

IA : Facebook partage fastText, ses travaux sur le langage naturel

Le laboratoire de recherche en intelligence artificielle de Facebook (le FAIR) vient de verser fastText en Open Source. FastText est une bibliothèque d'applications conçue pour construire des solutions évolutives de représentation et de classification de textes. Autrement dit, des outils d'intelligence artificielle pour interpréter le langage humain. Les chercheurs du réseau social ne se contentent pas d'ouvrir librement leurs sources logicielles au plus grand nombre ([sur GitHub](#)), ils se proposent également de partager leurs travaux de recherche ([ici](#) et [là](#)).

L'apprentissage de la sémantique est l'un des plus grands défis auquel est aujourd'hui confrontée l'intelligence artificielle. Il s'agit de faire comprendre et interpréter le sens des expressions, écrites ou orales, d'un humain par une machine. « *Mais c'est un besoin essentiel, estiment les chercheurs Armand Joulin, Edouard Grave, Piotr Bojanowski et Tomas Mikolov dans une [contribution de blog](#). Le traitement de texte automatique constitue un élément clé de l'interaction au jour le jour avec votre ordinateur; c'est un élément essentiel que ce soit pour la recherche sur le Web, le classement des contenus ou le filtrage du spam, et quand cela fonctionne bien, c'est complètement invisible pour vous. Avec la quantité croissante de données en ligne, il y a un besoin d'outils plus souples pour mieux comprendre le contenu de très grands ensembles de données, afin de fournir des résultats de classification plus précis.* »

FastText et l'approche par classement

D'où l'idée de partager les recherches internes avec la communauté scientifique dans l'espoir d'accélérer les développements dans le domaine du traitement informatique du langage humain. Une démarche similaire à celle de l'Open Compute Project (OCP) du même Facebook pour améliorer le design hardware des éléments du datacenter. Les travaux de fastText s'attachent notamment à traiter des représentations de phrases composées de « sacs de mots » et d'informations issues de mots secondaires, et à les partager. L'objectif immédiat visant à classer efficacement les textes et à « apprendre » les mots par des représentations vectorielles.

L'approche de la solution de Menlo Park permettrait ainsi de palier les limites que les réseaux neuronaux informatiques rencontrent avec les grands ensembles de données. En utilisant des classificateurs hiérarchiques au lieu d'une structure plane dans laquelle les différentes catégories sont organisées en arbre. « *Cela réduit le temps [de traitement] dû à la complexité de formation et de tests des classificateurs de texte [qui passent d'un mode] linéaire à [un mode] logarithmique en respectant le nombre de classes* », souligne les contributeurs du billet.

Quelques secondes contre plusieurs heures

Résultat, là où les solutions comme Char-CNN ou VDCNN mettent plusieurs heures à interpréter du contenu Yahoo, fastText n'aurait besoin que de 5 secondes. Avec un taux de pertinence d'interprétation similaire, voire supérieur, aux offres concurrentes, selon Facebook. « *FastText peut également classer un demi-million de phrases parmi plus de 300 000 catégories en moins de cinq minutes* »,

assurent les chercheurs. Et cela dans différentes langues occidentales, dont l'anglais, l'allemand et le français. L'objectif final étant de construire des solutions qui sauront interpréter le langage naturel aussi bien que les humains. Un doux rêve qui relève aujourd'hui encore de la science-fiction.

Lire également

[Yann LeCun, Facebook : l'intelligence artificielle est son amie](#)

[Aera 404 de Facebook : un lab hardware en mode collaboratif](#)

[Machine Learning : langage naturel et traduction sur le Cloud de Google](#)

Photo credit: [eston](#) via [VisualHunt](#) / [CC BY-NC-SA](#)