

Introduction to Image caption

做一些图像描述的入门任务

综述

一些benchmarks

- BLEU-n系列, 源自NLP中的n-gram模型, 对句子进行滑窗处理:

$$BLEU_n(infer, groundTruth) = \frac{\text{推理结果中以}n\text{为滑窗得出与真值对应相同的子串数}}{\text{真值以}n\text{为滑窗得出的子串总数}} \quad (1)$$

公开datasets

- Flickr30k
- SBU
- Coco

经典之作

Show&Tell, CVPR2015, arxiv:1411.4555

- basic idea:

- *maximize the likelihood $P(\text{Sentence}|\text{Image})$*
- *CNN, as decoder \rightarrow embedding Image to a fixed-length vector, a kind of middle representation \rightarrow RNN, as encoder*

$k = 1$

前沿