

Machine learning : AWS et Facebook lancent TorchServe

AWS, en partenariat avec Facebook, a annoncé la disponibilité du modèle serveur [TorchServe](#) dans le cadre du projet open source de bibliothèque de machine learning PyTorch.

« Avec TorchServe, les [utilisateurs de PyTorch](#) peuvent désormais déployer leurs modèles en production plus rapidement, sans avoir à écrire de code personnalisé », a expliqué dans un [billet de blog](#) Julien Simon, AI Evangelist chez Amazon Web Services (AWS).

« En plus de fournir une API de prédiction à faible latence, TorchServe intègre des gestionnaires par défaut pour les applications les plus courantes telles que la détection d'objets et la classification de texte », a-t-il ajouté.

TorchServe prend en charge tout environnement d'apprentissage automatique, dont Amazon SageMaker, les services de conteneurs et Amazon Elastic Compute Cloud (EC2).

TorchElastic sur Kubernetes

Parallèlement, les équipes d'AWS et de Facebook [ont annoncé](#) une nouvelle contribution open source : TECK ([TorchElastic Controller for Kubernetes](#)). Cette bibliothèque est une implémentation native de l'interface PyTorch Elastic qui gère automatiquement le cycle de vie des modules et services Kubernetes requis pour l'apprentissage.

Avec TorchElastic Controller for Kubernetes, [a indiqué](#) AWS, les développeurs peuvent « réduire le temps et les coûts d'apprentissage distribués » en limitant les ressources sous-utilisées de cluster et l'apprentissage sur les instances Spot Amazon EC2.