Overview

For this assignment I struggled to figure out the correct structure due to my misinterpretation of the assignment. Once learning it was intended for me to implement S2VT specifically instead of my own seq2seq model I was able to get things completed more efficiently.

My model outputs grammatically flawless captions of identical format to the training set, however, sometimes the model will be confused and not know how to describe something. For instance, there is a video of a polar bear and the model thinks it is a white dog. Or a video of a woman mixing melted butter is described in detail by the model as actually being mango pulp and whipped cream. Regardless of its inaccuracies I'm incredibly proud of its ability to produce confident, grammatically accurate captions.

Structure

The model is trained in `train.py`, this script expects the MSVD dataset to be located in a directory called `data`, and it also expects a word_tokens.json file to exist. This file is generated by running `gen_json.py` and contains the mapping of an id to a word token. The model itself is in `module.py`, and helper functions for converting the model output into an actual human readable string are contained in `convert_caption.py`.

I've saved a few models but the best functioning one is `s2vt_params.pkl`. `s2vt_params_oops.pkl` was trained the longest and probably has captions that fit the subject matter more often, but while training that model I accidentally removed the first word of every caption from the caption dataset, so the caption it outputs might be lacking a starting word like "A" or "A woman".

Instructions

To run the script `cd` into the `hw2` directory, and run `hwseq2seq.sh` with the first parameter being the directory containing the dataset, and the second parameter containing the filename of the desired output.

Example:

```
› ./hw2_seq2seq.sh ./data output.txt
```