



Thibault MAHO

- 25 ans
- thibault.maho@gmail.com
- 07.70.13.55.36
- 4 rue Carnot 59370
Mons-en-Baroeul
- Titulaire Premiers Secours
- Titulaire Permis B

COMPÉTENCES

Programmation



Autres Langages: PostgreSQL, Java, C
Environnement: Git, Linux, Windows

Techniques

Machine Learning, Deep Learning, Natural
Language Processing, Traitement du Signal
et d'Image, Analyse de Données

Qualités

Autonome, Curieux, Passionné, Esprit d'
équipe

LANGUES

Français: Langue Maternelle
Anglais: niveau B2 (Bulats)

LOISIRS

Domotique, Conception d'un drone

Sports: Crossfit, Tennis (30/2), Running,
football (12 ans)

Musique: Guitare (15 ans), Piano (1 an)

Doctorant en Machine Learning

FORMATION

2014 - 2017 | Phelma, Grenoble (38000)

Spécialisé en Traitement du Signal et Image, Programmation et
Électronique. Première année généraliste en physique.

2012 - 2014 | Lycée Faidherbe, Lille (59000)

Classe préparatoire en Maths/Physiques (MP)

Formation en autodidacte:

- *cours de Machine Learning* (Coursera) de Stanford,
- *cours de Deep Learning* (Youtube) de CMU,
- *cours de SQL* (Udemy)

EXPERIENCES PROFESSIONNELLES

Depuis 2017 | Ingénieur R&D en IA | xBrain, Lille (59000)

- Développement en **Python** de modèles de Machine Learning et
de Deep Learning avec utilisation de frameworks: **PyTorch**,
Keras, Scikit-learn, Gensim, Numpy, ..
- Mise à disposition des projets par **API REST**.
- **Text Mining**, mené en binôme:
 - **Objectif:** créer un moteur de recherche à partir de
sources de données textuelles variées (structuré et
non structuré).
 - **Données** issues du code français.
 - Création d'un embedding juridique. Vecteur du texte
obtenu par la moyenne des vecteurs de mots,
pondérés par leurs iDFs .
 - Création d'une base de tests issue de forums et de
sites officiels.
 - Utilisation d'un premier modèle **k-Nearest Neighbors**
sous GPU. Modèle choisi pour sa simplicité et rapidité.
Très bons résultats dus à une bonne représentation.
 - **Siamese Network** pour améliorer l'embedding.
Donnée d'apprentissage par annotation faible.
- **Clustering et Classification d'images (réalisé en autonomie)**
 - **Objectif:** identifier les différentes classes de pièces
jointes reçues et entraîner un modèle supervisé.
 - Données issues de mails d'assurance.
 - **Clustering:** Normalisation des images. Test de
plusieurs modèles et sélection d'un modèle de
K-means. Identification de classes photos (voiture,
maison, logos, ...) et texte (constat, carte grise,
formulaire ...). Validation des classes et classification
d'images tests manuellement.
 - **Classification:** récupération du modèle **VGG16**
pré-entraîné sur ImageNet. **Finetuning** des dernières
couches sur nos données.

- **Clusterisation et classification de mails** (réalisé en autonomie)
 - **Objectif:** Identifier les classes existantes. Créer un classifieur.
 - **Donnée:** 500.000 mails du domaine bancaire et assurance
 - Récupération d'un embedding **FastText**, pré-entraîné sur Wikipédia. **Finetuning** sur les mails et de la donnée récupérée sur des sites bancaires.
 - Evaluation du nombre de clusters par le **Critère d'Information Bayésien (BIC)** avec un TFIDF pour vectorisation et un mélange de gaussienne pour modèle.
 - **Non supervisé:**
 - Evaluation de 2 modèles: **LDA** et **Embedding + GMM**. Critères basés sur le score de cohérence, la distribution des classes, et l'histogramme des scores d'élections.
 - **Construction d'un arbre de clusters** pour créer un enchaînement de modèles experts supervisés. Sélection des clusters à conserver par leurs scores de cohérence et leurs vocabulaires.
 - Active learning avec un modèle FastText pour nettoyer les données du cluster.
 - **Supervisé:**
 - Création d'une base de tests et de validation.
 - Plusieurs modèles sélectionnés en fonction de la quantité de donnée d'apprentissage du modèle expert: réseau de neurones récurrent (LSTM) bidirectionnel avec attention, classifieur FastText initialisé avec les poids du modèle entraîné sur toute la donnée, recherche de mots clés.

2016 (3 mois) | Assistant Ingénieur | Laboratoire Cristal, Lille (59000)

- Traitement des signaux d'une interface cerveau-machine à destination des personnes handicapées.
- Fusion des informations de 3 sources (Joystick, EMG, ECG) par la **théorie des fonctions de croyance**.

2016 - 17 | Cours Particuliers, niveau Lycée | Banque Postale, L'envol

Cours particuliers en Mathématiques et en Science de l'Ingénieur

PROJETS

2016 (4 mois) | Réalisation d'une voiture autonome (Fonctionnelle)

- Voiture munie d'une caméra, d'un radar à ultrason et d'une carte National Instruments pour le traitement d'image.
- **Suivi automatique** d'une et 2 ligne(s) par traitement d'images et **transformée de Hough**.
- Détection automatique de panneau par méthode statistique.

2014 (4 mois) | Programmation du jeu stratégique Corridor en C/C++ (Fonctionnel)

- Développement d'un **algorithme de recherche A*** pour créer un adversaire Ordinateur.

RÉFÉRENCES

Gregory Renard - gregory.renard@xbrain.ai

CoFondateur de xBrain

Intervenant au MIT, à Stanford, à Berkeley et à la Nasa

Conseiller pour *AI for Humanity*

Dr Gregory Senay - gregory.senay@xbrain.ai

Directeur scientifique de xBrain