

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 19

Crítica

Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks

El paper propone un algoritmo para aplicar *meta-learning* a cualquier tipo de modelo que utilice *SGD* para optimizar sus parámetros, permitiendo aplicar un *framework* general para *few-shot learning*.

La innovación de este aporte, respecto a otros trabajos de *meta-learning*, es que decide utilizar el método clásico para optimizar con *SGD*, pero en vez de aplicarlo en cada tarea de forma individual, se aplica buscando minimizar la pérdida conjunta de múltiples tareas distintas. La idea detrás de esto me hace mucho sentido, ya que se está forzando al modelo a aprender a generalizar sobre varias tareas juntas, en vez de especializarse en sólo una. Una vez logrando esta generalización de parámetros para varias tareas, se puede extrapolar que el modelo se podrá entrenar de manera eficiente para nuevas tareas (con *fine-tuning*), siempre y cuando estas nuevas tareas estén relacionadas a las que ya ha visto el modelo en entrenamiento (supuesto válido, ya que el contexto de aplicación del modelo será elegido según su uso esperado).

En mi opinión, uno de los aspectos más destacables de esta propuesta es el hecho de ser agnóstico a la arquitectura, dado que abre las puertas para investigar una forma general de entrenar modelos de inteligencia artificial que sean multi-propósito (es decir, similares a los humanos). Los mismos autores mencionan en el trabajo futuro que se espera este sea un primer paso en esa dirección, lo que también está respaldado por los favorables resultados obtenidos en los experimentos. Algo que si me parece que podría ser una limitante es el tener que preparar el entorno de entrenamiento. Esto debido a que se necesita un conjunto no menor de tareas que estén relacionadas, y para tareas que no hayan sido abarcadas mucho en la literatura es posible que sea necesario dedicar bastantes recursos y tiempo en prepararlas. Otra limitación es que, si se desea usar esta estrategia en modelos gigantes, sigue siendo un problema el entrenarlos.

Respecto a los experimentos, creo que, si bien la propuesta logra superar al estado del arte en ciertos *datasets* y tareas, esto no es realmente lo importante. En su lugar, lo importante es que el mismo algoritmo puede ser adaptado a tareas muy distintas, desde regresión hasta aprendizaje reforzado, pasando por clasificación. Esto demuestra que, efectivamente, es posible que la mejor forma de acercarse al aprendizaje humano sea unificando las ideas detrás de lo que significa aprender en cualquier tipo de tarea y contexto.

En general me pareció un artículo bien importante, que claramente inspiró un nuevo enfoque dentro del área en estos últimos años. A lo largo del curso hemos visto varias arquitecturas que utilizan alguna forma de *meta-learning*, y buscan entrenar modelos que sean buenos generalizadores más que especialistas (trabajando en problemas de *few-shot learning* y *zero-shot learning*, por ejemplo con el caso de *GPT-3*).

Todo esto refleja un cambio de enfoque en el área, dado que se han empezado a encontrar limitantes con el *approach* clásico de utilizar modelos cada vez más grandes. Entre estas limitantes se encuentra el hecho de que es demasiado demandante en tiempo y recursos el entrenar estos modelos gigantes, por lo que debemos buscar otra forma más eficiente de hacer esto. Aquí es donde me parece que una posible solución es buscar formas de aprendizaje basadas en los humanos, que es justamente lo que se ha estado investigando últimamente (arquitecturas composicionales, *meta-learning*, *few-shot learning*, *prompting*).