



UNIVERSIDAD DE CHILE

Deep Learning

Deeper, Better, _____, Stronger than Machine Learning

Valentin Barriere

Universidad de Chile – DCC

CC6204, Primavera 2024

Aprendizaje profundo – Addicional

Redes Recurrentes

Outline : Redes Recurrentes

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

Outline : RNN

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

Redes recurrentes

Interés

Permiten tratar datos secuenciales, como texto, audio, video...

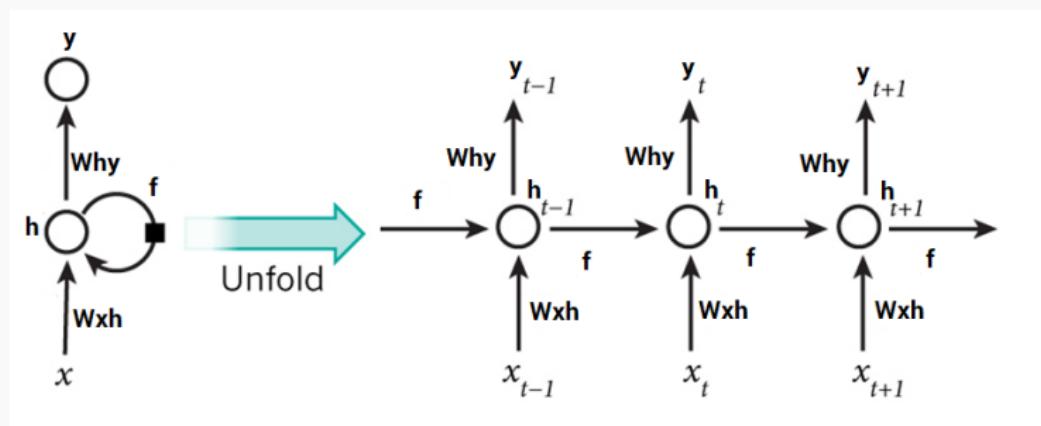
Existen diferentes maneras de etiquetar una secuencia:

- Dar una etiqueta única a toda la secuencia (análisis de sentimientos en un texto)
Me llamo Valentín y estoy muy contento → **Positivo**
- Dar una etiqueta a cada elemento de la secuencia (*sequence labelling*, para reconocimiento de entidades nombradas)
Me llamo Valentín y estoy muy contento → Entidad: Valentín
- Producir una secuencia de tamaño variable (traducción automática)
Me llamo Valentín y estoy muy contento → My name is Valentin
and I am really happy

Recurrencia

La propagación no es solo hacia adelante porque h_{t-1} influye en h_t

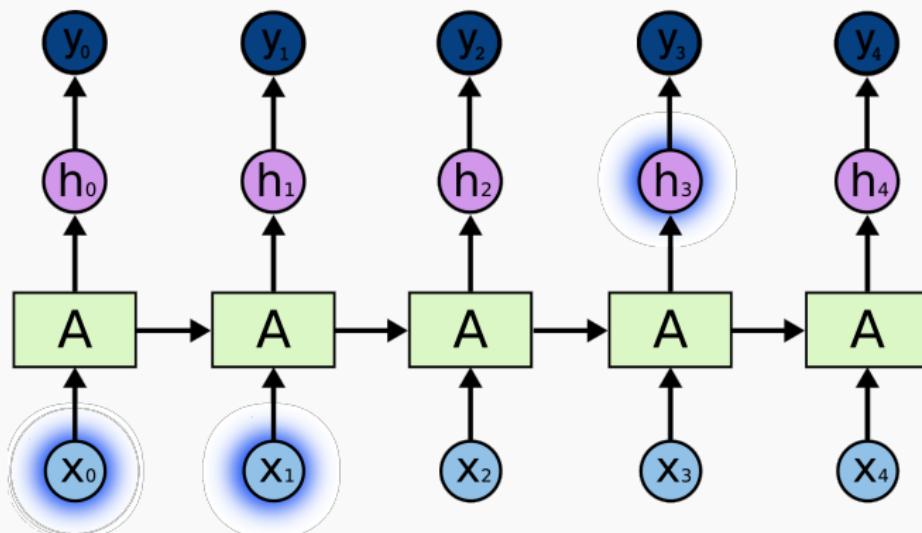
- Influencia de los valores de activación obtenidos en tiempos anteriores
- Coherencia entre las diferentes predicciones de una misma secuencia



Modelado de dependencias cortas

Ventaja

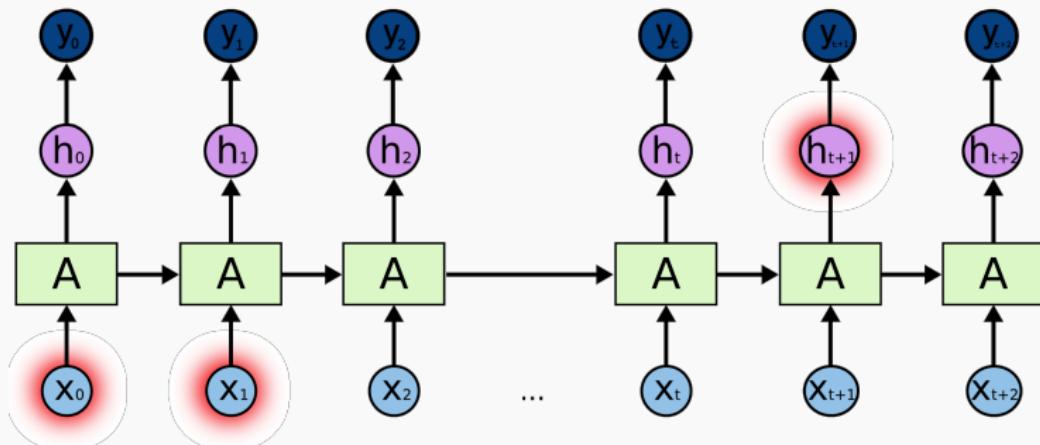
Gracias a esta estructura secuencial, el RNN puede modelar dependencias cortas entre las observaciones de entrada y las salidas de la red. De hecho, el estado oculto se actualiza en función de las entradas que llegan secuencialmente.



Modelado de dependencias largas

Desventaja

Sin embargo, si la dependencia es larga, el RNN no logra mantener en memoria las observaciones y establecer la relación entre la entrada y una salida.



Outline : LSTM

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

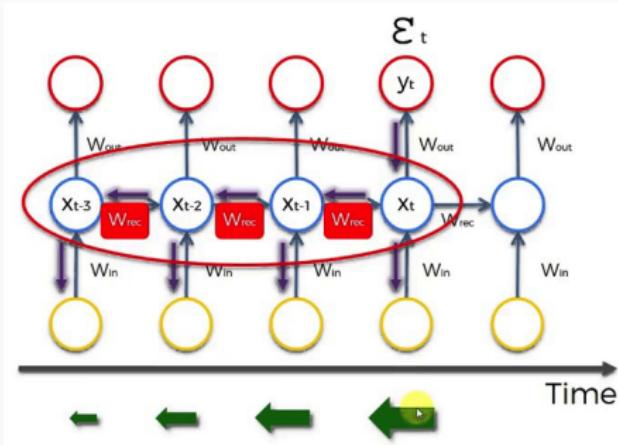
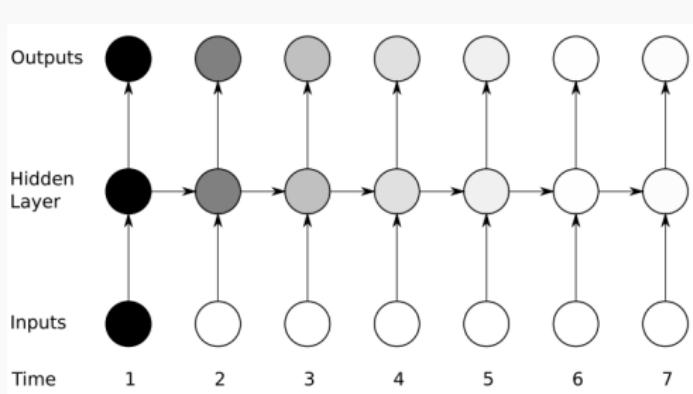
Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

Desvanecimiento del gradiente y dependencias largas

Problema del desvanecimiento del gradiente (Vanishing Gradient)

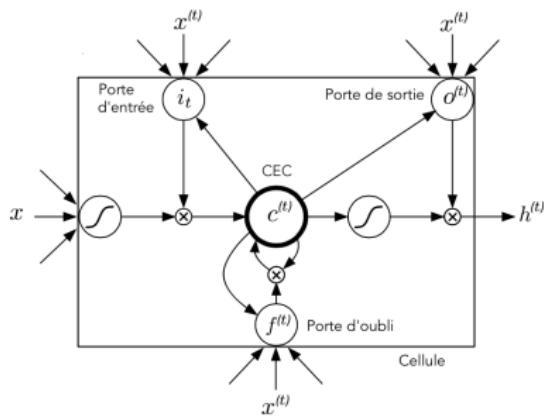
De la misma manera que la sensibilidad disminuye, el descenso de gradiente no es **adecuado para secuencias demasiado largas**. Si la secuencia modelada es demasiado larga, entonces los enlaces entre las salidas del LSTM y una observación que llega mucho antes en la secuencia se vuelven muy débiles.



Desvanecimiento del gradiente: LSTM

Solución a estos problemas: Long Short Term Memory

- Unidad lineal central
- Protección de esta unidad y de las predicciones mediante puertas de entrada y salida
- Posibilidad de reiniciar la memoria de la unidad lineal



$$\forall t \in [1; T], \quad i^{(t)} = \sigma(W_i x^{(t)} + U_i h^{(t-1)} + V_i c^{(t-1)} + b_i)$$

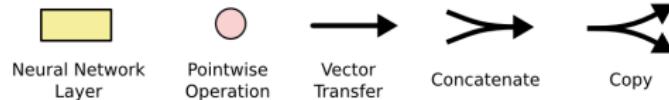
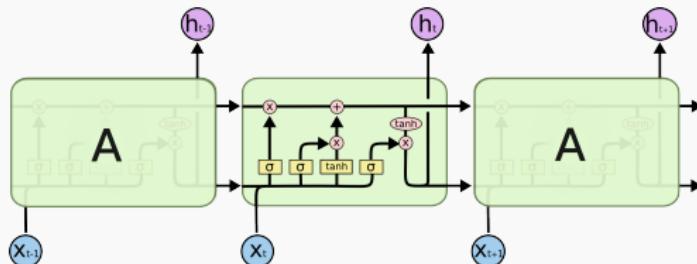
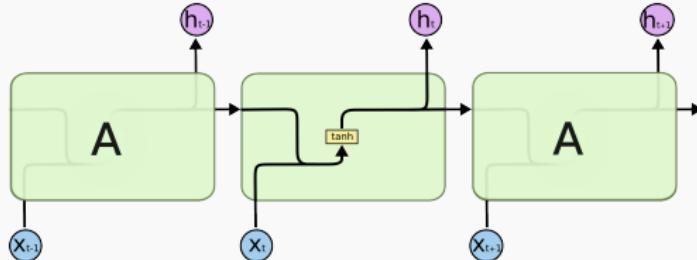
$$\forall t \in [1; T], \quad f^{(t)} = \sigma(W_f x^{(t)} + U_f h^{(t-1)} + V_f c^{(t-1)} + b_f)$$

$$\forall t \in [1; T], \quad c^{(t)} = f^{(t)} c^{(t-1)} + i^{(t)} \tanh(W_c x^{(t)} + U_c h^{(t-1)} + b_c)$$

$$\forall t \in [1; T], \quad o^{(t)} = \sigma(W_o x^{(t)} + U_o h^{(t-1)} + V_o c^{(t)} + b_o)$$

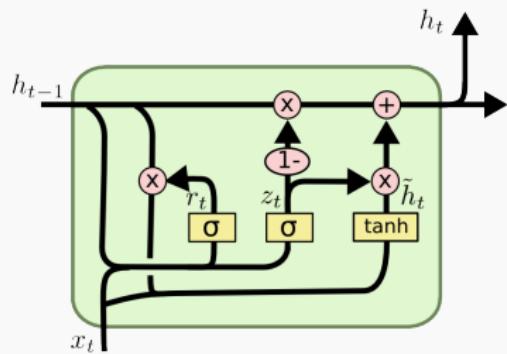
$$\forall t \in [1; T], \quad h^{(t)} = o^{(t)} \tanh(c^{(t)})$$

Comparación



[Más información](#)

Arquitectura GRU



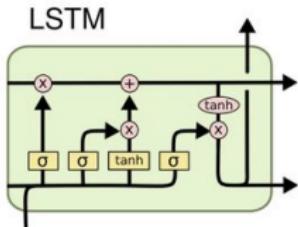
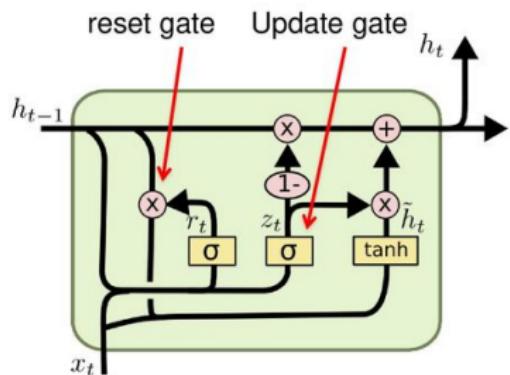
$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh (W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

GRU – gated recurrent unit (more compression)



$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh (W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

It combines the **forget** and **input** into a single **update gate**.

It also merges the cell state and hidden state. This is simpler than LSTM. There are many other variants too.

CNN, Arte y GANs

Outline : CNN, Arte y GANs

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

Outline : Deep Dream

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

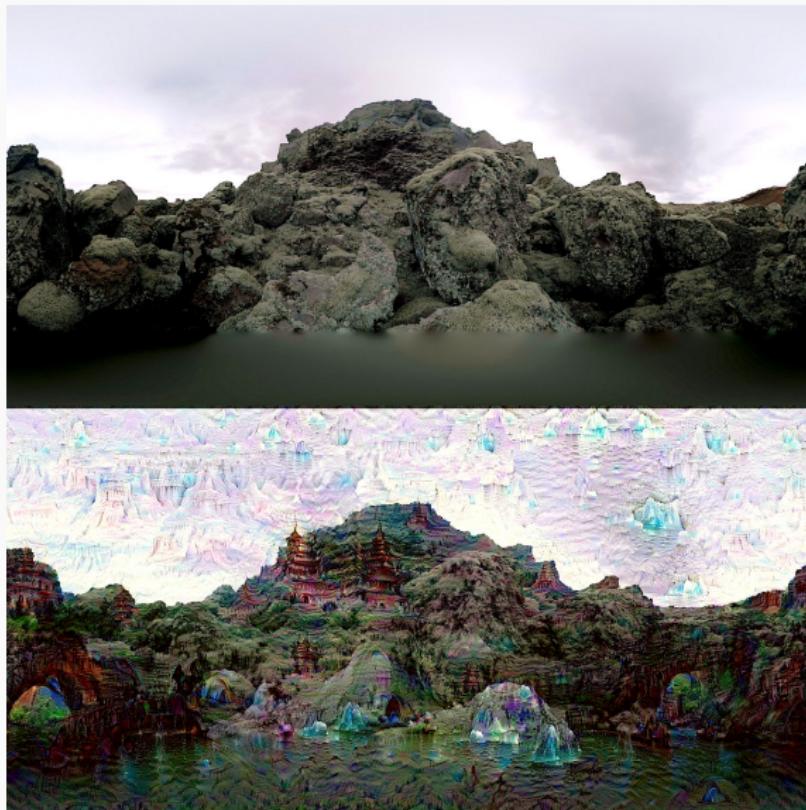
Auto-codificadores y embeddings

AE

Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

CNN y Arte: Deep Dream I



CNN y Arte: Deep Dream II

Principio

Transforma las partes de la imagen donde la red detecta un patrón conocido para hacerlo aparecer de manera más fuerte. Esto se hace de manera iterativa: se modifica un poco, luego se vuelve a introducir la imagen modificada, etc.



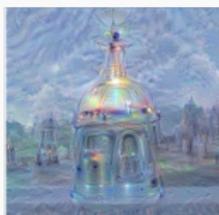
Horizon



Trees



Leaves



Towers & Pagodas

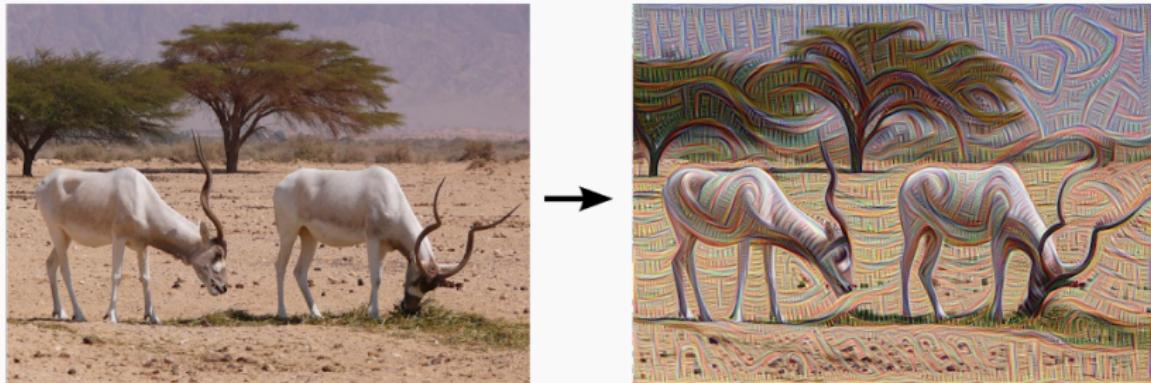


Buildings



Birds & Insects

CNN y Arte: Deep Dream III



Se puede elegir qué capas van a modificar la imagen. Por ejemplo, las capas de bajo nivel tienden a producir trazos o patrones ornamentales psicodélicos, ya que estas capas son sensibles a características de bajo nivel como los contornos y sus orientaciones.

[Más información sobre Deep Dream](#)

[Notebook de Google para implementación](#)

Outline : GANs

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

CNN y Arte: Transferencia de estilo neuronal

Principio

Aplica el estilo de una imagen a otra separando el fondo (representación de alto nivel) y la forma (representación de bajo nivel).

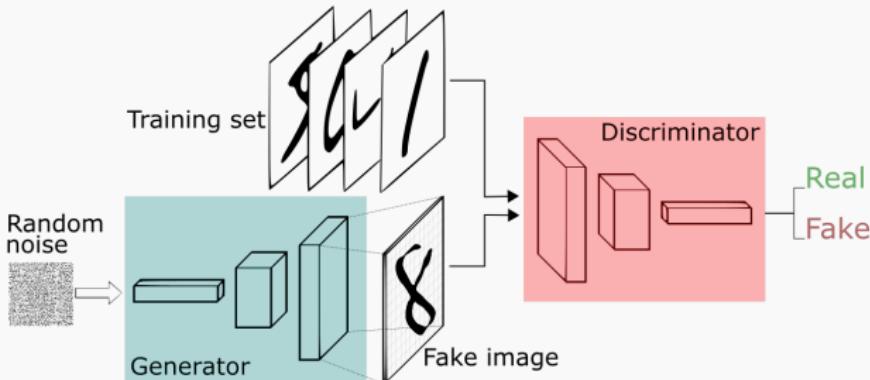


Redes Generativas Antagónicas (Generative Adversarial Network; GAN)

Principio

- Enseña a un modelo (**Generador**) a generar datos que otro modelo (**Discriminador**) no debe ser capaz de reconocer como artificiales.
- El discriminador

Al final se obtienen imágenes (por ejemplo) generadas por una máquina que se asemejan a las de nuestro conjunto de datos.



Redes Generativas Antagónicas (Generative Adversarial Network; GAN)



[Explicaciones completas sobre los GANs](#) [Tutorial de código](#) ;
[Aplicaciones de GANs](#) ;

Principio

Mismo principio que el GAN, pero en 2 etapas:

- La primera para discriminar si es una imagen falsa o real: esto permite **crear una imagen que se asemeje a la distribución de datos del conjunto de datos** (es decir, que parezca real).
- La segunda para discriminar cuál es el estilo artístico de la imagen: esto permite **crear una imagen que no esté asociada a ninguna de las clases conocidas**. Se obtiene así una imagen que parece real pero no corresponde a ningún estilo artístico conocido.

CNN y Arte: Redes Antagónicas Creativas (CAN)



Implementación en Tensorflow ; [Artículo](#)

Auto-codificadores y embeddings

Outline : Auto-codificadores y embeddings

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

Outline : AE

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

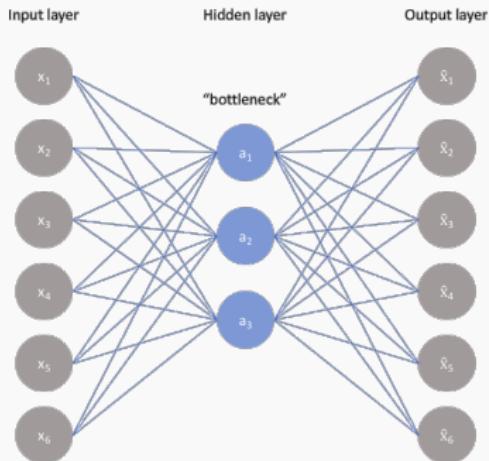
Codificador-Decodificador
secuencial: Seq2Seq

Embeddings de Contexto

Auto-codificador (Auto-Encoder; AE): Principio

Principio

El principio de un auto-codificador es encontrar una representación distribuida de menor dimensión a que permita recuperar la representación inicial \mathbf{X} , por lo tanto contiene tanta información como el original.



Auto-codificador: Aprendizaje

Aprendizaje no supervisado

El auto-codificador se aprende de manera no supervisada teniendo una función de costo sobre la reconstrucción $\hat{\mathbf{X}}$ de la variable inicial \mathbf{X} :

$$\ell(\mathbf{X}) = \|\mathbf{X} - \hat{\mathbf{X}}\|_2.$$

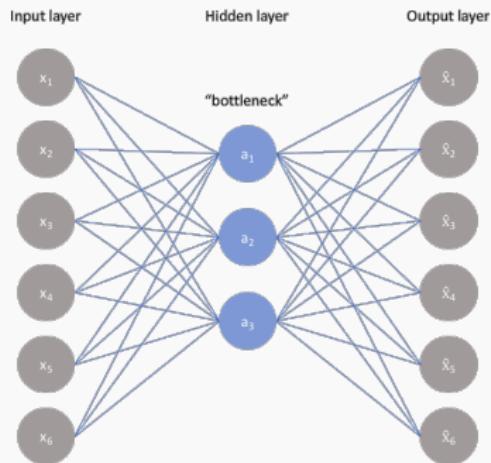
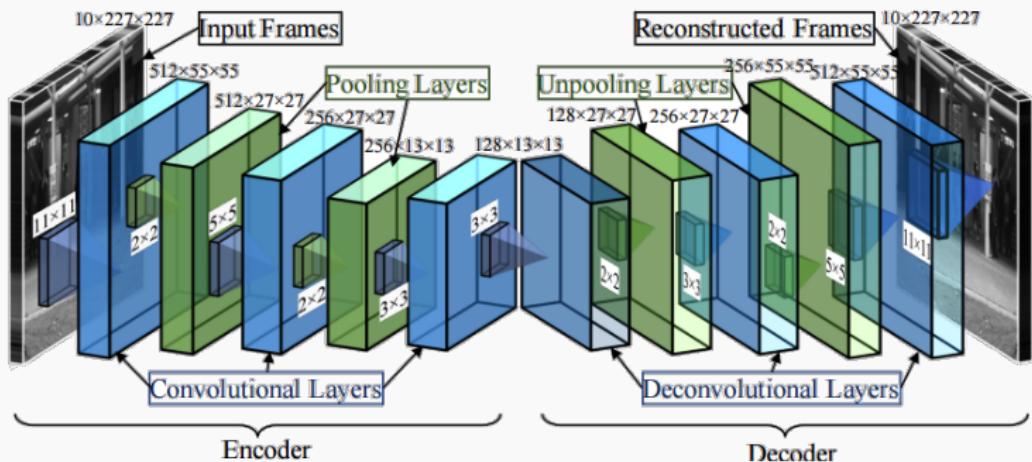


Figure 1: Ejemplo de AE de una capa, puede haber tantas capas como se desee

Auto-codificador: Tipo de codificador

Cualquier tipo de codificador

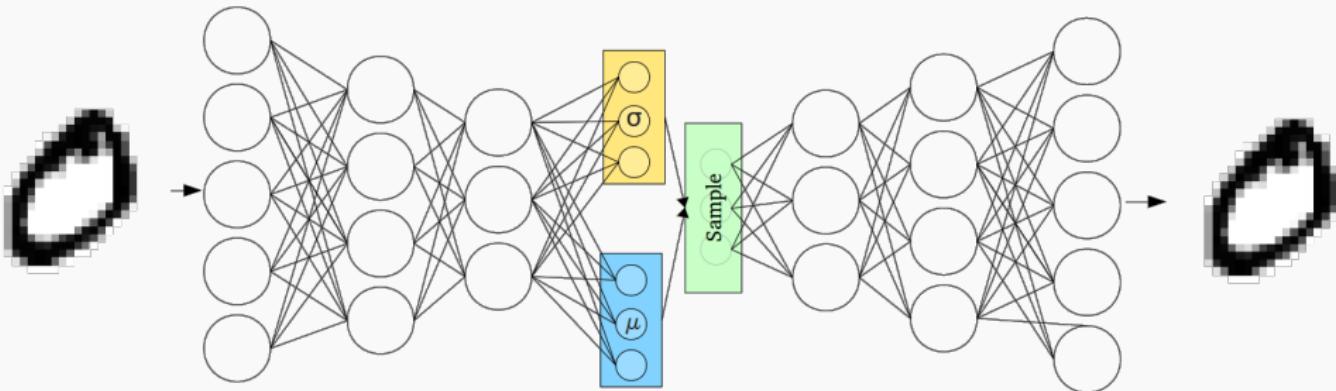
El codificador y el decodificador pueden ser algo más que perceptrones multicapa, como CNN o RNN



Auto-codificador variacional

Principio

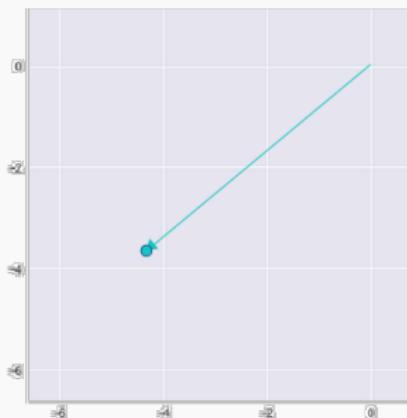
El VAE ya no codifica los datos como un punto en el espacio, sino como una gaussiana con un centro μ y una desviación estándar σ dada.



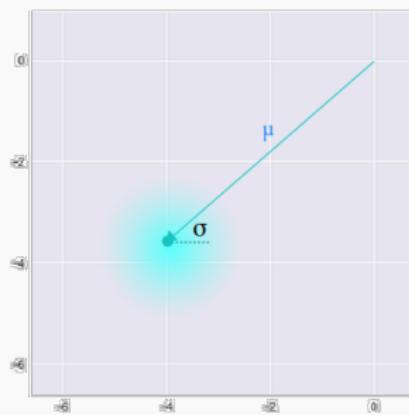
Auto-codificador variacional

Interés

Esto permite "suavizar" el espacio entre diferentes puntos y así pasar de un punto a otro para generar datos que parezcan coherentes.

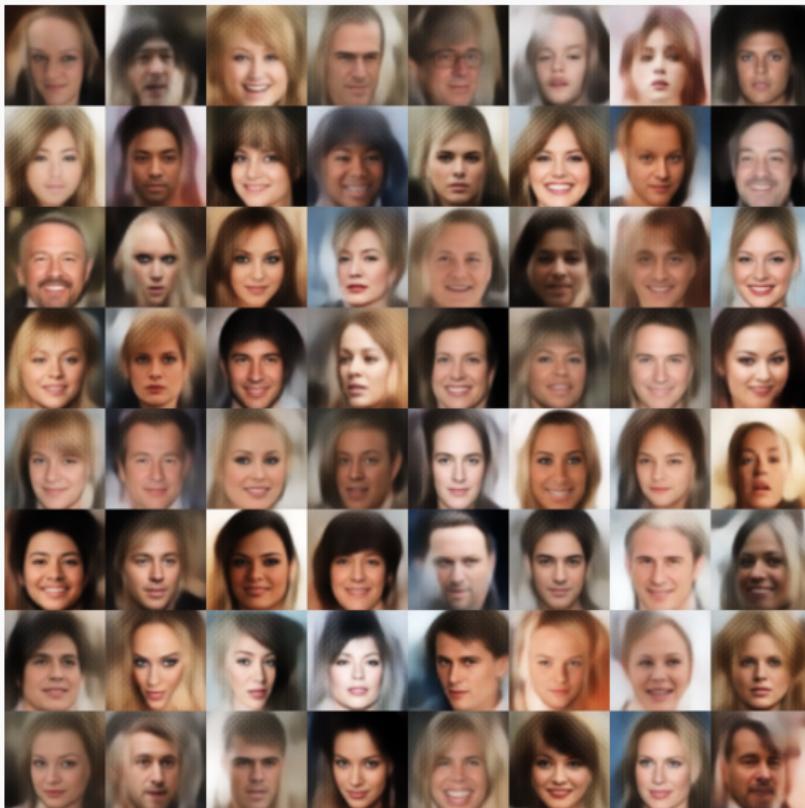


Standard Autoencoder
(direct encoding coordinates)



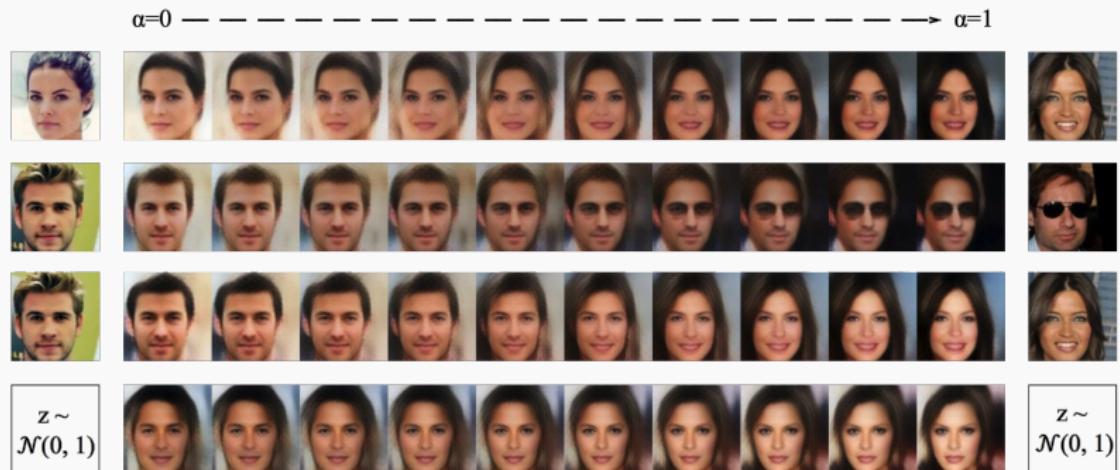
Variational Autoencoder
(μ and σ initialize a probability distribution)

Auto-codificador variacional: Ejemplo de uso



Auto-codificador variacional: Ejemplo de uso

Paso fluido de una imagen a otra:



Viajes en el espacio de las frases

Outline : Codificador-Decodificador secuencial: Seq2Seq

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

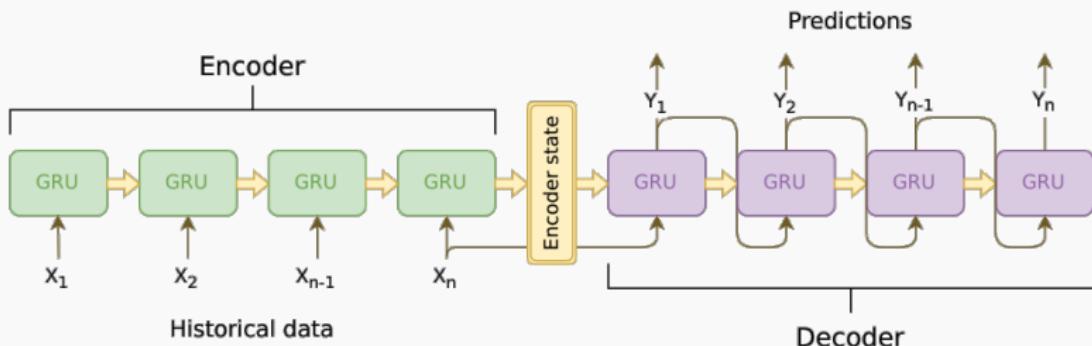
AE

Codificador-Decodificador
secuencial: Seq2Seq

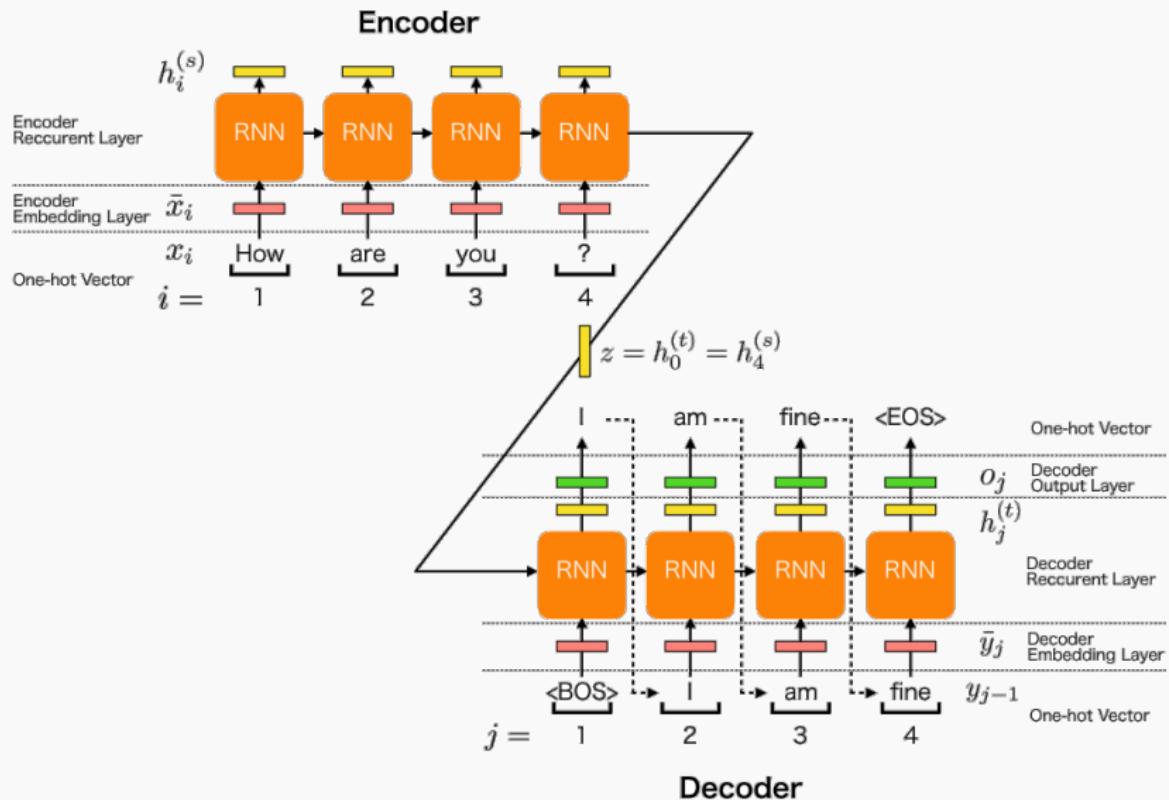
Embeddings de Contexto

Codificador-Decodificador secuencial (Seq2Seq)

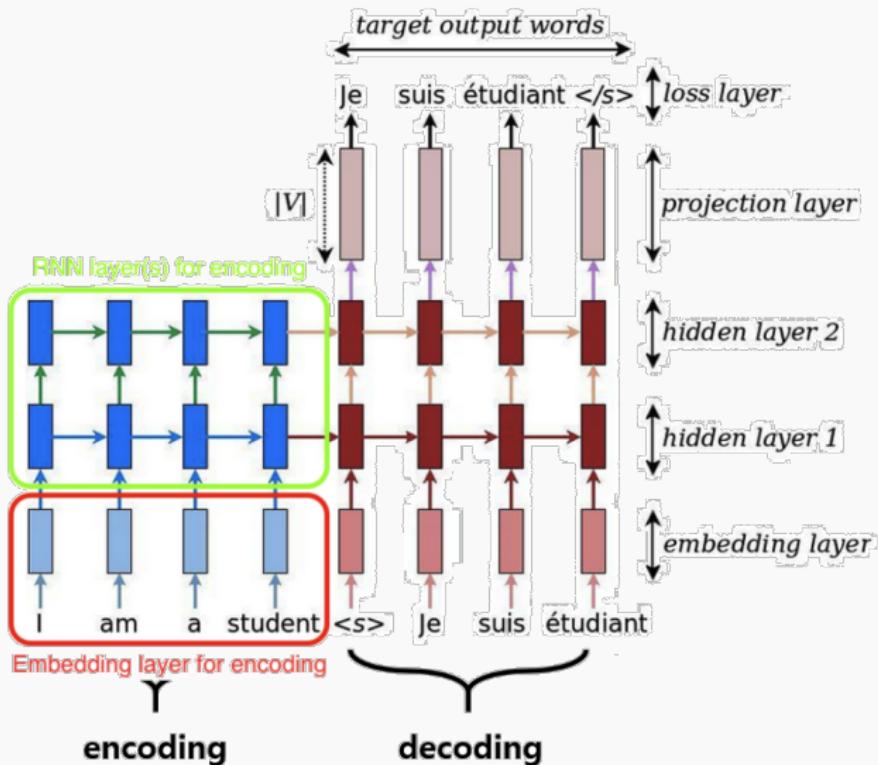
El Seq2seq es un codificador-decodificador secuencial que permite generar una secuencia en función de una secuencia de entrada.



Ejemplo seq2seq: Chatbot



Ejemplo seq2seq: Traducción automática neuronal



Outline : Embeddings de Contexto

Redes Recurrentes

RNN

LSTM

CNN, Arte y GANs

Deep Dream

GANs

Auto-codificadores y embeddings

AE

Codificador-Decodificador
secuencial: Seq2Seq

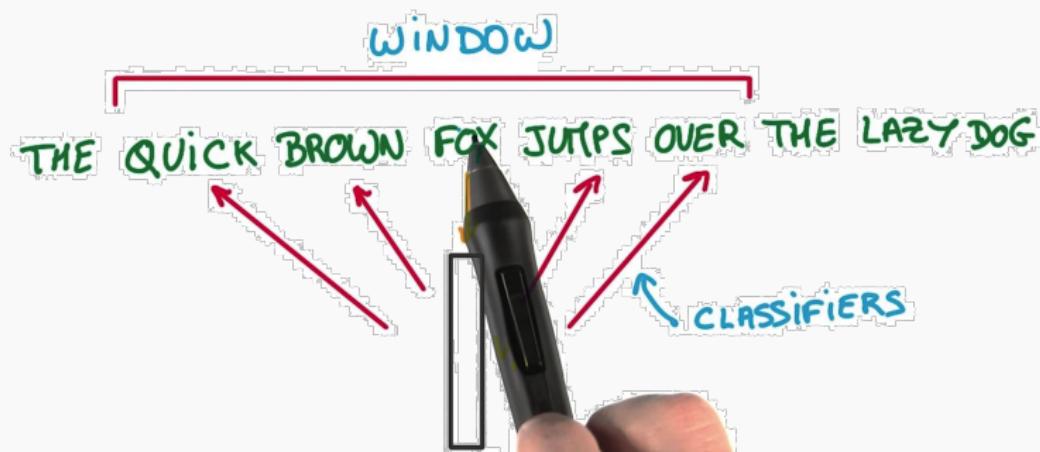
Embeddings de Contexto

Word Embeddings

Vectores densos que representan palabras en un espacio de baja dimensión. Una palabra se define por el contexto en el que se utiliza.

J.R. Firth

"Conocerás una palabra por la compañía que mantiene."



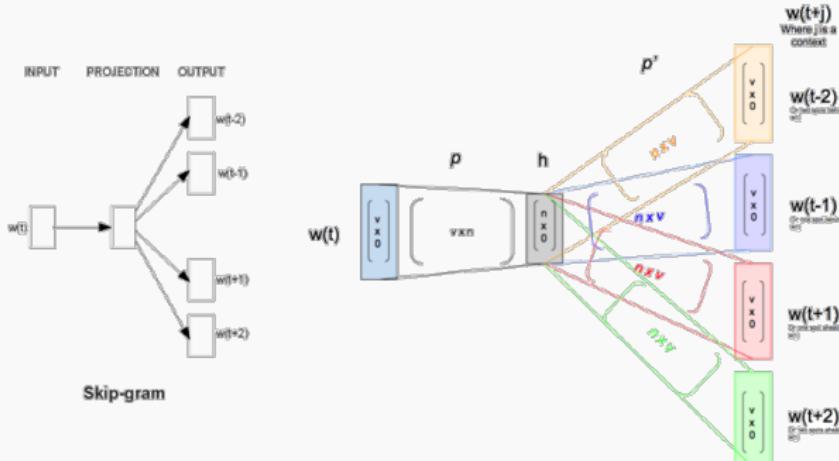
Dos técnicas de aprendizaje

- Skip-Gram: dado una ventana de contexto, se obtienen representaciones útiles para encontrar la palabra central.
- CBOW: dado la palabra, se obtienen representaciones útiles para encontrar las palabras del contexto.

Word Embeddings

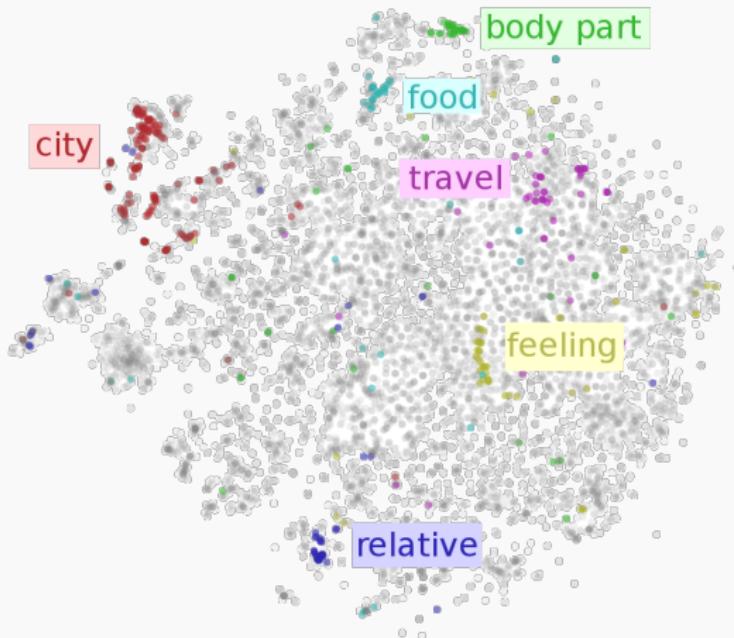
Reducción de la dimensión del espacio de vocabulario:

- De un vector disperso a un vector denso (por ejemplo, de 3,000,000 a 300 dimensiones).
- Las palabras están asociadas a una "ubicación geográfica" en el espacio de 300 dimensiones.
- Permite que el modelo aprenda conceptos (semántica) en lugar de cadenas de caracteres (palabras).



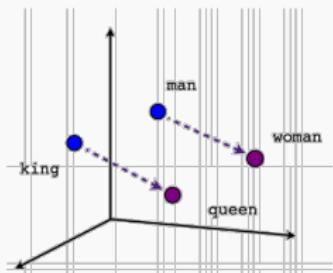
Word Embeddings

Palabras que aparecen en contextos similares tendrán vectores cercanos en el espacio de embeddings.

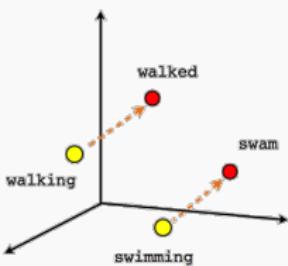


Word Embeddings

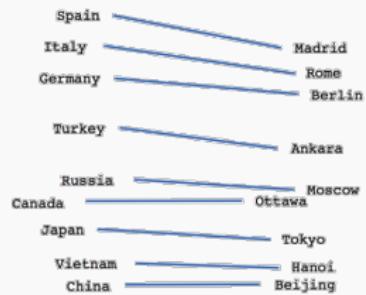
La semántica y los conceptos están integrados en los vectores, existen relaciones lineales entre ellos.



Male-Female



Verb tense



Country-Capital

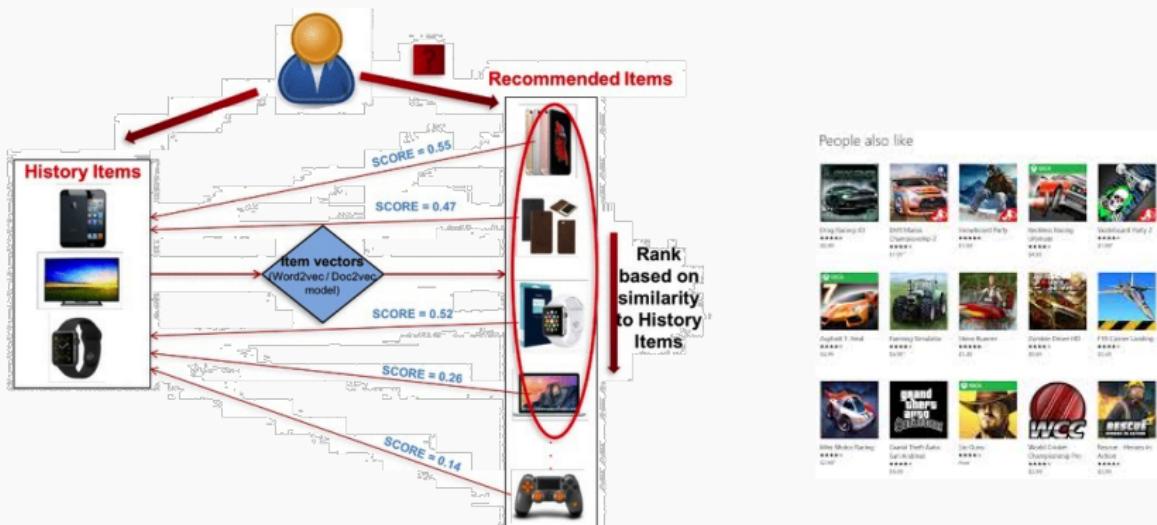
Toolbox gensim para el uso de word2vec

[Artículo](#)

Embeddings Generales

Las embeddings también se pueden adaptar a muchos otros problemas:

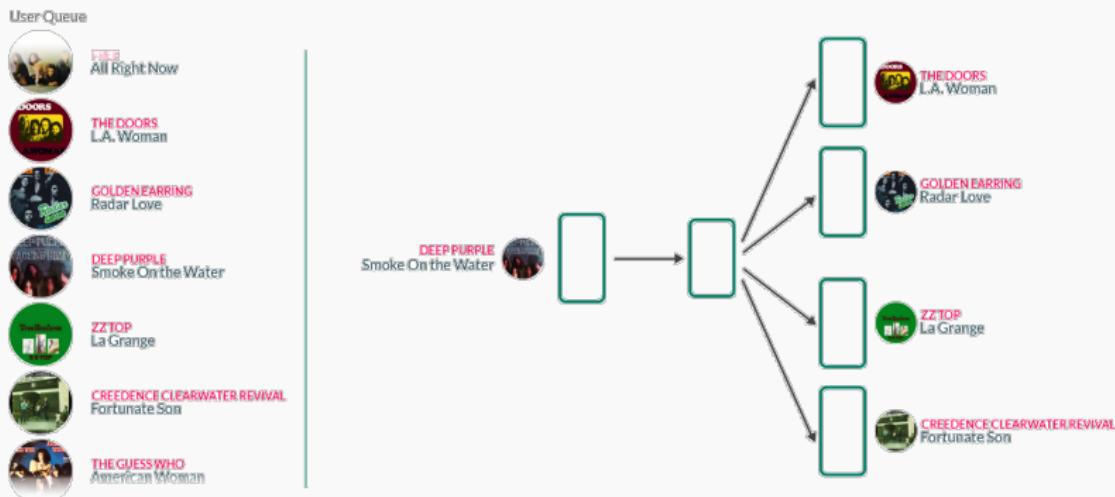
- Items



Embeddings Generales

Las embeddings también se pueden adaptar a muchos otros problemas:

- Items
- Música



Questions?