

TP Deap Learning

Céline Jacqueton, Eloïse Giroud

12/05/2020

Contents

Introduction TP Deap Learning	2
1. D'après le contexte, à quoi correspondent les variables prédictives et cible?	2
2. Récupération d'images sur image_search_scraper.py : remarques	3
3. Proposez une méthode pour réaliser ce travail, puis affichez et interprétez le script Python dataset_splitter.py	4
4. Observez et commentez la structure de ce fichier source	5
5. Lancement de l'apprentissage sur le corpus d'apprentissage précédemment créé: remarques:	6
6. Testez ces différents modèles et interprétez les données du fichier models.csv au moyen de la technique de votre choix (par exemple avec R)	7

Introduction TP Deep Learning

L'objectif de ce TP est de prendre le contrôle d'un moteur de recherche (ici FireFoxe) et de créer un réseau neuronal pour que le système apprenne à repérer des images cohérentes de manière précise.

Nous avons choisi d'étudier le domaine "Animaux de la ferme" avec les catégories suivantes: vache charolaise, vache aubrac, cochon, cheval, chèvre

1. D'après le contexte, à quoi correspondent les variables prédictives et cible?

Les variables prédictives correspondent aux input (de x_1 à x_n) = training set (= un ensemble d'observation).
Les variables cibles correspondent aux output = y c'est à dire ce que le pc renvoie.

2. Récupération d'images sur `image_search_scraper.py` : remarques

L'IA permet de produire un dossier « raw set » dans lequel on trouve 100 images de chaque catégorie correspondant à notre domaine d'étude, trouvées sur le moteur de recherche FireFoxe. Ensuite nous trions manuellement ces images brutes pour enlever les images incohérentes.

En effet, nous remarquons que les images ne correspondent pas toutes à ce qu'on aurait voulu. On ne reçoit pas forcément l'animal vivant mais aussi des dessins et ce n'est pas ce qu'on souhaite apprendre à l'IA. Ce sont ces images qu'il faut supprimer pour préciser notre demande.

Lorsque la photo comprenait une partie du corps de l'animal, plusieurs animaux ou un animal avec un homme nous avons supprimées les images pour ne pas induire l'IA en erreur. Il nous reste à ce stade x images de chèvres, Ensuite on sépare aléatoirement notre raw set en deux.

3. Proposez une méthode pour réaliser ce travail, puis affichez et interprétez le script Python `dataset_splitter.py`

Il faut entraîner le système à la reconnaissance d'images avec le « training-set » pour qu'il reconnaisse précisément la demande chèvre en tant que chèvre par exemple. Pour tester qu'il a bien appris, nous utilisons le test set pour voir combien d'images il reconnaît comme étant réellement une chèvre dans x images de chèvres.

4. Observez et commentez la structure de ce fichier source

5. Lancement de l'apprentissage sur le corpus d'apprentissage précédemment créé: remarques:

Résultats obtenus avec le domaine étudié:

CNN multiclass (6 catégories) avec 10 passes cheval : 5 / 10 (50 %) chevre : 3 / 14 (21 %) cochon : 4 / 11 (36 %) poule : 12 / 17 (70 %) vache_aubrac : 9 / 18 (50 %) vache_charolaise : 4 / 16 (25 %) Global : 37 / 86 (43 %)

CNN multiclass (6 catégories) avec 25 passes cheval : 5 / 10 (50 %) chevre : 5 / 14 (35 %) cochon : 4 / 11 (36 %) poule : 11 / 17 (64 %) vache_aubrac : 8 / 18 (44 %) vache_charolaise : 5 / 16 (31 %) Global : 38 / 86 (44 %)

Nous remarquons en étudiant les statistiques obtenues que le pourcentage de réussite global n'a augmenté que de 1% en faisant + de 2x + de passes. La solution a préférer est donc peut-être d'augmenter le corpus d'image de base (réponse en 6)

6. Testez ces différents modèles et interprétez les données du fichier models.csv au moyen de la technique de votre choix (par exemple avec R)

On fait une ACP pour déterminer quels facteurs comptent le + pour avoir le meilleur pourcentage d'apprentissage. Ensuite selon la réponse de l'ACP on peut tracer des graphiques du pourcentage de réussite en fonction de la/des variables sélectionnées .