

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 21

Crítica

Representation Learning with Contrastive Predictive Coding

El paper propone un modelo universal para aprendizaje no supervisado denominado *Contrastive Predictive Coding* (*CPC*), que se basa en aprender las características de los datos mediante predicciones del futuro que se puedan derivar del contexto temporal de la información de entrada.

La idea principal de este enfoque es aprovechar el contexto temporal que existe en los datos, buscando aprender características que son trascendentes a distintas partes de la secuencia de entrada. Esto de cierta forma requiere que efectivamente existan estas características en los datos, lo que no siempre será cierto. Por ejemplo, en los experimentos realizados en el contexto de un agente de aprendizaje reforzado, se encontró que una de las tareas no se beneficiaba de usar *CPC*, justamente porque la tarea a realizar no dependía del conocimiento temporal. Por lo tanto, es importante saber de ante mano, al menos de forma intuitiva, si es que la actividad de interés puede explotar el contexto temporal.

El otro aspecto que hace funcionar este enfoque es la función de pérdida utilizada. Los autores proponen una pérdida contrastiva basada en ruido, que básicamente busca optimizar la relación entre el contexto actual y el de los siguientes pasos temporales al mismo tiempo que busca evitar predicciones erróneas (usando muestras de posibles futuros, donde sólo una es correcta y las otras son ejemplos negativos). Este método de optimización es la principal innovación del artículo, dado que el resto de la propuesta se basa en utilizar arquitecturas ya existentes para codificar los datos y realizar auto-regresión. Esta función de pérdida es adecuada a la formulación de la propuesta, lo que se ve evidenciado en su buen rendimiento en distintas tareas, tales como reconocimiento de voz, clasificación de imágenes, lenguaje natural y aprendizaje reforzado en entornos 3D. Entre estos, me gustaría destacar el de aprendizaje reforzado, dado que los buenos resultados obtenidos (en términos de acelerar el aprendizaje del agente), reflejan la utilidad de usar la función de pérdida contrastiva por sí sola.

La principal ventaja de este enfoque respecto a otros de aprendizaje no supervisado es el hecho de que es muy flexible y sencillo de adaptar a prácticamente cualquier tarea y arquitecturas asociadas. Esto nuevamente indica que el interés actual en el campo de la inteligencia artificial está enfocado en simplificar las formas de aprendizaje, buscando modelos transferibles y fácilmente generalizables.

En términos generales, me parece una interesante propuesta para tareas que puedan aprovechar el contexto temporal, como lo son el campo del audio, los vídeos, lenguaje natural y simulaciones con aprendizaje reforzado. A mi parecer, la propuesta es presentada como un paso temprano dentro de una posible línea de

investigación, por lo que aún hay mucho que explorar dentro de esta idea de predecir el futuro al entrenar estos modelos. Por ejemplo, una primera idea para explorar en el futuro podría ser la auto-generación o extensión de videos (también audio), dado que esto podría utilizar la misma idea de predecir el futuro para aprender características y luego usar esas características para generar contenido.