

# Alpha Zero

## Alpha Zero

Autor 1: David Arcila Parra

Autor 2: Liz Dahianna Noreña Giraldo

Computación Blanda, Universidad Tecnológica de Pereira, Pereira, Colombia

Correo-e: [davidnon@utp.edu.co](mailto:davidnon@utp.edu.co).

[lizgiraldo@utp.edu.co](mailto:lizgiraldo@utp.edu.co)

**Resumen—** Hace veinte años muchos dudaban que una máquina fuese capaz de derrotar a los humanos en una partida de ajedrez. La victoria de Deep Blue, el ordenador creado por IBM, en 1997 sobre el campeón del mundo Gary Kasparov supuso un cambio de paradigma. Desde ese momento, la historia del deporte-ciencia cambió para siempre. Sin embargo, hace un tiempo llegó el holocausto absoluto. Un programa de inteligencia artificial destrozó de forma humillante al campeón de los ordenadores ajedrecistas, cambiando para siempre la disciplina y abriendo una nueva era a nivel científico.

A diferencia de otros programas, Alpha Zero, la inteligencia artificial creada por Deep Mind, propiedad de Google desde 2014, no está basado en el conocimiento humano. Su comprensión sobre el ajedrez, más allá de las reglas básicas, proviene únicamente de su capacidad de autoaprendizaje. Tras jugar casi cinco millones de partidas durante cuatro horas contra sí mismo, Alpha Zero obtuvo el mismo conocimiento que los humanos en casi 1.400 años.

**Palabras clave—** Algoritmo, Alpha Zero, Ajedrez, Autoaprendizaje, Conocimiento, Humano, Inteligencia artificial.

**Abstract—** Twenty years ago many doubted that a machine was capable of defeating humans in a game of chess. The victory of Deep Blue, the computer created by IBM, in 1997 over world champion Gary Kasparov was a paradigm shift. From that moment on, the history of sports science changed forever. However, a while ago the absolute holocaust arrived. A program of artificial intelligence humiliate the champion of chess computers, forever changing discipline and opening a new era at the scientific level.

Unlike other programs, Alpha Zero, the artificial intelligence created by Deep Mind, owned by Google since 2014, is not based on human knowledge. His understanding of chess, beyond the basic rules, comes only from his capacity for self-learning. After playing almost five million games for four hours against himself, Alpha Zero gained the same knowledge as humans in almost 1,400 years.

**Key Word —** Algorithm, Alpha Zero, Ajedrez, Learning, Knowledge, Human, Artificial intelligence.

## I. INTRODUCCIÓN

En muchos aspectos, los robots han demostrado ser "más duros, mejores, más rápidos, más fuertes". Ahora, su desafío es ser más elegantes, o por lo menos eso es lo que deja entrever la computadora AlphaZero, una de los desarrollos más recientes de DeepMind, la empresa de Google dedicada a la inteligencia artificial. [1]

Hace un año AlphaZero venció de manera contundente a otras máquinas entrenadas para jugar ajedrez; shogi, conocido como el ajedrez japones; y go, uno de los juegos de tablero más complejos que existen. [1]

Muchos teóricos se han encargado durante años de repetir que aunque las máquinas superen a los humanos, la creación de belleza nos pertenecía. Alpha Zero también ha puesto eso en duda. Y es que para muchos esta inteligencia artificial realiza un ajedrez único, que sacrifica piezas de forma humana, que se mantiene con desventaja material durante decenas de jugadas y que es capaz de especular como un humano en los finales de partida.[2]

El camino que Alan Turing, padre de la informática, comenzó en 1947 parece haber alcanzado un nuevo nivel. Los padres de esta monstruosa criatura aseguran que estamos ante “uno de los descubrimientos científicos de nuestra era porque permitirá entender los misterios del universo y abordar algunos de nuestros desafíos más apremiantes en el mundo real”. Dicen desde Deep Mind que, “de tener éxito, se avanzaría más rápido en la prevención del cambio climático o en la necesidad de una atención médica radicalmente mejorada. Con la inteligencia artificial como multiplicador del ingenio humano, esas soluciones estarán a nuestro alcance”, explican. El futuro parece que ya se ha instalado en nuestro tiempo. [2]

## II. CONTENIDO

### 1) Historia de la inteligencia artificial en los juegos de aprendizaje autónomo

La historia inicia con un ingeniero llamado Arthur Samuel, quien sería pionero en el desarrollo de la inteligencia artificial y especialmente en la creación de la “Machine Learning” o aprendizaje automático. [3]

Al final de 1948 Samuel leyó el famoso artículo de Shannon sobre cómo escribir un programa para jugar al ajedrez. Él decidió entonces escribir uno que jugara a las damas inglesas, que consideró sería más fácil de programar; nada más alejado de la verdad. Samuel dijo: “Pensé que sería un trabajo trivial programar una computadora para jugar a las damas inglesas”. [3]

Pero lo cierto es que tuvo que hacer todo desde cero ya que el artículo escrito por Shannon sólo prototipaba la idea, pero no daba escribió código alguno. Para fortuna de las generaciones que vendrían luego (nosotros entre ellos) Samuel se obsesionó con la idea y después de varios años de trabajo y ardua investigación pudo desarrollar el juego que tanto anhelaba [3]

En 1959 publica su primer artículo sobre aprendizaje de las computadoras bajo el título: “Some Studies of Machine Learning Using the Game of Checkers”, en donde presentó su programa usando la poda alpha-beta, la cual utilizan aún hoy varios programas para jugar Ajedrez. [3]

Por ese entonces Samuel tuvo la dicha de trabajar en IBM en la computadora 701 en la que pudo probar su programa. Colocaba 4 computadoras a jugar miles de partidas de damas entre sí, cuyos datos de los resultados de las partidas se iban acumulando, permitiéndole a las máquinas “Aprender de la experiencia” y en consecuencia mejorar su desempeño, con lo cual nació la “Machine Learning” o aprendizaje automático. [3]

### 2) Cómo funciona Alpha Zero

AlphaZero está conformado por tres etapas que corren en paralelo:

Self Play (Juego contra sí mismo): en esta etapa se crea un set de entrenamiento, en el caso de AlphaZero se sabe que jugó 4 de partidas contra sí mismo en 3 días. [3]

En cada movimiento, se almacena información conocida como “estado del juego”, lo que le permite al programa analizar las probabilidades de búsqueda a través de su MCTS (Búsqueda Montecarlo) [3]

#### Red de readiestramiento:

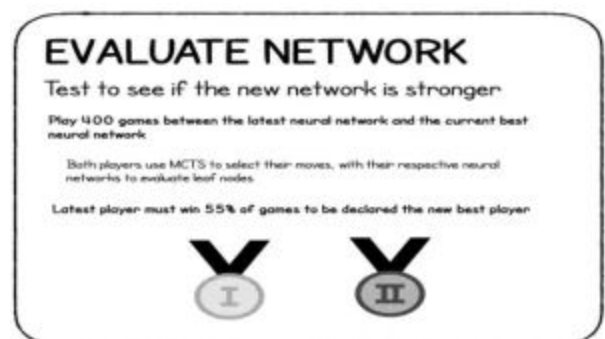
Esta red utiliza como datos de entrada los estados de juego mencionados anteriormente y realiza un bucle reentrenamiento con un lote de las mejores posiciones extraídas de las últimas 500.000 partidas. [3]

#### Optimización de los pesos de la red

Esta etapa además contiene una Función de pérdida, que compara las predicciones de la red neuronal con las probabilidades de búsqueda para elegir las jugadas que considera ganadoras. [3]

Después de cada 1,000 bucles de entrenamiento, la red se autoevalúa para predecir mejor según el juego de su rival. [3]

Evaluación de la red: tal y como lo muestra claramente la imagen, esta etapa hace partidas internas con sus redes sobre las variantes que se estén presentando en el tablero para deducir las mejores jugadas. [3]



Prueba para ver si la nueva red es más fuerte.

Juega 400 partidas entre la última red neuronal y la mejor red neuronal actual

Ambas redes jugadoras usan MCTS (Búsqueda Montecarlo) para seleccionar sus movimientos, con sus respectivas redes neuronales para evaluar los nodos de las hojas. El último jugador debe ganar el 55% de las partidas para ser declarado el nuevo mejor jugador entre estas dos redes. [3]

Se sabe entonces que el algoritmo AlphaZero funciona en una única computadora dotada con cuatro Unidades de Procesamiento de Tensor (TPU, en inglés), circuitos integrados desarrollados específicamente para el aprendizaje de máquinas. De hecho Google ha liberado su biblioteca "TensorFlow" la cual ha sido su principal biblioteca para el trabajo de Aprendizaje Automatizado y redes neurales en Inteligencia Artificial, (pronto haré un vídeo donde muestro como aprovechar esta biblioteca mediante un programita que use IA). [3]

### 3) El árbol de análisis

Los módulos de ajedrez usan una estructura de árbol para calcular variantes y usan una función de evaluación para asignar a la posición al final de una variante un valor como +1,5 (la ventaja de las blancas equivale a un peón y medio) o -9,0 (la ventaja de las negras equivale a una dama). El enfoque de AlphaZero con respecto del cálculo de variantes y de la evaluación de posiciones es radicalmente distinto a lo que hacen otros módulos. [4]

Todos los módulos de ajedrez populares se basan en el algoritmo minimax, que es un nombre sofisticado que simplemente significa que escoges el movimiento que te da la mayor ventaja independientemente de lo que juegue tu oponente. El minimax se mejora invariablemente con una poda alfa-beta, que se usa para reducir el tamaño del árbol de variantes a examinar. Aquí tienes un ejemplo exagerado de cómo funciona esta poda: Digamos que un módulo está considerando un movimiento y ve que su oponente tiene veinte respuestas factibles. Una de esas respuestas lleva a un jaque mate forzado. Entonces el módulo puede abandonar (o "amputar") el movimiento que estaba considerando,

independientemente de lo bien que funcionaría contra cualquiera del resto de 19 respuestas.[4]

Otro aspecto es que si un módulo poda movimientos que solo parecen malos, por ejemplo aquellos que pierden material, no logrará tener en cuenta ningún tipo de sacrificio, que es en parte el motivo de que los primeros módulos fueran tan materialistas. En módulos actuales como Stockfish, la poda alfa-beta está combinada con una serie de otras mejoras específicas del ajedrez como el tanteo de movimientos mortales (un movimiento fuerte en otra variante similar es probable que sea fuerte aquí), el tanteo de contramovimientos (algunos movimientos tienen respuestas naturales independientemente de la posición: seguro que a menudo has respondido a axb5 con axb5, ¿no?) y muchos otros. [4]

AlphaZero, en cambio, usa el Árbol de Búsqueda Monte Carlo (MCTS, por sus siglas en inglés). Monte Carlo es famoso por sus casinos, así que, cuando veas este término en un contexto informático, significa que sucede algo arbitrario. Un módulo que use la MCTS para evaluaría una posición generando una serie de secuencias de movimientos (o "desarrollos") a partir de esa posición de manera arbitraria y haciendo una media de los resultados finales (victoria/tablas/derrota) que genera. Este enfoque puede parecer demasiado simple, pero si lo piensas te darás cuenta de que es una forma bastante plausible de evaluar una posición. [4]

AlphaZero crea 800 desarrollos en cada movimiento. También aumenta la MCTS pura prefiriendo movimientos que no ha probado (mucho) aún, que parecen probables y que parecen llevar a "buenas" posiciones, es decir, donde la función de evaluación (hablaremos más de ella en el siguiente artículo) les da un valor alto. Realmente crea desarrollos semi arbitrarios, líneas que parecen apropiadas a su función de evaluación en mejora constante. ¿No es bastante parecido a cómo calculas tú? ¿Centrándote en líneas de juego plausibles?[4]

Observa que hasta ahora no hay absolutamente nada específico del ajedrez en lo que hace AlphaZero. En mi próximo artículo, cuando veamos cómo aprende AlphaZero a evaluar posiciones de ajedrez, veremos que tampoco tiene nada específico de ajedrez. [4]

AlphaZero, igual que un bebé recién nacido, vino al mundo con poco conocimiento, pero está tremendamente equipado para aprender. Una debilidad de la MCTS es que, dado que se basa en crear desarrollos semi arbitrarios, puede equivocarse totalmente en posiciones tensas en las que hay una línea precisa de juego óptimo. Si no selecciona esta línea de forma arbitraria, es muy probable que cometa un error grave. Esta ceguera seguramente es la que hizo que el predecesor de

AlphaZero que jugaba a Go, AlphaGo, perdiera una partida contra el 18 veces campeón mundial de Go, Lee Sedol. Sin embargo, no parece que esto fuera un problema en el match con Stockfish.[4]

La MCTS se ha utilizado previamente para juegos de dos jugadores, pero había dado un rendimiento mucho peor que el ya establecido enfoque de minimax con alfa-beta. En AlphaZero, la MCTS combina muy bien con la función de evaluación basada en una red neural. Explicaré más sobre esta red neural y sobre el fascinante modo en el que aprende, de forma autónoma, a evaluar posiciones de ajedrez. También describiré el hardware en el que opera AlphaZero y haré algunas predicciones sobre el impacto que tendrá todo esto en el ajedrez tal y como lo conocemos. [4]

[3] Historia de la inteligencia artificial, Disponible en: <https://thezugzwangblog.com/desvelamos-los-secretos-del-juego-alpha-zero/>

[4] ¿Cómo juega Alpha Zero?, Disponible en: <https://www.chess.com/es/article/view/como-juega-alpha-zero-al-ajedrez>

### III CONCLUSIONES

- Los grandes maestros tendrán que estudiar las partidas de AlphaZero y aprender de su sorprendente intuición ajedrecística.
- El algoritmo de Montecarlo tipo alfa-beta de Stockfish analiza unos 70 millones de posiciones por segundo, mientras que el de AlphaZero solo analiza 80 mil posiciones. Para compensar este déficit usa su red de neuronas profunda (multicapa) para seleccionar las posiciones más prometedoras (lo que alguna gente ha bautizado como su “intuición tipo humana”).
- AlphaZero es la primera máquina de ajedrez que no ha aprendido del hombre y que, por tanto, carece de prejuicios inculcados. Quién sabe si desarrollará los suyos propios. Es un robot (casi) sin dios, que aprende de sí mismo y de la observación, de su propia e insustituible experiencia. Nadie le ha dicho que la Tierra es plana y, sin ideas preconcebidas, descubre más pronto que nosotros que es redonda.

### REFERENCIAS

[1] "AlphaZero nos supera de manera profunda": la computadora que genera su propio conocimiento y juega como un "superhumano", Disponible en: <https://www.bbc.com/mundo/noticias-46705149>

[2] Alpha Zero, el programa que revoluciona el ajedrez y puede cambiar el mundo, Disponible en: <https://www.lavanguardia.com/deportes/otros-deportes/20171214/433624379301/alpha-zero-deep-mind-gary-kasparov-ajedrez-inteligencia-artificial.html>