

# Lectura 1

## Curso: Inteligencia artificial

Tomado desde The Deep Learning Revolution de Terrence J. Sejnowski The MIT Press Cambridge, London England 2018.

"El presente texto ha sido preparado de manera exclusiva para los alumnos del Curso Inteligencia Artificial, que forma parte de la Plan de Estudio de la Escuela de Ciencia de Computación, según el artículo 44 de la Ley sobre el Derecho de Autor, D.L. N°822. Queda prohibida su difusión y reproducción por cualquier medio o procedimiento, total o parcialmente fuera del marco del presente curso".

---

## 1 El renacimiento de la inteligencia artificial

Marvin Minsky fue un matemático brillante y fundador del Laboratorio de Inteligencia Artificial (MIT AI Lab) . Los fundadores establecieron la dirección y la cultura de un campo y, gracias en gran parte a Minsky, la inteligencia artificial en el MIT en la década de 1960 fue un bastión de inteligencia. Rebotante de más ideas por minuto que cualquier otra persona que supiera, podría convencerte de que su decisión sobre un problema era correcta, incluso cuando el sentido común te decía lo contrario. Admiré su audacia y su astucia, pero no la dirección que tomó la AI.

### 1.1 Juego de Niños

Blocks World es un buen ejemplo de un proyecto que surgió del MIT AI Lab en la década de 1960. Para simplificar el problema de la visión, Blocks World consistía en bloques de construcción rectangulares que podrían apilarse para crear estructuras. El objetivo era escribir un programa que pudiera interpretar un comando, como "Encontrar un bloque amarillo grande y colocarlo encima del bloque rojo" y planificar los pasos necesarios para que un brazo robot lleve a cabo el comando. Esto parece un juego de niños, pero se tuvo que escribir un programa grande y complejo, uno que se volvió tan engorroso que no se pudo depurar fácilmente y se abandonó de manera efectiva cuando el estudiante que escribió el programa, Terry Winograd, abandonó el MIT.

Este problema aparentemente simple era mucho más difícil de lo que nadie pensaba que sería, e incluso si hubiera tenido éxito, no había un camino directo desde Blocks World al mundo real, donde los objetos vienen en muchas formas, tamaños y pesos, y no todos los ángulos son ángulos rectos. En comparación con un entorno de laboratorio controlado donde la dirección y el nivel de iluminación pueden ser fijos, en el mundo real, la iluminación puede variar dramáticamente de un lugar a otro y de vez en cuando, lo que complica enormemente la tarea de reconocimiento de objetos para las computadoras.

En la década de 1960, el MIT AI Lab recibió una gran subvención de una agencia de investigación militar para construir un robot que pudiera jugar al ping-pong. Una vez escuché una historia que el investigador principal olvidó pedir dinero en la propuesta de subvención para construir un sistema de visión para el robot, por lo que asignó el problema a un estudiante graduado como un proyecto de verano. Una vez le pregunté a Marvin Minsky si la historia era cierta. Me contestó que estaba equivocado: "Asignamos el problema a los estudiantes de pregrado". Un documento de los archivos en el MIT confirma su versión de la historia. Lo que parecía que sería un problema fácil de resolver resultó ser una solución rápida. Borró a una generación de investigadores en visión por computador.

## 1.2 ¿Por qué la visión es un problema difícil?

Rara vez tenemos dificultades para identificar qué es un objeto a pesar de las diferencias en la ubicación, el tamaño, la orientación y la iluminación del objeto. Una de las ideas más tempranas en la visión por computadora fue hacer coincidir una plantilla del objeto con los píxeles en la imagen, pero ese enfoque falló porque los píxeles de las dos imágenes del mismo objeto en diferentes orientaciones no coinciden.

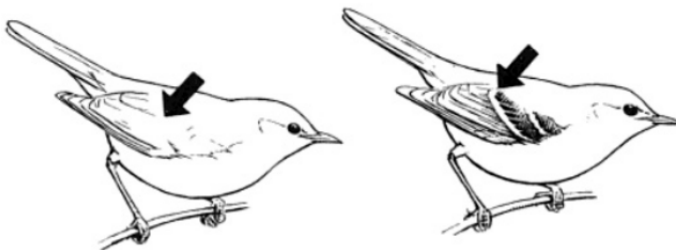
Por ejemplo, considere dos aves.



Si cambia la imagen de un ave sobre la otra, puede hacer que una parte coincida, pero el resto está fuera de registro; pero puede obtener una buena coincidencia con una imagen de otra especie de ave en la misma postura.

El progreso en la visión por computador se hizo enfocándose no en los píxeles sino en las características. Por ejemplo, los observadores de aves tienen que convertirse en expertos para distinguir entre diferentes especies que pueden diferir en solo unas pocas marcas sutiles. Un libro práctico y popular sobre la identificación de aves tiene solo una fotografía de un ave, pero muchos dibujos esquemáticos señalan las diferencias sutiles entre ellas.

Una buena característica es una que es única para una especie de ave, pero debido a que las mismas características se encuentran en muchas especies, lo que hace posible identificar un ave es la combinación única de varias marcas de campo, como barras del ala, rayas de ojos y cuando estas marcas de campo son compartidas por especies estrechamente relacionadas, hay llamadas y canciones que distinguen una de otra. Los dibujos o pinturas de aves son mucho mejores para dirigir nuestra atención a las características distintivas relevantes que las fotografías, que están llenas de cientos de características menos relevantes.



El problema con este enfoque basado en características no es solo que requiere mucho trabajo desarrollar detectores de características para los cientos de miles de objetos diferentes en el mundo, sino que, incluso con los mejores detectores de características, surgen ambigüedades de imágenes de objetos que están parcialmente ocluidas, lo que hace que el reconocimiento de objetos en escenas desordenadas sea una tarea desalentadora para las computadoras.

En la década de 1960, nadie sospechaba que llevaría cincuenta años y un millón de veces más de potencia informática antes de que la visión informática alcanzara rendimientos de niveles humanos. La intuición engañosa de que sería fácil escribir un programa de visión artificial se basa en actividades que nos resultan fáciles de realizar, como ver, escuchar y moverse, pero eso llevó a la evolución en millones de años en hacerlo bien. Para su disgusto, los primeros pioneros de la IA consideraron que el problema de la visión de la computadora era extremadamente difícil de resolver. En contraste, encontraron mucho más

fácil programar computadoras para probar teoremas matemáticos, un proceso que se cree que requiere los niveles más altos de inteligencia, porque las computadoras resultan ser mucho mejores en lógica que nosotros.

Ser capaz de pensar lógicamente es un desarrollo tardío en la evolución e, incluso en los seres humanos, requiere entrenamiento para seguir una larga línea de proposiciones lógicas hasta llegar a una conclusión rigurosa, mientras que, para la mayoría de los problemas, debemos necesitar resolver para sobrevivir, las generalizaciones de experiencias anteriores funcionan bien para nosotros la mayor parte del tiempo.

### 1.3 Sistemas expertos

Popular en los años 70 y 80, los sistemas expertos de inteligencia artificial se desarrollaron para resolver problemas como el diagnóstico médico mediante un conjunto de reglas. De este modo, se desarrolló un sistema experto temprano, MYCIN, para identificar las bacterias responsables de enfermedades infecciosas como la meningitis. Siguiendo el enfoque del sistema experto, los desarrolladores de MYCIN primero tuvieron que recopilar datos y reglas de los expertos en enfermedades infecciosas, así como los síntomas y las historias clínicas. de los pacientes, luego ingresar éstos en la computadora del sistema y finalmente programar la computadora para hacer inferencias usando la lógica.

Sin embargo, los desarrolladores tuvieron dificultades para recopilar los hechos y las reglas de los expertos, especialmente en los dominios más complejos, donde los mejores diagnosticadores no confían en las reglas sino en el reconocimiento de patrones basado en la experiencia, que es difícil de codificar, y donde se encuentran con un sistema que tuvo que actualizarse continuamente a medida que se descubrían nuevos hechos y las reglas antiguas quedaban obsoletas. Y encontraron más dificultades para recopilar e ingresar los síntomas e historial médico de los pacientes en la computadora del sistema, un proceso que podría demorar media hora o más por paciente, más tiempo del que podía permitirse un médico ocupado. No en vano, MYCIN nunca se utilizó clínicamente. Aunque muchos sistemas expertos se escribieron para otras aplicaciones, como la gestión de derrames tóxicos, la planificación de misiones para vehículos autónomos y el reconocimiento de voz, pocos están en uso hoy en día.

Los investigadores probaron muchos enfoques diferentes en las primeras décadas de la IA, pero sus enfoques eran más inteligentes que prácticos. No solo subestimaron la complejidad de los problemas del mundo real, sino que las soluciones que propusieron se escalaron mal. En dominios complejos, la cantidad de reglas puede ser enorme y, a medida que se agregan nuevos hechos a mano, no es práctico hacer un seguimiento de las excepciones e interacciones con otras reglas.

Douglas Lenat, por ejemplo, comenzó un proyecto llamado "Cyc" en 1984 para codificar el sentido común, que en ese momento parecía una buena idea, pero resultó ser una pesadilla en la práctica. Damos por sentado un número ilimitado de hechos sobre la forma en que funciona el mundo, la mayoría de los cuales se basan en la experiencia. Por ejemplo, un gato que cae de 40 pies probablemente evitará el daño, pero un humano que cae de la misma altura probablemente no lo hará.

Otra razón por la que el progreso en la IA inicial fue tan lento fue que las computadoras digitales eran increíblemente primitivas y la memoria era extremadamente costosa para los estándares actuales. Pero como las computadoras digitales son altamente eficientes en las operaciones lógicas, la manipulación de símbolos y la aplicación de reglas, no es demasiado sorprendente que estas primitivas computacionales se vean favorecidas en el siglo veinte. Así, en 1955, Allen Newell y Herbert Simon, dos científicos de la computación de la Universidad Carnegie Mellon, pudieron escribir un programa informático llamado "Logic Theorist" que podría probar los teoremas lógicos de Principia Mathematica de Alfred North Whitehead y Bertrand Russell en el intento de sistematizar todo matemáticas. Había grandes expectativas en estos primeros días de que las computadoras inteligentes estaban a la vuelta de la esquina.

A los pioneros de la IA que intentaban escribir programas de computadora con la funcionalidad de la inteligencia humana no les importaba cómo el cerebro en realidad lograba un comportamiento inteligente. Cuando le pregunté a Allen Newell por qué, me dijo que personalmente había estado abierto a las ideas de la investigación del cerebro, pero que en ese momento no se sabía lo suficiente como para ser de mucha utilidad. Los principios básicos de la función cerebral recién estaban surgiendo en la década de 1950, liderados por el trabajo de Alan Hodgkin y Andrew Huxley, quienes explicaron cómo las señales del cerebro son transportadas a larga distancia por picos eléctricos en los nervios de todo o nada, y Bernard

Katz, quien descubrió pistas sobre cómo estas señales eléctricas se convierten en señales químicas en las sinapsis, que se comunican entre las neuronas.

Aunque, en la década de 1980, se conocía más sobre el cerebro y era más accesible fuera del campo de la biología, el cerebro mismo se había vuelto irrelevante para la nueva generación de investigadores de IA, cuyo objetivo era escribir un programa que fuera funcionalmente equivalente a cómo el cerebro funciona. En filosofía, esta postura se llamó funcionalismo, que para muchos fue una buena excusa para ignorar los detalles desordenados de la biología. Pero un pequeño grupo de investigadores de la IA que no formaban parte de la corriente principal creía que un enfoque de la inteligencia artificial inspirado por la biología real del cerebro y llamado de varias maneras "redes neuronales", "conexionismo" y "procesamiento distribuido en paralelo" podría eventualmente resolver problemas difíciles que habían eludido la IA basada en la lógica.

Yo era uno de ese grupo.

## 1.4 En la guarida del león

En 1989, Michael Dertouzos, jefe del Laboratorio de Ciencias de la Computación del MIT, me invitó a dar una distinguida conferencia en el MIT sobre mi enfoque pionero de la inteligencia artificial basada en redes neuronales. Al llegar allí, fui recibido calurosamente por Dertouzos, quien, mientras viajábamos juntos en el ascensor, me dijo que era una tradición del MIT que el distinguido profesor tomara cinco minutos para iniciar una discusión con profesores y estudiantes sobre su tema durante el almuerzo. "Y", agregó mientras se abrían las puertas del ascensor, "odian lo que haces".

La sala estaba repleta de unas cien personas, lo que sorprendió incluso a Dertouzos. Los científicos se pararon en círculos a tres filas de profundidad: la primera fila para el profesorado superior, el profesorado de secundario en la segunda fila y los estudiantes en las filas más allá de ellos. Y yo estaba en el centro, estacionado frente al buffet, el plato principal. ¿Qué podría decir en cinco minutos que podría hacer una diferencia para una audiencia que odiaba lo que estaba haciendo?

Improvisé: "Esa mosca en la comida tiene un cerebro con solo 100,000 neuronas; "Pesa un miligramo y consume un milivatio de potencia", dije, volando. "La mosca puede ver, puede volar, puede navegar y puede encontrar comida. Pero lo verdaderamente notable es que puede reproducirse. MIT posee una supercomputadora que cuesta 100 millones: consume un megavatio de energía y se enfría con un enorme acondicionador de aire. Pero el mayor costo de la supercomputadora es el sacrificio humano en forma de programadores para alimentar su voraz apetito por los programas. Ese superordenador no puede ver, no puede volar, y aunque se comunica con otras computadoras, no puede aparearse o reproducirse. ¿Qué hay de malo con este enfoque?"

Después de una larga pausa, un miembro de la facultad dijo: "Porque aún no hemos escrito el programa de visión". (El Departamento de Defensa recientemente invirtió 600 millones en su Strategic Computing Initiative, un programa que se desarrolló desde 1983 hasta 1993 pero quedó corto en la construcción de un sistema de visión para guiar un tanque de conducción automática. "Buena suerte con eso", fue mi respuesta. Gerald Sussman, quien hizo varias aplicaciones importantes de la IA a problemas del mundo real, incluido un sistema para la integración de alta precisión para la mecánica orbital, defendió el honor del enfoque del MIT a la IA con una apelación al trabajo clásico de Alan Turing, quien había demostrado que la máquina de Turing, un experimento mental, podría calcular cualquier función computable.

"¿Y cuánto tiempo tomaría?" Pregunté. "Será mejor que calcules rápidamente o te comerán", agregué, luego crucé la habitación para servirme una taza de café. Y ese fue el final del diálogo con la facultad.

"¿Qué está mal con esta imagen?" Es una pregunta que todos los estudiantes de mi laboratorio pueden responder. Pero las dos primeras filas de mi audiencia a la hora del almuerzo quedaron perplejas. Finalmente, un estudiante en la tercera fila ofreció esta respuesta: "La computadora digital es un dispositivo de propósito general, que puede programarse para computar cualquier cosa, aunque sea ineficiente, pero la mosca es una computadora de propósito especial que puede ver y volar pero puede que no equilibre mi chequera ". Esta fue la respuesta correcta. Las redes de visión en el ojo de la mosca evolucionaron a lo largo de cientos de millones de años, y sus algoritmos de visión están integrados en las propias redes. Esta es la razón por la que puede hacer una ingeniería inversa trabajando con el diagrama de cableado y el flujo de información a través de los circuitos neuronales del ojo y por qué no puede hacer eso para

una computadora digital, donde el hardware por sí solo necesita software para especificar qué problema se está resolviendo

Reconocí a Rodney Brooks sonriendo en la parte de atrás de la multitud, alguien a quien había invitado una vez a un taller sobre neurociencia computacional en Woods Hole en Cape Cod, Massachusetts. Brooks es de Australia y, en la década de 1980, fue miembro de la facultad junior en el MIT AI Lab, donde construyó insectos robóticos ambulantes utilizando una arquitectura que no dependía de la lógica digital. Eventualmente se convertiría en el director del laboratorio y fundaría iRobot, la compañía que fabrica Roombas. La sala donde di mi conferencia esa tarde era enorme y estaba llena de un gran contingente de estudiantes de pregrado, la próxima generación mirando hacia el futuro en lugar del pasado. Hablé sobre una red neuronal que aprendió a jugar al backgammon, un proyecto en el que colaboré con Gerald Tesauero, físico del Center for Complex Systems Research de la Universidad de Illinois en Urbana-Champaign. El backgammon es una carrera hasta el final entre dos jugadores, con piezas que avanzan en función de cada tirada de dados, pasando una sobre la otra en el camino. A diferencia del ajedrez, que es determinista, el backgammon se rige por el azar: la incertidumbre con cada tirada de dados hace que sea más difícil predecir el resultado de un movimiento en particular. Es un juego muy popular en el Medio Oriente, donde algunos se ganan la vida jugando al backgammon de alto riesgo.

En lugar de escribir un programa basado en la lógica y la heurística para manejar todas las posiciones posibles del tablero, una tarea imposible dado que hay  $10^{20}$  posibles puestos en el tablero de backgammon, la red aprendió a jugar a través del reconocimiento de patrones al ver jugar a un maestro. Gerry continuó para crear el primer programa de backgammon que se jugó en los niveles de campeonato mundial haciendo que la red de backgammon juegue así mismo.

Después de mi conferencia, me enteré de que había un artículo de primera plana en el New York Times esa mañana sobre cómo las agencias gubernamentales estaban recortando los fondos para la inteligencia artificial. Aunque este fue el comienzo de un invierno de inteligencia artificial para los investigadores principales, no me afectó ni a mí ni al resto de mi grupo, para quienes la primavera de la red neuronal acababa de comenzar.

Pero nuestro nuevo enfoque hacia la IA llevaría veinticinco años para entregar aplicaciones del mundo real en visión, el habla y lenguaje. Incluso en 1989, debería haber sabido que llevaría tanto tiempo. En 1978, cuando era un estudiante graduado en Princeton, extrapolaba la ley de Moore para el aumento exponencial en el poder de cómputo, aplicándose cada 18 meses, para ver cuánto tardaría en alcanzar los niveles cerebrales del poder de cómputo y concluí que sucedería en 2015, afortunadamente, eso no me impidió seguir adelante.

Mi creencia en las redes neuronales se basaba en mi intuición de que si la naturaleza hubiera resuelto estos problemas, también podríamos aprender de la naturaleza cómo resolverlos. Los veinticinco años que tuve que esperar no fueron ni siquiera un abrir y cerrar de ojos en comparación con los cientos de millones de años que tomó la naturaleza.

Dentro de la corteza visual, las neuronas están dispuestas en una jerarquía de capas. A medida que la información sensorial se transforma capa cortical por capa cortical, la representación del mundo se vuelve más y más abstracta. A lo largo de las décadas, a medida que aumentó el número de capas en los modelos de redes neuronales, su desempeño continuó mejorando hasta que finalmente se alcanzó un umbral crítico que nos permitió resolver problemas que solo podíamos soñar con resolver en los años ochenta. El aprendizaje profundo automatiza el proceso de encontrar buenas características que distingan diferentes objetos en una imagen, y es por eso que la visión por computadora es mucho mejor hoy que hace cinco años.

Para 2016, las computadoras se habían vuelto un millón de veces más rápidas y la memoria de las computadoras había aumentado mil millones de veces de megabytes a terabytes. Se hizo posible simular redes neuronales con millones de unidades y miles de millones de conexiones, en comparación con redes en la década de 1980 que tenían solo cientos de unidades y miles de conexiones. Aunque todavía son pequeñas para los estándares de un cerebro humano, que tiene cien mil millones de neuronas y un millón de mil millones de conexiones sinápticas, las redes de hoy son ahora lo suficientemente grandes como para demostrar pruebas de principio en dominios reducidos.

El aprendizaje profundo en redes neuronales profundas ha llegado. Pero antes de que existieran redes profundas, tuvimos que aprender a entrenar redes poco profundas.