News Summarization: Mandate-3

R Prasannavenkatesh
IMT2019063
IIIT Bangalore
prasannavenkatesh.ramkumar@iiitb.ac.in

M Mani Nandadeep
IMT2019051
IIIT Bangalore
mani.nandadeep@iiitb.ac.in

***Abstract*—The goal of this project is to build a text summarizer and use it to summarize the news articles. People are looking for shortcut methods to learn ideas in less time. Text summarizers are also helping people to decide whether a book, a research paper, or an article is worth reading or not. A text summarizer is a NLP based tool that wraps up text to a short length. It condenses a long article to main points. The need for text summarizers are increasing in everyday life since people want news quick and to the point rather than long and shabby walls of text. Social media sites like reddit also use a text summarizer called autotldr which is a bot that uses SMMRY to automatically summarize long reddit submissions.**

I. Pytorch Lightning

In this mandate submission we have implemented the Text summarizer with the help of T5 transformer and Pytorch Lightning.

PyTorch Lighting is a lightweight PyTorch wrapper for high-performance AI research that aims to abstract Deep Learning boilerplate while providing you full control and flexibility over your code.

A. Dataset

All datasets that represent a map from keys to data samples should subclass it.

- **getitem**
  All subclasses should overwrite __getitem__(), supporting fetching a data sample for a given key.
- **len**
  Subclasses could also optionally overwrite __len__(), which is expected to return the size of the dataset by many Sampler implementations and the default options of DataLoader.

B. LightningDataModule

To define a DataModule the following methods are used to create train/val/test/predict dataloaders:

- **prepare_data**
  Downloading and saving data with multiple processes (distributed settings) will result in corrupted data. Lightning ensures the **prepare_data()** is called only within a single process, so you can safely add your downloading logic within.
- **setup**
  There are also data operations you might want to perform on every GPU. Use setup() to do things like:
  - count number of classes
  - create datasets
  - perform train/val/test splits
  - etc...
- **train_dataloader**
  Use the train_dataloader() method to generate the training dataloader(s). Usually you just wrap the dataset you defined in setup. This is the dataloader that the Trainer fit() method uses.
- **val_dataloader**
  Use the val_dataloader() method to generate the validation dataloader(s). Usually you just wrap the dataset you defined in setup. This is the dataloader that the Trainer fit() and validate() methods uses.
- **test_dataloader**
  Use the test_dataloader() method to generate the test dataloader(s). Usually you just wrap the dataset you defined in setup. This is the dataloader that the Trainer test() method uses.

II. Model

- Model used: t5-base
- Optimizer used: AdamW
- Learning Rate used: $1e-4$
- Batch size used: 8
- Maximum Text Token Length used: 512
- Maximum Summary Token Length used: 128

III. Rouge score

ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It is essentially a set of metrics for evaluating automatic summarization of texts as well as machine translations.

A. Rouge-1

ROUGE-1 refers to overlap of unigrams between the system summary and reference summary.

B. Rouge-2

ROUGE-2 refers to the overlap of bigrams between the system and reference summaries.

C. Rouge-l

measures longest matching sequence of words using LCS. An advantage of using LCS is that it does not require consecutive matches but in-sequence matches that reflect sentence level word order. Since it automatically includes longest in-sequence common n-grams, you don't need a predefined n-gram length.

D. Recall

The recall counts the number of overlapping n-grams found in both the model output and reference, then divides this number by the total number of n-grams in the reference.

$$RECALL = \frac{no\_of\_overlapping\_words}{total\_words\_in\_reference\_summary}$$

E. Precision

The precision counts the number of overlapping n-grams found in both the model output and reference, then divides this number by the total number of n-grams in the system summary.

$$PRECISION = \frac{no\_of\_overlapping\_words}{total\_words\_in\_system\_summary}$$

F. F1 Score

$$F1Score = \frac{2 * precision * recall}{precision + recall}$$

A python lightning trainer with number of Epochs 1 and accelerator "gpu" ran for around 15 minutes (wall clock time).
The F1 score was approximately 40-45%.
A python lightning trainer with number of Epochs 2 and accelerator "gpu" ran for around 30 minutes (wall clock time).
The F1 score was approximately 50 - 60%.

IV. Summarization Results

Sample Text:

'Days after Tesla CEO Elon Musk said that artificial intelligence (AI) was the biggest risk, Facebook has shut down one of its AI systems after chatbots started speaking in their own language defying the codes provided.According to a report in Tech Times on Sunday, the social media giant had to pull the plug on the AI system that its researchers were working on ?because things got out of hand?. ?The AI did not start shutting down computers worldwide or something of the sort, but it stopped using English and started using a language that it created,? the report noted.Initially the AI agents used English to converse with each other but they later created a new language that only AI systems could understand, thus, defying their purpose.This led Facebook researchers to shut down the AI systems and then force them to speak to each other only in English.

⋮

Elon Musk, philanthropist Bill Gates and ex-Apple founder Steve Wozniak have also expressed their concerns about where the AI technology was heading.Interestingly, this incident took place just days after a verbal spat between Facebook CEO and Musk who exchanged harsh words over a debate on the future of AI.?I?ve talked to Mark about this (AI). His understanding of the subject is limited,? Musk tweeted last week. The tweet came after Zuckerberg, during a Facebook livestream earlier this month, castigated Musk for arguing that care and regulation was needed to safeguard the future if AI becomes mainstream.?I think people who are naysayers and try to drum up these doomsday scenarios – I just, I don?t understand it. It?s really negative and in some ways I actually think it is pretty irresponsible,? Zuckerberg said.Musk has been speaking frequently on AI and has called its progress the ?biggest risk we face as a civilisation?.?AI is a rare case where we need to be proactive in regulation instead of reactive because if we?re reactive in AI regulation it?s too late,? he said.'

Reference Summary:

'Facebook has shut down one of its Artificial Intelligence systems after chatbots programmed to converse in English started speaking in their own language. Using machine learning algorithms, the bots were allowed to converse freely to strengthen their conversational skills. However, they began to deviate from the scripted norms and started speaking a new language created without human input. '

Predicted Summary:

'Facebook has shut down one of its AI systems after chatbots started speaking in their own language defying the codes provided. The AI agents used English to converse with each other but later created a new language that only AI systems could understand, thus, defying their purpose. This led Facebook researchers to shut down the AI systems and then force them to speak to each other only in English.'

V. Outcomes mapping to mandate contributions

- CO1: **
- CO2: **
- CO3: ****
- CO4: ***
- CO5: ****
- C06: ***

VI. References

- Documentation of T5
- Simple abstractive text summarization with pretrained T5
- ROUGE SCORE
- Lightning Module