

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 10

Crítica

Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks

Este artículo se centra en el tema de las habilidades composicionales, que permiten a los humanos entender y producir una cantidad prácticamente infinita de conocimiento mediante combinaciones de información previamente aprendida. En la introducción se presenta la hipótesis de que el estado del arte en *RNNs*, si bien logra un buen rendimiento y generalización, requiere de patrones a gran escala sobre los datos para poder entrenarse y aprender correctamente, a diferencia de las personas, que son capaces de inferir mucho conocimiento a partir de pocos datos (gracias a sus habilidades composicionales). Para realizar experimentos, se presenta un *dataset* denominado *SCAN*, que asocia comandos en lenguaje natural a secuencias de acciones de navegación. Los modelos a ser probados corresponden a variantes de *RNNs Seq2Seq* (que permiten codificar la información de entrada y luego decodificarla para generar un *output*). A continuación se explicaron los experimentos realizados, que incluían generalizar muestras aleatorias de comandos, generalizar comandos con secuencias largas de acciones, generalizar composición sobre comandos primitivos y composición aplicada a traducción de idiomas. De los resultados se puede concluir que la hipótesis es medianamente correcta, ya que en ciertos casos las redes funcionan bien al generalizar cosas similares a las vistas en su entrenamiento, pero fallan catastróficamente cuando la conexión entre los datos de entrenamiento y testing depende de la extracción de reglas sistemáticas, donde las redes buscan patrones superficiales sobre los datos y no profundizan lo suficiente como para encontrar reglas lógicas de la forma que lo hacen los humanos. Esto último es un tema de investigación que me parece muy importante, dado que si se logra dotar a estos modelos de la habilidad de encontrar reglas sistemáticas a partir de pocos datos, se podrían llegar a entrenar sin requerir de *datasets* gigantescos para conseguir un buen rendimiento (hoy en día los mejores modelos para generalizar son también increíblemente demandantes en recursos computacionales). El *paper* me pareció bien útil como punto de partida y motivación para abrir una nueva línea de investigación, que se encargue de explorar las habilidades sistemáticas que podrían integrarse en estas arquitecturas para aprovechar la lógica inherente que compone la semántica del lenguaje humano.