Learning Domain-General Reasoning by Exclusion with Neural Networks

Aaron Traylor

April 2019

1 Introduction

The ability to reason by exclusion– given A or B, and not A, predict B– is a fundamental element of logical inference. Adult humans can trivially demonstrate this reasoning pattern across a wide range of modalities and extend inference to previously unknown objects and words. However, logical inference using novel symbols is challenging for machine learning methods that do not explicitly manipulate variables, particularly neural networks, despite the number of recent advances in the field of deep learning. Generalizing algebraic rules to withheld symbols is a classic difficulty for neural networks [4][5][3], and approaches that combine neural networks with symbolic reasoning are an active area of research.

While this reasoning process is trivial for adults, infants do not immediately demonstrate reasoning by exclusion across domains, and have been found to display generalization of logical inference across modalities in a staggered manner. Infants first demonstrate competence in a word learning with novel objects task at 18 months of age[2]. However, they are able to leverage extremely similar skills in the physical domain to reason logically about object search tasks as early as 12 months of age[1]. It may be the case that first learning the generalizable process in the world grounded by physical limitations leads to its use in more open-ended language use. We hypothesize that "scaffolding" neural networks without specific architecture for logical inference by first training on concrete out-of-domain tasks will increase the speed of generalization to novel symbols on more abstract logical inference tasks.

2 Methods

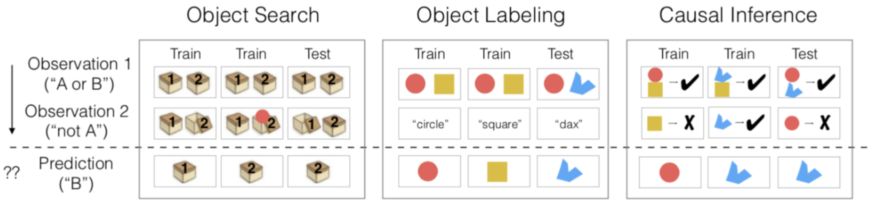
We define three subtasks of increasing abstract challenge, as seen in Figure 1. These subtasks model experiments used to demonstrate competency of 12-18 month old children at reasoning tasks. In each experiment, information is shown to the child sequentially- whether this information must be input to the neural network sequentially or all at once will be determined later.

The first subtask is object search. Given that a ball is in one of two closed boxes, and one box is then opened and revealed to be empty or to contain an alternative object, the infants successfully search the closed box when prompted. To parallel this, we may be able to abstract away from the physical or visual domains so long as the *restriction* is clear within the domain. In this problem setting, the concrete limitations of physical space lead to the disjunctive syllogism.

The second task is word learning- when shown a ball and an novel object simultaneously, and told to ”look at the ’dax’”, the child will attend to the previously unseen object, thus implying assumption of reference, despite having heard the word before. For the neural network, it may be necessary to define the input space such that word knowledge can be easily hand-embedded or to define another subtask where the meaning of some words is learned to precede this task.

The third subtask is a causal inference task. The network is first presented with two objects, and then with one of the specified objects and another truth statement pertaining to that object specifically, and must determine which is the ”correct” output by making deductions.

Figure 1: Visual depiction of the three subtasks.



We are interested in using basic neural networks– specifically feedforward networks, and then potentially recurrent neural networks or convolutional neural networks if sequential or visual information are eventually required by the subtasks respectively.

We will perform transfer learning by first observing that the network trained on a subtask generalizes successfully to withheld input symbols in the test set, and then using the learned weights as initialization for a network that will be trained on the next task.

3 Evaluation

We will evaluate the full scaffolding pipeline by comparing the number of epochs that the model takes to generalize to the word learning and causal inference test data with random initialization versus with the weights learned from the object search task as an initialization. The hypothesis will be said to be supported if there is a significant decrease in training batches with the pretrained weights compared to random initializations.

A model of one of the subtasks will be said to generalize successfully if it achieves a 100% accuracy rate on test data containing symbols that are not available to it at training time, potentially alongside symbols from the training data.

4 Anticipated Challenges

Generalizing to novel symbolic input is an unsolved problem in the field of neural network methods. The most challenging task will be defining the input representation for each task. If each input symbol is represented as a singular bit, the model will not be able to represent new unseen symbols at test time. However, the fact that the model is presented with unique *objects* must be represented in some capacity– it is potentially unsatisfying to simply embed a symbol in a feature space without maintaining its comparative uniquity. Experimenting with different methods of representing objects as dense input to the neural network and then configuring input representations such that weights can be shared across all three subtasks will likely take the bulk of the time on this project.

5 Timeline

- September 2019 - November 2019: Create synthetic dataset for object search task and examine generalization to novel symbols.

  It is unclear if the network will be able to perform this task satisfactorily. If it is not, during this time, I will examine why this phenomenon occurs, find the minimum structure necessary to succeed at the object search task by generalizing to novel symbols, and change the course of the research if need be. This step is expected to be the most challenging.

- November 2019 - January 2020: Create the data and evaluate the same neural network structure on the word learning and causal inference tasks.

- February 2020 - March 2020: Create transfer learning setups and evaluate if scaffolding improves speed of generalization.

- Defend in April 2020.

References

[1] Nicoló Cesana-Arlotti et al. “Precursors of logical reasoning in preverbal human infants”. In: *Science* 359.6381 (2018), pp. 1263–1266.

[2] Justin Halberda. “The development of a word-learning strategy”. In: *Cognition* 87.1 (2003), B23–B34.

[3] Junkyung Kim, Matthew Ricci, and Thomas Serre. “Not-So-CLEVR: visual relations strain feedforward neural networks”. In: (2018).

[4] Gary F Marcus et al. “Rule learning by seven-month-old infants”. In: *Science* 283.5398 (1999), pp. 77–80.

[5] Paul Tupper and Bobak Shahriari. “Which Learning Algorithms Can Generalize Identity-Based Rules to Novel Inputs?” In: *arXiv preprint arXiv:1605.04002* (2016).