Final Report: NLP generation of Sci-fi stories

Team members: Rachel Goeken

Summary:

My goal was to develop an application utilizing NPL techniques that could generate a new and unique sci-fi story using input from example sci-fi stories and books. The first step was to stream together words to make an accurate sentence. From there I would hopefully be able to generate a cohesive story.

If the application can be developed to the full extent it can either write or help in the generation of short sci-fi stories. The ultimate goal was to have the application write stories that show up on the New York time best sellers list. Currently there are some book generators, poem generators, and most famously tv/movie script generators. These show up from time to time on the internet for their humorous output.

Method:

Background research I did:

https://www.researchgate.net/publication/313874773_NLP_based_Poetry_Analysis_and_Generation

This research paper gave me insight into various different models that could be used and methods to combine the models in order to improve results (Markov chains, Context Free Grammar and Recurrent Neural network/Deep learning). It also went into what it means for the generated text to be considered a poem. This provided me some points to look at when evaluating my application's output.

https://arxiv.org/pdf/2110.15723.pdf

This paper was helpful for approaches when the input/output is in a language other than English. I believe that this could help make my application better with the long-term utilization of the application. In the article the authors were able to generate poems that had better cohesion and the same or better quality. They also gave some details about their scoring model and how they evaluated the generated text.

https://aclanthology.org/2020.acl-main.223.pdf

This article goes into detail about how they were able to take non-poetic text as input and generate (with constraints) a poem. They were also able to achieve this in two different languages. In the future this might be helpful or a feature I would want to investigate.

https://arxiv.org/abs/1808.10122

This article discussed some of the problems with specific models. They were able to show similar results with a neural generation system with a hidden semi-Markov model decoder. This knowledge could help me to be more open to other model combinations as options beyond the standard neural networks. This is a feature that I may wish to investigate in the future.

https://arxiv.org/abs/1711.09534

If I use neural networks, this article will hopefully give me some guidance to be able to work through some expected but undesired behavior.

https://www.sciencedirect.com/science/article/pii/S1319157820303360

This article was an overview of research that has been done from 2015 to the time of the paper in the area of NLP. They look at various models and some of their benefits and detriments when it comes to NLP generative text problems.

Implementation:

From these methods and insights, I initially started out using a LSTM (long short-term memory) model that would be supervised using the input data in the text generation. I also tested out a Markov model that would be unsupervised in the text generation. Both of these methods were able to generate sentences that vary in length but they were somewhat clunky in the 'story' that was output.

For Markov using the Kaggle sci-fi story (https://www.kaggle.com/jannesklaas/scifi-stories-text-corpus) I was able to generate the following output:

```
The Voion were breaking in and...

Though he did was to be faced if they were not such things existed.

What better time to expire, but she was right, too.

Nevertheless, Ren continued swiftly, having no idea that present a
publishing program.

The Ballard quotation above was struggling to his position to raise my
arm, then pressed a button and the flag.

St George found his mind to be making a great director could see why.

He plans to solve Aleuker's IQ test, and unfortunately that had had to
be done that often.

The next day Zuliani met Rdph during his fourth so far.

The beings from a vantage point of view.
```

(For this input I didn't run it against the LSTM data. This was because initially I thought I was running on a model utilizing both. I didn't go back and run against this data since the input was not what I had hoped for.)

Based on the output, I went back to look at the Kaggle document and saw that the story was not in a good format (grammatically or spelling). But even with this ill formatted story I was able to generate sentences. From there I went looking for multiple stories that I could utilize in the training and generation of sci-fi stories. In my research I found: https://blog.reedsy.com/short-stories/science-fiction/. This website is a forum where people can upload stories of different genres and have them liked. The highest liked stories will show up first for any given genre. (This can be utilized in future training of weights and models).

With LSTM I kept getting repeated words, so I chose not to move forward with that model.

```
She tried to think back, but Alham's intensity was making it
impossible to think around the fog of feelings. Alham's eyes bore into
her. It was the eyes that spoke to her, spoke deep into her soul. The
eyes continued to paint a scene that she could not reliably recall
without them.  ‘Ah, ‘Ah, ‘Ah, ‘Ah, ‘Ah, ‘Ah, ‘Ah, ‘Ah, ‘Ah, ‘Ah,
```

This is most likely due to how I am building the dataset.

I tried some other model layer structures but they took more than 10+ hours to train. This led me to focus on the Markov models.

From there I worked on model enhancements to Markov since I was getting good results with that model. This approach generated better individual sentences and sentences that improve the flow when threaded together. I did this by incorporating GPT-2 as a secondary model. It utilized the sentences that were generated by the Markov model as the starter sentence to the GPT-2 117M model. This allows for a more cohesive and better organized story.

The issue I now ran into was that the story would veer off down a path that was not relevant to the initial intent.

Markov:

```
(Photo: Getty Images, File)

Clinton's first book, the book she wrote for the first time in 1996
about her time as secretary was called "Clinton: A History of the
Clintons" (published in 1998), a memoir about the Clintons.
```

Or

```
Well, we do have a lot of work to go. We are working with the
government to develop a new technology that is more accurate and
efficient than traditional grids and we're looking at ways of doing
this." This article was first published at the New York Times, a
website that is a hub for the world's leading news organizations.

In a recent column for The Washington Examiner, David Ignatiev, the
author and former chief White House press secretary to former
Secretary Hillary Rodham Clinton, argued for an end to "the Obama
administration's policy of 'regulatory opacity' that has allowed the
federal government to dictate how and where it operates and to impose
its own set of rules."
```

…

```
(LAUGHTER) And I thought about what it was about to be the second most
diverse movement in the history.

RANK 2 0 1 0.0 1.0 No votes. 1 1.5 0.0 No votes. 1 1 1 0.0 No votes. 1
1.5 1.5
```

This led me to a path to try to train the GPT-2 model with my training input. I did this by following these instruction manuals:

https://pypi.org/project/keras-gpt-2/

https://lambdalabs.com/blog/run-openais-new-gpt-2-text-generator-code-with-your-gpu/

https://www.geekslop.com/features/technology-articles/computers-programming/2020/what-is-gpt-2-and-how-do-i-install-configure-and-use-it-to-take-over-the-world

I had to pull bits from all three since I don't have access to my GPU.

This allowed for a somewhat more cohesive output:

Markov:

```
Charlie stood up and looked out at the room without a key on her
shoulder to me for a door on the streets, I see them again.

 His black hair that clung to the mountains was a nice one.

 He was a good guy, but he didn't look like he had any real power.

 He looked more of a boy, a little boy with his own personality, but a
little boy.

 He was not like a normal man, but I knew he was a boy, I knew that he
would grow up to be the best man in this room and he would be my best
friend.

 I looked at his hands, and then I saw him.

 He looked like he had been shot.

 He had his eyes fixed upon a large white man, with long, dark-green
beard. He was wearing an open shirt, a white coat with black sleeves.
His eyes looked down at the man.
```

As we can see there are still some issues though with the model:

Markov:

```
(The New York Times, April 19, 2013). The study was published online
May 6, 2013 in the journal Nature Medicine.
```

```
. The study is funded in part with the U.S Department of Defense
Office of Naval Research. This article is a disambiguation article for
The Legend of Zelda series

 "I don. I. don't. want. To. Want." —Goro to Link, after he defeats
Link's army in a fight.[src][src]. [5][src][5] In the manga adaptation
of this arc,[6], Goro's father, a former soldier and soldier of King
Kagu, is a member. In this story Goto's mother, a former soldier of
the Kagu, was also a Kage. In the manga, Link is the first of the
three to meet Link, but he does not meet him until he meets Link's
father.[6].
```

This was similar but very different from my baseline implementation. For my baseline I thought I was utilizing both LSTM and Markov in one model. This was not the case. I was only utilizing the Markov in my results. Once I realized this, I wanted to look at how the two compared and which, if either, of the models performed better.

I also was seeing that the Markov output was not as accurate as I had hoped but its output was closer to being properly formed. Therefore, I used the Markov generated sentences as input into the GPT-2 model. This was taken a step further by training the GPT-2 model to give output that hopefully correlated more to sci-fi rather than all genres.

Experiments:

Many of the experiments I did were in response to subjective evaluation of the output that was generated.

For the LSTM model I adjusted the number of hidden layers in the network during training to see if I could correct the repeated output.

I ran experiments to include syllables and syllable weights in the training for the LSTM model, again to help with the repeated output.

Also, with the Sequential model, I tested what the different options were for the hidden layers (Dense, Bidirection, and Embedding) to see if the repeated output would decrease.

For both LSTM and Markov, I was not thrilled with the output I was getting so I researched other models that I could utilize alongside them. I selected GPT-2 for additional improvements.

When I started to utilize GPT-2, I played around with top_k values of 1, 2, and 3. While 3 takes longer to generate values it produces less repetition. When top_k had a value of 1 there was a lot of repetition, 2 was better but there was still too much repetition. Since I was breaking up my sentence generation, I tried 3 and had good success.

For all the experiments I initially used: https://www.kaggle.com/jannesklaas/scifi-stories-text-corpus. I found this data to be not well normalized. But once I determined I needed different data I utilized stories from: https://blog.reedsy.com/short-stories/science-fiction/.Both inputs were saved in the form of a text file.

The final result of utilizing output generated from the Markov models in the GPT-2 model gave (subjectively) much better cohesive output to an English reader.

Metrics for this type of system are difficult to enumerate and therefore tend to be subjective. What constitutes a good story? I started to include metrics that would help determine if there was repetition in a sentence. This simple, rudimentary number can be applied to this problem to figure out if you were redundant. This helped me to determine a top_k value for GPT-2 that decreases the repetition in the output. It is currently not directly used in training the system.

Conclusions:

The NLP models that read TV shows and other books take a lot of effort to create. There are many ways to implement this, and the training is key here. Subjective problems inherently have the problem of not having numeric evaluations.

GPT-2 had well trained models. One of the issues that I ran into was that sometimes with the malformed input from Markov the output would start off a bit strange, but it would come around to a cohesive output. The other major issue I ran into was that the models for GPT-2 were trained on general input – it wasn't specific to a genre or field of research. This led to the story going off into the weeds of things that are relevant to the present but not to the genre that I was trying to generate.

Finally, I noticed the sample data that was generated when I was training the GPT-2 model was really good. I would suggest investigating to see if I could get to that same output.

For example:

```
Generating samples...

======== SAMPLE 1 ========

 computer-dependent. They did know what I said, and they did it
slowly, like I'm an electrician. Why was I confused?

Slowly the truth dawned on me. I wasn't alone. The crew of the
spacecraft had communicated in space for the last twenty years. Since
the failure of the Friendship 8 mission, an SOS had been launched from
Earth's atmosphere, programmed to return anyone found to the surface
if they were found. They had been successful in finding people, but
had not been able to find information on the cause of the toxic
sludge.

It was only a theory, a fantasy, a desperate fiction. I had been away
for two years, and my mom had barely had the chance toregnuff her baby
on a first birthday. She had found herself pulling multiple
pregnancies out of a birth certificate just to delay the procedure.
How could she be bothered to answer a question like that?

I was missing one more piece. This was the best clue as to what had
happened. I had been out of earshot for almost a year, and probably
would be out for much longer, had Tim chosen to come home. Even if he
```

```
didn’t return, I can’t help but wonder if he was trying to kill
himself. It seemed like a reasonable thing to do, given that he’s been
out for a while. But if this was the start of a long tail, how could
one person be so easily swept away?
```

Also, I propose reevaluating the LSTM and other models to see if they give any better output. In addition, I suggest trying to properly train the GPT-2 model. As well as getting more into updating the weights of the data (more weight to highly ranked blog postings, more weight to specific topics, etc.).