Paper Review

Qilong Wang

**Paper Method:**

The paper mainly uses attention mechanism to achieve the communication between picture features and question features. It has shown very good results and collaborative attention has also performed well in image and text representations. However, this collaborative attention mechanism can only learn rough interactions between multiple modalities, and the collaborative attention is not enough to infer the relationship between images and question keywords. In order to solve the problem that cooperative attention cannot fully interact between multiple modalities, co-attention models can be connected in tandem to solve deeper visual reasoning. The authors believe that the reason is that the models do not model dense self-attention in every modality at the same time.

We intended to use BAN-VQA as our basic model at first, but we failed to use its code. After that, we decided to us another well-performed model, MCAN-VQA, to achieve our original goal.

**Development Process:**

It contains two processes: BAN-VQA and MCAN-VQA.

The development process of BAN-VQA is stated below. It’s easy to import the github code. We met the first problem when I found the memory in SCC is not enough for installing the libraries and dataset.

After contacting with our professor two times, the library problem is solved. There are pre-installed libraries in SCC so we can use them directly by using the following command:

$ module load python

Besides, we can use the following command to check which version of the library can be loaded:

$ module av python

Use the following command to unload the library

$ module unload python

After solving the library problem, we met another problem that the dataset is much larger than 16GB which means that we cannot use the personal space to run the code. The situation we encountered at that time was: command wget did not report error, but the downloaded file was damaged. I tried several times, finding that the file size was same, but much smaller than the size It should be. Finally, I found that there is a limitation of self-holder.

After contacting with IT help center, we found the solution that we can use the class space. I duplicated the code repository to the class folder, and it can download the dataset successfully.

There was still problems when we tried to run the code. The situation was stated as follows:

文本

描述已自动生成

It seems that a dictionary in python was null, causing the error. I think it may be caused by the mistakes during the importing process. I did not find how to solve it, so we had to change the model we used.

Development process of MCAN-VQA:

We found another well-performed model, MCAN-VQA. I downloaded the dataset from a Baidu-Netdisk. This time I run the code on my own computer (Windows System).

I changed some file path in the code into windows path. When running the code, there was an error in run.py that yaml.load(f) reported an error. After searching on the internet, I found that the new version of pyyaml did not support this method. I changed the code into yaml.safe\_load(f). It can run successfully.

The final result is stated as follows.

**Result:**

In this project, we successfully gave a mcan-vqa model which can offer answers to the questions about the pictures. The result is pretty well, which achieved 81.22% accuracy.

文本

描述已自动生成

**Poster:**

During the poster display, there were about 10 people who were interested in our work. Most of them cared more about the method of model. Some of them were more interested in what the VQA is and where it can be used.