Creating Visual Language Reasoning Data Set using ShapeWorld Package

Chi Duan

INTRODUCTION

This script uses a package named ShapeWorld [1] to generate a visual language reasoning data set. The package ShapeWorld can be accessed from GitHub [1]. The package generates abstract images consisted by shapes (square, circle, triangle, etc.) in different colors. A (language) caption about the image is also generated. The package can generate images in batches. We first generate the images and then a create manually curated text paragraph, question and answer choices based on each created image. A sample generated by the script is like the following:
Image:

Paragraph: If someone add one more circle to the image, there will be two circles in total in the image.
Question: Determine whether the passage is a valid argument for the given image.
Answer: A. True; B. False
Correct Answer: True

METHODOLOGY

The manually created formats are used to generate text paragraph, question, answer choices. The formats can be categorized into five main types based on the methods to generate the texts from the image. Currently we generate arbitrary 207 instances in the data set. However, the number of data instances in each category can be adjusted in the script by the user (details could be found in the script's comment).

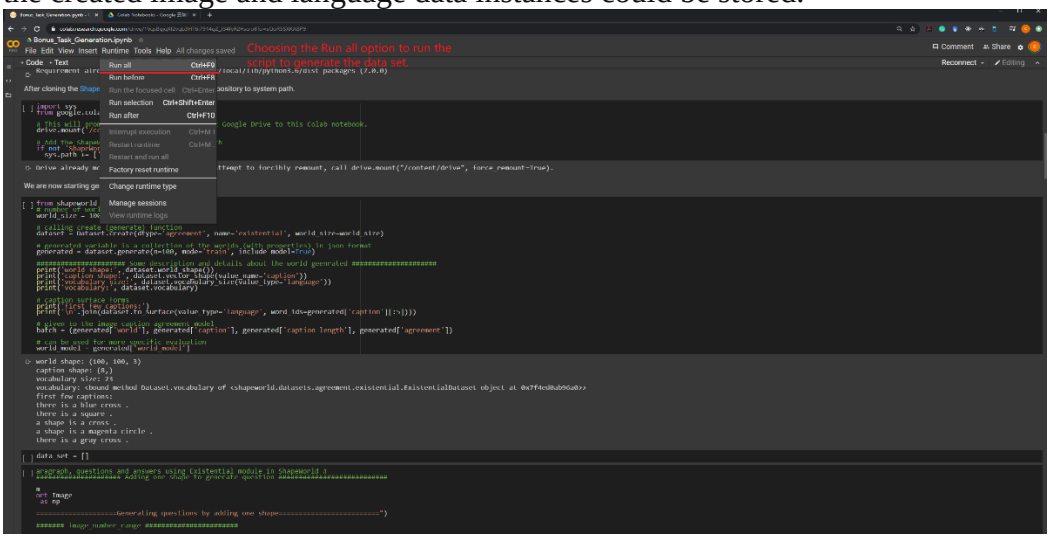
1. Generation by adding one more shape:
   With this generation method, we add one more quantity of a shape to the image and generate two types of paragraph. One is correctly describing the image marked with true after the addition and the other is falsely describing the image marked with false. We created the paragraph by manually rephrasing the wording in many ways. The question is to ask whether the paragraph is correct.
2. Generation by adding one more color:
   This generation method follows the same way as generation by adding one more shape but in a way to add one more color to the image.
3. Generation by replacing one shape with another shape:
   This generation method generates text by assuming that one shape is replaced by another shape in some quantities in the image. The text generation then follows different rephased formats that either give a true or false description of the image after replacing the shape.
4. Generation by replacing one color with another color:
   This generation method generates text by assuming that one color is replaced by another color in some quantities in the image. The text generation then follows different rephased formats that either give a true or false description of the image after replacing the shape.

5. Generation using relational caption:
   In this method, we use the caption ground truth generated by the ShapeWorld package [1] as a comparing target to generate kinds of true or false statement through various formats.

In order to simplify the dataset generation process due to time limitation, all the answers in this generated data set has a uniform two-option format of ["True", "False"] instead of four-options case.

USAGE OF THE SCRIPT

The script is created using Google Colab. Colab is a cloud-based Python notebook service provided by Google. A Colab tutorial could be found here https://www.geeksforgeeks.org/how-to-use-google-colab/. In order to run the script, one needs to register for a Google account which will be used to log in to Colab. Then one needs to mount Google Drive as the source drive where the created image and language data instances could be stored.



If the Google Drive in not mounting correctly and error may be given. For details of mounting Google Drive to Colab note book, please refer https://colab.research.google.com/notebooks/io.ipynb#scrollTo=BaCkyg5CV5jF.

IMPLEMENTATION DETAILS

The script is implemented in Python. The following lists all Python packages used in this script.

1. numpy
2. pandas
3. pillow
4. random
5. os
6. shapeworld

Please make sure all packages are installed before running the script.

CONCLUSION

This project using an automated method to generate a visual language data set. The creation process is fulfilled using a Python script. The GitHub Repository is at the address [2]. The script runs in Google Colab environment. There are five types of data instances corresponding to five types of creation method. The images are created using the Python package ShapeWorld [1]. The

text paragraph and questions are created using manually curated formats. User can adjust the number of data instances in each format in the script as instructed in the script comment. We saved all documents in a GitHub repository [2].

REFERENCE

1. ShapeWorld package: https://github.com/AlexKuhnle/ShapeWorld
2. The GitHub link for this bonus task script: https://github.com/duanchi1230/Automated-Vision-Language-Reasoning-Data-Set-Creation