

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 16

### Crítica

#### Prototypical Networks for Few-Shot Learning

El paper presenta una arquitectura denominada *prototypical networks*, que permite entrenar redes de forma que sean capaces de rendir bien en un contexto de *few-shot learning* posterior.

El problema que atacan los autores es de clasificación, donde su idea principal es encontrar un prototipo para cada clase en entrenamiento, de forma que la red aprenda a agrupar rápidamente los ejemplos de entrada según sus características. La estrategia que utilizan es similar a realizar *clustering*, ya que la red se va configurando para aprender a juntar ejemplos de la misma clase, usando la media entre sus representaciones obtenidas como el punto central que representa a cada prototipo de clase. Esta idea funciona muy bien para este tipo de problemas discriminatorios, ya que las características que definen a cada clase se pueden aprender mediante la función de *embedding* de la red neuronal, y la forma de hacerlo es minimizando el error generado cuando ejemplos de una misma clase están demasiado separados.

Uno de los aspectos más relevantes en la propuesta es la forma en la que se van entrenando estas redes. Los autores proponen un entrenamiento basado en episodios, lo que permite simular de forma fiel un escenario de *few-shot learning* (que sería cada uno de estos episodios). Para lograr esto, en cada episodio se eligen ciertas clases y ciertos ejemplos de ellas como si fueran las pocas muestras que el modelo recibe en un contexto *few-shot*, y luego el resto de ejemplos de estas clases sirven para que la red aprenda a generalizar con información no vista en entrenamiento. Todo esto me pareció muy intuitivo, ya que permite regularizar lo aprendido siguiendo un procedimiento similar al que se utilizará posteriormente para evaluar al modelo (pasarle pocos ejemplos nuevos y pedirle que sea capaz de generalizar).

Esta arquitectura también permite competir en escenarios de *zero-shot learning*, donde la red sólo recibe meta-datos que caracterizan nuevas clases y debe aprender a clasificar muestras de estas clases sin ver nunca un ejemplo. Simplemente se propone tener otro *embedding* separado para aprender a generar los prototipos en base a sólo los meta-datos. Dado que la idea central es la misma, es de esperar que este *approach* se comporte bien, ya que la red aprenderá a generar prototipos que se sabe que funcionan bien una vez creados (por el caso de *few-shot learning*).

Los experimentos muestran que la propuesta logra ser estado del arte (o competir contra éste) en varios escenarios de clasificación de imágenes. Además, es un enfoque mucho más simple en comparación a los existentes (que utilizan por ejemplo *meta-learning*). El éxito de esta arquitectura se atribuye a su simplicidad: la idea de sólo utilizar un prototipo por clase permite que el aprendizaje sea más general y ocurra poco

*overfitting*, lo que en mi opinión es perfecto para atacar el problema de *few-shot learning*.

En general, lo que encontré más interesante fue la forma de entrenar estos modelos, dado que me hizo mucho sentido el tener episodios donde cada uno simule una situación de evaluación post-entrenamiento, lo que permite ajustar funciones que aprendan a agrupar ejemplos usando distancias y *clustering* (que son métodos ampliamente utilizados y probados en *machine learning* clásico).