Lamiah Khan 12/01/2024
Final Project: Story Generation

I. Project Description

In this project, pretrained-transformer models were utilized from Hugging Face to create a story generator. The datasets, which were imported from Kaggle, included experimentation within three different writing styles: historical classics, modern gothic folktales, and children's textbooks. The three datasets featured stories from various regions, including Russia, Japan, Scandinavia, etc. Overall, the datasets are combined into the text-file for this project, named `storyTextData.txt`. The dataset also contained a variety of authors, including Willian Shakespeare, Jane Austen, Phoebe Cary, Issac Watts, etc..

II. Experimentation

The project explored two transformer architectures, GPT-2 and T5, to fine-tune them for optimal story generation. Key hyperparameters investigated included top_p and temperature. It was observed that increasing the temperature above 0.9 led to outputs that were increasingly incoherent and random. In contrast, experimenting with lower top_p values (e.g., 0.3 to 0.6) resulted in generated stories that were notably repetitive. Additionally, dataset truncation was employed, limiting the dataset to 20,000 examples in the final model. This experimentation revealed a clear trend: larger datasets tend to enhance generalization. However, including shorter stories often caused increased repetitiveness in outputs, prompting the implementation of a minimum story length threshold of 600 characters. For training, a batch size of 4, two epochs, and a learning rate of 2e−5 were employed for both models. For GPT-2 specifically, smaller batch sizes were found to stabilize gradient updates, resulting in reduced loss. Limiting training to two epochs effectively mitigated overfitting. Furthermore, the choice of a lower learning rate contributed to improved performance in terms of quantitative evaluation metrics. Lastly, for GPT-2, setting a max-length of 200-300 value to each generated prompt also decreased repetitiveness, randomness, and provided more structure to each story output.

For GPT-2, I evaluated the model's performance using quantitative metrics such as BLEU and ROUGE. Although BLEU is primarily designed for machine translation and is not ideally suited for assessing story generation—given its focus on literal accuracy rather than creativity—it provides a measure of how closely the generated text aligns with the reference text at the word level. My model achieved a BLEU score of approximately 14%. While this score does not directly reflect the success of the model in generating compelling stories, I interpreted it as indicative of greater variability and creativity in the output. However, BLEU's emphasis on precision makes it less relevant for evaluating story generation, where creativity and narrative coherence are more critical.

Conversely, ROUGE is widely regarded as a more appropriate metric for story generation, as it prioritizes recall by focusing on capturing key information and narrative details. Since effective storytelling often hinges on the accurate representation of essential plot points and narrative structure, ROUGE is better suited for assessing whether the generated story preserves the core elements of the reference story, even when the wording differs. ROUGE's tolerance for linguistic variation further enhances its applicability to creative text generation. For this project, I employed the n-ROUGE method described in the referenced literature and observed significantly higher performance compared to BLEU,

with a mean ROUGE score of approximately 73%. The ROUGE metric was also employed to evaluate T5's performance, yielding a score of approximately 33%. These quantitative assessments closely align with the qualitative evaluation of the stories generated by the two models.

III. Results

Overall, the results indicate that GPT-2 outperforms T5 in story generation when evaluated using similar hyperparameters and training metrics. As demonstrated, GPT-2 produced narratives with a more coherent structure, greater variation aligned with the genre of the given prompt, and reduced repetition compared to T5.

Stories Generated by GPT-2:

```
print(tell_a_tale("Hannah went into a scary haunted house, where the floors were", model2, max_length=280))
```

```
Hannah went into a scary haunted house, where the floors were all covered with all sorts of little things.  She thought she had seen al
l kinds of things, but she could not tell if any were real.  Her curiosity grew and grew until she came to a very dark place.  She saw
a large tree with a large branch hanging from the trunk.  She climbed up and looked down.  There was nothing but black and white and ve
ry dark things in the wood.  She was not surprised to find a lot of blood, but she did not know if it was real.  It was a strange kind
of blood.  Hannah had never seen anything like it before. "I should like to know where it is," said Hannah.  "I've seen it a long tim
e." Hannah felt very anxious, and looked at the branch she was climbing up.  She wanted to tell the story. She could not tell how she h
ad come to the tree, but she felt very
```

```
print(tell_a_tale("How art thy on this morning", model2, max_length=200))
```

```
How art thy on this morning? Thou art so very beautiful!” cried the goddess; and s
he saw her husband's face with astonishment. “Ah, thou art so beautiful,” she crie
d, “that thy wife’s eyes will not rest upon thee.” “Then,” answered the goddess, “
the sun, too, shall be shining on thee,” and she began to weep, “that I have been
so long away from thee, and that thou hast not come to me.” But when she had cried
thus, she drew her hands out of the ground, and began to cry again, “Ah, thou art
so beautiful,” and she began to weep again, “that I have been so long away from th
ee, and that thou hast not come to me.” But when she had cried thus, she drew her
hands out of the ground, and began to cry again,
```

```
print(complete_prompt("It's 1897, and huge diamonds are discovered in ", model2, max_length=200))
```

```
It’s 1897, and huge diamonds are discovered in          The original stone was a piece of          Gold that had been quarried at the
quarries in the Middle East. It was found         and was very much admired. It was very valuable. The gold pieces were called gold in Indi
a,          and the Indian       Gold was valued at an           Thousand-year-old       Gold. It was so precious that         the     King
of          He had a goldmine, and he said he wanted to find it. He did. It is said that the gold in the mine was
```

```
print(tell_a_tale("One upon a time, in a distant", model2, max_length=350))
```

```
One upon a time, in a distant land, there lived a little man called Aylor. He lived a little in a forest, where no trees could grow, an
d no wood was needed. When he was a boy he went to play in a field, and had the best of everything in the world. The trees had all thei
r leaves cut off, but he did not care about them. He said to himself, "If I could cut my own leaves out of the forest, what would I hav
e to do with it?" He made up his mind that he would not care what the people said about him, for he knew that he was a pretty boy, and
that his father would be very happy to see him, for he would be the happiest man in the world. So Aylor went out into the fields and be
gan to cut his own leaves. One day he was sitting on a branch and his father said to him, "I will cut your leaves out of the tree, and
you will be my husband, and you shall have a wife of your own." Aylor said, "Well, then," said his father, "I will go and take care of
my own child, and you shall have a wife of your own." A little boy was lying on the ground in the middle of the field. He looked up and
saw his father, sitting by the window, and he said, "My father, my son, my dear, what can you do for me?" "My father," said Aylor, "I w
ill take care of you, and you shall have a wife of your own." "Well, then," said his father, "I will take care of you, and you shall ha
ve
```

Stories Generated by T5:

```
# Example prompt for T5 model
prompt = "The girl sat in the snow and threw "
generated_story = generate_story(prompt, model, tokenizer)
print(generated_story)
```

```
The girl sat in the snow and threw a threw a threw a threw a threw a threw a threw a threw a threw a threw a threw a threw a threw a th
rew a threw a threw a threw a threw a threw a threw a threw
```

```
# Example prompt for T5 model
prompt = "It's 1897, and huge diamonds are discovered in "
generated_story = generate_story(prompt, model, tokenizer)
print(generated_story)
```

```
It's 1897, and huge diamonds are discovered in the Middle Ages. It's 1897, and huge diamonds are discovered in the Middle Ages in 1897,
and huge diamonds are discovered in the Middle Ages.
```

```
# Example prompt for T5 model
prompt = "Once upon a time, in a distant land there lived  "
generated_story = generate_story(prompt, model, tokenizer)
print(generated_story)
```

Notebook editor cells

```
Einmal upon a time, in a distant land there lived a time, in a distant land there lived a time, in a distant land there lived a time, u
pon a time, in a distant land there lived a time, upon a time, in a distant land there lived a time, upon a time, in a distant land the
re lived a time, upon a time, in a distant land there lived a time, upon a time, in a distant
```

IV. Additional Information

Due to GPT-2's inherent variability in text generation, it is recommended to run the same prompt multiple times to obtain a selection of stories, as each execution generates a unique output. Prompts should be specific and clearly convey the desired style and tone to guide the model effectively in producing stories aligned with the intended narrative framework. Additional details on the hyperparameters employed, as well as the BLEU and ROUGE metrics and alternative references, are thoroughly documented within the accompanying Google Colab notebook.