

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 7

Crítica

The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

El paper se enfoca en materias de optimización de espacio y rendimiento computacional de redes neuronales. En la introducción se habla del surgimiento de técnicas de *pruning* (poda), que son utilizadas para eliminar una proporción de los parámetros totales de la red, sin afectar en gran medida a la calidad del modelo y a su capacidad para generalizar. La novedad de artículo corresponde a una hipótesis, la que indica que dentro de redes neuronales densas se encuentran sub-redes más pequeñas que son capaces de alcanzar el mismo rendimiento (o incluso mejor en algunos casos), consumiendo menos espacio de almacenamiento/memoria y demorando menos en entrenar (estas sub-redes son denominadas *winning tickets*). El cuerpo del artículo detalla una variedad de experimentos, realizados sobre distintas arquitecturas (redes densas, convolucionales), los que evidencian la existencia de los *winning tickets*, y como su capacidad de aprendizaje está estrechamente ligada tanto a su estructura como a los valores iniciales de sus parámetros. De esta forma, se propone una interesante estrategia de entrenamiento: encontrar las sub-redes podando la red original, y luego re-entrenar solamente la sub-red pero partiendo de los mismos valores iniciales que tenían sus parámetros durante el entrenamiento de la red original. Se muestra que con estas estrategias se logra una capacidad de generalización igual o mejor que la red original, demorando menos tiempo en entrenar y ocupando menos espacio. También se logra corroborar el hecho de que las redes más grandes aprenden mejor porque en su interior contienen más combinaciones posibles de sub-redes que son potenciales candidatos para *winning tickets*. Toda esta sección me llamó bastante la atención, ya que nunca había pensado en lo importante que pueden llegar a ser los valores iniciales de los parámetros de las redes (al entrenar), dado que se escogen de forma aleatoria. El descubrimiento principal aquí es que al decrecer la densidad de las redes (a tamaños que normalmente no logran muy buenos resultados), los valores iniciales de los parámetros cobran mucha mayor importancia que en sus contra-partes más densas, puesto que ya no tienen el lujo de poseer una gran cantidad de combinaciones en su interior que podrían llegar a ser favorables. Finalmente, se concluye que aún queda mucho trabajo en esta vía, principalmente en pulir las técnicas de análisis y descubrimiento de *winning tickets* en arquitecturas cada vez más complejas, lo que encuentro perfectamente razonable. En general, encontré muy interesante el artículo, da harto de qué pensar desde un ángulo nuevo. Lo único que encontré algo denso al leer fue la parte de los experimentos, que me pareció un poco desordenada, en particular por explicar resultados y experimentos de forma textual a través de varios párrafos seguidos, lo que no aportaba más que resumirlo en algunas tablas (y así dejar más texto para las ideas y la discusión).