# VQA Memnet

Kevin, Atef

October 2017

## 1 Progress

### 1.1 Dataset

- Created dataset with about 50,000 binary (Yes/No) questions and answers. About 120,000 captions across 200 species.
- Maximum number of words in sentence: 66
- Vocabulary size: 6009

We should consider reducing the maximum number of words in a sentence by taking out long captions.

### 1.2 Model

1. Implemented end-to-end memnet in Pytorch, about 5% worse than existing implementation for bAbI tasks. Discrepancy is probably due to small tricks in training.

### 1.3 Experiments

The following experiments use the home-made model rather than existing repos.

#### 1.3.1 Using All Captions

We train on each question-answer pair using all 120,000 captions. This is not viable because GPU runs out of memory or training is slow.

#### 1.3.2 Filtering Captions by Species

We train on each question-answer pair using the 200 captions relevant to the species. Model doesn't seem to learn anything from the captions and just overfits on questions. Training accuracy starts at 73% and peaks at 83%. Testing error is about equal to training error, suggesting nothing is being learned. The same embedding layer was used for each of the 200 species, maybe different ones should be used?

### 1.3.3 Using InferSent and Average Specie Embedding

We reduce the number of captions to 200 by using InferSent to embed captions and averaging over all embeddings for a single species. The problem is InferSent embeddings are size 4096 while memnet usually uses embedding sizes on the order of 30 for questions. To calculate the attention weights, one usually does a dot product of caption embedding and question embedding. This means the question has to be embedded to 4096 dimensions which likely will make the model not work. Model didn't seem to learn anything in training.

### 1.3.4 Using InferSent and PCA

We can reduce the InferSent embeddings' size using PCA. I took the first 100 principle components, however singular values suggestion we may be losing significant information. Captions are now size 100 embeddings. Model didn't seem to learn anything in training.