

# En Profundidad

## Deep learning

**Autor:** David García Dolla, Área de Planificación y Control, SDGPLATIN.

**Palabras clave:** redes neuronales artificiales, big data, reconocimiento de imágenes, análisis de voz, GPU.

**Metas tecnológicas relacionadas:** MT 2.5.1; MT 6.1.2; MT 6.1.3.

Durante los últimos años se ha producido un enorme crecimiento de la capacidad para adquirir ingentes cantidades de datos procedentes de todo tipo de sensores distribuidos y de fuentes abiertas. Adicionalmente se han dado importantes pasos en la capacidad para gestionar todos esos datos y ponerlos a disposición de los usuarios en todo momento y lugar, proporcionándoles, entre otras, capacidades de búsqueda. No obstante, continúa existiendo un cuello de botella en el análisis e interpretación automática de esos datos, en especial, cuando se trata de imágenes o vídeos, lo cual supone un problema importante, dado la enorme dimensión de los datos disponibles.

En general, la capacidad de los ordenadores para analizar e interpretar la información sensorial dista mucho de las de los humanos, en particular, si es necesario realizar tareas de reconocimiento, generalización o abstracción. La complejidad del análisis varía mucho según factores tales como el dominio de la aplicación, la finalidad del análisis, las

condiciones y elementos presentes en la escena, el punto de vista, etc., lo que hace que las soluciones existentes sean válidas solamente para abordar problemas muy concretos.

La solución de este tipo de problemas se aborda frecuentemente desde una de las disciplinas vinculadas a la inteligencia artificial. Se trata de las técnicas de aprendizaje automático, que persiguen el desarrollo de la capacidad de los ordenadores para aprender de los datos disponibles, siendo capaces de proponer soluciones ante nuevos datos, normalmente en problemas relacionados con su clasificación en categorías o con la predicción de valores numéricos asociados a alguna variable (problemas de regresión numérica).

Si bien desde hace décadas se han venido produciendo notables avances con el desarrollo y utilización de nuevos algoritmos, muy pocos han captado el interés que desde hace unos pocos años están teniendo las denominadas técnicas de aprendizaje profundo (*deep learning*, en inglés).

Las tecnologías de aprendizaje profundo son una mejora de las redes neuronales artificiales, que alcanzaron una gran popularidad en los años ochenta y noventa. El fundamento de las redes neuronales es el de recrear unas estructuras de procesamiento de datos similares a las del cerebro animal de forma que, ante un estímulo de entrada, un sistema de

elementos muy sencillos (neuronas) interconectados entre sí, colaboran para producir un estímulo de salida. En principio, ninguna de las neuronas tiene asignada una tarea concreta. Más bien, las conexiones que las neuronas van creando y reforzando durante el proceso de aprendizaje permiten representar modelos matemáticos a través de la arquitectura de la red y los múltiples parámetros asociados a las conexiones. El resultado que se busca en la red es el de ser capaz de ir aprendiendo en base a la experiencia, construyendo representaciones de problemas altamente no lineales con capacidad para responder de forma correcta ante estímulos de entrada nunca antes recibidos.

A lo largo de los años se han desarrollado múltiples topologías de redes neuronales. La arquitectura de las redes neuronales consiste en la organización y disposición de las neuronas formando capas más o menos alejadas de la entrada y salida de la red. Los parámetros fundamentales de la red son fundamentalmente el número de capas, el número de neuronas por capa, el grado de conectividad y el tipo de conexiones entre neuronas. No obstante, factores tales como la necesidad de disponer de una importante capacidad de procesamiento para entrenar las redes, la necesidad de ir ajustando la red para asegurar la convergencia en el aprendizaje y evitar problemas de excesivo ajuste de la red a los datos de entrena-

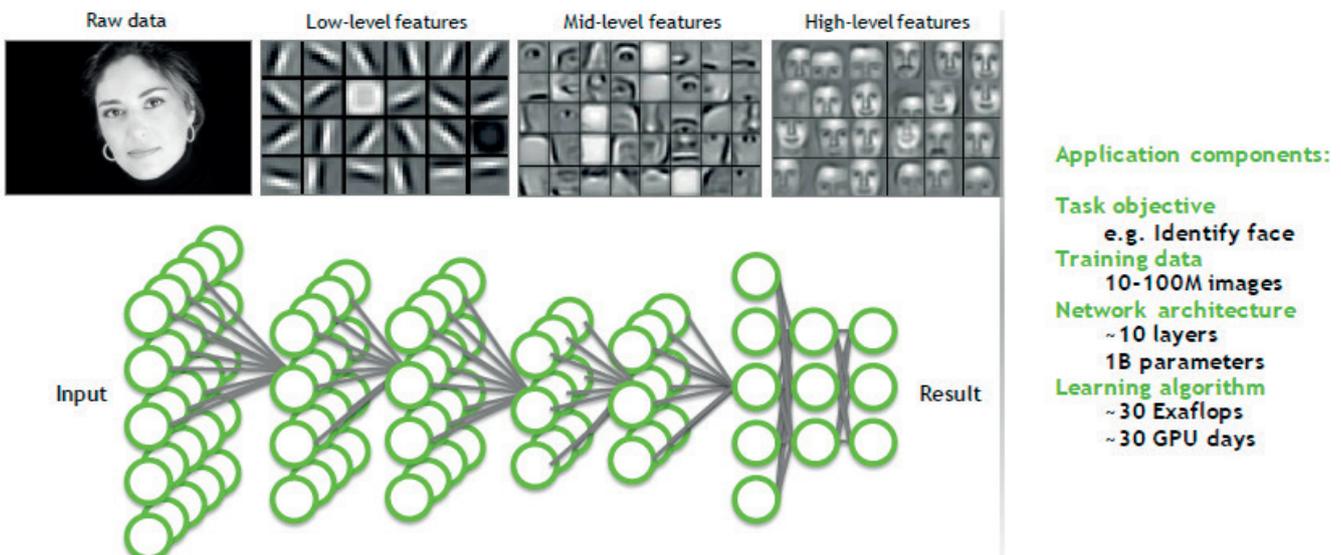


Fig. 1. Ejemplo de una CNN. (Fuente: NVIDIA).



Fig. 2. Empleo de DNN en reconocimiento de objetivos. (Fuente: NVIDIA - <https://blogs.nvidia.com/blog/2016/06/29/deep-learning-6/>).

miento (*overfitting*) ha dado lugar a que a menudo las soluciones que se han desarrollado tengan solamente unas pocas capas internas que no lograban alcanzar las expectativas creadas.

**¿Qué es lo que está cambiando para que se vuelva a hablar de las redes neuronales artificiales, o más en particular, del aprendizaje profundo?**

La existencia de múltiples capas internas es donde se encuentra uno de los elementos diferenciadores de esta tecnología, al permitir representar diferentes niveles de abstracción de los datos de entrada, mejorando la capacidad de predicción de la red. Y es precisamente en el desarrollo de nuevas técnicas que permiten trabajar con redes dotadas de más capas internas (varias decenas) con mayores conexiones donde las cosas más están cambiando. De ahí el nombre *Deep Neural Network* (DNN) que se vienen utilizando para referirse a este tipo de redes.

Algunos autores apuntan a dos factores como habilitadores de estos avances. En primer lugar, la posibilidad de

disponer de enormes cantidades de datos para entrenamiento, gracias al desarrollo del *big data*. Y en segundo, los avances en cuanto a *hardware* de procesamiento de datos y, muy especialmente, el desarrollo de las GPU (*Graphics Processor Unit*), es decir, los coprocesadores que desde hace años se incluyen en los ordenadores para realizar el procesamiento de gráficos o las operaciones de coma flotante, aliviando así la carga de trabajo de las CPUs de los ordenadores, sobre todo si trabajan con videojuegos o aplicaciones 3D interactivas. Dado que las GPUs están optimizadas para trabajar con grandes cantidades de datos y realizar las mismas operaciones, una y otra vez, es posible transferir a estas unidades una parte importante de las operaciones que se llevan a cabo en las redes neuronales durante el aprendizaje, mejorando notablemente las prestaciones y a muy bajo coste [1].

De esta manera, en los últimos años, los fabricantes de GPUs han venido adecuando sus desarrollos para facilitar su empleo en aplicaciones de aprendizaje, permitiendo que hoy en día se pueda

hablar de problemas cuyo análisis es masivamente paralelizable. Adicionalmente se han venido desarrollando conjuntos de librerías *software* (Caffee, CNTK, Theano, Tensorflow, Torch, etc.) que simplifican enormemente el desarrollo y despliegue de aplicaciones de aprendizaje profundo que permiten explotar al máximo las capacidades de una o múltiples GPUs.

Otra notable ventaja que presentan las DNN frente a las técnicas clásicas de aprendizaje automático es la simplificación del proceso de aprendizaje y la mejora en robustez frente a variaciones naturales en los datos de entrada. Así, por ejemplo, en aplicaciones que trabajan con datos de entrada complejos como imágenes o vídeos, normalmente es necesario analizar de forma manual los datos de entrada y determinar las características (elementos específicos en la escena) en torno a los que se debe centrar el aprendizaje, siendo además necesario implementar una etapa de extracción de esas características de forma que sean únicamente esos datos los que alimenten

el algoritmo de aprendizaje. La elección y extracción de estas características puede suponer un problema complejo y costoso que condicione las posibilidades del aprendizaje, además de estar condicionado por cambios en las condiciones de los datos (p.e. variaciones en iluminación). Pues bien, en el caso de las DNN, se introducen directamente en el algoritmo los datos de entrada originales (p.e. los píxeles de las imágenes) o con solo cierto preprocesamiento, siendo responsabilidad de la red, el analizar esos datos y aprender automáticamente las características que sirven como medio de discriminación

Otra línea en la que se está trabajando es en la de mejorar la capacidad de reutilización de las redes ya entrenadas para ser empleadas en problemas similares con pequeñas modificaciones (problema denominado *transfer learning*). La finalidad es reducir los tiempos de aprendizaje o solventar las situaciones en las que no se disponga de grandes cantidades de datos para entrenamiento en los nuevos problemas a analizar.

De todos los tipos de DNN que se han venido desarrollando en los últimos años, cabe destacar algunos de ellos por su singularidad. En particular, en el campo del procesamiento de imágenes desde aproximadamente el año 2012 se ha venido utilizando las denominadas redes neuronales convolucionales (CNN - *Convolutional Neural Networks*), que consisten en redes con varias capas internas interconectadas entre sí, que permiten extraer características de las imágenes con diferentes niveles de abstracción. Para ello, combina capas que aplican filtros sobre las imágenes con otras que submuestran para reducir su tamaño, hasta llegar a las últimas que se encargan de clasificar. Se están utilizando fundamentalmente en aplicaciones tales como reconocimiento de objetos, clasificación de imágenes, etc.

Por su parte, en aplicaciones de reconocimiento del habla o de escritura, generación de texto o voz, traducción automática, etc., se utiliza un tipo de red neuronal diferente (RNN - *Recurrent Neural Networks*), más orientada a analizar información secuencial (a diferencia de las CNN que son más aptas para analizar información espacial). En este tipo de aplicaciones, existe dependencia en el orden de las

palabras, lo que hace que la presencia de una, condicione las anteriores y las siguientes.

Todos estos avances están motivando que las grandes empresas dedicadas a la tecnología de gestión y análisis de datos (Google, Facebook, Baidu, Microsoft, etc.), así como numerosas *startups* estén desplegando aplicaciones basadas en DNN dedicadas a reconocimiento del habla, detección y reconocimiento de imágenes, análisis de lenguaje no estructurado, etc., en variados ámbitos como el de la medicina, redes sociales, conducción autónoma, observación de la tierra, etc. Asimismo, prestigiosas revistas, como *MIT Technology Review*, la han considerado en 2016 como una de las diez adelantos tecnológicos que más impacto van a tener en los próximos años.

### ¿Y cómo afecta esto al ámbito de defensa y seguridad?

El ámbito de defensa y seguridad es uno de los que puede beneficiarse de estos avances, dada la necesidad de analizar las crecientes cantidades de información sensorial que las agencias de inteligencia y ejércitos de cada país adquieren a diario, buena parte del cual, lo realizan operadores que analizan horas de vídeo en busca de objetivos.

Como ejemplo del interés que algunos organismos vinculados a defensa están teniendo en esta tecnología, cabe apuntar al estudio que la EDA ha licitado recientemente para profundizar en las aplicaciones que la tecnología de aprendizaje profundo puede tener en el ámbito militar.

También son varios los autores que apuntan posibles aplicaciones de esta tecnología a los ámbitos de defensa y seguridad. Así, en [2] se destaca como principal interés la posibilidad de llevar las DNN a la "palma de la mano del combatiente". Si se logra miniaturizar el *hardware* necesario para ejecutarlas, sería posible embarcar en pequeños drones el equipo necesario para adquirir las imágenes y realizar el reconocimiento de objetos a bordo, sin necesidad de enviarlas a la estación en tierra por analistas humanos o centros de procesamiento de datos. También permitiría a equipos de operaciones especiales o cuerpos de seguridad realizar algo similar mediante drones, cámaras portátiles u otros dispositivos, de forma que se pudiese

realizar en el sitio la identificación automática de personas u objetos, distribuyendo esa información inmediatamente en la red.

Es por ello, por lo que resulta especialmente interesante los avances que están logrando en el Instituto Tecnológico de Massachusetts [3] en el proyecto *Eyeriss* que tiene por objetivo el desarrollo de un *hardware* para acelerar la ejecución de CNN en el estado del arte (múltiples capas, millones de pesos, diferentes arquitecturas, etc.) que además sea energéticamente eficiente para su uso en dispositivos portátiles. Se trata de un proyecto que cuenta entre otros con financiación de la agencia americana de investigación DARPA. Además, el hecho que un *hardware* como éste pueda llegar a ser integrado en el futuro en los teléfonos inteligentes [4] está haciendo que otros fabricantes ya estén trabajando en esta misma línea pensando en múltiples aplicaciones para uso civil, con la consiguiente movilización de inversiones y aceleración de los desarrollos, lo que hace pensar que quizás, a no muy largo plazo, se trate de una realidad.

En resumen, se trata de una tecnología que puede aportar importantes mejoras en la capacidad de los centros de procesamiento de datos para llevar a cabo análisis automáticos de grandes volúmenes de datos procedentes de sensores. Adicionalmente, la combinación de DNN con *hardware* miniaturizado parece ser la puerta a nuevas aplicaciones que, en el caso de desarrollarse, pueden ser ciertamente disruptivas para el ámbito de defensa y seguridad.

### Referencias

- [1] *Deep Learning on GPUs*. NVIDIA. Marzo 2016.
- [2] *New microchip could increase military intelligence powers exponentially*. Patrick Turner. Febrero 2016. <http://www.defenseone.com>
- [3] *Energy-friendly chip can perform powerful artificial-intelligence tasks*. <http://news.mit.edu/2016/neural-chip-artificial-intelligence-mobile-devices-0203>
- [4] *A new MIT computer chip could allow your Smartphone to do complex AI tasks*. <http://www.theverge.com/2016/2/5/10922646/mit-eyeriss-chip-deep-learning-neural-networks>