

NOMBRE: Benjamín Farías Valdés

N.ALUMNO: 22102671



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

IIC3692 — Tópicos Avanzados en Inteligencia Artificial — 2' 2022

Lectura 11

Crítica

Compositional Attention Networks for Machine Reasoning

Este trabajo presenta un nuevo modelo de redes neuronales, que combina los conceptos de literatura previa en ámbitos de memoria, atención y composición. En la introducción se presenta el *framework* sobre el que se trabajará, que corresponde a *CLEVR*, un *dataset* de *VQA* (responder preguntas a partir de la información proporcionada por una imagen). Los autores indican que los modelos actuales (redes modulares altamente supervisadas, *stacks* de *CNNs* y *LSTMs*, entre otras), si bien logran gran precisión en el *dataset*, lo hacen aprendiendo patrones superficiales sobre los datos y no aprenden a razonar de forma lógica y estructurada como lo hacen los humanos. El grueso del artículo consiste en la explicación detallada de la nueva arquitectura propuesta, denominada red *MAC*. La red recibe de entrada la información sobre la pregunta y la imagen (procesadas con redes recurrentes y convolucionales, respectivamente), y luego entra al núcleo de la arquitectura: una red recurrente compuesta por celdas *MAC*, para finalmente utilizar la salida de dicha red (y la pregunta) al momento de elegir la respuesta con una capa densa. La gran innovación del *paper* viene en las celdas *MAC*, donde cada una separa internamente un estado de control y uno de memoria, de forma que la red sea capaz de trabajar con la información que va aprendiendo (secuencialmente), pero sin mezclar el acceso a la información memorizada con el razonamiento sobre la pregunta. De esta forma se logra emular más de cerca el razonamiento humano, que al enfrentarse a un problema, va atacándolo parte por parte y accediendo sólo a la información necesaria desde la memoria en cada paso. Al mirar los resultados de los experimentos, es completamente cierto que el aprendizaje es mucho más eficiente y efectivo que las otras redes del estado del arte, puesto que logra superar la precisión de los mejores rivales con un tiempo mucho menor de entrenamiento. Esto se puede interpretar como que la red realmente está condicionada para razonar de forma estructurada y composicional, incluso logrando buenos resultados usando tan sólo el 10% de los datos de entrenamiento. La propuesta también trae la ventaja de que posee mayor explicabilidad, puesto que es posible visualizar el comportamiento de sus capas de atención y verificar las zonas de la imagen y de la pregunta que están siendo utilizadas en cada paso de la red recurrente. Este fue uno de los artículos que más me gustó en lo que va del curso, me parece muy interesante el cambio desde modelar redes basadas en los sentidos de las personas a modelarlas basándose en el razonamiento inteligente que realiza el cerebro. De cierta forma se puede extrapolar que la investigación en el área se dirige hacia el objetivo de simular el pensamiento humano, directamente inspirando cada nuevo modelo a partir de algún aspecto empíricamente determinado sobre la forma en que funcionan nuestras propias neuronas.