

Redes Recurrentes

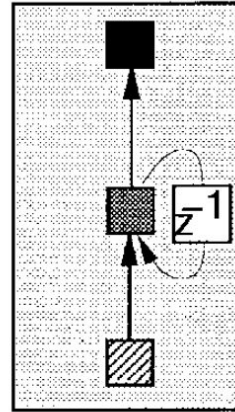
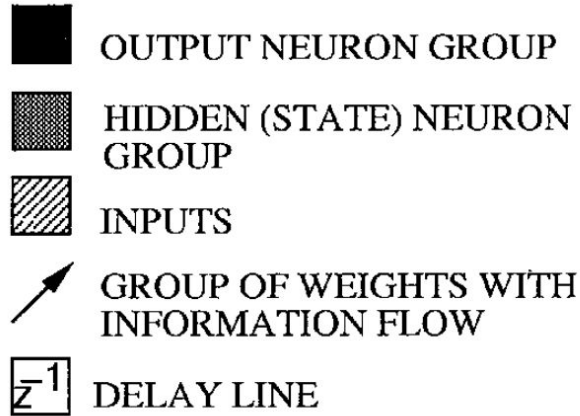
Orlando Ramos Flores

Contenido

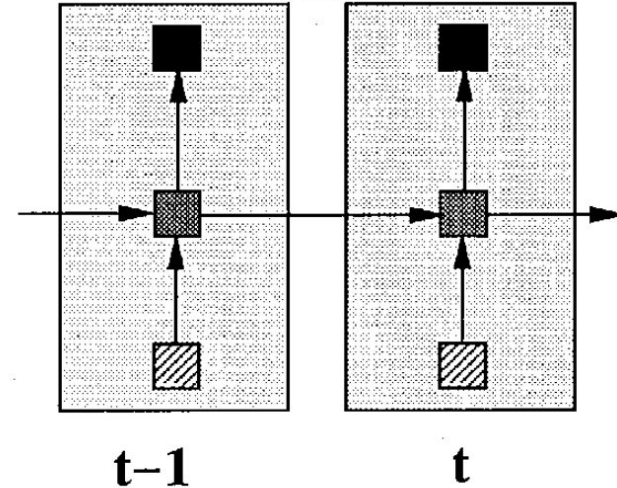
- Introducción
- Bidirectional Long Short-Term Memory (Bi-LSTM)
 - Arquitectura de una Bi-LSTM
 - Resumen

Introducción

Introducción: Red Recurrente



(a)



(b)

General structure of a regular unidirectional RNN shown (a) with a delay line and (b) unfolded in time for two time steps.

Introducción:

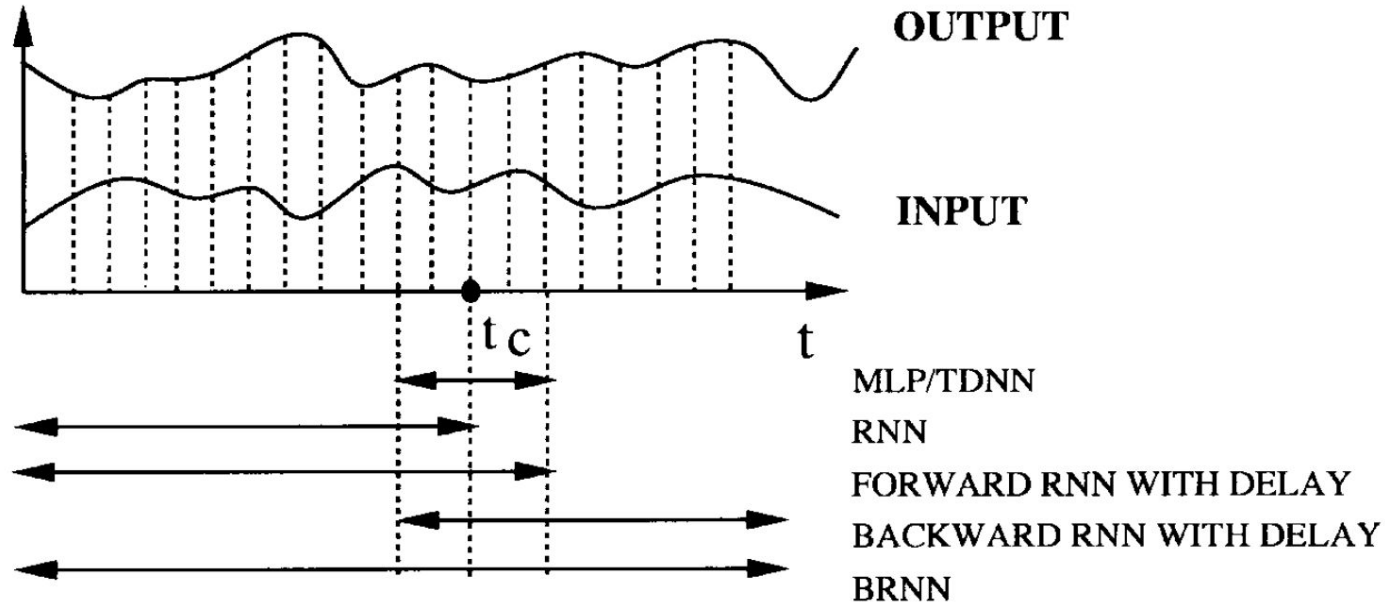
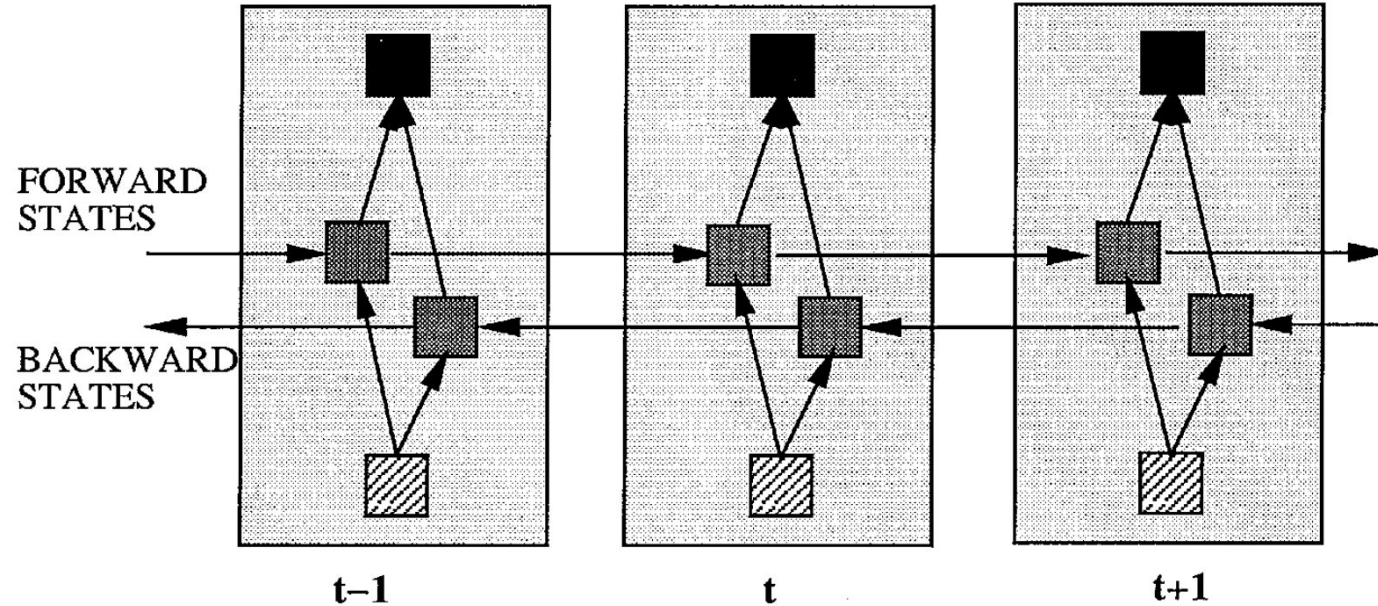


Fig. 2. Visualization of the amount of input information used for prediction by different network structures.

Introducción:

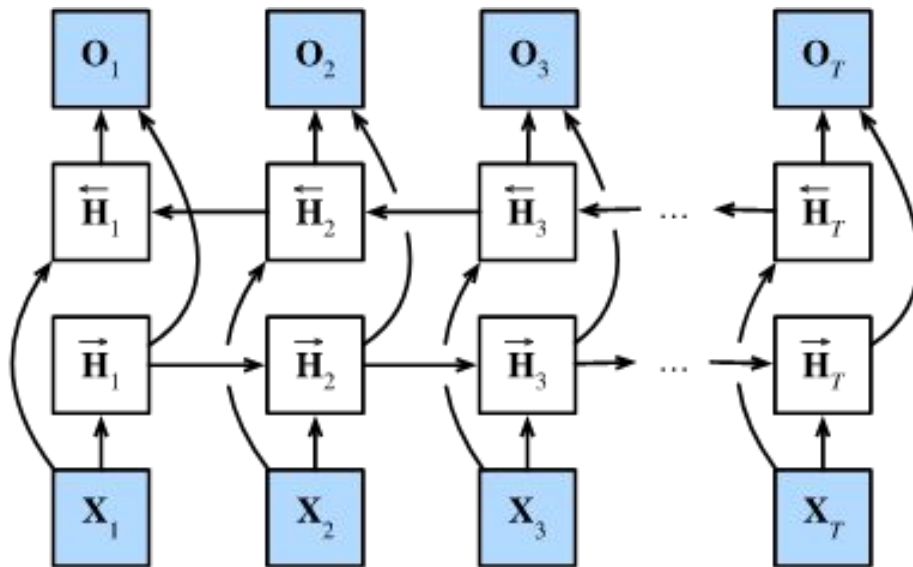


3. General structure of the bidirectional recurrent neural network (BRNN) shown unfolded in time for three time steps.

Bidirectional Long Short-Term Memory

Bidirectional Long Short-Term Memory: Bi-LSTM

- El modelo LSTM bidireccional (Bi-LSTM) mantiene dos estados separados para las entradas hacia adelante y hacia atrás generadas por dos LSTM diferentes.
- La primera LSTM es una secuencia regular que comienza desde el principio de la oración, mientras que en la segunda LSTM, la secuencia de entrada se alimenta en el orden opuesto.
- La idea detrás de la red bidireccional es capturar información de las entradas circundantes.
- Por lo general, aprende más rápido que el enfoque unidireccional, aunque depende de la tarea.



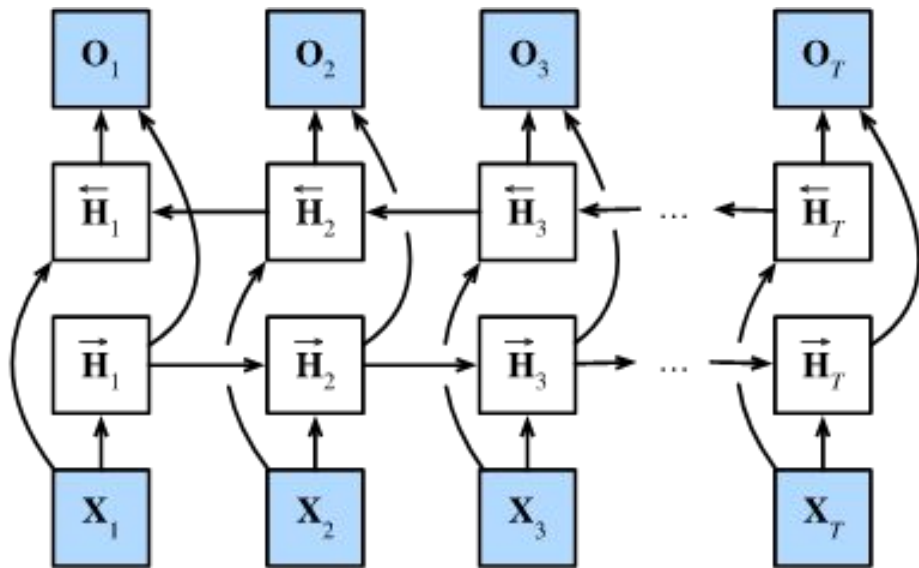
Bi-LSTM

- Para cualquier paso t dado un mini-batch de entrada $X_t \in \mathbb{R}^{n \times d}$, donde n es el número de muestras y d es el número de entradas en cada muestra. Y ϕ sea la función de activación de la capa oculta.
- Los estados ocultos hacia adelante y hacia atrás para este paso de tiempo son:

$$\vec{H}_t \in \mathbb{R}^{n \times h} \text{ y } \overleftarrow{H}_t \in \mathbb{R}^{n \times h}$$

donde h es el número de unidades ocultas.

- Las actualizaciones de los estados ocultos hacia adelante y hacia atrás son los siguientes: $\vec{H}_t = \phi(X_t W_{xh}^{(f)} + \vec{H}_{t-1} W_{hh}^{(f)} + b_h^{(f)})$,
 $\overleftarrow{H}_t = \phi(X_t W_{xh}^{(b)} + \overleftarrow{H}_{t+1} W_{hh}^{(b)} + b_h^{(b)})$

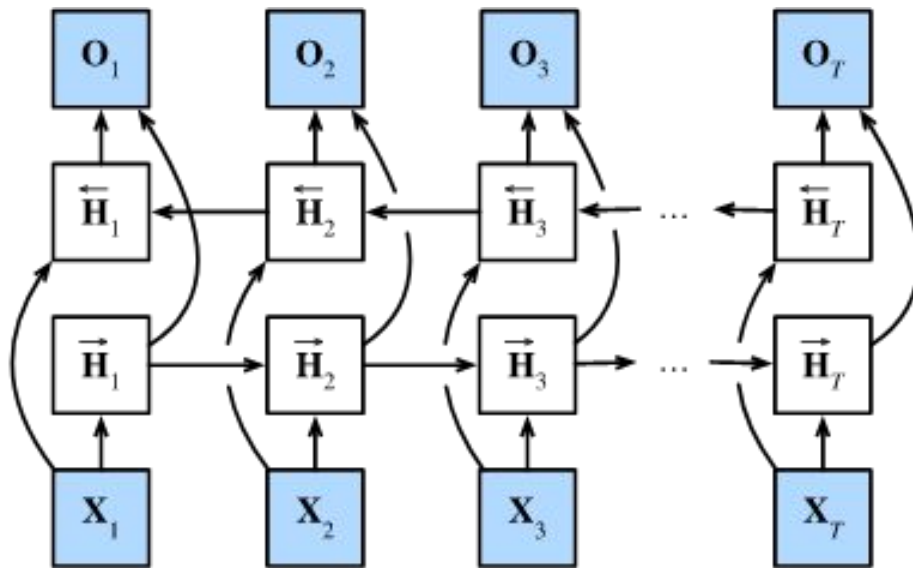


$$W_{xh}^{(f)} \in \mathbb{R}^{d \times h}, W_{hh}^{(f)} \in \mathbb{R}^{h \times h}, W_{xh}^{(b)} \in \mathbb{R}^{d \times h}, \text{ and } W_{hh}^{(b)} \in \mathbb{R}^{h \times h}$$

$$b_h^{(f)} \in \mathbb{R}^{1 \times h} \text{ y } b_h^{(b)} \in \mathbb{R}^{1 \times h}$$

Bi-LSTM

- Después, se concatenan los estados ocultos hacia adelante y hacia atrás $\vec{\mathbf{H}}_t$ y $\overleftarrow{\mathbf{H}}_t$ para obtener el estado oculto $\mathbf{H}_t \in \mathbb{R}^{n \times 2h}$ con este alimentar a la capa de salida.
- En Bi-LSTM profundos con múltiples capas ocultas, dicha información se transmite como entrada a la siguiente capa bidireccional.
- Por último, la capa de salida calcula la salida $\mathbf{O}_t \in \mathbb{R}^{n \times q}$, donde q es el número de salidas (targets).



$$\mathbf{O}_t = \mathbf{H}_t \mathbf{W}_{hq} + \mathbf{b}_q$$

$$\mathbf{W}_{hq} \in \mathbb{R}^{2h \times q} \quad \mathbf{b}_q \in \mathbb{R}^{1 \times q}$$

Bi-LSTM: Resumen

- Una de las características clave de una RNN bidireccional es que la información de ambos extremos de la secuencia se usa para estimar la salida.
- Es decir, usamos información de observaciones futuras y pasadas para predecir la actual.
- Sin embargo, las RNN bidireccionales también son extremadamente lentas.
- Las razones principales de esto son que la propagación hacia adelante requiere recurrencias tanto hacia adelante como hacia atrás en capas bidireccionales y que la propagación hacia atrás depende de los resultados de la propagación hacia adelante.
- Por lo tanto, los gradientes tendrán una cadena de dependencia muy larga.