```
! pip install datasets
! pip install transformers
! pip install rouge-score
! pip install nltk
# py7zr

import nltk
nltk.download("punkt")
```

```
Collecting datasets
  Downloading datasets-1.18.2-py3-none-any.whl (312 kB)
     |████████████████████████████████| 312 kB 5.3 MB/s
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (fro
Collecting aiohttp
  Downloading aiohttp-3.8.1-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manyli
     |████████████████████████████████| 1.1 MB 53.6 MB/s
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: pyarrow!=4.0.0,>=3.0.0 in /usr/local/lib/python3.7/di
Collecting fsspec[http]>=2021.05.0
  Downloading fsspec-2022.1.0-py3-none-any.whl (133 kB)
     |████████████████████████████████| 133 kB 16.9 MB/s
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: dill in /usr/local/lib/python3.7/dist-packages (from
Collecting xxhash
  Downloading xxhash-2.0.2-cp37-cp37m-manylinux2010_x86_64.whl (243 kB)
     |████████████████████████████████| 243 kB 48.0 MB/s
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: multiprocess in /usr/local/lib/python3.7/dist-package
Collecting huggingface-hub<1.0.0,>=0.1.0
  Downloading huggingface_hub-0.4.0-py3-none-any.whl (67 kB)
     |████████████████████████████████| 67 kB 2.9 MB/s
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-package
Collecting yarl<2.0,>=1.0
  Downloading yarl-1.7.2-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux
     |████████████████████████████████| 271 kB 56.3 MB/s
Collecting asynctest==0.13.0
  Downloading asynctest-0.13.0-py3-none-any.whl (26 kB)
Collecting aiosignal>=1.1.2
  Downloading aiosignal-1.2.0-py3-none-any.whl (8.2 kB)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-packag
Collecting async-timeout<5.0,>=4.0.0a3
  Downloading async_timeout-4.0.2-py3-none-any.whl (5.8 kB)
Collecting frozenlist>=1.1.1
  Downloading frozenlist-1.3.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.man
     |████████████████████████████████| 144 kB 72.6 MB/s
```

```
Collecting multidict<7.0,>=4.5
  Downloading multidict-6.0.2-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.
     |████████████████████████████████| 94 kB 3.4 MB/s
Requirement already satisfied: charset-normalizer<3.0,>=2.0 in /usr/local/lib/python
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/di
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (f
Installing collected packages: multidict, frozenlist, yarl, asynctest, async-timeout
Successfully installed aiohttp-3.8.1 aiosignal-1.2.0 async-timeout-4.0.2 asynctest-0
Collecting transformers
  Downloading transformers-4.16.2-py3-none-any.whl (3.5 MB)
```

```
import datasets
from transformers import AutoTokenizer, AutoModelForSeq2SeqLM, DataCollatorForSeq2Seq, Seq2Se
import nltk
import numpy as np

raw_dataset = datasets.load_dataset("cnn_dailymail", '3.0.0', split=["train[:15%]","test","va
metric = datasets.load_metric("rouge")

raw_datasets = datasets.DatasetDict({"train":raw_dataset[0],"test":raw_dataset[1], "validatic
print(raw_datasets)
model_checkpoint = "facebook/bart-base"
tokenizer = AutoTokenizer.from_pretrained(model_checkpoint)


max_input_length = 512
max_target_length = 128

def preprocess_function(examples):
    inputs = [doc for doc in examples["article"]]
    model_inputs = tokenizer(inputs, max_length=max_input_length, truncation=True)

    # Setup the tokenizer for targets
    with tokenizer.as_target_tokenizer():
        labels = tokenizer(examples["highlights"], max_length=max_target_length, truncation=T

    model_inputs["labels"] = labels["input_ids"]
    return model_inputs

tokenized_datasets = raw_datasets.map(preprocess_function, batched=True)

model = AutoModelForSeq2SeqLM.from_pretrained(model_checkpoint)

batch_size = 4
args = Seq2SeqTrainingArguments(
    "bart_summarization_cnn",
    evaluation_strategy = "epoch",
    learning_rate=2e-5,
    per_device_train_batch_size=batch_size,
```

```
        per_device_eval_batch_size=batch_size,
        gradient_accumulation_steps=2,
        weight_decay=0.01,
        save_total_limit=2,
        num_train_epochs=1,
        predict_with_generate=True,
        # fp16=True,
    )

    data_collator = DataCollatorForSeq2Seq(tokenizer, model=model)

    def compute_metrics(eval_pred):
        predictions, labels = eval_pred
        decoded_preds = tokenizer.batch_decode(predictions, skip_special_tokens=True)
        # Replace -100 in the labels as we can't decode them.
        labels = np.where(labels != -100, labels, tokenizer.pad_token_id)
        decoded_labels = tokenizer.batch_decode(labels, skip_special_tokens=True)

        # Rouge expects a newline after each sentence
        decoded_preds = ["\n".join(nltk.sent_tokenize(pred.strip())) for pred in decoded_preds]
        decoded_labels = ["\n".join(nltk.sent_tokenize(label.strip())) for label in decoded_label

        result = metric.compute(predictions=decoded_preds, references=decoded_labels, use_stemmer
        # Extract a few results
        result = {key: value.mid.fmeasure * 100 for key, value in result.items()}

        # Add mean generated length
        prediction_lens = [np.count_nonzero(pred != tokenizer.pad_token_id) for pred in predictio
        result["gen_len"] = np.mean(prediction_lens)

        return {k: round(v, 4) for k, v in result.items()}


    trainer = Seq2SeqTrainer(
        model,
        args,
        train_dataset=tokenized_datasets["train"],
        eval_dataset=tokenized_datasets["validation"],
        data_collator=data_collator,
        tokenizer=tokenizer,
        compute_metrics=compute_metrics
    )

    trainer.train()
    trainer.evaluate()
    trainer.save_model("/content/bart_summarization_cnn/model")
```

```
Downloading:                                    9.35k/? [00:00<00:00, 255kB/s]

Downloading:                                    9.50k/? [00:00<00:00, 274kB/s]

Downloading and preparing dataset cnn_dailymail/3.0.0 (download: 558.32 MiB, gener

100%                                            5/5 [00:24<00:00, 3.36s/it]

Downloading:                                    159M/159M [00:06<00:00,

100%                                            28.2MB/s]

Downloading:                                    376M/376M [00:12<00:00,

100%                                            45.3MB/s]

Downloading:                                    2.11M/? [00:00<00:00, 26.4MB/s]

Downloading:                                    46.4M/? [00:00<00:00, 80.5MB/s]

Downloading:                                    2.43M/? [00:00<00:00, 27.6MB/s]

100%                                            5/5 [01:25<00:00, 46.19s/it]

    286821/0 [00:52<00:00, 5761.59 examples/s]

    12781/0 [00:02<00:00, 5671.32 examples/s]

    11369/0 [00:02<00:00, 4826.77 examples/s]

Dataset cnn_dailymail downloaded and prepared to /root/.cache/huggingface/datasets

100%                                            3/3 [00:00<00:00, 56.89it/s]

Downloading:                                    5.61k/? [00:00<00:00, 159kB/s]

DatasetDict({
    train: Dataset({
        features: ['article', 'highlights', 'id'],
        num_rows: 43067
    })
    test: Dataset({
        features: ['article', 'highlights', 'id'],
        num_rows: 11490
    })
    validation: Dataset({
        features: ['article', 'highlights', 'id'],
        num_rows: 13368
    })
})

Downloading:                                    1.65k/1.65k [00:00<00:00,

100%                                            51.4kB/s]

Downloading:                                    878k/878k [00:00<00:00,

100%                                            3.09MB/s]

Downloading:                                    446k/446k [00:00<00:00,

100%                                            2.38MB/s]
```

```
Downloading:                                    1.29M/1.29M [00:00<00:00,
100%                                            3.37MB/s]
100%                                            44/44 [01:16<00:00, 1.26s/ba]
100%                                            12/12 [00:21<00:00, 1.60s/ba]
100%                                            14/14 [00:25<00:00, 1.53s/ba]
Downloading:                                    532M/532M [00:10<00:00,
100%                                            52.3MB/s]
The following columns in the training set  don't have a corresponding argument in
/usr/local/lib/python3.7/dist-packages/transformers/optimization.py:309: FutureWar
  FutureWarning,
***** Running training *****
  Num examples = 43067
  Num Epochs = 1
  Instantaneous batch size per device = 4
  Total train batch size (w. parallel, distributed & accumulation) = 8
  Gradient Accumulation steps = 2
  Total optimization steps = 5383
```

```
[5383/5383 1:25:45, Epoch 0/1]
```

Epoch	Training Loss	Validation Loss	Rouge1	Rouge2	Rougel	Rougelsum	Gen Len
0	2.166800	2.001869	23.669400	10.667800	19.127700	22.105700	20.000000

```
Saving model checkpoint to bart_summarization_cnn/checkpoint-500
Configuration saved in bart_summarization_cnn/checkpoint-500/config.json
Model weights saved in bart_summarization_cnn/checkpoint-500/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-500/tokenizer_con
Special tokens file saved in bart_summarization_cnn/checkpoint-500/special_tokens_
Saving model checkpoint to bart_summarization_cnn/checkpoint-1000
Configuration saved in bart_summarization_cnn/checkpoint-1000/config.json
Model weights saved in bart_summarization_cnn/checkpoint-1000/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-1000/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-1000/special_tokens
Saving model checkpoint to bart_summarization_cnn/checkpoint-1500
Configuration saved in bart_summarization_cnn/checkpoint-1500/config.json
Model weights saved in bart_summarization_cnn/checkpoint-1500/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-1500/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-1500/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-500] due to args.save
Saving model checkpoint to bart_summarization_cnn/checkpoint-2000
Configuration saved in bart_summarization_cnn/checkpoint-2000/config.json
Model weights saved in bart_summarization_cnn/checkpoint-2000/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-2000/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-2000/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-1000] due to args.sav
Saving model checkpoint to bart_summarization_cnn/checkpoint-2500
Configuration saved in bart_summarization_cnn/checkpoint-2500/config.json
Model weights saved in bart_summarization_cnn/checkpoint-2500/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-2500/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-2500/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-1500] due to args.sav
Saving model checkpoint to bart_summarization_cnn/checkpoint-3000
```

```
Saving model checkpoint to bart_summarization_cnn/checkpoint-3000
Configuration saved in bart_summarization_cnn/checkpoint-3000/config.json
Model weights saved in bart_summarization_cnn/checkpoint-3000/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-3000/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-3000/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-2000] due to args.sav
Saving model checkpoint to bart_summarization_cnn/checkpoint-3500
Configuration saved in bart_summarization_cnn/checkpoint-3500/config.json
Model weights saved in bart_summarization_cnn/checkpoint-3500/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-3500/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-3500/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-2500] due to args.sav
Saving model checkpoint to bart_summarization_cnn/checkpoint-4000
Configuration saved in bart_summarization_cnn/checkpoint-4000/config.json
Model weights saved in bart_summarization_cnn/checkpoint-4000/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-4000/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-4000/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-3000] due to args.sav
Saving model checkpoint to bart_summarization_cnn/checkpoint-4500
Configuration saved in bart_summarization_cnn/checkpoint-4500/config.json
Model weights saved in bart_summarization_cnn/checkpoint-4500/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-4500/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-4500/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-3500] due to args.sav
Saving model checkpoint to bart_summarization_cnn/checkpoint-5000
Configuration saved in bart_summarization_cnn/checkpoint-5000/config.json
Model weights saved in bart_summarization_cnn/checkpoint-5000/pytorch_model.bin
tokenizer config file saved in bart_summarization_cnn/checkpoint-5000/tokenizer_co
Special tokens file saved in bart_summarization_cnn/checkpoint-5000/special_tokens
Deleting older checkpoint [bart_summarization_cnn/checkpoint-4000] due to args.sav
The following columns in the evaluation set  don't have a corresponding argument i
***** Running Evaluation *****
  Num examples = 13368
  Batch size = 4
```

```
from transformers import AutoTokenizer, AutoModelForSeq2SeqLM

tokenizer_pre = AutoTokenizer.from_pretrained("/content/bart_summarization_cnn/model")
model_pre = AutoModelForSeq2SeqLM.from_pretrained("/content/bart_summarization_cnn/model")

def summarize(text, max_length):
  input_ids = tokenizer_pre.encode(text, return_tensors="pt", add_special_tokens=True)

  generated_ids = model_pre.generate(input_ids=input_ids, num_beams=2, min_length=500, max_le

  preds = [tokenizer_pre.decode(g, skip_special_tokens=True, clean_up_tokenization_spaces=Tru

  return preds


text = """
What is a Computer Network?

Computer Network is a group of computers connected with each other through wires, optical fib
The aim of the computer network is the sharing of resources among various devices.
```

```
In the case of computer network technology, there are several types of networks that vary fro

Components Of Computer Network:

Major components of a computer network are:
NIC(National interface card)
NIC is a device that helps the computer to communicate with another device. The network inter
There are two types of NIC: wireless NIC and wired NIC.

Wireless NIC: All the modern laptops use the wireless NIC. In Wireless NIC, a connection is m
Wired NIC: Cables use the wired NIC to transfer the data over the medium.


Hub
Hub is a central device that splits the network connection into multiple devices. When comput

Switches
Switch is a networking device that groups all the devices over the network to transfer the da

Cables and connectors
Cable is a transmission media that transmits the communication signals. There are three types

Twisted pair cable: It is a high-speed cable that transmits the data over 1Gbps or more.
Coaxial cable: Coaxial cable resembles like a TV installation cable. Coaxial cable is more ex
Fibre optic cable: Fibre optic cable is a high-speed cable that transmits the data using ligh


Router
Router is a device that connects the LAN to the internet. The router is mainly used to connec

Modem
Modem connects the computer to the internet over the existing telephone line. A modem is not

Uses Of Computer Network

Resource sharing: Resource sharing is the sharing of resources such as programs, printers, ar
Server-Client model: Computer networking is used in the server-client model. A server is a ce
Communication medium: Computer network behaves as a communication medium among the users. For
E-commerce: Computer network is also important in businesses. We can do the business over the

"""
# print(text)

new = " ".join( text.splitlines())

print(summarize(new, 1000))
```


```
       "encoder_layers": 6,
       "eos_token_id": 2,
       "forced_eos_token_id": 2,
       "gradient_checkpointing": false,
       "id2label": {
         "0": "LABEL 0",
```

```
    "1": "LABEL_1",
    "2": "LABEL_2"
  },
  "init_std": 0.02,
  "is_encoder_decoder": true,
  "label2id": {
    "LABEL_0": 0,
    "LABEL_1": 1,
    "LABEL_2": 2
  },
  "max_position_embeddings": 1024,
  "model_type": "bart",
  "no_repeat_ngram_size": 3,
  "normalize_before": false,
  "normalize_embedding": true,
  "num_beams": 4,
  "num_hidden_layers": 6,
  "pad_token_id": 1,
  "scale_embedding": false,
  "task_specific_params": {
    "summarization": {
      "length_penalty": 1.0,
      "max_length": 128,
      "min_length": 12,
      "num_beams": 4
    },
    "summarization_cnn": {
      "length_penalty": 2.0,
      "max_length": 142,
      "min_length": 56,
      "num_beams": 4
    },
    "summarization_xsum": {
      "length_penalty": 1.0,
      "max_length": 62,
      "min_length": 11,
      "num_beams": 6
    }
  },
  "torch_dtype": "float32",
  "transformers_version": "4.16.2",
  "use_cache": true,
  "vocab_size": 50265
}

loading weights file /content/bart_summarization_cnn/model/pytorch_model.bin
All model checkpoint weights were used when initializing BartForConditionalGeneratic

All the weights of BartForConditionalGeneration were initialized from the model chec
If your task is similar to the task the model of the checkpoint was trained on, you
['Computer Network is a group of computers connected with each other through wires,
```

✓ 44s completed at 6:02 PM