



Caption Generator Obscura

An Exploration of Caption Generation
with a Small and Varied Dataset

Contents

1. The Problem
2. The Data
3. The Caption Generator
4. The Results
5. The Examples
6. The Recommendations
7. The Future
8. The Credits



The Problem

Have you ever tried to post your travel photographs, and had a brain freeze about how to caption them? Or have you ever just lacked the creativity to make anything but the most literal caption?

With my program, you can upload your photograph and it will generate a creative suggested caption for you.

The Data

I downloaded the data for training the model from the Atlas Obscura website using Scrapy, a Python tool for pulling data from websites.

I deleted images without captions, images that wouldn't display, and short captions of 2 words or less.

19,455 images and corresponding captions remained after cleaning.

The Caption Generator

The caption words were processed by how similar they were using a pre-trained word characterization model from Stanford University (GloVe 6B 200d).

The photographs were processed using a pre-trained image recognition model from Google (Inception v3).

The Caption Generator

Both the words and photographs were then run through my model, which predicts the next word of the caption based on the previous words and the photograph.

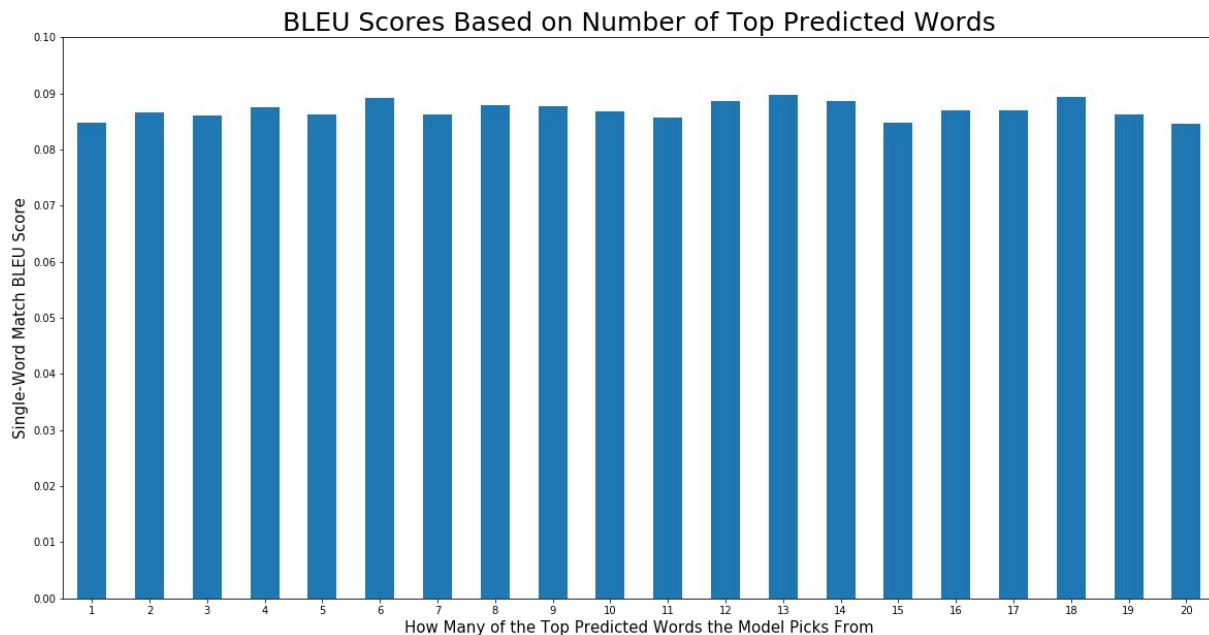
Normally, models like this work with literal captions (e.g., dog on the grass) not creative ones (e.g., a museum full of wonders and curiosities sits atop a historical ancient cemetery).

The Results

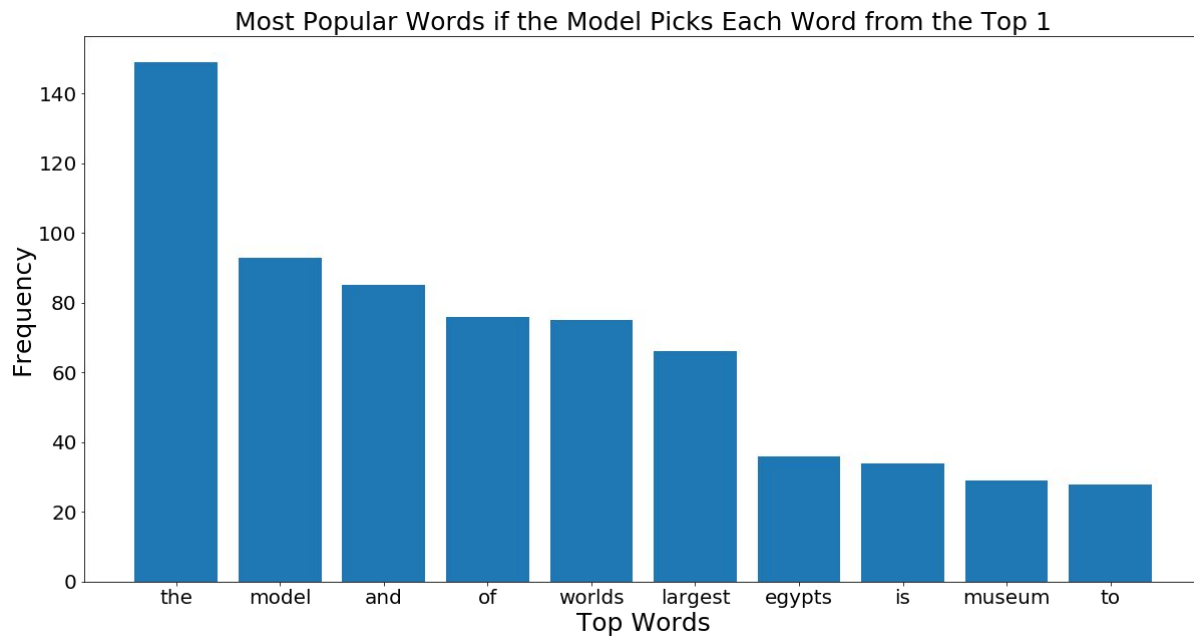
BLEU Score: A measure of how many of the model's predicted caption words match the test caption. Like an R-squared or accuracy for captions. Usually looks at groupings of words too, not just single words, but I looked at single words for this caption generator. On a scale of 0 to 1.

Number of Top Predicted Words: Whether the model picks from the top 1, 2, 3, or more, predicted words to use as the next word in the caption. I added this to the model to allow for more variation in the captions. I refer to this as the caption's "unpredictability" on the website.

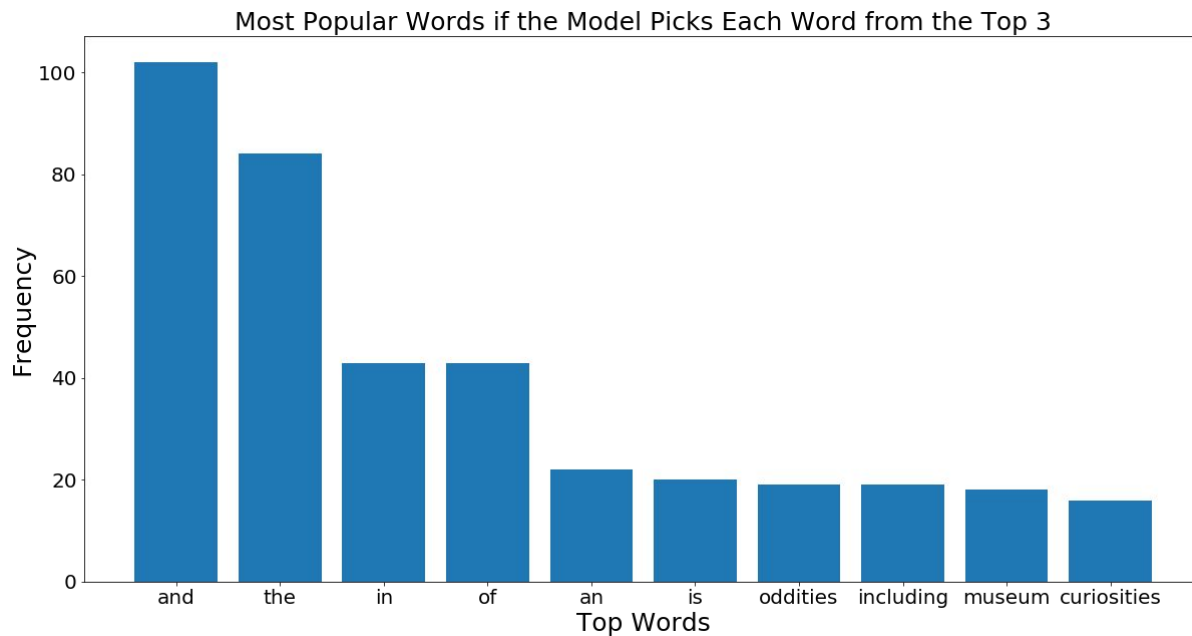
The Results



The Results



The Results



The Examples



the worlds largest model model
museum is home to the worlds most
famous flowers and the remains of
the worlds largest model of egypt's
egypt's and the dead flesh of the
shooting bug lovecraft and the
occasional hungry and

The Examples



the most disputed bridge is one mans
lifetime of oldest of modern in san
franciscos austin city that dates from
to all over million glass bottles in the
philippines has old postcards with
direct sculptures by his childhood
mark

The Examples



the worlds largest model model
museum is home to the worlds
largest model of egypt's history and
orleans and patriotism of the french
poet and the dead flesh of the
proportions and of the proportions
and of the sea

The Examples



the remains stranger of its buzzing
history oddities and oddities models
from antique equipment and plant
that seems up in continental national
library is now one and largest
collection in new hampshire and with
globe in this historic cemetery

The Examples

<https://caption-generator-obscura.herokuapp.com/>



The Recommendations

- Train on captions from more travel websites, e.g., Lonely Planet
- Try more types of models for pre-processing images (before passing the data to the caption model I created), e.g., ResNet by Microsoft
- Add more layers or make other adjustments to the caption model I created

The Future

- Train and use for predicting captions from a larger variety of websites:
 - Cooking blogs
 - Real estate catalogs
 - Social media
 - Music blogs
- Generate creative image captions for blind users for images that don't have good alt text or a caption

The Credits

- Dr. Jason Brownlee on <https://machinelearningmastery.com/>
- Harshall Lamba on <https://towardsdatascience.com/>
- The DataFlair team on <https://data-flair.training/>
- Mohd Sanad Zaki Rizvi on <https://www.analyticsvidhya.com/>
- Hafsa Jabeen on <https://www.datacamp.com/>
- David Membrives on <https://letslearnabout.net/>
- User “buildwithpython” on <https://www.youtube.com/>
- Alexandre Wrg on <https://towardsdatascience.com/>

The Credits

- Moez Ali on <https://towardsdatascience.com/>
- Krish Naik on <https://www.youtube.com/>
- The Cambridge Spark team on <https://blog.cambridgespark.com/>
- User “Tech With Tim” on <https://www.youtube.com/>
- Julian Nash on <https://www.youtube.com/>

The Credits

My classmates, Aman Hafez, Wayne Chan, Kavan Pandya, Mariam Javed, and Shawn Syms, who were amazing to collaborate with throughout the course and always super helpful

Our very supportive and quotable instructor, Musfiqur “Sumit” Rahman

My wife, who wrote unbiased test captions and suggested I use Atlas Obscura for my training data