

Aprendizaje profundo

APRENDIZAJE POR TRANSFERENCIA EN REDES RECURRENTE

Gibran Fuentes Pineda

17 de octubre de 2023

Motivación

- Muy efectivo para redes convolucionales y tareas de visión por computadora

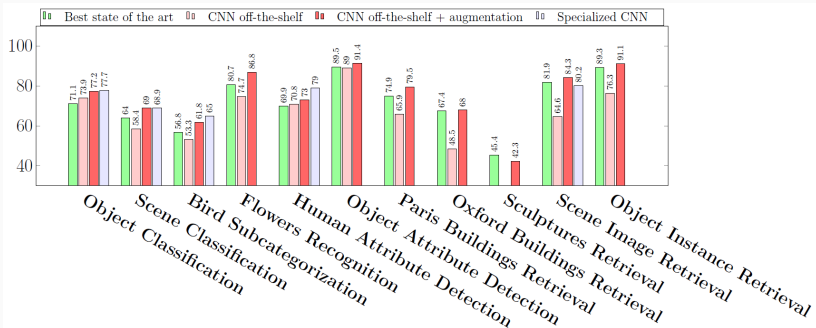
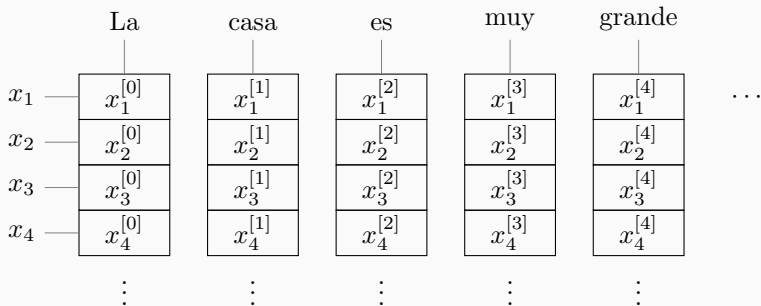


Imagen tomada de Razavian et al. CNN Features off-the-shelf: an Astounding Baseline for Recognition, 2014

Representación de elementos discretos: vectores 1-de-K



Representación de elementos discretos: vectores densos

- Encajes de palabra

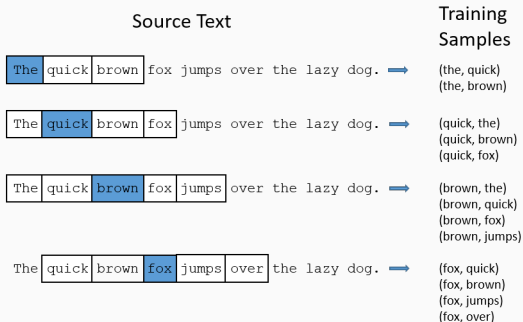


Imagen tomada de McCormick. Word2Vec Tutorial – The Skip-Gram Model, 2016.

Preprocesamiento: palabras como vectores densos

- Encajes de palabra
- Word2Vec

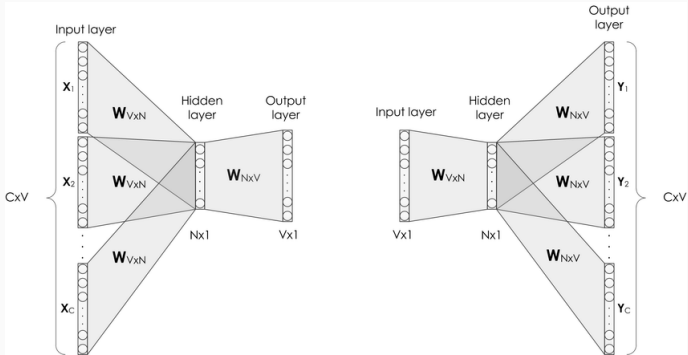


Imagen tomada de Tutubalina y Nikolenko. Demographic Prediction Based on User Reviews about Medications, 2017.

- Pre-entrenamiento de modelo de lenguaje de dominio general y estrategia para ajuste fino

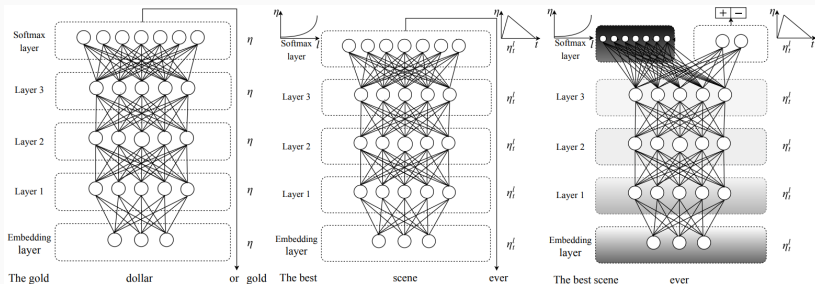


Imagen tomada de Howard y Ruder. Universal Language Model Fine-tuning for Text Classification, 2018.

Preentrenamiento

- Modelo de lenguaje basado en AWS-LSTMs
- Conjunto de datos de dominio general: Wikitext-103
 - 28,595 artículos de Wikipedia, 103 millones de palabras

= Gold dollar =

The gold dollar or gold one @-@ dollar piece was a coin struck as a regular issue by the A gold dollar had been proposed several times in the 1830s and 1840s , but was not initi Gold did not again circulate in most of the nation until 1879 ; once it did , the gold d

= Super Mario Land =

Super Mario Land is a 1989 side @-@ scrolling platform video game , the first in the Sup At Nintendo CEO Hiroshi Yamauchi 's request , Game Boy creator Gunpei Yokoi 's Nintendo Initial reviews were laudatory . Reviewers were satisfied with the smaller Super Mario B

= = = Sinclair Scientific Programmable = = =

The Sinclair Scientific Programmable was introduced in 1975 , with the same case as the It had 24 @-@ step programming abilities , which meant it was highly limited for many pu However , included with the calculator was a library of over 120 programs that that perf

Imagen tomada de Howard y Ruder. Universal Language Model Fine-tuning for Text Classification, 2018.

- Aprendizaje autosupervisado: predecir la siguiente palabra

Ajuste fino discriminativo

- Entrenamiento del modelo de lenguaje se hace en diferente grado dependiendo de la capa
 - Se emplean distintas tasas de aprendizaje $\alpha^{\ell-1} = \frac{\alpha^{\ell}}{2.6}$
- Programación de la tasa triangular

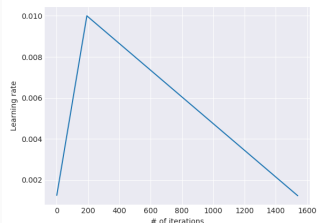


Imagen tomada de Howard y Ruder. Universal Language Model Fine-tuning for Text Classification, 2018.

- Se van descongelando capas de la final hacia atrás cada cierto número de épocas durante el entrenamiento

- Se agregan bloques con capas densas al modelo del lenguaje
- Se concatena el último estado y el sub-muestreo máximo y promedio de todos los estados

$$\mathbf{h}_c = [\mathbf{h}^T, \text{maxpool}(\mathbf{H}), \text{meanpool}(\mathbf{H})]$$