Accessible Learning with Recursive Neural Networks

Patrick Green

Fourth Year Project Report
School of Informatics
University of Edinburgh

2021

Abstract

Table of Contents

0.1 Introduction

Before the eruption of Youtube studying different disciplines in academia was mainly practiced through reading and writing. Youtube has given a greater incite to how we learn new concepts which would take considerably longer to understand from text based descriptions. In 2005 it was though that 65% [1] of the population were visual learners. So its no surprise that the curious of today search for the description of a concept on Youtube.

Work has been done on training machines to parse images to text whilst conserving the semantics. However a reverse of this procedure for schematics and diagrams would aid the learning of that concept to 65% of people. Furthermore previous work on summarizing text could be enhanced as analogous situations would resemble similar schematics and diagrams. In the subject of physics, analogies give a gateway to the understanding of new concepts allowing people apply prior experience. With a large enough set of different analogies easier concepts to grasp subjectively could be used to help aid the understanding of almost anything, whether it be in a article or a textbook.

The current proposal suggests directions to explore the use of machine learning to generate schematics and diagrams for a given written description in such a way that the message that its putting across is easier to consume. It summarizes previous work and existing methods in the area and then proposes the steps to gather data to train a model.

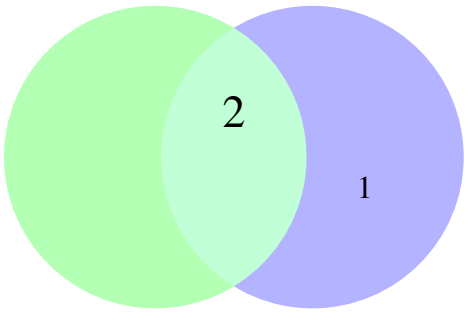
0.2 Background and Related Work

A large amount of work by B.Z Yao et al [3] has been done visualizing the knowledge in an image. The method consists of first visually identifying separate objects and

second constructing an And-Or Graph from the relation of the objects. This extend to its spacial position, how it bounds with the others and also the fact that its there in relation to all the other objects. Using the source Wordnet, categorical and hierarchical relations between the objects was used to derive the use that one imposes on the other. From which a written description of the image could be derived. Later a paper by O Vinyals et al [2] demonstrated that a convolutional neural network gave a much faster method for extracting entities. By having a trainable kernel they had the ability to eliminate potential objects between steps as increasingly more detail was gained from the image. Furthermore a image description could be derived either by appending words of some connectivity as they were found. Or by eliminating from a list, implausible sentences as the details increased.

0.3 Proposal

This proposal suggests exploring the use of machine learning to construct diagrams from a collection of entities with a written description on how they interact. The method would entail collecting a large data set of diagram and text pairs in which both the diagram and the text are sufficient on there own to describing the topic in question. I would then go about indexing the entities and key words used in both the diagram and the text.



Unlike the 1, the 2 can live on the land and ocean.

Detected:
$Circle(210:1.2):[inside(2), outside(1)...]$,
$Circle(330:1.2):[inside(1,2), outside()...]$

Finally I would then train a Neural Network on the placement labeling and connectivity of the entity and keyword placeholders by analyzing the shapes drawn in the diagram and semantics of the description in the text.

Bibliography

[1] Pearson. Why the blank stare? strategies for visual learners, 2005.

[2] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2015.

[3] Benjamin Z Yao, Xiong Yang, Liang Lin, Mun Wai Lee, and Song-Chun Zhu. I2t: Image parsing to text description. *Proceedings of the IEEE*, 98(8):1485–1508, 2010.