

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

Propuesta de Proyecto

Introducción

En los años recientes, la aparición de modelos con grandes cantidades de parámetros (como *GPT-3*, que tiene varios billones) ha traído consigo tanto ventajas como desventajas al momento de aplicarlos en la práctica. Las ventajas apuntan a un gran rendimiento en tareas de generalización sobre los datos que aprenden, evidenciando el potencial de tener arquitecturas a gran escala. Por otro lado, las principales limitaciones aparecen respecto a la gran cantidad de información, memoria y tiempo de procesamiento que estos modelos demandan al momento de ser entrenados. En este contexto surge la técnica conocida como *prompting*, que será el enfoque de este proyecto.

Prompting

Esta técnica consiste en condicionar un modelo pre-entrenado para que aprenda a realizar una nueva tarea, tan sólo mostrándole ejemplos sobre esta tarea como parte de la entrada y salida del modelo. Es decir, existe un cambio de enfoque: en vez de que el modelo se ajuste a los datos (entrenamiento y *fine-tuning*), ahora ajustamos los datos al modelo generalizador. Dada la naturaleza de este enfoque, se puede clasificar como una forma de *few-shot learning*, ya que mediante unos pocos ejemplos, se pueden extender las funcionalidades de un modelo pre-entrenado. En cuanto a los *approaches* existentes, se tiene que para modelos no tan grandes (*BERT*, *RoBERTa*), se busca aplicar *fine-tuning* basado en *prompts*. Para modelos enormes (*GPT-3*, *T5*), dado el gran costo de ajustar sus parámetros internos, el enfoque está en congelarlos y añadir una cantidad pequeña de parámetros en la entrada y salida, los que serán utilizados para modular el modelo pre-entrenado.

Multimedia

Actualmente, existen modelos pre-entrenados capaces de realizar tareas de alto rendimiento en los ámbitos de *NLP* y visión. Para dar algunos ejemplos:

- *BERT* y similares en tareas de clasificación sobre texto.
- *GPT-3* (y *GPT-2*) en modelamiento del lenguaje.
- *StyleGAN* en generación artificial de imágenes.
- *CLIP* en modelamiento conjunto del lenguaje e imágenes.

Por otro lado, en el ámbito del audio, también existen variadas aplicaciones, tales como reconocimiento de voz, generación de audio a partir de texto y clasificación de audio.

Propuesta

Para este proyecto, se propone investigar diversos modelos pre-entrenados como los anteriormente mencionados, y buscar formas de adaptar sus capacidades a nuevas tareas utilizando *prompting*. En particular, es de interés seguir enfoques multimodales, donde sea posible relacionar distintos dominios mediante nuevas tareas, las que aprovecharán una *setting* de *prompt learning* para no tener que entrenar los parámetros internos de ninguno de los modelos. Algunas ideas:

- Aprovechar modelos como *CLIP* para adaptarlos a nuevas tareas que relacionen imágenes con texto.
- Combinar modelos de texto con *StyleGAN* para generar imágenes a partir de alguna inferencia sobre lenguaje natural.
- Probar nuevas tareas que relacionen audio con texto, o audio con imágenes.
- Combinar audio, texto e imágenes para alguna tarea de interés, utilizando *prompting* en algún paso de la cadena.

La propuesta general será trabajar con modelos pre-entrenados y aplicar *soft prompting* para poder adaptarlos a nuevas tareas sin tener que re-entrenarlos o hacerles *fine-tuning*. A medida que se avance en la investigación, se decidirá por perseguir un tema más específico entre las ideas mencionadas, dependiendo de la viabilidad de cada una. Lo ideal sería perseguir un tema multimodal, pero si esto prueba tener un *scope* demasiado alto, se optará en su lugar por un enfoque en un sólo dominio (posiblemente *NLP*).

Bibliografía

- [1] Baevski, A. et al. (2020). *wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations*.
- [2] Brown, T. et al. (2020). *Language Models are Few-Shot Learners*.
- [3] Karras, T., Laine, S., Aila, T. (2018). *A Style-Based Generator Architecture for Generative Adversarial Networks*.
- [4] Lester, B., Al-Rfou, R., Constant, N. (2021). *The Power of Scale for Parameter-Efficient Prompt Tuning*.
- [5] Radford, A. et al. (2021). *Learning Transferable Visual Models From Natural Language Supervision*.
- [6] Wang, Z. et al. (2021). *Learning to Prompt for Continual Learning*.