CSC401 Homework Assignment #2
Analysis

Kai Wang
Student number: 1005422213
UTORid: wangka76
kaixxx.wang@mail.utoronto.ca

1 Training Results

1.1 Training Loop Printout

Place your training loop printout in this section. The numbers shown are for illustration purposes only and the loss and BLEU values presented are fabricated.

Model with Pre-layer Normalization

```
[Device:cuda] Epoch 1 Training ====
[Device:cuda] Epoch 1 Validation ====
Epoch 1: loss=6.352157288847405, BLEU: skipped until epoch 4, time=00:01:43
[Device:cuda] Epoch 2 Training ====
[Device:cuda] Epoch 2 Validation ====
Epoch 2: loss=2.8295345301729147, BLEU: skipped until epoch 4, time=00:03:28
[Device:cuda] Epoch 3 Training ====
[Device:cuda] Epoch 3 Validation ====
Epoch 3: loss=1.9428371553292532, BLEU: skipped until epoch 4, time=00:05:14
[Device:cuda] Epoch 4 Training ====
[Device:cuda] Epoch 4 Validation ====
Epoch 4: loss=1.5872223453895664, BLEU-4: 35.7068 BLEU-3: 42.7892, time=00:07:25
[Device:cuda] Epoch 5 Training ====
[Device:cuda] Epoch 5 Validation ====
Epoch 5: loss=1.3464618606371903, BLEU-4: 36.5916 BLEU-3: 43.6124, time=00:09:33
Finished 5 epochs
...
```

Model with Post-layer Normalization [Device:cuda] Epoch 1 Training ==== [Device:cuda] Epoch 1 Validation ==== Epoch 1: loss=6.349785398607812, BLEU: skipped until epoch 4, time=00:01:41 [Device:cuda] Epoch 2 Training ==== [Device:cuda] Epoch 2 Validation ==== Epoch 2: loss=2.7990091258839276, BLEU: skipped until epoch 4, time=00:03:25 [Device:cuda] Epoch 3 Training ==== [Device:cuda] Epoch 3 Validation ==== Epoch 3: loss=1.9596696510149114, BLEU: skipped until epoch 4, time=00:05:08 [Device:cuda] Epoch 4 Training ==== [Device:cuda] Epoch 4 Validation ==== Epoch 4: loss=1.626192694199574, BLEU-4: 35.5586 BLEU-3: 42.6348, time=00:07:15 [Device:cuda] Epoch 5 Training ==== [Device:cuda] Epoch 5 Validation ==== Epoch 5: loss=1.397774521276897, BLEU-4: 36.3736 BLEU-3: 43.3222, time=00:09:22 Finished 5 epochs

1.2 Test Set BLEU Score

This section lists the test set BLEU score reported on the test set for each model in table 1.

Model	BLEU-4	BLEU-3
Model Pre-layer Normalization	41.82	49.23
Model Post-layer Normalization	41.78	49.16

Table 1: The BLEU score reported on the test set for each model.

Is there a difference between BLEU-3 and BLEU-4? What do you think could be the reason behind the differences?

BLEU-4 measures longer sequences (4-grams), making it harder to achieve high scores compared to BLEU-3, which focuses on shorter sequences (3-grams). The difference suggests that the model struggles more with accurately generating longer, coherent phrases.

2 Translation Analysis

2.1 Translations

List all of your translations in this section.

My Model

»> translate("Trudeau et Colbert se lient autour du statut partage de gars qui etaient cool il y a dix ans")
'<s> trudeau would be subjected to the sharing of the marie bq'

»> translate("Un collegue que tout le monde deteste est surpris de ne pas pouvoir convaincre ses collegues de faire ce qu'il veut")
'<s> a fellow hate to be surprised by his colleagues'

»> translate("Jagmeet Singh affirme son independance en faisant exactement ce que Pierre Poilievre lui a dit de faire")
'<s> immigration is saying exactly what he said'

»> translate("Les conservateurs promettent que s'ils sont elus, vos parents se reuniront, votre emission de television preferee ne sera pas annulee et McDonald's ramenera la pizza")
'<s> the tories were elected to your parents will not be coming from'

»> translate("Trudeau rend une visite surprise a Ottawa")
'<s> trudeau wrote a surprise to ottawa'

»> translate("Le Canada souffre d'une crise de productivite, affirme que le gouvernement ne fait rien")
'<s> canada suffers from productivity is saying that the government is doing nothing'

»> translate("Le gouvernement federal annonce que l'O Canada commencera desormais avec 15 secondes de publicites incontournables")
'<s> the federal government announced that canada will now begin with seconds'

»> translate("Tout ce que les conservateurs epris de liberte ont interdit aux Canadiens de faire ces dernieres annees")
'<s> all the tories have banned freedom to canadian horses in the last few years'

T5 MT

Trudeau and Colbert bond over their shared status as guys who were cool ten years ago.

A colleague whom everyone hates is surprised that he can't convince his coworkers to do what he wants.

Jagmeet Singh asserts his independence by doing exactly what Pierre Poilievre told him to do.

The Conservatives promise that if elected, your parents will reunite, your favorite TV show won't be canceled, and McDonald's will bring back pizza.

Trudeau makes a surprise visit to Ottawa.

Canada is suffering from a productivity crisis, and it is claimed that the government is doing nothing.

The federal government announces that O Canada will now start with 15 seconds of must-watch commercials.

Everything that freedom-loving Conservatives have forbidden Canadians from doing in recent years.

Google

Trudeau and Colbert bond over shared status as guys who were cool a decade ago

A colleague that everyone hates is surprised not to be able to convince his colleagues to do what he wants

Jagmeet Singh asserts his independence by doing exactly what Pierre Poilievre told him to do

Conservatives promise that if elected, your parents will reunite, your favorite TV show won't be canceled and McDonald's will bring back pizza

Trudeau makes a surprise visit to Ottawa

Canada suffers from productivity crisis, says government is doing nothing

The federal government announces that O Canada will now begin with 15 seconds of must-see commercials

Everything freedom-loving Conservatives have forbidden Canadians from doing in recent years

2.2 Discussion

In this section, write a brief discussion on your findings. Describe the quality of those sentences. How's your model compared with Google Translate or ChatGPT?

Google's translation model provides the best quality, accurately capturing both meaning and structure. My model performs the worst, often failing to convey the intended meaning.

The differences in quality are likely due to the amount of pre-training data, the model architecture, and fine-tuning. Google benefits from vast, high-quality datasets and advanced optimization.

My model handles simple, short sentences better, translating phrases like "Trudeau rend une visite surprise à Ottawa" more accurately. It struggles significantly with complex, idiomatic, or sarcastic sentences, leading to errors in meaning. The model seems to perform poorly with nuanced language or when context is crucial.

Google is more consistent, but still can face challenges with idioms or sarcasm. The pattern persists but is less severe with these advanced models, highlighting the importance of extensive pre-training and task-specific fine-tuning.