

# Crítica: Inferring and Executing Programs for Visual Reasoning

En este paper, los autores afirman que los métodos de razonamiento visual que intentan hacer un mapeo directo de input a output principalmente aprenden a explotar los sesgos presentes en los datos por sobre aprender el razonamiento necesario para la tarea. Dentro de este ámbito, el principal aporte del paper es una red entrenable *end to end*, con un módulo generador de programas y otro que representa un motor de ejecución.

Es muy interesante que incorporen representaciones simbólicas a una red neuronal, para luego formar un árbol sintáctico usado para responder la pregunta asociada. Me gustaría recalcar como ponen énfasis en la analogía a crear un lenguaje de programación, pero para ser escrito y usado por una red neuronal, donde esta misma aprende las distintas funciones del lenguaje. Creo que este es un concepto muy poderoso que puede acercarnos a cosas como “computadores neuronales”, que pueden generar una secuencia de operaciones para resolver un problema dado.

Aun así, una limitante importante que veo de su modelo es que la traducción directa de la pregunta a una secuencia de instrucciones puede no funcionar muy bien si la pregunta no se ve estructurada como una instrucción en lenguaje natural (como el quinto ejemplo de la figura 7 del paper), ya que implicaría ir de una oración sin una estructura marcada a otra que requiere de una estructura específica. Esto mismo es algo que se ve (pero desde el punto de generalización sistemática) en el paper “Generalization without systematicity” [1], y el módulo generador de programas puede sufrir de los mismos problemas que se presentan en este paper.

Por otra parte, es impresionante como su modelo es capaz de generalizar bien con poca supervisión de los programas que genera. Esto lleva a pensar que puede ser muy útil poder supervisar una parte específica de la red (aunque sea con pocos datos), y luego hacer *finetuning* de esa parte de forma *end to end*. Un procedimiento así suena similar a *curriculum learning*, pero en vez de entregar los datos progresivamente según nivel de dificultad, se entregan los datos según que parte de la red se está entrenando, lo que podría lograr un aprendizaje más profundo del razonamiento necesario.

A pesar del problema descrito, los resultados de este paper son increíbles. El hecho de que hayan logrado superar el rendimiento humano en esta tarea muestra que su metodología es muy relevante para este tipo de problemas. Ahora solo queda ver si se puede extender a otros dominios, y como se traduce su rendimiento en ellos.

- [1] B. Lake and M. Baroni, “Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks,” *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 7, pp. 4487–4499, 2018.