**Intro**

I’m Joshua Banks Mailman. For my Metis Data Science Bootcamp final project, I chose to make a streamlit web app exploiting machine learning, specifically Neural Nets, and Word Embeddings, to classify and then caption images in an amusing way. Here’s an example geerated by my Streamlit app.

Guitar

This caption is apparently based on the idiom ‘Close, but no cigar’ The neural net detected the guitar, but, through the idiom draws our attention to the singer who isn’t playing guitar but IS in the proximity of someone who is. Perhaps she wishes she played guitar. Or: If *only* she did play guitar, she’d be more successful, and so on.

**Motivation**

I had three inspirations:

New Yorker cartoon caption contest, which publishes an uncaptioned image asking reader to invent captions which are evaluated and voted on. Now people do amongst their friends with all kinds of photos. The most amusing captions enlist a reference to something not actually pictured.

As I’ve written about elsewhere composer Milton Babbitt titles his pieces with puns, which use idioms to indirectly reference musical features: *It Takes 12 to tango* based on 12 tones in a tango rhythm, *Whirled Series,* based on a pitch series that loops around, and *Joy of more Sextets,* an intimate coupling of only violin and piano but based on six tones.

Rene Magritte’s titled his surrealist paintings so the meaning comes not just from the image, but rather from how we relate it to the title. He isn’t painting the object he sees with his eyes, but rather what it prompts him to anticipate, seeing the egg’s future is to be a bird.

And Cockney rhyming slang is a clever code based on phrase that rhyme with what they reference: ‘Bees and honey’ for ‘money’; ‘Borrow and beg’ for ‘egg’ and so forth.

All these entail some *indirect* reference that the viewer or listener supplies based on shared knowledge. When something comes to mind that isn’t literally present, we enjoy that *thinking around the corrner,* and so I aimed my app at this.

Although it could upload any image, currently I have it set so you choose one of 5500 images with a slider. Let’s see…

Hmmm, Perhaps you don’t have time for a *full* breakfast this morning?

**Examples**

Beetle

Many of the generated captions are merely cute, but still suggestive. PUASE. Rembering the idiom ‘let the genie out of the bottle,’ it’s funny to think the ladybug has some special secret, or a magical power we don’t yet know about.

Suit

When humans are pictured, the captions often have a touch of irony. PAUSE.

Through the idiom ‘Don’t shoot yourself in the foot,’ we think: one of these two—perhaps him because he’s wearing the suit, or her also because he’s wearing the suit—is about to do something he or she will later regret, some melodramatic of gender dynamics. By referencing the idiom, the algorithm prompts us to draw our own conclusions.

**Data**

The data came from various sources. 5500 images from Imagenet. 1500 English Idioms webscraped with Beautiful Soup, which I colloqualized with certain text substituions. To contextualize and transform image labels, my algos rely on the SenitWordNet corpus of 117,000 sentences, the Gensim Google News and Text 8 Word2Vec corpuses, Carnegie Mellon corpus of English words, and the International Phonetic Alphabet.

**Nuance/Subteley**

The bees knees or Shangra la of this project, and my original intent, is for the algortithm to generate captions that don’t directly reference what’s pictured, but rather do so indirectly.

I succeeded in tailoring the algorithm to do this.

Screeching

As we know owls are known to screech the caption amusuingly *inverts* the original meaning of the idiom ‘preaching to the choir’. Notice that screeching is not actually contained in the image, but rather is brought in by the algorithm, and we make the connection between this caption and what’s pictured.

**Algorithms**

The algorithm I developed, and its component are worth explaining. From the 5500 images in IMAGENET, choose any one. The pre-trained Inception Neural Network categorizes it with one of 1000 labels, in this case *owl*.

Then three algorithms supply related words: The NLTK Synset API supplies synonyms if any.

Term frequency-Inverse Document Frequency correlation matrix I computed, containing over 3 Billion datapoints, answers the question: what words are most likely to occur in a sentence with owl.

Gensium, Word2Vec word embedding, works by analogy. I used human action as a model, so my algo asks the Word2Vec, based on your modeling of text corpuses: if a ‘man walks’ what does an owl do?

Together these generate a Semantic Family of words

Then Phonetics algorithms come in. These translate each Semantic Family word into a standardized phonetic spelling and then compute the Levenshsetin edit-distance to phonetic spellings all the words in the English dictionary, to find close matcheds. and I tailored it it to prioritizes rhymes and assonances. This results in a set of phonetic families of words.

When the app is running in its *nuanced* or *subtelty* mode, the original label and synonyms are supressed, so direct reference to the image content are avoided.

1500 English Idioms are searched for a matches of each phonetic family word. For speed, I precomputed this as a lookup dictionary. From semantic and phonetic similarity, suitability scores are computed.

These are sorted, to inspect all the high scores. The idiom ‘Preaching to the choir’ ranks best, so it’s selected, and the rhyming word ‘screeching’ is subsituted, to produce the caption, which is then placed under the image.

Here are some of my favorites so far.

**Final examples**

peel

Here the algo settled on ‘The real McCoy’, meaning the only real thing. The photo with caption conjur the peeling an apple, and thereby indirectly suggest that perhaps apples, or some specific varieties of apple, are the only real fruit, or the best fruit, or maybe it’s apple computers or New York City.

Deisel

This one might be a headline for a news article about the geopolitics of renewable energy.

Trow

sers

This one one’s a touch ironic. Based on the idiom: ‘All talk and no trousers, which means something like ‘talk is cheap,’’all talk an no action’, or ‘you’ve talked the talk but can you walk the walk?’ which is funny because the photo presents a bunch of prominant signage in a laundreyroom, indicating where to install money for soap and commanding “DO NOT…” do this or that, yet without much evidencc of any launderying going on—or trousers for that matter.

Bones

Sometimes I was momentarily fooled into thinking the algorithm really does have a mischevous sense of humor, with perhaps, more than ourselves, an insightful appreciation of the seemingly carefree quality of invertibrate animals.

Thanks for watching.