Instructions for *ACL Proceedings

Anonymous ACL submission

Abstract

1 Introduction

TODO:

- how

2 Background and previous work

3 Materials and methods

TODO:

- creation of dataset (CLEVR)
  - multiple 'real' objects in scene
  - 3 attributes (color, size, shape) differentiate objects
  - using 'dale' setup to uniquely identify target object
- building a language game using EGG
- based on feature extractors ResNet/VGG
-
- setup of discriminating game of objects in image
- message encoder/decoder is auto-encoder
-

4 Results

TODO:

- DaleTwo:
  - small hidden/embedding dims, small vocab -> high accuracy
  - high hidden/embedding dims, small vocav -> low accuracy
- DaleFive:
  - small hidden/embedding dims, small vocab -> low accuracy
  - small hidden/embedding dims, bigger vocab -> higher accuracy
- ... test different hidden/embedding dims
- ... test 3/4 objects

5 Discussion

TODO:

- maybe calculation of loss (multiplicating instead of summing loss per token), unlikely, since sequence length short -> shouldn't result in big differences
- reducing dims of image better the increasing dims of message, increasing dims is not learnable for models
- Vocab:
  - vocabulary could describe attributes of target image (non-discriminative) or describe only differences (discriminative)
  - in second case, two images is a far easier task than five images. Hence, much lower accuracy

6 Conclusions and further work

Acknowledgements

A Example Appendix