



hackAtech

Shake science. Shape innovation.

#metric

#machinelearning

#metriclearning

#similarity

METRIC LEARNING

Une mesure de distance adaptée à la tâche

Inria

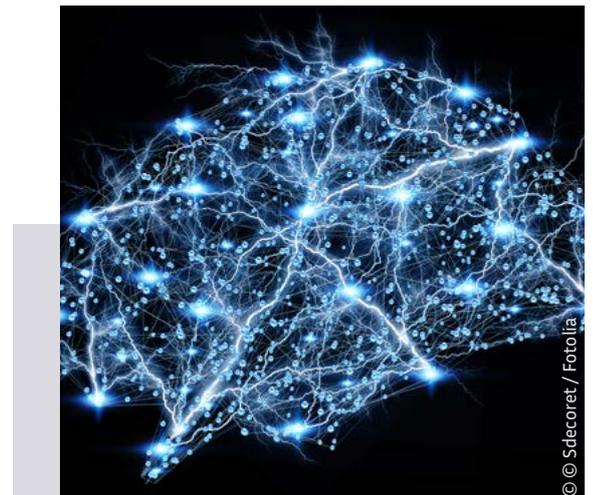
CARACTÉRISTIQUES

La plupart des algorithmes de machine learning s'appuient sur une métrique (*mesure de distance*) permettant de comparer les données deux à deux. L'apprentissage de métrique (*metric learning*) consiste à ajuster automatiquement la métrique à partir d'un ensemble de paires de points que l'on sait similaires ou dissimilaires afin qu'elle soit adaptée à la tâche cible.

La métrique apprise peut, par exemple, être utilisée pour trouver dans une base de données (*texte, image, ...*) les éléments les plus proches sémantiquement d'un élément cible (*Information retrieval*). Elle peut aussi être intégrée à d'autres algorithmes de machine learning, par exemple pour améliorer les modèles de plus proches voisins, dans des tâches de classification, régression ou détection d'anomalie. Le metric learning peut aussi être vu comme un moyen de réduire les dimensions des données de façon (*faiblement*) supervisée.

QUELS AVANTAGES ?

- Le Metric Learning peut être appliqué au delà du cas supervisé classique, dans des situations où les données sont « *faiblement supervisées* » (par exemple, des paires de points similaires/dissimilaires).
- La librairie *metric-learn* implémente en Python plusieurs algorithmes populaires de metric learning supervisés et faiblement supervisés. Elle est compatible avec *scikit-learn*, bibliothèque de référence pour le machine learning en Python, permettant de combiner *metric-learn* avec les outils et chaîne de traitements de machine learning.



© Sdecoret / Fotolia

USE CASES

On peut apprendre une métrique adaptée à la reconnaissance faciale en l'entraînant sur un ensemble de paires d'images représentant la même personne et de paires représentant des personnes différentes.

- **Recherche d'information** : trouver les documents/ images d'une base les plus similaires à une requête.
- **Vision par ordinateur** : vérification faciale,
- **Détection automatique de doublons dans les bases**
- **Recommandation de contenus**
- **Clustering** : profils clients, documents, etc.
- **Preprocessing** : dans la plupart des algorithmes de machine learning.



FICHE IDENTITÉ

- Licence : MIT
- Langage de programmation: : Python

FONCTIONNALITÉS GÉNÉRIQUES

En fonction du type de supervision disponible pour les données d'entraînement, metric-learn permet d'appliquer deux catégories d'algorithmes :

1. Apprentissage supervisé

L'apprentissage se fait à partir d'un ensemble de points associés à des classes (labels) comme dans tout problème de classification standard. L'objectif est alors d'apprendre une distance (ou de manière équivalente, une transformation des points dans un nouvel espace) qui rapproche les points d'une même classe alors qu'il éloigne les points de classes différentes.

2. Apprentissage faiblement supervisé

La supervision n'existe qu'au niveau de paires de points (ou triplets ou quadruplets). Un exemple classique est un ensemble de paires positives et négatives. Dans ce cas, l'objectif est de rapprocher les points des paires positives et d'éloigner les points de paires négatives.

Après entraînement, on peut alors transformer de nouveaux points dans ce nouvel espace ou calculer les distances entre des paires de points. Ce Transformer est compatible avec les pipelines et les routines de sélection de modèle de scikit-learn (grid search, cross validation, ...). La nouvelle représentation ou distance peut donc être facilement utilisée dans les algorithmes classiques de machine learning pour la classification, la régression, le clustering, la détection d'anomalie, etc.

READ ME

Documentation : <http://metric-learn.github.io>

Github : <https://github.com/scikit-learn-contrib/metric-learn>

