



Juan José
García Cedeño
juajogar@espol.edu.ec

Resumen:

Las redes neuronales son modelos de *machine learning* con una capacidad de generalización extraordinaria y misteriosa. Pueden resolver tareas complejas como reconocer actividades en un video. Sin embargo, estos modelos usualmente carecen de una justificación matemática en su diseño. Esta tesis se desvía de la tendencia y diseña una red neuronal, para reconocer actividades humanas, con sustento matemático. La arquitectura propuesta, denominada D-RNN, logra resultados de clasificación comparables a modelos más complejos, pero con una fracción de los parámetros. Convirtiéndose en una alternativa viable cuando el porcentaje de error puede intercambiarse por un modelo más liviano.

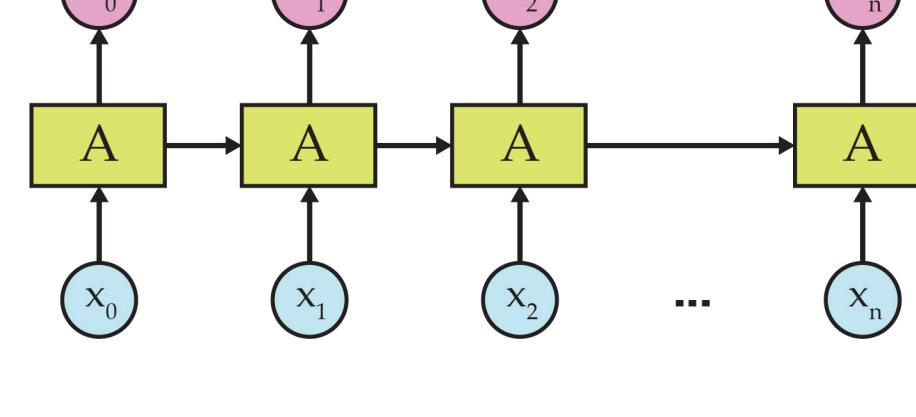
Introducción

Las redes neuronales recurrentes son arquitecturas utilizadas para clasificar secuencias. Son la base para resolver tareas como reconocimiento de actividades y traducción. Sin embargo, no son muy utilizadas porque se les dificulta aprender dependencias temporales extensas.

Implementaciones actuales acomplejan el modelo para solucionar el problema (e.g. LSTM). Esta tesis propone una solución más simple, y la evalúa en el contexto del reconocimiento de actividades.

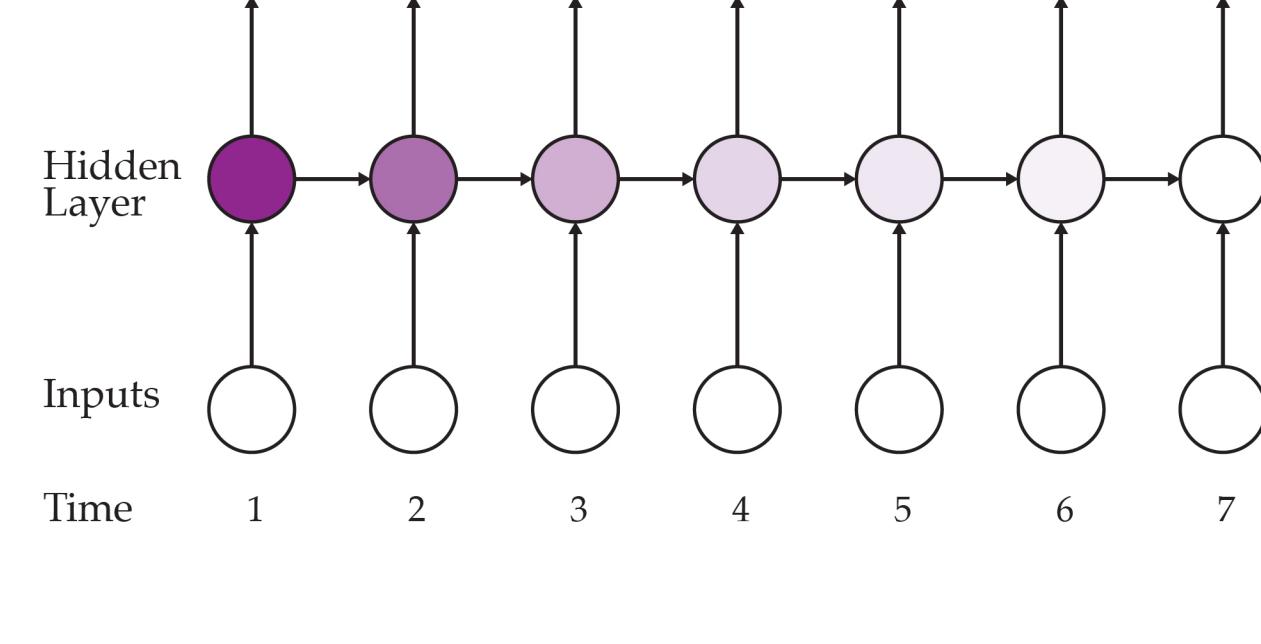
Redes neuronales recurrentes (RNN)

Es una variación de las redes neuronales, diseñada para procesar secuencias. Su característica es la unidad de recurrencia, y es lo que conecta la salida de la red con las entradas anteriores.



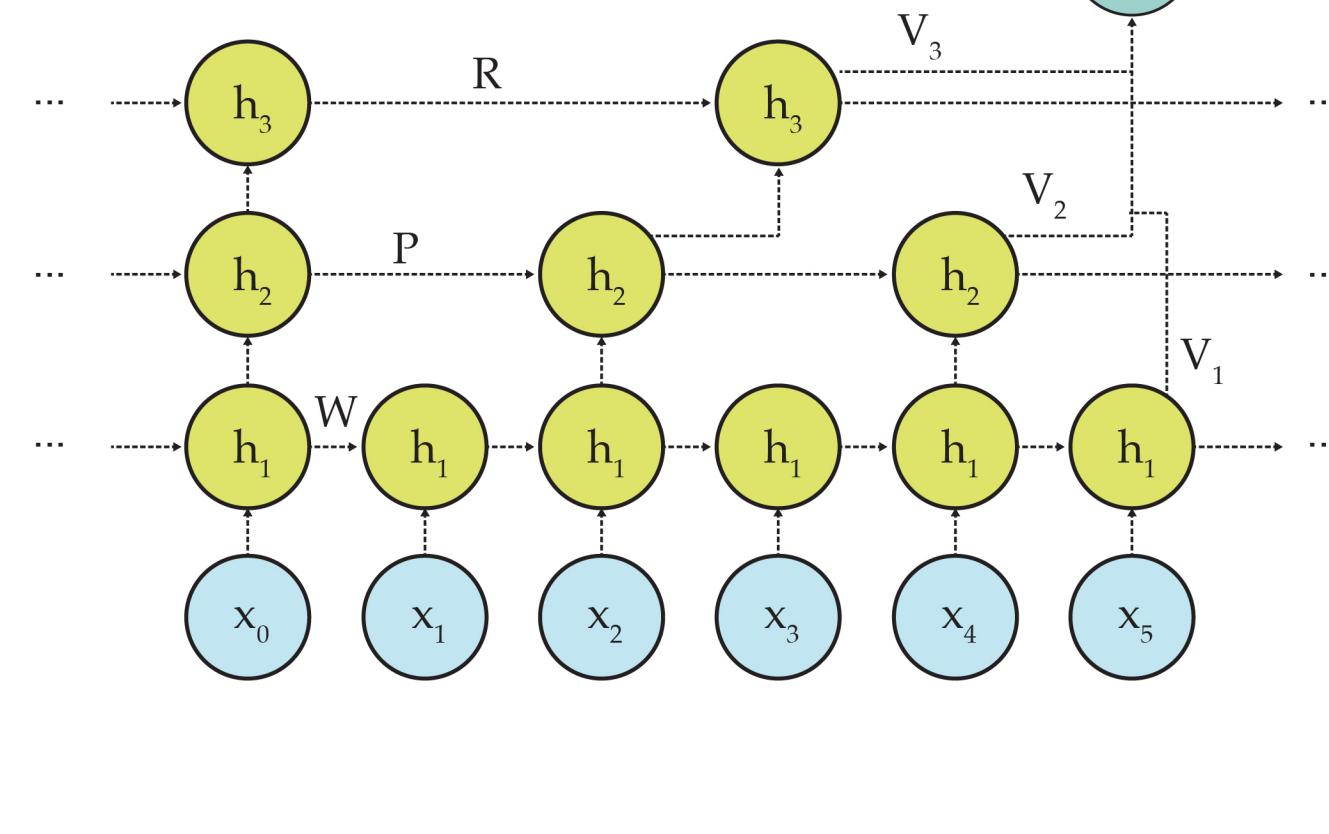
Problema de las RNN

La unidad recurrente es un factor que multiplica todas las entradas conforme son procesadas. Esto causa que la salida tenga una dependencia exponencial con respecto a las entradas.



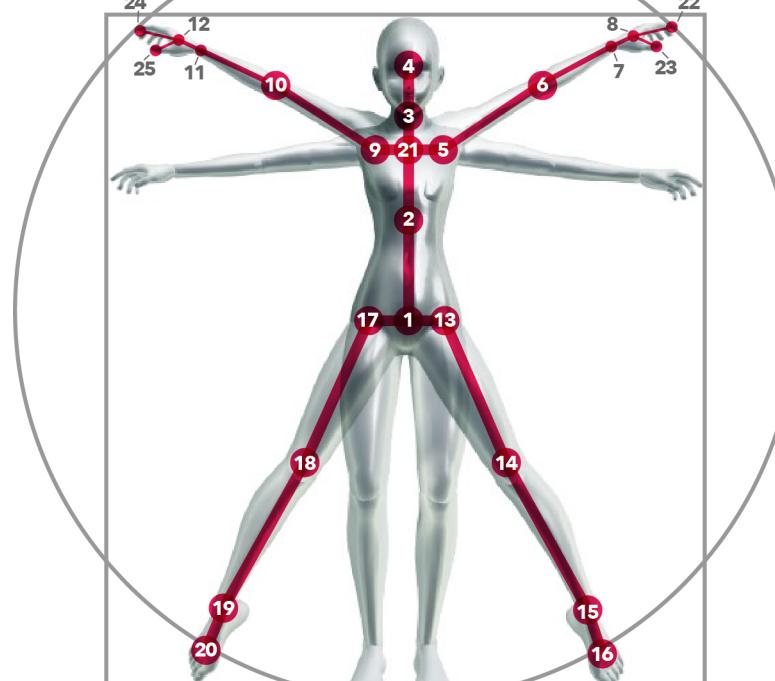
Solución propuesta (DRNN)

Utilizar unidades recurrentes que procesan sus entradas a diferentes escalas de tiempo; reduciendo así el número de transformaciones entre las entradas y la salida.



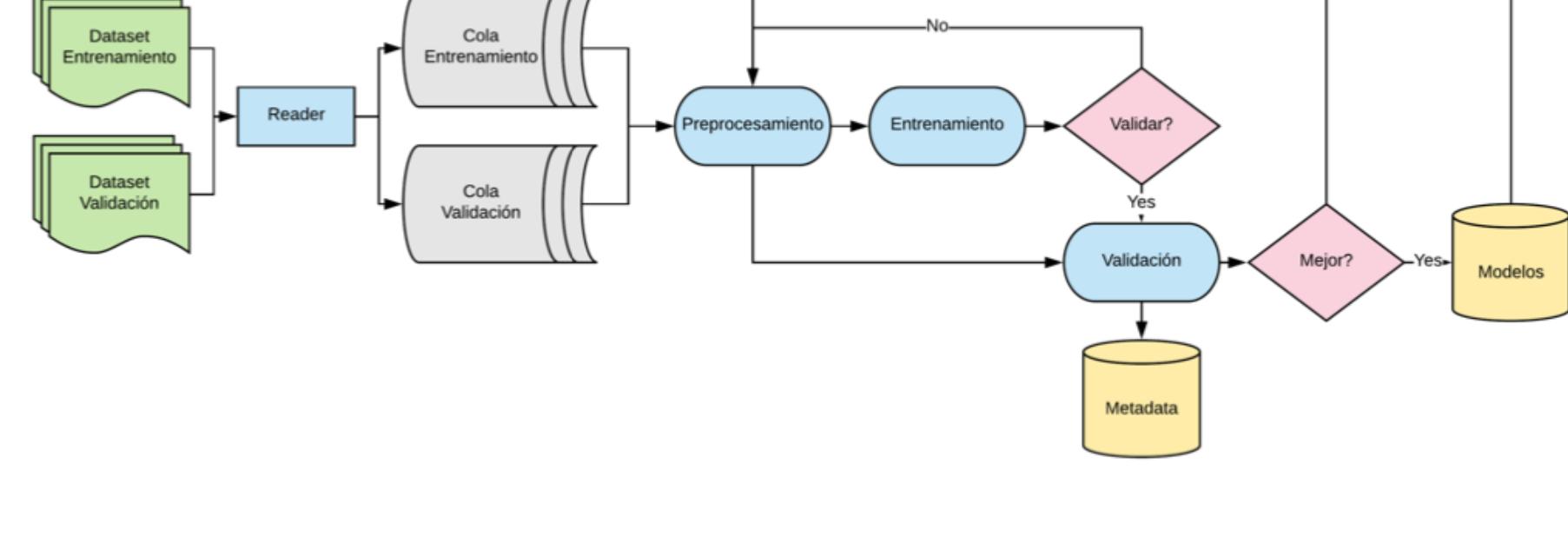
Evaluación de la solución

Se entrena tres modelos recurrentes para clasificar las actividades que suceden en un video. Los modelos son: La red neuronal recurrente (RNN), la solución propuesta (DRNN), y la arquitectura más utilizada para reconocer actividades (LSTM). El dataset de entrenamiento consta de 60 actividades, filmadas con tres cámaras Kinect V2, desde tres ángulos distintos. Todos los modelos son entrenados con la perspectiva de dos cámaras y evaluados con la perspectiva de la tercera. La entrada de los modelos es la posición espacial (i.e. coordenadas x,y,z) de 25 junturas del cuerpo. Los tres modelos se comparan con las siguientes métricas: porcentaje de clasificaciones correctas, tiempo de convergencia y número de parámetros.



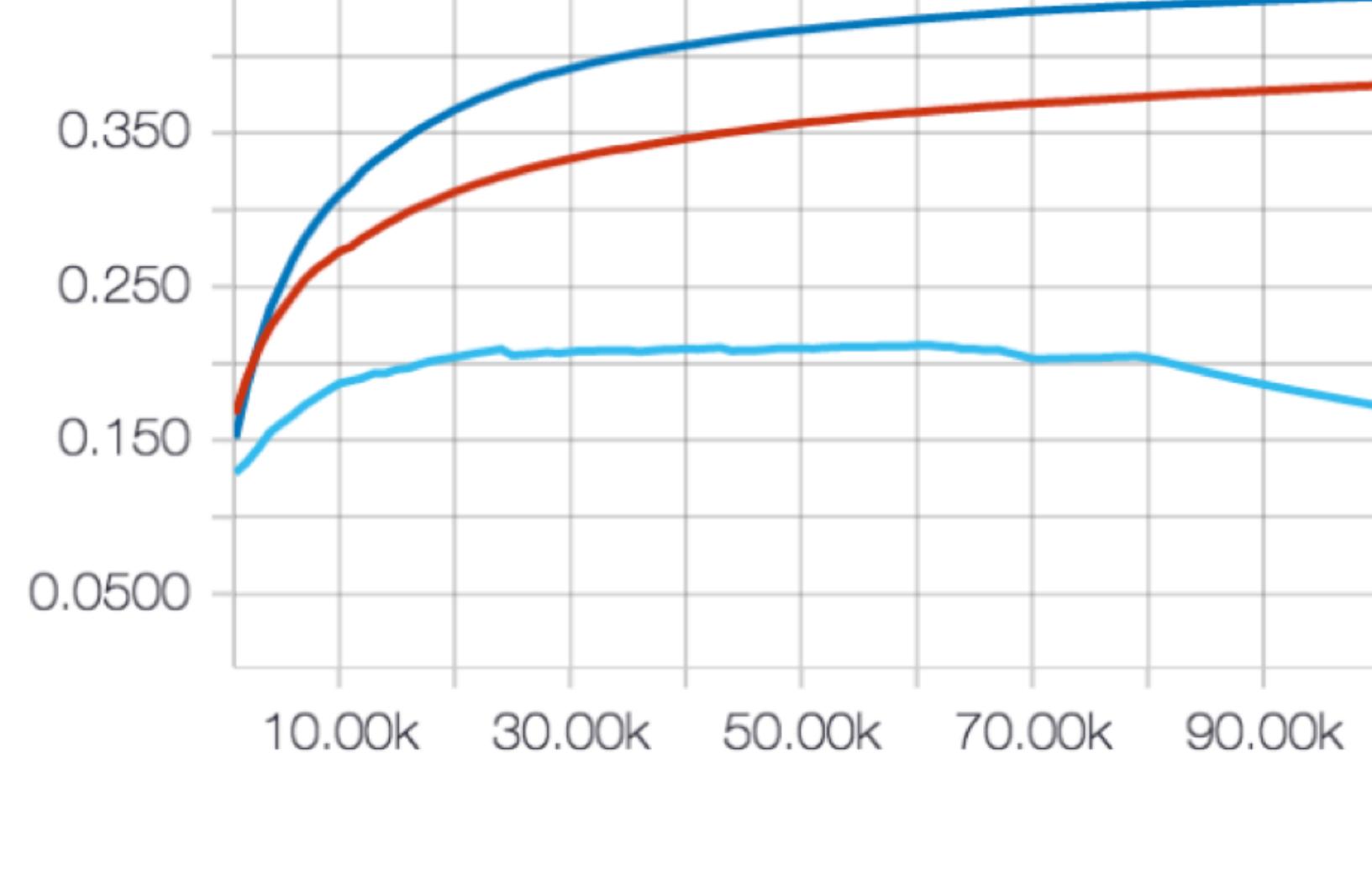
Implementación

Los modelos son implementados en TensorFlow, y entrenados con un GPU NVIDIA Tesla K80 de 12Gb.

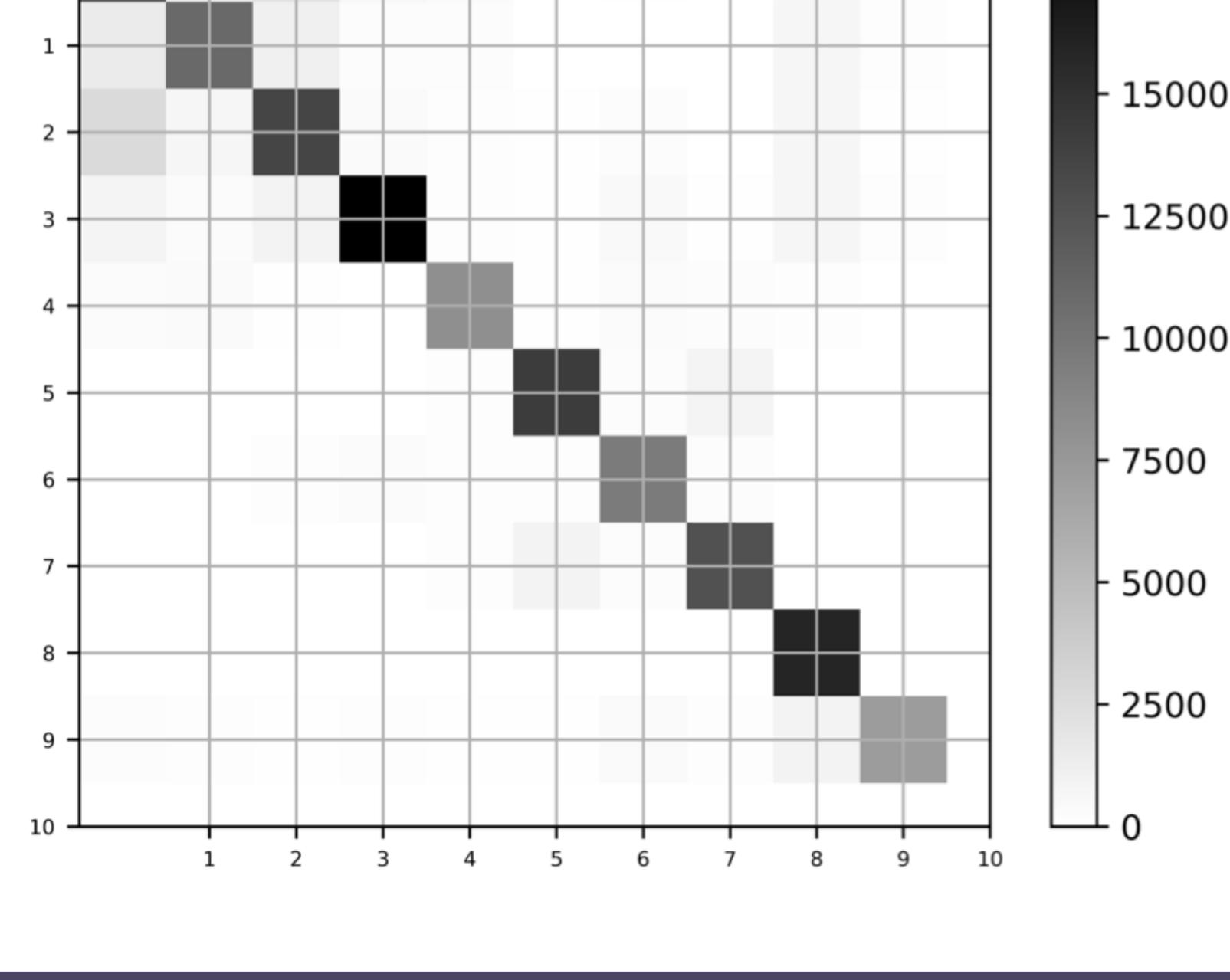


TensorFlow™ Google Cloud Platform

Resultados



# parámetros	Tiempo	Clasificación
LSTM	1441800	43,89%
D-RNN	9150	38,11%
RNN	36015	17,24%



Conclusiones y recomendaciones

- D-RNN mejora la clasificación de una RNN en un 20%, con 76% de los parámetros.
- D-RNN tiene un 6% más error que una LSTM, pero con 0,68% de los parámetros de una LSTM.
- D-RNN logra identificar correctamente la mayoría de actividades.

En trabajos futuros se recomienda:

- Explorar el impacto que las unidades recurrentes con retardos distintos tienen en los gradientes del modelo.
- Evaluar a DRNN con otros datasets.
- Comparar DRNN con otras arquitecturas que mitigan la inestabilidad de los gradientes.