分工

1．Flickr上数据读入

2．图片特征补零

3．使用Faster R-CNN提取图片特征

4．注意力机制的可视化

5．Web平台搭建的相关接口定义

阶段1 对ATT-NIC模型进行参数调优

1 在Flickr8k上不同优化算法损失下降曲线如图所示。

SGD Moment SGD RMSProp Adam损失下降曲线

2 在Flickr8k和Flickr30k上不同网络结构对模型性能的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | Beam size | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr8k | VGG |  |  |  |  |  |  |  |
| ResNet-50 |  |  |  |  |  |  |  |
| ResNet-101 |  |  |  |  |  |  |  |
| Flickr30k | VGG |  |  |  |  |  |  |  |
| ResNet-50 |  |  |  |  |  |  |  |
| ResNet-101 |  |  |  |  |  |  |  |

3 beam size对生成语句的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | Beam size | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr8k | 1 |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |
| Flickr30k | 1 |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |

4 Flickr30k上字典大小对模型性能的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 字典大小 | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr30k | 4000 |  |  |  |  |  |  |  |
| 5000 |  |  |  |  |  |  |  |
| 6000 |  |  |  |  |  |  |  |

5 词向量编码长度对模型性能的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 编码长度 | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr30k | 400 |  |  |  |  |  |  |  |
| 500 |  |  |  |  |  |  |  |
| 600 |  |  |  |  |  |  |  |

6 LSTM隐层神经元数目对模型性能的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 神经元数目 | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr8k | 3 |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |
| Flickr30k | 3 |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |

7 LSTM初始状态生成中，隐层神经元数目

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 神经元数目 | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr8k | 300 |  |  |  |  |  |  |  |
| 400 |  |  |  |  |  |  |  |
| 500 |  |  |  |  |  |  |  |
| Flickr30k | 400 |  |  |  |  |  |  |  |
| 500 |  |  |  |  |  |  |  |
| 600 |  |  |  |  |  |  |  |

8 对LSTM输出进行解码生成单词中，隐层神经元数目

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 神经元数目 | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr8k | 256 |  |  |  |  |  |  |  |
| 512 |  |  |  |  |  |  |  |
| 1024 |  |  |  |  |  |  |  |
| Flickr30k | 256 |  |  |  |  |  |  |  |
| 512 |  |  |  |  |  |  |  |
| 1024 |  |  |  |  |  |  |  |

阶段2 对ATT-NIC模型进行改进

1对图像特征补一列0，这样模型在生成单词时并不总是需要参考图片

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 模型 | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr8k | ATT-NIC |  |  |  |  |  |  |  |
| Ours |  |  |  |  |  |  |  |
| Flickr30k | ATT-NIC |  |  |  |  |  |  |  |
| ours |  |  |  |  |  |  |  |
| MS COCO | ATT-NIC |  |  |  |  |  |  |  |
| ours |  |  |  |  |  |  |  |

2图像特征提取部分采用Faster R-CNN，选最可能有物体的前29个区域连同原图一起生成特征向量

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 模型 | Blue-1 | Blue-2 | Blue-3 | Blue-4 | METOR | ROUGE | CIDEr |
| Flickr8k | ATT-NIC |  |  |  |  |  |  |  |
| Ours |  |  |  |  |  |  |  |
| Flickr30k | ATT-NIC |  |  |  |  |  |  |  |
| ours |  |  |  |  |  |  |  |
| MS COCO | ATT-NIC |  |  |  |  |  |  |  |
| ours |  |  |  |  |  |  |  |

日志

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 完成时间 | 任务描述 | 完成人 |
| 1 | 4月2日 | Flickr上数据读入相关代码，包括Flickr8k和flickr30k | 成思邈 |
| 2 | 4月3日 | 完成evaluate相关代码中因python版本不同导致的错误，包括numpy.int64无法使用json.dump(改为numpy.int64.item())；使用subprocess.Popen时，stdin.write()后stdout无返回导致的死锁(stdin.flush())；以及python2和python3编码字符串的方式不同导致的str转bytes操作(decode()/encode())；default dict类型在python2中使用iteritem()，pytho3中改为item() | 吴黄子桑 |