



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
IIC3697 APRENDIZAJE PROFUNDO

PROYECTO DE CURSO

Fecha Máxima de Entrega: 10 de Julio

Estimad@s, llegó la hora de comenzar el proyecto de curso, en el cual podrán poner en práctica los conocimientos que han adquirido este semestre sobre aprendizaje profundo o *deep learning* (DL). Dado que Aprendizaje Profundo es un curso nivel 3000, es decir, nivel postgrado, como parte del curso también deben ejercitar su capacidad para realizar estudio independiente e investigación en esta área.

En esta oportunidad podrán desarrollar su proyecto en grupos de 2 o 3 integrantes. Tienen libertad para formar los grupos, si alguien necesita ayuda, por favor, enviar email al profesor.

1 Descripción

Recientemente, grandes modelos de language basados en IA han demostrado avanzadas capacidades cognitivas. Algo similar ha ocurrido en el ámbito del reconocimiento visual, donde modelos de aprendizaje profundo han demostrado un alto grado de versatilidad y exactitud. Estos modelos están basados en la codificación de información (conocimiento) contenida en la representación simbólica del language o en los píxeles de las imágenes utilizadas en la fase de entrenamiento. Una de las controversias existentes en la comunidad de IA es si la representación aprendida por estos modelos se basa en la memorización de patrones superficiales [Geirhos et al., 2020] o en la inferencia de reglas generales (gramática) que determinan el comportamiento o asociación de estos patrones [Chan et al., 2022]. El proyecto de curso está dedicado a estudiar esta disyuntiva. Específicamente, el proyecto de centrará en estudiar como, bajo distintas condiciones de entrenamiento y test, un modelo de aprendizaje profundo resuelve una problemática de aprendizaje, ya sea en base a memoria, extracción de reglas, o una combinación de ambos.

Para el desarrollo del proyecto usted deberá elegir un modelo a su elección (solo uno), y optar por el uso de uno de los siguientes datasets:

- Scan: <https://github.com/brendenlake/SCAN>
- Omniglot: <https://github.com/brendenlake/omniglot>

Si bien pueden elegir el modelo que deseen, se recomienda lo siguiente:

- Scan es un dataset en base a texto, por ende, modelos basados en LSTMs o Transformer pueden ser apropiados. Puede ser un modelo entrenado desde cero o un modelo pre-entrenado que se refina (fine-tuning) para la tarea de SCAN.
- Omniglot es un dataset de imágenes, por ende, modelos basados en CNNs o Visual Transformer pueden ser apropiados. Puede ser un modelo entrenado desde cero o un modelo pre-entrenado que se refina (fine-tuning) para la tarea de Omniglot.

Existen diversas formas de medir el tipo de aprendizaje que logra un modelo de aprendizaje profundo, las cuales pueden servir para realizar su proyecto de curso. Como guía, acá hay 3 de ellas (más detalles en clases):

- **Transfer Learning:** memorización vs aprendizaje de reglas se puede medir mediante el grado de generalización del modelo a ejemplos novedosos no vistos durante entrenamiento. Específicamente, el conocimiento aprendido en base a reglas debería generalizar mejor a estos casos.
- **Contextual Prompt:** memorización vs aprendizaje de reglas se puede medir en términos del conocimiento incorporado durante test en el prompt. Específicamente, si el prompt no contiene suficiente información para resolver el problema, este debe estar memorizado en los pesos de la red.
- **Proof classifier:** el tipo de conocimiento guardado en distintas partes del modelo puede ser accesado mediante el uso de un clasificador simple que opere sobre el embedding objetivo.

2 Lecturas recomendadas

Como punto de partida es importante que realicen una revisión bibliográfica que les permita familiarizarse con la temática del proyecto de curso y así definir en qué enfocar su trabajo. En particular, acá unas lecturas que deben revisar y cuyo resumen crítico debe ser parte de su informe de avance:

- R. Geirhos et al., Shortcut learning in deep neural networks. Nature Machine Intelligence, 2020. <https://arxiv.org/abs/2004.07780>.
- B. Lake and M. Baroni, Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. Proceedings of ICML 2018. <https://arxiv.org/abs/1711.00350>.
- S. Chan et al., Transformers generalize differently from information stored in context vs in weights, ArXiv, 2022. https://memari-workshop.github.io/papers/paper_21.pdf.

3 Nivel de Complejidad Esperado

El proyecto es una parte importante del curso, tanto en términos de su metodología como de su sistema de evaluación. IIC3697 es un curso de 10 créditos semanales, de acá en adelante parte sustancial de este tiempo debe ser dedicado al desarrollo de un buen proyecto de curso. A grandes rasgos es posible identificar 4 posibles niveles de complejidad para el proyecto. En orden creciente de complejidad:

- (a) Implementar y evaluar algún método existente, ya sea algún algoritmo o modelo visto en clases o alguno que ustedes hayan encontrado en su revisión bibliográfica. En su reporte debe incluir un análisis de como este modelo codifica parte de su solución en base memoria o extracción de reglas.
- (b) Proponer, implementar y evaluar alguna mejora a algún método existente. En este caso, la comparación sería con respecto a cómo la modificación cambia la forma como el modelo codifica su solución usando memoria o extracción de reglas.
- (c) Combinar varios métodos con objeto de crear un sistema más robusto (ensamble) y ver como esto afecta la forma de codificar la solución usando memoria o extracción de reglas.
- (d) Realizar una innovación relevante a un método existente o desarrollar un nuevo modelo y ver como esto afecta la codificación de la solución usando memoria o extracción de reglas.

Lograr un buen nivel en cualquiera de los puntos anteriores garantizará una buena calificación. En todos los casos, especialmente en los puntos a) y b), será fundamental respaldar el estudio con un adecuado análisis de los resultados obtenidos, en términos de análisis de sensibilidad respecto a distintas condiciones de operación, y tipos de instancias de entrenamiento y test. En este sentido, limitar el proyecto de curso a bajar una herramienta de software de la web y probar su rendimiento sin un análisis crítico adecuado, no es suficiente y será evaluado en forma deficiente.

Importante: si bien el proyecto de curso se establece en modo desafío (challenge), es decir, todos trabajan sobre un mismo problema, tengan presente que el obtener un buen resultado cuantitativo es una característica deseada pero no el objetivo central del proyecto de curso. Lo central es que desarrollen un trabajo que aporte a su formación en las temáticas del curso, por ende, el énfasis es en efectuar un adecuado análisis de su implementación y resultados del proyecto. Así, un trabajo que obtenga un excelente resultado cuantitativo, pero consista en simplemente copiar una implementación pública, será penalizado con una nota deficiente.

4 Fechas Importantes

Informe de Avance. Fecha entrega: Lunes 19/06

Este informe de avance no tiene nota pero será tomado en cuenta al momento de evaluar el informe final. El objetivo de este informe es demostrar que el desarrollo del proyecto se ha realizado en forma sistemática durante el período de tiempo asignado. El formato de entrega es un breve documento en pdf indicando los avances a la fecha. Como mínimo, este informe debe incluir: i) Un breve resumen de las lecturas indicadas en la sección 2, y ii) Un archivo con los avances del código implementado a la fecha, no importa que aún no sea funcional.

Entrega Final. Fecha entrega: Lunes 10/07

El informe final debe ser documentado en formato paper. En la página web del curso se publicarán templates en latex y MS-Word que pueden ser usados para este fin. El documento final debe incluir las siguientes secciones:

- Breve introducción describiendo las aristas principales de su propuesta.
- Revisión bibliográfica que incluya al menos 2 trabajos relevantes distintos a los indicados en la sección 2.
- Descripción del modelo propuesto, incluyendo un diagrama de bloques y una breve descripción de cada bloque.
- Resultados experimentales.
- Descripción de las principales conclusiones de su trabajo.
- Detalle de las referencias en formato paper.

Adicionalmente, deben entregar un jupyter notebook con su código y los archivos de su modelo.