

Data Science “CaptionMe”

ENSEIGNANT : Wandrille DOMAIN

Étudiants : Merieme BOURENANE, Chifaa WEHBE, Paulius MICKUNAS

UE : Data Science

Année : 2020-2021

Sommaire

1. Introduction

2. Choix du projet

- a. Comparaison et choix final
- b. Problématique et dataset

3. Mise en place de projet

- a. Démarche de résolution générale
- b. Définition des outils utilisés
- c. Architecture logicielle - matérielle
- d. Documentation et veille technologique sur le ML

4. Démarche de résolution itérative - 2 méthodes

- a. Exploration
- b. Pre-processing data
- c. Machine Learning Model architecture and development
- d. Training Model

5. Résultats

- a. Comparaison résultat attendu/prédiction

6. Conclusion

- a. Les points d'amélioration

1. Introduction

1. Introduction

— — —

Consigne initiale :



Vous allez devoir
découvrir la Data
Science au
travers d'un projet
libre !

2. Choix du Projet

a. Comparaison et choix final

— — —

Talking AI



- ✓ Faire le meilleur choix de réponse
- ✗ Trop complexe, dataset mal organisé

Image captioning



- ✓ Plusieurs domaines d'app : médecine, administration
- ✗ Complexe : traitement image, NN

Image colorization



- ✓ Recoloriser des photos des grands parents
- ✗ Mathématiques avancées - fait l'objet de recherches

b. Problématique et dataset

Qu'est ce que CaptionMe ?

- Un analyseur d'image
- Un analyseur et générateur de texte, e.g de phrases



Un générateur de descriptions

“CaptionMe”

Quelles sont les données de travail?

Dataset Flickr30k :

- Taille : 4Go
- Nb d'images : \approx 30k
- Nb de colonnes : 3

	image_name	comment_number	comment
0	1000092795.jpg	0	two young guys with shaggy hair look at their ...
1	1000092795.jpg	1	two young , white males are outside near many ...
2	1000092795.jpg	2	two men in green shirts are standing in a yard .
3	1000092795.jpg	3	a man in a blue shirt standing in a garden .
4	1000092795.jpg	4	two friends enjoy time spent together .

— — —

3. Mise en place du projet

a. Démarche de résolution générale

1

1ère méthode :

- 1) Exploration et pre-processing des données
- 2) Extraire les caractéristiques des images
- 3) Traiter les séquences de mots de la description
- 4) Création du modèle permettant de générer des séquences de mots
- 5) Entrainement du model
- 6) Validation du model
- ~~7) Test du model~~

2

2ème méthode :

- 1) Exploration et pre-processing des données
- 2) Extraire les caractéristiques des images
- 3) Conversion des descriptions en vecteurs spatiaux
- 4) Création du modèle permettant de générer des séquences de mots
- 5) Entrainement du model
- 6) Test du model

— — —

b. Définition des outils utilisés

— — —

>>> Langages

Python



>>> Frameworks

Tensorflow



>>> Bibliothèques

Keras



>>> Notebook

Google Colaboratory



c. Architecture logicielle - matérielle

Ressources logicielles :

- Google Colaboratory **Pro**

Ressources matérielles :

- Accès à un **GPU**
- **Mémoire RAM** plus importante
- **Mémoire Disque** plus importante

```
Sun May 30 10:57:16 2021
+-----+
+ NVIDIA-SMI 465.19.01    Driver Version: 460.32.03    CUDA Version: 11.2    +
+-----+-----+-----+-----+-----+-----+
+ GPU  Name            Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
+ Fan  Temp            Perf   Pwr:Usage/Cap|     Memory-Usage | GPU-Util  Compute M. |
+-----+-----+-----+-----+-----+-----+
+   0   Tesla V100-SXM2...    Off      | 00000000:00:04:0 Off |   0          0      |
+ N/A   33C             P0      25W / 300W |  0MiB / 16160MiB |   0%      Default  |
+-----+-----+-----+-----+-----+-----+
+
+ Processes:
+ GPU   GI    CI          PID    Type   Process name                      GPU Memory
+   ID   ID     ID              |                 | Usage
+-----+-----+-----+-----+-----+-----+
+ No running processes found
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+
+ Your runtime has 27.3 gigabytes of available RAM
+
+ You are using a high-RAM runtime!
```

d. Documentation et veille technologique sur le ML

Supports d'informations - Bibliographie :

- Documentation
- Papiers de recherche
- Tutoriels

Exemples de sujets :

- Convolutional Neural Networks - CNN
- Image detection and object localization
- Natural Language Processing - NLP
- ...

54	Données utiles		
55	Sujet	Lien	
56	Image processing Machin Learning Image Classification Prediction using the trained Model	https://towardsdatascience.com/in-depth-machine-learning-image-classification-with-tensorflow-2-0-a76526b32af8	
57	Reseau de Neuronal Convolutifs (CNN) Reconnaissance d'image avec TensorFlow	https://www.tensorflow.org/tutorials/images/cnn	
58	Caption Generator	https://www.analyticsvidhya.com/blog/2020/11/create-your-own-image-caption-generator-using-keras/	
59			
60		https://towardsdatascience.com/building-an-image-captioning-model-with-keras-ebccaadb98b9	
61		https://towardsdatascience.com/creating-vgg-from-scratch-using-tensorflow-a998a5640155	
62			
63	Link clustering after embedment	https://towardsdatascience.com/how-to-cluster-in-high-dimensions-4ef693bacc6	
64		https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html	
65		https://scikit-learn.org/stable/modules/clustering.html	
66	Sklearn map	https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html	
67	NLP	https://towardsdatascience.com/natural-language-processing-nlp-for-machine-learning-d44498845d5b	

4. Démarche de résolution itérative

! [Les slides qui suivront présenteront les deux méthodes.]

a. Exploration

Dataset Flickr30k :

- Taille : 4Go
- Nb d'images : $\approx 30k$
- 5 commentaires par image $\rightarrow 158915$ échantillons

	image_name	comment_number	comment
0	1000092795.jpg	0	two young guys with shaggy hair look at their ...
1	1000092795.jpg	1	two young , white males are outside near many ...
2	1000092795.jpg	2	two men in green shirts are standing in a yard .
3	1000092795.jpg	3	a man in a blue shirt standing in a garden .
4	1000092795.jpg	4	two friends enjoy time spent together .

Test et erreurs !

- ➔ Mauvaise exploration et compréhension du dataset
- ➔ Images entraînées redondantes

- ➔ Meilleure appropriation de dataset après erreurs

b. Pre-processing data - 1

Essai 1 : utiliser la totalité du dataset



Ressources mémoire RAM et Disque insuffisantes pour contenir les données matrices spares

Essai 2 : utiliser 1 commentaire par image



Métriques d'apprentissage très basses

Essai 3 : utiliser 8000 images - (5xC)



Amélioration des métriques d'apprentissage et pas de problèmes d'espace Disque

Essai 1 : utiliser 8000 images - (5xC)



Amélioration des métriques d'apprentissage et pas de problèmes d'espace Disque

b. Pre-processing data - 2

Cleaning and formatting comments :

- Text en minuscule
- Retirer la ponctuation
- Ajout de marqueurs de début et fin de chaînes
“ blablabla” → “startseq ... blablabla ... endseq”

comment_text_clean
two young guys with shaggy hair look at their ...
several men in hard hats are operating a giant...
a child in a pink dress is climbing up a set o...
someone in a blue shirt and hat is standing on...
two men one in a gray shirt one in a black s...



comment_text_surrounded
startseq two young guys with shaggy hair look ...
startseq several men in hard hats are operatin...
startseq a child in a pink dress is climbing u...
startseq someone in a blue shirt and hat is st...
startseq two men one in a gray shirt one in a ...

b. Pre-processing data - 3

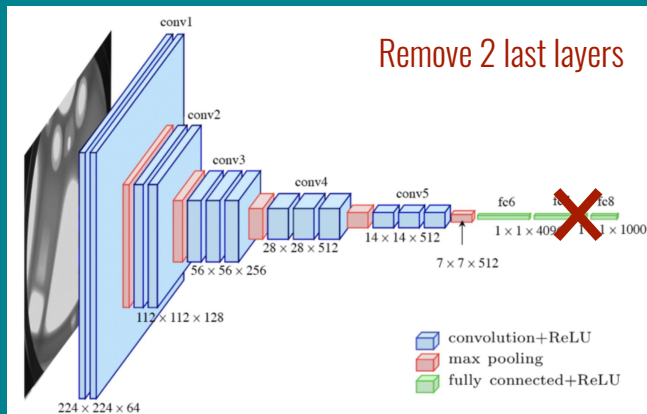
Extraction de caractéristiques des images :

Model pré-entraîné : VGG16

- 6800 images
- ≈ 40 mins de prédictions

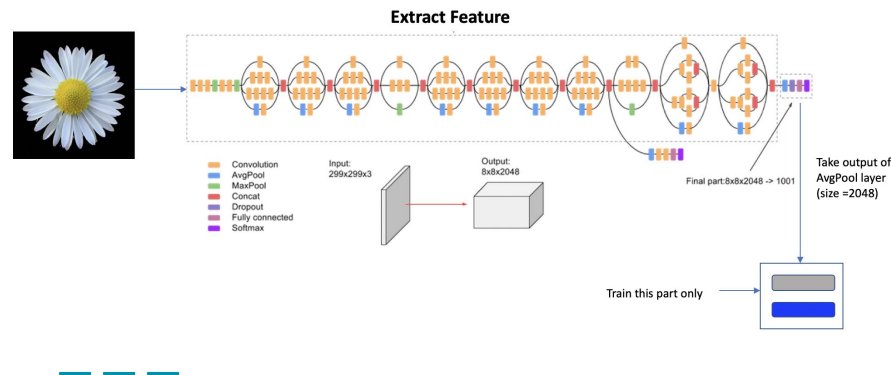
Découper les données en sets :

- 70 % : training
- 15 % : validation
- 15 % : test



Model pré-entraîné : InceptionV3

- 6800 images
- ≈ 35 mins de prédictions
- Moins de features détectables



Processing des commentaires :

Construction de 3 types de données :

- training
- validation

X1 : image features

[illegible]

Y : the next sequence

0.0,0.0,1.03331,0.0,0.189431,0.0,0.943515
0.0,0.0,0.0,0.0,0.0,0.0,0.3932184,0.0,0.54153
0.0,0.0,0.834477,0.0,0.0,0.0,0.2,6927682,0.0,0.0,0.0
0.0,0.914382,0.0,0.2,7951675,0.0,0.0,0.0,0.0,0.0,0.0
0.1,0.486216,0.0,0.0,0.0,0.0,0.4,0.5549097,0.5226562,
0.0,0.0,0.744874,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
0.0,0.3210545,0.6345081,0.0,0.0,0.1,0.6998686,0.0,
0.0,0.0,0.0,0.0,0.4,0.6307626,0.325189,0.0,0.0,
0.0,0.0,0.0,0.1,0.2198475,0.0,0.0,0.0,0.0,0.0,0.11
5502,0.931355,0.8941671,0.0,0.0,0.0,0.0,0.1,0.46
2,0.7828449,0.9252959,0.0,0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
0.0,0.0,0.545282,0.0,0.0,0.0,0.6359756,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,0.0,0.1,0.2566407,0.0,2.5
36,0.0,0.0,0.0,0.16648636,0.26770976,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,0.0,0.2597669,0.2737093,0.0,
0.0,0.0,0.2,0.6437445,0.0,0.0,0.1,0.5686823,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
0.1752594,0.0,0.4588936,0.0,0.0,0.1,0.7417683,0.0

X2 : sequences

[illegible]

Construction de 3 types de données :

- training

c. Machine learning model and architecture - 1

Vectorisation de mots :

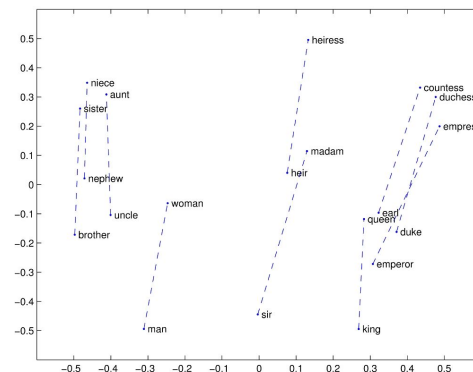


Vecteurs de mots pré-construit : Glove6d

- A chaque mot correspond un vecteur (1, 200)

```
dict(string : list(float32)) glove6d
```

- Représentation spatiale des mots : Mots ayant un sens similaire sont regroupés



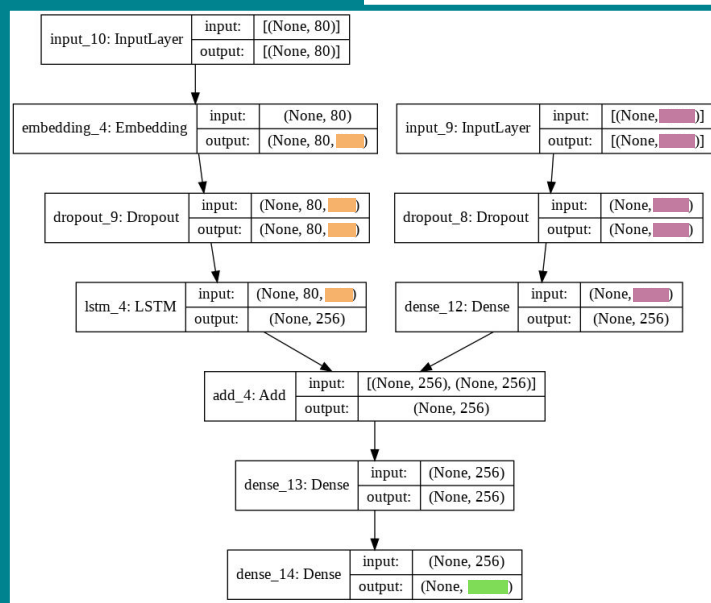
c. Machine learning model and architecture - 2

Modèle en couche de génération de description:

256

4096

10085



200

2048

2101

b. Training model

Analyse des métriques d'entraînement :

Paramètres du modèle :

Epoch : 60 (seuls 30 derniers sur les graphiques)

Optimizer : Adam

Learning rate : 0.001

Interprétation :

- **Loss** : valeurs linéaires décroissantes
- **Accuracy** : valeurs linéaires croissantes avec sursauts



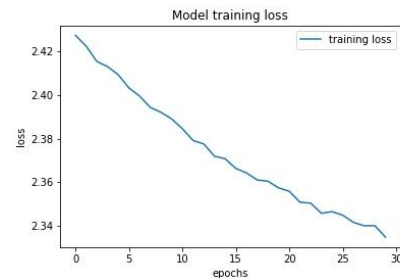
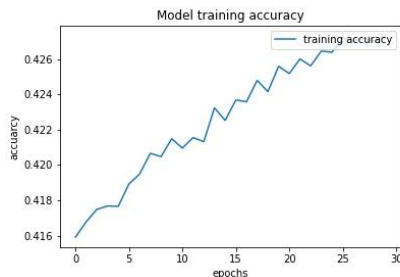
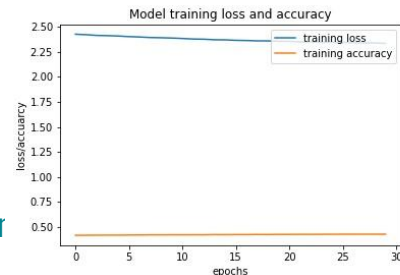
The model is underfitting

Paramètres du modèle :

- **Epoch** : 60 (seuls 30 derniers sur les graphiques)
- **Optimizer** : Adam
- **Learning rate** : 0.001

Interprétation :

- **Loss** : valeurs linéaires décroissantes
- **Accuracy** : valeurs linéaires croissantes avec sursauts



5. Résultats

a. Comparaison résultat attendu/prédiction

Modèle retenu :

- ✓ Architecture n°2
- ✗ Architecture n°1



Simulation !

Interprétation :

- ✓ Grammaire et conjugaison
- ✓ Phrases finies
- ✗ Extraction des objets



6. Conclusion

a. Les points d'amélioration

— — —

Concernant le dataset :

- Ajout d'un set de validation
- Tester sur le dataset connu Flickr8k
- Tester d'autres vectorisations de mots (ex : Word2Vec)

Concernant l'extraction des features des images :

- Tester d'autres modèles pré-entraînés (ex : ResNet, VGG avec la 2ème méthode)

Concernant le modèle :

- Tuning : faire varier les paramètres du modèle (ex : learning rate, batch size)

