Week 4 Practical Exercise of Lesson 4: Baseline Machine Translation - Translate
Evaluate [60 Mins]

Your task

In this hands-on exercise, you have to use Hugging Face translation pipeline to translate a small English→French parallel corpus. Then you have to measure inference speed and compute a baseline BLEU score for later comparison.

1 Environment Setup

```
pip install transformers torch sacrebleu datasets time

import time
import sacrebleu
from datasets import load_dataset
from transformers import pipeline
```

2 Load the Translation Hugging Face Pipeline

```
translator = pipeline("translation_en_to_fr",model="Helsinki-NLP/opus-mt-en-fr")
```

3 Prepare a Small Parallel Test Set

Use the datasets library to load a public small test split, or define your own list:
# Example: WMT16 English-French test

```
dataset = load_dataset("wmt16", "ro-en", split="test[:100]") # adjust to en-fr if available

# If no en-fr, define custom:

pairs = [

{"en": "The weather is nice today.", "fr": "Il fait beau aujourd'hui."},

{"en": "I love reading science fiction.", "fr": "J'adore lire de la science-fiction."},

# ... add 20 more

]
```

4 Batch Translation & Latency Measurement

Run the translation pipeline in batches of 4 and time the translations:

```
# Timed run
start = time.perf_counter()

translations= translator([p["en"] for p in pairs],batch_size=4)

elapsed = time.perf_counter() - start

print(f"Avg latency:{elapsed/len(pairs):.3f}s per sentence")
```

5 Compute Baseline BLEU Score

```
refs = [[p["fr"] for p in pairs]]# list-of-list for sacrebleu

sys = [t for t in translations]

bleu = sacrebleu.corpus_bleu(sys, refs)

print(f"Baseline BLEU: {bleu.score:.2f}")
```

6 Save Outputs for Analysis (5 min)

Write a JSONL file with source, reference, and prediction:

```
import json

with open("mt_baseline.jsonl", "w", encoding="utf-8") as f:

    for src, ref, pred in zip([p["en"] for p in pairs], [p["fr"] for p in pairs], translations):

record = {"source": src, "reference": ref, "prediction": pred}

f.write(json.dumps(record, ensure_ascii=False) + "textbackslash{}n")
```

7 Reflection

In a markdown cell or at the end of your script, answer:

- Which sentences scored poorly (review BLEU details)?
- What types of translation errors did you observe (e.g., word order, missing articles)?

Deliverable: A notebook or script performing all steps above, printed latency and BLEU results, and a brief reflection summarizing translation quality and next steps.