Rapport sur la conférence AI Demystification Schlumberger

Laurent Butré, 23 ans de carrière (responsable IA chez Schlumberger, Clamart)

Filières : TSIA, éventuellement SD…

Note de l’auteur

Le conférencier (Laurent B) a comme formation un master en mathématiques appliquées (celui de l’ENSIMAG) et a vu beaucoup d’horizons depuis le début de sa carrière : Support (Paris), VIP IT Support (NY), Business System Portfolio Manager (Paris, aurait dû poser un brevet pour ce dropbox interne qui a particulièrement plu), GeoMarket Manager (Australie), Desktop Services & Mobility Manager (Houston), Innovation Manager (San Francisco) (c’est un remerciement : il y a développé du machine learning, du chatbot… est resté 2 ans dans la Silicon Valley !) et enfin…

A pris la direction (started from scratch) du laboratoire pour sa création et sa gestion à Paris (AI Lab)

Ca fait une grande carrière chez Schlumberger…

Résumé

Comment créer une IA, formater les données correctement, les challenges…

Mots clés : IA, Data Processing…

Rapport sur la conférence AI Demystification Schlumberger

# Slide 1 : Definition de l’IA

L’IA se découpe en plusieurs sous-parties : Machine Learning, Machine perception, Intelligent robots, Automated planning, Knowledge Representation & Reasoning, Natural Language Processing… On n’a pas toujours le luxe de la donnée ! On ne fait pas toujours du BruteForce… Il y a donc dans chacun de ces domaines principaux d’autres sous-domaines.

## Les étapes fondamentales dans la création de l’IA :

### Workflow, Digitalization, Automation, PUIS AI.

On fait évoluer également l’état des données en parallèle. D’abord on la collecte, on la nettoie, on l’explore/analyse, on la modélise (on Model la data), et enfin on l’interprète.

### On peut pour cela rentrer des données structurées (BDD SQL par exemple, ou des str découpées…), ou non structurées (pdf)

Egalement la sortie peut se faire de manière structurée ou non. Il y a deux phases lors du Machine learning : Training, Inference.

##### Training.

Raw Data + Class => Properties => Statistical Model Class = f(propriétés)

##### Inference.

Raw Data + Properties => Class? = f(Propriétés)

Machine Learning : c’est l’humain qui extrait les données et les exploite pour le redonner à la machine, au réseau de neurones… contrairement au Deep Learning qui laisse à la machine le devoir de s’adapter elle-même, de s’autocorriger pour nous rendre le meilleur output possible à partir d’un output. Le Feature Extraction + Classification se fait par la machine. Ce sont els deux façons d’extraire des propriétés.

On a également deux types d’erreurs : False Positive / False Negative, ou encore deux types d’Output : Classification / Regression… (cf png)

# Slide 2 : Data Preparation

**Sensitive Data** – Quelles données utilisées, dans quel pays, comment les récupérer ?

**Data Preprocessing –** Création du dataset avec plusieurs sources, nettoyer la donnée et la polir, puis la transformer avant de l’équilibrer. Respectivement : formater, ajuster, aligner, tout mettre en même résolution, remettre aux mêmes durées… puis combler les valeurs manquantes (max, min, moy…), les valeurs fausses et les extrêmes… ensuite un traitement type standardisation, normalisation et autres transformations. Step 4 : Balance Data, étape intéressante. Data augmentation / Oversampling / Undersampling. Exemples de biais : Celui sur la résolution des caméras pour la détection de valises abandonnées dans des gares (startup qui avait acheté pour 10K de valises. Le problème était que les 95% de réussite au départ étaient corrects pour des **caméras d’entraînement à résolution plus haute que dans la gare,** où la résolution est beaucoup plus faible... l’efficacité est tombée à 5%), ou celui de détection de « criminel » juste en étudiant la photo d’identité. Problème ? Le jeu de données utilisé était sur des prisonniers qui portaient tous la même chemise, ce qui a évidemment été détecté par le réseau de neurones… Voilà deux exemples de biais. Il faut faire très attention aux jeux de données que l’on utilise.

Comment se rendre compte qu’un réseau de neurones, qu’une IA est biaisée ? Il faut d’abord être conscient que cela peut arriver, c’est déjà bien. Exemple du bot automatique de réponse Twitter, qui est devenu raciste suite à des insultes… Mais c’est très difficile.

Data augmentation : l’art de grossir une base de données de la bonne manière. Copies légèrement modifiées de données existantes, données synthétiques, données traitées comme par exemple saturées, retournées pour des images, en changeant le temps et l’ordre des mots pour du texte, …

Les slides suivantes sont sur Schlumberger (pétrole, donc) en particulier. Comment se servent-ils de l’IA ? Augmenter, Accélérer, Automatiser. Exemple : trouver des rivières souterraines qui sont souvent des lieux de migration du pétrole. Bon exemple pour opposer le Data Scientist et le Domaine Expert -> Le DataScientist préfère avoir 90% de réussite et amputer les données de pointe. Le domain expert préfère largement l’autre cas : il préfère les cas extrêmes ! Par exemple, un jeu de données bien uniforme à 99% sur l’étude des sols peut par exemple révéler une seule ou deux images (le 1%) où l’algorithme échoue, parce qu’il y a une rivière souterraine qui fausse l’image… Ce que veut le D.E. !

Problématiques de l’énergie : stockage, transport, restitution. Ajd, le pétrole est la seule énergie à être facile à stocker, transporter, à restituer… Modèle économique de Schlumberger pérenne à 10 ans OK, à 50 ans, non… Mais l’entreprise applique toute sa technologie pour prendre le virage. L’hydrogène remplit les trois critères, mais gros problème de génération. Les renouvelables ? Problème de stockage et de transport… Askip, selon Laurent B., le problème n’est pas tant le pétrole que la demande énergétique des humains aujourd’hui. Assez bonne présence en France, avec 6 centres de technologie, plus 5-6 autres (trop vite). Prix d’un puits de pétrole : 50K/jour sur la terre, 1M/jour en mer.

Un exemple pratique : complétion d’images pour l’étude des appareils qu’on envoie sous terre (pour les formes, pour bien les coller aux parois). Un autre (juste mentionné) : image completion en temps réel pendant le forage pour extrapoler sur ce qu’on sait pour cartographier les fonds. Encore un exemple : Vision AI Classifier. Des cônes de diamant noir. C’est pour le forage. Pour savoir lesquelles remplacer, on utilise comme un dentiste, qui va leur faire passer un test (good, bad state… il faut reconnaître les usures de dents). L’IA permet d’identifier lesquelles n’ont pas besoin d’être observées, ou lesquelles ont besoin d’être remplacées etc…

**Digital Hardware – Lionel Beneteau**

**Partie 2**

Introduction : l’accès à l’énergie pas chère : le pétrole est la ressource phare. Surtout dans les pays émergents. Schlumberger accompagne dans la transition déjà vers le gaz, qui est déjà mieux en tout cas sur la génération de gaz à effet de serre. Autre remarque : Schlumberger se diversifie en dehors du pétrole !

Un ancien de Telecom : diplôme en 1987, doctorat à telecom en électronique, y est rentré à l’époque en électronique, a commencé dans les compteurs (à 200°C), puis a fait de la gestion de projet, à Houston, Oslo… Est revenu à Clamart pour de l’électronique haute température. Maintenant, s’occupe de la transformation digitale. Sur la Data. L’extraction de données aujourd’hui ne se fait pas de façon optimale. Réduire l’usure, le temps de traitement, réduire les déchets, augmenter la durée de vie, réduire l’émission de gaz à effet de serre.

**Le digital hardware dans un futur proche**

L’adaptabilité, la customizabilité des équipements en fonction du client et de la zone géographique : très important.

La modélisation, le digital permet de progresser sur la gestion de l’usure des équipements sur place par exemple, pour des raisons budgétaires et envirionnementales, et aussi de temps, de ne pas passer trop de temps à remplacer ou à vérifier qu’il faut remplacer etc…

L’IA s’applique bien là où ce qu’on lui demande est facile est répétable. Pas compliqué et qu’on ne sait pas faire ! High Level Automation = algorithmes d’automatisation pour que les équipements prennent en compte leur environnement et les perturbations, les anomalies, et s’adaptent. Comment faire du digital hardware enhancement via Modeling ? On fait des modèles, statistiques, des avatars, pour consommer intelligemment les données qui remontent, interconnecter les modèles et ensuite : voir s’il y a des risques de pollution, de panne, d’usure, de performance… La détection du risque c’est extrêmement important. Le taux de fiabilité du modèle doit être >95% au vu des risques. On veut créer des modèles hybrides : Experts + Data Driven. Excellent niveau de fiabilité dans un temps raisonnable qui permet de prendre une décision business. Travail conjoint entre Data Engineers et Experts Domaines est d’une importance capitale.

L’erreur est une catastrophe pour Schlumberger. Leurs travaux ont beaucoup d’importance, et rien n’est toléré. Etre fiable est CAPITAL. Encore une fois deux mondes s’opposent : le domaine mathématique des Data Scientist Vs domaine physico-géologique des domain experts. On va beaucoup plus loin si on arrive à faire le lien entre ces mondes. (Exemple : Ah bah on aurait pas dû vous couper le pied, mon réseau de neurones était finalement biaisé, vous aviez pas de cancer… sorry my bad !)

Attention : on n’a quasiment jamais réutilisé un modèle fait pour un projet sur un autre projet. Le model transfert est extrêmement compliqué, sinon impossible. Il faudra travailler le modèle, réapprendre pour attendre la finalité précise de ce qu’on veut résoudre. D’où la nécessité de prendre des Data Scientists. Il y a une vraie value à capter les données, analyser les données, et enfin (domaine prédictif) sortir un output de la table de données.

The 4-Hour WorkWeek.

Comment postuler à des stages de fin d’année ? Sur le site spp.com ? VOIR AVEC MARCO LES STAGES COMMUNIQUES… EN IA, CLOUD COMPUTING… Go voir sur le site de l’entreprise.

Ma question sur la double-compétence : ctftimes.org ? Il faut se faire expliquer le domaine par le domain expert. Il faut rester ouvert d’esprit. (Lionel exemple contact privilégié avec l’entreprise qui est très vivifiante, et explique qu’il a eu l’impression de faire plein de métiers différents).

MOOCS ou sites webs : cours de stanford en ligne sur l’IA, sur funmoocs aussi, tagueule ??? exemples de codes, de résolution… cours de Andrew Engie pour stanford ?

Références

Nom, p. d. (Année). Titre de l’article. *Titre du journal*, Pages de - à.

Nom, p. d. (Année). *Titre du livre.* Nom de la ville: Nom de l’éditeur.

Notes de bas de page

1[Ajoutez des notes de bas de page, s’il y en a, dans leur page figurant après les références. Pour les besoins de mise en forme APA, vous pouvez facilement entrer vos propres notes et appels de notes de bas de page. Pour mettre en forme un appel de note de bas de page, sélectionnez le numéro, puis, sous l’onglet Accueil, dans la galerie Styles, cliquez sur Appel de note de bas de page. Le corps d’une note de bas de page, tel que cet exemple, utilise le style Normal. (Remarque : si vous supprimez cet exemple de note de bas de page, n’oubliez pas de supprimer également sa référence dans le texte. Elle se trouve à la fin de l’exemple de paragraphe Titre 2 dans la première page du contenu de ce modèle.)]

Tableaux

Tableau 1

[Titre du tableau]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| En-tête de colonne | En-tête de colonne | En-tête de colonne | En-tête de colonne | En-tête de colonne |
| En-tête de ligne | 123 | 123 | 123 | 123 |
| En-tête de ligne | 456 | 456 | 456 | 456 |
| En-tête de ligne | 789 | 789 | 789 | 789 |
| En-tête de ligne | 123 | 123 | 123 | 123 |
| En-tête de ligne | 456 | 456 | 456 | 456 |
| En-tête de ligne | 789 | 789 | 789 | 789 |

Remarque : [Placez tous les tableaux de votre document dans une section dédiée aux tableaux figurant après les références (et, le cas échéant, les notes de bas de page). Commencez une nouvelle page pour chaque tableau, incluez un numéro et un titre de tableau pour chacun d’eux, comme illustré dans cette page. Le texte explicatif figure dans une note qui suit le tableau, comme celle-ci. Utilisez le style Tableau/Figure, disponible sous l’onglet Accueil dans la galerie Styles pour définir l’espacement entre le tableau et la note. Les tableaux au format APA peuvent utiliser un interligne simple ou 1,5. Incluez un titre pour chaque ligne et colonne, même si le contenu peut sembler évident. Un style de tableau par défaut a été configuré pour ce modèle correspondant à des instructions de l’APA. Pour insérer un tableau, sous l’onglet Insertion, cliquez sur Tableau.]

Titre des figures :

Figure 1. [Incluez tous les chiffres dans leur propre section après les références (et les notes de bas de page et tableaux, le cas échéant). Incluez une légende numérotée pour chaque figure. Utilisez le style Tableau/Figure pour ménager un interligne entre la figure et la légende.]

Pour plus d’informations sur tous les éléments de la mise en forme APA, voir le guide APA Style Manual, 6th Edition.