

Amazon Review Generator

Feng Zhang, Jiawei Wang
{feng1013, jwang96}@uab.edu

Overview

In this project, we fine-tuned two different models (RNN, GPT-2) for text generation using Amazon Review Data, to see if a well-trained model can generate human-like product reviews. As an evaluation, we first investigated the result with our own, then evaluate them with an automatic grammar checker Grammarly. After analysis, the output reviews from 1.0 to 5.0 score, it's clear that 5.0-score reviews are better than 1.0-score reviews. Comparing the GPT-2 model with the RNN model, GPT-2 performs much better than RNN. Since in a long sentence, previous information may get corrupted in the LSTM structure, but not in the Attention model (as used in GPT-2).

Dataset

We are using Amazon Review Data from <https://nijianmo.github.io/amazon/index.html> and extracted 1M reviews to train the model.

Books 1M sample:

```
{"overall": 5.0, "verified": false, "reviewTime": "03 30, 2005", "reviewerID": "A1REUF3A1YCPHM", "asin": "0001713353", "style": {"Format": "Hardcover"}, "reviewerName": "TW Ervin II", "reviewText": "The King, the Mice and the Cheese by Nancy Gurney is an excellent children's book. It is one that I well remember from my own childhood and purchased for my daughter who loves it.\n\nIt is about a king who has trouble with rude mice eating his cheese. He consults his wise men and they suggest cats to chase away the mice. The cats become a nuisance, so the wise men recommend the king bring in dogs to chase the cats away. The cycle goes on until the mice are finally brought back to chase away the elephants, brought in to chase away the lions that'd chased away the dogs.\n\nThe story ends in compromise and friendship between the mice and the king. The story also teaches cause and effect relationships.\n\nThe pictures that accompany the story are humorous and memorable. I was thrilled to discover that it is back in print. I \"highly\" recommend it for children ages 2 to 7.", "summary": "A story children will love and learn from", "unixReviewTime": 1112140800}
```

Electronics 1M sample:

```
{"overall": 5.0, "verified": false, "reviewTime": "12 21, 2011", "reviewerID": "A1TOXCG4QWQLHA", "asin": "B00001WRSJ", "reviewerName": "Ranjan Chari", "reviewText": "My first experience with these headphones exceeded expectations. You can hear stuff you normally would not hear with cheaper headphones. The bass is incredible with this set and the phones are easily driven using standard computer headphone jacks as well as portable MP3 players. Toss out whatever cheap headphone you currently use and go for these. My only regret was not buying this pair earlier.", "summary": "Godly Sound - This is what you need for a quality listening experience.", "unixReviewTime": 1324425600}
```

RNN(baseline)

Model Structure

Layer (type)	Output Shape	Param #	Connected to
input (InputLayer)	(None, 40)	0	
embedding (Embedding)	(None, 40, 100)	46500	input[0][0]
rn_1 (LSTM)	(None, 40, 128)	117248	embedding[0][0]
rn_2 (LSTM)	(None, 40, 128)	131584	rn_1[0][0]
rn_concat (Concatenate)	(None, 40, 356)	0	
attention (AttentionWeightedAve)	(None, 356)	356	rn_concat[0][0]
output (Dense)	(None, 465)	166005	attention[0][0]

Total params: 461,693
Trainable params: 461,693
Non-trainable params: 0

Output Example

1.0: Great book we see that it was approvenes to see how the most presented to the ship readable. The situation, this is a creation of the Narnia series. The only and come to read this book in the series. I recommend this book to my very well written book accounts it words.

5.0: The Harry Potter book is a great book at all. I was remembered. I read this book since I was the older side of this book that is God up a world of the story by the best book in the series. I did not read it as a said and I was a fantastic story because it was disappointed.

5.0: The Twilighcal Church has been a difficult great manager. I should have enjoyed , and the story does not be a few times because of the story of a bit of a simply adult thingre and thought that the book or time was born, and the book is a really amazing read...

Evaluation

As we can see at the output example, RNN generated text is not good.

- In terms of relatedness, these generated results lost track of contents.
- In terms of grammar, we use Grammarly (Automated Grammar checker) to check its correctness; they received an overall score of 65/83/59.

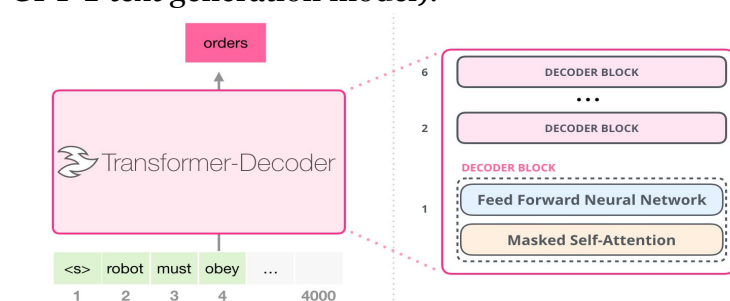
Summary

As a baseline, the RNN model succeeds in generating review texts that look like a reviewer. However, it does not perform well in terms of content relatedness. Maybe we need a model that understands better in context, such as a Transformer-related model.

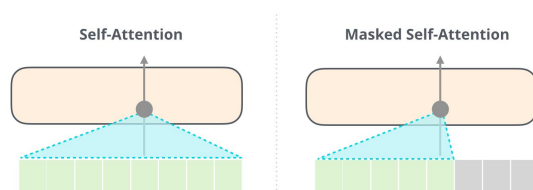
GPT-2

Model Introduction

GPT-2 is a pre-trained language model which can use for various NLP tasks, such as: Text generation. Language translation. In our project, we used model from a simple Python package named gpt-2-simple (A simple Python package that wraps existing model fine-tuning and generation scripts for OpenAI's GPT-2 text generation model).



A normal self-attention (what BERT uses) block allows a position to peak at tokens to its right. But Masked self-attention (what GPT-2 uses) prevents that from happening.



Output Example

❑ Electronic Dataset: 1.0 and 5.0 score review

1.0: I have had this for over a year and it works perfectly. I have had several other products. This one works great as well. It comes with a case, instructions and a usb mouse. It works with my 1st gen mouse. Its a little heavier than I expected, but I still use mine for the most part. I use it for school. Its a good deal.

5.0: I have always liked the sound of the JBLs for the money. They are very crisp and clear. They are very portable and come with all the hardware needed to get them out the door of a car. They are also very good for the price.

❑ Prefix = '5.0: The Harry Potter'

5.0: The Harry Potter author has created a perfect world that is so unique. The characters are so unique and so original that they bring out the food at the table. Your children will love the books. This book is a must read!

❑ Prefix = '5.0: The Twilight'

5.0: The Twilight Zone is the second best book of all time. It's the new kind of medium. It's the same old medium that plays out in movies. It's the same old medium that weaves into the music of the 1950s. And the results are amazing. The Twilight Zone is the one book with the signal to the world that the world has turned into a movie. Will this be the next book? I don't know. It's too good to be true and I wish I had read it years earlier. I'm not a fan of the movie, but it's still a great book.

Evaluation

The outputs of GPT-2 model are much better than RNN above.

- In terms of relatedness, these generated results are excellent. It's easy to understand what is talking about, and it describes products precisely.
- The model is terrible at discriminating reviews between low and high scores. 5.0 review is not bad, while 1.0 review does not make sense.
- With the help of Grammarly, the scores of 4 examples are 62, 96, 90, and 78 correspondingly. But it is still much better than the results of RNN.

Summary

GPT-2 does well in the text generation process. The generated text is readable and easy to understand. It could generate precisely content with prefix provided, although some sentences get low scores with Grammarly due to tiny grammar mistakes (e.g., its should be it's). However, the performance on discrimination of scores is terrible. We think this is because 1.0 reviews are insufficient in training data.

Reference

- Ni, J. 2019. Amazon review data. [nijianmo.github.io](https://nijianmo.github.io/amazon/index.html). <https://nijianmo.github.io/amazon/index.html>.
- Alammr, J. 2019. The Illustrated GPT-2 (Visualizing Transformer Language Models).
- [Jalammr.github.io](https://github.com/nijianmo/gpt2). <http://jalammr.github.io/illustrated-gpt2/>.
- Woolf, M. 2019. minimaxir/textgenmn. [GitHub](https://github.com/minimaxir/textgenmn). <https://github.com/minimaxir/textgenmn>.
- Woolf, M. 2019. minimaxir/gpt-2-simple. [GitHub](https://github.com/minimaxir/gpt-2-simple). <https://github.com/minimaxir/gpt-2-simple>.
- Write your best with Grammarly. 2019. [Grammarly.com](https://grammarly.com/). <https://grammarly.com/>.