

Reinforcement Learning: FlappyBird

Kenneth Obando Rodríguez, *Instituto Tecnológico de Costa Rica*
Alejandro Pacheco Quesada, *Instituto Tecnológico de Costa Rica*

Resumen—

Index Terms—neural network, gradient descent, softmax, cross-entropy, relu, mnist, machine learning, image classification

I. INTRODUCCIÓN

EXISTEN problemas computacionales que requieren que un programa realice una acción ante un estado determinado, con la agravante que estos estados pueden tener muchas variables diferentes que lo convierten en una tarea prácticamente imposible para una aproximación estática de la solución. Ante estos problemas, surge toda una rama del campo de la Inteligencia Artificial que se llama Aprendizaje Por Reforzamiento (“*Reinforcement Learning*”), que consiste, en términos generales, en entrenar un algoritmo utilizando premios (“*reward*”) positivos o negativos, según sea el resultado de la acción tomada en un estado dado.

Esta investigación tiene como objetivo implementar un algoritmo de *Reinforcement Learning* conocido como “Policy Gradient” para entrenar un modelo computacional para que juegue con éxito “FlappyBird”, para ello se utilizan herramientas como PyTorch,

II. TRABAJO RELACIONADO

Un dato interesante es que el famoso visionario Turing (1948,1950) propuso una aproximación del aprendizaje por reforzamiento aunque no estaba totalmente convencido de su efectividad: “... *el uso de castigos y premios puede, en el mejor de los casos, ser parte de un proceso de enseñanza.*” [1], [2]. Se sospecha que la primera investigación exitosa en este campo es la de Arthur Samuel, el cual contiene una buena parte de las ideas modernas en aprendizaje por reforzamiento, incluyendo la aproximación utilizando una función [3].

Desde principios de la década de los 90 se realizaron aportes en el campo del aprendizaje por reforzamiento utilizando “*Policy Gradient*” inicialmente gracias a Ronald Williams que desarrolló esta familia de algoritmos. Trabajos posteriores hicieron grandes aportes, aunque fue con Papavassiliou y Russel que se describe un nuevo tipo de reforzamiento que converge con cualquier tipo de aproximador de función [4].

[5]

III. METODOLOGÍA

[6]

III-A. Entrenamiento y testing

III-B. Red neuronal

REFERENCIAS

- [1] A. Turing, “Intelligent machinery,” *National Physical Laboratory*, 1948.
- [2] —, “Computing machinery and intelligence,” *Mind*, vol. 59, pp. 433–460, 1950.
- [3] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, ser. Prentice Hall series in artificial intelligence. Prentice Hall, 2010. [Online]. Available: <https://books.google.co.cr/books?id=8jZBksh-bUMC>
- [4] V. A. Papavassiliou and S. Russell, “Convergence of reinforcement learning with general function approximators,” in *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2, 1999, pp. 748–755. [Online]. Available: <https://www.ijcai.org/Proceedings/99-2/Papers/014.pdf>
- [5] J. Peters, “Policy gradient methods,” *Scholarpedia*, vol. 5, no. 11, p. 3698, nov 2010. [Online]. Available: http://www.scholarpedia.org/article/Policy_{_}gradient_{_}methods
- [6] B. Siegmann and T. Jarmer, “Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data,” *International Journal of Remote Sensing*, vol. 36, no. 18, pp. 4519–4534, 2015. [Online]. Available: <http://dx.doi.org/10.1080/01431161.2015.1084438>

ANEXOS