Proposal for CSC413 Final Project

Author: Yulin Wang, Tian Ze Jia, Hongwei Wen, Shijia Wu

Model tasks:

The task aims to train a RNN model that receives a picture as input then generates a reasonable caption for that picture. Also to investigate whether grouping the image first helps with the quality of caption.

Since we want to challenge the capacity of the model, we want to build a model that is not limited to the elements in the pictures. Therefore, we choose images that contain creatures, and images that only contain items instead of creatures. We hope the model we build can tell the difference between different images accurately.

Model explained:

We will train two different models based on two different dataset, one is for creatures, the other is for items. Each model will contain a couple of CNN layers followed by an attention mechanism to embed our image, followed by an RNN layer(transformer). Then we will train another model to classify whether the input image is a creature or an item, then put it to the correct model described above to generate the final caption.

We will also combine the two dataset to form a single huge dataset and try to train a single model based on this large dataset to compare the results with the separate one.

Dataset outline:

After discussion, we decided to do Image Caption. We went to github and wanted to find a dataset which has a big amount of images. After we reviewed a couple of datasets, we had decided on two datasets which contain different objects. The pictures in one of the datasets are all about creatures. The other one is about items.

In each dataset, we can divide them into three groups, which are training set, validation set and testing set. Therefore, we can use an exact dataset to train the model or test the model. It can help us build a better model which is more efficient.

There are two folders in each dataset. The dataset regarding creatures, one of the folders includes 8091 images that we need to use. The other one contains the text that each image stands for. Also, in the text folder, each phrase corresponds to the serial number of the associated photo. The dataset regarding items, image folder includes 15807 images. And other one folder includes the texts that correspond to those photos.

Dataset

The first dataset contains a folder with 8091 photographs in JPEG format and a text file with their corresponding captions. All these photographs contain at least a person or an animal. Among these 8091 photographs, 6000 images are divided for training and 1000 for testing and 1000 for validation.

The second dataset contains 15807 photographs in JPEG format. All these photographs contain merchandise like carpets, stamps, and jewelry. Among these 1000 images are used for testing and 2000 images are used for validation and the rest will be used as training.

Ethical implications:

First of all, image caption is a function that we use in people's usual lives. Although machine learning nowadays is powerful enough to do captions for many images, it is hard for a model to take parameters that it never learned about or something abstract. Therefore, we certainly do not recommend anyone who purely and fully relies on this model for commercial use. This is a disclaimer to all people that we are not responsible for any kind of personal business loss over this machine learning model.

People may use this model to generate useless comments on social media like Instagrams and Twitter, this will make the poster confused and waste their time on replying to this.

Project tasks division:

Tian Ze Jia: Forming graphs and pictures with the model prediction results and analyzing, relate the partial results to real-world phenomena and demonstrate the power of the model by using examples.

Yulin Wang: Analyzing and summary to write README document, writing proposal, Testing data and providing information to teammates about the dataset. testing and analysis parameters.

Hongwei Wen: Build model, tune parameters, research on how to use RNN package and relate to course contents, compare different RNN models with results.

Shijia Wu: collecting data, cleaning and adjusting data, analyzing the results come through the model to determine whether the models we build fit on these datasets.