traitement du signal, automatique et informatique, Mehdi a une solide expérience dans le domaine de la maintenance prédictive, de l'intelligence artificielle et de l'apprentissage automatique.



La data science (ou science des données) est une discipline à la croisée de l'informatique, du métier client et de la modélisation mathématique. Elle vise à construire des modèles statistiques et d'apprentissage automatique (*machine learning*) en utilisant le big data, l'IoT et la puissance de calcul... L'objectif est ainsi d'explorer et d'analyser les données brutes (structurées ou non) afin de transformer ces données en informations pertinentes et en connaissance répondant à une problématique business.

A titre d'exemple, en maintenance prédictive, il faut déjà comprendre comment fonctionne l'équipement (ses modes de fonctionnement et ses principaux modes de défaillances) afin de choisir les capteurs les mieux appropriés. En prenant un parallèle avec la médecine, il s'agit de la phase d'auscultation qui a pour but d'isoler les symptômes d'une pathologie. Le data scientist modélise le comportement du système à l'aide des données de surveillance (quel est l'état de santé de mon système ?) et en extrait des connaissances utiles à l'aide de modèles mathématiques ou d'intelligence (quel est le composant dont l'état de fonctionnement n'est pas nominal ? Comment va-t-il évoluer dans le temps pour préconiser, avec une meilleure précision, la date à laquelle il faudra le remplacer avant qu'il ne tombe en panne et qu'il occasionne une indisponibilité de l'infrastructure...).

“ La data science est à la croisée de l'informatique, des métiers client et de la modélisation mathématique. Elle vise à construire des modèles statistiques et d'apprentissage automatique.”

Quatre typologies de données à structurer et analyser

Aujourd'hui, dans l'industrie, mais pas seulement, il existe quatre types de données. **Les séries temporelles** : celles issues des capteurs principalement, des données météorologiques, des données financières (performance sur un chantier ou un projet...); **les données en deux ou trois dimensions** : images, vidéos, maquettes numériques; **les documents « texte »** et **les données opérationnelles** : ce sont des données structurées (ou formatées/data value types), stockées dans une table de base de données (ERP), comme l'historique de la production ou l'historique de la qualité des pièces...

Ces quatre typologies de données répondent aux problématiques du secteur industriel, avec quatre solutions.

La maintenance prédictive

Sur les problématiques rencontrées avec les séries temporelles, on peut y répondre par des services de maintenance prédictive, le but étant de faire du juste-à-temps, c'est-à-dire remplacer des pièces ou réaliser des tâches de maintenance au moment opportun avant leur défaillance. *« Cela nécessite de mettre en place un système de surveillance à travers des solutions d'IoT permettant l'extraction d'indicateurs de santé de l'équipement pour la détection d'éventuelles anomalies. Ensuite, on peut mettre en place une procédure de diagnostic de l'équipement (d'où vient l'anomalie, sur quel composant, à quelle fréquence et pourquoi ?) afin d'identifier le mode défaillant et dans un second temps, prédire l'évolution de la dégradation du système ou, mieux, faire du pronostic. »*



A l'inverse de la prédiction, le pronostic est non déterministe : on fait une projection dans le temps avec une incertitude. « Par exemple, j'ai installé un capteur sur mon système (qui reproduit dans le temps son état de santé). A partir de là, je vais pouvoir dire que le roulement de ce système va atteindre un état critique dans les trois mois suivants, plus ou moins 2 jours... Cet exemple concret montre qu'il faut, pour en arriver là, bien maîtriser les trois domaines que sont l'informatique, la compréhension du métier client et la modélisation. »

Pour la mise en place d'une approche de maintenance prédictive, Assystem s'appuie sur la démarche « PHM » (*Prognostics and Health Management*), qui apporte les outils nécessaires pour les différents modules de maintenance prédictive (pour la capture de données, la détection d'anomalies, le diagnostic de défaillance, le pronostic de défaillance, la décision et les actions à mettre en place). « Par exemple, dans le cadre d'une installation nucléaire, nous avons eu à surveiller les entrées d'eau sur la centrale, au niveau du captage de l'eau dans l'océan. L'eau passe au travers de grilles qui peuvent s'obstruer par des algues ou divers déchets amenés par la mer. Un problème qui peut réduire fortement les flux d'entrée d'eau dans le circuit primaire de refroidissement de la centrale. Dans ce cas précis, nous avons développé la démarche PHM pour cibler les systèmes ou composants critiques qui vont générer le plus fréquemment des défaillances et temps d'arrêt... »

Cette méthode permet de faire de la détection d'anomalies qui, couplées à l'analyse de l'historique de la maintenance des grilles, a conduit

“ Dans les industries réglementées, on observe une volumétrie conséquente de documents et de textes. Le « Document to Data » est la 1^{ère} étape de structuration des données.”

à la mise en place d'une approche de diagnostic fiable. Nous avons également trouvé une corrélation entre les données météorologiques et les données d'exploitation. En fonction de la marée, on arrive aujourd'hui à préconiser des modes de fonctionnement sur le système de nettoyage des grilles et ainsi prolonger le cycle de vie du système, ce qui se traduit par un gain significatif en termes de temps passé et de coût.

Le text mining et le Traitement du Langage Naturel (TLN) : le « Google » des systèmes industriels

La partie « Document to Data » est celle où Assystem est le plus sollicité. « Nous y répondons avec des solutions de Text Mining pour retrouver l'information dans les documents ; et de NLP (*Natural Language Processing*

ou *traitement automatique du langage naturel*) pour comprendre le sens ou la thématique du texte. »

Dans l'industrie et plus particulièrement dans les industries réglementées (nucléaire, transport, industrie pharmaceutique), on observe une volumétrie conséquente de documents et de textes (rapports d'incidents, exigences, cahiers des charges de conception, livrables de projets...), nous créons ainsi des moteurs de recherche adaptés à un métier pour rendre les opérations plus efficaces. Les applications sont très larges. Nous avons, par exemple, travaillé sur des audits de gares ferroviaires à partir de différents documents (Word, Excel, mails, Power Point, PDF, plans, images...) fournis par la SNCF. Notre système a ensuite indexé automatiquement ces documents par catégories en fonction de la problématique métier rencontrée.



Dans le cas du démantèlement d'une base nucléaire, le client nous a fourni plus de 50 ans de documents de la vie de la centrale, soit des milliers de documents à scanner et à analyser par nos experts du démantèlement. *« Dans ce cas précis, nous avons développé une méthodologie originale de collecte, traitement et analyse de ce corpus de données hétérogène qui s'articule autour de trois solutions »*. D'abord, un moteur de recherche qui intègre l'ontologie du démantèlement nucléaire, ce qui permet d'indexer les documents en fonction du cycle d'un démantèlement. Ensuite, un module d'OCR (Reconnaissance Optique de Caractères) qui, grâce à l'utilisation des techniques d'apprentissage profond (*deep learning*), permet d'extraire automatiquement tous les tableaux des documents PDF. Enfin, un module de NER (Name Entity Recognition) pour reconnaître les noms relatifs à des systèmes et sous-systèmes dans ces tableaux... *« Nous avons ainsi reconstruit la description de l'installation et son architecture ; vérifié tous les incidents liés à ces systèmes et retracé l'historique de ces incidents. »*

Le moteur de recherche développé s'appuie sur un algorithme d'intelligence artificielle développé par Facebook (Dr. Q&A), qui répond à des questions ouvertes à partir d'un corpus documentaire (Wikipédia). *« Nous avons repris cette solution Open Source, que l'on a réadaptée au français sur Wikipédia, nous l'avons ensuite appliqué à l'industrie nucléaire autour du démantèlement, via un chatbot. Nos clients et nous-mêmes l'utilisons en interne pour gagner en efficacité lors d'un démantèlement d'une infrastructure. Grâce à cette solution, nous arrivons*

à une fiabilité des réponses de l'ordre de 80 % pour l'instant, ce qui nous permet d'améliorer notre efficacité dans l'exécution des projets pour le compte de nos clients. »

Le traitement des images et des vidéos

Concernant les images ou les vidéos dans le domaine de la vidéosurveillance, des approches de *Deep Learning* permettent la détection d'intrusion sur des sites sensibles (nucléaire, site du domaine de la défense, site industriel, gare ferroviaire, barrage...). Par exemple, sur certaines de ces installations, les systèmes actuels rencontrent beaucoup de fausses alarmes, dues à des erreurs de diagnostic (confusion entre un animal qui entre sur un site et une personne notamment). *« Avec les approches de Deep Learning, nous arrivons à baisser le taux de fausses alarmes et à apporter une aide à la décision aux services sécurité des acteurs chez qui nous intervenons. »*

L'analyse des données opérationnelles

Pour les données structurées (qualité, production, planning...), nous utilisons des solutions provenant du domaine de la recherche opérationnelle. Assystem modélise un processus pour l'optimiser. *« Nous travaillons sur ces sujets en partenariat avec la société Cosmo Tech pour analyser le déroulement et l'ordonnement des diverses tâches dans le temps, de façon à avoir une meilleure planification et optimiser, par exemple, en amont les arrêts de tranches d'une centrale, face à la ressource humaine locale disponible. Ceci permet d'anticiper le retard éventuel de certaines opérations et d'en diminuer les coûts. »*

Nous pouvons aussi développer la démarche RCA (*Root Cause Analysis* ou analyse de cause racine) pour trouver l'origine de problèmes de qualité ou de production, principalement dans l'industrie nucléaire, le transport,

“ L'apport de l'intelligence artificielle et la puissance de calcul ont bouleversé nos méthodes de travail ainsi que l'exécution des projets industriels.”



l'industrie manufacturière et de process... Par exemple, un fabricant de médicaments avait un différentiel de diamètre des pilules sur une même production. Nous avons analysé toutes les données disponibles de production, afin de trouver la causalité entre qualité et diamètre du comprimé en fonction des paramètres du système mis en place. *« Nous avons trouvé la cause et, en réponse, mis en place un modèle mathématique intégrant l'hygrométrie et la température pour prédire le diamètre et ainsi mieux paramétrer l'outil de production. »*

La structuration des données est en cours dans l'industrie

L'apport de l'intelligence artificielle et la puissance de calcul ont bouleversé nos méthodes de travail ainsi que l'exécution des projets industriels. Néanmoins, l'accès à des données structurées reste rare. La structuration des données est pourtant la première étape à mettre en place quand on se lance dans la data science. Cette étape de structuration passe également par la bonne compréhension des métiers du client et la mise en place d'algorithmes permettant de traiter cette masse de données textuelles. La connaissance des métiers industriels est ainsi primordiale pour aboutir à une solution adaptée à chacun des processus et spécificités du monde industriel.

« Aujourd'hui, nous voyons s'opérer la massification de la structuration des données... Pour autant, la culture digitale chez les industriels doit

encore être renforcée. L'erreur dans ce secteur est de croire que la masse de données dont on dispose est immédiatement exploitable... ou alors que l'on détient des données que l'on n'a pas... »

“ L'erreur dans ce secteur est de croire que la masse de données dont on dispose est immédiatement exploitable... ou alors que l'on détient des données que l'on n'a pas...”

A souligner que cette volumétrie de données est souvent sujette à la génération de fausses alarmes en forte augmentation. Aussi, pour pallier à ce phénomène, les industriels investissent dans le domaine du jumeau numérique, qui combine des modèles physiques, des connaissances d'experts et des données pour créer un modèle « in vivo » dynamique d'un équipement ou d'un système avec une fidélité plus importante en ce qui concerne ses propriétés. ■