

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 4

Crítica

Language Models are Few-Shot Learners

Este trabajo toma un enfoque más general dentro del área de procesamiento del lenguaje natural, buscando probar distintas hipótesis respecto del comportamiento de modelos generales del lenguaje (agnósticos a sus posibles aplicaciones) que tienen una gran cantidad de parámetros. El modelo utilizado para representar esta clase de arquitectura es *GPT-3*, que tiene 175 billones de parámetros. La introducción está bien organizada, presenta la nueva tendencia de buscar modelos NO específicos a las tareas deseadas, sino que más bien generales, que luego puedan ser acomodados a distintas aplicaciones sin entrenarlos desde cero. Se menciona que la técnica a utilizar en el artículo (para acomodar el modelo a distintas tareas) será la de *few-shot learning* (y otras variantes como *one-shot* y *zero-shot*). No se describe en mucho detalle la arquitectura del modelo, lo que no es problema ya que el enfoque del paper no está en el modelo mismo (y se incluye esa información en el anexo). Tras presentar el *dataset*, se explican los diversos experimentos realizados, evaluando la habilidad del modelo para encargarse de muchas tareas distintas sin necesidad de entrenarlo nuevamente o hacer *fine-tuning*. Entre estas tareas destacan el completar oraciones, generar texto dentro de un contexto dado, responder preguntas, traducción y actividades de comprensión de lectura. Todos los experimentos, así como sus resultados, están bien documentados, logrando un trabajo completo que muestra cómo la propuesta se comporta bastante bien en casi todos los casos (en algunos incluso supera al estado del arte). Se evidencia el hecho de que los modelos de lenguaje en el futuro probablemente se beneficien sustancialmente de tener una gran cantidad de parámetros, de forma que puedan aprender mejor a generalizar las reglas y semánticas de dicho lenguaje. Algunas secciones me parece que se podrían haber organizado un poco mejor, ya que son muy cortas respecto del resto del artículo (por ejemplo la parte de memorización de los datos y las limitaciones podrían estar juntas en algún capítulo más extenso). Finalmente, me gustaría mencionar que me pareció muy interesante la parte final del paper, donde se habla sobre el impacto social que tendrán estos tipos de modelos en el futuro cercano, así como de los problemas que surgen (sesgos dañinos en los datos, falsificación de documentos, etc...), lo que representa un aspecto importante que no suele ser mencionado en este tipo de artículos.