Learning Programs for Modeling Strategy Differences in Visuospatial Reasoning

James Ainooson
Department of Computer Science, Vanderbilt University, 400 24th Ave S
Featheringill Hall, Room 254
Nashville, TN 37212
james.ainooson@vanderbilt.edu

Introduction

You are presented with the three cards shown in Figure 1(a), and without any specific instructions, or selection criteria you are asked to pick a card. What choice will you make? Now, let us assume a second set of cards, shown in Figure 1(b), are subsequently presented to you. You may start reorganizing your thoughts and realize the task was for you to pick the card bearing the odd image, and likewise after the third set of cards from Figure 1(c).

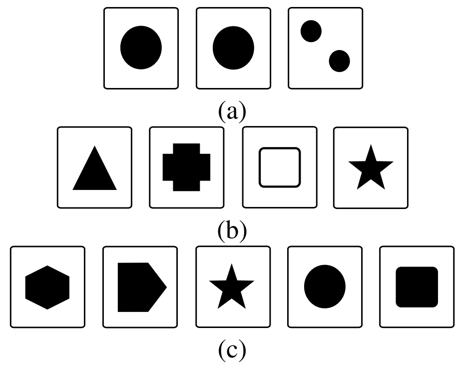
The task we just worked through can be considered as an instance of the "odd-one-out" task. In this particular formulation, one thing stands out: although there is no verbal communication of the goal, anyone performing the task, in most cases, will be able to form a strategy that leads to a goal after observing the sequence of the problems presented.

But another question remains, can this strategy of identifying the odd one in the image be the only way to solve this problem? Certainly not. Maybe you may have noticed that each "odd" item in the list also has a characterstic that relates it to the number of cards presented. For example, in Figure 1(b) there are four items and the odd card bears a square.

Features of this task, as presented, exemplify significant aspects of my dissertation work. Succinctly, the problem I am working on can be stated as follows: Humans can form strategies for novel tasks, seemingly without effort (Mumford et al. 1993), but current AI systems cannot, at least with the level of flexibility shown by people (including young children in many cases). How can we gain further insights about factors that affect strategy choices in intelligent systems from models generated through program synthesis?

Research Questions

As expressed in the problem statement above, I am studying the possible mechanisms people may be using when they face certain novel visuospatial reasoning tasks. My work is using ideas from program synthesis to equip intelligent agents with the ability to form generalizable strategies for reasoning through the tasks they face. These systems are built to rely on visual imagery as their core knowledge representation, making image operations (such as rotations, translations, and scaling) the basis of reasoning. With the tools to synthesize programs that represent strategies for these systems in place, I will further explore the possibility of fitting programs to human performance to help understand the factors that drive a person's choices as in the performance of the tasks I am studying.



(a)

(b)

(c)

Figure 1: A simple task presented as a series of cards

My dissertation is organized around answering the following research questions.

1. For intelligent systems that are to solve specific visuospatial reasoning tasks using imagery as representations, what kinds of operations and strategies are sufficient for various levels of performance?
2. Provided a program synthesis approach is taken as a means to generate strategies for intelligent systems that reason through specific visuospatial problems, what kinds of search techniques can be applied to ensure the production of strategies?
3. How can the techniques adopted from question 2 above be applied to fit programs on human performance for the purposes of understanding their strategies on standard visuospatial reasoning tasks?

Methods

The agents produced in this work will be evaluated on standardized visuospatial reasoning tasks. I will be using the Punched-hole paper folding task (Ekstrom and Harman

1976), the Block Design Test (Kohs 1920), The Leiter International Intelligence Scale-Revised (Leiter-R) (Roid and Miller 1997), which is a battery of 20 different visuospatial reasoning subtests, and the Abstract Reasoning Corpus (ARC) (Chollet 2019), which is a collection of 1000 different reasoning tasks.

A bulk of my work goes into exploring the structure of search spaces that represent programs—which can also be considered as strategies—agents use for reasoning about the tasks in my task domain. These spaces are expressed as programs in a state machine based language I am developing, called Visual Imagery Reasoning Language (VIMRL). With programs in VIMRL being state machines, every program contains sets of states, edges for transitioning between states, and rules that govern transitions. Further, each state has an attached list of instructions, known as the state script, which are executed whenever the state is active. Transitions between states are conditioned on the side-effects of actions performed in these state scripts.

Machines in VIMRL can be considered as nodes in a search graph. Through different sets of rules that systematically add new VIMRL elements (like states, instructions, and transitions) to an existing state machine, successor nodes can be generated from a given machine. I have used traditional search algorithms like breadth first search to explore the dynamics of this search space on simplified problems from the block design test. I have additionally applied the more advanced Monte-carlo tree search algorithm to searching for programs around the ARC task.

Current Work

Currently, I have performed the following activities:

1. I have conducted a number of experiments on the sufficiency of imagery as a representation for reasoning. Much of this work has been through experiments on the punched hole paper folding task (Ainooson and Kunda 2017), the block design task and the Leiter-R (Ainooson et al. 2020).
2. I have taken an early look into how strategy differences in intelligent agents can be studied through experiments on the block design task and the Leiter-R. For these experiments I built an integrated environment for conducting visuospatial reasoning experiments, called the Visuospatial Reasoning Environment for Experimentation (VREE) (Ainooson et al. 2020). This is an environment in which virtual agents can reason about and interact with virtual objects.
3. Considering how expressive the VIMRL search space is, any exploration quickly explodes in nodes explored. To keep this explosion in check, I have worked on a collection of logical rules that are able to strip any malformed and invalid machines from the search space. Additionally, I have worked on an ordering algorithm that makes it possible to compare any two programs that are logically equivalent, regardless of how they are structured. Results of these improvements can be seen in Figure 2, which shows the progress made in terms of nodes expanded at increasing search depths for different combinations of pruning and ordering.

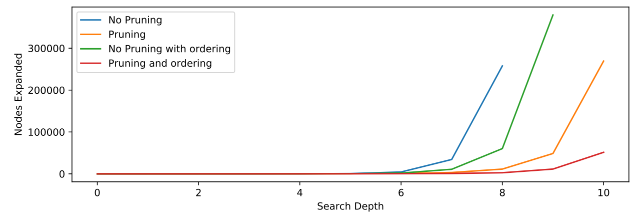


Figure 2: A plot showing improvements in program synthesis search performance with pruning rules and ordering.

Proposed work

1. Finish my implementation of VIMRL, including releasing all my code and documentation for use by the broader AI research community.
2. Improving the search algorithms used for synthesizing programs in VIRML, with the goal of having decent scores on the ARC and performance on block design.
3. I will be investigating factors that affect strategy on the Block Design Test by running ablation experiments on agents in VREE.
4. Finally, I intend to investigate how programs can be generated from traces of human performance on the Block design Test. From these traces one goal will be to find programs that are consistent with the behavioral pattern of an individual (or a group of individuals). My methods for analyzing human performance using program synthesis will be evaluated on data collected using virtual assesments (as part of another project ongoing in my research lab) from large samples (n=500 in each group) of neurotypical adults and adults on the autism spectrum.

References

Ainooson, J.; and Kunda, M. 2017. A Computational Model for Reasoning About the Paper Folding Task Using Visual Mental Images. In *CogSci*.

Ainooson, J.; Michelson, J.; Sanyal, D.; Palmer, J. H.; and Kunda, M. 2020. Strategies for Visuospatial Reasoning: Experiments in Sufficiency and Diversity. 20.

Chollet, F. 2019. On the Measure of Intelligence.

Ekstrom, R. B.; and Harman, H. H. 1976. *Manual for kit of factor-referenced cognitive tests, 1976*. Educational testing service.

Kohs, S. C. 1920. The Block-Design Tests. *Journal of Experimental Psychology*, 3(5): 357–376.

Mumford, M. D.; Baughman, W. A.; Threlfall, K. V.; Uhlman, C. E.; and Costanza, D. P. 1993. Personality, Adaptability, and Performance: Performance on Well-Defined Problem Solving Tasks. *Human Performance*, 6(3): 241–285.

Roid, G. H.; and Miller, L. J. 1997. Leiter international performance scale-revised (Leiter-R). *Wood Dale, IL: Stoelting*.