

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 8

### Crítica

#### Overcoming Catastrophic Forgetting Using Sparse Coding and Meta Learning

El paper ataca el problema conocido como *Catastrophic Forgetting* (olvido catastrófico, *CF*), que se refiere a la pérdida de rendimiento de una red sobre una tarea previamente entrenada, al momento de entrenarla sobre otra tarea nueva. En la introducción se presentan las soluciones del estado del arte, donde se destacan estrategias que congelan partes de la red para reservarlas para tareas ya aprendidas, además de técnicas de *Meta Learning* para alentar el uso de pesos compartidos por varias tareas en la red (generalizar a tareas futuras). Esta primera parte del artículo es bien clara y me ayudó a entender el problema a pesar de no haberlo visto antes en ningún otro trabajo. A continuación se presenta la solución propuesta, denominada *GoCaT*, que combina lo mostrado en la introducción de forma que se logren reservar ciertos pesos de la red para tareas específicas, y al mismo tiempo, tratar de generalizar otros pesos a varias tareas (transferir el conocimiento). Esta sección la encontré muy interesante, me gustó como se fue presentando todo de forma ordenada e incremental, comenzando por explicar en palabras y figuras los conceptos aplicados, y luego entrar en los detalles matemáticos posteriormente (así el lector puede prestar más atención a la idea en vez de perderse analizando fórmulas). En la última parte del artículo, se muestran los resultados para varios experimentos realizados sobre *datasets* de *Continual Learning* en el contexto de reconocimiento de imágenes, evidenciando el gran poder de este nuevo *approach* comparado con el estado del arte (al que supera en todos los casos). Encontré buena la agrupación de los gráficos y tablas con los resultados, de forma que no queden separados entre medio de párrafos del texto que podrían estar hablando de otras cosas (como pasa con algunos papers). En general el artículo me gustó bastante, recordando un poco a trabajos previamente leídos sobre *few-shot learning*, pero agregando ahora mayor robustez al entrenar a los modelos de forma secuencial. También me llamó la atención lo mencionado en la conclusión, sobre la posibilidad de entrenar varias redes específicas a cada tarea y combinarlas mediante otra red que sea capaz de generalizarlas, lo que quizás podría tener cercanía con la forma en la que nuestro propio cerebro maneja el conocimiento, memoria y experiencias a través del tiempo, llevando a aprendizaje continuo y natural.