

Projet final

Reconnaissance d'images

Guillaume Wisniewski
(d'après une idée d'A. Allauzen)
`guillaume.wisniewski@limsi.fr`

21 février 2017

Résumé L'objectif du projet est de concevoir un système de reconnaissance d'images. Il s'agit d'un projet de recherche dont les contours sont volontairement flous : l'objectif est que vous puissiez laisser libre court à votre imagination tout en nous montrant que vous avez bien compris les grands principes de l'apprentissage statistique.

Le projet est à réaliser en binôme. Il sera évalué par rapport d'une vingtaine de pages et une soutenance orale.

1 Aperçu

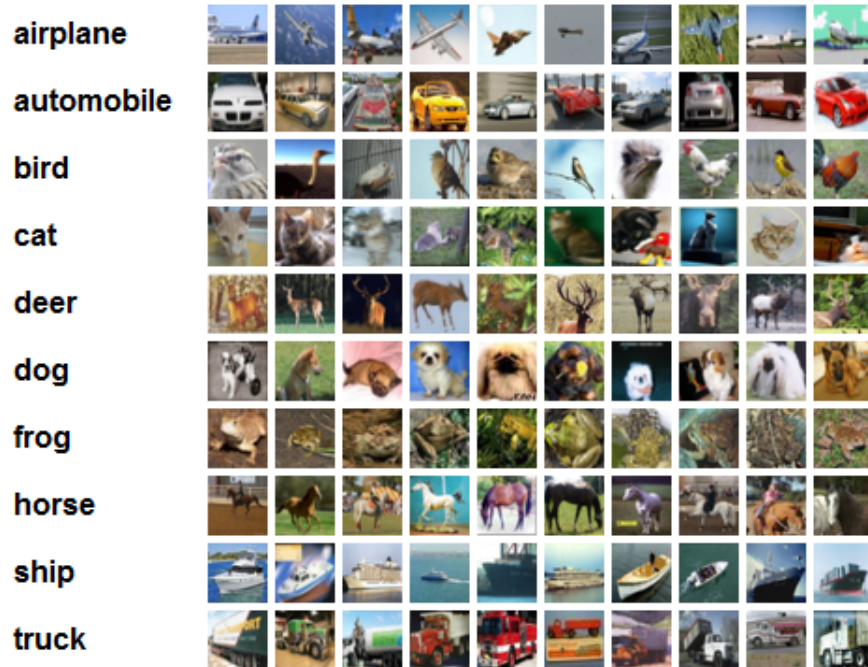


FIGURE 1 – Exemple d’images à identifier

L'objectif de ce projet est de développer un système de reconnaissance d'image capable d'identifier automatiquement l'objet représenté sur une image. La Figure 1 donne plusieurs exemples des images à catégoriser et des classes correspondantes. La principale idée du projet est d'utiliser la méthodes des k -moyennes pour déterminer automatiquement la représentation la plus pertinente des images ; une méthode de classification comme le perceptron permet alors déterminer la classe. Cette approche a été initialement proposée dans l'article :

An Analysis of Single-Layer Networks in Unsupervised Feature Learning, A. Coates, H. Lee et A. Y. Ng, Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 2011

qui est disponible à l'URL :

<http://jmlr.csail.mit.edu/proceedings/papers/v15/coates11a/coates11a.pdf>

2 Le base de données Cifar-10

Nous utiliserons, dans nos expériences, la base de données CIFAR-10 qui regroupe des images représentant 10 types d'objets différents (chien, camion, avion, ...). Cette base est disponible à l'url : <http://www.cs.toronto.edu/~kriz/cifar.html>.

Vous devez commencer par vous familiariser avec le format de la base (comment sont stockées les données), le type des images représentées (est-ce que la tâche est difficile pour un humain ? pour un ordinateur ? quel taux d'erreur peut-on espérer atteindre ? ...) et les statistiques « générales » du jeu de données (combien y a-t-il de classes ? d'exemples par classe ? ...).

Votre rapport final devra comporter une section rapportant l'ensemble de ces observations, ainsi qu'une discussion sur l'intérêt de la tâche aussi bien du point de vue des applications que des méthodes statistiques mise en œuvre.

La base CIFAR-10 est un des principaux jeux de données utilisées pour évaluer les performances des nouvelles méthodes d'apprentissage développées ces dernières années. Une simple recherche sur l'Internet vous permettra de trouver de nombreux articles scientifiques permettant d'obtenir des taux d'erreurs particulièrement faibles. Quel est l'ordre de grandeur que les méthodes d'apprentissage « moderne » permettent d'obtenir ? Pensez-vous être capable d'atteindre un tel niveau de performance ?

Remarque : CIFAR-10 est une base nettement plus grande que celles que nous avons utilisées en TP. Nous vous conseillons fortement d'utiliser, dans vos développements, une base de données plus petite (par exemple en ne considérant qu'une partie des exemples et en commençant par essayer de distinguer deux types d'objets).

3 Approche naïve

Une manière naïve de traiter la tâche est d'utiliser une méthode de classification considérant comme caractéristique directement la « valeur » de chaque pixel.

Comment représenter les pixels ? Quelles sont les performances obtenues ? Quelles sont les classes qui sont bien reconnues ? quelles sont celles qui sont systématiquement mal reconnues ? Pourquoi ?

Il est également possible d'utiliser l'algorithme des k -moyennes pour partitionner les images. Quelle distance peut-on utiliser ? Peut-on utiliser ce partitionnement pour classer les objets ? Si oui, quelle performance obtient-on ?

4 Apprentissage de caractéristiques

Nous allons maintenant mettre en œuvre la méthode proposée par A. Coates *et al.* (cf. référence dans la section 1). Cette méthode est résumée par la Figure 1 de l'article qu'il est important de bien comprendre.

L'idée essentielle est de construire automatiquement une représentation de l'image par un vecteur de réels. Pour cela, nous commençons par extraire du corpus d'apprentissage un (très) grand nombre de *patches*. Un patch est simplement une petite partie d'une image. En multipliant les patches, on peut espérer capturer certaines caractéristiques importantes (par exemple une oreille ou un œil). Il est alors possible de représenter une image à partir des patches qui la compose.

La première étape de l'approche consiste à construire un *dictionnaire de N patches* :

- chaque image est divisée en 4 patches ;
- l'ensemble des patches du corpus d'apprentissage est collecté ;
- on utilise un algorithme des k -moyennes pour déterminer N représentants, décrivant un *dictionnaire* de patches.

En général, on considère des dictionnaires comportant plusieurs centaines d'éléments.

Il est alors possible de construire le vecteur de caractéristiques d'une nouvelle image de la manière suivante :

- divisez l'image en 4 patches ;
- calculer la distance entre chaque patches et les N clusters ;
- chaque patch est représenté par un vecteur de N caractéristiques binaires ; la i^e caractéristique est non nulle si et seulement si le centre du i^e cluster est le plus proche du patch
- le vecteur de caractéristiques de l'image est obtenu en concaténant les 4 vecteurs correspondant à chaque patch.

Finalement, cette représentation est utilisé comme entrée d'un algorithme d'apprentissage.

5 Travail à réaliser

Le projet est à réaliser en binômes. Il vous est demandé de :

- répondre aux différentes questions du sujet ;
- implémenter et tester l'approche naïve ;
- implémenter et tester l'approche proposée ;
- tester l'impact des différents paramètres de la méthode, analyser les résultats obtenus et les dictionnaires trouvés.

L'ensemble de votre travail donnera lieu à un rapport de 20 pages maximum et une soutenance orale.

Les rapports devront être envoyé aux différents chargés de TP (guillaume.wisniewski@limsi.fr, veronique.ventos@lri.fr et aurelien.decelle@lri.fr) avant le 14 avril minuit. Les soutenances auront lieu la semaine du 17 avril. L'ordre de passage des différents groupes et la date précise vous seront transmis rapidement.