El algoritmo AlphaZero basado en una red neuronal que aprende por reforzamiento ha destrozado a la mejor máquina de ajedrez existente. En tan solo cuatro horas AlphaZero ha recorrido la historia de la humanidad en el ajedrez

La noticia

AlphaZero vapulea a Stockfish, el mejor programa de ajedrez hasta la fecha.

El 5 de diciembre de 2017 el equipo de DeepMind, empresa especializada en inteligencia artificial propiedad de Google, publicó la noticia de que su programa AlphaZero había derrotado al programa Stockfish por 28 partidas ganadas, 72 tablas y ninguna derrota.

El logro podría considerarse una simple victoria de una máquina sobre otra o una repetición de las, cada vez más frecuentes, victorias de la máquina sobre el humano. Sin embargo, varias de las características del encuentro llevan a una seria reflexión sobre el ajedrez, la inteligencia artificial (IA) y el papel del ser humano en el mundo que viene.

Lo más notable es que AlphaZero aprendió desde la nada, tabula rasa, y en tan solo 4 horas logró un conocimiento del ajedrez superior al que toda la historia de la humanidad había conseguido durante 1.500 años máquinas incluidas.

Más aún, todo parece indicar que el esfuerzo humano por dotar a las máquinas de habilidades en el ajedrez había sido inútil. El conocimiento humano parece ser erróneo y contraproducente. Un prejuicio.

El juego

El ajedrez es un juego milenario. Durante el siglo XX alcanzó una enorme popularidad en Occidente y después en todo el mundo. Está lleno de brillantes figuras e históricos encuentros. Desde el cubano Raúl Casablanca a toda una pléyade de jugadores rusos. El duelo del estadounidense Boby Fisher frente al ruso Boris Spasky puso de moda el deporte del tablero. Los enfrentamientos entre Karpov y Kasparov, la pérdida del cetro humano por Kasparov frente a la máquina Deep Blue y los logros de la rutilante estrella noruega, Magnus Carlsen son pocos de los miles de apasionantes hitos que llenan las páginas de la historia del ajedrez.

El ajedrez es un juego finito pero en la práctica puede considerarse infinito y no resuelto. Es decir, las posibilidades son tantas que ningún superordenador está cerca de resolverlo aunque teóricamente ese momento llegará.

En el primer movimiento, las blancas tienen 20 jugadas legales (cada peón una o dos casillas -8×2=16- y cada caballo dos casillas -2×2=4-). A su vez, las negras pueden jugar otras 20 jugadas. Al final de la primera jugada hay 400 posibles posiciones en el tablero. El venerable Deep Blue (hoy casi un juguete) calcularía los tres primeros movimientos (tres jugadas por bando, es decir, seis jugadas) en menos de un segundo. En cambio, los cuatro primeros movimientos los calcularía en ¡dos horas! y los cinco primeros, en ¡853 horas! Más adelante el juego se complica y las jugadas posibles son alrededor de 35 por bando. El mejor ordenador actual tardaría millones de años en resolverlo.

AlphaZero aprendió en solo cuatro horas y logró un conocimiento del ajedrez superior al que la humanidad había conseguido en 1.500 años

Sin embargo, hay una parte resuelta: los finales de siete piezas o menos. El logro no fue empresa menor. Los finales de seis piezas estaban resueltos en 2005 (mediante las tablas de Nalimov). En 2012 el superordenador Lomonosov de la Universidad de Moscú logró resolver los finales de siete piezas. Estos se hallan en unas tablas. Es decir, ya no hay que calcular y basta con buscar la posición para obtener todas las variantes posibles con una certeza del 100%. La base de datos ocupa la friolera de 140 terabytes (TB). La posición más complicada requiere 545 jugadas para dar mate. Se calcula que por cada pieza adicional el esfuerzo de computación se multiplica por 100. Resumiendo: queda mucho tiempo hasta que el ajedrez sea un juego resuelto.

La otra parte del juego donde la teoría ha avanzado más es la apertura. Millones de horas y miles de jugadores, partidas y programas se han dedicado a establecer cuáles son las mejores líneas de juego en su inicio. Una de las más populares y antiguas es la apertura Ruy López que toma el nombre del clérigo que la popularizó y que describió en su libro publicado en Alcalá de Henares (Madrid) en 1561. Después, un largo rosario de aperturas, defensas y sus variantes se han ido sumando a la teoría de aperturas. Todos los grandes jugadores estudian aperturas ayudados de los programas informáticos y ocasionalmente desarrollan nuevas variantes.

El medio juego es la parte donde menos teoría hay. Las variantes son virtualmente infinitas, las posiciones se repiten poco y priman ideas generales como la estrategia y los recursos tácticos.

El rival

Stockfish es uno de los motores de ajedrez más potentes del mundo, mucho más fuerte que cualquier Gran Maestro[1] humano. Los otros rivales son Houdini y Komodo, entre otros. Una buena parte de su fuerza reside en que es *open source*, de código abierto. Esto significa que cualquiera puede mejorar el código y de hecho hay cientos de desarrolladores que trabajan en él. Una vez que un *software* es de código abierto y se hace popular, las marcas comerciales tienen pocas posibilidades de competir. Es el caso de Wikipedia que ha barrido a todas las enciclopedias privadas. Pero también el de otro mucho código, desde las bases de datos a los servidores web.

En 1997 Deep Blue ganó a Kasparov y desde entonces las máquinas ganan a los humanos. Stockfish es heredero de la línea que abrió Deep Blue y en realidad los programas no han cambiado tanto. Se basan en un potente hardware (mucho más en el caso de Stockfish que en el de Deep Blue ya que 20 años no han pasado en vano) y en dos conjuntos de algoritmos.

El algoritmo Minimax evalúa las posiciones. Es algo natural en los ajedrecistas. Por ejemplo, se evalúa el material. Una dama vale más que un peón y una torre más que un caballo. El material es muy importante, pero no es el único criterio: un peón a punto de coronar y hacerse dama vale mucho; un caballo en el centro, también; y una dama en una esquina tiene menos valor.

Los programas de ajedrez tienen todos los criterios humanos escritos en su código y el trabajo del programador es facilitarle los patrones de evaluación

Los programas de ajedrez tienen todos estos criterios humanos escritos en su código y una parte esencial del programador es darle los mejores patrones de evaluación. El programa no evalúa jugadas sino posiciones. Si una jugada lleva a una posición mejor, entonces la jugada es buena.

Como hemos visto, no es posible calcular todas las jugadas, de modo que hace falta un algoritmo que pode todo el árbol de movimientos. Este algoritmo se llama poda alpha beta y en realidad es un conjunto de algoritmos. Se basan en detectar la línea más prometedora y evaluarla a fondo descartando otras líneas mediante heurística. Es evidente que el programa puede descartar la mejor jugada por mal diseño y por ello la calidad de la poda alpha beta es esencial. Todas estas heurísticas son mecanismos que siguen criterios humanos y que los programadores se han esforzado por mejorar e implementar en la máquina.

La clasificación de ajedrez se mide por un sistema de puntuación llamado ELO[2] . El mejor jugador actual, Magnus Carlsen, tiene un ELO de 2.837 mientras que Stockfish tiene un ELO de 3.228 puntos. Un abismo. Los Grandes Maestros ya no juegan contra máquinas, pero las usan para entrenarse y aprender.

El alienígena

En estas estaba el mundo del ajedrez cuando llegó un ser que parecía venido de otro mundo: AlphaZero.

El origen de AlphaZero está en el juego del Go. Después de ganar al ajedrez, la inteligencia artificial se planteó otros juegos. El que más se resistió fue Go, otro juego milenario. Recientemente DeepMind creó un programa llamado AlphaGo que derrotó al vigente campeón mundial humano. AlphaGo tenía mucho conocimiento de Go imbuido. Google decidió a probar un software que no supiera nada de Go. De modo que creó AlphaGo Zero y lo enfrentó a la anterior versión, AlphaGo. La nueva versión Zero, que no sabía nada de Go machacó a la primera que sí tenía conocimientos del juego.

Después, DeepMind generalizó el algoritmo de AlphaGo Zero para llegar a AlphaZero que se ha usado para ganar a Go, ajedrez y shogi (ajedrez japonés).

AlphaZero usa una red neuronal que aprende. Las redes neuronales son una idea tan antigua como la informática, que data de los años 50 del siglo pasado. Se basa en una red de neuronas artificiales conectadas y organizadas en capas. Cada neurona recibe estimulación (input) de algunas neuronas de las capas iniciales y ella misma estimula a algunas neuronas de las capas posteriores (output). La red aprende cambiando la fuerza de estas conexiones.

Las redes neuronales eran un invento interesante pero no solucionaban ningún problema real. Hasta que hace unos cinco años todo ha cambiado. Hay tres elementos fundamentales que han impulsado las redes neuronales. Uno es la potencia del hardware que permite velocidades enormes. Otro son los algoritmos (como el que nos ocupa AlphaZero) que son mucho más complejos. Por último, la ingente cantidad de datos disponibles con las que se pueden entrenar (aunque en AlphaZero no se aplica).

El caso es que las redes neuronales que aprenden están empezando a ser ubicuas. Por ejemplo el reconocimiento de voz o la traducción se basan en redes neuronales. No es de extrañar que Google comprara DeepMind y que realice experimentos como AlphaZero. Es su negocio. AlphaZero es una red neuronal que no tiene ningún conocimiento de Go, ajedrez ni shogi. No sabe nada de aperturas ni finales, de algoritmos Minimax ni poda alpha beta. Para cada desafío se le dan las reglas del juego y el resultado (gana, empata o pierde). Técnicamente es un algoritmo general de aprendizaje por reforzamiento.

Ya en 2015, DeepMind había preparado el camino. Entonces anunció que un software había sido capaz de aprender sin ayuda a jugar a los míticos juegos de la videoconsola de Atari. El programa solo recibía como input los píxeles de la pantalla y el resultado del juego. Este aprendizaje por reforzamiento no supervisado es un gigantesco avance. Las máquinas aprenden cosas para las que no habían sido entrenadas solo con darles el resultado final.

AlphaZero ha utilizado para aprender un hardware muy poderoso. Más de 5.000 TPUs[3]. Jugando contra sí mismo, empezó sin saber nada de ajedrez y en 4 horas logró el nivel ELO de Stockfish. Después siguió jugando hasta completar 24 horas.

El enfrentamiento

Una vez entrenado, AlphaZero se enfrentó a Stockfish en un *match* de 100 partidas. La paliza fue sensacional. Jugando con blancas ganó la mitad, 25 partidas y las otras 25 fueron tablas. Con negras ganó tres y empató 47. En términos ajedrecísticos, donde las tablas entre Grandes Maestros son muy habituales, el resultado es abultadísimo.

El hardware usado para jugar es casi ridículo comparado con el usado en el entrenamiento. Tan solo cuatro TPUs en el juego frente a las 5.000 del entrenamiento. Comparando la potencia de cálculo de los dos rivales, las cifras son también escandalosas. Stockfish realizaba 70 millones de evaluaciones por segundo mientras AlphaZero solo 80.000, un 99,89% menos.

La conclusión relativa a la potencia de cálculo es clara: una red neuronal requiere muchos recursos para aprender pero una vez entrenada apenas los consume.

Además de ganar al ajedrez, AlphaZero también ganó a los mejores programas de Go y shogi aunque fueron instancias separadas para cada juego con sus reglas y su aprendizaje específico distinto para cada una.

DeepMind publicó un artículo con varias explicaciones del experimento en el que se incluye una somera descripción del algoritmo, las condiciones experimentales y un conjunto de 10 partidas ganadas por AlphaZero.

Las reacciones en el mundo del ajedrez han sido de completo asombro. A estas alturas hay que considerar a Stockfish del lado humano ya que tiene mucho conocimiento de ajedrez inculcado por los programadores y además es muy usado por los jugadores para aprender y mejorar.

El conocimiento humano es un

prejuicio. Es inútil, erróneo y un lastre. Las máquinas aprenden por sí solas y todas nuestras ayudas y conocimientos solo sirven para confundirlas

Si el apabullante resultado causó estupor en los expertos, las partidas publicadas lo hicieron aún más. En estos momentos hay miles de ajedrecistas repasando las partidas y esperando a que se publiquen más. Simplemente AlphaZero hacía cosas que nadie podía esperar incluyendo los mejores equipos hombre máquina.

Entregar una pieza en un sacrificio es común. El contrario no suele aceptarla porque suele ser un ofrecimiento envenenado. Sin embargo, AlphaZero además de sacrificarla, no tiene ninguna prisa en recuperarla lo que provoca el desconcierto de programas y humanos. En varias ocasiones la jugada realizada por AlphaZero ni siquiera era considerada por las máquinas. Más aún, la mayoría de los programas de ajedrez dan erróneamente por perdido a AlphaZero hasta que la partida está muy avanzada y la victoria es inminente.

Maxime Vachier-Lagrave es el sexto jugador del mundo. Sus palabras son elocuentes: «Por supuesto, el resultado es realmente impresionante; yo no podría ni soñar con ganarle una partida a Stockfish. El resultado, especialmente con las blancas, es realmente impresionante. Si la gente tuviera acceso a AlphaZero en lugar de Stockfish o Houdini, la teoría cambiaría. En la mayoría de las partidas, AlphaZero estaba ganando y Stockfish estaba como: todo está bien, todo está bien, todo está bien durante unas 20 jugadas, eso es bastante antelación en cuanto a módulos, pero luego ya no estaba bien y poco después las cosas se iban de las manos. He visto cosas muy impresionantes».

Algunas de las pegas que se han puesto al experimento son importantes, pero apenas modifican las conclusiones finales. Se ha dicho que la versión de Stockfish no es la mejor, que no incluye aperturas o finales y que el sistema de gestión de tiempo es arbitrario (un minuto por jugada, lo que es muy extraño). También que el hardware sobre el que corría no es el mejor.

El experimento es un gran logro científico, pero el documento no sigue los estándares científicos. No es reproducible. Solo se muestran diez de las 100 partidas y ninguna de las partidas de entrenamiento en que AlphaZero jugó contra sí mismo. Pero, naturalmente, el orden de prioridades de Google, en sentido inverso, es: ajedrez, ciencia y cuenta de resultados. Google trabaja para los accionistas y publicará lo que le interese.

El conocimiento humano es prejuicio

AlphaZero empezó desde cero, se insiste. Haciendo movimientos aleatorios se entiende que los iniciales fueron muy malos. Poco a poco fue mejorando. En tan solo cuatro horas había alcanzado el nivel del mejor programa de ajedrez existente. Esto lleva a algunas conclusiones escalofriantes.

AlphaZero recorrió los 1.500 años de historia del ajedrez en cuatro horas. Toda la experiencia y saber humanos acumulados durante millones de horas se resumieron en solo 240 minutos. Lo que aprendimos desde que Ruy López publicó su libro hasta la última versión de Stockfish se vio superado en un instante.

Como es lógico, AlphaZero fue jugando algunas de las aperturas más conocidas. En el documento se analizan las aperturas que fue descubriendo. Hay miles de libros de aperturas. AlphaZero comienza jugando todas ellas pero las va descartando. ¿Significa eso que las aperturas consideradas buenas por los expertos no lo son? Así parece. AlphaZero descarta las populares siciliana, inglesa, española, Caro-Kann…

Al alienígena AlphaZero no le gustan esas cosas que hacen los humanos y sus máquinas. Más aún, en todas menos una de las partidas publicadas en las que juega con blancas empieza con el mismo movimiento d4.

Parece entonces que el principal problema del rival, Stockfish es que es demasiado humano. Los programadores han puesto en la máquina mucho conocimiento humano y esto es contraproducente. El conocimiento humano es un prejuicio. Es inútil, erróneo y un lastre. Las máquinas aprenden por sí solas y todas nuestras ayudas y conocimientos solo sirven para confundirlas.

Generalización

La ciencia ficción acostumbra a plantearnos escenarios en los que una inteligencia artificial toma el control de los humanos. Una característica de esa IA es que es general, una inteligencia artificial general (IAG) que resuelve cualquier tipo de problemas. La realidad es muy distinta. Hoy todo el software es de propósito específico: solo sirve para lo que se ha diseñado. Estamos muy lejos de una IAG. Sin embargo, el mismo algoritmo de AlphaZero es capaz de aprender a jugar a tres juegos distintos (como su predecesor, el de los videojuegos de Atari lo fue). Esto es una generalización muy importante ya que abaratará el desarrollo de software y acelerará el progreso. Pero el temor a una IAG que controle el mundo sigue siendo infundado.

Máquinas que aprenden solas y generalizan su comportamiento. Grandes noticias. El progreso se acelera.