Culture Scientifique

Conférence : Deep Learning

21506749
Master 1 Informatique Romain Ambroise

L’apprentissage profond traduit de l’anglais « deep learning », est un ensemble de méthodes d’apprentissage automatique. Cette technologie intervient dans le domaine de l’intelligence artificielle, la bioinformatique, la sécurité, etc. Ces techniques ont permis des progrès importants dans la perception visuelle et la perception auditive pour une machine. Elle a notamment permis de contourner les problématiques de la reconnaissance faciale.
Comment a évolué cette technologie qui s’est développée en quelques années? Nous verrons dans un premier temps les impacts que cette technologie a engendré et comment fait-on pour l’évaluer. Puis nous détaillerons son mode de fonctionnement. Pour finir nous aborderons les principaux frameworks qui se distinguent sur le marché.

Le deep learning a eu un impact médiatique fort grâce aux frontières qu’il a repoussé. En effet, lorsque la machine de la filiale de Google : DeepMind a remporté une partie du jeu de Go contre le sud corréen Lee Sedol cela a démontré que rien n’était impossible. La complexité de développer une intelligence artificielle pour ce jeu avec un nombre de configuration important est de l’ordre de $3 * 10^{684}$, bien supérieur à la complexité du jeu d’échec.
Du côté de la reconnaissance faciale, Facebook annonce en juin 2014 ses recherches nommés « DeepFace » . Cet outil a pour but de reconnaître un visage, quelle que soit son orientation. Nvidia quant à lui propose du matériel toujours plus performant pour répondre aux besoins du deep learning avec notamment la carte graphique TitanX. Google se penche désormais sur des problématiques plus fondamentales avec la recherche sur le cancer.
On ne connaît pas encore les limites de cette nouvelle technologie. Comment l’évalue-t-on ?Nous parlerons par la suite de deux benchmarks : PASCAL VOC 2007 pour la classification d’images et LFW pour la reconnaissance faciale.

Pour un ordinateur, une image est un tableau possédant des valeurs numériques. Ce tableau peut être plus ou moins important selon la résolution de l’image. L’objectif de la classification est d’assimiler à ses valeurs numériques une ou un ensemble de catégorie. Ainsi, le programme intelligent doit émettre des prédictions si une catégorie d’objet est apparente dans une image. C’est un problème qui peut s’avérer extrêmement difficile. En effet, les images peuvent avoir des points de vue différent (par rapport à la position de l’appareil photo) ainsi que des variations d’échelle, des occultations, des masquages et des déformations. Les difficultés ne s’arrêtent pas là : l’image peut être riche et peut contenir plusieurs autres catégories. L’ordinateur doit pouvoir les différencier. La méthode à suivre est basée sur l’apprentissage à partir d’une base de données colossales contenant plusieurs milliers de photos représentant une catégorie. Cette méthode est donc basée sur des statistiques.
Ce programme va posséder un pipeline de classification. Il prendra en entrée un ensemble de N images chacune marquée d’une catégorie. Ces données seront appelées données d’entraînement. L’apprentissage consiste à apprendre à partir de ces

données à quoi ressemble une catégorie en comparant les valeurs d'une image à 100 000 autres appartenant à la catégorie. Enfin, on va évaluer le système en le soumettant à des images qu'il n'a jamais analysé. On va comparer les vraies étiquettes données à ces images à celle que le système va donner. On attribue une valeur nommée « average precision » qui sera positive si l'image contient l'objet de la catégorie et négative sinon. On établira une moyenne des performances du système sur l'ensemble des catégories du benchmark. Ainsi il y aura quatre issues au test : le vrai positif, le faux positif, le vrai négatif et le faux négatif.
La progression des systèmes sur le benchmark PASCAL VOC 2007 ne fait qu'évoluer depuis 2008 on passe de 54,48 % de moyenne average precision à 82,42 % en 2014. Les systèmes possèdent des couches appelés « maxpool », « input » et bien d'autres.

Un autre benchmark appelé « Challenge Labeled Faces in the Wild » soumet les systèmes à d'autres épreuves. Ce test propose une banque d'image montrant des Hommes sous différents points de vue : angle différent, vêtements différents (avec lunettes, chapeaux ou non). Ce benchmark impose les mêmes difficultés que le précédent si ce n'est plus. Des systèmes proposent de découper les images et d'envoyer chaque partie dans des couches (CNN-H1, CNN-H2) afin de les analyser. On a pu constater que la machine a battu la perception de l'oeil humain sur ce test. Le système MM-DFR-JB a obtenu une précision de 99.02 %.
Ces progressions s'expliquent par l'évolution du Big Data, les systèmes peuvent s'entraîner sur des banques de données gigantesques (14 millions d'images). Par exemple, certains catégories du benchmark se sont enrichies : 2,8 millions d'images d'animaux, 1 million d'images de fleurs ou encore 1 million d'images de nourriture. Les bases de visage quant à elles sont tout aussi colossales : 4,4 millions de photos pour le dataset SFC. Ces datasets de visages ne sont cependant pas toujours accessibles à tous et demeurent privés.
On remarque également un important travail d'annotation d'image effectué par des personnes. Les ingrédients de ce succès sont donc : des jeux de données très riches, des codes de calcul performants, des systèmes de calculs (GPU) efficaces. Mais aussi l'accessibilité open-source des systèmes développés autour de cette technologie. On retrouve CNTK par Microsoft, TensorFlow par Google mais encore Caffe par Facebook.

On retrouve différentes types de tâches traitées par les machines : la regression, la classification, le renforcement, le clustering et la réduction de dimensionalité. On va essayer de minimiser les erreurs faites par une machine avec une fonction de coût pour la classification par exemple. L'objectif de l'apprentissage est de réduire ces risques.

On peut voir des architectures possédant de nombreuses couches : max pooling, fully connected ou encore des couches de convolution. Cette dernière consiste en un empilage multicouche de perceptrons. Le perceptron a été inventé en 1957 par Frank Rosenblatt. C'est un algorithme d'apprentissage supervisé. Il s'agit d'un neurone formel qui permet de déterminer automatiquement les poids

synaptiques de manière à séparer un problème d'apprentissage supervisé. Si le problème est linéairement séparable, un théorème assure que la règle du perceptron permet de trouver une séparatrice entre les deux classes. Ceci a été testé sur des jeux de données d'images représentant des iris. Le programme a comparé les longeurs des pétales et des sépales. Afin d'émettre des prédictions basées sur des statistiques.
« Back propagation » ou traduit en rétropropagation du gradient est une technique pour calculer le gradient d'une erreur pour chaque neurone d'un réseau de neurones.

L'architecture Le Net-5 conçu par Yann Le Cun en 1998 est un réseau de convolution (CNN) pour la reconnaissance de codes postaux. Il est basé sur une configuration à 7 couches. Il est capable d'apprendre des filtres de convolution pour le traitement spatial des images. On distingue trois hyper-paramètres contrôlant la taille de la couche de sortie : la profondeur (nombre de couche appliqué), la foulée (saut spatial entre deux masques) et le rembourrage (la gestion des bords).

En conclusion, le deep learning est une science très complexe qui a permis de nombreuses prouesses. Quelles sont ses limites ?

Le deep learning est une science émergente qui est à la pointe de technologie. Les chercheurs s'appuyent sur des alliés conséquents comme le Big Data pour le faire évoluer. Au début de la conférence, le professeur nous a demandé d'imaginer une solution pour reconnaître une chaise dans une image. J'ai proposé l'approche symbolique : on découpe l'objet en éléments qui le compose. Cette conférence m'a beaucoup apporté et m'a permis de voir qu'il existe toujours une solution même aux problèmes qui à priori semblent les plus complexes. Je me pose cependant une question : comment les entreprises ont autant fait évoluer leur banque d'images annotés ?
Facebook s'est servis de la fonction « identifier quelqu'un » sur son réseau social. Ainsi les utilisateurs annotaient les images à la place du géant. C'est incroyable et dénoncable la façon d'agir. On peut se poser une réelle question sur l'utilisation des données par les autres géants de l'informatique. Grâce au deep learning, on a pu voir précédemment qu'une simple recherche google de l'identité d'une personne permettait de voir quelques photos de la personne en question. Va-t-on vers un monde où les droits à l'image et à la vie privée sont bafoués ?
Tous les benchmarks visant à évaluer la reconnaissance faciale proposent des images de personne. Est-il possible qu'un système puisse un jour identifier une personne sur une photo où elle est beaucoup plus jeune et une photo actuel ?