

LinkedIn rend Open Source FeatureFu, boîte à outils du Machine Learning

LinkedIn rend Open Source FeatureFu. Écrite en Java, cette boîte à outils est destinée à l'extraction de caractéristiques à partir de données utilisées dans l'apprentissage automatique, dont les moteurs de décision et la modélisation statistique (classification, clustering et normalisation). FeatureFu peut servir différents objectifs. Dans un [billet de blog](#) du 4 septembre, Bing Zhao, ingénieur logiciel chez LinkedIn, donne l'exemple du réseau social professionnel qui veut isoler et classer les préférences des membres pour différents flux apparaissant en page d'accueil. Et ce en comptant le nombre de 'j'aime' (likes) dans l'historique et de commentaires laissés par le membre dans chaque type de flux.

Java, Expr et FeatureFu

Dans l'exemple ci-dessus, les chiffres bruts doivent être combinés à un ratio 'j'aime-par-commentaire', avant de pouvoir être utilisés comme un élément stable, avec une formule mathématique du type : $(1 + \text{j'aime}) / (10 + \text{commentaires})$, commente l'ingénieur. Normalement, la formule doit être codée en ligne et tout changement apporté à cette formule implique une modification du code... Le processus est assez chronophage. « Avec Expr [une bibliothèque légère écrite en Java] et FeatureFu, nous aurons simplement besoin d'écrire la formule comme une expression symbolique '(/ (+ 1 j'aime) (+ 10 commentaires))' et de l'inclure dans le fichier de configuration du modèle », explique-t-il. Avant d'ajouter : « toute modification future de la formule ne nécessitera qu'un changement de configuration de l'expression symbolique elle-même : '(- (log2 (+ 10 commentaires)) (log2 (+ 1 j'aime)))' ». Cette option est bien plus « flexible et agile » que la précédente, selon lui.

LinkedIn et l'écosystème

Le code source de FeatureFu est disponible sous licence Apache 2.0 sur [Github](#). FeatureFu est le dernier d'une longue série d'outils rendus Open Source par LinkedIn, parmi lesquels : Azkaban, Kafka, Samza, Voldemort et, plus récemment, le [framework Pinot d'analyse de données](#) en temps réel.

Lire aussi :

[LinkedIn, des investissements qui inquiètent les analystes](#)

[Machine Learning contre statistiques « classiques » : qui remportera le match ? \(tribune\)](#)

crédit photo © isak55 - Shutterstock