Reproduction of a Pragmatic Model by Nie et al. on CUB and A3DS Datasets

Project for the class "Neural Pragmatic NLG" by M. Franke - Winter Semester 2023, University of Tübingen

Uliana Vedenina
uliana.vedenina@student.uni-tuebingen.de

Hoa Do
hoa.do@student.uni-tuebingen.de

Inessa Iliadou
inessa.iliadou@student.uni-tuebingen.de

1 Introduction

The combination of probabilistic models of pragmatic reasoning with natural language generation algorithms makes it possible to achieve enhanced flexibility in solving neural pragmatic language generation tasks, as it considers both the context and communicative goals of the outputted message. The pragmatic-based approaches can be introduced directly during the training phase or during inference. As part of the latter approach, Nie et al. (2020) introduced an Issue-Sensitive Image Captioning algorithm, which follows pragmatic needs by specifying the task-relevant information. The model includes three pragmatic agents: the speaker (image-captioning algorithm), and both sensitive-to-issue extended pragmatic listener and speaker (specifically, considering image partitions). The algorithm was evaluated on the CUB dataset (Welinder et al., 2010).

2 Re-implementation of the original study

As part of our course project, our aim was to reproduce the ISIC model on the CUB dataset. The time taken for the sentence classifier training on free Google Colab GPU totaled about 8 hours, and the pragmatic captioner took about 3 hours to train. The results of the default evaluation metrics are the following:

$Bleu_1$	0.8477
$Bleu_2$	0.6706
$Bleu_3$	0.5055
$Bleu_4$	0.3772
$METEOR$	0.2914
$ROUGE_L$	0.5759
$CIDEr$	0.5196
$SPICE$	0.2147

Because the authors of the original paper did not include performance evaluation on the metrics present in their code, it is not possible to compare our results with theirs. Based on Bleu scores, the model predicts the unigrams rather well, while the results significantly drop with the increase of n-grams lengths. Among other metrics, the model performs well on ROUGE and CIDEr, signifying the relative closeness of the generated caption to the reference. Surprisingly, SPICE shows low results, the reason for it may be the incomplete description of the scene graph provided in the output.

3 Test of the ISIC model on the A3DS data set

After testing the model on the original data set, we attempted to adapt the ISIC algorithm to work with the A3DS data set by Burgess and Kim (2018). The following data set contains the images of 3D objects located in the colored space and related captions. The implementation required several changes to be made. Firstly, a *3d* option was added to the `arg_parser.py` file. Next, the `data_prep.py` file was modified to include an elif statement that accepted the option "3d" for `dataset_name`. Additionally, adjustments were made to the `get_dataset()` and `get_dataloader()` methods to accommodate the changes in the constructor of the A3DSDataset class. In the `model_loader.py` file, a couple of lines were removed since the `set_label_usage()` method was not implemented in the A3DSDataset class. Minor changes were also made to `train_epoch()`the function of the `sentence_classifier_trainer.py`.
It was decided that it was better to only return the wordtargets, lengths, and labels from the `get_item()` function in the A3DSDataset class

(they were actually the only values used in the `train_epoch()`). Some tensor sizes in the `train_step()` function were also adjusted to be more suitable for the criterion() method.

To facilitate the training process, the `a3ds_dataset.py` file was created. This file is called when the dataset and dataloader objects are created in the main.py file. It contains the `get_item()` and `collate_fn` functions that are necessary during the training process.

Due to frequent disconnections of the Google Colab runtime, only a sentence classifier was trained for 10 epochs. Unfortunately, a checkpoint recovery was not implemented, making it impossible to continue training from a specific checkpoint. Attempts were made to train a pragmatic captioner using the A3DSDataset class and the sentence classifier that was trained, but multiple issues were encountered, including a batch size that was too large. Due to time constraints, the decision was made to stop attempting to train the pragmatic captioner and continue with providing the repository and the report on what was done.

References

Chris Burgess and Hyunjik Kim. 2018. 3d shapes dataset. https://github.com/deepmind/3dshapes-dataset/.

Allen Nie, Reuben Cohn-Gordon, and Christopher Potts. 2020. Pragmatic issue-sensitive image captioning. *arXiv preprint arXiv:2004.14451*.

Peter Welinder, Steve Branson, Takeshi Mita, Catherine Wah, Florian Schroff, Serge Belongie, and Pietro Perona. 2010. Caltech-ucsd birds 200.