Universitat Oberta de Catalunya

Predicción de tráfico en redes móviles mediante Deep Learning

José Manuel Gran Josa

José López Vicario
Xavi Vilajosana Guillén

Enero 2019

Índice

- Objetivos
- Contexto
- Virtualización
- 5G
- Cloud RAN
- Machine Learning
- Deep Learning
- Python, Pandas, Numpy, Matplotlib.
- TensorFlow
- Otros trabajos y contribución
- Algoritmos predicción de trafico
- Tratamiento de datos
- Regresión lineal
- Red neuronal
- Comparativa
- Conclusiones

Objetivos

1. Introducción conceptos tecnológicos, virtualización, 5G…
2. Conocer y asimilar arquitectura Cloud RAN
3. Conceptos de programación, Python, librerías de datos…
4. Inteligencia artificial, Machine Learning, Deep Learning.
5. Algoritmos de predicción.
6. Tratamiento de datos para algoritmos predictivos de Deep Learning.
7. Valorar predicciones según MSE.
8. Obtener predicción de datos de llamadas realizadas sobre red móvil.
9. Interpretación de datos para optimizar Cloud RAN y conseguir mejores prestaciones de cara al 5G.

Contexto

De cara al próximo estándar 5G, se necesitan conseguir prestaciones exigentes en ancho de banda, número de dispositivos conectados y sobre todo en latencia.

La implementación del Cloud-RAN (RAN virtualizado) es un paso previo para 5G y obtener las latencias requeridas.

En el momento que se mejoren las latencias se podrán implementar servicios de Voz sobre IP (VoIP) de forma masiva en la redes de telecomunicaciones.

Se pretende predecir el número de llamadas realizadas en una red móvil, con el objetivo de optimizar elementos virtualizados en arquitectura Cloud RAN.

Virtualización

La virtualización es una tecnología que nos permite crear a través de software, versiones virtuales de recursos tecnológicos.

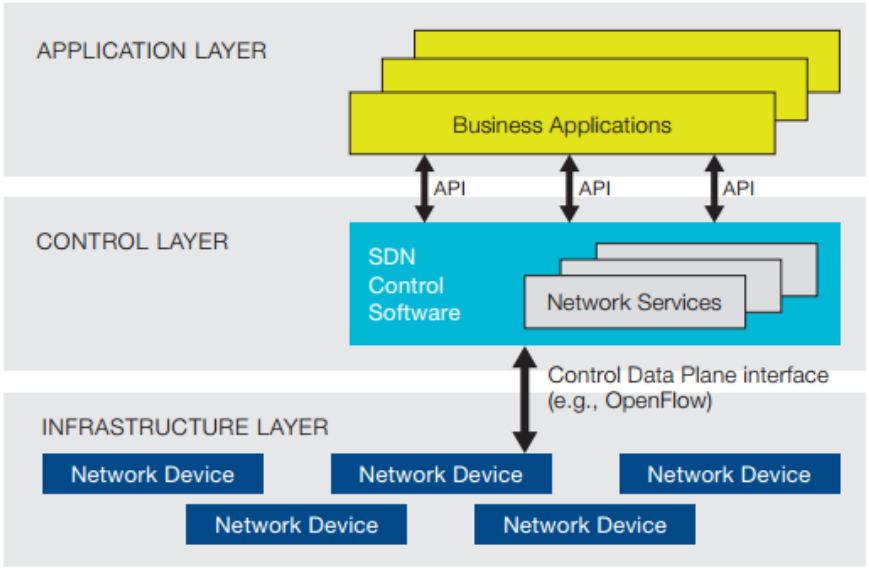
El software que se encarga de gestionar los recursos físicos se llama hipervisor.

SDN, redes definidas por software, es el concepto de aplicar la virtualización a la redes de telecomunicaciones.

Nos permite disponer de redes más flexibles.

Se crean redes de manera modular (infraestructura, control y aplicación).

NFV, virtualización de las funciones que utilizamos en un servicio de red



5G

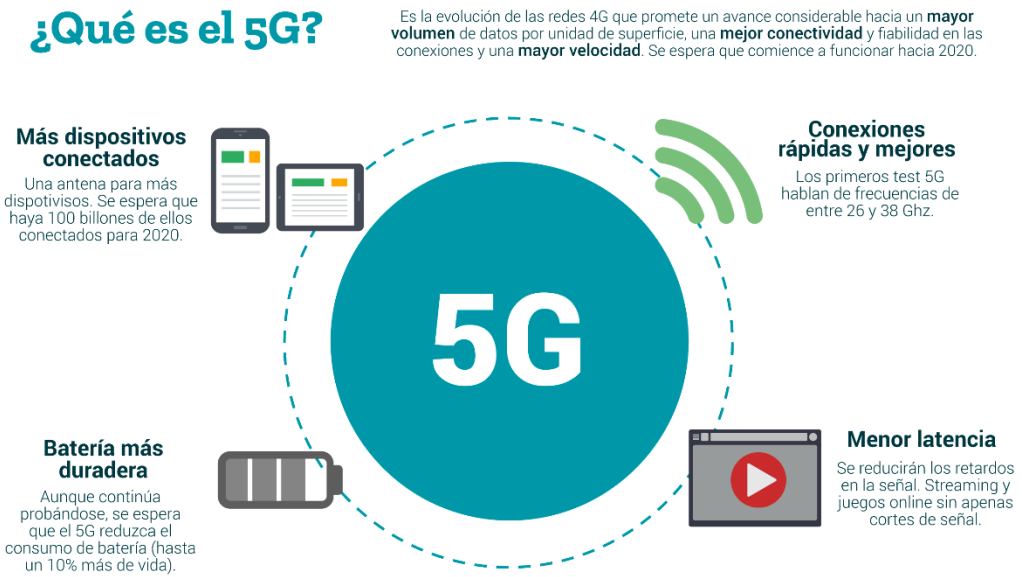
5G es la quinta generación tecnológica de comunicaciones móviles, evolución de la red 4G (LTE).

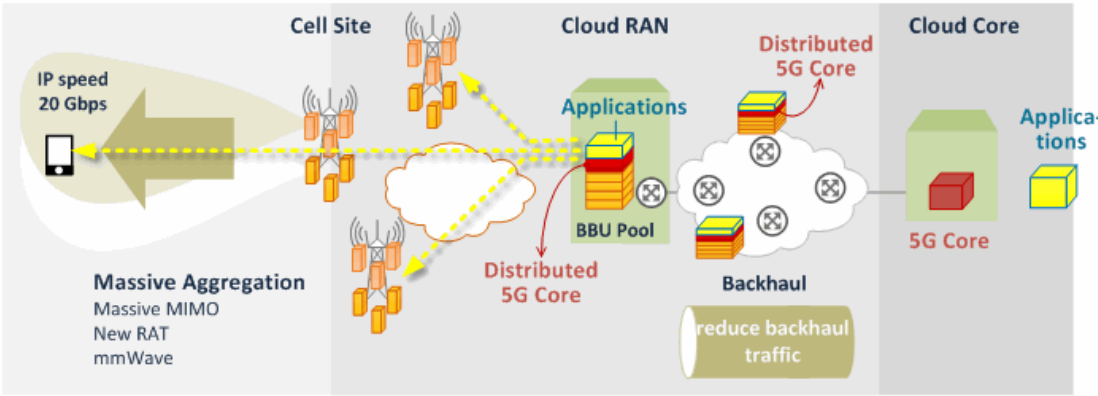
Actualmente en desarrollo pero ya se han entregado las primeras especificaciones.

Se pretende incrementar el número de dispositivos conectados, aumentar el ancho de banda, optimizar la energía consumida y disminuir la latencia extremo a extremo.

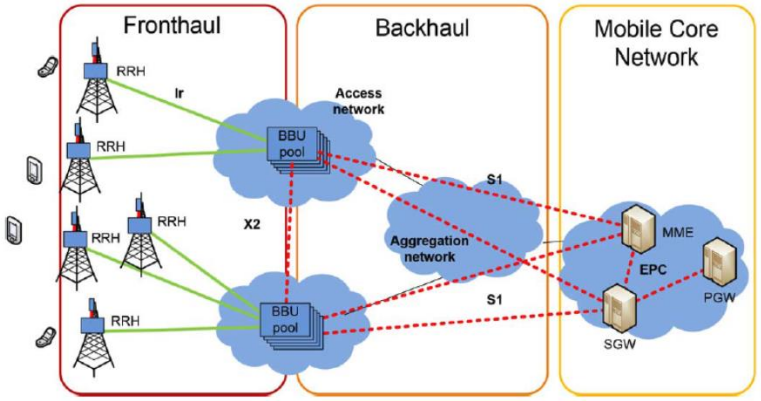
Orientado a dispositivos móviles, streaming de video, IoT.

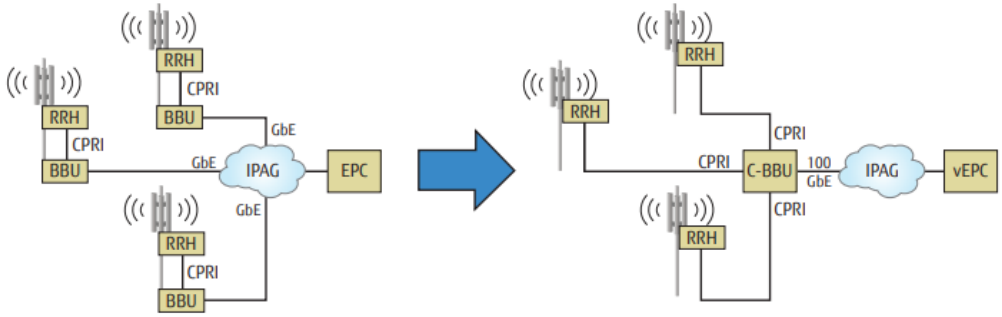
Arquitectura 5G con Cloud RAN y core distribuido)





Cloud RAN





Cloud-RAN es una arquitectura de acceso para redes móviles.

La red de acceso se divide en fronthaul, backhaul. La primera es la más cercana al usuario y la última más próxima al núcleo de la red móvil.

Cloud-RAN centraliza y virtualiza las BBUs, separándolas de las RRH. Ya no se necesita una estación bases de dicada en cada RH.

Así se colocan las BBUs más próximas al núcleo de red.

Proporciona una simplificación en la red y permite gestionar los recursos de forma centralizada.

Se consiguen mejores prestaciones de cara al 5G.

Machine Learning

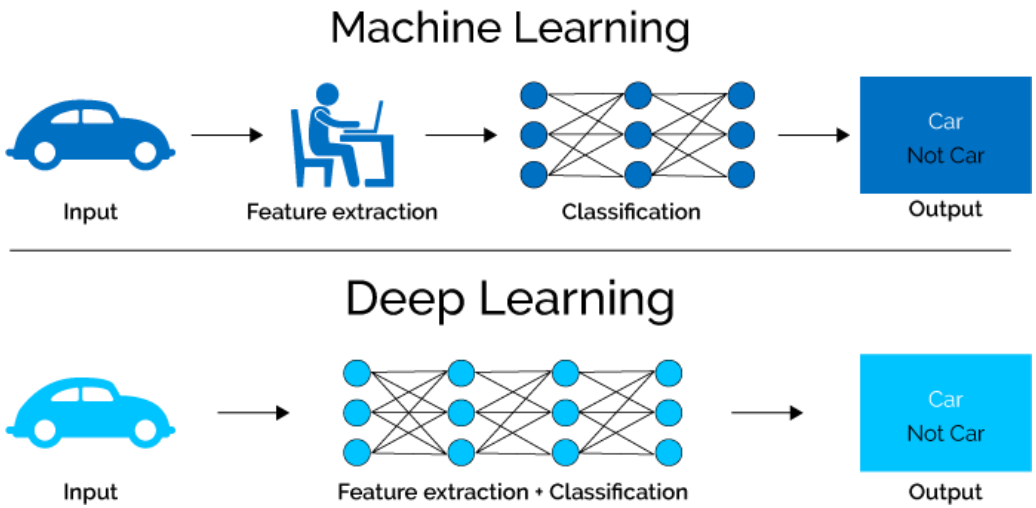
Machine Learning o aprendizaje automático es un subcampo de la inteligencia artificial.

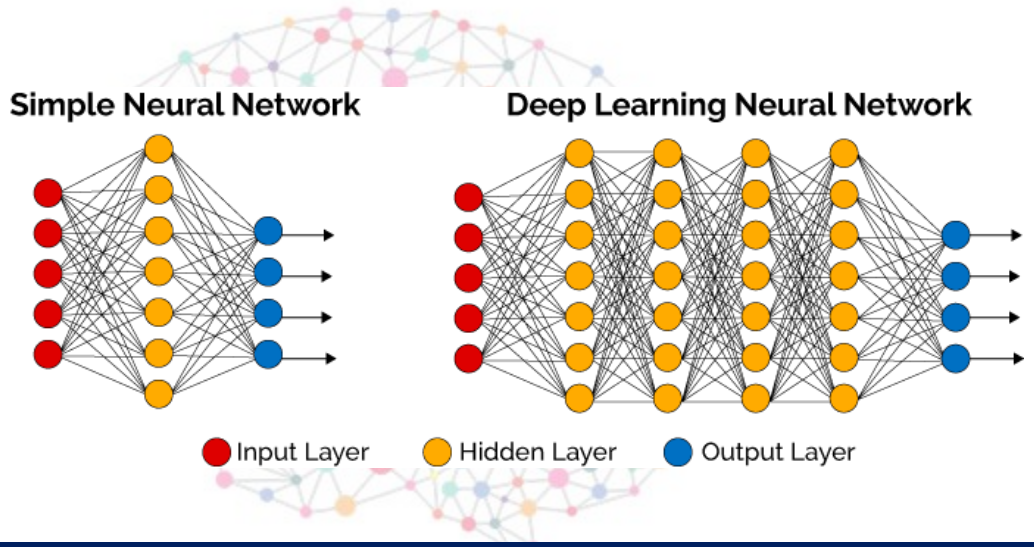
La Inteligencia Artificial (IA) es la combinación de algoritmos diseñados para crear máquinas que demuestren similares capacidades que las de las personas.

Los algoritmos de Machine Learning se clasifican en aprendizaje supervisado y no supervisado.

Destacamos aprendizaje por refuerzo, semi supervisado y profundo.

Deep Learning





Deep Learning no necesita realizar la extracción de las características de los datos, puede realizar el proceso conjuntamente con la clasificación.

Permite distintos niveles de abstracción.

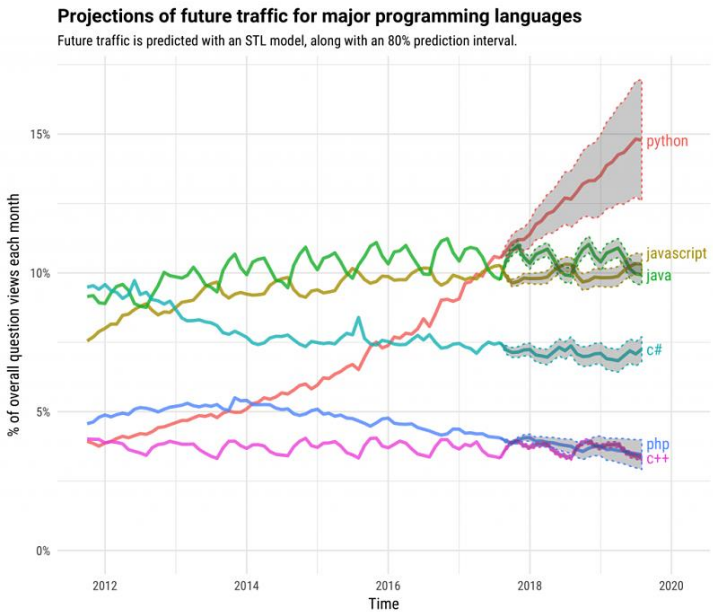
Modelos compuestos con varias capas ocultas.

Redes neuronales artificiales (ANN). Neuronas y conexiones.

Clasificación de imágenes, reconocimiento de voz, predicción de datos

Python, Pandas, Numpy...

Python es un lenguaje de programación interpretado. Creado por Guido van Rossum en los 90s.



Soporta programación orientada a objetos, programación imperativa y programación funcional. Todo en Python es un objeto.

Gran comunidad (previsión creciente de uso). Zen de Python: Bonito es mejor que feo…

Pandas es una librería open source para el operar con estructura de datos. DataFrames.

Numpy es un paquete que permite manipular largos arrays multidimensionales realizar operaciones matemáticas sobre estos de forma sencilla.

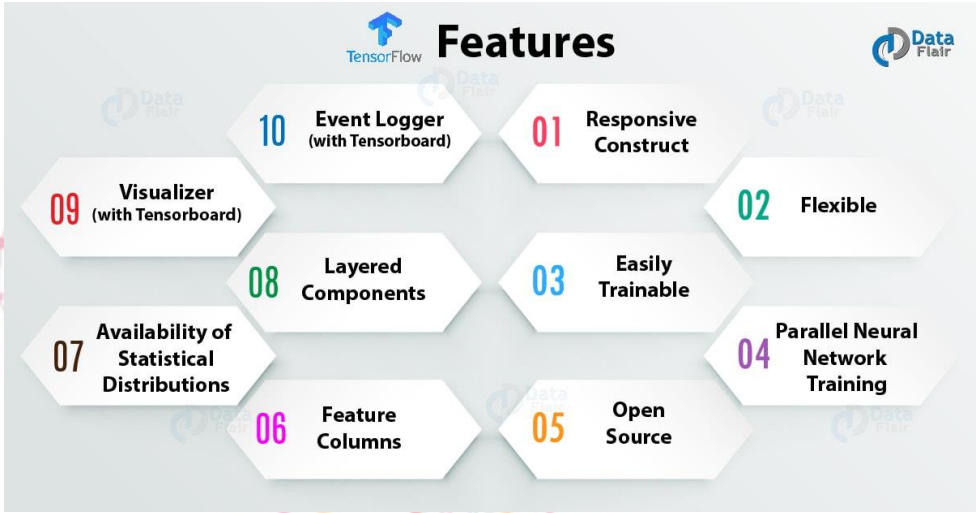
Matplotlib es una librería con la que podemos graficar en 2D arrays de datos.

TensorFlow

TensorFlow es una biblioteca de código abierto para realizar cálculos numéricos mediante diagramas de flujos de datos utilizada para Machine Learning.

Desarrollada por Google para construir y entrenar redes neuronales, descifrar patrones de forma parecida al razonamiento del ser humano.

Un tensor es un objeto geométrico que describe relaciones lineales entre vectores geométricos, escalares, etc.



Playgraund. Herramienta web para aprender conceptos básicos sobre redes neuronales artificiales.

TensorBoard. Herramienta para visualizar los grafos y observar las métricas de la ejecución del grafo.

Otros trabajos y contribución

El salto cualitativo de Deep Learning en problemas de clasificación

Solución de rastreo y detección para conocer y analizar los individuos que cohabitan en un mismo lugar. Se analizan métodos para la calificación y se construye una red neuronal para clasificar imágenes de banco de peces. Concluye con éxito la viabilidad de la red neuronal implementada para el sistema de clasificación.

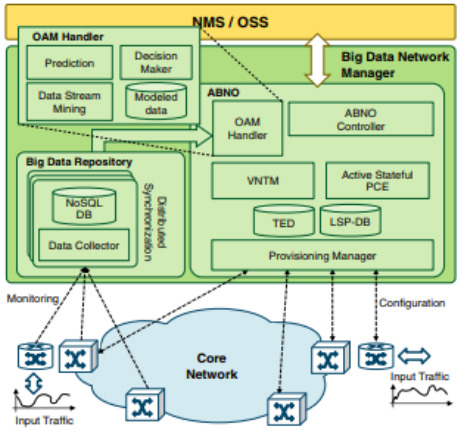
Detector predictivo de conexiones fraudulentas

Diseñar e implementar un detector predictivo de conexiones fraudulentas de manera eficiente. Se basa en herramientas de almacenamiento de datos NoSQL y herramientas de predicción de Deep Learning como TensorFlow. Finaliza con éxito la construcción de un clasificador predictivo eficiente capaz de distinguir conexiones intrusivas de las lícitas.

Applying Deep Learning Approaches for Network Traffic Prediction

Presenta una discusión de los principales métodos de predicción basados en Deep Learning para la predicción de tráfico de red. Se ponen en práctica y se realiza una comparativa con los resultados obtenidos.

Algoritmo	MSE
FFN	0.091
RNN	0.067
LSTM	0.042
GRU	0.051
IRNN	0.059

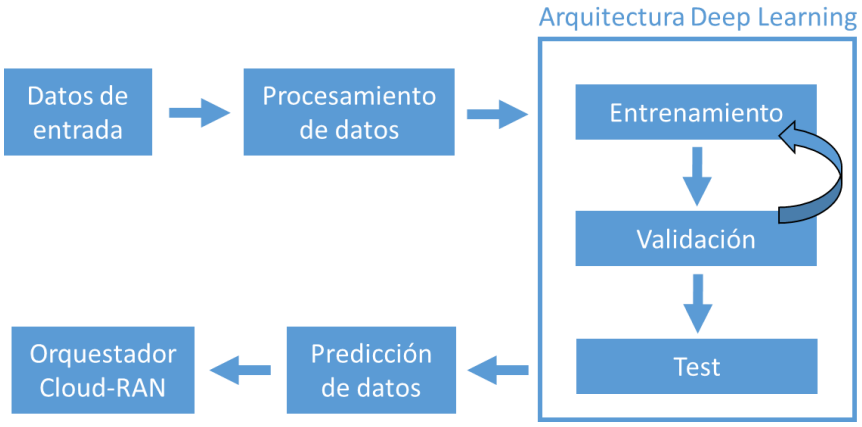


Virtual Network Topology Adaptability Based on Data Analytics for Traffic Prediction

En base a los datos predictivos, basados en algoritmos de Machine Learning, se proporciona una arquitectura para reconfigurar parámetros de red en base a dichas predicciones. Concluye con la demostración del intercambio de mensajes generados para la adaptabilidad de la red.

Contribuciones: Contribuir de forma personal con conocimiento en el área de Deep Learning. Enfoque distinto (regresión en lugar de clasificación). Proporcionar datos predictivos para adaptar elementos virtualizados (Cloud-RAN).

Algoritmos predicción tráfico



Datos de entrada que requieren un procesamiento para arquitectura Deep Learning.

Después del correspondientes entrenamiento y prueba del algoritmo se obtienen la predicciones de datos de tráfico de red.

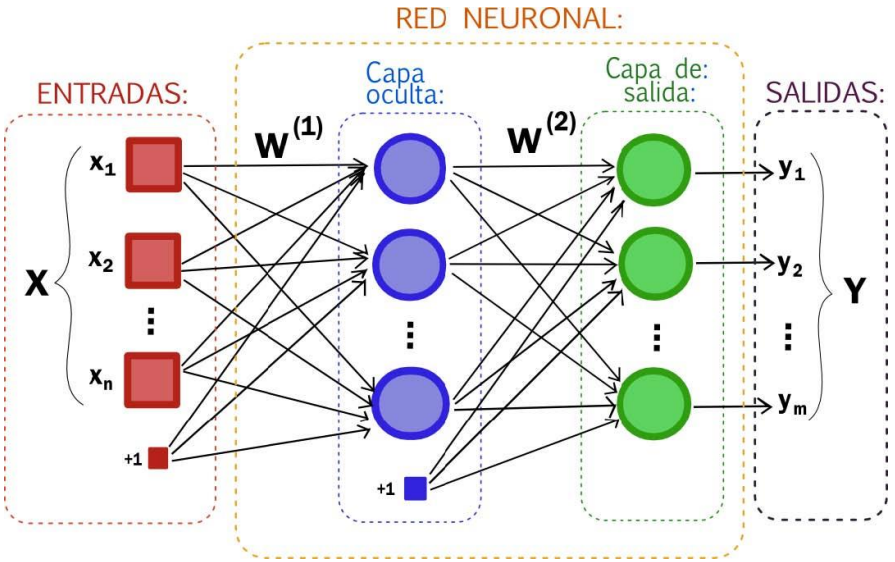
Con las soluciones obtenidas se optimizan los recursos de los elementos del Cloud RAN a través del orquestador de su arquitectura.

Redes Neuronales Artificiales (ANN)

Modelo computacional basado en neuronas artificiales con comportamiento similar al cerebro de los seres vivos.

Permite trabajar con grandes conjuntos de entrada.

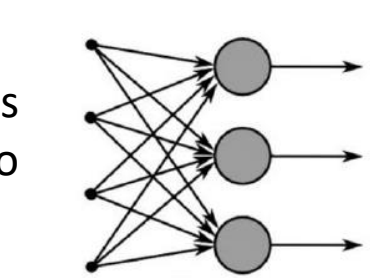
Principalmente diferenciamos dos tipos. FNN y RNN.



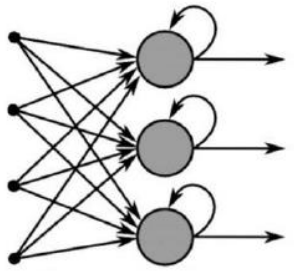
Algoritmos predicción tráfico

Redes Neuronales Feed-forward (FNN)

Primeras y más sencillas redes neuronales. Las conexiones entre ellas no forman un ciclo. Sólo propagación en un sentido.



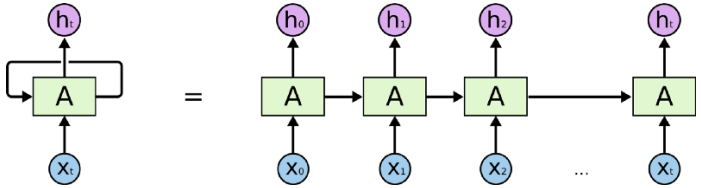
Feed-Forward Neural Network



Recurrent Neural Netwok

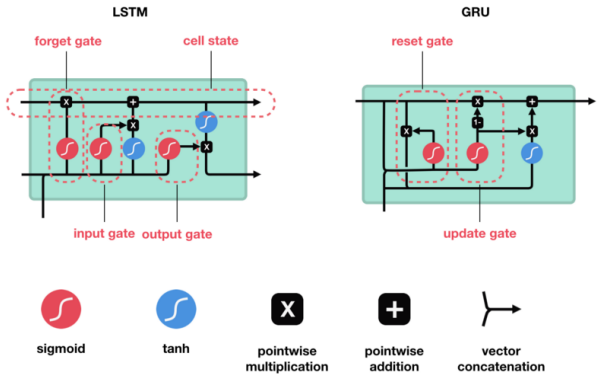
Redes Neuronales Recurrentes (RNN)

Tipo de redes neuronales que permiten conexiones hacia atrás. Facilita el aprendizaje con dependencias temporales. Propagación a través del tiempo (BPTT), propagación hacia adelante, propagación hacia atrás.



Long short-term memory (LSTM)

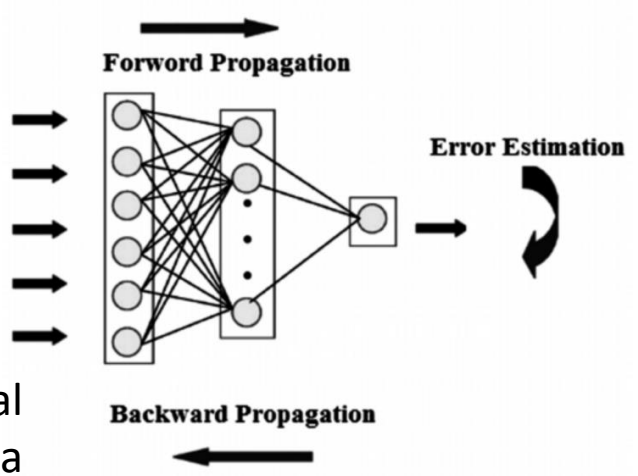
Modelo de red basado en RNN que extiende su memoria. Celda formada por 3 puertas (input, forget y output).



Resuelven problemas de vanishing y exploding gradient

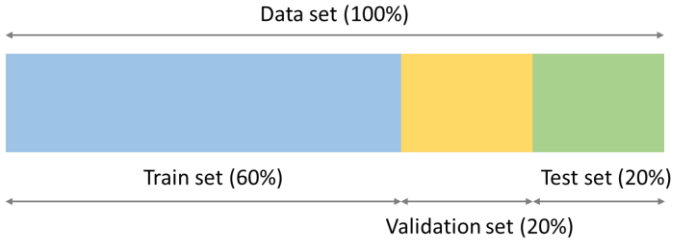
Gated recurrent unit (GRU)

Reducen el coste computacional de las LSTM. Carecen de puerta de salida. Reset y update

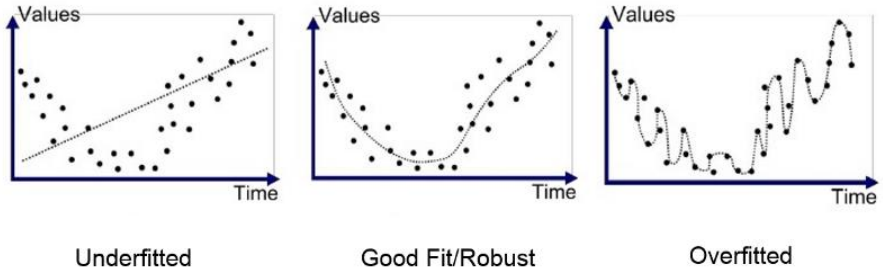


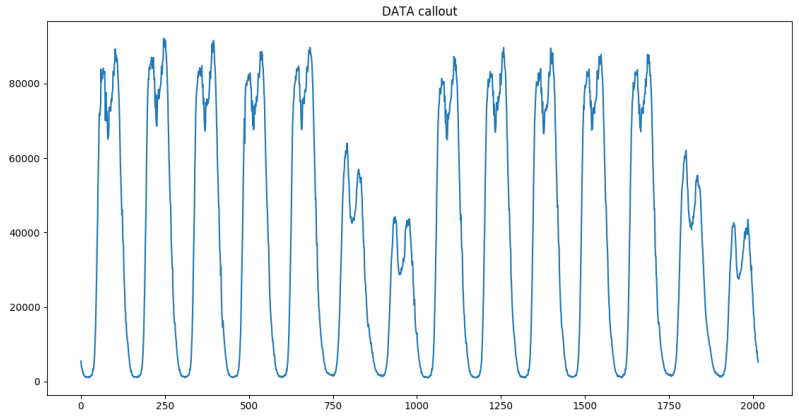
Tratamiento de datos

Datos de entrada separados en 3 conjuntos. Datos de **entrenamiento**, datos de **validación** y datos de **test**.

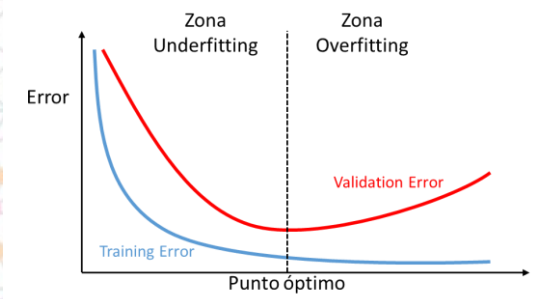


Se debe iterar todo el conjunto de datos de entrenamiento sobre la red neuronal (**epoch**).





Se produce **underfitting**, cuando no entrenamos suficientemente la red. En contra tenemos **overfitting** cuando se produce un sobreentrenamiento.



Necesidad de escalado de los datos. Standard scale vs MinMax scale.

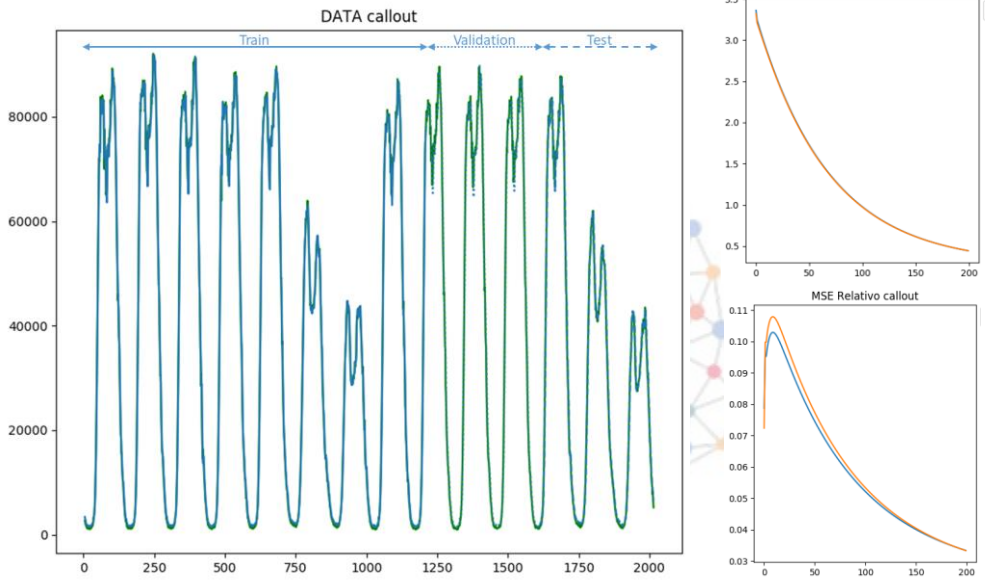
Datos proporcionados en el artículo *"A multi-source dataset of urban life in the city of Milan and the Province of Trentino"*. Generados en la red móvil de la ciudad de Milán. El conjunto de datos que utilizamos corresponde a las llamadas realizadas en 14 días.

Regresión lineal

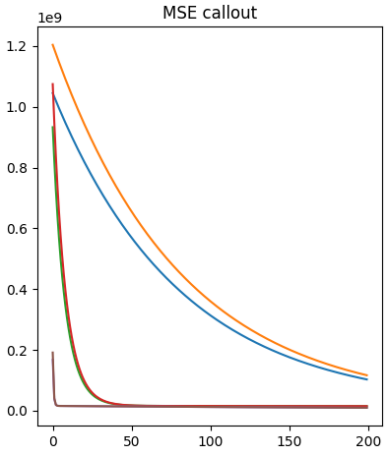
Sistema de predicción de datos basado en regresión lineal multivariable. Modelo sencillo de implementar. Permite la variación del ratio de aprendizaje y del número de características a utilizar.

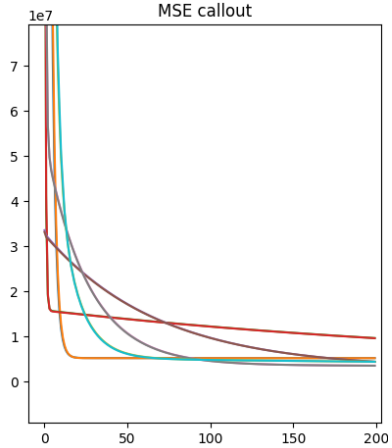
Se varia el *learning rate* entre 0,001, 0,01 y 0,1 y el número de características entre 1, 3, 5, 7 y 9.

Apreciamos como se necesitan menos *epochs* al aumentar el *learning rate*.



$$Y = X * W + b$$





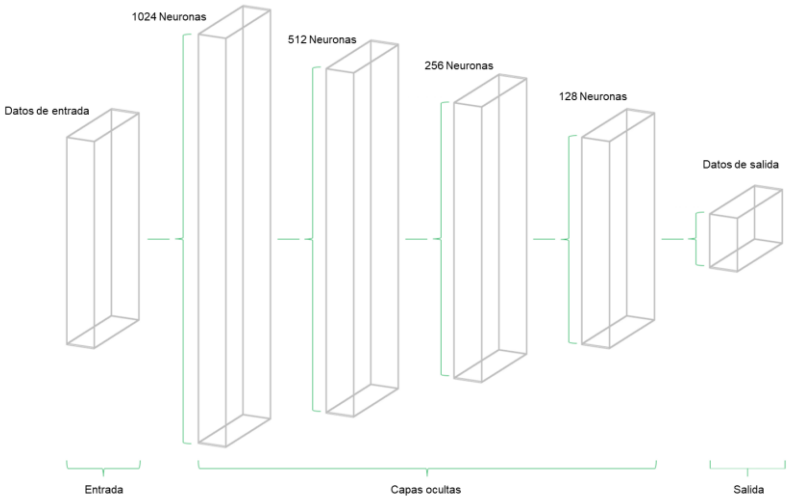
Se realiza la predicción con un ratio de aprendizaje de 0,01 y 5 características.

Se observa las predicciones obtenidos, el MSE y el MSE relativo.

Comprobamos como disminuye el MSE en cada iteración (epoch)

El modelo finaliza con éxito con una precisión del MSE relativo del 3,33 %

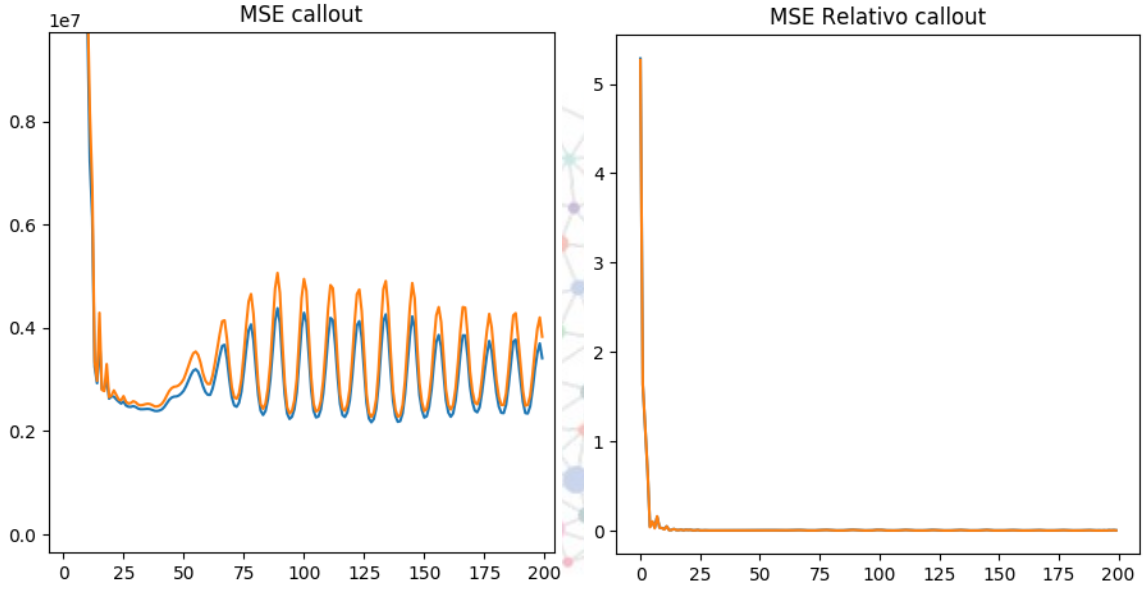
Red neuronal



Construcción de una red neuronal con 6 capas. Una capa de entrada, 4 capas ocultas y una de salida.

En las capas ocultas se utilizan 1024, 512, 256 y 128 neuronas respectivamente.

Red con función de activación ReLu, por lo que se escalan los datos entre 0 y 1.



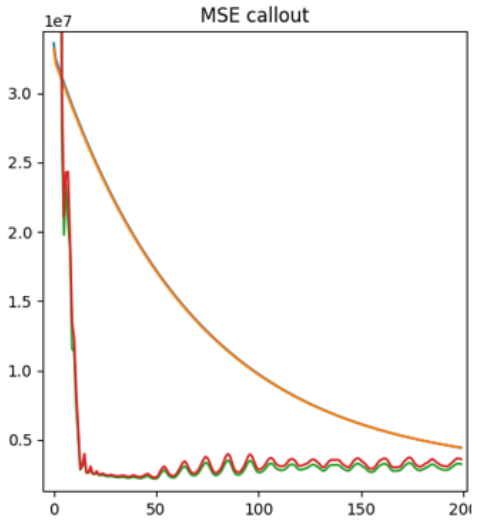
Simulamos con 5 características y un tamaño de batch de 256.

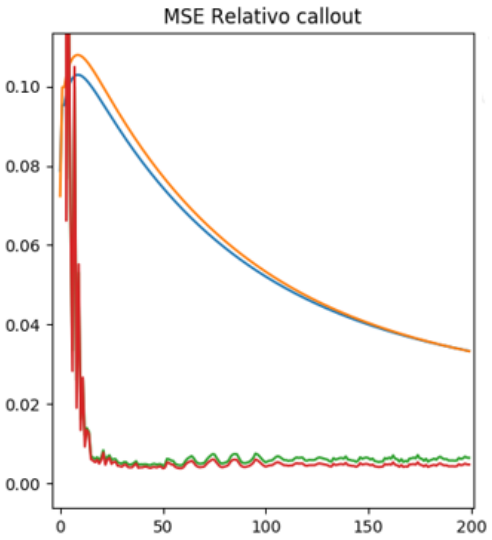
Se observan fluctuaciones en el MSE a partir del epoch 40.

Obtenemos con éxito una predicción con un MSE relativo de 0,7%

Comparativa modelos

- Se observa como obtenemos mejores resultados en un modelo de predicción basado en redes neuronales.
- Se obtienen resultados más precisos con menores iteraciones. Es más eficiente respecto al número de epochs necesitados.
- En contra, el primer modelo es más sencillo y necesita utilizar menos recursos por iteración.
- Resultados de MSE relativo de 3,33% en regresión lineal frete al 0,7% obtenido con redes neuronales.
- Desarrollo concluido con éxito en ambos modelos. Se obtienen ambos valores relativos satisfactorios.





Conclusiones

- Dada la importancia de las nuevas tecnologías que están surgiendo, se ha comprendido en detalle temas novedosos como la virtualización y el próximo estándar 5G.

- Cloud RAN es la arquitectura de acceso de radio que virtualiza las BBUs. Es un paso previo al 5G en el que se consiguen prestaciones que este requiere (mejoras en latencia). Entendemos que optimizar la arquitectura Cloud RAN a partir de los datos obtenidos en la predicción de datos, es una buena solución ya que nos permitirá ajustar los recursos de la red en función de sus necesidades futuras.

- Se consigue introducir aspectos de Machine Learning y Deep Learning, así como detallar los principales aspectos de los modelos de redes neuronales más utilizados.

- Se implementan dos modelos de predicción de datos, que cuentan como entrada las llamadas realizadas en la ciudad de Milán. Se observa como se consiguen mejores soluciones en el modelo basado en redes neuronales.

- Con las predicciones satisfactorias obtenidas, se podrían aplicar llevando a cabo la optimización real de los elementos virtualizados del Cloud RAN.

Universitat Oberta de Catalunya

Predicción de tráfico en redes móviles mediante Deep Learning

¡Muchas gracias!

José Manuel Gran Josa

José López Vicario
Xavi Vilajosana Guillén

Enero 2019