浙江大學

ZHEJIANG UNIVERSITY

Visual Question Answering

DL4NLP Final Project

蒋景伟

Data: 2022-06-26

目录:

1 Project Introduction

1.1 选题

- Visual Question Answering

1.2 工作简介

- 以一张图片和一个关于图片内容的自然语言形式的问题作为输入，要求输出正确答案
- 在数据集`VQAv2`上进行训练与测试
- 属于一种多标签分类的问题，计算损失的时候采用多标签损失。

1.3 开发环境

- 开发工具：`ModelArts Ascend Notebook`环境，选用`Ascend910`芯片作为训练芯片
- 开发包、开源库：
    1. Mindspore1.3.0
    2. numpy
- 系统运行要求：

  `python3.7.5`与可运行`Mindspore1.3.0`的开发环境

2 Technical details

VQA整体pipeline

2.1 DataLoader

2.1.1 数据集简介

使用`VQAv2`数据集，分为`image`，`question`和`annotation`三个大的数据集，官网数据量如下：

	train	val	test
image	82783	40504	81434
question	443757	214354	447793
annotation	4437570	2143540	—

我们使用的是课程要求的数据集，数据量如下：

- train: 44375，validation: 21435，test: 21435

数据集的构成如下：

- 1张图片有大概5个问题
- 1个问题有10个答案
- test没有annotation文件

下面具体介绍问题与回答的数据构成：

1 Question

根据官网的解释可知，**question**被保存为`JSON`文件的格式，其具有数据结构如下：

```
question{
"question_id" : int,  #问题id
"image_id" : int,     #问题对应的图片id
"question" : str      #具体的问题
}
```

2 Annotation

根据官网的解释可知，**Annotation**也被保存为`JSON`文件的格式，其具有数据结如下：

```
annotation{
"question_id" : int,
"image_id" : int,
"question_type" : str,          #问题类型
"answer_type" : str,            #答案类型
"answers" : [answer],
"multiple_choice_answer" : str
}
------------------------------
answer{
"answer_id" : int,
"answer" : str,                 #具体答案
"answer_confidence": str
}
```

2.1.2 数据预处理

1 数据对齐

首先是对问题与答案进行对齐，剔除没有一一对应关系的问题或答案，该部分代码在`match_align/align.py`中，运行后发现所有问题答案均已对齐。

2 图片与问题对齐

前面提到，一张图片对应5个问题左右，需要将图片与问题对齐，剔除没有一一对应关系的问题或图片，该部分代码在`match_align/match.py`中，运行后发现所有训练集中有部分问题没有对应图片，因此去除该部分的问题，重新整理后写回`JSON`文件。

2.1.3 数据集加载

由任务可知，我们需要把数据集加载成`img`，`question`，`answers`的形式，数据类型与预处理如下：

1. `img`：图片为三通道`RGB`模式，加载成三维的`Tensor`即可
2. `question`：预处理，进行词形还原，大小写转换等，再通过预训练的`Tokenizer`进行`one-hot`编码，扩充成定长向量输出。
3. `Answers`：预处理，进行词形还原，大小写转换等，自己构造词汇表进行`one-hot`编码

自定义数据集需要重载两个函数：

1. `__getitem__`：根据输入的下标选取对应的数据
2. `__len__`：获取数据集长度

完成后加载为`DataLoader`格式，调用相关函数即可，这里可以灵活设置`batch_size`与图片的增强模式（在`config.py`中完成定义即可）。

最终每一个`batch`的数据如下所示：

2.2 img特征提取

内容包括：

（1） 工程实践当中所用到的理论知识阐述

（2） 具体的算法，请用文字、示意图或者是伪代码等形式进行描述（不要贴大段的代码）

（3） 程序开发中重要的技术细节，比如用到了哪些重要的函数？这些函数来自于哪些基本库？功能是什么？自己编写了哪些重要的功能函数？等等

介绍使用的网络等

2.3 text特征提取

2.4 特征融合与预测网络

3 Experiment Results

3.1 实验结果

系统界面、操作说明、运行结果

3.2 总结说明

4 References