Capita Selecta AI - Topic 3: Application study

Initial idea: Visual Question Answering

Jorrit Willaert *(r0652971)*

DeepProbLog offers the ability to integrate probabilistic knowledge with deep neural networks. This way, the strength of neural networks (system 1: typical subconscious tasks such as visual recognition, the processing of languages, ...) is complemented with the strength of rule-based mechanisms (system 2: slow, sequential thinking such as the derivation of a proof). I propose an application that requires the integration of both systems.

The Sort-of-CLEVR dataset is a simplified version of the CLEVR dataset [2]. This simplified dataset is composed of 10 000 images with per image 20 accompanied questions. An image consists of spread out objects, with randomly chosen shapes and colors. The questions are divided in two categories: non-relational and relational questions. Non-relational questions ask for example about the shape, the horizontal or vertical location of the colored object. Relational questions, on the other hand, ask about the shape of the object which is closest (or furthest) to a certain colored object, or ask about the number of objects with the same shape [1].

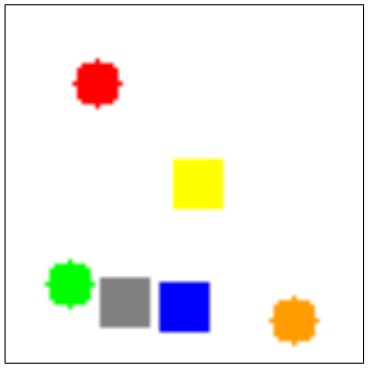

Figure 1: A sample image from the Sort-of-CLEVR dataset [1]

A sample image is given in Figure 1. With this sample image, an answer on a non-relational question such as "What is the shape of the blue object?" would be: "square", while an answer on a relational question such as "How many objects have the same shape as the blue one?" would be: "3".

I propose to generate data based on the Sort-of-CLEVR dataset, with the simplification that objects may not be entirely randomly located, but instead only in certain positions on a grid. A 5-by-5 grid could for example be used. The generation of new images and accompanied questions is rather straightforward, which could probably even be done at training time.

I would use 25 convolutional neural networks, one for each grid position. The output would, naturally, indicate if an object is located in that position, and what its shape and color is. Once the location, shapes and colors of the objects are known, rule-based mechanisms can be used to answer the accompanied questions.

References

[1] Kim Heecheol. kimhc6028/relational-networks: Pytorch implementation of "A simple neural network module for relational reasoning" (Relational Networks).

[2] Justin Johnson, Li Fei-Fei, Bharath Hariharan, C. Lawrence Zitnick, Laurens Van Der Maaten, and Ross Girshick. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:1988–1997, 2017.