
Intelligence artificielle

Cours 2 – Réseaux de neurones convolutifs



Table des matières

Pytorch	2
Les tenseurs	2
Les couches	2
La couche Linear	2
La couche Conv2d	3
La couche MaxPool2d	3
Points clés	3
L'explosion de la popularité des réseaux de neurones convolutifs	4
Les couches convolutives	4
L'augmentation de la puissance de calcul	4
L'augmentation de la quantité de données	5
Des résultats impressionnants	6
Et ChatGPT dans tout ça?	6

Pytorch

La librairie Pytorch est une librairie d'apprentissage machine (machine learning) basée sur la librairie Torch. Elle est principalement utilisée pour le traitement de données et l'entraînement de réseaux de neurones.

Pour installer Pytorch, il faudra utiliser la commande suivante :

```
1 pip3 install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cpu
```

Les tenseurs

Les tenseurs sont des tableaux multidimensionnels qui peuvent être manipulés par des opérations mathématiques.

Cette librairie ressemble beaucoup à la librairie numpy mais permet d'effectuer des calculs sur des cartes graphiques.

Les couches

Il existe plusieurs types de couches dans Pytorch, par exemple :

- Linear
- Conv2d
- MaxPool2d
- ...

La couche Linear

La couche Linear est une couche qui permet de calculer une fonction linéaire qui est donnée par la formule :

$$y = x \cdot W^T + b \quad (1)$$

où x est un vecteur d'entrée, W est une matrice de poids et b est un vecteur de biais.

La dimension de W est $[n_{out}, n_{in}]$ et la dimension de b est $[n_{out}]$.

Une multiplication matricielle est une opération qui prend en entrée une matrice A de dimension $[n, m]$ et une matrice B de dimension $[m, p]$ et qui renvoie une matrice C de dimension $[n, p]$ telle que :

$$C_{i,j} = \sum_{k=1}^m A_{i,k} \cdot B_{k,j} \quad (2)$$

La couche Conv2d

La couche Conv2d est une couche qui permet de calculer une convolution 2d.

Soit I une image de dimension $[n, m]$ et F un filtre de dimension $[k, l]$, alors la convolution C de I par F est une image de dimension $[n - k + 1, m - l + 1]$ telle que :

$$C_{i,j} = \sum_{a=1}^k \sum_{b=1}^l I_{i+a-1, j+b-1} \cdot F_{a,b} \quad (3)$$

La couche MaxPool2d

La couche MaxPool2d est une couche qui permet de réduire la taille d'une image en prenant la valeur maximale dans une fenêtre de taille $[k, l]$.

Généralement, on utilise une fenêtre de taille $[2, 2]$ et un pas de $[2, 2]$ ce qui signifie qu'un seul pixel sur quatre sera conservé.

Elle permet de réduire l'utilisation de mémoire et de calculer plus rapidement les convolutions.

Points clés

- Le but d'un réseau de neurones est d'approximer une fonction f à partir de neurones. Le même réseau de neurones peut approximer différentes fonctions.
- Plus un réseau contient de neurones, plus il sera capable d'approximer des fonctions complexes.
- La qualité et la quantité des données d'entraînement sont très importantes pour entraîner un réseau de neurones.
- Un réseau de neurones possède énormément de paramètres, il est généralement impossible d'expliquer son fonctionnement. C'est un modèle boîte noire.
- Il est impossible de prédire comment un réseau de neurones va réagir à des données qu'il n'a jamais vu.
- L'entraînement d'un réseau de neurones est un processus itératif qui consiste à modifier les paramètres du réseau de neurones afin de minimiser la fonction de coût. C'est un processus dont l'élégance et l'efficacité sont discutables.

L'explosion de la popularité des réseaux de neurones convolutifs

Depuis quelques années, les réseaux de neurones convolutifs sont devenus très populaires et se sont immiscés dans de nombreux domaines.

Ce succès est dû à plusieurs facteurs.

Les couches convolutives

Entraîner un réseau de neurones utilisant des couches Conv2d est beaucoup plus efficace que d'entraîner un réseau de neurones avec des couches Linear. En effet, une couche Conv2d utilise beaucoup moins de paramètres qu'une couche Linear.

De plus, une couche Conv2d prend en compte la structure de l'image et l'aspect local des données : dans une couche linear toutes les entrées sont analysées en même temps.

Les couches convolutives exploitent la localité de l'information et sont particulièrement efficaces pour :

1. L'analyse d'images
2. L'analyse de texte
3. L'analyse audio
4. ...

Dans ce genre de données, les données sont souvent structurées et l'information est locale, en effet dans une phrase, les mots proches sont plus liés que les mots éloignés.

Chat-GPT produit par exemple du texte en produisant un mot à la fois, chaque mot doit être cohérent avec les mots précédents.

L'augmentation de la puissance de calcul

La puissance de calcul des ordinateurs augmente régulièrement, ce qui permet d'entraîner des réseaux de neurones plus complexes ou les mêmes réseaux de neurones plus rapidement.

Mais le facteur le plus important est l'utilisation de cartes graphiques pour effectuer des calculs. En effet, la carte graphique (Graphics Processing Unit en anglais (GPU)) est une unité de calcul très efficace pour effectuer des calculs en parallèle, ce qui est possible dans les réseaux de neurones convolutifs.

Pour ces calculs, les cartes graphiques sont beaucoup plus efficaces que les processeurs.



FIGURE 1 – Une carte graphique avec laquelle il n'est pas possible de jouer à des jeux vidéos

Dans la figure 1, nous pouvons voir une carte graphique, une Nvidia Tesla, qui n'a même pas de sortie vidéo.

Il est aussi intéressant de noter que les cartes graphiques sont également utilisées pour le minage de cryptomonnaies.

L'augmentation de la quantité de données

Les données sont devenues de plus en plus accessibles, en effet, il est possible de trouver des données sur internet, de les collecter, de les acheter, etc.

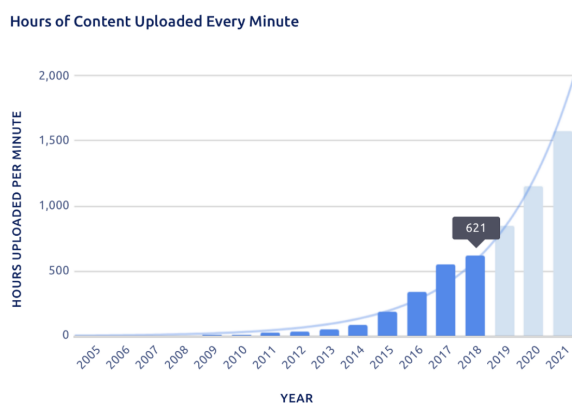


FIGURE 2 – Heures de contenu mis en ligne par minute en fonction des années

Dans la figure 2, nous pouvons voir l'augmentation de la quantité de données sur Youtube.

Les entreprises comme Google, Facebook, Amazon, etc. possèdent énormément de données et les utilisent pour "améliorer leurs services".

Ils utilisent également ces données pour générer des profils utilisateurs et les vendre à des entreprises tierces cherchant à faire de la publicité ciblée.

Ils utilisent également ces données pour entraîner de nouveaux modèles de réseaux de neurones, par exemple Chat-GPT. Nous utilisons la suite office de Microsoft dans les gymnases vaudois, Microsoft a un partenariat avec

OpenAI, l'entreprise qui a développé GPT-2 et Chat-GPT.

Il n'est pas impossible que vos données soient utilisées par des entreprises sans que vous le sachiez.

Des résultats impressionnants

Les CNNs ont explosé en popularité, car la performance de ces modèles dans diverses tâches a été démontrée :

1. [Conduite autonome](#)
2. [Analyse d'image médicale](#)
3. [Jeu de Go](#)
4. [Deepfakes 1](#)
5. [Deepfakes 2](#)
6. [ChatGPT 1](#)
7. [ChatGPT 2](#)
8. [Midjourney](#)
9. [Midjourney](#)
10. [Reconnaissance de visage](#)
11. ...

Et ChatGPT dans tout ça ?

GPT signifie Generative Pre-trained Transformer, la famille de modèles de langage à laquelle appartient ChatGPT.

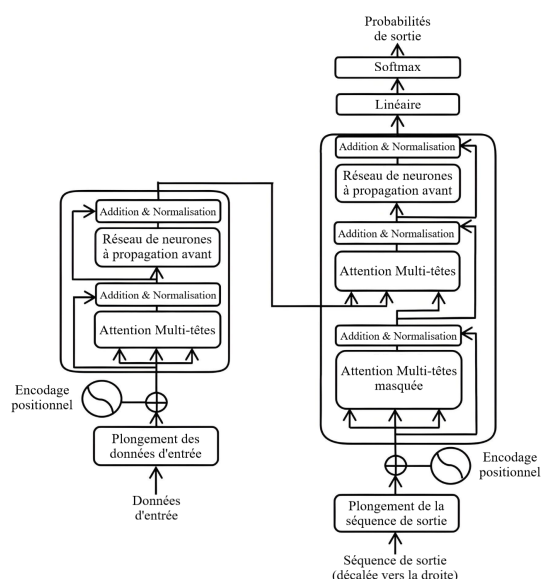


FIGURE 3 – L'architecture d'un réseau de neurones de type transformeur, comme ChatGPT

ChatGPT est « simplement » un réseau de neurones convolutifs spécialisé. Pour plus d'information, vous pouvez commencer par lire l'article Wikipedia sur les réseaux transformeurs en [français](#) ou [anglais](#) qui est plus détaillé.