Bigfoot Assignment 2 Report: Team 9
DSCI 550: Spring 2024

Large Scale Data Extraction and Analysis for Bigfoot

In this assignment, we furthered our analysis of the Bigfoot Field Researchers Organization (BFRO) dataset and the data we added in the first assignment through image, location, and entity analysis.

Image Analysis through Captioning and Object Detection

Image generation was performed on leveraged text from each report followed by image captioning and object identification. The image captions represent the generated images of particular objects such as vehicles (car, truck) and natural landscapes (forest, park, road, river, water) with high accuracy. Additionally, colors were easily identified and described, for instance, many captions include color descriptions such as “black and white”, “green”, and “red” with accuracy.

Other types of images which were not accurately captioned. The most common caption was “a red fire hydrant sitting in the middle of a forest” with a value count of 197 and “a fire hydrant in the middle of the forest” with a value count of 161. Upon inspecting the associated images for these captions, none of them actually depict a fire hydrant. Tika Dockers generated this caption for images that only contain environmental objects and no other identifiable objects. Additionally, creatures including elephant, gorilla, humans, giraffe, and bear were included in the captions with low accuracy. For some images, an animal will be identified merely because some shapes in the images look like an animal (e.g. tree brunched-giraffe). Similarly, if an animal or object is included in the image, the image captions are likely to include and describe the behaviors of said animal or object. However, these behavioral descriptions have low accuracy and seem to be based on assumptions of the relationships between two objects. For example, a bear might sit on a branch and an elephant might stand on the ground.

The objects detected in the image were verified by calculating the percentage of objects in the sighting text that was used for generating the images. This resulted in a finding of 5.3%. Similarly, the inclusion rate of detected objects in the image caption was 5.2%. The percentage of detected objects included in both the image captions and the text used to generate the images was notably low, at just 0.26%. The overall proportion of detected objects included either in the texts used for image generation or in the image captions was 10.3%. These findings suggest a significant difference between the text used for image generation and the image captions. If the sighting description text and captions had similar content, we could expect a higher rate of overlap in the detected object strings between them. Several factors could contribute to this discrepancy. For instance, the image generation process might lead to the generation of images which differ from the original text. The image generation model could be prioritizing certain tokens at a higher frequency than others or some images could be based purely on a limited amount of tokens. For instance we observed an image depicting plastic bottles. This minute detail from the description of the bigfoot sighting was present in the image caption possibly due to the plastic bottles which were prioritized by the image generation model instead of the entire textual description which seem more significant to humans. The context of the original description was found to be lost in the captions. This highlights a need for caution while using data obtained by employing large scale image generation and captioning methods, since it could result in substantial reduction or alteration of the content.

The trend that stood out the most in the image captions and identified objects was how images with bigfoot creatures were identified and captioned. While all of the rows have descriptions of bigfoot sightings, not all the descriptions or images are related to what bigfoot would look like; however, there may be interesting things to learn about Tika Dockers from looking at the trends of images that depict bigfoot creatures. For images 1594 and 2401 shown below for example, the caption for both images is: “a large brown bear sitting on top of a tree.” The identified objects for image 1594 were “orangutan, gorilla,

chimpanzee" and for image 2401 the identified object was "gorilla." Similarly, we noticed several image captions such as "a group of animals that are standing in the grass" (image_33), "a large brown bear walking through a lush green forest" (image_1951), "a black and white photo of a dog in a field" (image_1681), and "a black and white photo of a baby giraffe" (image_1011).

 [1]
 [2]
 [3]
 [4]

It seems from these trends that the Tika Dockers image caption and object identification models stem from machine learning or deep learning models where the training data likely did not include bigfoot images. Rather, we may presume that it detects what these unknown images are closest to from what it has been trained on such as bears, dogs, gorillas, and giraffes. It is particularly interesting that it adds descriptions of actions and environmental factors such as "sitting on top of a tree" where that is not the case and the creature may not even be in close proximity to a tree. This may suggest that this was common for images of bears that the model was trained on.

From looking at the most common image captions, we can see similar trends. In addition to the captions for fire hydrants having the highest value counts, the following captions had the third, fourth, and fifth highest value counts: "a wooden bench sitting in the middle of a forest" (77), "a train traveling down tracks next to a forest" (74), and "a large brown bear walking across a lush green field" (62). For the identified objects, the five values with the highest counts were: "mountain tent" (179), "brown bear" (82), "gorilla" (58), "gazelle, impala" (53), and "impala, gazelle" (35); however, when inspecting images depicting bigfoot rather than other animals, "gorilla", "chimpanzee", and "orangutan" appear to be the most common values. It is interesting that the captions seem to most commonly identify bigfoot creatures as bears while the object identifier most commonly detects bigfoot as a gorilla. While the image captions and object identifications were more accurate on images depicting natural landscapes, wildlife, or humans, we noticed images of forests without fire hydrants where that is in the caption. As that is also the caption with the highest value count, this suggests that there is a similar trend in how Tika Dockers detects other items based on the data it was trained on for natural landscapes as well, even though this was more apparent for images depicting bigfoot creatures.

Geographic Analysis with GeoTopicParser

In Task 7, we retrieved location information for each report using the GeoTopicParser. While analyzing the locations generated by GeoTopicParser we were surprised that even though they were a close second, the most common locations weren't states but rather national parks. More specifically, the most common National Parks were Mount Rainier' (306 instances) and 'Great Smoky Mountains National Park' (257 instances). Besides these, 'Colorado Springs' (168 instances) was the most common city and strangely enough 'River Nile' (68 instances) was the most common natural landmark. Although this is just a guess, the former seems to be a default location generated by GeoTopicParser, since it appears when the word 'River' is mentioned multiple times without any specific name attached to it such as Mississippi River. We were also able to notice that sightings belonging to 'Class A' and 'Class B' seem to have, on average, 60% more locations per sighting than 'Class C', with the rates '1.77', '1.88' and '1.13'

[1] image_1595
[2] image_2401
[3] image_33
[4] image_1011

respectively. This tells us that being specific and providing more information about the sighting location plays a significant role in the 'Class' classification.

Finally, while relating locations with sightings, we calculated that there is an average euclidean distance of '14.84' between the sighting location (column of the original dataset) and the nearest location generated by GeoTopicParser. This will be an estimate of 1020 miles which is 1640 km, a considerably large distance that make us think that the locations detected by GeoTopicParser may not be fully relied on, since even if the location name is correct, the geographical coordinates might be inaccurate due the possible extension of a location such as a river or a state.

Named Entity Recognition Analysis with spaCy

Named Entity Recognition was conducted using the free and open source natural language processing python library SpaCy. It could detect entities of various types such as 'Date', 'Geopolitical', 'Organizations', 'Cardinal', 'Time', 'Person', 'Quantity', 'Location', 'Nationalities, Religious, or Political Groups', 'Ordinal', 'Percent', 'Facility', 'Work of Art', 'Product', 'Event', 'Money', 'Law', and 'Language'. These entities provide a deeper understanding of the reports by extracting meaningful data out of the text which can be easily glanced to understand the relation or summary of the reported events. For example, just a quick look at the Time, Location, and Person entity tells us who saw the bigfoot, where, and at what time. Extracting this information manually from such a huge dataset of 5500+ reported sightings would have been very time-consuming.

The number of witnesses detected by the number parser and spaCy was inconsistent with each other in some cases. For instance, the original description could mention 4 people but the witness count by number parser would detect 5 people and spaCy would identify names of 2 people mentioned in the report as Person entity and the number 4 as cardinal entity. Even though spaCy could detect the numbers, it did not provide context in most cases. For instance, if spaCy detected '4', we won't know if the 4 refers to people or trees. It would be more appropriate to use number parser for total count of people and just refer the Person entities for names of people mentioned in report, instead of relying on cardinal entities. The person entity includes names of all the people whose names were mentioned in the report such as 'Susan' and 'Jerry', and the cardinal entity includes counts of people described numerically such as 'at least six', 'only two', and '3'. The description could also contain cardinal values describing unrelated things such as 'two' present in 'two large spruce trees' could be detected as a cardinal entity. Therefore, one cannot solely rely on numbers detected as cardinal entities. Additionally, it would fail to detect 'My hunting buddy and I' or 'Me and my husband' but tagged 'berry bushes', 'wolf', 'looking', and 'strawberry' as Person Entities. Bush and Wolf could be confused by spaCy since they are commonly used last names such as for the names George W. Bush and Robert Wolf. However, words like 'looking' and 'strawberry' were unexpected to be found in the Person entity.

SpaCy further confirmed our findings such as the most common time for sighting bigfoot was between late evenings and early mornings. Complex terms described in either numerical or word forms are also easily identified, for example the 'Quantity' entity includes useful information such as 'about seven foot' or 'approximately 75 yards', and the 'Time' entity has data such as 'Friday Night', 'forty minutes', or 'about 12:00 Midnight/full moon'. Misidentification of time was another common occurrence. For instance, 7 P.M. would be categorized separately sometimes where '7' would be a cardinal entity and 'P.M.' would be a NORP entity. Also the majority of the entities detected as 'Work of Art', 'Product', 'Event', 'Law', and 'Language' were incorrect. Eg: 'Sasquatch' was detected as a language entity, 'Hurricane Rd' was detected as an Event Entity, 'Longitude W142' was detected as a Product entity, and 'Hairy Man' was detected as a Work of Art entity.

We also verified whether it was possible to validate the national parks associated with each report through the entities detected by spaCy. As a result, There are correlations between the national parks and identified entities. For instance, for the observations with “Mount Rainier NP” which is a national park located in Washington state, there are entities such as “WA” or “Washington” identified in a large number of cases. Other entities, related to the city or specific addresses, are somewhat correlated with national parks, e.g. name of the city close to the park. However, there are still some unrelated to the national parks entities, such as “California” for the same “Mount Rainier NP”, or even “Vietnam”. Still, as we have observed the nearest national park - “nearest” might mean a lot of miles away from the actual bigfoot observation - the usefulness of that information would depend on the type of analysis needed.

Reflections on Machine Learning and Deep Learning Tools

GeoTopicParser

In practice, GeoTopicParser has proven to be a fast and effective tool to extract location, latitude and longitude from natural text. Although the final result created is satisfactory, we did encounter some issues getting it to work. First, the whole installation process was chaotic and getting the parser to work was much tougher than using it. We only managed to use the parser in a very specific environment which is detailed in README. Finally, the results provided lacked a bit of consistency, providing different results in its classification (Geographical/Optional) for the geographic location of every execution.

SpaCY

The entities detected by spaCy organized the useful information present in the data. However, the results still include some noise and need to be sorted and cleaned. Another difficulty we faced was having to adapt the code for calculating the scores to our other parts of code. In addition, we spent around 1.5 hours running the code to produce the final version of the dataset. This way, if using even larger datasets, it is needed to somehow decrease that time, for instance by parallel processing.

Tika Image Captioning & Object Detection

For those familiar with Docker, Tika Image Captioning and Tika Object Detection are very user-friendly. However, since both image captioning and object detection use the same port, the instructions in the readme are insufficient for launching both services simultaneously and accessing them from Python. We resolved this issue by using the docker-compose command to map the common port number to different port numbers. A suggestion for improvement would be to distribute a Docker Compose YAML file on GitHub, designed for launching multiple services simultaneously, as it would be more user-friendly.

Diffusers

Finally, we used the Diffusers text-to-image model for generating the images which was very easy to implement. The main challenge was realizing we needed to work in a non-local environment to utilize the GPU available in Google Colab in order for it to not take 30 minutes per image. A limitation to the model is that it only takes a total of 77 tokens. After leveraging text columns, we noticed that certain observations in the BFRO data had a significant amount of text with more than a 1000 tokens and some only had 10. For the observations with a lot of text, this presented the challenge of optimally selecting 77 tokens prior to running the image generation. While we used keywords to intentionally do so, there is always the chance that a different part of the text would have produced a better image and from comparing text columns to images it seems like the model performed best on text columns with up to a sentence of text rather than a large amount of tokens.