



Recuperación de Información Multimedia

Deep Learning (texto, multimodal)
Anexo: Tendencias y ética

CC5213 – Recuperación de Información Multimedia

Departamento de Ciencias de la Computación

Universidad de Chile

Juan Manuel Barrios – <https://juan.cl/mir/> – 2020



Vectorización

- Para poder utilizar texto en una red neuronal es necesario “vectorizar” el texto
- Primero se debe segmentar el texto en palabras (Tokenizer)
 - Ver capítulo de Bag-of-Words
- Convertir cada palabra (token) en un vector:
 - One-hot encoding
 - Word embedding



One-Hot Encoding

- Primero se debe obtener el vocabulario (lista de palabras conocidas)
- Para un vocabulario de n palabras, la i -ésima palabra se codifica con un vector de n dimensiones, con un 1 en la i -ésima coordenada y 0 en el resto:

$$(0, \dots, 0, 1, 0, \dots, 0)$$

- Es una codificación sparse (muchos ceros)
- Alta dimensionalidad
- Todas las palabras son igualmente distintas



Word Embeddings

- Representar palabras con vectores cuya distancia se ajuste a su diferencia en significado
- Es una codificación densa
- Menor dimensionalidad que one-hot (ej.: 300-d)
- Similitud entre palabras se debe a similitudes en su contexto
- Se entrena una conversión desde vectores one-hot a vectores densos usando una MLP
 - Es posible usar vectores pre-entrenados para vocabularios conocidos o entrenarlos para cada problema a resolver



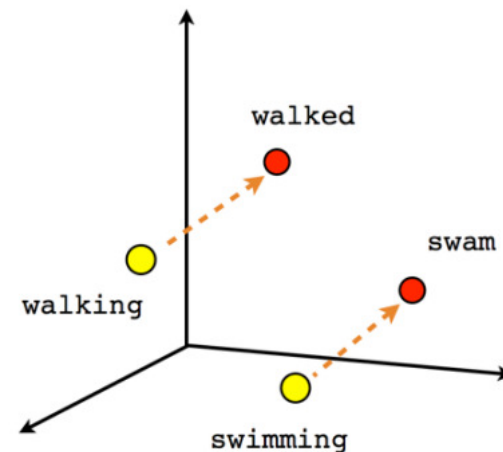
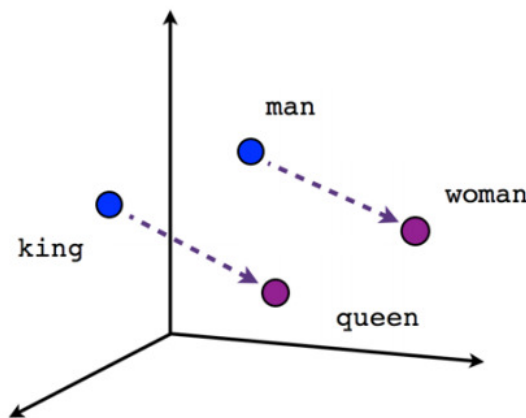
Word Embedding Space

- Se espera que el espacio de las palabras tenga propiedades como:
 - Palabras que son sinónimos estén asociadas a vectores muy cercanos entre si (distancia euclidiana cercana a cero)
 - Las direcciones en el espacio tengan algún significado de operación con las palabras:
 - singular a plural, masculino a femenino, sustantivo a adverbio, infinitivo a participio, presente a pasado, etc.
 - “el día soleado” \leftrightarrow “los días soleados”
 - “el gato negro” \leftrightarrow “la gata negra”
- El espacio de las palabras depende del idioma y también del uso (legal, técnico, popular)

Word Embedding Space

- Permiten resolver analogías:

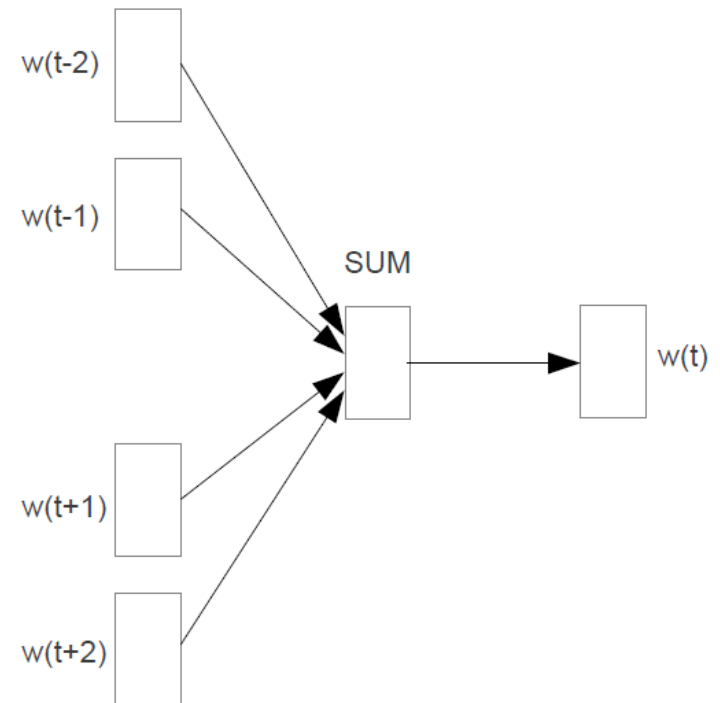
- ☐ (sintáctico) “Aparente” es a “Aparentemente” como “Evidente” es a ...
- ☐ (semántico) “Atenas” es a “Grecia” como “Oslo” es a ...



Entrenamiento Word2Vec

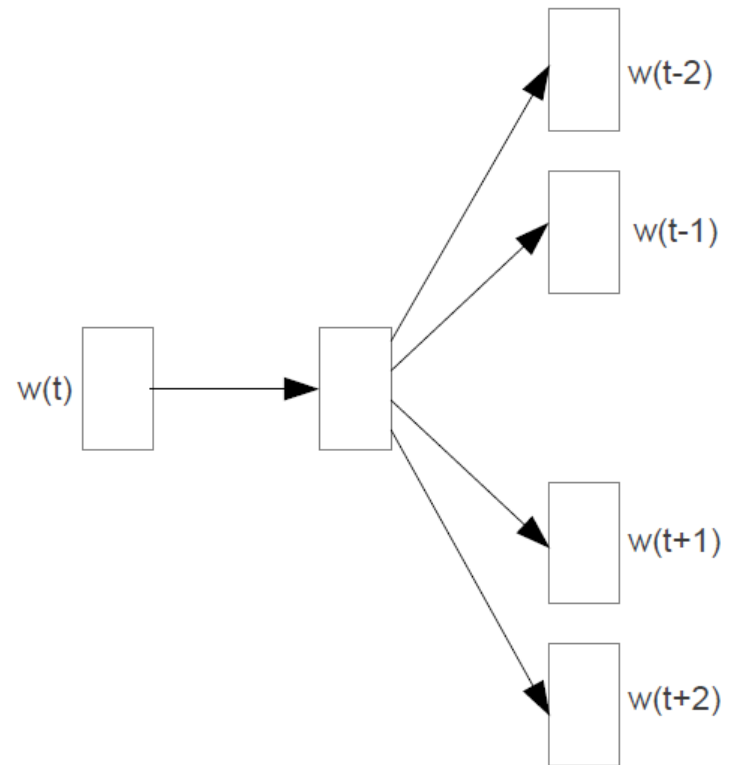
■ Modelo Continuous Bag-of-Words

- Predecir una palabra dadas sus palabras de contexto
- Produce vectores con mejor resultado en predicción sintáctica



Entrenamiento Word2Vec

- Modelo Continuous Skip-gram
 - Dada una palabra predecir sus palabras de contexto
 - Produce vectores con mejor resultado en predicción semántica





GloVe

- <https://nlp.stanford.edu/projects/glove/>
- Se basa en factorizar una matriz de co-ocurrencia de palabras
- Muy similar en idea a Latent Semantic Analysis
 - Ver capítulo de Bag-of-Words y LSA



FastText

- <https://github.com/facebookresearch/fastText>
- Calcula vectores para secuencias de caracteres y los suma para crear el vector de cada palabra
- Permite generar un vector para palabras desconocidas



Sentence Embedding

- Calcular un vector para una frase.
 - Promedio de los word vectors
 - Eliminar stop-words
 - Promedio ponderado por IDF

$$v_s = \frac{1}{|s|} \sum_{w \in s} \text{IDF}_w v_w$$
$$\text{IDF}_w := \log \frac{1 + N}{1 + N_w}$$

- Doc2Vec: Entrenar word2vec incluyendo un id del documento (sentence)

Arora et al. A simple but tough-to-beat baseline for Sentence Embeddings. 2017.
Le, Mikolov. Distributed Representations of Sentences and Documents. 2018.

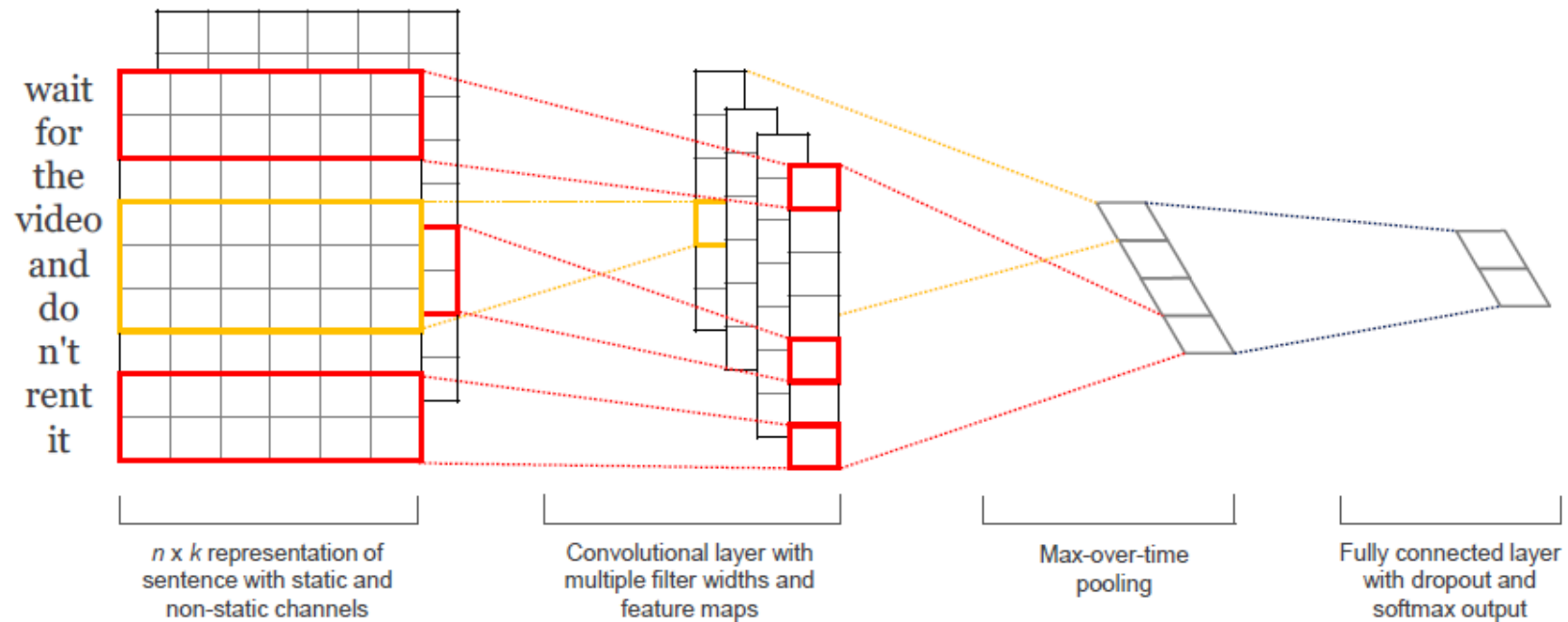


Otros Vector Embeddings

- **Node2Vec**: Vectorizar un grafo calculando un vector por nodo al medir nodos vecinos
- **Item2Vec**: Calcular vectores de ítems para sistemas recomendadores

CNN para Texto

- Conv1D es similar a N-Grams



CNN para Texto

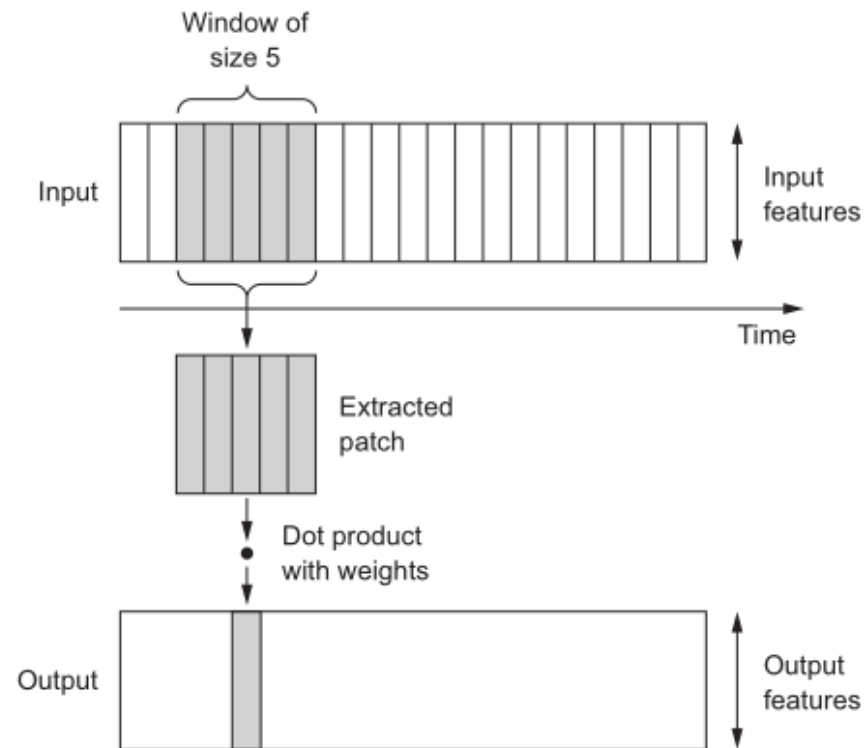
- Input: n palabras, cada palabra es un vector del word embedding de dim k (ej. $k=300$)
- Convolución: Un filtro de tamaño h corresponde un vector de $h*k$ que se usa como producto punto con una ventana de h palabras

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b) \quad \mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}]$$

- Max-Pooling en el tiempo $\hat{c} = \max\{\mathbf{c}\}$
- Se concatenan varios filtros para formar un vector
- 100 filtros de tamaño 3, 4, 5

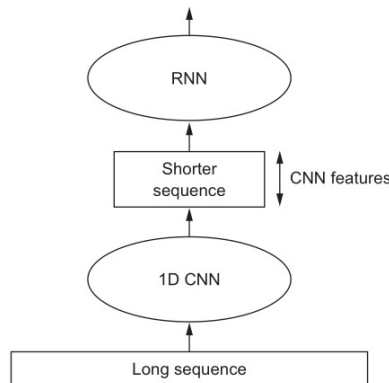
CNN para Texto

- Convolución 1D es el producto punto entre vectores



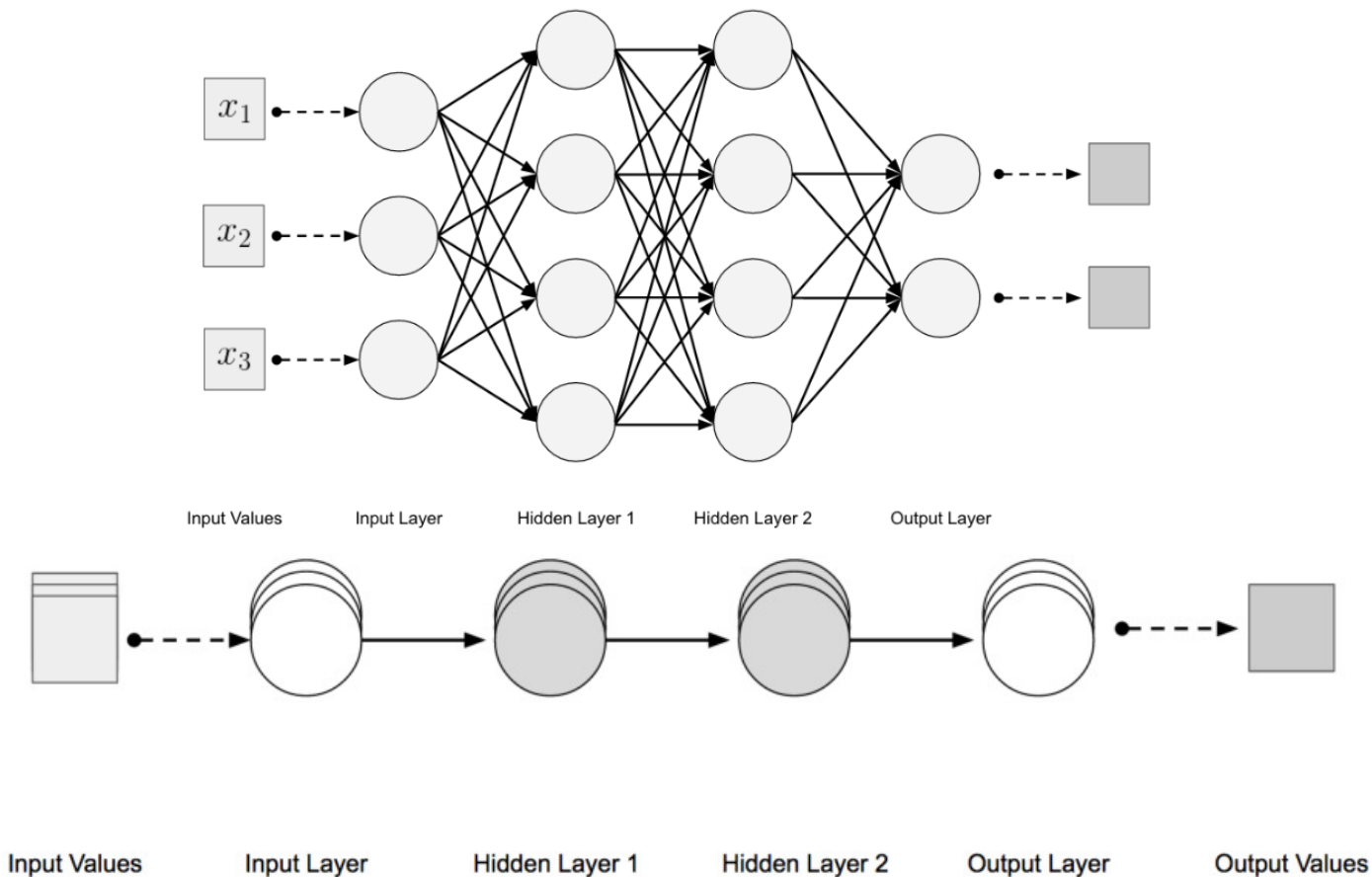
CNN para Texto

- Conv1D permiten hacer detección de grupos de palabras como n-grams
- Las convoluciones no ven la secuencias en el tiempo
- El mejor resultado se obtiene con una primera capa de Conv1D y luego una red recurrente que vea la secuencia temporal.



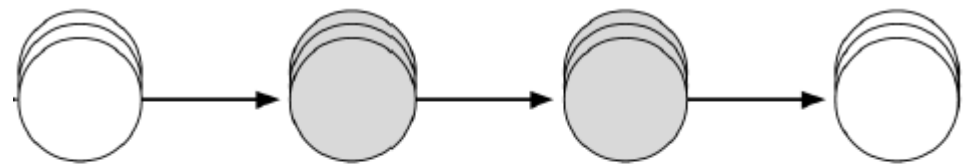
Redes Recurrentes

■ Feed-Forward vs Recurrent

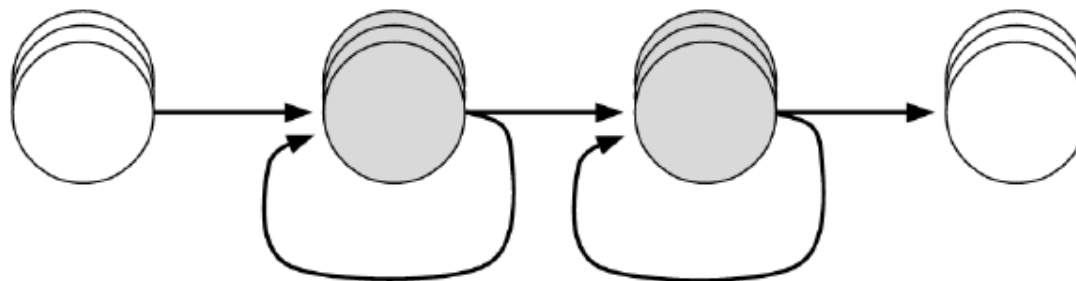


Redes Recurrentes

■ Feed-Forward vs Recurrent



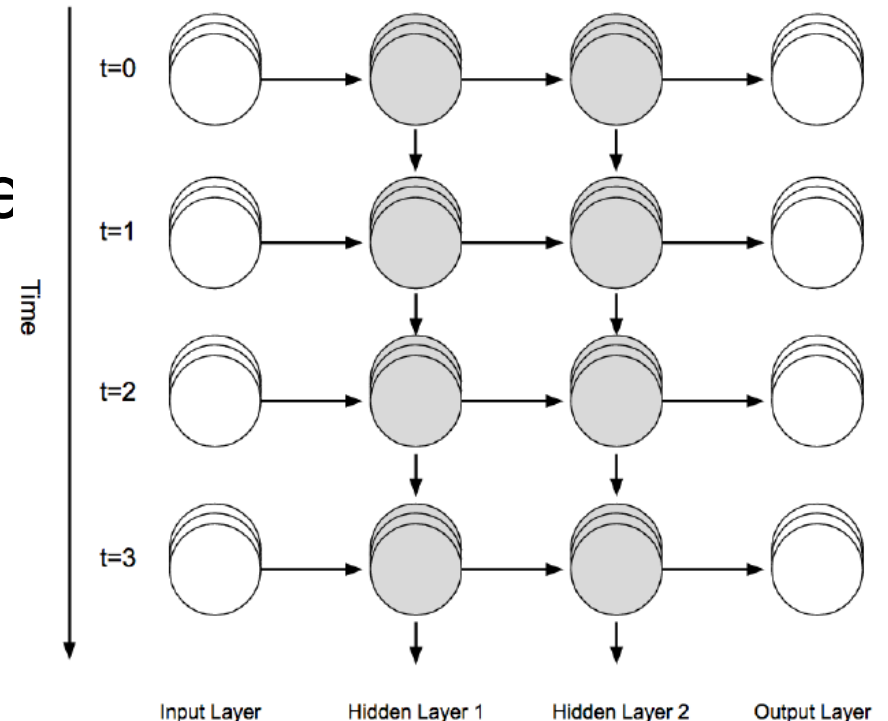
Input Layer Hidden Layer 1 Hidden Layer 2 Output Layer



Input Layer Hidden Layer 1 Hidden Layer 2 Output Layer

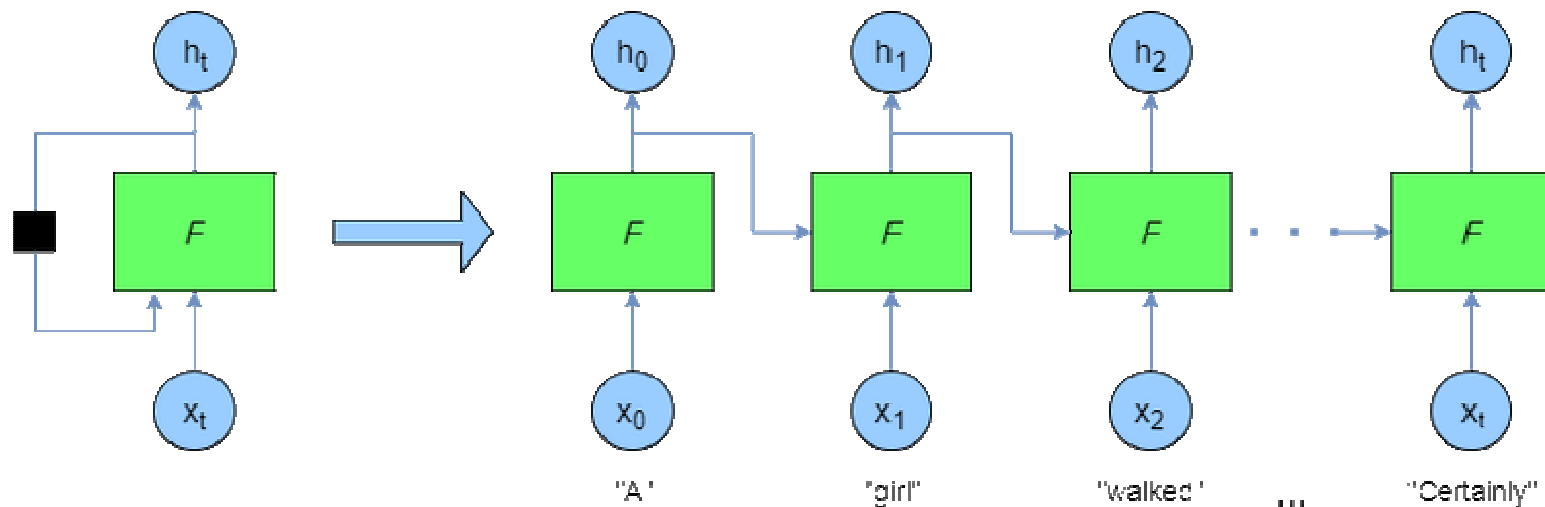
RNN

- Redes Recurrentes se usan para procesar datos con dimensión temporal, donde importa el orden de los datos
- RNN simples (“vanilla”) sufren del Vanishing Gradient para mantener información entre varias ventanas



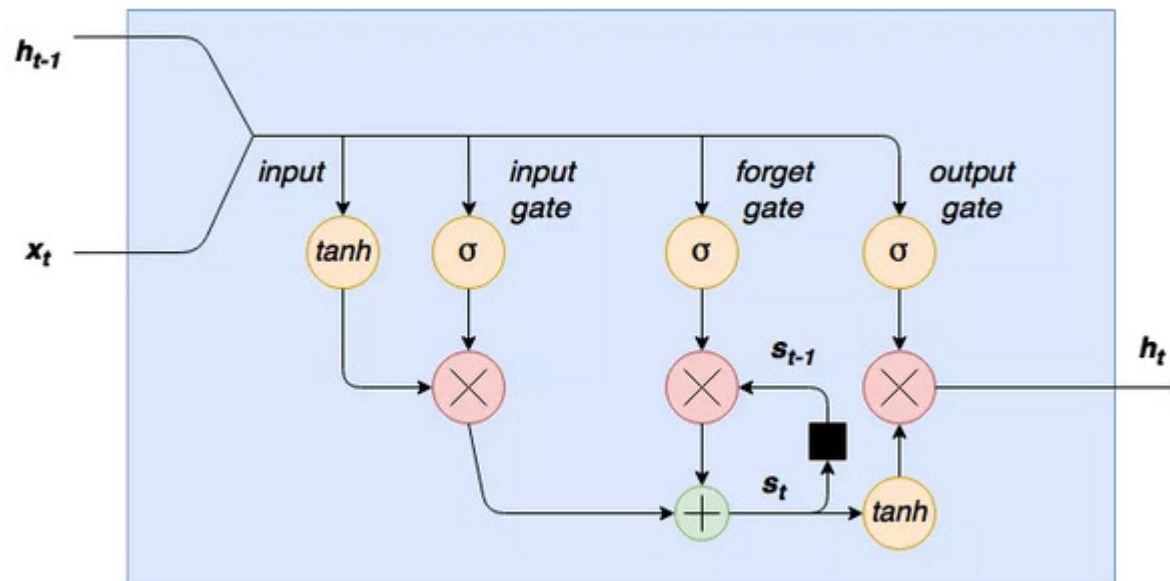
RNN

- El input de la red es una secuencia de vectores x_0 a x_t de la misma dimensión que se consumen uno a uno
- Cada entrada produce una salida intermedia h_i
- La entrada es el vector x_i junto con el estado anterior h_{i-1}
- El output final de la red es el estado final h_t



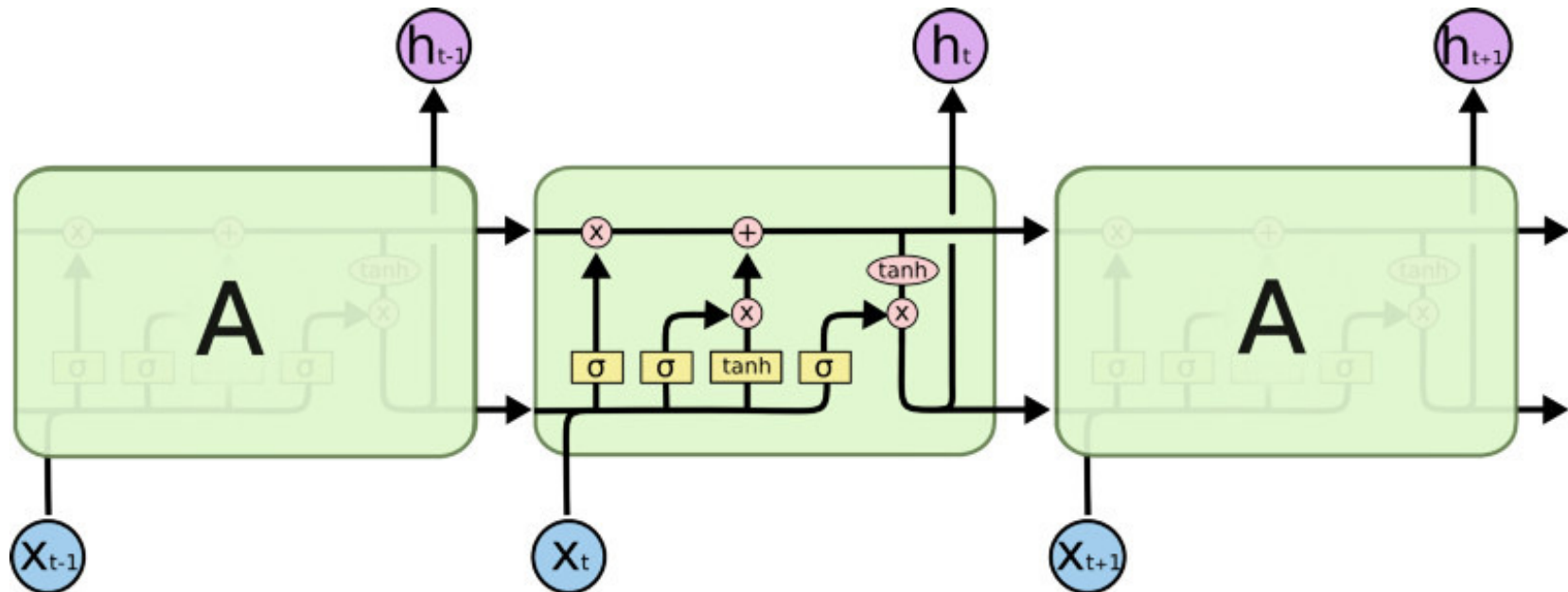
Long-Short Term Memory (LSTM)

- Contiene gates para decidir que valores se leen del estado anterior y se generan a la salida (tanh)



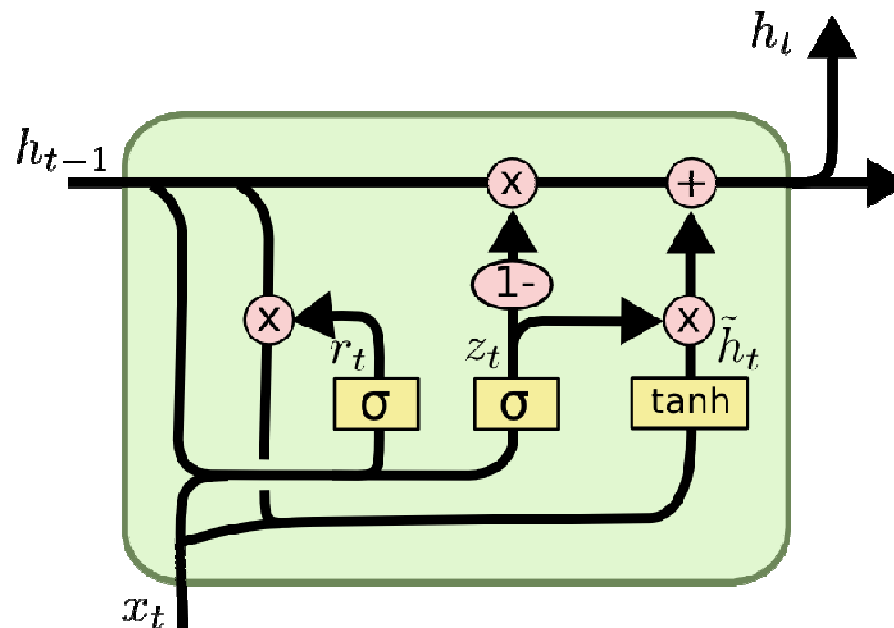
LSTM

- Long-Short Term Memory puede guardar información por periodos largos y cortos gracias a las compuertas para guardar/olvidar



RNN

- GRU (Gated Recurrent Unit)
 - Variante de LSTM con menos gates
 - Es más simple y rápida de entrenar





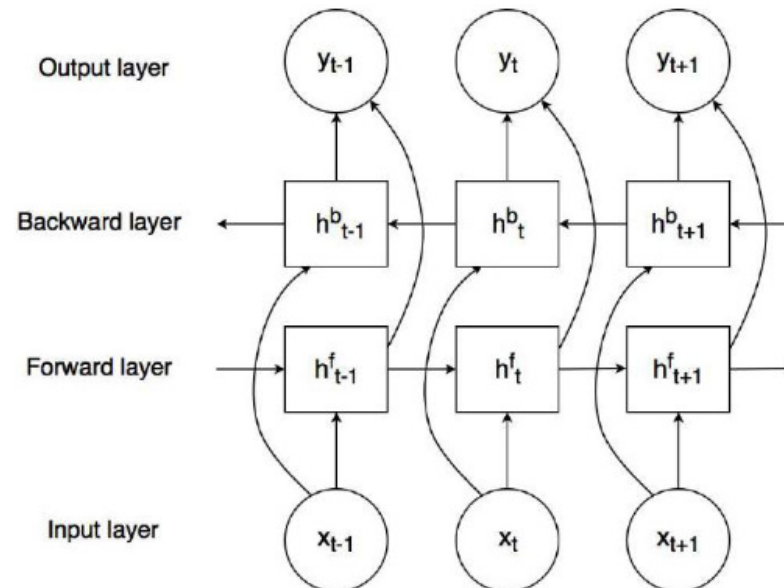
Generación de texto con RNN

- Se debe entrenar con secuencias de largo fijo con cada valor de entrada y su valor de salida correspondiente
 - Por ejemplo, para entrenar una red que genere texto se usa:

Entradas		Salidas
[puedo, escribir, los, versos]		[más]
[escribir, los, versos, más]	→	[tristes]
[los, versos, más, tristes]		[esta]
[versos, más, tristes, esta]		[noche]

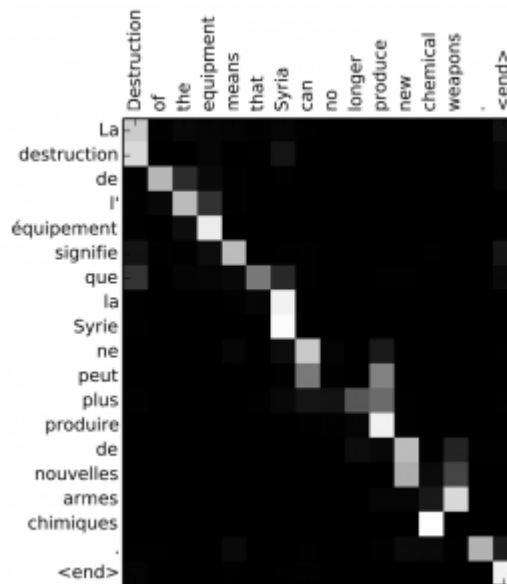
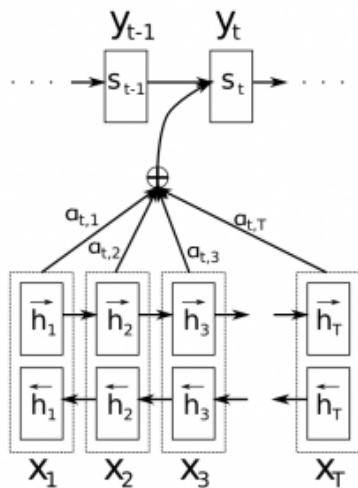
Bi-Direccional

- Para reducir la influencia de los últimos valores en el resultado final
- Se concatenan las salidas de ambas direcciones



RNN con Zona de Atención

- Crear una zona donde se guarda la relación entre inputs y outputs
- Permite encontrar la causa de una decisión



by *ent270* , *ent223* updated 9:35 am et , mon march 2 , 2015
(*ent223*) *ent63* went familial for fall at its fashion show in
ent231 on sunday , dedicating its collection to `` mamma "
with nary a pair of `` mom jeans " in sight . *ent164* and *ent21* ,
who are behind the *ent196* brand , sent models down the
runway in decidedly feminine dresses and skirts adorned
with roses , lace and even embroidered doodles by the
designers ' own nieces and nephews . many of the looks
featured saccharine needlework phrases like `` i love you ,
...

X dedicated their fall fashion show to moms



Combinación con Imágenes

Combinar Texto con Imágenes

- Dataset COCO tiene ~80 mil imágenes cada una con 5 descripciones (~400 mil descripciones)



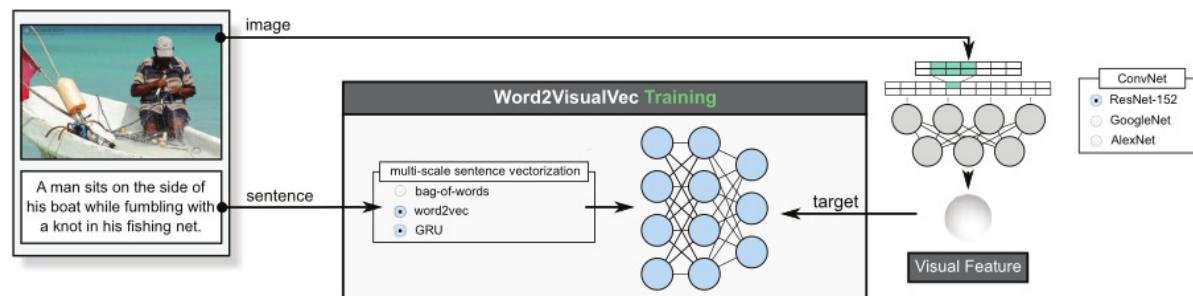
- A white lawn chair laying on top of a sandy beach
- A sun shade sitting out on the beach
- Empty beach chair under an umbrella while different boats are out in the ocean



- A group of people are surfing and swimming in the ocean
- Sunsets over a surfer and other people enjoying the ocean beach
- A child walking and watching a surfer at sunset

Buscar imágenes sin etiquetar

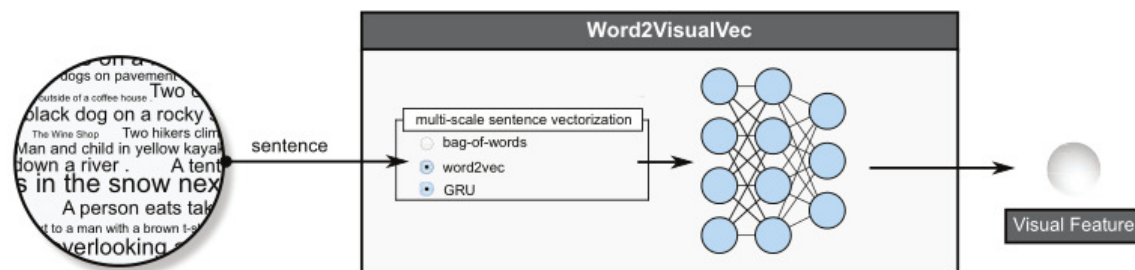
- Con datos de entrenamiento (COCO) calcular un vector visual para cada imagen y un vector textual para su descripción correspondiente
- Entrenar MLP para regresión, con entradas los vectores textuales y salidas los vectores visuales correspondientes
 - Conversión de espacios de características textual a visual (embedding)



Word2VisualVec

Dong, Li, Snoek. Predicting Visual Features from Text for Image and Video Caption Retrieval. 2018
<https://github.com/danieljf24/w2vv>

- Búsqueda de texto libre:
 - Calcular los vectores visuales de las imágenes del dataset
 - Calcular el vector textual de la consulta
 - Usar la red MLP (entrenada con COCO) y obtener su conversión a vector visual
 - Buscar los vectores visuales más cercanos en las imágenes del dataset

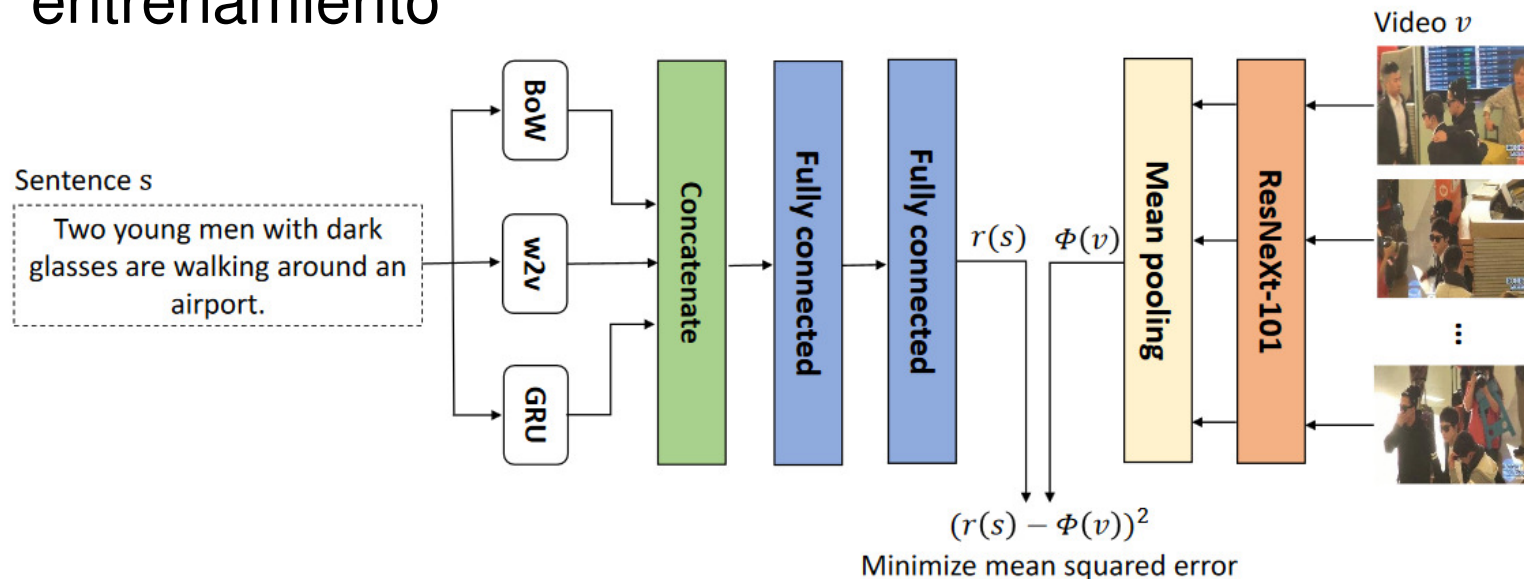


Word2VisualVec++

Li et al. Word2VisualVec++ for Ad-hoc Video Search. 2018

<https://www-nlpir.nist.gov/projects/tvpubs/tv18.slides/rucmm.avs.slides.pdf>

- En vez de convertir un tipo de descriptor en el otro, generar un Espacio Combinado al que ambos espacios se proyectan
- Se requiere una función de distancia que se desea minimizar
- Selector de los mejores pares para mejorar el entrenamiento

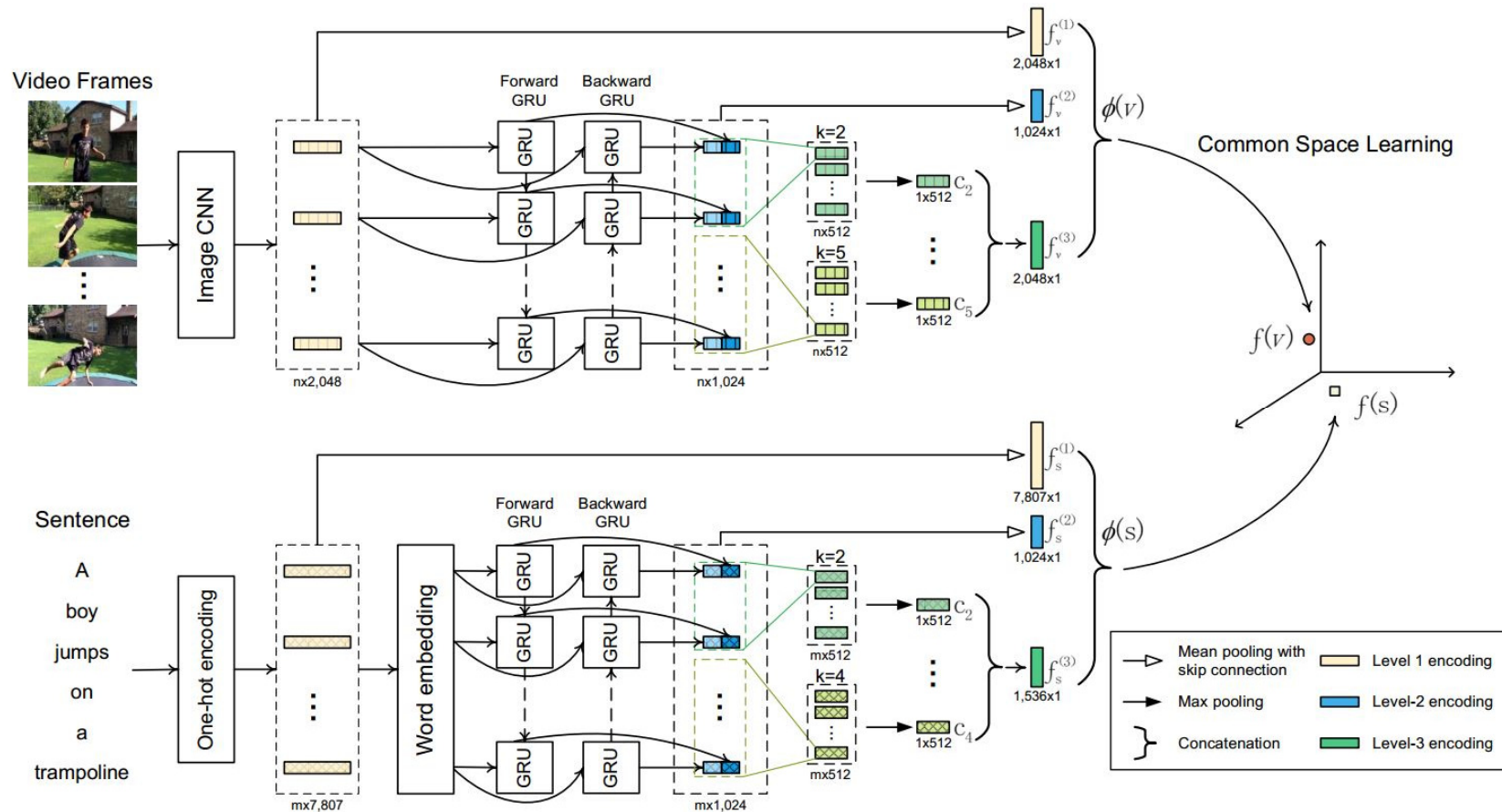


Dual Encoding

Dong et al. Dual Encoding for Zero-Example Video Retrieval. 2019

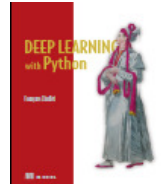
https://github.com/danieljf24/dual_encoding

- Combinar imagen y texto de forma simétrica:

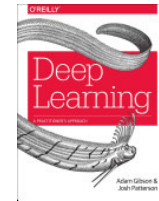


Bibliografía

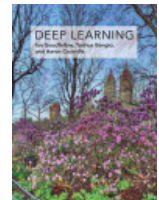
- **Deep Learning with Python.** Chollet. 2018.



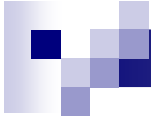
- **Deep Learning: A Practitioner's Approach.** Patterson, Gibson. 2017.



- **Deep Learning.** Goodfellow, Bengio, Courville. 2016.



- Curso de Stanford (<http://cs231n.github.io/>)
 - <http://cs231n.github.io/neural-networks-1/>
 - <http://cs231n.github.io/neural-networks-2/>



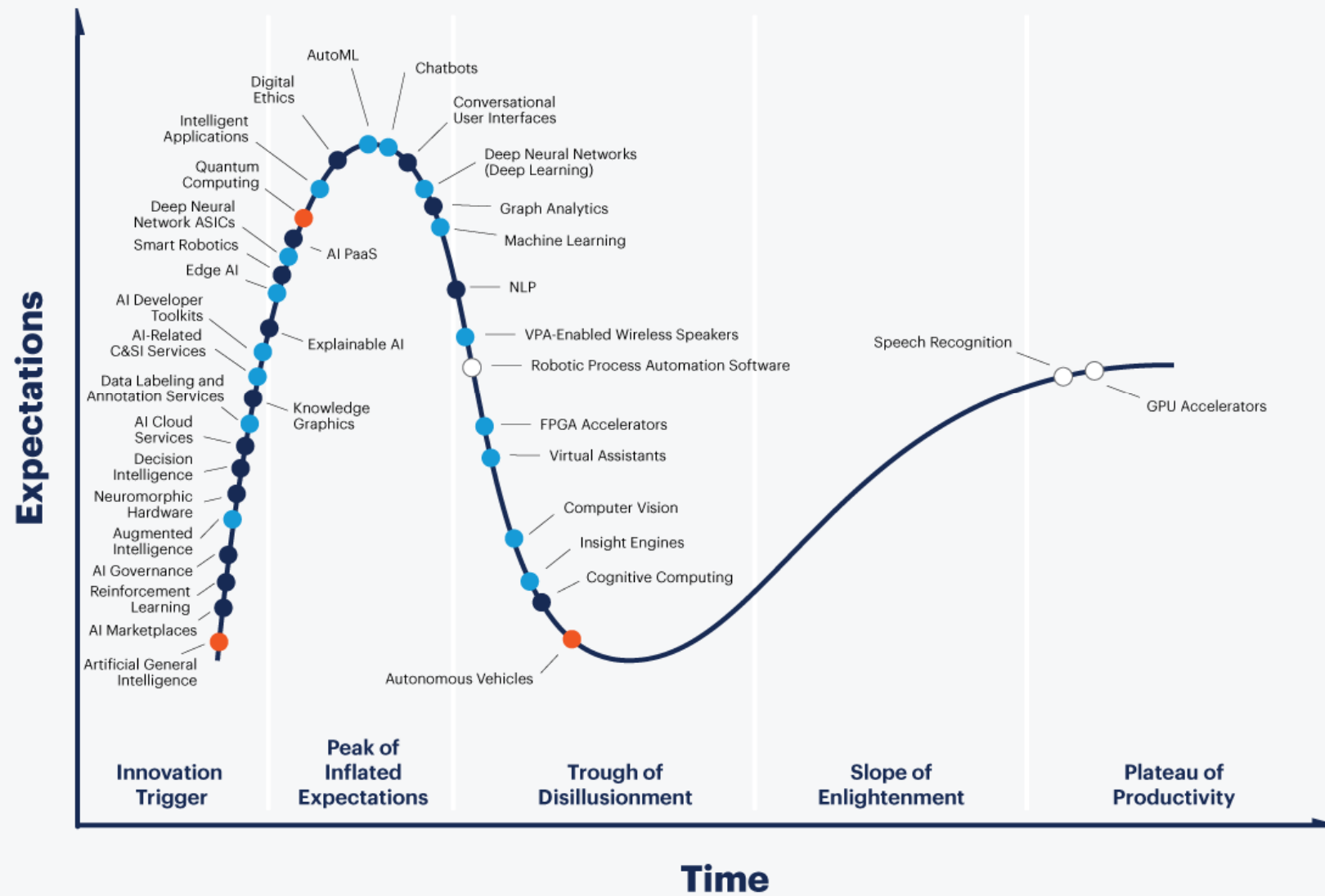
Temas para Discusión

Gartner Hype Cycle for Artificial Intelligence, 2019

Gartner

gartner.com/SmarterWithGartner

Source: Gartner
© 2019 Gartner, Inc. and/or its affiliates. All rights reserved.



Plateau will be reached:

○ less than 2 years

● 2 to 5 years

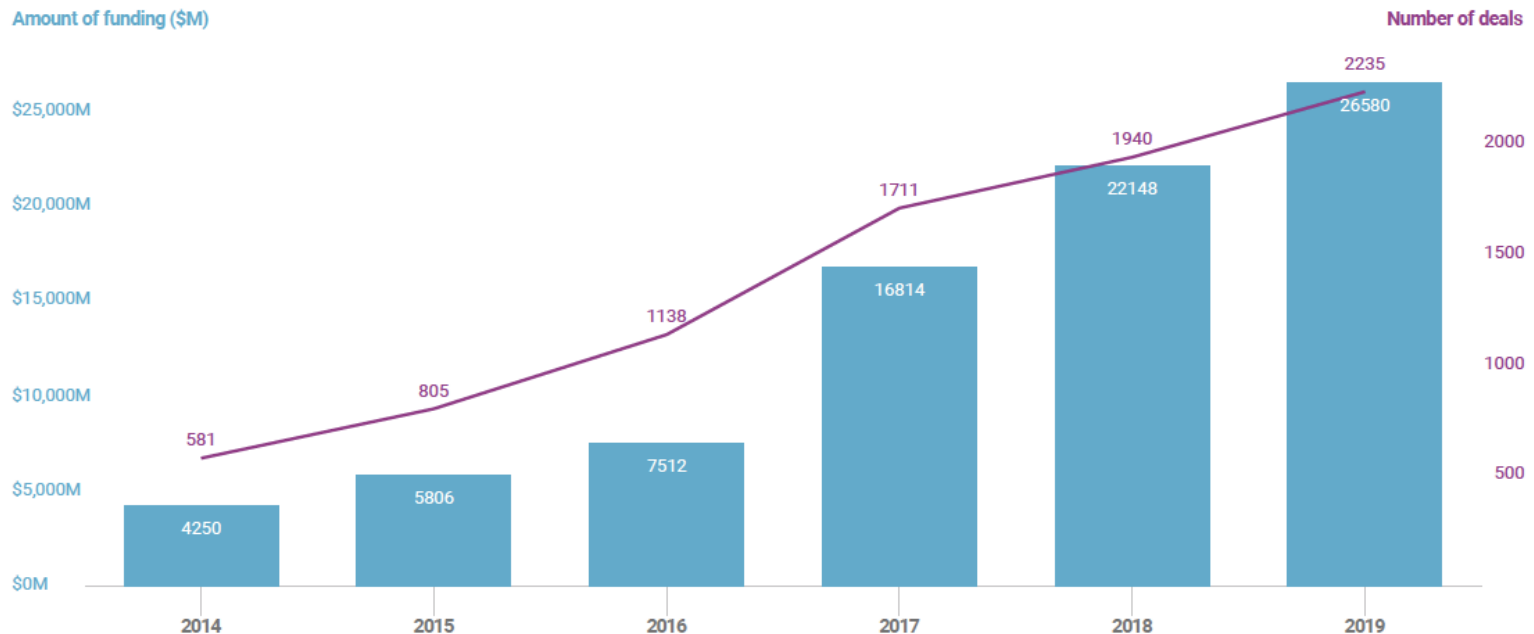
● 5 to 10 years

● more than 10 years

● obsolete before plateau

As of July 2019

2019 sees record funding to AI startups at \$26.6B



Created with:  CBINSIGHTS

Source: CB Insights

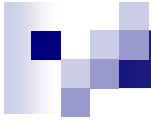
6

Forty percent of 'AI startups' in Europe don't actually use AI, claims report

Companies want to take advantage of the AI hype

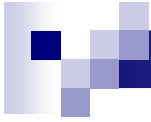
By James Vincent | Mar 5, 2019, 8:14am EST

<https://www.theverge.com/2019/3/5/18251326/ai-startups-europe-fake-40-percent-mmc-report>



Para debatir...

- Artificial Intelligence representa la nueva revolución industrial (?)
¡Es la computación!
- Artificial Intelligence eliminará y creará muchos puestos de trabajo (?)
- Presente y futuro de:
 - ☐ Artificial Intelligence
 - ☐ IoT
 - ☐ Block Chain
 - ☐ Quantum Computing



Altas Expectativas de la IA

- IBM has a Watson dilemma
 - <https://www.wsj.com/articles/ibm-bet-billions-that-watson-could-improve-cancer-treatment-it-hasnt-worked-1533961147>
- Layoffs at Watson health reveal IBM's problem with AI
 - <https://spectrum.ieee.org/the-human-os/robotics/artificial-intelligence/layoffs-at-watson-health-reveal-ibms-problem-with-ai>
- Why everyone is hating on IBM Watson - Including the people who helped make it
 - <https://gizmodo.com/why-everyone-is-hating-on-watson-including-the-people-w-1797510888>
- IBM Watson reportedly recommended Cancer treatments that were 'unsafe and incorrect'
 - <https://gizmodo.com/ibm-watson-reportedly-recommended-cancer-treatments-tha-1827868882>



Vehículos Autónomos

- Accidentes fatales (5 conductores y 1 peatón a marzo 2020)
 - https://en.wikipedia.org/wiki/List_of_self-driving_car_fatalities
- Ver el historial del accidente del vehículo Uber, en Marzo 2018 en Tempe, Arizona:
 - <https://www.theverge.com/2018/3/28/17174636/uber-self-driving-crash-fatal-arizona-update>
 - En 19-nov-2019 la investigación federal culpa al conductor, al peatón, a Uber (por falta de procedimientos y monitoreo) y al Estado de Arizona (por su poca regulación)
 - En 24-mayo-2018 el reporte del accidente dice:

“At 1.3 seconds before impact, the self-driving system determined that an emergency braking maneuver was needed to mitigate a collision. According to Uber, **emergency braking maneuvers are not enabled while the vehicle is under computer control, to reduce the potential for erratic vehicle behavior.** The vehicle operator is relied on to intervene and take action.”



Cambridge Analytica

- Influencia en votaciones de EE.UU., UK y otros.
 - Recopilar datos de personas por medio de apps (encuestas, juegos)
 - Generar perfiles psicológicos de personas
 - Influnciar votantes dudosos por medio de publicidad y noticias (falsas) dirigidas a cada uno
- Facebook dice: “Los usuarios cedieron su información; no hubo infiltración en los sistemas y no hubo robo de contraseñas ni de información sensible”
 - <https://www.bbc.com/mundo/noticias-43472797>
- The Great Hack (Documental en Netflix)
 - https://en.wikipedia.org/wiki/The_Great_Hack
- Las personas no saben que son manipuladas **¡No hay regulación!**



Artículos para Leer

- El caso de motores Diesel de Volkswagen
 - https://es.wikipedia.org/wiki/Esc%C3%A1ndalo_de_emisiones_contaminantes_de_veh%C3%ADculos_Volkswagen
- Sesgos en rostros. Amazon's face recognition misidentifies 28 members of Congress as suspected criminals:
 - <https://gizmodo.com/amazons-face-recognition-misidentifies-28-members-of-co-1827887567>
- Columna de opinión: Has AI surpassed humans at translation? Not even close!
 - https://www.skynettoday.com/editorials/state_of_nmt
- Columna de opinión: AI winter is well on its way
 - <https://blog.piekniowski.info/2018/05/28/ai-winter-is-well-on-its-way/>
- Carta de investigadores: Research Priorities for Robust and Beneficial Artificial Intelligence
 - http://futureoflife.org/misc/open_letter

Científicos piden que el desarrollo de la inteligencia artificial beneficie a la humanidad

Una agrupación de más de 700 expertos de todo el mundo firmó una carta donde se solicita que todo desarrollo debe incluir un estudio sobre cómo se beneficia a la sociedad.

AFP

Lunes, 12 de Enero de 2015, 17:49

Expertos en inteligencia artificial hacen llamado para frenar desarrollo de armas autónomas

El científico Stephen Hawking y el cofundador de Apple Steve Wozniak, están en la lista de quienes firmaron una carta donde advierten sobre los peligros de este tipo de tecnología bélica.

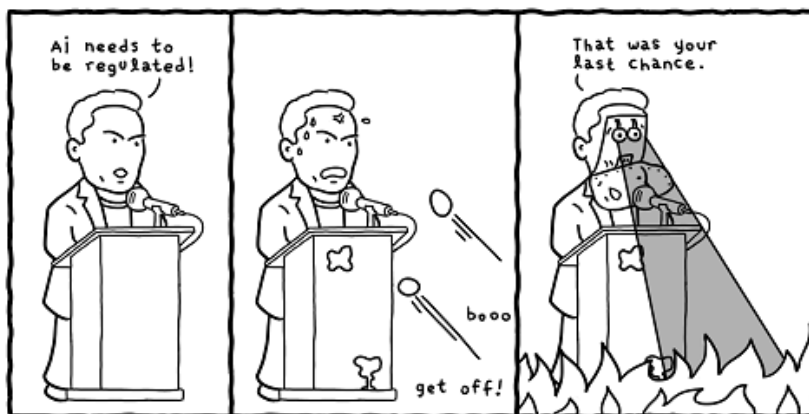
Emol

lunes, 27 de julio de 2015 11:59



WASHINGTON.- Más de 700 científicos de todo el mundo, entre ellos el físico británico Stephen Hawking y el fundador de SpaceX Elon Musk, pidieron este fin de semana que los avances en materia de inteligencia artificial sirvan para beneficiar a la humanidad.

"Los progresos realizados en inteligencia artificial son una buena ocasión para concentrar nuestras investigaciones en aquellos trabajos que no sólo hacen de las tecnologías herramientas cada vez más poderosas, sino además más beneficiosas para la sociedad", escribieron los expertos en una carta abierta.



Daniel Stori {turnoff.us}

<http://turnoff.us/geek/ai-regulation/>