

LONG SHORT-TERM MEMORY NETWORK

- Jack Christopher Huaihua Huayhua
- Angel Tomas Concha Layme
- Jean Pierre Chavez Guevara

CONTENIDO

01

Artículo

02

Estructura

03

Datos de
entrenamiento

04

Áreas de
aplicación

05

Ventajas y
desventajas

05

Conclusiones





| 01

Artículo

LONG SHORT TERM MEMORY

Año: 1997

Medio de publicación: Journal Neural
Computation.

Autores: Sepp Hochreiter y Jürgen
Schmidhuber.

LONG SHORT-TERM MEMORY

NEURAL COMPUTATION 9(8):1735–1780, 1997

Sepp Hochreiter
Fakultät für Informatik
Technische Universität München
80290 München, Germany
hochreit@informatik.tu-muenchen.de
<http://www7.informatik.tu-muenchen.de/~hochreit>

Jürgen Schmidhuber
IDSIA
Corso Elvezia 36
6900 Lugano, Switzerland
juergen@idsia.ch
<http://www.idsia.ch/~juergen>

MOTIVO

Superar problemas de desvanecimiento y explosión de errores en RNN tradicionales.

IMPACTO

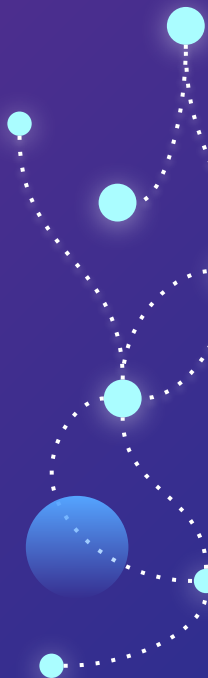
Aplicaciones en reconocimiento de voz, procesamiento del lenguaje natural y predicciones secuenciales.

PUBLICACIÓN

Revista científica especializada en inteligencia artificial, aprendizaje automático o procesamiento de lenguaje natural.

RESULTADOS

Capacidad para resolver problemas complejos con largos retrasos temporales.





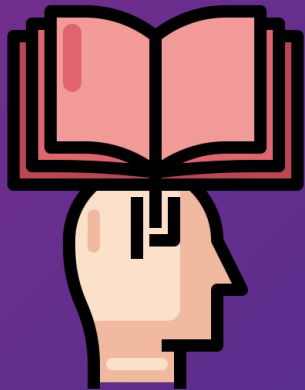
02

Estructura

Idea Principal

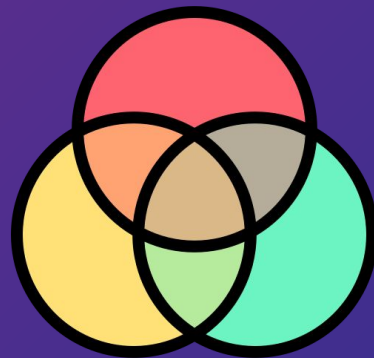
Memorización

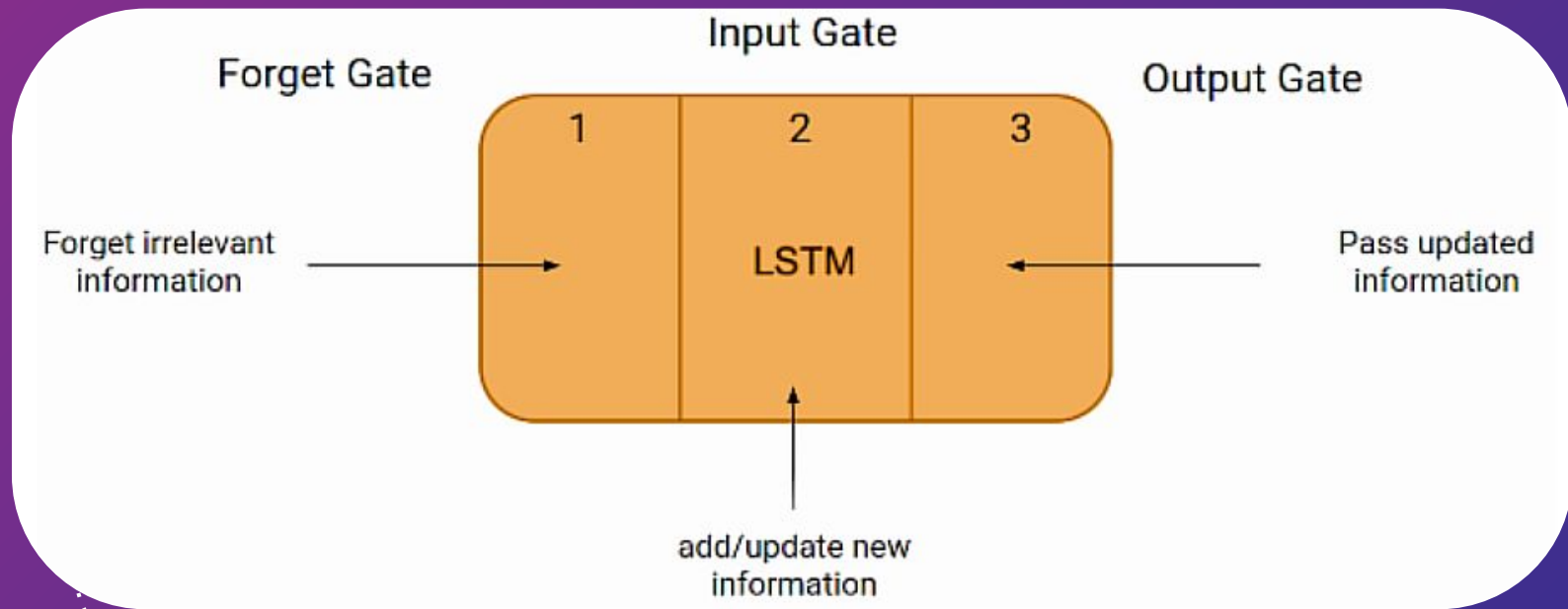
Se gana información → se almacena para uso futuro

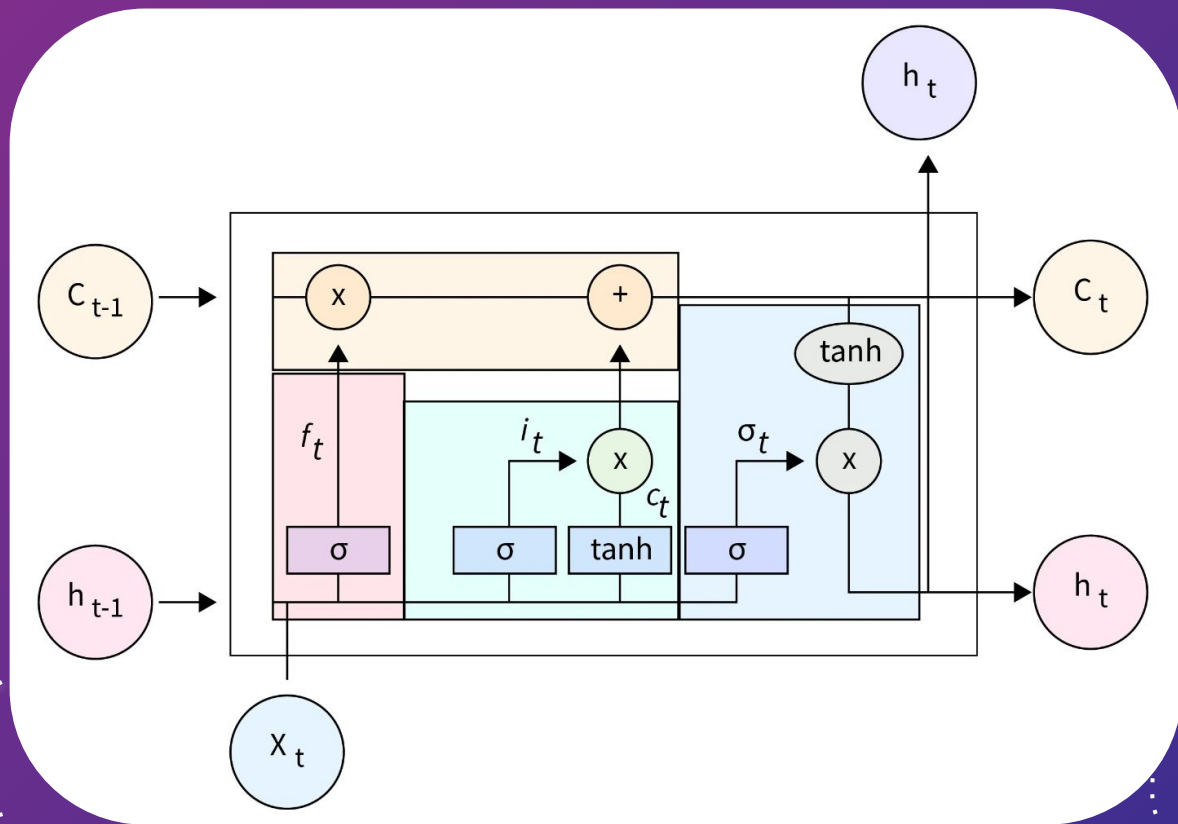


Combinación

Información almacenada + habilidades analíticas





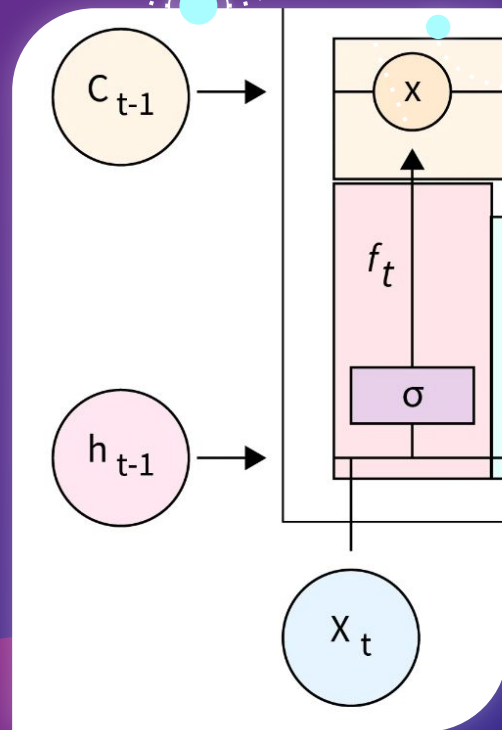


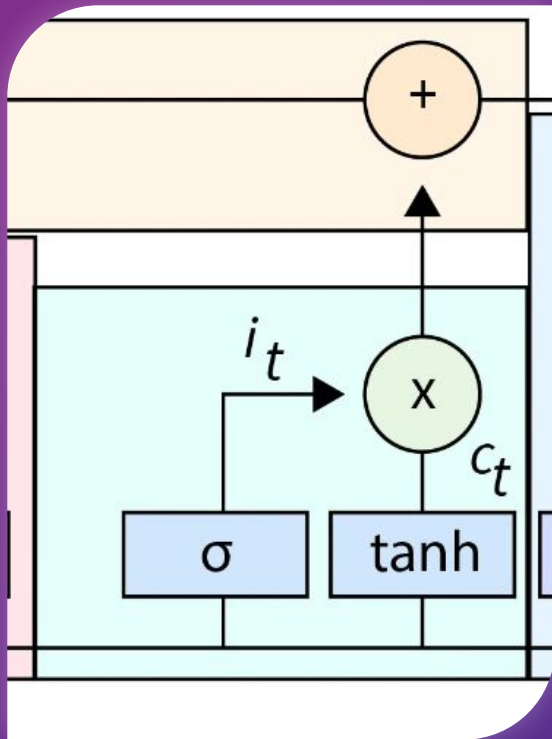
FORGET GATE

Decide qué información debe eliminarse del **estado de la celda**.

Toma el **estado oculto** anterior y la entrada actual y lo pasa a una función de activación Sigmoid.

Genera un valor entre 0 y 1, donde 0 significa olvidar y 1 significa mantener.





INPUT GATE

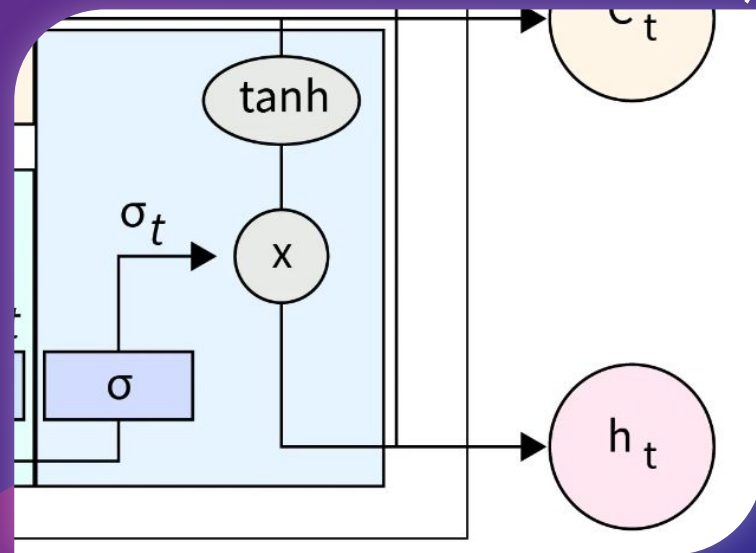
Considera la **entrada actual** y el **estado oculto** anterior para actualizar el valor del **estado de la celda**.

- Función de activación de Sigmoide → Decide qué porcentaje de la información se requiere.
- Función de activación de Tanh → Mapea los datos entre -1 y 1, luego la multiplica por la función Sigmoide.

OUTPUT GATE

Devuelve el **estado oculto** para la próxima ocasión.

- Función de activación Sigmoide → Decide el porcentaje de información relevante requerida.
- Función de activación Tanh → Toma el **estado de la celda** actualizada y la multiplica por la función de activación Sigmoide.

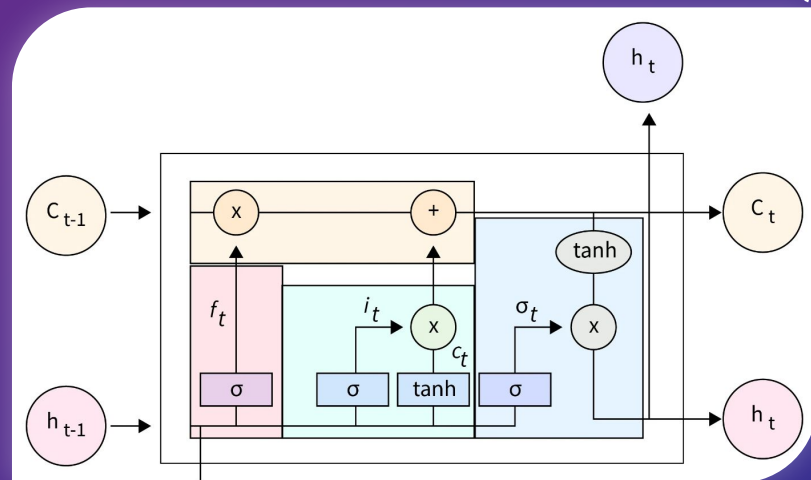


CELL STATE

La **puerta de olvido** y la **puerta de entrada** actualizan la **celda de estado**.

La **celda de estado** anterior se multiplica por la salida de la **puerta de olvido**, se suma con la salida de la **puerta de entrada**.

Este valor se usa luego para calcular el estado oculto en la puerta de salida.





03

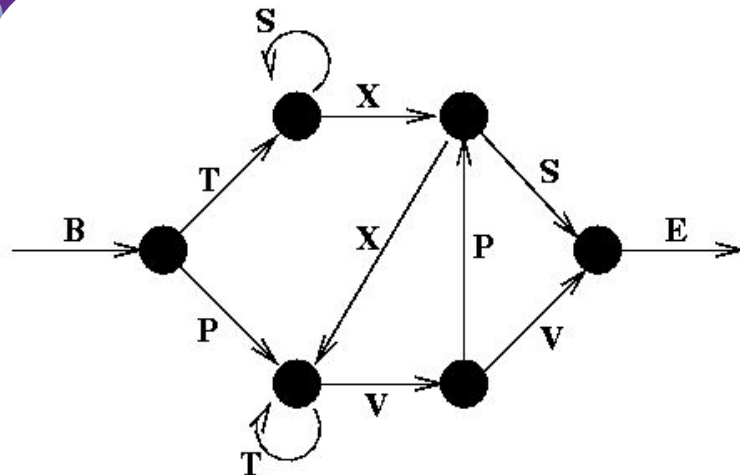
Datos de
entrenamiento

Embedded Reber

Es una gramatica

Genera una secuencia de caracteres

Sigue una logica determinista

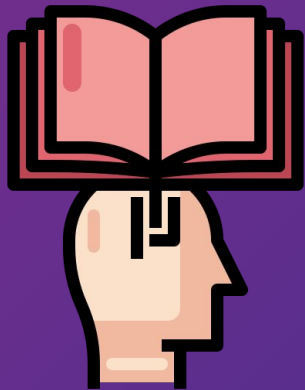


LSTM - REBER



APRENDER

Las reglas que usaba la gramática



PREDECIR

Los caracteres que siguen



Otros datos

Secuencias de datos como texto, audio, video.

Corpus de noticias, libros, artículos científicos, sitios web

Datos de entrenamiento de Stanford Question Answering Dataset (SQuAD)





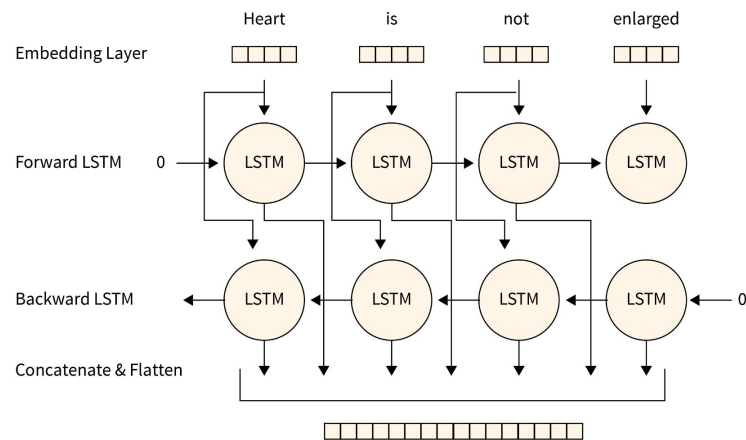
| 04

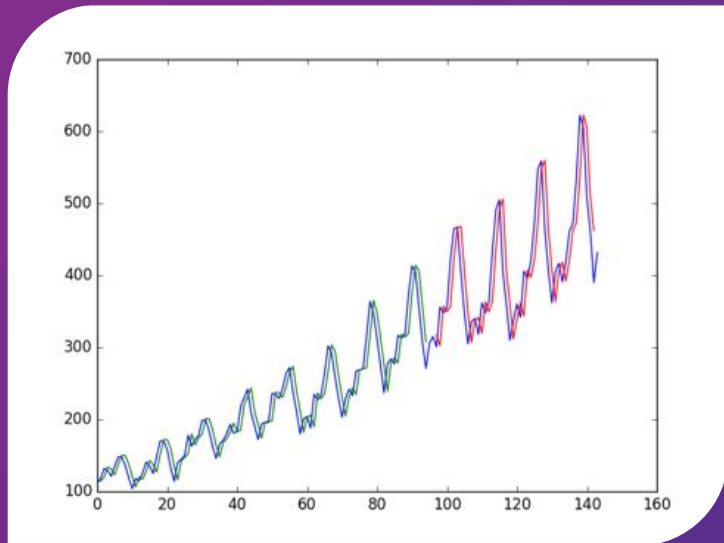
Áreas de aplicación

Modelado de lenguaje

LSTM es capaz de: crear modelos de lenguaje

Son usados en: sistemas de traducción automática o chatbots





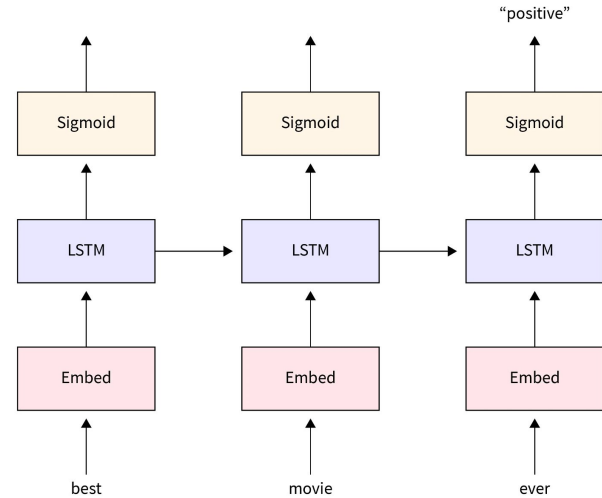
Predicción de series temporales

LSTM es capaz de: modelar datos de series temporales y predecir valores futuros en la serie

Son usados para: predecir precios de acciones o patrones de tráfico.

Análisis de sentimientos

LSTM es utilizado para: analizar y clasificar las emociones expresadas en un texto.





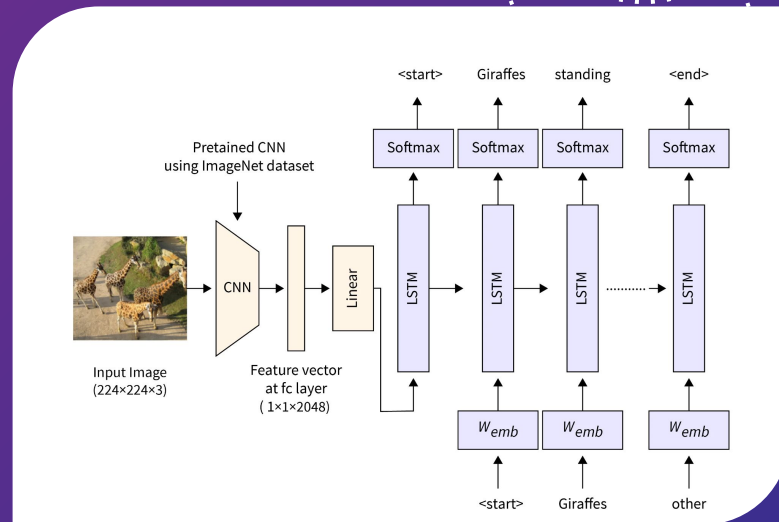
Reconocimiento de voz

LSTM es capaz de: construir sistemas de reconocimiento de voz

Son usados los siguientes casos: asistentes virtuales , automatización de tareas, accesibilidad y seguridad

Subtítulos de imágenes

LSTM es utilizado para: generar subtítulos descriptivos para imágenes, como en motores de búsqueda de imágenes





05

VENTAJAS Y DESVENTAJAS

VENTAJAS

Captura de dependencias a largo plazo.

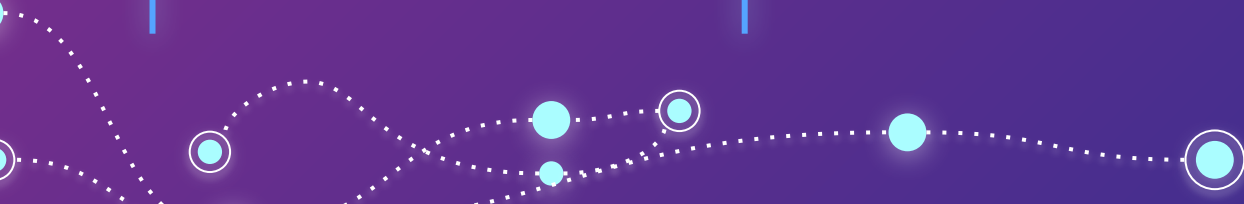
Manejo de secuencias de longitud variable.

Aprendizaje de representaciones jerárquicas.

Memoria de largo plazo.

Resistencia al ruido en los datos.

Adaptabilidad a diferentes dominios.



DESVENTAJAS

Mayor
complejidad
computacional

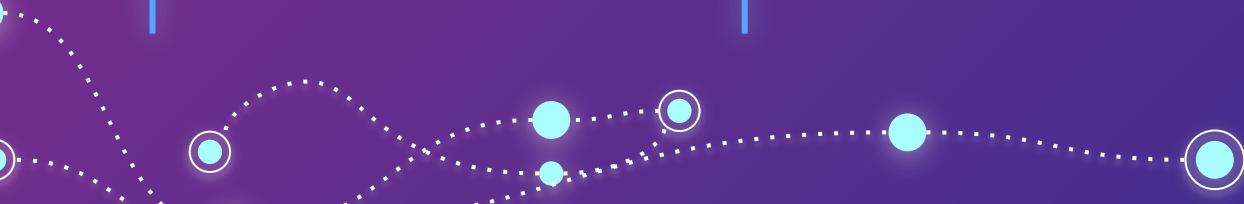
Mayor
dificultad de
interpretación.

Requiere
tiempo y
experiencia
para su
configuración.

Mayor
consumo de
recursos.

Riesgo de
sobreajuste.

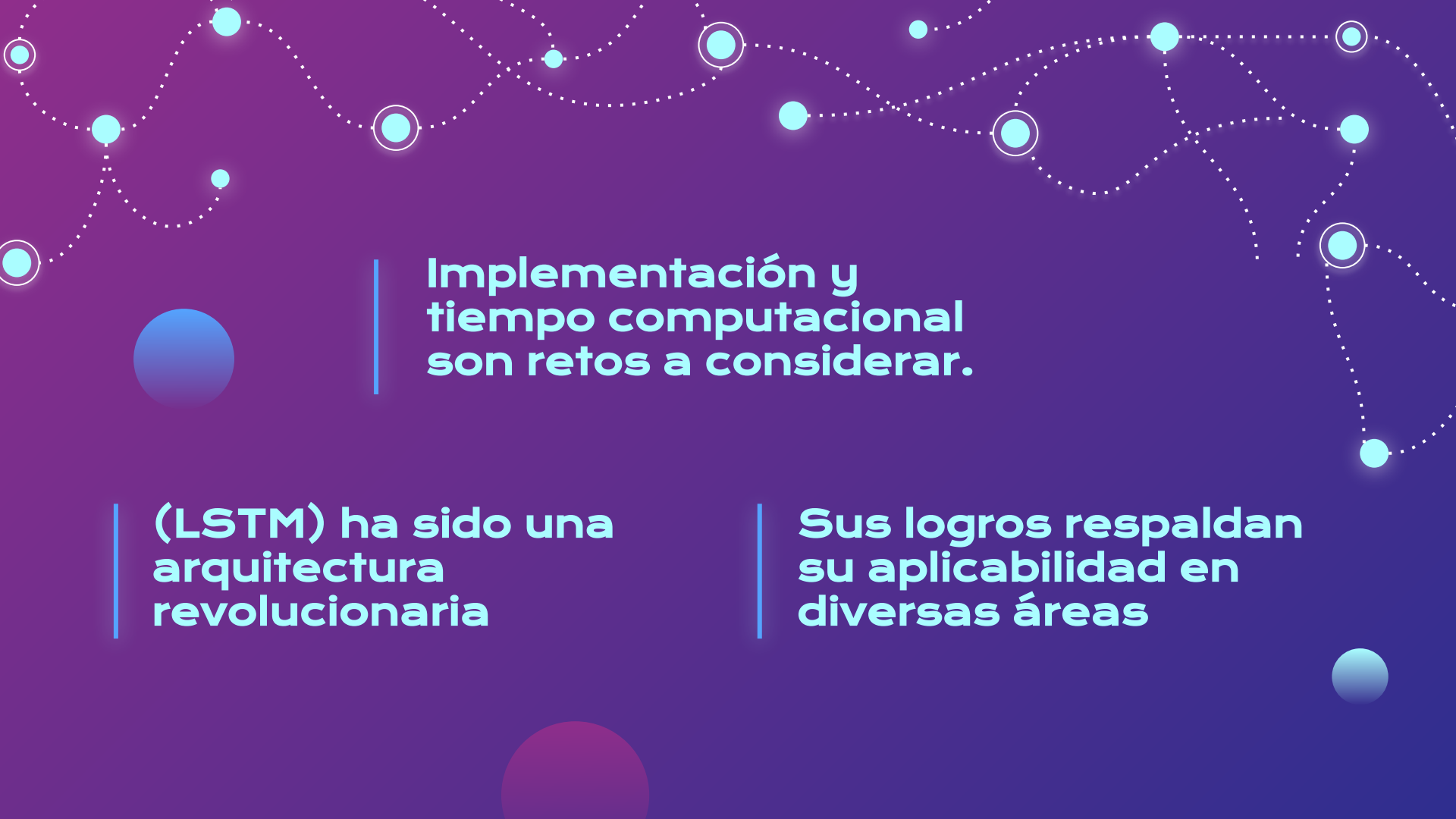
Sensibilidad a la
selección de
hiperparámetros





06

CONCLUSIONES



**Implementación y
tiempo computacional
son retos a considerar.**

**(LSTM) ha sido una
arquitectura
revolucionaria**

**Sus logros respaldan
su aplicabilidad en
diversas áreas**