ÍNDICE

Prefacio

No cabe la menor duda de que los conceptos abordados en el libro constituyen una sólida base de los avances, ya en marcha y de futuro, en Inteligencia Artificial donde el Aprendizaje Profundo es uno de sus máximos exponentes actuales.

Son muchos los ámbitos docentes, profesionales e industriales donde el término Aprendizaje Profundo suscita un inusitado interés, refiriéndose a él, la mayoría de las veces, como *Deep Learning* en su acepción inglesa.

Es evidente que los avances conceptuales y, sin duda, los tecnológicos han contribuido a dicho interés y desarrollo. Naturalmente, soportado por el buen desempeño que la puesta en práctica de los conceptos proporciona una vez instalados en múltiples y diversos dispositivos, que incorporan sistemas automáticos con diferentes propósitos. Robots y sistemas inteligentes en general, aplicaciones en telefonía móvil, reconocedores automáticos, entre otros, son buenos ejemplos de ello. Por otra parte, es necesario reconocer al respecto la contribución realizada por el desarrollo de tecnologías y modelos dentro del paradigma conocido como Internet de las Cosas (IoT, *Internet of Things*).

La base del Aprendizaje en general y del Profundo, en particular, es el procesamiento de los datos provenientes de diferentes sensores o de distintas fuentes de información. Ese procesamiento, en lo que respecta al Aprendizaje Profundo, se lleva a cabo mediante la aplicación de los conceptos abordados en el libro. Son innumerables los avances a nivel internacional al respecto, que han propiciado la aparición de ingentes publicaciones, y desarrollos en múltiples ámbitos y bajo todos los formatos posibles, digitales y convencionales. A este respecto, existe un problema evidente, que es el hecho de la dispersión de los conceptos bajo las diferentes apariencias mencionadas. Este libro trata de unir y aglutinar el mayor número posible de ellos bajo un mismo contenedor. Se facilita así, la posibilidad de iniciación o profundización en la materia, por parte de quienes tienen interés en ella,

tanto a nivel docente e investigador como profesional en el sector industrial. Es justamente aquí donde radica el valor añadido del libro, que proporciona, con suficiente claridad, los conceptos esenciales en sendos ámbitos con tal finalidad. Sin duda, esta es la razón fundamental por la que se publica la obra, siendo conscientes de la necesidad de un texto con las características reunidas por este. En este sentido, además, el libro se ha planteado de forma que resulte autosuficiente, de suerte que el lector, incluso sin ser experto en los conceptos expuestos, puede abordar sin dificultad los contenidos en el mismo. Sin duda, este hecho otorga un valor añadido de interés a la obra.

En definitiva, desarrolladores, ingenieros, investigadores o estudiantes universitarios encontrarán en el libro una referencia de base de suma utilidad para abordar los aspectos conceptuales y de implementación en el desarrollo de aplicaciones basadas en Aprendizaje Profundo, particularmente para quienes se inicien en la materia por su carácter didáctico y autocontenido.

Sobre los autores

Gonzalo Pajares Martinsanz es profesor en la Facultad de Informática de la Universidad Complutense de Madrid, en el Departamento de Ingeniería del Software e Inteligencia Artificial y miembro del Instituto de Tecnología del Conocimiento de la misma Universidad. Ha desarrollado una extensa actividad profesional durante más de tres lustros en la industria aplicando tecnologías software y de Inteligencia Artificial. Además, está ampliamente involucrado en tareas de investigación en el ámbito de la Inteligencia Artificial mediante distintos proyectos nacionales e internacionales de investigación, durante más de dos décadas, con transferencia tecnológica a la industria. Es autor y editor de varios libros sobre visión por computador, inteligencia artificial, tecnologías sensoriales, así como autor de numerosas publicaciones en dichas áreas, incluyendo técnicas de reconocimiento de patrones y estructuras en varios ámbitos industriales, con participación de empresas situadas en la vanguardia tecnológica. Ha publicado numerosos artículos en revistas especializadas de prestigio internacional, a la vez que ha colaborado y colabora como editor invitado y asociado en varias revistas con alto índice de impacto donde las técnicas de Aprendizaje Profundo constituyen un aporte importante.

Pedro Javier Herrera Caro es profesor en la Escuela Técnica Superior de Ingeniería Informática de la Universidad Nacional de Educación a Distancia, en el Departamento de Ingeniería de Software y Sistemas Informáticos. Ha desarrollado su actividad investigadora y docente en la Universidad Complutense de Madrid, el Centro de Automática y Robótica (CSIC-UPM) y la Universidad Francisco de Vitoria. Ha realizado estancias de investigación en el Centro de Investigación Forestal del

Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, y en el AutoNOMOS LAB de la Universidad Libre de Berlín. Ha colaborado con diversos organismos y centros de investigación, universidades y empresas a través de proyectos internacionales, nacionales y autonómicos, y contratos con empresas. Fruto de ello son numerosas publicaciones en revistas, libros y conferencias internacionales. Su investigación se orienta hacia la Visión por Computador, el Reconocimiento de Patrones, la Inteligencia Artificial y la Robótica. Aplicando técnicas relacionadas con el procesamiento de imágenes, la clasificación de patrones, toma de decisiones, visión estereoscópica, teledetección y minería de datos.

Eva Besada Portas es profesora de la Sección Departamental de Arquitectura de Computadores y Automática de la Facultad de Ciencias Físicas de la Universidad Complutense de Madrid. Ha desarrollado una extensa actividad docente e investigadora durante los últimos veinte años, esta última centrada en el desarrollo y aplicación de técnicas de inteligencia artificial, optimización y fusión multisensorial a diferentes problemas reales. Es autora de numerosas publicaciones en estos ámbitos en revistas especializadas y congresos de prestigio internacional, y ha participado durante todos estos años en proyectos de investigación y contratos de transferencia tecnológica a la empresa relacionados con las mismas. Finalmente, indicar que ha realizado estancias de investigación tantos predoctorales como post-doctorales en centros extranjeros, como la Universidad de Londres, la Universidad Carnegie Mellon, la Universidad de Nuevo México o Google.

Agradecimientos

Es digno reconocer la contribución activa o pasiva a todas las personas que han tenido que sufrir la sustracción de tiempo por culpa de este libro, familiares cercanos son los más afectados y se sitúan en primera línea. Un recuerdo especial a quienes siempre disfrutaron con el esfuerzo de sus allegados.

Gracias también a las industrias, por su confianza depositada en la aplicación con éxito de algunos conceptos en sus productos de mercado. Mención especial merece Lector Vision, como empresa tecnológica puntera en el procesamiento de imágenes. Gracias, también, a los respectivos Departamentos y al Instituto de Tecnología del Conocimiento de la Universidad Complutense de Madrid por la oportunidad brindada para la aplicación de los conceptos vanguardistas del libro en diversos ámbitos.

Finalmente, un especial agradecimiento a la editorial RC Libros por su soporte para la publicación y difusión de la obra. Gracias a José Luis, por su dedicación e incansable apoyo continuo a la difusión de los conceptos técnicos, en este caso de vanguardia tecnológica.

1 INTRODUCCIÓN

1.1 PRELIMINARES

El Aprendizaje Profundo constituye una rama muy poderosa del Aprendizaje Automático (AA), enmarcándose dentro de la Inteligencia Artificial (IA) y abriendo un campo de conocimiento muy prometedor en continuo auge, en parte gracias a los avances tecnológicos, que permiten el procesamiento de ingentes cantidades de datos con estructuras complejas, cuyo exponente máximo son las redes neuronales, y más específicamente las catalogadas como *profundas*. Es evidente que hoy en día la IA constituye un campo prometedor con múltiples y diversas aplicaciones prácticas, manteniendo un área de investigación muy activa, en la que han florecido abundantes sistemas inteligentes que facilitan y permiten automatizar el trabajo rutinario. Dentro de la IA destaca un área de gran relevancia, cual es el AA (Pajares y Cruz, 2010), que comprende una amplia gama de métodos y aplicaciones capaces de procesar ingentes cantidades de datos para extraer la máxima información posible subyacente en los mismos. En términos generales, el esquema global de aprendizaje consta de tres módulos principales, a saber:

a) G*enerador*, que proporciona las entradas convenientemente estructuradas para su procesamiento, generalmente en forma de vectores como características, cuyas componentes son los atributos de los datos; por ejemplo, si los datos son imágenes, las características pueden ser los píxeles y los atributos los valores de color de estos.

b) *Entrenamiento*, de forma que a cada vector se le asigna una categoría de salida, constituyendo el objetivo para ajustar una serie de parámetros, que son el resultado del aprendizaje, finalizando con el modelo correspondiente convenientemente estructurado o aprendido. La mencionada asignación puede ser de naturaleza supervisada o no supervisada, en función de que haya un instructor o no, respectivamente, dirigiendo el proceso. Los vectores de entrada se denominan patrones (*patterns*), casos (*cases*), entradas (*inputs*), instancias (*instances*) u observaciones (*observations*). Las categorías de salida se denominan etiquetas (*labels*), objetivos (*targets*), salidas (*outputs*) y a veces también observaciones (*observations*).

c) *Decisión*, que asigna una categoría dada a una nueva muestra de entrada aplicando los parámetros aprendidos del modelo. La *clasificación*, la *regresión* o la *optimización*, entre otras, son tareas que pertenecen a esta fase.

Dentro del AA se encuadra el conocido como *Aprendizaje Profundo* (AP). En la literatura especializada a nivel internacional es muy común referirse al AP por sus términos en inglés, esto es, *Deep Learning*, que es el núcleo central del presente libro. Por otra parte, y al hilo de esta cuestión, conviene reseñar que son muchos y diversos los términos en inglés utilizados para definir y describir los conceptos involucrados bajo el paradigma del AP. Muchos de los cuales no poseen una clara traducción conceptual al español, razón por la cual los considerados bajo esta situación se mantienen a lo largo del libro con el fin de que el lector pueda fácilmente identificarlos en la literatura especializada escrita en inglés. Solo se han traducido aquellos conceptos que no admiten discusión, manteniendo en todo caso su expresión original en inglés.

Bien es cierto que desde los años 50 del siglo pasado, la IA a veces se ha sobrevalorado y se ha considerado como muy prometedora en diversas ocasiones, eso pese a que no se han llegado a alcanzar las perspectivas iniciales. Por otra parte, no es menos cierto que en los últimos años se están viendo avances importantes gracias al AP. Ello a pesar de que todavía es relativamente frágil de cara a su generalización y adaptación a entornos o escenarios cambiantes, principalmente por falta de datos suficientes que capturen los cambios de dicho entorno, pudiendo aparecer ciertos sesgos por falta de la información necesaria extraíble de los datos.

Algunos autores como Marcus y Davis (2019), expresan, no sin cierta razón, algunos aspectos relativos a las ventajas e inconvenientes de los procesos de AP, achacándoles que muchas arquitecturas basadas en redes neuronales hacen cosas increíbles, sin ser conscientes por parte de quien las aplica del conocimiento real sobre lo que están haciendo, y por ello, no son sistemas totalmente inteligentes.

Aunque en parte, esto último puede verse desde esta perspectiva, no es menos cierto que los desarrollos basados en AP son capaces de conseguir resultados importantes, siendo esta la perspectiva desde la que se abordan y plantean los temas del presente libro.

Como sostienen Goodfellow y col. (2016), en los comienzos de la IA, se abordaron rápidamente problemas intelectualmente difíciles para los seres humanos, pero relativamente sencillos para las computadoras, todo ello mediante una lista de reglas matemáticas formales. A partir de ahí, el verdadero desafío para la IA se tornó en resolver tareas fáciles de realizar para las personas, pero difíciles de describir formalmente. Aquí se incluyen tareas tales como el reconocimiento de objetos en imágenes, palabras o acciones en los movimientos. No cabe duda de que en este aspecto el AP ha conseguido ya logros muy relevantes. Es en este rasgo, y más concretamente en la exposición de una serie de técnicas orientadas a tal fin, donde se centra el presente libro.

En definitiva, se trata de exponer una serie de técnicas para resolver problemas, por decirlo de alguna manera, intuitivos para el ser humano, con el uso de las computadoras, que aprenden mediante los métodos y algoritmos diseñados a partir de los datos suministrados y sin necesidad de que los humanos especifiquen formalmente todo el conocimiento requerido por la computadora. En cualquier caso, y siguiendo también la teoría expuesta en Goodfellow y col. (2016), la jerarquía de conceptos permite que la computadora aprenda conceptos complicados al construirlos a partir de otros más simples, todos ellos estructurados en múltiples capas, razón por la cual a este enfoque se le denomina con el término ya indicado de AP. En cualquier caso, como una técnica específica del AA, estos procedimientos se encaminan a extraer patrones determinados a partir de los datos.

Existe una diferencia fundamental en lo que respecta a la extracción de las mencionadas características entre las técnicas clásicas, por llamarlas de alguna manera, de AA y las específicas del AP. Por ejemplo, considérese un ejemplo sencillo biclase, en el que se trata de separar el cielo y la hierba en una imagen de color de un paisaje de campo. Los datos disponibles en este caso son valores de color de los píxeles, de forma que en el caso del cielo predominan las tonalidades azules, mientras que en la hierba son las verdes. Un método simple tal como *naive* Bayes puede separar los patrones en dos clases diferentes teniendo en cuenta que los mismos están definidos, por lo que se conoce como características. La extracción de características en este caso es esencial. Por otro lado, y siguiendo en el ámbito de las imágenes, estas se caracterizan por poseer información espacial, y en vídeos, también temporal. Las *redes neuronales profundas* pueden captar perfectamente ambos tipos de información. En el primer caso, los *filtros de convolución* son

responsables de la captura espacial, a diferencia de lo que ocurre con otros modelos de red, tal como las de *retropropagación*, en las que las características de las imágenes se transforman en vectores que se suministran a la entrada, perdiendo las relaciones espaciales. Por ejemplo, una imagen de dimensión 3x3 se transforma en un vector con 9 componentes. En el caso de las características temporales las redes recurrentes tienen la habilidad de realizar tal captura.

No obstante, para muchas tareas no resulta fácil extraer características. Por ejemplo, supóngase que a partir de una imagen se quieren identificar peatones cuando un vehículo autónomo navega en un entorno urbano. Una persona puede identificarse por poseer cabeza, tronco y extremidades. Se podría pensar en detectar la presencia de extremidades o del cuerpo y la cabeza o todas, lo cual no resulta trivial debido a que no es fácil establecer las características de dichas partes. Los brazos y las piernas son alargados, el tronco tiene una forma más rectangular, pero en todos los casos nunca están exentos de elementos perturbadores, tales como el uso de distintos tipos de ropa, sombras, oclusiones totales o parciales, entre otros.

Una solución a este problema consiste en utilizar el AA para descubrir no solo la proyección de la representación a la salida, sino también la representación misma. Este enfoque se conoce como *aprendizaje de representación* según Goodfellow y col. (2016). Las representaciones aprendidas a menudo resultan en un rendimiento mucho mejor que el que se puede obtener con representaciones diseñadas a mano. También permiten que los sistemas de IA se adapten rápidamente a nuevas tareas, con una mínima intervención humana. Un algoritmo de aprendizaje de representación puede descubrir un buen conjunto de características para una tarea simple en minutos o para una tarea compleja en horas y meses. El diseño manual de características para una tarea compleja requiere una gran cantidad de tiempo y esfuerzo humanos, pudiendo llevar incluso décadas. Un ejemplo por excelencia de un algoritmo de aprendizaje de representación es el *autoencoder* (autocodificador), que convierte los datos de entrada en una representación diferente para luego poder devolverla a la representación original mediante el correspondiente *decoder* (decodificador).

Cuando se diseñan algoritmos para aprender las características, el objetivo consiste en separar los factores de variación que expliquen los datos observados. El concepto *factor* se refiere a abstracciones que ayudan a distinguir entre la alta variabilidad de los datos observados; así, en el reconocimiento de los peatones los factores de variación hacen referencia a la posición de las extremidades con respecto al tronco, la posición con respecto a la cámara, la ropa con la que van vestidos, las oclusiones de las extremidades, las posibles sombras proyectadas sobre sus cuerpos o la intensidad de la luz con la que se ha obtenido la imagen, entre otros. La mayoría

de las aplicaciones exigen separar los factores de variación descartando aquellos que no interesan. A la vista de lo cual, resulta francamente difícil obtener una representación tal que permita resolver el problema. Es precisamente aquí donde entra en acción el aprendizaje profundo, ya que permite introducir representaciones que se expresan en términos de otras representaciones a distintos niveles, que van estructurando convenientemente la información. Por ejemplo, la figura 1-1 muestra un sistema basado en AP, concretamente una Red Neuronal Convolucional (RNC) o en terminología inglesa *Convolutional Neural Network* (CNN), donde se representa el concepto de una imagen de una taza combinando conceptos más simples, tales como bordes, contornos o partes de los objetos hasta llegar a su clasificación, en este caso como taza. La idea de representación en múltiples capas es lo que determina una de las perspectivas del AP. De esta forma, se puede decir con carácter general, que las primeras capas de las redes profundas extraen características de bajo nivel, de modo que estas se van tornando en más complejas con características de mayor nivel hasta llegar a las capas superiores, en las que las características extraídas de la imagen son del más alto nivel. Esta información concatenada permite identificar un objeto (taza), a pesar de que pueda presentar diferentes características tales como, por ejemplo, color, forma o tamaño, lo que permite claramente diferenciar el AP del AA.

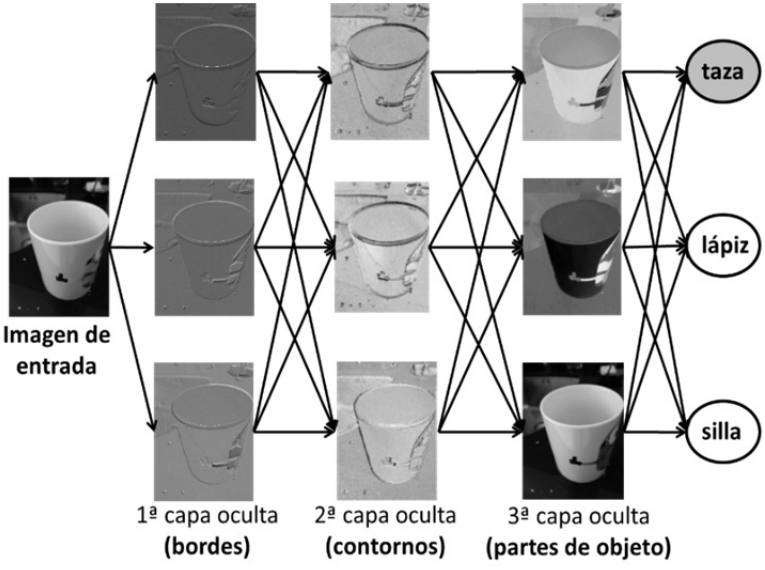

Fig. 1-1 Modelo de aprendizaje profundo

Otra idea para determinar el concepto de *profundidad* es la también establecida por Goodfellow y col. (2016), en el sentido de que la profundidad se determina como el estado del computador para aprender un programa computacional multipaso, de forma que cada capa de la representación puede verse como el estado de la

memoria del computador después de ejecutar otro conjunto de instrucciones en paralelo. Las redes con mayor profundidad pueden ejecutar más instrucciones en secuencia. Las instrucciones secuenciales ofrecen un gran poder porque las instrucciones posteriores pueden referirse a los resultados de instrucciones anteriores. Según esta visión del AP, no toda la información en las activaciones de una capa codifica necesariamente factores de variación que explican la entrada. La representación también almacena información de estado que ayuda a ejecutar un programa que puede dar sentido a la entrada. Esta información de estado podría ser análoga a un contador o puntero en un programa de computación tradicional. No tiene nada que ver con el contenido de la entrada específicamente, pero ayuda al modelo a organizar su procesamiento. Existen dos formas principales de medir la profundidad de un modelo. La primera se basa en el número de instrucciones secuenciales que deben ejecutarse para evaluar la arquitectura. Se puede pensar en esto como la longitud de la ruta más larga a través del diagrama de flujo que describe cómo calcular cada una de las salidas del modelo dadas sus entradas. Otro enfoque, utilizado por modelos probabilísticos profundos, considera que la profundidad de un modelo no es la profundidad del gráfico computacional sino la profundidad del gráfico que describe cómo se relacionan los conceptos entre sí.

En este caso, la profundidad del diagrama de flujo de los cálculos necesarios para computar la representación de cada concepto puede ser mucho más profunda que la gráfica de los conceptos en sí mismos. Esto se debe a que la comprensión del sistema de los conceptos más simples puede refinarse dando información sobre los conceptos más complejos. Por ejemplo, siguiendo también a Goodfellow y col. (2016), un sistema inteligente que observa una imagen de una cara con un ojo en la sombra puede ver inicialmente únicamente un ojo. Después de detectar la presencia de una cara, el sistema puede inferir que probablemente también esté presente un segundo ojo. En este caso, la gráfica de conceptos incluye sólo dos capas, una capa para *ojos* y una capa para *caras*, pero la gráfica de cálculos incluye dos capas si se refina la estimación de cada concepto dadas las otras *n* veces. Debido a que no siempre está claro cuál de estos dos modelos (la profundidad del gráfico computacional o la profundidad del gráfico de modelado probabilístico) es más relevante, y debido a que distintas personas eligen diferentes conjuntos de elementos más pequeños a partir de los cuales construir sus gráficos, no existe un único valor correcto para la profundidad de una arquitectura, así como tampoco hay un valor correcto único para la duración de un programa computacional. Tampoco existe un consenso acerca de la profundidad que un modelo requiere para calificarlo como “profundo”. Sin embargo, el aprendizaje profundo puede considerarse, con seguridad, como el estudio de modelos que implican una mayor cantidad de composición de funciones aprendidas o conceptos aprendidos que el aprendizaje automático tradicional.

En definitiva, el aprendizaje profundo es un tipo particular de aprendizaje automático que consigue un gran poder y flexibilidad al representar al mundo como una jerarquía anidada de conceptos, donde cada concepto se define en relación con conceptos más simples y representaciones más abstractas calculadas en términos de conceptos menos abstractos.

1.2 BREVE REVISIÓN HISTÓRICA

Tal y como sostienen Goodfellow y col. (2016), el concepto de aprendizaje profundo (AP) data de los años 40 del siglo pasado a pesar de su actual relevancia y auge, lo que ciertamente parece sorprendente. Ello se debe al hecho de que durante cierto tiempo las aplicaciones basadas en este paradigma no consiguieron el éxito que se les suponía inicialmente.

Una primera etapa corresponde a la época en la que el AP se enmarca en el contexto cibernético, aproximadamente entre los años 1940 y 1960 y coincidiendo con la época del desarrollo de las teorías biológicas del aprendizaje como las desarrolladas por McCulloch y Pitts (1943) o Hebb (1949) a las que siguieron las implementaciones de los primeros modelos, entre los que destaca el perceptrón (Rosenblatt, 1958), que es la unidad básica en las redes neuronales o ADALINE (Widrow y Hoff, 1960). Por tanto, en esta fase prevalecen los modelos lineales, de forma que a partir de un conjunto de *n* valores de entrada, la salida es una combinación lineal en la que intervienen los pesos de un modelo que se ajustan durante la fase de aprendizaje. No obstante, la incapacidad para resolver el problema XOR con modelos lineales llevó, en parte, a la decadencia en la popularidad de las redes neuronales.

Las redes neuronales artificiales fueron uno de los nombres que recibió el AP, evolucionando posteriormente a un principio más general de aprendizaje en el sentido de configurarse como múltiples niveles de composición. No obstante, a diferencia de aquellos modelos iniciales basados en la neurociencia, ahora en el AP la neurociencia es la inspiración, pero no el modelo, entre otras razones por el desconocimiento del cerebro humano para considerarse una guía, ya que incluso se desconocen operaciones del cerebro elementales (Olshausen y Field, 2005).

No obstante, la neurociencia proporciona una perspectiva avanzada en el ámbito del AP proporcionando la idea de que quizás un único algoritmo en este campo puede ser válido para resolver distintas tareas, tal y como sostienen Goodfellow y col. (2016). Esto es así a raíz del descubrimiento por parte de Von Melchner y col. (2000), de que los hurones pueden llegar a ver cosas con la región auditiva de su cerebro. Esto ha permitido que diversas disciplinas converjan hacia desarrollos y