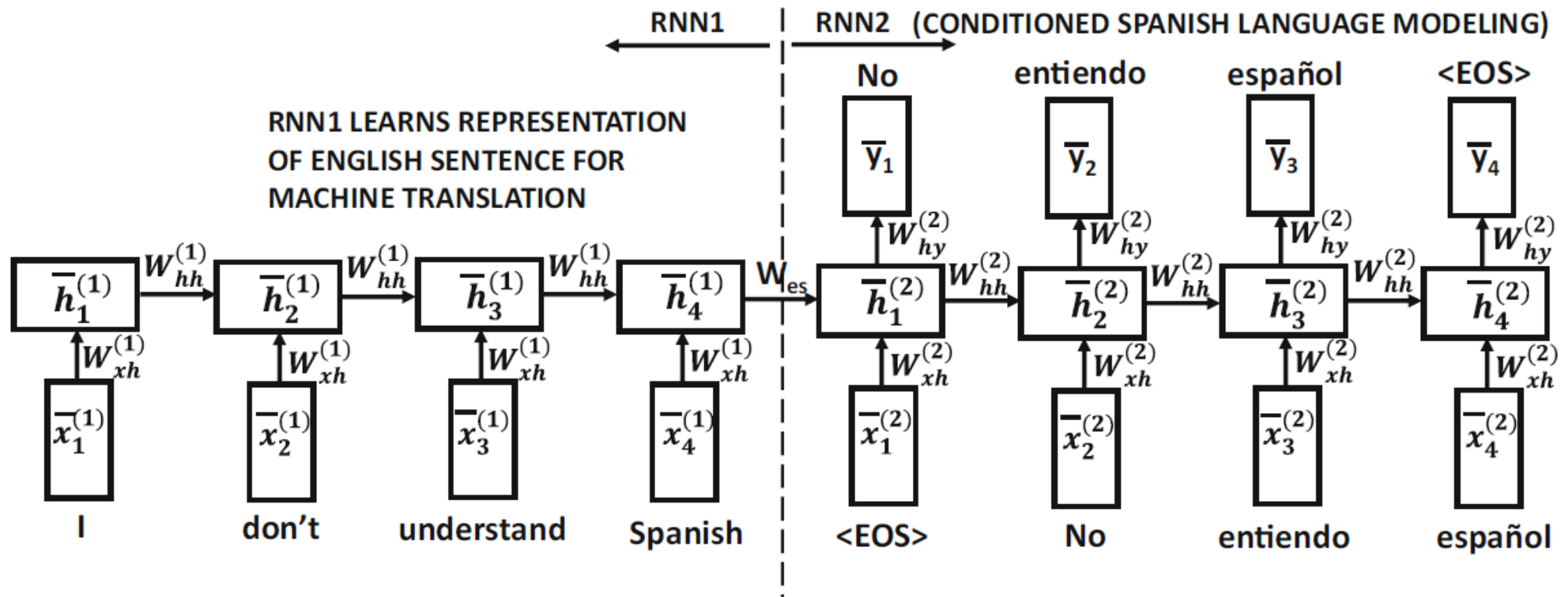


Consideré muy interesante cómo cambia la ar  
atención comparada con la forma que no la tiene.  
View, también es soluble con una rec

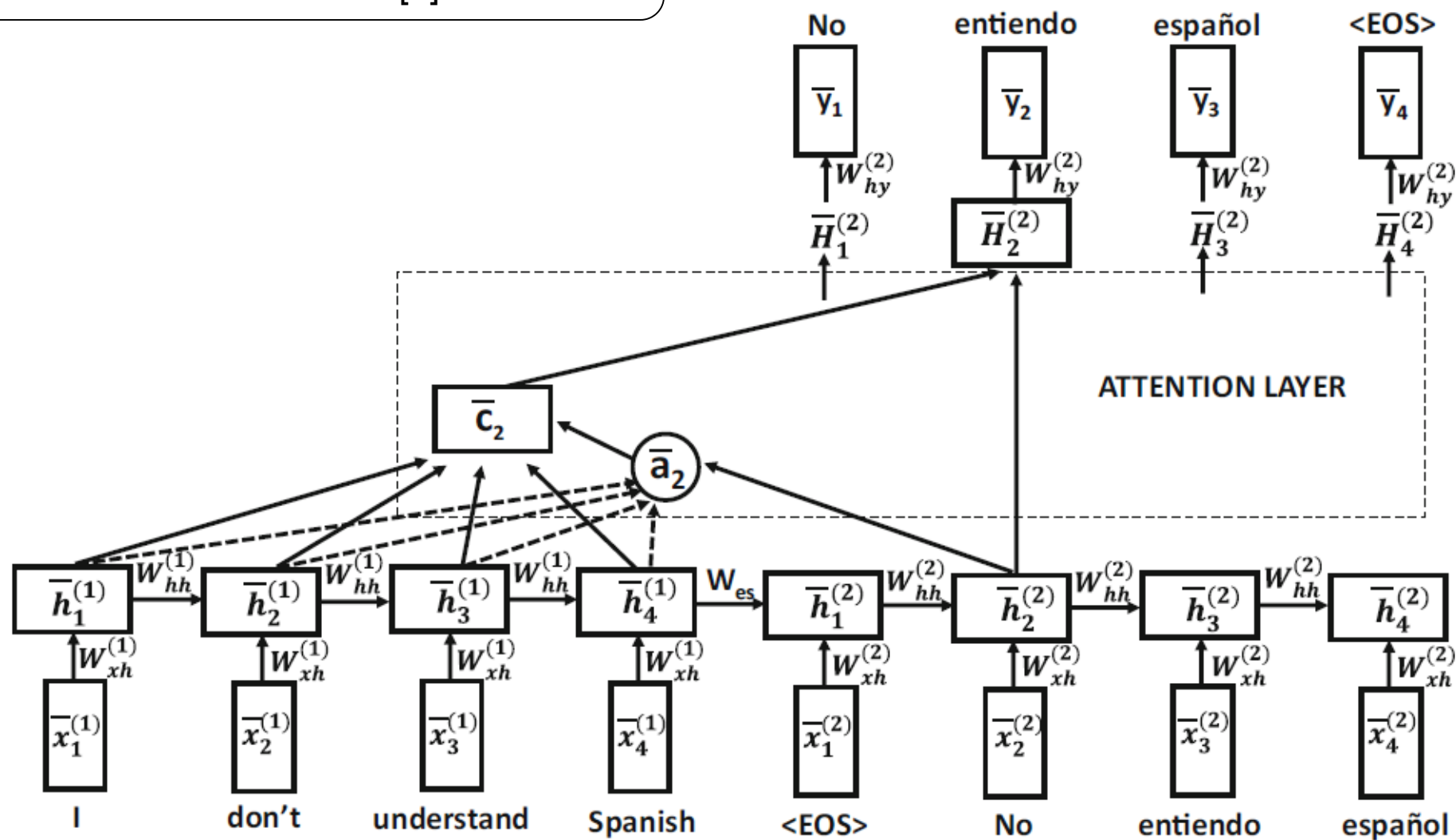


(a) Machine translation without attention

Como en realidad, es una clase de redes, y usa un mecanismo,

Sin embargo, claro que una red concreta tiene algoritmo y

arquitectura cuando se implementa la capa de  
De hecho, el ejemplo de los números de Street  
d sin el mecanismo de atención [1].



(b) Machine translation with attention

no tiene como tal un algoritmo

pertenece a la clase de atención

La selección mediante articulaciones de norma  $l_{2,1}$  de minimización es usada en bioinformática para seleccionar características y desechar las que no nos sirven, su algoritmo es el siguiente:

Entrada:  $A \in \mathbb{R}^{(n \times m)}$ ,  $Y \in \mathbb{R}^{(n \times c)}$

Resultado:  $U \in \mathbb{R}^{(m \times c)}$

Sea  $t = 0$ . Inicializar  $D(t) \in \mathbb{R}^{(m \times m)}$  como la matriz identidad

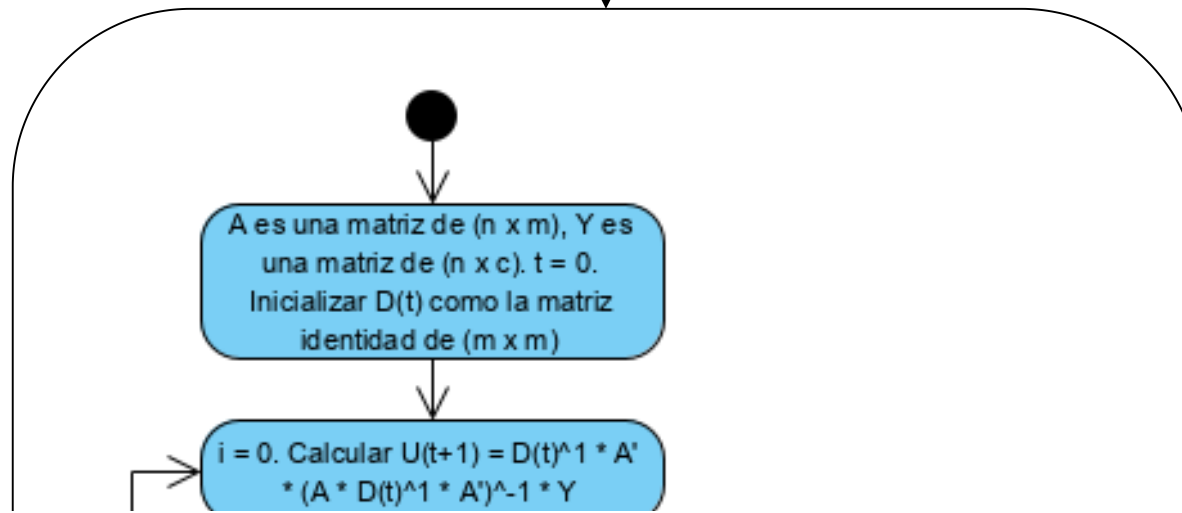
Repetir:

Calcular  $U_{t+1} = D_t^{-1} A^T (A D_t^{-1} A^T)^{-1} Y$

Calcular la matriz diagonal  $D(t+1)$  donde en  $i, i$  son  $\frac{1}{2 \|u_{t+1}^i\|_2}$

$t = t + 1$

Hasta Converger ( $D(t-2) == D(t-1)$ )



Se compara a la resolución del ojo humano para tener más resolución que la del ojo humano

La fovea en el ojo humano tiene una resolución: a lo largo de la línea directa poner

Mientras que el resto del ojo tiene una menor resolución, se presta atención a lo que

[1],

Algoritmo

Orígenes

Redes

os humanos de  
ción en el centro  
a sus orillas

o tiene la máxima  
o que miramos  
mos atención

esto del ojo tiene  
ción, Poniendo  
que ve directo

[5]

enes

es de

¿Qué es?

Representan más una clase de  
redes en vez de una arquitectura  
particular.

Emplean mecanismos de  
atención para prevenir que la  
información se desvanezca.

El sensor de vislumbramiento  
(glimpse sensor) es un  
mecanismo para imágenes

Se refiere a que en una imagen,  
tenga en muy alta resolución un  
pequeño fragmento

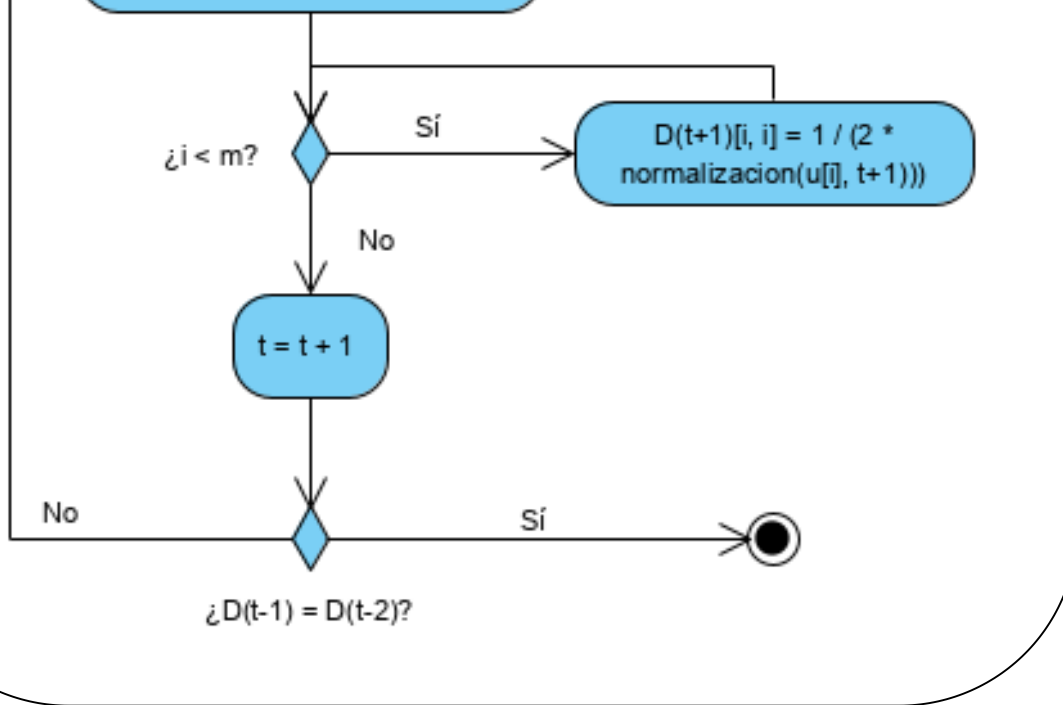
Lo logra guardando estados de la  
red previos y cambiando entre  
ellos.

En realidad son arquitecturas que  
tenga el mecanismo de atención

Una red de vislumbramiento  
puede ganarle a una red  
convolucional en clasificación

Es común mezclarlo con  
aprendizaje reforzado, para que  
se recompense de lo que hace

[1], [2]



[1], [5], [6]

Referencias

[1] Aggarwal, C., 2019. Neural Networks And Deep Learning. New York: Springer, cap. 10.

[3] Jaderberg, M., Simonyan, K., et al. 2015. Spatial transformer networks. Londres

[5] Mnih, V., Heess, N., & Graves, A. (2014). Recurrent models of visual attention.

[2] Leijnen, S., Veen, F. V. 2020. The Neural Network Zoo. MDPI

[4] Jaderberg, M., Simonyan, K., et al. 2015. Spatial transformer networks. Londres

[6] Nie, F., Huang, H., Cai, X., & Ding, C. (2010). Efficient and robust feature selection via joint  $\ell_2, 1$ -norms minimization.

ción

Ejemplo

a qué  
sa?

Puede dar interesantes datos en correlaciones de la entrada y las predicciones.

También puede servir en escritos donde se requiere atención al contexto: como la traducción

Igual para aplicaciones móviles como una cámara que escanea documentos, aunque en sí es un mecanismo y no una red concreta

[1], [3]

Implementación en Python de RAM  
<https://github.com/jlindsey15/RAM>

Las imágenes en Google Streetview, el sistema necesita conectar casas con su número

Pero no puede usar el número porque es un dato sensible que debe permanecer borroso

¿Puede encontrar una forma de conectar las imágenes sin el número de la casa?

Sí, con aprendizaje profundo y atención. Se fija en detalles para comparar y encadenar fotos.

Es como si un humano dijera ¿Dónde miro para recordar esta casa?

[1], [3]