深度學習Deep Learning (15)

112-1

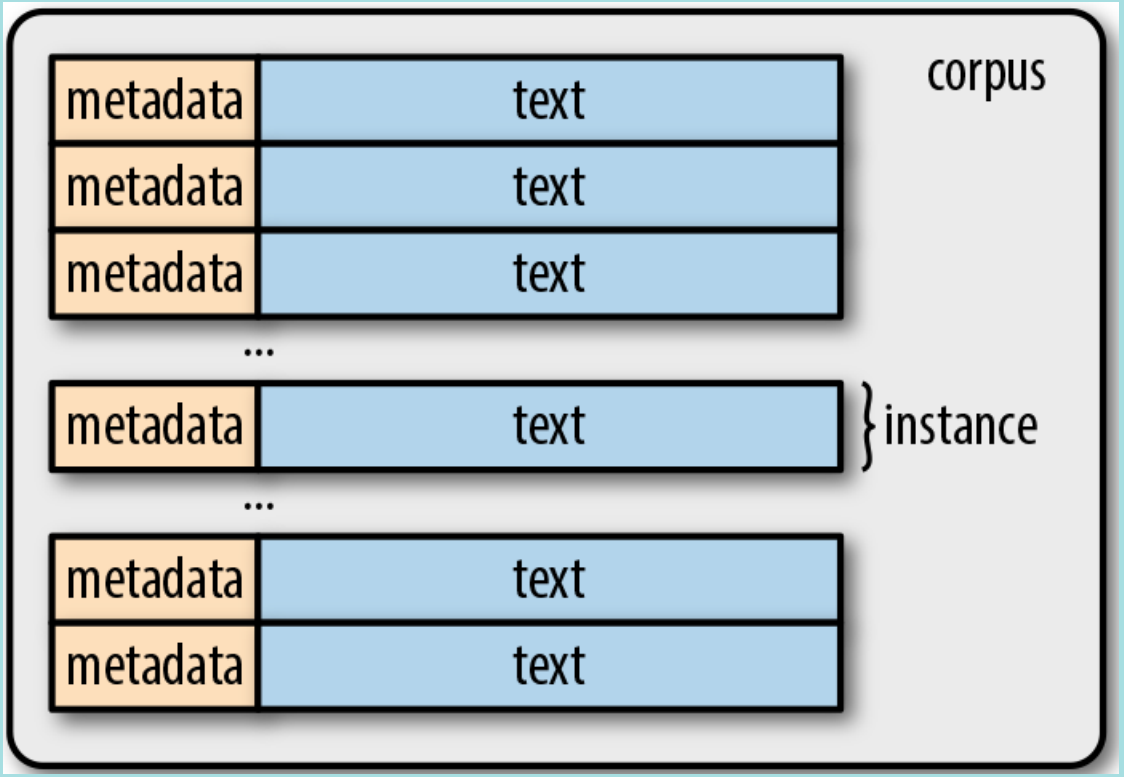
朱學亭老師

課程大綱

- W1-課程介紹/Introduction
- W2-Python/Colab and TensorFlow
- W3-Numpy/Pandas and PyTorch
- W4-Sklearn and 機器學習
- W5-神經網路, TensorFlow, PyTorch
- W6-載客熱點預測
- W7-自動光學檢查(AOI)-1
- W8-自動光學檢查(AOI)-2
- W9-Midterm presentation
- W10-**RNN**
- W11-YoloV5
- W12-AICUP 1
- W13-AICUP 2
- W14-GAN
- W15-NLP
- W16-Final presentation(1)
- W17-元旦
- W18-交報告

NLP(自然語言處理)

The corpus: the starting point of NLP tasks



The process of breaking a text down into tokens is called *tokenization*.

Spacy and NLTK

- spaCy is a free open-source library for Natural Language Processing in Python.
- It features NER, POS tagging, dependency parsing, word vectors and more.
- NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical.
- Gensim: Topic modelling for humans

SpaCy

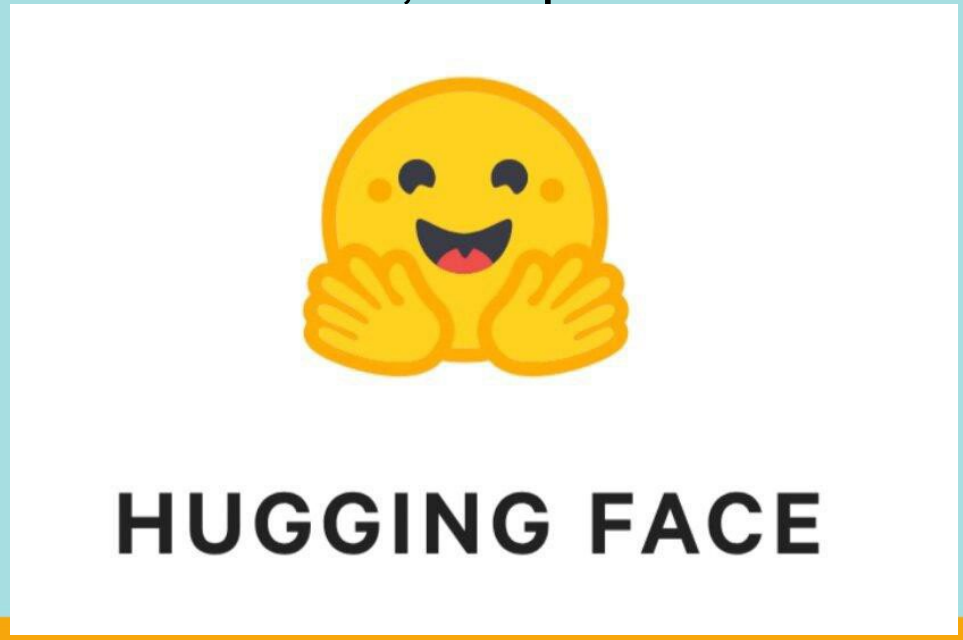
- SpaCy 支援深度學習, 利用自身的機器學習模組 Thinc 做為後端可連接由 TensorFlow, PyTorch, 與 MXNet 等框架訓練出來的統計模型, 利用卷積神經網路 (CNN) 進行詞類標註 (POS tagging), 相依性剖析 (dependence parsing), 文章分類, 以及命名實體辨識 (NER recognition) 等, 預建的統計式神經網路模型可對 17 種語言進行這些 NLP 任務;
- 斷詞 (tokenization) 方面, SpaCy 可在超過 65 種語言上讓使用者以自己的資料集訓練客制化的模型. 此外, SpaCy 也內建例如詞向量 (word vector) 等高級的 NLP 功能, 而 NLTK 則需要第三方套件例如 Gensim 來支援.

Hugging Face

AI Research Organization:

Hugging Face is an artificial intelligence technology company established in 2016. Their most well-known product is the Hugging Face Hub, which has the largest open-source machine learning model library in the world. Anyone can share their trained models in this library, promoting model sharing and collaboration.

Currently, it aggregates more than 320,000 pre-trained models.



Transformers Library

- **Framework Purpose:**
  - Hugging Face Transformers is designed to facilitate the use of pretrained models for various natural language processing (NLP) tasks, including text classification, language generation, and more.
- **Reduced Compute Costs:**
  - Using pretrained models from Hugging Face can significantly reduce the computational resources needed for training, as these models have already been trained on vast datasets.
- **Wide Adoption:**
  - The framework has gained popularity within the machine learning and NLP communities due to its user-friendly APIs and the availability of a wide range of pretrained models.

Hugging Face Hub

- Models
  - Provides a large number of pre-trained open source models, including natural language processing, computer vision, and motimodal models.
- Datasets
  - Provides datasets for various tasks, accrossing text, image and multimodal.
- Spaces
  - Offer a simple way to host ML demo apps directly on public cloud.

The contribution of Hugging Face

- Hugging Face's mission is to make artificial intelligence technology more accessible.
- Through their open-source framework and pre-trained models, they have significantly **lowered the barriers** to model training, enabling more people to easily apply artificial intelligence technology

Models

- Tasks
  - Multimodal, Computer Vision, Natural Language Processing, Audio, Tabular, Reinforce Learning
- Libraries
  - Pytorch, Tensorflow, Keras, etc.
- Datasets
  - Wikipedia, IMDB, etc.
- Languages
  - English, Chinese, French, etc.

Naming Rules

Organization

Base model

Version

facebook/bart-large-cnn

Repository

Datasets

- Tasks
  - Multimodal, Computer Vision, Natural Language Processing, Audio, Tabular, Reinforce Learning
- Sizes
- Sub-tasks
  - Name Entity Recognition, Sentiment Classification, etc.
- Languages
  - English, Chinese, French, etc.

Spaces

- Access to various trendy ML applications (mainly for demo purpose)
- Example: Llama, Comic Factory, Leaderboard
  - Prompt 1: I want to eat something healthy for my lunch. Do you have any suggestions?
  - Prompt 2: A bear running into a house.

How to Use the Models on the Hugging Face?

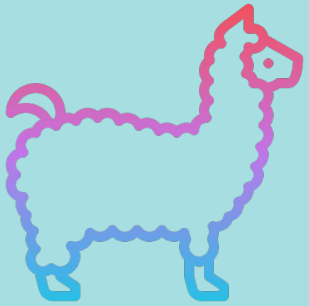
How to Call Models

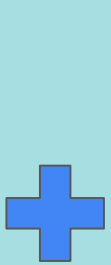
- Inference API
  - Send RESTful API to obtain the model output.
  - Need to get access tokens: Profiles (top-right corner) -> Settings -> Access tokens
- Transformer API
  - Download and host model on your local machine.
  - Pipeline library
- Spaces
  - Web UI deployed by others

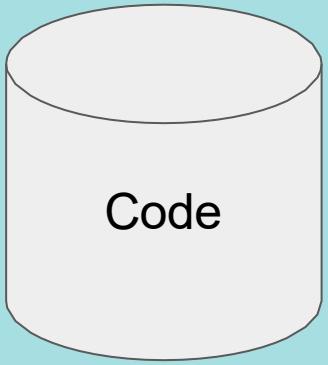
Use “Code Llama” to Learn

Code Llama

- Llama is an open large language model (like GPT)
- Code llama is the fine-tuned model based on llama using code datasets
- Powerful at generating codes







What Can I Do with Code Llama

Learning programming knowledge

What is the difference between stochastic gradient descent and batch gradient descent?

Answer explanation

Can you answer the following question?
“Why isn't a larger learning rate better in Gradient Descent?”

Example code

Can you write a function of gradient descent in Python?

Generate questions

I would like to practice my machine learning skills. Can you generate questions related to regression?

How to Utilize Hugging Face?

How to Utilize Hugging Face?

- When you face a task, find a model for that task on HF.
  - Find existing models for the task.

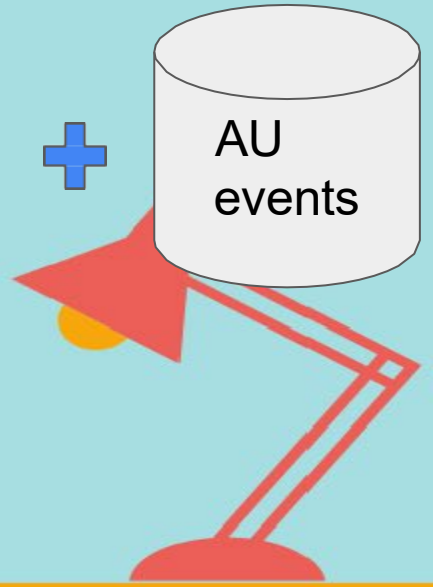
Task: summarization →  → facebook/bart-large-cnn

How to Utilize Hugging Face?

- When you face a task, find a model for that task on HF.
  - Find existing models for the task.
  - Fine-tune a model by yourself.

Task: summarize events in AU

facebook/bart-large-cnn



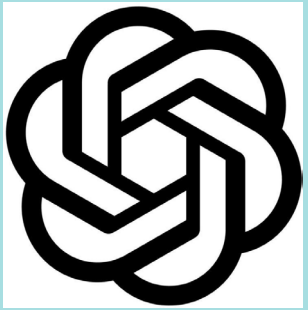
fine-tune

AUBart

How to Utilize Hugging Face?

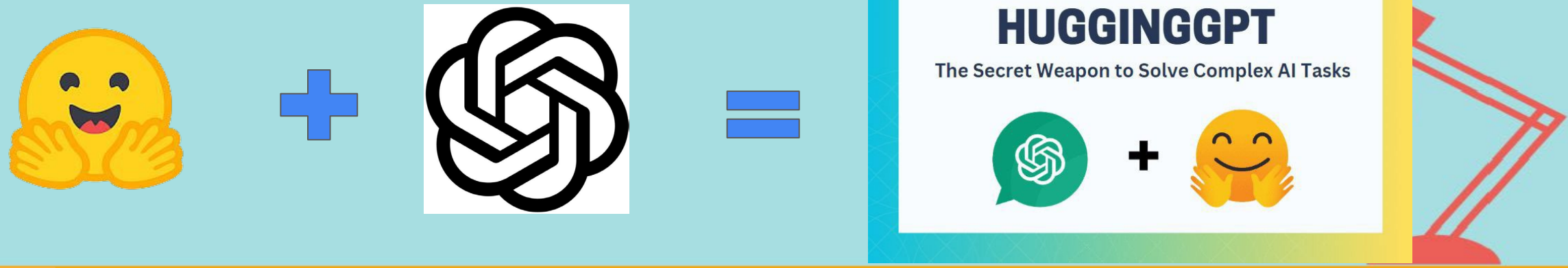
- When you face a task, find a model for that task on HF.
  - Find existing models for the task.
  - Fine-tune a model by yourself.
- What if your task involve multiple sub-tasks?
  - Summarize AU events in comic form.






How to Utilize Hugging Face?

- When you face a task, find a model for that task on HF.
  - Find existing models for the task.
  - Fine-tune a model by yourself.
- What if your task involve multiple sub-tasks?
  - Summarize AU events in comic form.



Thanks!

Q&A