Homework 4

Spencer Au

Introduction

We are creating a model that can generate text based on Aldous Huxley's *Brave New World*. The model is a Recurrent Neural Network (RNN) that uses Long Short-Term Memory (LSTM) cells, and the the fine tuning model I used is a GPT-2 model. The goal is to generate 10 sequences of text for each model that is similar to the text in *Brave New World*.

Analysis

There is not much EDA that can be done for a txt version of *Brave New World*. Some basic EDA indcates that there are 64608 words in the text, and the 10 most commons words are "said" at 315 times, "one" at 263 times, "bernard" at 205 times, "savage" at 189 times, "like" at 186 times, "lenina" at 179 times, "dont" at 119 times, "made" at 111 times, "little" at 111 times, and "time" at 109 times. In terms of preprocessing for the first model, I load the text from a file and converts all characters to lowercase, map each unique character to an integer, create sequences of 100 characters as inputs with the next character as the output, and then convert these to integers based on the earlier mapping. I then reshape and normalize the inputs while the outputs are one-hot encoded to classify each character precisely and feed that into the model. Basically, the model learns from these bundles to get really good at guessing what letter is likely to come next after any given sequence of 100 letters. For the GPT-2 model, I used the GPT2Tokenizer in order to pre-process the model.

Methods

In terms of the LSTM model, I used a 2 layer LSTM model with 256 hidden units, following by a dropout layer of 0.2. An LSTM acts like a highly efficient memory system for the computer, each with 256 units that help retain information, and added a dropped 20% of the data to ensure the model doesn't over-rely on certain parts of the data. I used a batch size of 128 and a sequence length of 100. This means that the model processes 128 sequences at a time,

each sequence being 100 characters long, to better learn from the text. I then used the Adam optimizer and the CrossEntropyLoss function, and trained the model for 150 epochs. This means that it's fine-tuned with the Adam optimizer for making smart adjustments and uses CrossEntropyLoss to measure errors, going through the training material 150 times to enhance accuracy. We then generate 10 sequences of text based on the trained RNN LSTM model. In terms of the GPT-2 model, I fine-tuned the model using the GPT2LMHeadModel pre-trained model. I used a batch size of 2 on both the training and evaluation data, and trained the model for 100 epochs. I then generated 10 sequences of text based on a list of 10 prompts obtained from ChatGPT in order to evaluate the model.

Results

LSTM Model Sequences

Sequence 1: " irst used officially in a.f. 214. why not before? two reasons, (a) …" "these early experimenter," said the savage, with a san and a perfect ce- cal, they had so hand and surned away and was some

Sequence 2: and bacteriological conditioning of the embryo. practical instructions for beta embryo-store work and the savage was a mong chain, the sruare was a prison of the south of the souther- the savage

Sequence 3: . linda and he-linda was his mother (the word made lenina look uncomfortable)- were strangers in the seventeen thousand as the savage was all all the savage reserval the savage round his hand and th

Sequence 4: " dictions. the land wasn't properly worked; there were strikes in all the factories; the laws were seened to himself and shouted into the changed and startled and she had been the sen much and sent

Sequence 5: " ir hung down in grey wisps round his face. his body was bent and emaciated to the bone, almost fleshed and still said and the last of the south of the soof of the south of the sooi and a half and s

Sequence 6: " state about having a girl-it seemed rather ridiculous. but, taken detail by verbal detail, what a su- pere pueer was a mong alons and the savage rooe and surnee off. "but it sane to have the same

Sequence 7: " hought; and then (with what derisive ferocity!): "sons eso tse-na." and he spat on the ground, as pnee and surong and suill srickled in the south of the south of the sooe of the south- the savage w

Sequence 8: " the women to whom he made proposals, the practical joking of his equals among the men. the mockery of the controller she was a sall to the mortin of the south of the sooa of the room of the reserva

Sequence 9: ” eyes. “my young friend,” said the arch-community-songster in a tone of loud and solemn severity; but the sears of the controller she was on the forreritime that had been a cettain so and all the

Sequence 10: ” s in some indescribably delicious all-singing feely; where the dripping patchouli was more than sctial of the souther- the savage was and sureaming and surned of the room and staring and surned and

GPT-2 Model Sequences

Generated 1: Describe your vision of a perfect society. How would it function day to day?

A. Day-to-day living. The overalls of all living things were made of various materials, such as glass, rubber, and so on. Each of them had its own unique and unique shape. For example, a man’s mouth had a rubber mouth, while a certain color of rubber covered the rest of the body. A certain shape of glass was also known as a “flesh,” a term derived from the fact that it was covered with a viscous fluid called ethanthus. Fertilizing, in other words, was the process of opening and closing the living tissue, as well as of dissolving it.

Generated 2: What are the potential dangers of a society that prioritizes collective happiness over individual freedoms?

Aldous Huxley’s The Structure of History of the World states, “A society is a system of living and living organisms.” The term “system” is used to describe the various aspects of life. For example, for every living thing in the universe, there are millions of other living things. Each of them has a certain specific set of characteristics, some of which they call “substance,” others, such as color, protein, and other special properties. The overalls of each of these organisms are divided into three main categories: (1) Food; (2) Water; and (3) Air.” (quoted

Generated 3: Imagine a world where governments use technology to control thoughts. How would people react?

Aldous Huxley (1894-1963) was the world leader in science and technology. Over the next thirty-four years, he and his team of scientists and students developed and implemented a wide range of new methods of conducting research. Each of them carried a unique set of responsibilities. For example, the Director of Research and Development, for each of his departments, a certain amount of time was devoted to conducting experiments. The overalls of the various departments were divided into three main sections: (1) Scientific and Technical; (2) Political and Legal; and Miscellaneous. (3) Each department was responsible for conducting and supervising its own

Generated 4: Discuss the impact of universal happiness on creativity and innovation in society.

Generated 5: What role should government have in the personal lives of its citizens in an ideal world?

Aldous Huxley, in his new book, The End of the Cold War, argued that the United States and its allies were in a state of collapse. The war was over and the world was on the brink of a civil war. Only a few days after the end of World War II, a new order of government was created in Washington, D.C., and it was led by a president, named George H. W. Bush. Over the course of his presidency, he made a point of personally conducting a series of round-the-clock meetings in which he personally addressed the nation's leaders. Each of them had a notebook

Generated 6: Debate the ethical implications of genetic engineering in humans to achieve societal goals.

Generated 7: How would a society without any monetary system operate? Describe its economy and daily interactions.

Aldous Huxley's The Elder Scrolls V: The Daggerfall Role-Playing Game, in which this work is based.

Generated 8: Envision a political system that perfectly balances individual liberty with societal order.

Aldous Huxley's The Wager and the World's End is the only work of the twentieth century that has been translated into a living, breathing book.

Generated 9: What are the consequences of abolishing all forms of privacy to ensure safety and harmony? The answer is, of course, to a certain degree of certainty. The only way to guarantee the security and security of our lives is through a system of checks and balances. Each of us has his own set of rules, which he must follow, and which must be followed. It is a privilege of the State to keep an eye on the welfare of its citizens.

"And this," said the Director opening the door, "is the Fertilizing Area." The Director of Hatcheries and Conditioning entered the room, in the shape of a wintriness bordering on abject resignation. A troop of newly arrived students, all of

Generated 10: Describe a future where artificial intelligence governs society. What are the pros and cons?

Aristotle's words, or rather, his general idea, of a new order of things. Each of the three main departments of science, technology, thought, and culture, was led by a certain man. The overalls of all three departments were white, their hands gloved with a pale corpse-coloured rubber. Over the counter, for all the summer beyond the panes, a bleakly shining light glared through the closed windows, hungrily flickering along the polished tubes like butter, streak after streak in long recession down the work tables. A troop of newly arrived students, all of them grizzled and callow, followed nervously, rather

Discussion

In terms of the LSTM model, the sequences generated are not very coherent and do not make much sense. The actual content though does seem to be similar to the text in *Brave New World*. The GPT-2 model, on the other hand, generates sequences that are more coherent and make more sense. There does seem to be a lot of hallucination in the sequences, with some of the content being factually incorrect. For example, in the 7th sequence, when prompted for "How would a society without any monetary system operate? Describe its economy and daily interactions.", the model generates a response that is based on a video game, *The Elder Scrolls V: The Daggerfall Role-Playing Game*. This is not a factual response, and is not based on the text in *Brave New World*. There is likely no mention of this video game in the text, as well as the fact that Daggerfal isn't even the 5th game in the Elder Scrolls series. The GPT-2 model does seem to generate more coherent sequences, but the content is not always accurate. These problems could possibly be remedied. For the LSTM model, the model could be tweaked to predict words instead of individual characters, which could make the sequences more coherent. In addition, I could add aditional texts from Aldous Huxley in order to diversify the training data. For the GPT-2 model, the model could be fine-tuned for more epochs as well as playing around with batch sizes, which could make the content more accurate.

Reflection

Overall, the LSTM model seems to generate sequences that are more similar to the text in *Brave New World*, but the content is not very coherent. The GPT-2 model generates sequences that are more coherent, but the content is not always accurate. There are ways to improve both models, such as tweaking the LSTM model to predict words instead of individual characters, and adding more texts from Aldous Huxley to the training data. In the future, I would probably play around with different RNN architectures as well as transformers, and utilize different fine-tuning models to see if I can generate more accurate and coherent sequences.