

Embedding Projector : Google place sa dataviz en Open Source

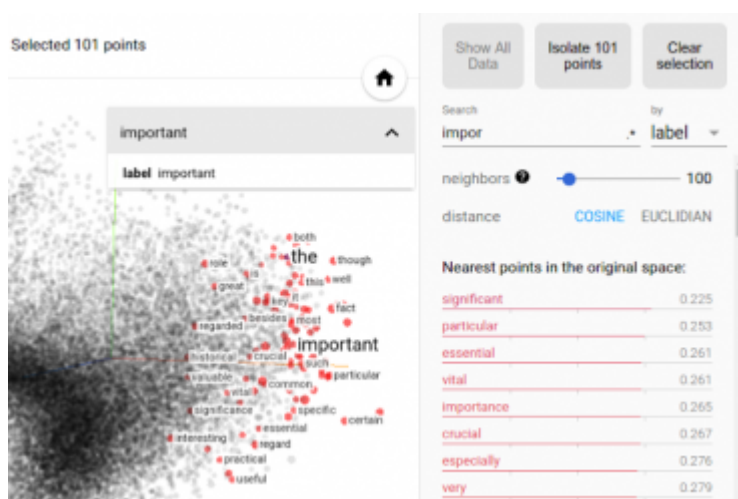
Encore un élément de plus mis en Open Source dans la corbeille des services autour de l'intelligence artificielle (IA) et plus particulièrement du machine learning (ML). Google a annoncé [dans un blog](#), la mise à disposition d'un outil de visualisation des données, baptisé Embedding Projector. Il s'agit d'une composante du programme plus général de machine learning, nommé [Tensor Flow](#). Ce dernier est utilisé dans plusieurs domaines comme la reconnaissance d'image, la traduction naturelle ou le diagnostic médical.

Or avec l'adoption rapide de l'apprentissage automatique, les scientifiques sont confrontés à un problème d'intégrer des centaines voire des milliers de dimensions au sein des données. Elles nécessitent des outils spécifiques. Et c'est là qu'intervient Embedding Projector, une application web de visualisation de données multidimensionnelles. A noter qu'elle peut tourner indépendamment de Tensor Flow.

Voir comment le machine learning apprend

Les données nécessaires à l'apprentissage automatique existent dans un format non reconnu par un ordinateur. Il faut donc les convertir en algorithme. Pour cela, Google utilise un vecteur Embedding (structure mathématique d'une instance contenue dans une autre instance, comme un groupe et un sous-groupe), pour intégrer plusieurs dimensions d'une donnée. Embedding Projector permet donc de réaliser une visualisation de cette représentation multidimensionnelle.

Google donne l'exemple du langage naturel en réalisant une cartographie des mots qui se rapprochent d'une requête (dans le cas cité, il s'agit du mot important). Les spécialistes de l'IA peuvent voir en 2D, 3D, zoomer ou tourner des nuages de points et analyser les corrélations réalisées par le système de machine learning. Embedding Projector s'aide de solutions de réduction de dimensionnalité comme [PCA](#) ou [t-SNE](#) pour faciliter la visualisation.



Cette initiative de Google s'inscrit dans la démarche plus globale des grands acteurs de l'IT dans le

domaine de l'intelligence artificielle. La majorité d'entre eux a décidé de placer en Open Source leurs plateformes de machine learning. Même Apple va sauter le pas en autorisant ses chercheurs à publier leurs travaux. Il faudra attendre encore un peu pour que Cupertino cède aux sirènes de l'Open Source.

A lire aussi :

[Google veut rendre le machine learning moins discriminant](#)

[Comment le Machine Learning aide à débusquer les failles de sécurité](#)