



**UNIVERSIDAD
DE GRANADA**

**Sistemas Inteligentes para la Gestión de la
Empresa**

**Trabajo sobre el Algoritmo AlphaGo sin
presentación**

Curso 2019-2020

Máster en Ingeniería Informática

Lidia Sánchez Mérida

Introducción

AlphaGo es un algoritmo perteneciente al campo de la Inteligencia Artificial que ha sido desarrollado por *Google DeepMind* cuya finalidad consiste en jugar a un juego de mesa denominado *Go* [1]. El objetivo de este juego es conseguir dominar una mayor cantidad de territorio que tu oponente [2]. Para ello han elaborado y entrenado un modelo predictivo basado en redes neuronales convolucionales dentro del ámbito del *Deep Learning* con hasta 13 capas para obtener características que le ayude a encontrar el conjunto de movimientos ganador [4].

El primer objetivo de este modelo consiste en decidir el siguiente movimiento dado un determinado estado [3]. Con el objetivo de tomar esta decisión incluyeron un segundo algoritmo que utiliza las denominadas estructuras en árbol para representar las distintas posibilidades existentes en una partida de un determinado juego. Cada nodo representa el estado del juego en un momento dado y su objetivo es determinar el movimiento más prometedor, es decir, aquel por el cual se asegure, con una cierta probabilidad, la victoria. Si bien existen una gran variedad de algoritmos pertenecientes a esta familia, el que se ha integrado en el AlphaGo es el denominado **Monte Carlo Tree Search**. La principal razón de esta elección reside en la gran complejidad que entraña analizar todos los posibles movimientos, dado un estado, para escoger el mejor camino hacia la victoria. Mientras que otros algoritmos como el *Minimax* se detienen a estudiar todos los posibles movimientos de un estado, el algoritmo *Monte Carlo Tree Search* implementa un sistema por el cual prioriza ciertas ramas del árbol para intentar ser lo más eficiente posible [1]. Para ello realiza varias simulaciones de la partida para un estado determinado, maximizando una función concreta que calcula la probabilidad de ganar la partida para un conjunto de movimientos. Aquellos caminos con un mayor valor serán los que primero explore el algoritmo puesto que parecen ser los más prometedores [5].

Para el entrenamiento del modelo, en primer lugar, comenzaron enfrentado el algoritmo contra una gran variedad de jugadores. El objetivo consistía en recopilar información acerca de las tácticas de cada uno para que el algoritmo fuese capaz de aprender las estrategias que utilizaban los jugadores. Técnicamente el algoritmo analizaba hasta 30 millones de posiciones diferentes del tablero para encontrar aquella que le acercase a la victoria. Con el objetivo de ajustar sus parámetros se aplicaba el método conocido como **backpropagation**, por el cual propagaba el error desde la capa de salida hacia capas anteriores con el objetivo de obtener mejores valores que le permitiese mejorar su precisión. Si bien se invirtieron hasta tres semanas en esta primera fase del entrenamiento su tasa de acierto fue del 57%. Con el fin de reducir el tiempo de entrenamiento emplearon un clasificador *softmax* más sencillo que el que habían utilizado en un principio y como consecuencia los tiempos disminuyeron considerablemente pero también su tasa de acierto [4].

Para mejorar su precisión comenzaron una nueva fase de entrenamiento en la que utilizaron una técnica conocida en el ámbito del *Machine Learning* como **aprendizaje por refuerzo** [1][3][4], por la cual incluyeron un sistema de recompensas para alentar al algoritmo a escoger los movimientos más prometedores. En este caso sus contrincantes eran las diferentes versiones de sí mismo, que se iban generando con cada entrenamiento que se realizaba, con el objetivo de que pudiese aprender de sus errores [3][4].

Campeonatos

AlphaGo participó en su primera competición en 2015 contra el tricampeón europeo Mr Fan Hui, al que venció en todas las partidas que se disputaron. Posteriormente se enfrentó al 18 veces campeón mundial de este juego particular y de nuevo el algoritmo fue capaz de alcanzar la victoria en la gran mayoría de partidas. Con el paso del tiempo fue acumulando un mayor número de victorias que le llevaron a alcanzar el primer galardón otorgado a un algoritmo tanto por las partidas ganadas como por los movimientos realizados. A partir de este momento se desarrollaron otra serie de versiones del algoritmo como *Master*, la cual se encuentra disponible online además de varios formatos específicamente creados para la convención *The Future of Go Summit*, en la cual también venció a otro oponente, en este caso se trataba de Ke Jie, el mejor jugador del mundo [3].

Variantes

Tras el éxito de este algoritmo se ha continuado desarrollando diversas variantes del mismo con el objetivo de mejorar su habilidad tanto en el particular juego *Go* como en otros similares, como el ajedrez. Uno de los casos más destacados es el de **Alpha Go Zero**, el cual se diferencia del original en los siguientes aspectos principales.

- En primer lugar este algoritmo fue exclusivamente entrenado a partir de un gran número de simulaciones en las que el algoritmo luchaba contra sí mismo. Al comienzo no tenía ningún tipos de conocimientos, sin embargo conforme jugaba más partidas iba combinando los modelos entrenados resultantes para combinar el conocimiento adquirido en cada partida y mejorar de cara a las siguientes.
- La segunda diferencia con respecto al original es la reducción del número de *inputs* que consideraba el modelo. En el caso de esta nueva versión se reducían únicamente a las posiciones de las piedras blancas y negras, lo cual aumentaba la eficacia además de la eficiencia.
- Como tercera discrepancia se encuentra la fusión de las dos redes neuronales capaces de escoger el siguiente movimiento y predecir el ganador dado el estado resultante, respectivamente. De este modo se mejora tanto el entrenamiento como la validación del modelo predictivo [6].
- Por último, desplaza al Monte Carlo Tree Search por un nuevo algoritmo de entrenamiento que le permite a la propia red neuronal realizar sus propias evaluaciones acerca de los movimientos más prometedores dado el estado actual de la partida [7].

Este algoritmo, además de vencer con bastante ventaja a las primeras versiones del AlphaGo también fue capaz de incorporar nuevo conocimiento que provocó el desarrollo de nuevas estrategias con movimientos innovadores y difícilmente interpretables por sus contrincantes [6].

Uno de los siguientes pasos que dió la empresa fue trasladar la idea de su algoritmo a un videojuego más complejo, como es el *StarCraft II*. Se trata de un juego de estrategia en tiempo real basado en la ciencia ficción que se caracteriza por la existencia de múltiples estrategias, la inclusión de incertidumbre y su complejidad pues no en todos los movimientos se puede obtener toda la información necesaria para elegir el siguiente. Para participar en las competiciones asociadas a este videojuego implementaron una variante denominada **AlphaStar**. De nuevo su modelo se encuentra basado en redes neuronales profundas que han sido entrenadas aplicando aprendizaje supervisado y por refuerzo, característica que comparte con las versiones originales. Como entradas recibe la información asociada a la interfaz gráfica del jugador y como salida devuelve las mejores acciones posibles a realizar.

En relación a su entrenamiento, se comenzó una primera fase en la que se entrenó el modelo a través de aprendizaje supervisado por una persona experimentada en el juego. De este modo el algoritmo aprendía las estrategias de su maestro. En una segunda fase se organizó un campeonato de algoritmos en el cual luchaban entre sí para continuar con sus entrenamientos. Sus estrategias se basaban en el aprendizaje por refuerzo, tal y como ocurría con las primeras versiones del AlphaGo. Conforme aumentaba el número de partidas y de contrincantes, el *AlphaStar* fue capaz de desarrollar nuevas estrategias que le permitían hacerse con la victoria contra prácticamente cualquier participante. De este modo se consiguió un modelo predictivo capaz de ganar a uno de los mejores equipos profesionales en *StarCraft II*.

Si bien el desarrollo de este tipo de algoritmos se orientó hacia los juegos de mesa con el paso del tiempo se han ido aplicando a otras disciplinas como la científica. En este caso se trata de un modelo denominado **AlphaFold** cuya principal tarea consiste en predecir el modelo en 3D de una proteína en función de su secuencia genética. El hecho de intentar representar una proteína gráficamente es sumamente costoso puesto que la gran mayoría de ellas se encuentran formadas por una cantidad masiva de cadenas con estructuras irregulares. Uno de los principales aspectos por lo que es relevante la forma de las proteínas es porque generalmente suelen explicar el papel que tienen en un ser vivo. Por lo tanto, la incorporación de este algoritmo al ámbito científico permite invertir los costes temporales y monetarios que se gastaban en modelar las proteínas en la capacidad de predecir su influencia en relación a enfermedades con el objetivo de hallar mejores curas [9].

Bibliografía

1. Analytics Vidhya, Ankit Choudhary, *Introduction to Monte Carlo Tree Search: The Game-Changing Algorithm behind DeepMind's AlphaGo*, 2019, <https://www.analyticsvidhya.com/blog/2019/01/monte-carlo-tree-search-introduction-algorithm-deepmind-alphago/>
2. American Go Association, *What is Go?*, <https://www.usgo.org/what-go>
3. DeepMind, *AlphaGo*, <https://deepmind.com/research/case-studies/alphago-the-story-so-far>
4. Medium, Jonathan Hui, *AlphaGo: How it works technically?*, 2018, https://medium.com/@jonathan_hui/alphago-how-it-works-technically-26ddcc085319
5. Medium, Kung-Hsiang, Huang (Steeve), *Introduction to Various Reinforcement Learning Algorithms. Part I (Q-Learning, SARSA, DQN, DDPG)*, 2018, <https://towardsdatascience.com/introduction-to-various-reinforcement-learning-algorithms-i-q-learning-sarsa-dqn-ddpg-72a5e0cb6287>
6. DeepMind, David Silver, Demis Hassabis, *AlphaGo Zero: Starting from scratch*, 2017, <https://deepmind.com/blog/article/alphago-zero-starting-scratch>
7. Silver, D., Schrittwieser, J., Simonyan, K., *Mastering the game of Go without human knowledge*, Nature, 2017, <https://www.nature.com/articles/nature24270>
8. DeepMind, The AlphaStar team, *AlphaStar: Mastering the Real-Time Strategy Game StarCraft II*, 2019, <https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii>
9. DeepMind, Andrew Senior, John Jumper y Demis Hassabis, *AlphaFold: Using AI for scientific discovery*, 2018, <https://deepmind.com/blog/article/alphafold-casp13>